Aspects of the ergodic, qualitative and statistical theory of motion

G. Gallavotti, F. Bonetto and G. Gentile

Roma 2003, version 3.1

20/*novembre*/2011; 22:18

Giovanni Gallavotti
Department of Physics and I.N.F.N.
I^{th} University of Roma, La Sapienza
P.le A. Moro 2,
00185 Roma, Italia

Federico Bonetto
Department of Mathematics
Georgia Institute of Technology
Atlanta, Georgia 30322, USA

Guido Gentile
Department of Mathematics
Università di Roma Tre
L.go S. Leonardo Murialdo,
00146 Roma, Italia

Preface (2003)

This book started as a translation of an earlier Italian book by one of us. The present book is not only a substantial revision of the earlier book but it also deals with a number of problems that were not treated there. The main novelty is the systematic treatment of a few characteristic problems of ergodic theory by a unified method in terms of convergent power series expansions. The methods of resummation necessary to deal with such series have become familiar as "renormalization group methods" and in our opinion provide the simplest and most intuitive approach to the problems that we discuss (like KAM theory and Anosov maps).

A substantial number of questions, some even more interesting than the ones in the main body of the book, are treated in the form of guided problems. This is not done to save space (in spite of the importance of such saving) but mainly because we feel that it constitutes, for the interested reader, a more stimulating way of presenting the matter.

We are grateful to Dr. Alessandro Giuliani for his many critical comments and suggestions. And we warmly thank Professor Wolf Beiglböck for his continuous support and encouragement.

version 3.1.

Roma 18 july 2003.

Giovanni Gallavotti
Federico Bonetto
Guido Gentile

Preface to the Italian book (1980)

In recent years books on problems of ergodic theory appeared with the common theme of stressing the close relations observed between it and the theory of Gibbs states.

In these lectures I collect the contents of several courses and seminars that I presented in various occasions (mainly Courant Institute, (1972), EPF of Lausanne (1979), Scuola di perfezionamento in Fisica di Roma (1980)).

The organization and the choice of the arguments has been largely influenced by the monographs of R. Bowen and D. Ruelle: this book mainly deals with the ergodic theory of maps (one dimensional ergodic theory). My intention has been to exhibit the existence of a tread connecting apparently different problems by always selecting, for illustration purposes, simple examples and by often avoiding formulating results in the widest generality: this collection of lectures is intended for beginners in ergodic theory and it follows a particular point of view: that of keeping always contact with statistical mechanics (from which the problems originated but which no longer appears clearly as the source of the main problems, ideas and conjectures).

The problems at the end of every section are an essential complement to the text and they are not always obvious or of easy solution. In using this book as a textbook for a course it would be useful to solve many of the problems leaving to the students the task of studying the main part of the textbook: it is only through the problems that the student can reach and go beyond the formal (hence superficial) level of understanding and to master the technical difficulties and to reach at the roots of the ideas. Almost all problems propose a proof of a classical result that, although I consider it not very useful to describe explicitly in the text, nevertheless cannot be neglected of skipped if one is attempting to study the subject of ergodic theory.

At the same time the problems provide a guide to the systematic study of related results or of results traditionally pertaining to other disciplines; the guide is meant to put within the frame of ergodic theory a certain number of them and to show their connection with some basic achievements in ergodic theory. Such a guide is necessary because the extraordinary variety of techniques and methods that constitutes one of the reasons of the fascination that ergodic theory exerts on mathematicians becomes also the main difficulty for a beginner.

The lectures on Gibbs states concern *only* one–dimensional systems: nevertheless a good part of the results extends with obvious modifications to systems in higher dimensions (which are more directly related to the statistical mechanics) and this holds in particular for the §(6.1)§(6.2) and to a fair extent for §(5.1), §(5.2), §(6.1), §(6.4), §(6.4). Therefore I think that what is discussed in this book could also serve as an introduction to the theory of Gibbs states and of stochastic processes in more dimensions.

In the references the publication years refer to the quoted edition rather than to the original edition.

While writing up these lectures I benefitted from several discussions with students and collaborators or colleagues whom I wish to thank because without their help and encouragement this work would have been impossible. In particular I thank G. Benfatto, M. Campanino, F. Ledrappier, G. Pianigiani.

I am also indebted to several institutions for invitations to give courses and seminars providing me the chance and the means to learn and to organize the topics treated here: among them in particular Institut Hautes Etudes Scientifiques (IHES), Scuola Matematica Internazionale (SMI), Istituto Nazionale di Alta Matematica (INDAM), Scuola di Perfezionamento in Fisica di Roma.

Finally the encouragement to start writing these notes from Unione Matematica Italiana and Carlo Pucci has been essential.

Roma 15 giugno 1980 Giovanni Gallavotti

Table of contents

CHAPTER I

General qualitative properties

§1.1 Historical note

We begin with a few comments on the meaning of the word *ergodic*.

At the beginnings of Statistical Mechanics a "collection of time evolution invariant probability distributions on phase space" or a "collection of *stationary distributions* on phase space" for a given Hamiltonian system, was called by Boltzmann, see [Bo09] and pag. 85 of [Bo96], a *monode*: the modern abbreviated locution describing this notion is a "statistical ensemble" (or simply an "element of a stationary ensemble").

The word *monode*, one among many coined by Boltzmann who clearly loved to invent this kind of words, is of greek origin: it is composed by μόνος , unique, and by εἶδος , aspect, appearance. Probably because this suggests the image of a collection of copies of the considered system which keeps, as a collection, its appearance as time evolves: each copy being subjected to a motion whose only global effect is a permutation of the various copies.

Boltzmann, then, called *ergode* a monode characterized by a uniform distribution on a surface of constant energy: a monode is *ergodic* if it is an ergode.

Hence it looks difficult or, better, impossible to attribute a literal meaning to *ergodic theory*, a locution that has not been used by Boltzmann. The origin of the word seems to go back to the Ehrenfests who, in the their important work of interpretation and popularization of Boltzmann's ideas, called "ergodic" any mechanical systems whose surfaces of constant energy consist of a single trajectory (which had been named, instead, *isodic* by

Boltzmann, from ἴσος , same, ὁδός , road, path), see footnote 93 in [EE59]. They called *quasi-ergodic* any mechanical system whose trajectories invaded densely the surface Σ_E of constant energy E, see footnote 98/99 in [EE59], and concluded with words of discomfort and doubt on the actual existence of systems endowed with the ergodic property, see pag. 25, line 15, of [EE59]. In this respect see also the footnote 97 in [EE59], where one cannot avoid being surprised by the depth of Boltzmann's intuitions. The meaning of these intuition, we feel, still escaped the Ehrenfests in 1912. Indeed Boltzmann chooses, in quoting an example of an ergodic system, a system which is still today considered a possible example. The work of the Ehrenfests generated renewed efforts, by many mathematicians, to interpret formally Boltzmann's ideas. Thus the true and proper ergodic theory began and the first fundamental result has been the classical formulation by Birkhoff of the *ergodic hypothesis*: a precise mathematical translation of a nice formula by Boltzmann, see formula (34) at pag. 25 of [EE59].

Such a hypothesis also called *metric transitivity hypothesis*, supposes that for "interesting" mechanical systems the trajectories of almost all points of the surface Σ_E invade it densely and, furthermore, spend in each of its parts $\Lambda \subset \Sigma_E$ a fraction of time proportional to its Liouville measure:

e1.1.1
$$\frac{\int_\Lambda \delta(H(\underline{p},\underline{q})-E)d\underline{p}d\underline{q}}{\int_{\Sigma_E} \delta(H(\underline{p},\underline{q})-E)d\underline{p}d\underline{q}} \tag{1.1.1}$$

having denoted with $\underline{p}$, $\underline{q}$ the canonical coordinates of the system and with $H(\underline{p},\underline{q})$ its Hamiltonian function.

Ergodic theory thus acquired the precise meaning of "theory of the ergodic hypothesis": *i.e.* it became the complex of mathematical propositions connected, or held as connected, with attempts of showing the validity (or the falsity) of the hypothesis in the case of various mechanical systems.

As a curiosity it can be interesting to note that, oddly, the Ehrenfests derive the ethimology of the word *ergode* from ἔργον , energy, and ὁδός , road, rather than from ἔργον and εἶδος , as it seems to be beyond doubt: by reading Boltzmann the word *ergode* comes out as a natural abbreviation of the word (more clearly explicative of the concept but lengthier hence, perhaps, less satisfactory) *ergomonode*, see footnote 93 in [EE59].

Mathematicians and physicists made efforts (with increasing vigor, particularly after the 1950's) directed to obtain a proof of the validity of the ergodic hypothesis in particular mechanical systems: although the efforts did not lead to a solution of the original problem, at least not in a generality relevant for applications, they led to extensions of the problem and to a vast class of mathematical results that encompass fields which at a first sight are rather inhomogeneous (like number theory or information theory). It is to this set of results, a mixture of results on diverse fields and in continuous expansion, that we refer today when we talk of ergodic theory which is, therefore, a name characterizing rather unprecisely each of its arguments.

In this book we shall select and illustrate several different problems that represent aspects of ergodic theory, trying to stress the common features that justify their classification as elements of a single mosaic.

Note to §1.1

(1) The use of the *axiom of choice* seems to have influenced in an essential way the interval of time that has been necessary for the correct mathematical formulation of the metric transitivity hypothesis: this appears quite clearly from the footnote 87 of the quoted book of the Ehrenfests, [EE59], that seems to consider as an insuperable obstacle to the otherwise natural formulation of the hypothesis a (probably unconscious) application of the axiom.

Indeed the $\sigma_0(p,q)$ of the quoted footnote would be, in a metrically transitive system, a nonmeasurable function whenever really varying from a G–path to another and it would not be possible to construct it without using the axiom of choice, that is therefore implicitly used, see problem [2.2.54] for an example of the construction of such a function along the argument of [EE59].

This instance is likely to be the only one in which the axiom of choice has exerted its sinister influence over a fundamental question of physical and applicative interest.

(2) A detailed critical analysis, and an exegesis in contemporary language, of Boltzmann's work would be very interesting. Until now such an enterprise has not been really undertaken, obviously because of the prohibitive amount of work that it implies. Nevertheless the literature on Boltzmann is large and rich of ideas and proposals for a deeper understanding, as the fascination of his personality allows anyone (even a profane) to predict, see [Ce99]. A first bibliography of studies on Boltzmann is given here although it is possibly seriously incomplete.

Bibliographical note to §1.1

On the foundations of statistical mechanics many of the original works are interesting. Among these we mention those of Boltzmann, Gibbs, Maxwell; see [Bo09], [Bo02], [Bo03],[Gi60], [Ma65]

Among the works of critique on the foundations see for instance Ehrenfest and Ehrenfest, [EE59], and Krylov, [Kr79]. More recent discussions can be found in [Br99], [Ga00].

The life and scientific achievements of Boltzmann are molded together in the books [Ce99], [Li01]: the first being more informative on the scientific aspects of Boltzmann's figure while the second gives a very clear picture of the aspects of his personality that made his life look to himself rather unhappy.

§1.2 Examples and some definitions

We begin our analysis by illustrating simple examples of "*dynamical systems*": dynamical systems, indeed, constitute the fundamental mathematical entity of ergodic theory, [AA68].

E1.2.1 *Example (1.2.1)* : Let $\underline{f} \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ be a function on $\mathbb{R}^d$ with values in $\mathbb{R}^d$ that we shall suppose to be uniformly bounded together with its first derivatives. We denote by $t \to \underline{x}(t) \equiv S_t(\underline{\xi})$ the solution of the differential equation

e1.2.1
$$\dot{\underline{x}} = \underline{f}(\underline{x}) \qquad \underline{x}(0) = \underline{\xi}\ , \qquad \underline{\xi} \in \mathbb{R}^d. \tag{1.2.1}$$

Then $(S_t)_{t\in\mathbb{R}}$ is a group of maps of class C^∞ of $\mathbb{R}^d$ into itself and $(\mathbb{R}^d, (S_t)_{t\in\mathbb{R}})$ is a dynamical system called *the flow on* $\mathbb{R}^d$ *generated by the differential equation (1.2.1)* or *by the vector field* $\underline{f}$.

Often $\underline{f}$ has other remarkable properties. For example sometimes $\underline{f}$ has *zero divergence*:

e1.2.2
$$\sum_{i=1}^d \frac{\partial f_i}{\partial \xi_i}(\underline{\xi}) = 0 \qquad \text{for all } \underline{\xi} \in \mathbb{R}^d, \tag{1.2.2}$$

if $\underline{f} = (f_1, \dots, f_d)$. If λ denotes the Lebesgue measure on $\mathbb{R}^d$ one has, in the latter case,

e1.2.3
$$\lambda(E) = \lambda(S_t E) \qquad \text{for all } t \in \mathbb{R}, \tag{1.2.3}$$

for every λ–measurable set E: *i.e.* the map S_t *preserves the Lebesgue measure.*

Sometimes $\underline{f}$ generates a group that leaves a closed set Ω invariant. For instance this happens if Ω is a sphere or a torus or, more generally, a domain with C^∞–regular boundary $\partial\Omega$ and, denoting by $\underline{n}(\underline{\xi})$ the external normal to $\partial\Omega$ in $\underline{\xi}$,

e1.2.4
$$\underline{f}(\underline{\xi}) \cdot \underline{n}(\underline{\xi}) \le 0 \qquad \text{for all } \underline{\xi} \in \partial\Omega, \tag{1.2.4}$$

holds everywhere on $\partial\Omega$. In such a case $S_t(\Omega) \subset \Omega$, for all $t > 0$, and one can consider the *flow generated, for* $t \ge 0$*, by equation (1.2.1) in* Ω, that will be denoted $(\Omega, (S_t)_{t\in\mathbb{R}_+})$.

Invariant sets can also be constructed from *first integrals*, also called *constants of motion*, of equation (1.2.1), when they exist. Indeed if $F \in C^\infty(\mathbb{R}^d, \mathbb{R})$ is a first integral for equation (1.2.1), *i.e.* if $F(S_t\underline{\xi}) \equiv F(\underline{\xi})$, for all $\underline{\xi} \in \mathbb{R}^d$ and for all $t \in \mathbb{R}$, given $\gamma \in \mathbb{R}$ the sets

e1.2.5
$$\Omega^<_\gamma = \{\underline{\xi} | F(\underline{\xi}) < \gamma\},\ \Omega^>_\gamma = \{\underline{\xi} | F(\underline{\xi}) > \gamma\},\ \Omega_\gamma = \{\underline{\xi} | F(\underline{\xi}) = \gamma\} \tag{1.2.5}$$

are obviously invariant under the flow generated by equation (1.2.1).

From example (1.2.1) that we have just discussed one derives as particular cases, or as generalizations, the examples below (1.2.2) (Hamiltonian flows) or (1.2.3) (flows on differentiable manifolds), respectively.

E1.2.2 *Example (1.2.2)* : (*Hamiltonian flows*) Let $H \in C^\infty(\mathbb{R}^{2N}, \mathbb{R})$ and $E \in \mathbb{R}$ be such that

(i) $\Omega_E^{\leq} = \{(\underline{p}, \underline{q}) | (\underline{p}, \underline{q}) \in \mathbb{R}^{2N},\ H(\underline{p}, \underline{q}) \leq E\}$ is a bounded set.
(ii) $\Omega_E = \{(\underline{p}, \underline{q}) | (\underline{p}, \underline{q}) \in \mathbb{R}^{2N},\ H(\underline{p}, \underline{q}) = E\}$ is a bounded smooth surface of dimension $2N - 1$ with no *equilibrium points*, *i.e.* with no points such that $\underline{\partial} H = 0$, where $\underline{\partial}$ denotes the gradient.

Under such conditions the Hamilton equations

e1.2.6
$$\dot{p}_i = -\frac{\partial H}{\partial q_i}, \qquad \dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad i = 1, \ldots, N, \tag{1.2.6}$$

allow us to define a group $(S_t)_{t \in \mathbb{R}}$ of maps of class C^∞ on $\mathbb{R}^{2N}$ such that for all $E \in \mathbb{R}$:
(1) the sets $\Omega_E^{\leq}$ and Ω_E are invariant.
N1.2.1 (2) the maps preserve the Lebesgue measure λ restricted to $\Omega_E^{\leq}$, or to Ω_E.[1]

The *dynamical systems* $(\Omega_E^{\leq}, (S_t)_{t \in \mathbb{R}})$, $(\Omega_E, (S_t)_{t \in \mathbb{R}})$ are called *Hamiltonian flow* with energy $\leq E$ or with fixed energy E, respectively.
The hypotheses (i) and (ii) allow us to define the normalized Lebesgue measures $\overline{\lambda}_E$ on $\Omega_E^{\leq}$ or on Ω_E obtained by normalizing to 1 the restriction of the Lebesgue measure to $\Omega_E^{\leq}$ or to Ω_E: the normalization factor turns out to be, indeed, $< +\infty$.

The pairs $(\Omega_E^{\leq}, \overline{\lambda}_E)$ or $(\Omega_E, \overline{\lambda}_E)$ are, in the language of probability theory, *probability distributions* on $\Omega_E^{\leq}$ and on Ω_E. Such distributions are invariant with respect to the Hamiltonian flows on $\Omega_E^{\leq}$ or on Ω_E, because the Hamilton equations have zero divergence. When considered together with the corresponding Hamiltonian flows, they are denoted with the symbol $(\Omega_E^{\leq}, \overline{\lambda}_E^{\leq}, (S_t)_{t \in \mathbb{R}})$ or $(\Omega_E, \overline{\lambda}_E, (S_t)_{t \in \mathbb{R}})$, respectively, and their collections, as the energy E varies, constitute two examples of *invariant statistical ensembles* or *stationary ensembles*. In Boltzmann's nomenclature they are an example of a *monode* (the first collection) and of *ergode* (the second), see Sec.1.1.

E1.2.3 *Example (1.2.3)* : Let V be a compact differentiable Riemannian manifold of class C^∞ and let $\underline{f}$ a vector field tangent to V. The differential equation

e1.2.7
$$\dot{\underline{x}} = \underline{f}(\underline{x}), \qquad \underline{x}(0) = \underline{\xi}, \tag{1.2.7}$$

[1] The Lebesgue measure restricted to Ω_E is, the measure $\lambda_E(d\underline{p}d\underline{q}) = \delta(H(\underline{p}, \underline{q}) - E)d\underline{p}d\underline{q} = \frac{d\sigma_E}{|\underline{\partial} H|}$ if $d\sigma_E$ is the surface element on Ω_E. This is locally finite in all dimensions $N > 1$ if $\underline{\partial} H$ does not vanish to a too high order at the equilibrium points, if any. See also Appendix (1.2).

allows us to define a group of maps $(S_t)_{t\in\mathbb{R}}$ of V into itself as we did in the example (1.2.1). Such maps are of class C^∞ and the pair $(V,(S_t)_{t\in\mathbb{R}})$ will be the *flow generated on* V *by* (1.2.7).

We shall say that $(S_t)_{t\in\mathbb{R}}$ preserves the volume on V if the volume measure μ on V is invariant under the action of S_t, *i.e.* if for each $E\in\mathcal{B}(V)=\{$Borel sets of $V\}$ (see Appendix (1.2)) one has

e1.2.8
$$\mu(E)=\mu(S_tE)\qquad\text{for all}\, t\in\mathbb{R}. \tag{1.2.8}$$

The measure μ will be always considered normalized (note that the volume of V is certainly finite) and the probability distribution (V,μ) will be a stationary distribution for the flow $(V,(S_t)_{t\in\mathbb{R}})$ when (1.2.8) holds. In the last case μ will be called a *stationary distribution* for the flow $(V,(S_t)_{t\in\mathbb{R}})$.

A concrete case associated with the above example is described in the following one.

E1.2.4 *Example (1.2.4)* : (*Quasi-periodic flows*) Let $\mathbb{T}^d$ be the standard d–dimensional torus (*i.e.* $\mathbb{T}^d$ is $[0,2\pi]^d$ with "opposite sides identified" or, more precisely, $\mathbb{T}^d=\mathbb{R}^d/\mathbb{Z}^d$) and consider the differential equation on $\mathbb{T}^d$

e1.2.9
$$\dot{\underline{\varphi}}=\underline{\omega}, \tag{1.2.9}$$

where $\underline{\omega}\in\mathbb{R}^d$. It defines the group $(S_t)_{t\in\mathbb{R}}$:

e1.2.10
$$S_t\underline{\varphi}=\underline{\varphi}+\underline{\omega}t=(\varphi_1+\omega_1t,\ldots,\varphi_d+\omega_dt)\ \mathrm{mod}\ 2\pi \tag{1.2.10}$$

The flow $(\mathbb{T}^d,(S_t)_{t\in\mathbb{R}})$ is called a *rotation flow* or a *quasi-periodic flow* of the torus $\mathbb{T}^d$ with *velocities* $\underline{\omega}\in\mathbb{R}^d$. It preserves the volume measure λ on $\mathbb{T}^d$ (thought of as a flat Riemannian manifold with the natural metric)

e1.2.11
$$\lambda(d\underline{\varphi})=\frac{d\underline{\varphi}}{(2\pi)^d}. \tag{1.2.11}$$

If $\underline{\rho}\in\mathbb{R}^d$ the map

e1.2.12
$$S\underline{\varphi}=\underline{\varphi}+\underline{\rho}=(\varphi_1+\rho_1,\ldots,\varphi_d+\rho_d)\ \mathrm{mod}\ 2\pi \tag{1.2.12}$$

will be called a *rotation map* of $\mathbb{T}^d$ with *rotation vector* $\underline{\rho}$. Obviously $S=S_1$ if $(S_t)_{t\in\mathbb{R}}$ is the rotation flow of $\mathbb{T}^d$ with velocities $\underline{\rho}$.

Many important examples of maps S of manifolds do not correspond to differential equations on the same manifold (in the sense that $S=S_1$ if $(S_t)_{t\in\mathbb{R}}$ is the flow associated with a suitable differential equation). A classical case

N1.2.2 is the following.[2]

[2] Another one which is widely studied in literature is the *standard map* which will be discussed in Sec. (9.2).

E1.2.5 *Example (1.2.5)* : (*Arnold's cat map*) Let S be a $d \times d$ matrix with integer entries and define, for all $\underline{\varphi} \in \mathbb{T}^d$,

e1.2.13
$$(S\underline{\varphi})_i = \sum_{j=1}^{d} S_{ij}\varphi_j \mod 2\pi, \qquad i = 1, \ldots, d. \tag{1.2.13}$$

Then S maps the torus $\mathbb{T}^d$ into itself and is of class C^∞. Furthermore, if $\det S = \pm 1$ then S is invertible and preserves the volume.

In general even if $\det S = 1$, it is not possible to find a differential equation on $\mathbb{T}^d$ generating a flow $(S_t)_{t\in\mathbb{R}}$ that interpolates S (*i.e.* such that $S_1 = S$).

A typical and important example is provided by the map of $\mathbb{T}^2$ associated, via (1.2.13), with the matrix:

e1.2.14
$$S = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}; \tag{1.2.14}$$

see problem [1.2.9].

If the map S defined by (1.2.13) has $\det S \neq \pm 1$ then it is not invertible: every $\underline{\psi} \in \mathbb{T}^d$ is an image of $|\det S|$ pairwise distinct points φ of $\mathbb{T}^d$. In this case the pair $(\mathbb{T}^d, S)$ is a *noninvertible dynamical system.*

E1.2.6 *Example (1.2.6)* : (*Lorenz' equation*) Noninvertible systems have usually origin in connection with the theory of equations, or maps, that model dissipative phenomena. Consider the equation in $\mathbb{R}^3$

e1.2.15
$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= -\sigma x - y - xz, \\ \dot{z} &= -bz + xy - \alpha; \end{aligned} \tag{1.2.15}$$

then the solutions of this equation exist globally for $t \ge 0$ for every initial datum if, as we shall suppose, σ, b, $\alpha > 0$. If $(S_t)_{t\ge 0}$ is the semigroup that solves the (1.2.15) it is, furthermore, true that the sphere

e1.2.16
$$\Omega_0 = \Big\{\underline{\xi} \mid \underline{\xi} \in \mathbb{R}^3,\ |\underline{\xi}| \le \rho_0 = \frac{2\alpha}{\min\{1, \sigma, b\}}\Big\} \tag{1.2.16}$$

is invariant (as a consequence of (1.2.4), see also problem [1.2.11]): $S_t\Omega_0 \subset \Omega_0$, for all $t \ge 0$. The pair $(\Omega_0, (S_t)_{t\ge 0})$ is a noninvertible dynamical system or a noninvertible flow on Ω_0, because S_t fails to be surjective from Ω_0 to Ω_0. Note however that S_t is invertible as a map from $\mathbb{R}^3$ into itself, see problem [1.2.11].

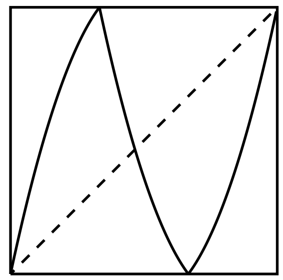
Note that the divergence of the second member of (1.2.15) is $-(1+b+\sigma) < 0$ and, hence, the measure of every (measurable) set A contracts according to

e1.2.17
$$\lambda(S_t A) = e^{-(1+b+\sigma)t}\lambda(A) \qquad t \ge 0 \tag{1.2.17}$$

i.e. the Lebesgue measure is not invariant under the flow $(S_t)_{t\ge0}$ and there
N1.2.3 can be no invariant measure on Ω_0 equivalent to it.[3] If a measure μ is absolutely continuous with respect to the Lebesgue measure or with respect to the volume measure on a manifold one often omits the reference measure and one says simply that μ is absolutely continuous.

The equation (1.2.15) is usually written using $x, y, z + r + \sigma$, instead of x, y, z, and is called *Lorenz' equation*, see equations (25),(26),(27) in [Lo63].

E1.2.7 *Example (1.2.7)* : (*Interval maps*) Another example of noninvertible dynamical system is provided by a continuous, piecewise C^∞, map $S : [0, 1] \to [0, 1]$ with a graph of the form



F1.2.1 **Fig.(1.2.1)** : *A map of the interval* $[0, 1]$.

In general the map S will not be invertible and the pair $([0, 1], S)$ will form, therefore, a noninvertible dynamical system.

The above are examples of the mathematical entities called *dynamical systems*. Because of their obvious interest, they motivate the following definitions that summarize and put in abstract form their main properties.

D1.2.1 **(1.2.1) Definition:** (Topological dynamical systems and metric dynamical systems)
Let Ω be a compact separable metric space and let S be a continuous map of Ω into itself. The pair (Ω, S) will be said a discrete topological dynamical system on Ω. *(Ω, S) will be said* invertible *if S has a continuous inverse S^{-1}.*

Let μ be a complete probability measure defined on a σ–algebra $\mathcal{B}$ of sets of Ω and let $N \in \mathcal{B}$ be a set with zero μ–measure, and suppose that S is a map measurable with respect to $\mathcal{B}$ outside N, see Appendix (1.2). If

e1.2.18
$$\mu(A) = \mu(S^{-1}A) \qquad \textit{for all } A \in \mathcal{B} \cap N^c, \tag{1.2.18}$$

*where $N^c = \Omega/N$ is the complement of N, we shall say that (Ω, S) is μ–*preserving *or that μ* is S–invariant. *The triple (Ω, S, μ) will be called a*

[3] Given two measures μ and ν defined on the same σ–algebra $\mathcal{B}$, see Appendix (1.2), ν is absolutely continuous with respect to μ if there exists $f \in L_1(\mu)$ such that $\nu = f\mu$. The measures μ and ν are equivalent if μ is absolutely continuous with respect to ν and ν is absolutely continuous with respect to μ.

(discrete) metric dynamical system mod 0 *defined outside* N. *If* $N=\emptyset$ *the triple* (Ω,S,μ) *will be called a* (discrete) metric dynamical system.

Often the set N of "singularities of S" is not explicitly mentioned and one simply talks of a metric dynamical system (Ω,S,μ). We shall try to avoid this practice because of the risks of confusion it generates. Analogously:

D1.2.2 **(1.2.2) Definition:** (Topological flows and metric flows)
Let Ω *be a compact separable metric space. Let* $(S_t)_{t\in\mathbb{R}}$ *or* $(S_t)_{t\in\mathbb{R}_+}$ *be a group or a semigroup, homomorphic to* $\mathbb{R}$ *or to* $\mathbb{R}_+$*, of maps that act with*
N1.2.4 *continuity on* Ω.[4]
The pair $(\Omega,(S_t)_{t\in\mathbb{R}})$ *or* $(\Omega,(S_t)_{t\in\mathbb{R}_+})$ *will be called respectively an* invertible topological flow *or a* topological flow *on* Ω.
Let μ *be a complete probability measure on a* σ*–algebra* $\mathcal{B}$ *and let* $(S_t)_{t\in\mathbb{R}}$ *be a group of* μ*–measurable maps homomorphic to* $\mathbb{R}$ *and preserving* μ*. Suppose that the function* $(t,x)\to S_tx$ *defined on* $\mathbb{R}\times\Omega$ *with values in* Ω *is* μ*–measurable with respect to the* σ*–algebra generated by the sets in* $\mathcal{B}(\mathbb{R})\times\mathcal{B}$*. Then the flow* $(\Omega,(S_t)_{t\in\mathbb{R}},\mu)$ *will be called a* invertible metric flow *on* Ω.
If, mutatis mutandis, we replace $(S_t)_{t\in\mathbb{R}}$ *by a semigroup* $(S_t)_{t\in\mathbb{R}_+}$, $\mathcal{B}(\mathbb{R}_+)\times\mathcal{B}$*–measurable, we obtain the notion of* metric flow.

Flows will also be called, sometimes, *continuous dynamical systems* because their time "elapses continuously".
Speaking of dynamical systems one often omits the qualifications (topological, metric, discrete, continuous, invertible,...) that are usually supposed to be understandable from the context. One can (generously) even consider the (here seldom used) notions of *abstract discrete dynamical system* (Ω,S) with Ω being a "space" and S a "map" acting on it, as well as the similar notion of *abstract flow*.
Before proceeding to analyze the structure of certain classes of dynamical systems it is convenient to set up the notion of *isomorphism* between dynamical systems.

D1.2.3 **(1.2.3) Definition:** (Isomorphisms)
If (Ω,S) *and* (Ω',S') *are two abstract discrete dynamical systems, we shall say that they are* isomorphic*, or* conjugated*, when there exists an invertible map* $I:\Omega\leftrightarrow\Omega'$ *such that*

e1.2.19
$$IS=S'I. \tag{1.2.19}$$

If the systems (Ω,S) *and* (Ω',S') *are discrete topological dynamical systems we shall say that they are topologically isomorphic if they are isomorphic and if the isomorphism* I *can be chosen bicontinuous.*
Let (Ω,S,μ) *and* (Ω',S',μ') *be two discrete metric dynamical systems* mod 0*, defined outside the sets* N,N' *of zero* μ,μ' *measures and defined on the* σ*–algebras* $\mathcal{B}_\mu$ *and* $\mathcal{B}_{\mu'}$ *respectively. If* $(\Omega/N,S)$ *and* $(\Omega'/N',S')$ *are*

---

[4] This means that the functions $(t,x)\to S_tx$ of $\mathbb{R}\times\Omega$ or of $\mathbb{R}_+\times\Omega$ into Ω are continuous.

isomorphic and if I can be chosen bimeasurable outside N, N' with respect to the σ–algebras $\mathcal{B}_\mu$ and $\mathcal{B}_{\mu'}$ and if, furthermore,

e1.2.20
$$\mu'(IA) = \mu(A) \qquad \textit{for all } A \in \mathcal{B}_\mu \tag{1.2.20}$$

we shall say that the systems are isomorphic mod 0.

In the latter case there exist two zero measure sets $N \in \mathcal{B}$ and $N' \in \mathcal{B}'$ such that $S^{-1}N = SN = N$ and $S'^{-1}N' = S'N' = N'$ and furthermore $(\Omega/N, S, \mu)$ and $(\Omega'/N', S', \mu')$ are isomorphic in the sense that there is a bimeasurable map I of Ω/N to Ω'/N' verifying $IS = S'I$ on such sets and mapping μ onto μ'. One can formulate, in an analogous fashion, various notions of isomorphism between flows.

Appendix 1.2: *Basic definitions of measure theory*

We recall here some basic defintitions of measure theory. A collection of sets among which is the empty set and that is closed under the operations of complementation and of finite union is called an *algebra*; if it is also closed under countable union is called a *σ–algebra*. The smallest σ–algebra containing a given family $\mathcal{F}$ of sets is called the σ–algebra generated by $\mathcal{F}$. The σ–algebra generated by the open sets of a topological space Ω is called the *Borel σ–algebra* $\mathcal{B}(\Omega)$ and its elements are called *Borel sets*. A *measure* μ on a σ–algebra $\mathcal{B}$ is a countably additive non-negative function of the sets of $\mathcal{B}$, which are called *μ–measurable* sets. Given a σ–algebra $\mathcal{B}$ and a measure μ we can consider the σ–algebra $\mathcal{B}_\mu$ generated by $\mathcal{B}$ and all the sets contained in a μ–measurable set with zero μ measure. The σ–algebra $\mathcal{B}_\mu$ is called the *μ–completion* of $\mathcal{B}$. A measure μ can always be extended uniquely to a measure on the σ–algebra $\mathcal{B}_\mu$. If $\mathcal{B} = \mathcal{B}_\mu$ the measure μ is called *complete*. A measure μ which is defined on the Borel sets of a topological space is called a *Borel measure*, while a measure that is the completion of its restriction to the Borel sets is called a *complete Borel measure*.

The usual Riemann measure on $\mathbb{R}^d$ is defined over the algebra of sets which are approximable from inside and from outside by unions of rectangles with an arbitrarily small volume in between. It is not countably additive but it can be extended to the Borel sets and then completed defining in this way the Lebesgue measure. On Riemannian manifolds the Riemann measure is defined in the same way by using local charts and the volume form generated by the metric. The usual Lebesgue measure and, more generally, the volume measures λ on Riemannian manifolds will be regarded as complete measures defined on the Borel sets $\mathcal{B}(\Omega)$ and extended to $\mathcal{B}_\lambda(\Omega)$, see problem [1.2.19].

Given a function S from Ω to Ω' the inverse image $S^{-1}(E)$ of a set $E \subset \Omega'$ is the set of all $x \in \Omega$ such that $S(x) \in E$. A function (map) S between two spaces Ω and Ω' is measurable with respect to the σ–algebras $\mathcal{B}$ and $\mathcal{B}'$, defined respectively on Ω and Ω', if the inverse image of every set $E \in \mathcal{B}'$ is an element of $\mathcal{B}$. A function S measurable with respect to the completion $\mathcal{B}_\mu$ of $\mathcal{B}$ relative to a measure μ is called μ–measurable. A function S is called *bimesurable* if S^{-1} exists and if both S and S^{-1} are measurable. If $\Omega = \Omega'$ and $\mathcal{B} = \mathcal{B}'$ we simply say that S is measurable with respect to $\mathcal{B}$. Given a set N in $\mathcal{B}$ we call $\mathcal{B}(\Omega) \cap N^c$ the σ-algebra generated by the sets $A\underline{\nu}$ for $A \in \mathcal{B}$. If N, N' are sets in $\mathcal{B}, \mathcal{B}'$ and the map S of $\Omega \setminus N$ into $\Omega' \setminus N'$ is such that $S^{-1}E' \in \mathcal{B}(\Omega) \cap N^c$ for all $E' \in \mathcal{B}(\Omega') \cap N'^c$ then S is said to be measurable with respect to $\mathcal{B}, \mathcal{B}'$ outside N, N'. If both S and S^{-1} are defined and both are $\mathcal{B}, \mathcal{B}'$ measurable outside N, N' then the map S is said *bimeasurable* with respect to $\mathcal{B}, \mathcal{B}'$ outside N, N'.

Problems for §1.2

Q1.2.1 **[1.2.1]:** Show that the differential equation (1.2.1) admits, as stated in example (1.2.1), global solutions under the assumptions considered there.

Q1.2.2 **[1.2.2]:** (*Liouville theorem*)
Show that equation (1.2.2) implies equation (1.2.3).

Q1.2.3 **[1.2.3]:** Show that equation (1.2.4) implies the invariance of Ω.

Q1.2.4 **[1.2.4]:** Construct an example of a non-Hamiltonian differential equation in $\mathbb{R}^2$ possessing first integrals in the sense of example (1.2.1), in $C^\infty(\mathbb{R}^2)$.

Q1.2.5 **[1.2.5]:** Show that $\dot{\underline{x}} = -\underline{x}$ is a differential equation with no non-constant C^∞ first integrals.

Q1.2.6 **[1.2.6]:** Consider a differential equation in $\mathbb{R}^d$ of the type considered in example (1.2.1) and suppose that for every $\underline{x} \in \mathbb{R}^d$ the limit $\lim_{t\to+\infty} S_t\underline{x}$ exists and takes only a finite number of values as $\underline{x}$ varies in $\mathbb{R}^d$. Show that the differential equation does not admit non-constant C^∞ first integrals.

Q1.2.7 **[1.2.7]:** Find some other criterion of non-existence of first integrals for a differential equation inspired by that of problems [1.2.5], [1.2.6].

Q1.2.8 **[1.2.8]:** Show that, in the case of equation (1.2.13), if $\det S = \pm 1$ then S is invertible and it is of class C^∞ together with its inverse.

Q1.2.9 **[1.2.9]:** (*A map not embeddable into a flow*)
Show the non-existence of a differential equation on the torus $\mathbb{T}^2$ such that $S_1 = S$ where S is the map of $\mathbb{T}^2$ into itself defined by (1.2.14). (*Hint:* Show that if $\dot{\underline{\varphi}} = \underline{f}(\underline{\varphi})$ were such an equation we would have: $\sum_j S_{ij} f_j(\underline{\varphi}) \equiv f_i(S\underline{\varphi})$ and, more generally: $\sum_j (S^n)_{ij} f_j(\underline{\varphi}) \equiv f_i(S^n\underline{\varphi})$ for all $n \in \mathbb{Z}$. This is absurd. In fact since the matrix S is Hermitian with eigenvalues $(3 \pm \sqrt{5})/2$, one greater and one smaller than 1, $\limsup_{|n|\to\infty} |S^n\underline{v}| = +\infty$, for all $\underline{v} \in \mathbb{R}^2 \setminus \{\underline{0}\}$, while $|\underline{f}(S^n\underline{\varphi})| \le \max_{\underline{\psi}} |\underline{f}(\underline{\psi})| < \infty$).

Q1.2.10 **[1.2.10]:** (*Toral maps*)
Show that if S is a matrix with integer entries and if $N = |\det S| \neq 0$ then S is a continuous map of the torus into itself and every point of $\mathbb{T}^d$ is the image of N points of $\mathbb{T}^d$. Furthermore S is of class C^∞ on $\mathbb{T}^d$.

Q1.2.11 **[1.2.11]:** (*Lorenz' equation*)
Show that Lorenz' equation (of example (1.2.6)) admits global solutions (in the past and in the future); and show that Ω_0 is invariant only in the future. (*Hint:* Multiply scalarly the (1.2.15) by (x, y, z) and deduce an *a priori* bound on the solutions of (1.2.15) by noting that the cubic terms obtained in this way disappear. Use criterion (1.2.4) to conclude.)

Q1.2.12 **[1.2.12]:** (*Expanding interval maps*)
Let $Sx = 2x$, $0 \le x \le 1/2$, $Sx = 2(1-x)$, $1/2 \le x \le 1$, cf. example (1.2.7). Show that

the Lebesgue measure λ on $[0,1]$ is S–invariant ($\lambda(E)=\lambda(S^{-1}E)$, for all $E\in\mathcal{B}([0,1])$).

Q1.2.13 **[1.2.13]:** (*The equation for invariant density*)
Consider a continuous piecewise smooth (in general noninvertible) map T of $[0,1]$ and suppose that the measure $\mu_T(dy)=f(y)dy$, absolutely continuous with respect to the Lebesgue measure, is T–invariant, see definition (1.2.1): then the equation for its density is $f(y)=\sum_{z:\,Tz=y}|T'(z)|^{-1}f(z)$.

Q1.2.14 **[1.2.14]:** Let S and T be two smooth maps of $[0,1]$ which are *conjugated* or *isomorphic*, *i.e.* such that there exists an invertible $C^1([0,1])$ map $F:[0,1]\to[0,1]$ such that $Ty=FSF^{-1}y$, and suppose that $\mu_T(dy)=f(y)dy$ is T–invariant (see problem [1.2.13]). Check that setting $\mu_S(E)=\mu_T(F(E))$, *i.e.* $\mu_S(dx)=f(F(x))|F'(x)|dx$, yields an S–invariant measure μ_S.

Q1.2.15 **[1.2.15]:** (*Ulam–von Neumann map*)
Show that the Ulam-von Neumann map $x\to T(x)=4x(1-x)$ and the map $x=Sx$, where S is defined as in problem [1.2.12], are conjugated via the map $y=F(x)=(2/\pi)\arcsin\sqrt{x}$. The map $T(x)=ax(1-x)$, with $a\in\mathbb{R}$, is sometimes referred to as the *logistic map*.

Q1.2.16 **[1.2.16]:** It is known that every compact C^∞ Riemannian manifold V can be smoothly immersed into an Euclidean space of suitably large dimension n. Denoting by Y such an immersion show that there exists a differential equation $\dot{\underline{\xi}}=\underline{\varphi}(\underline{\xi})$ in $\mathbb{R}^n$, of class C^∞, that admits the manifold $\widetilde V=Y(V)$ as an invariant manifold and that induces on $\widetilde V$ a flow $(\widetilde S_t)_{t\in\mathbb{R}}$ that is the image of the flow on V generated by equation (1.2.7) so that, denoting the latter by $(S_t)_{t\in\mathbb{R}}$, it is $\widetilde S_tY\underline{x}=YS_t\underline{x}$, $t\in\mathbb{R}$, $\underline{x}\in V$.

Q1.2.17 **[1.2.17]:** (*Geodesic flow*)
Let V be a C^∞ Riemannian manifold, let g be its metric tensor and let W be the *tangent bundle*, *i.e.* the manifold consisting of the points (x,v) with $x\in V$ and with v tangent to V in x. Define on W the *geodesic flow* $(S_t)_{t\in\mathbb{R}}$ that associates with (x,v) the point $S_t(x,v)=(x_t,v_t)$ obtained by constructing the geodesic that starts in x in the direction v and by running on it with uniform velocity given by $|v|=\sqrt{\sum_{ij}g_{ij}v^iv^j}$ on a portion of length $|v|\,t$ (in the metric g) reaching in this way a point x_t with velocity v_t.

It is convenient to describe the geodesic flow as a flow on the space $\widehat W$ of the pairs (x,p) of points of V and of vectors *cotangent* to V in x. Recall that the vector p cotangent in x and corresponding to the vector v tangent in x is, by definition:

$$p_i=\sum_{j=1}^d g_{ij}(x)v^j,\qquad i=1,\ldots,d.$$

The geodesic flow on $\widehat W$ is naturally described by the correspondence $(x,p)\to(x_t,p_t)$ built by
(1) starting with $(x,p)\in\widehat W$ and constructing, via the preceding construction, the vector v tangent in x, namely $v=g^{-1}(x)p$,
(2) associating with (x,v) the point $S_t(x,v)=(x_t,v_t)$ and then
(3) defining $(p_t)_i=\sum_{j=1}^d g_{ij}(x_t)v_t^j$.

A celebrated proposition of geometry and of mechanics states that the geodesic flow is described on $\widehat W$, in every chart, by the Hamiltonian differential equations with Hamiltonian

$$H(x,p)=\frac12\sum_{j=1}^d(g^{-1}(x))^{ij}p_ip_j.$$

Deduce that the geodesic flow on W preserves the measure that on $\widehat W$ coincides with the Lebesgue measure $dx\,dp$. Deduce that the geodesic flow on W conserves a measure that is absolutely continuous with respect to the volume measure on W considered as a Riemannian manifold with the natural metric $\Big(\sum_{j=1}^{d}(g_{ij}(x)dx^i dx^i + g_{ij}(x)dv^i dv^j)\Big)^{\frac12}$ and express the density with respect to the measure that in a natural chart for W becomes the measure $dx\,dv$. Find its expression in terms of the metric tensor.

$Q1.2.18$ **[1.2.18]:** Deduce the differential equation of the geodesics on V and hence that of the geodesic flow on W and $\widehat W$ by using the definition of geodesic as an extremal curve for the line element $ds^2 = \sum_{i,j} g_{ij}(x)dx^i dx^j$ and the principles of Mechanics. (*Hint:* The geodesic flow takes place at constant speed and takes place over curves that minimize or make stationary $\int_{P_1}^{P_2}\sqrt{\sum_{i,j} g_{ij}dx^i dx^j}$ over all lines joining P_1, P_2.).

$Q1.2.19$ **[1.2.19]:** (*Incompleteness of the restriction of Lebesgue measure to Borel sets*) Assuming that there exist sets in $[0,1]$ that are not Lebesgue measurable show that there exist sets in $\mathbb{R}^2$ which are not Borel sets but that are contained in sets of 0–Lebesgue measure. Show in fact that if L is a not Lebesgue–measurable set of the segment $[0,1]$ then the set $L\times\{0\}\subset[0,1]\times[0,1]$ is an example of a set which is not a Borel set of the square $[0,1]^2$ and it is nevertheless contained inside the set $[0,1]\times\{0\}$ with 0 Lebesgue measure. (*Hint:* Just show that $L\times\{0\}$ cannot be a Borel set as a subset of the square $[0,1]\times[0,1]$: if so it could be obtained by a transfinite induction via operations of countable unions and intersections starting from open sets of the square. But the same transfinite construction performed with the intersections of the open sets with the segment $[0,1]\times\{0\}$ would lead to the set L which would therefore be a Borel subset of $[0,1]$.)

Bibliographical note to §1.2

Examples and definitions are taken from [AA68], [Si77], [Av76] where several other examples can be found.

§1.3 Harmonic oscillators and integrable systems as dynamical systems

We shall consider a few simple dynamical systems which provide us with a first illustration of the just introduced notions and establish a relation between the examples (1.2.2), (1.2.3), (1.2.4) of Section §1.2.

$P1.3.1$ **(1.3.1) Proposition:** (Harmonic oscillations)
Let $G=(g_{ij})$, $V=(v_{ij})$ be two positive definite symmetric matrices and let $G^{-1}=(g^{-1}_{ij})$ be the inverse matrix of G. Consider the function

$e1.3.1$
$$H = H(\underline{p},\underline{q}) = \frac12\Big(\sum_{i,j=1}^{r} g^{-1}_{ij}p_i p_j + \sum_{i,j=1}^{r} v_{ij}q_i q_j\Big) \tag{1.3.1}$$

$N1.3.1$ *on $\mathbb{R}^{2r}$, and let $\eta^{(1)},\ldots,\eta^{(r)}$ be G–orthonormal vectors*[1] *verifying*

$e1.3.2$
$$(-\omega_k^2 G + V)\underline{\eta}^{(k)} = \underline{0}, \qquad k=1,\ldots,r, \tag{1.3.2}$$

[1] *i.e.* such that $(G\underline{\eta}^{(i)}\cdot\underline{\eta}^{(j)}) = \sum_{h,k=1}^{r} g_{hk}\eta_h^{(i)}\eta_k^{(j)} = \delta_{ij}$.

where $\omega_1^2, \ldots, \omega_r^2$ *are the roots, ordered and repeated according to multiplicity, of the secular equation* $\det(-\omega^2 G + V) = 0$.

Define $\underline{x}$ *and* $\underline{\dot{x}}$ *in* $\mathbb{R}^r$ *so that*

e1.3.3
$$\underline{q} = \sum_{i=1}^{r} x^{(i)} \underline{\eta}^{(i)}, \qquad G^{-1}\underline{p} = \sum_{i=1}^{r} \dot{x}^{(i)} \underline{\eta}^{(i)}, \tag{1.3.3}$$

and consider the motion associated with the Hamiltonian H *in* $\mathbb{R}^{2r}$ *with given initial data* $\underline{p}$, $\underline{q}$. *This motion is described by* $t \to S_t(\underline{p}, \underline{q}) \equiv (\underline{p}(t), \underline{q}(t))$, *with*

e1.3.4
$$\begin{aligned} \underline{q}(t) =& \sum_{i=1}^{r} \Big(x^{(i)} \cos \omega_i t + \frac{\dot{x}^{(i)}}{\omega_i} \sin \omega_i t \Big) \, \underline{\eta}^{(i)}, \\ \underline{p}(t) =& G \frac{d\underline{q}}{dt}(t) \equiv G \underline{\dot{q}}(t). \end{aligned} \tag{1.3.4}$$

The proof of this well known proposition is, for instance, a simple check by substitution of (1.3.4) into the equations of motion.

Remark: It is convenient to remark that the parameters $\underline{x}$ and $\underline{\dot{x}}$ are determined by equation (1.3.3) simply by using the G–orthonormality of the vectors $\underline{\eta}^{(1)}, \ldots, \underline{\eta}^{(r)}$

e1.3.5
$$\begin{aligned} x^{(i)} &= G\underline{\eta}^{(i)} \cdot \underline{q} \equiv (\sqrt{G}\underline{\eta}^{(i)} \cdot \sqrt{G}\underline{q}), \\ \dot{x}^{(i)} &= \quad \underline{\eta}^{(i)} \cdot \underline{p} \equiv (\sqrt{G}\underline{\eta}^{(i)} \cdot \sqrt{G^{-1}}\underline{p}), \end{aligned} \tag{1.3.5}$$

In other words $(x^{(i)}, \dot{x}^{(i)})$ are the components of $\sqrt{G}\underline{q}$ and of $\sqrt{G^{-1}}\underline{p}$ on the orthonormal basis in $\mathbb{R}^r$ formed by the vectors $\sqrt{G}\underline{\eta}^{(i)}$. Note that the map
N1.3.2 $\underline{p} \to \underline{p}' = (\sqrt{G})^{-1}\underline{p}$ and $\underline{q} \to \underline{q}' = (\sqrt{G})\underline{q}$ is a linear canonical map.[2]

From (1.3.5) and from the above proposition an easy, but remarkable, corollary follows.

C1.3.1 **(1.3.1) Corollary:** (Action–angle coordinates for harmonic oscillations) *Under the assumptions of proposition (1.3.1) let, for* $(\underline{p}, \underline{q}) \in \mathbb{R}^{2r}$

e1.3.6
$$\begin{aligned} A_i &= \Big(x^{(i)2} + \big(\dot{x}^{(i)}/\omega_i\big)^2 \Big)^{1/2}, \\ \cos \varphi_i &= \frac{x^{(i)}}{A_i}, \qquad \sin \varphi_i = \frac{\dot{x}^{(i)}}{A_i \omega_i}, \end{aligned} \tag{1.3.6}$$

[2] We recall that a $2n \times 2n$ matrix $L = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ is called *canonical* (or also *symplectic*) if $L^{-1} = \begin{pmatrix} D^T & -B^T \\ -C^T & A^T \end{pmatrix}$, where T denotes transposition: a map defined on an open domain of $\mathbb{R}^{2n}$ into $\mathbb{R}^{2n}$ is called canonical (or symplectic) if its Jacobian matrix is canonical at every point of the domain.

where $x^{(i)}$ *and* $\dot x^{(i)}$ *are determined by (1.3.5).*
Let $W \subset \mathbb{R}^{2r}$ *be the set of the points* $(\underline{p},\ \underline{q}) \in \mathbb{R}^{2r}$ *for which* $A_i \neq 0$, $i = 1, \ldots, r$. *Define on* W, *via (1.3.6), the map* $I(\underline{p}, \underline{q}) = (\underline{A}, \underline{\varphi}) = (A_1, \ldots, A_r, \varphi_1, \ldots, \varphi_r) \in (\mathbb{R}_+ \setminus \{\underline{0}\})^r \times \mathbb{T}^r$.
(1) The map I *is invertible as a map between* W *and* $IW = (\mathbb{R}_+ \setminus \{0\})^r \times \mathbb{T}^r$*; furthermore it has a Jacobian determinant which does not vanish on* W,

e1.3.7
$$\Big|\frac{\partial I^{-1}(\underline{A}, \underline{\varphi})}{\partial \underline{A}\ \partial \underline{\varphi}}\Big| = (\det G) \prod_{i=1}^{r} \omega_i A_i, \tag{1.3.7}$$

and, therefore, I *is analytic and invertible on* W.
(2) In the coordinates $(\underline{A}, \underline{\varphi})$ *the motion (1.3.4) becomes*

e1.3.8
$$I(S_t(p, q)) = (\underline{A}, \underline{\varphi} + \underline{\omega} t) \bmod 2\pi. \tag{1.3.8}$$

Remarks: (1) For $\underline{a} = (a_1, \ldots, a_r) \in (\mathbb{R}_+ \setminus \{0\})^r$ we set

e1.3.9
$$\Omega_{\underline{a}} = \{(\underline{p}, \underline{q}) \in \mathbb{R}^{2n} | A_1 = a_1, \ldots, A_r = a_r\}, \tag{1.3.9}$$

and we denote with

e1.3.10
$$\lambda_{\underline{a}}(d\underline{p} d\underline{q}) = \frac{\prod_{i=1}^{r} \delta(A_i(\underline{p}, \underline{q}) - a_i) d\underline{p} d\underline{q}}{\text{“normalization to 1”}} = \prod_{i=1}^{r} \frac{d\varphi_i}{2\pi} \tag{1.3.10}$$

the Lebesgue measure $d\underline{p}\, d\underline{q}$ restricted to $\Omega_{\underline{a}}$ and normalized to 1. If S_t denotes the Hamiltonian evolution on $\mathbb{R}^{2r}$ corresponding to (1.3.4), we can say that $\Omega_{\underline{a}}$ is invariant, that $\lambda_{\underline{a}}$ is invariant and that the metric flow $(\Omega_{\underline{a}}, (S_t)_{t \in \mathbb{R}}, \lambda_{\underline{a}})$ is isomorphic to the rotation of $\mathbb{T}^r$ with velocity $\underline{\omega}$, *i.e.* to the dynamical system $(\mathbb{T}^r, (\widetilde{S}_t)_{t \in \mathbb{R}}, \lambda)$ with $\widetilde{S}_t \underline{\varphi} = \underline{\varphi} + \underline{\omega} t (\bmod 2\pi)$ and $\lambda(d\underline{\varphi}) = \prod_i \frac{d\varphi_i}{2\pi}$.
(2) This is an isomorphism to which one refers by stating that the Hamiltonian system (1.3.1) is a system of harmonic oscillators whose trajectories develop with uniform velocity on r–dimensional invariant tori parameterized by r parameters (each of which, of course, is a first integral).
(3) One says that the phase space W is *foliated* into analytic r–dimensional invariant tori: this refers to the relation $W = \cup_{\underline{a} \in (\mathbb{R}_+ \setminus \{0\})^r} \Omega_{\underline{a}}$ and to the fact that $\Omega_{\underline{a}}$ topologically is a torus, which by (1.3.6), depends analytically on r parameters.

The property of oscillators systems established by corollary (1.3.1) lead to their natural generalization expressed by the following definition that isolates a class of Hamiltonian systems whose phase space can be thought of as foliated into r–dimensional invariant tori parameterized by r parameters and such that the Hamiltonian flow on each of them reduces to a constant velocity flow, *i.e.* to a quasi-periodic motion.

D1.3.1 **(1.3.1) Definition:** (Hamiltonian integrability)
A Hamiltonian system of class C^∞ in $\mathbb{R}^{2r}$ is called integrable *in an open region $W \subset \mathbb{R}^{2r}$ of phase space if there exists a regular change of coordinates*
N1.3.3 [3] *I that transforms W into $V \times \mathbb{T}^r$, with V open subset of $\mathbb{R}^r$, and, denoting by $(S_t)_{t\in\mathbb{R}}$ the Hamiltonian flow on W and setting*

e1.3.11
$$(\underline{A}, \underline{\varphi}) = I(\underline{p}, \underline{q}), \tag{1.3.11}$$

one has

e1.3.12
$$I(S_t(\underline{p}, \underline{q})) = (\underline{A}, \underline{\varphi} + \underline{\omega}(\underline{A})t) \bmod 2\pi \tag{1.3.12}$$

where $\underline{\omega}(\underline{A}) = (\omega_1(\underline{A}), \ldots, \omega_r(\underline{A}))$ are r functions of class C^∞ on V, called velocities.

If the Hamiltonian is analytic and the map I and the functions $\underline{A} \to \underline{\omega}(\underline{A})$ are analytic on the respective domains of definition the system is called analytically integrable.

Finally if I is a canonical map, see footnote 2 above, we shall say that the system is canonically integrable.

Remark: (1) The linear oscillators are an *analytically integrable* system in the region W defined by proposition (1.3.1). The velocities $\underline{\omega}$ are in this case A–independent: this is the phenomenon of *isochrony* of harmonic oscillations.

(2) There exist several other examples, although not really many, of physically or mathematically interesting integrable systems. We refer to treatises on Rational Mechanics for their list and analysis (such systems are classically called *systems integrable by quadratures*). Here we mention only the system

e1.3.13
$$H(\underline{p}, \underline{q}) = \frac{1}{2}\underline{p}^2 - g/|\underline{q}| \qquad \underline{p}, \underline{q} \in \mathbb{R}^d \times \mathbb{R}^d, \qquad d = 2, 3, \tag{1.3.13}$$

known as the *Kepler system.* This system is integrable in the region in which $H < 0$ and $|\underline{q} \wedge \underline{p}| \neq 0$, but it is not isochronous (indeed the third Kepler's law shows that the period of a motion depends on parameters of the motion itself).

We shall come back later to integrable systems to analyze some of their aspects that are more interesting from a technical and conceptual point of view: the above brief introduction is motivated by the simplicity of their motions and their relevance for the discussion of complexity of a dynamical system. The latter notion, of obvious interest, will be the first, among several notions typical of ergodic theory: integrable systems will be the prototype of systems exhibiting simple motions. They are well suited to be discussed, from this point of view, as an introduction to the theory of more complex motions.

---

[3] This means that the change of coordinates is of class C^∞, that it is invertible and that it has non-zero Jacobian determinant.

In the next section we begin this analysis by studying the simplest property of regularity of a motion: namely the property of visiting given regions with well defined frequency.

Problems for §1.3

Q1.3.1 **[1.3.1]:** (*Analytical integrability of harmonic oscillations*)
Show that the system of harmonic oscillators of proposition (1.3.1) is also analytically canonically integrable (and not just analytically integrable): the analytic and canonical map that integrates it on W is $(\underline{p},\underline{q}) \to (\underline{B},\underline{\varphi})$ with $\underline{\varphi}$ defined by equation (1.3.6) and $B_i = \omega_i A_i^2/2$. In the new coordinates one has $H(\underline{p},\underline{q}) = \sum_{i=1}^{r} \omega_i B_i$. (*Hint:* (a) The map $(\underline{p},\underline{q}) \to (\dot{\underline{x}},\underline{x})$ is canonical as remarked after proposition (1.3.1); hence the problem is reduced to the case $r = 1$; (b) the map $(\dot{x}_i, x_i) \to (B_i, \varphi_i)$ is canonical for $i = 1, \ldots, r$
N1.3.4 because it has as generating function $F_0(x_i, \varphi_i) = \frac{1}{2}\omega_i x_i^2 \tan\varphi_i$.[4], cf. [Ga82].)

Q1.3.2 **[1.3.2]:** (*Analytical integrability of Kepler's system*)
Show that equation (1.3.13) is analytically integrable via a map which is canonical in the region $W = \{\underline{p},\underline{q} |\, |\underline{q} \wedge \underline{p}| \neq 0,\, H(\underline{p},\underline{q}) < 0\}$. Compute explicitly the variables $\underline{A}$, $\underline{\varphi}$ in the bidimensional case. (*Hint:* Just suitably interpret the classical results of the two-body Kepler problem).

§1.4 Frequencies of visit

Let (Ω, S) be a (discrete) invertible topological dynamical system. Motions $i \to S^i x$, $i \in \mathbb{Z}$, beginning in $x \in \Omega$, will be usually described by selecting a certain number of possible properties of a point of Ω can enjoy and, then, by listing which of such property is actually possessed by the points successively visited by x in its motion.

The mathematical model associated with the latter description is the *history* $\underline{\sigma}(x)$ of x on a partition $\mathcal{P} = \{P_0, P_1, \ldots, P_n\}$ of Ω into $n+1$, $n \geq 1$, pairwise disjoint sets where $\underline{\sigma}(x) = \{\sigma_i(x)\}_{i\in\mathbb{Z}} \in \{0, \ldots, n\}^{\mathbb{Z}}$ is the sequence such that

e1.4.1
$$S^i x \in P_{\sigma_i(x)} \qquad \text{for all } i \in \mathbb{Z}. \tag{1.4.1}$$

The set P_i, $i = 1, 2, 3, \ldots, n$, is the collection of the points of Ω that enjoy the property indicated with the label i, so that the union $\bigcup_{i=1}^{n} P_i$ is the set of points that enjoy anyone of the properties labeled with $1, 2, \ldots, n$ while $P_0 = \Omega \setminus \cup_{i=1}^n P_i$ is the set of the points that do not enjoy any of them.

We will always require that the elements of the partition $\mathcal{P}$ (often called *atoms* of $\mathcal{P}$) be Borel sets in $\mathcal{B}(\Omega)$ (see Appendix 1.2): one says that such a partition is a *Borel partition*. For the time being, we do not impose further regularity requirements on the atoms of $\mathcal{P}$.

---

[4] We recall that a sufficient condition for canonicity of a map $I : W \longleftrightarrow V \times \mathbb{T}^r$ is that for all $(\underline{p}_0, \underline{q}_0) \in W$ there exist a neighborhood U_0 of $(\underline{A}_0,\, \underline{q}_0)$ and a function $F_0 : (\underline{A},\underline{q}) \to F_0(\underline{A},\underline{q})$ in $C^\infty(\mathbb{R}^{2r})$ such that for all $(\underline{A},\underline{q}) \in U_0$ one has: $\underline{p} = \frac{\partial F_0}{\partial \underline{q}}(\underline{A},\underline{q})$ and $\underline{\varphi} = \frac{\partial F_0}{\partial \underline{A}}(\underline{A},\underline{q})$ if $(\underline{A},\underline{\varphi}) = I(\underline{p},\underline{q})$.

It is nevertheless clear that in the applications we shall be interested only on rather reasonable partitions. For example we often consider the following types of partitions

D1.4.1 **(1.4.1) Definition:** (Topological, analytically regular, C^∞-regular partitions)
(i) Given a compact metric space Ω we call topological partitions $\mathcal{P}$ those in which each P_i is either open or closed for all $i = 0, 1, \ldots, n$, and, furthermore, the closure $\overline{P}_i$ of P_i and the closure of its interior $\overline{int\ P_i}$ coincide: $\overline{P}_i = \overline{int\ P_i}$.
(ii) When Ω is a class C^∞ or analytic manifold one often considers topological partitions $\mathcal{P}$ with C^∞–regular or, respectively, analytically regular atoms, cf. *Appendix (1.4).*
We often briefly call any such partition a regular partition.

E1.4.1 *Example (1.4.1)* : Let $\mathbb{T}$ be the unit circle and S the rotation by an angle $\rho \in (0, 2\pi)$ (one also says a rotation with *rotation number* $\rho/2\pi$). We can consider the partion $\mathcal{P}$ for the dynamical system $(\mathbb{T}, S)$ formed by the sets $P_0 = [0, \pi/2]$, $P_1 = (\pi/2, \pi)$, $P_2 = [\pi, 3\pi/2]$ and $P_3 = (3\pi/2, 2\pi)$. This simple example will be used in the problems to illustrate some of the concepts discussed in what follows.

Before formalizing with a mathematical definition the notion of "observation" of a motion on a given partition of (Ω, S) and before introducing other possible properties and restrictions on $\mathcal{P}$ it is convenient to introduce some philosophical considerations that should elucidate such a notion avoiding a too abstract tone and form.

It is natural to think of the result of "real observations" of a motion starting at x as sequences $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ generated by making large the observation time N during which the motion visits successively the elements P_{σ_i}, $i = -N, \ldots, N$. For all N the string $\underline{\sigma}^{(N)} \in \{0, \ldots, n\}^{[-N,N]}$ has the property

e1.4.2
$$\bigcap_{k=-N}^{N} S^{-k} P_{\sigma_k} \neq \emptyset, \tag{1.4.2}$$

because $S^k x \in P_{\sigma_k}$ for all k, so that the intersection in (1.4.2) certainly contains at least x.

Therefore a motion, or better an observation of a motion, will appear as a sequence $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ that enjoys the *finite intersection property*, in the sense that

e1.4.3
$$\bigcap_{j \in J} S^{-j} P_{\sigma_j} \neq \emptyset \qquad \text{for all } J \subset \mathbb{Z}\ , \ |J| < \infty, \tag{1.4.3}$$

where $|J|$ indicates the number of elements of J.

Hence it will be natural to identify the space of the motions *observed on* $\mathcal{P}$ with the set of infinite sequences

e1.4.4
$$\widehat{\Omega} = \big\{ \underline{\sigma} \,|\, \underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}},\ \bigcap_{j \in J} S^{-j} P_{\sigma_j} \neq \emptyset \text{ for all } J \subset \mathbb{Z},\ |J| < \infty \big\}. \tag{1.4.4}$$

Furthermore, when possible, it will be convenient that the partition $\mathcal{P}$ be fine enough to be *S–separating*. This means that if x, x' have the same history $\underline{\sigma}$ then $x = x'$: if $\mathcal{P}$ is a separating partition it will be possible to identify the points of Ω with their histories.

We note, however, that even if $\mathcal{P}$ is S–separating it will not in general be possible to identify $\widehat{\Omega}$ with Ω: setting aside the important but trivial case in which $P_0, \ldots, P_n$ are all *closed and pairwise disjoint* and $\mathcal{P}$ is separating we must expect that $\widehat{\Omega}$ contains sequences that are not histories of points of Ω.

In fact it is quite generally possible to find $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ for which there is no $x \in \Omega$ with $\underline{\sigma} = \underline{\sigma}(x)$ but such that for any arbitrarily long prefixed time T we can find a point x whose history coincides with $\underline{\sigma}$ between $-T$ and T. The sequence $\underline{\sigma}$ is thus in $\widehat{\Omega}$ but is not the history of any point in Ω. The set of sequences in $\widehat{\Omega}$ that are not histories of points in Ω is, however, generally negligible is a sense that will be made clear in the following.

The above remark serves to clarify the interest of the space $\widehat{\Omega}$ and why in the study of the motions of (Ω, S) observed on $\mathcal{P}$ it is, in a certain sense, natural to identify motions with sequences of $\widehat{\Omega}$ rather than with trajectories of points of Ω.

In applications the requirement that $\mathcal{P}$ be separating appears to be often imposed via the requirement of *expansivity of S on* $\mathcal{P}$: S is *expansive* with respect to $\mathcal{P}$ if for all $\underline{\sigma} \in \widehat{\Omega}$ one has

e1.4.5
$$\lim_{N\to\infty} \big(\text{diam} \bigcap_{j=-N}^{N} S^{-j} P_{\sigma_j}\big) = 0. \tag{1.4.5}$$

In this case it is clear that if x, x' have the same history then they must coincide.

The above analysis is conveniently summarized into a precise definition that will be useful as a reference and as a basis for the future study of the structural properties of dynamical systems.

D1.4.2 **(1.4.2) Definition:** (Symbolic motions)
Let (Ω, S) be an invertible topological dynamical system and let $\{P_0, \ldots, P_n\} = \mathcal{P}$ be a partition of Ω into $n+1$, $n \geq 1$, Borel sets.
N1.4.1 *(1) Consider the set* [1]

e1.4.6
$$\widehat{\Omega} = \Big\{\underline{\sigma} \,|\, \underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}} \;,\quad \bigcap_{j=-N}^{N} S^{-j} P_{\sigma_j} \neq \emptyset \textit{ for all } N\Big\}. \tag{1.4.6}$$

We shall call $\widehat{\Omega}$ the set of $(\mathcal{P}, S)$–histories of the symbolic motions generated by S on Ω as seen from $\mathcal{P}$. *When $\mathcal{P}$ and S are clearly implicit in the text*

---

[1] Remark that (1.4.6) is equivalent to (1.4.4)

we shall simply call $\widehat{\Omega}$ *the* set of symbolic motions *(seen from* $\mathcal{P}$*).*
If $x \in \Omega$ *we shall call* $(\mathcal{P}, S)$–history of x *the element* $\underline{\sigma}(x)$ *of* $\widehat{\Omega}$ *such that*

e1.4.7
$$x \in S^{-j} P_{\sigma_j(x)} \qquad \text{for all } \; j \in \mathbb{Z}. \tag{1.4.7}$$

(2) If the relation $\underline{\sigma}(x) = \underline{\sigma}(x')$ *implies* $x = x'$ *we shall say that* $\mathcal{P}$ is S–separating. *If* $\mathcal{P}$ *and* S *verify (1.4.5) we shall say that* S is $\mathcal{P}$–expansive *or that it is* expansive on $\mathcal{P}$.
(3) The correspondence defined by the map $\Sigma : \Omega \to \widehat{\Omega}$ *that associates with every* $x \in \Omega$ *the sequence* $\underline{\sigma}(x) \in \widehat{\Omega}$ *will be called the* code of the symbolic dynamics of S with respect to $\mathcal{P}$.
(4) Finally when considering sets $J = (j_1, \ldots, j_q) \subset \mathbb{Z}$*, and sequences* $\underline{\sigma}_J \in \{0, \ldots, n\}^J$ *we shall often employ the notation*

e1.4.8
$$P^J_{\underline{\sigma}_J} \equiv P^{j_1 \ldots j_q}_{\sigma_1 \ldots \sigma_q} = \bigcap_{j \in J} S^{-j} P_{\sigma_j} \tag{1.4.8}$$

to denote the points whose history in J is specified by $\underline{\sigma}_J$.

Remarks: (1) If $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$, $S = \tau$ is the translation on $\{0, \ldots, n\}^{\mathbb{Z}}$, and $P_i = \{\underline{\sigma} \,|\, \sigma_0 = i\}$, $i = 0, 1, \ldots, n$, the sets $P^J_{\underline{\sigma}}$ will be denoted also $C^J_{\underline{\sigma}}$ and will be called *cylinders* of $\{0, \ldots, n\}^{\mathbb{Z}}$ with "base J and specification $\underline{\sigma}$", *i.e.*

e1.4.9
$$C^J_{\underline{\sigma}} = \{\underline{\sigma}' \,|\, \underline{\sigma}' \in \{0, \ldots, n\}^{\mathbb{Z}}, \sigma'_{j_k} = \sigma_k \quad \text{for all } \; k = 1, \ldots, q\}, \tag{1.4.9}$$

if $J = (j_1, \ldots, j_q)$ and $\underline{\sigma} = (\sigma_1, \ldots, \sigma_q) \in \{0, \ldots, n\}^J$.
(2) If $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$ one has $\widehat{\Omega} = \Sigma(\Omega)$, but this is essentially the only case; if Ω is a connected space it will be, in general, $\widehat{\Omega} \supset \Sigma(\Omega)$ and $\widehat{\Omega} \neq \Sigma(\Omega)$.
(3) It is natural to consider the set of symbolic motions $\widehat{\Omega}$ as a topological space with the topology that it inherits as a subset of $\{0, \ldots, n\}^{\mathbb{Z}}$ which, in turn, is always considered with the product topology of the discrete topologies on the factors $\{0, \ldots, n\}$. The set $\widehat{\Omega}$ is closed in $\{0, \ldots, n\}^{\mathbb{Z}}$, see problem [1.4.2].

If (Ω, S) is a topological dynamical system and $\mathcal{P}$ is an expansive partition we can "invert" the coding map Σ. Given a point $\underline{\sigma} \in \widehat{\Omega}$ we can consider the set $\mathcal{X}(\underline{\sigma}) \in \Omega$ defined by

e1.4.10
$$\mathcal{X}(\underline{\sigma}) = \bigcap_{j=-\infty}^{+\infty} S^{-j}\, \overline{P}_{\sigma_j}. \tag{1.4.10}$$

where $\overline{P}_j$ is the closure of P_j. The set $\mathcal{X}(\underline{\sigma})$ is not empty because Ω is compact.

P1.4.1 **(1.4.1) Proposition:** (Symbolic codes)
Let (Ω, S) *be an invertible topological dynamical system (*cf. *definition*

(1.2.1)); let $\mathcal{P} = \{P_0, \ldots, P_n\}$ be a topological partition of Ω on which S is expansive.
(i) The set $\mathcal{X}(\underline{\sigma})$ defined by (1.4.10) contains one and only one point so that we can define the map $X : \widehat{\Omega} \to \Omega$ by setting $X(\underline{\sigma})$ to be the unique point in $\mathcal{X}(\underline{\sigma})$ or, with a slight abuse of notation,

e1.4.11
$$\underline{\sigma} \to X(\underline{\sigma}) = \bigcap_{j=-\infty}^{+\infty} S^{-j}\, \overline{P}_{\sigma_j}. \tag{1.4.11}$$

(ii) X is a continuos map;
(iii) $\Sigma^{-1} X^{-1}(x) = x$ for every x of Ω;
(iv) X and Σ are the inverse of each other if considered as maps between Ω and $\Sigma\Omega$ and, respectively, between $\Sigma\Omega$ and Ω;
(v) $S(x) = X(\tau\Sigma(x))$, where $(\tau\underline{\sigma})_i = \sigma_{i+1}$ is the translation, by one time unit to the left, of the sequence $\underline{\sigma}$.

Remark: Note also that $\tau\widehat{\Omega} = \widehat{\Omega}$, hence $(\widehat{\Omega}, \tau)$ as well is an invertible topological dynamical system (indeed $\widehat{\Omega}$ is closed in $\{0, \ldots, n\}^{\mathbb{Z}}$ and τ is a continuous map).

Proof: It is clear that the set $\mathcal{X}(\underline{\sigma})$ defined in (1.4.10) is not empty because Ω is compact and the closed sets $\overline{P_{\sigma_j}}$ form a family with the property of non empty finite intersection, if $\underline{\sigma} \in \widehat{\Omega}$.

The S–expansivity on $\mathcal{P}$ guarantees that $\mathcal{X}(\underline{\sigma})$ consists of a single point; indeed, cf. definition (1.4.2) and (1.4.5), by assumption for all $\underline{\sigma} \in \widehat{\Omega}$:

e1.4.12
$$\text{diam}\,\Big(\bigcap_{-N}^{N} S^{-j}\overline{P_{\sigma_j}}\Big) \equiv \text{diam}\,\Big(\bigcap_{-N}^{N} S^{-j} P_{\sigma_j}\Big) \xrightarrow[N\to\infty]{} 0. \tag{1.4.12}$$

The continuity of $\underline{\sigma} \to X(\underline{\sigma})$ also follows from the S–expansivity. If $\underline{\sigma}_n \xrightarrow[n\to\infty]{} \overline{\underline{\sigma}}$ we have that for every T there exists N such that $\underline{\sigma}_n$ coincides with $\overline{\underline{\sigma}}$ from $-T$ to T, if $n > N$. This implies that eventually $X(\underline{\sigma}_n)$ is in $\cap_{j=-T}^{T} S^{-j}\,\overline{P_{\sigma_j}}$. The fact that the diameter of this set tend to 0 complete the proof of (ii).

Since the history of x is uniquely determined by x and it determines x the validity of (iii) follows. We can say that, among the various sequences that determine x via the (1.4.10), only one is in $\Sigma\Omega$. The validity of (iv) and (v) is also clear. ∎

One among the remarkable properties that a "simple" motion should have is that of spending a well determined fraction of time visiting an arbitrary "reasonable" set $E \subset \Omega$. This means that for the motion $(S^i x)_{i\in\mathbb{Z}}$ to be "simple enough" the limit

e1.4.13
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \chi_E(S^j x) = \nu_x(E) \tag{1.4.13}$$

should exist (here χ_E denotes the characteristic function of E): this is, by definition, the *frequency of visit* to E by the motion originating in x.

If E is an element of a partition $\mathcal{P} = \{P_0, \ldots, P_n\}$ of Ω, the existence of the limit (1.4.13) is directly deducible from the $(\mathcal{P}, S)$–history of x: if $E = P_j$ (1.4.13) becomes

e1.4.14
$$\nu_x(E) = \lim_{N\to\infty} N^{-1} \quad \{\text{number of labels } h \text{ between } 0 \text{ and } N-1 \text{ such that } \underline{\sigma}_h(x) = j\}. \tag{1.4.14}$$

From the history $\underline{\sigma}(x)$ of x one can obtain a more general information: one can indeed deduce the frequency with which certain groups of p symbols $\sigma_1, \ldots, \sigma_p$ appear in the history $\underline{\sigma}(x)$ of $x \in \Omega$, in sites which are translates of given sites $j_1, \ldots, j_p \in \mathbb{Z}$. Such frequency, if defined, is the limit for $N \to \infty$ of

e1.4.15
$$N^{-1}\{\text{number of labels } h \text{ between } 0 \text{ and } N-1 \text{ such that } \underline{\sigma}_{j_1+h}(x) = \sigma_1,\ \underline{\sigma}_{j_2+h}(x) = \sigma_2, \ldots, \underline{\sigma}_{j_p+h}(x) = \sigma_p\}, \tag{1.4.15}$$

or, equivalently, it is the limit

e1.4.16
$$p\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\Big|\underline{\sigma}(x)\Big) = \lim_{N\to\infty} N^{-1} \sum_{h=0}^{N-1} \chi_{P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}}(S^h x). \tag{1.4.16}$$

The number defined inside curly brackets in (1.4.15) will be called the *number of the strings homologue to* $\begin{pmatrix} J \\ \underline{\sigma} \end{pmatrix} \equiv \begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}$ *that appear in* $\underline{\sigma}(x)$ *between 0 and* $N-1$ and the value of the limit (1.4.15) will be the *frequency of appearance of the portion of history* $\begin{pmatrix} J \\ \underline{\sigma} \end{pmatrix}$ *in* $\underline{\sigma}(x)$.

Therefore we set a definition that will be useful in the following and that fixes more precisely the above notion.

D1.4.3 **(1.4.3) Definition:** (Sequences with defined frequencies)
Let $\widehat{\underline{\sigma}} \in \{0, \ldots, n\}^{\mathbb{Z}}$, $\{j_1, \ldots, j_p\} \subset \mathbb{Z}$ *and* $\sigma_1, \ldots, \sigma_p \in \{0, \ldots, n\}$.
(i) We define the number of strings homologue to $\begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}$ appearing in $\widehat{\underline{\sigma}}$ between 0 and $N-1$ *as the number of labels* h *between* 0 *and* $N-1$ *such that* $\widehat{\sigma}_{j_1+h} = \sigma_1, \ldots, \widehat{\sigma}_{j_p+h} = \sigma_p$. *We denote such number* $\mathcal{N}_N\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\Big|\widehat{\underline{\sigma}}\Big)$.
(ii) We define the frequency of appearance in $\widehat{\underline{\sigma}}$ of the string homologue to $\begin{pmatrix} j_1 \ldots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix}$ *as the limit, when it exists,*

e1.4.17
$$\lim_{N\to\infty} N^{-1}\mathcal{N}_N\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\Big|\widehat{\underline{\sigma}}\Big) = p\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\Big|\widehat{\underline{\sigma}}\Big). \tag{1.4.17}$$

(iii) A sequence $\widehat{\underline{\sigma}}$ will be said with defined frequencies *if the limit (1.4.17) exists for all $j_1, \ldots, j_p \in \mathbb{Z}$, for all $\sigma_1, \ldots, \sigma_p \in \{0, \ldots, n\}$, and for all $p = 1, 2, \ldots$.*

In the following chapter we shall show that having defined frequencies is a "not too rare" property for the sequences $\underline{\sigma}(x)$ that are $(\mathcal{P}, S)$–histories of $x \in \Omega$, for the dynamical system (Ω, S).

Appendix 1.4: *Analytically regular sets in $\mathbb{R}^n$ and $\mathbb{T}^n$*

An *analytic system of coordinates* defined on the open set $U \subset \mathbb{R}^n$ is a pair (Ω, Ξ), where Ω is an open set of $\mathbb{R}^n$ and Ξ is an invertible function from Ω to U, analytic together with its inverse.
If $x = \Xi(\underline{b})$, $\underline{b} \in \Omega$, we shall say that $\underline{b} = (b_1, \ldots, b_n)$, are the coordinates of x in (Ω, Ξ).
A surface $M \subset \mathbb{R}^n$ will be called *locally analytic in U* if $M \cap U$ can be covered by a finite family $(U_a)_{a \in A}$ of open set of $\mathbb{R}^n$ endowed with analytic systems of coordinates (Ω_a, Ξ_a) such that the points of $M \cap U_a$ have coordinates $b_i = \bar{b}_i$ for $i = 1, \ldots, s$, with $s > 0$ and $\bar{b}_1, \ldots, \bar{b}_s$ given; s is the *codimension* of M, *i.e.* $n - s$ is the dimension of M.
In the same way we can define a C^∞ *system of coordinates* (Ω, Ξ) and a *locally C^∞ surface M* of dimension k in $\mathbb{R}^n$.
A closed set $G \subset \mathbb{R}^n$ is said to be *locally analytic* if ∂G is a locally analytic surface.
A set $G \subset \mathbb{R}^n$ is said *analytically regular* in U if it can be constructed via a *finite number of operations of union and intersection* starting with sets which are locally analytic in U.
The torus $\mathbb{T}^n$ can be thought of as an analytically regular set in $\mathbb{R}^{2n}$ via the coordinate system:

eA1.4.1
$$\begin{aligned} &(\rho_1, \ldots, \rho_n, \phi_1, \ldots, \phi_n) \to (\rho_1 e^{i\phi_1}, \ldots, \rho_n e^{i\phi_n}) \to \\ &\to (\rho_1 \cos\phi_1, \rho_1 \sin\phi_1, \ldots, \rho_n \cos\phi_n, \rho_n \sin\phi_n) \end{aligned} \tag{A1.4.1}$$

by fixing $\rho_1 = \ldots = \rho_n = 1$. Then the analytically regular subsets of $\mathbb{T}^n$ can be naturally defined as the intersections between analytically regular subsets of $\mathbb{R}^{2n}$ and $\mathbb{T}^n$.
Analytically regular sets have, by definition, the property of being stable with respect to operations of union and intersection. The most relevant consequence of this is that the intersection of any pair E, F of them has the property of being Riemann–measurable with respect to the Riemann measure restricted to E or F: thus the intersection of an analytically regular set with an analytically regular surface is Riemann measurable with respect to the Riemann area measure on the surface. Replacing, in the previous definitions, "analyticity" with "C^∞–differentiability" the last property would not be true in general: see problems [1.4.12] and [1.4.13].
Therefore we shall prefer to define, inductively, *C^∞–regular of dimension k* a set G in $\mathbb{R}^n$ if:

(i) it is contained in a locally C^∞ surface M of dimesion k;
(ii) its closure $\overline{G}$ in M coincides with the closure of its interior $\overline{\text{int}\, G}$, relative to the topology on M;
(iii) its boundary ∂G in M is the union of a finite number $\widetilde{G}_i$, $i = 1, \ldots, n$ of C^∞–regular sets of dimension $k-1$, *i.e.* $\partial G = \cup_{i=1}^n \widetilde{G}_i$;
(iv) $\overline{G}/G = \cup_{j\in\mathcal{N}} \widetilde{G}_j$ where $\mathcal{N} \subset \{1\ldots, n\}$.

Finally we shall call C^∞–regular of dimension 0 a finite collection of points.

Remark that this class of sets is not closed with respect to operation of finite intersection, cf. problems [1.4.12], [1.4.13]. A C^∞–regular set G is Riemann measurable with respect to the Riemann measure restricted to the locally C^∞ surface $M \supset G$. However problem [1.4.13] shows that if G and G' are two C^∞–regular sets contained in two surfaces M and M' then $G \cap G'$ needs not to be measurable with respect to the Riemann measure on $M \cap M'$. This is the main reason why we shall try to *avoid* the use of C^∞–regular sets.

Problems for §1.4

Q1.4.1 **[1.4.1]:** (*Permutations as dynamical systems*)
Let $W = \{1, \ldots, n\}$ and let y be a permutation $i \to y(i)$ of the n elements. Let P be the partition of W into its points. Show that the (P, S)–history of every point of W is periodic and, hence, with defined frequencies. What is the meaning of the frequency in terms of the cycles that represent y?

Q1.4.2 **[1.4.2]:** ($\widehat{\Omega}$ *is closed*)
Show explicitly that $\widehat{\Omega}$ is a closed set in $\{0, \ldots, n\}^{\mathbb{Z}}$, for any choice of dynamical system (Ω, S) and of partition $\mathcal{P}$.

Q1.4.3 **[1.4.3]:** (*Non-expanding partitions*)
Find an example of a dynamical system (Ω, S) with $S \neq$ identity and such that no partition is S–separating for it.

Q1.4.4 **[1.4.4]:** Let $(\mathbb{T}, S)$ and $\mathcal{P}$ be the dynamical system and the partition of example (1.4.1). Show that $\mathcal{P}$ is S–separating if and only if the components of the vector $(\rho, 2\pi)$ are rationally independent (*i.e.* if $\rho/2\pi$ is irrational).

Q1.4.5 **[1.4.5]:** ($\widehat{\Omega} \neq \Sigma\Omega$)
Under the hypotheses of the previous problem with the components of the vector $(\rho, 2\pi)$ rationally independent find a sequence $\underline{\sigma} \in \widehat{\Omega}$ that is not the history on any point x in Ω. (*Hint:* Let $\underline{\overline{\sigma}}$ be the history of the point π. Clearly $\overline{\sigma}_0 = 1$. Consider the sequence $\underline{\sigma}$ identical to $\underline{\overline{\sigma}}$ but for $\sigma_0 = 2$.)

Q1.4.6 **[1.4.6]:** (*Arnold's cat map expansivity*)
Consider the dynamical system $(\mathbb{T}^2, S)$ defined in the example (1.2.5), cf. (1.2.13). Show that every partition $\mathcal{P}$ of $\mathbb{T}^2$ into Borel sets of the torus and with diameter small enough is such that S is expansive on $\mathcal{P}$.

Q1.4.7 **[1.4.7]:** (*Sequences with undefined frequencies*)
Find an example of a sequence $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ that does not have defined frequencies. (*Hint:* 0 followed by ten 1's followed by hundred 0's followed by thousand 0's,...)

Q1.4.8 **[1.4.8]:** Adapt this section definitions of $(\mathcal{P}, S)$–histories, frequencies of visit *etc* to the case of non-invertible dynamical systems. (*Hint:* "It suffices to replace $\mathbb{Z}$ with $\mathbb{Z}^+$".)

Q1.4.9 **[1.4.9]:** (*Interval maps and cylinders*)
Consider example (1.2.7) assuming that $[0, 1]$ can be thought of as the union $[0, 1] =$

$\cup_{\sigma=0}^{n-1}[a_\sigma, a_{\sigma+1}]$ where $0 = a_0 < a_1 < \ldots < a_n = 1$ and that S is strictly monotonic on $[a_\sigma, a_{\sigma+1}]$, for all $\sigma = 0, 1 \ldots, n-1$. Consider the partition $P_0 = [a_0, a_1), \ldots, P_{n-1} = [a_{n-1}, 1]$ of $[0,1]$. Show that the sets $P^{0,\ldots,N}_{\sigma_0,\ldots,\sigma_N}$ are intervals, for all N.

Q1.4.10 **[1.4.10]:** (*Tent map and binary expansion*)
If $S(x) = 2x$, $0 \le x \le 1/2$, and $S(x) = 2(1-x)$, $1/2 \le x \le 1$, consider $\mathcal{P} = \{[0,1/2), [1/2,1]\}$ and find the relation between the history of x, $\underline{\sigma}(x)$, on $\mathcal{P}$ and the sequence of digits of the binary development of $x : x = \sum_{k=1}^{\infty} \gamma_k 2^{-k}$, $\gamma_k = 0, 1$.

Q1.4.11 **[1.4.11]:** (*Expansive interval maps*)
Under the hypotheses of problem [1.4.9] suppose S of class C^1 in every interval $(a_\sigma, a_{\sigma+1})$, $\sigma =, 0, \ldots, n-1$, and suppose $|S'(x)| \ge \lambda > 1$, for all $x \in \cup_\sigma (a_\sigma, a_{\sigma+1})$. Show that $\mathcal{P}$ is S–separating and that S is expansive on $\mathcal{P}$ and, furthermore,

$$\mathrm{diam}\left(\bigcap_{i=0}^{N-1} S^{-i} P_{\sigma_i}\right) \le \lambda^{-N}\ , \quad \text{for all } \sigma_0, \ldots, \sigma_{N-1}\ , \quad \text{for all } N$$

Q1.4.12 **[1.4.12]:** Let $x_1, x_2, \ldots$ be an enumeration of the rationals in $[0,1]$. For every x_k consider the open interval of length 2^{-1-k} and center x_k. Show that the union A of such intervals is not Riemann measurable. (*Hint:* It has external measure ≥ 1 and internal measure $\le 1/2$.)

Q1.4.13 **[1.4.13]:** (*A C^∞ regular set with non-Riemann measurable intersection with another C^∞ regular set*)
With the notation of problem [1.4.12], consider a function $g(x)$ which is $C^\infty(\mathbb{R})$ and positive in the open interval $(-\frac{1}{2}, \frac{1}{2})$ and zero elsewhere. Set $f(x) = \sum_{k=1}^{\infty} k!^{-1} g(2^{k+1}(x - x_k))$. Show that f is C^∞ and that it is positive in A and vanishes elsewhere. Show that the set $\{x, y \,|\, y \ge f(x)\} \subset \mathbb{R}^2$ is a C^∞–regular set but its intersection with the x–axis is A which is not only not C^∞–regular but it is not even Riemann–measurable with respect to the Riemann measure on the x–axis ([Ga82], p. 337).

Bibliographical note to §1.4

The idea of studying motions by means of symbolic dynamics was essentially born with ergodic theory: it can be found, for instance, in [Mo21], [Bi35].

.

CHAPTER II

Ergodicity and ergodic points

§2.1 Quasi-periodic motions and integrability

The problem of determining which sets are visited with defined frequency by the motions of a dynamical system (Ω, S) can be satisfactorily solved in the case of particularly simple systems. For instance in the case in which $S = S_{t_0}$ and $(S_t)_{t\in\mathbb{R}}$ is a Hamiltonian flow which is analytically integrable on a region $W \subset \mathbb{R}^{2r}$ and $\Omega = W$, cf. definition (1.3.1). This means looking at motions observed at time intervals t_0. More precisely the following proposition holds.

P2.1.1 **(2.1.1) Proposition:** (Existence of frequency of visit for motions integrable by quadratures)
Consider a Hamiltonian system which is analytically integrable on a set $W \subset \mathbb{R}^{2r}$ and consider the flow $(W, (S_t)_{t\in\mathbb{R}})$ associated with it. Let $\mathcal{P} = \{P_0, \ldots, P_n\}$ be an analytically regular partition of W, i.e. such that each P_i is an analytically regular subset or the complement of an analytically regular subset of the interior int(W) *of W (*cf. *Appendix 1.4).*

Then every $x \in W$ has a $(\mathcal{P}, S_{t_0})$–history $\underline{\sigma}(x)$ with defined frequencies, for all $t_0 \in \mathbb{R}$, and hence the limits

e2.1.1
$$p\begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix} \Big| \underline{\sigma}(x)\Big) = \lim_{N\to\infty} N^{-1} \sum_{h=0}^{N-1} \chi_{P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}}(S^h_{t_0} x) \qquad (2.1.1)$$

exist for all $p > 0$, for all $\{j_1, \ldots, j_p\} \subset \mathbb{Z}$, for all $\sigma_1, \ldots, \sigma_p \in \{0, \ldots, n\}^p$ if $P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}$ is the set of points which at times $j_1 t_0, \ldots, j_p t_0$ are in the sets $P_{\sigma_1}, \ldots, P_{\sigma_p}$ respectively (see (1.4.8)).

Remarks: (1) If H is a Hamiltonian function on W which is analytically integrable there, such will also be tH, for all $t \in \mathbb{R}$: therefore S_{t_0} can be replaced by S_1 without loss of generality.
(2) As one can see from the proof it would be enough to assume that the sets P_i are Riemann measurable.

Proof: Set $S = S_1$. Note that $S^{-j_k} P_{\sigma_k}$ is an analytically regular set (being the image of such a set under an analytic map, cf. Appendix 1.4) hence Riemann measurable so that also $P^{j_1 \ldots j_p}_{\sigma_1 \ldots \sigma_p} = \cap_{k=1}^{p} S^{-j_k} P_{\sigma_k}$ is Riemann measurable.

By remark (1) it suffices to study the frequencies of appearance of strings homologue to $\begin{pmatrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}$ within the histories of points $\underline{\varphi} \in \mathbb{T}^r$ under the rotation S of $\mathbb{T}^r$ with rotation numbers $\underline{\omega} = (\omega_1, \ldots, \omega_r) \in \mathbb{R}^r$. We shall show that the rotation S of $\mathbb{T}^r$ with rotation numbers $\underline{\omega}$ is such that $(S^k \underline{\varphi})_{k \in \mathbb{Z}^+}$ visit with defined frequency every Riemann measurable set $E \subset \mathbb{T}^r$.

The proof of this statement is particularly simple in the case in which the
N2.1.1 $r+1$ numbers $(\omega_1, \ldots, \omega_r, 2\pi)$ are rationally independent.[1] This case will be treated first and the general case will be reduced to it later.

Since every analytically regular set E is Riemann-measurable, given $\varepsilon > 0$ there exist two $C^\infty(\mathbb{T}^r)$ functions f^-, f^+ such that

e2.1.2
$$\begin{aligned} & f^-(\underline{\varphi}) \le \chi_E(\underline{\varphi}) \le f^+(\underline{\varphi}), \\ & \int_{\mathbb{T}^r} \big(f^+(\underline{\varphi}) - f^-(\underline{\varphi})\big) \frac{d\underline{\varphi}}{(2\pi)^r} < \varepsilon. \end{aligned} \tag{2.1.2}$$

see proplem [2.1.9]. This implies that it will suffice to show that if $f \in C^\infty(\mathbb{T}^r)$ and $(\omega_1, \ldots, \omega_r, 2\pi)$ are rationally independent one has, for all $\underline{\varphi} \in \mathbb{T}^r$,

e2.1.3
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^j \underline{\varphi}) = \int_{\mathbb{T}^r} f(\underline{\psi}) \frac{d\underline{\psi}}{(2\pi)^r}, \tag{2.1.3}$$

to conclude from (2.1.2), that

e2.1.4
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \chi_E(S^j \underline{\varphi}) = \int_{\mathbb{T}^r} \chi_E(\underline{\psi}) \frac{d\underline{\psi}}{(2\pi)^r}. \tag{2.1.4}$$

In fact (2.1.3) can be seen as follows. If $\hat{f}_{\underline{\nu}}$ are the Fourier coefficients of

[1] A set H of real numbers is said to consist of rationally independent numbers if the relation $\sum_{\omega \in H} n_\omega \omega = 0$, with n_ω integer and $n_\omega = 0$ but for a finite subset of H, implies $n_\omega \equiv 0$.

$f \in C^\infty(\mathbb{T}^r)$, $\underline{\nu} \in \mathbb{Z}^r$, we shall have

e2.1.5

$$N^{-1}\sum_{j=0}^{N-1} f(\underline{\varphi}+j\underline{\omega}) = \sum_{\underline{\nu}\in\mathbb{Z}^r} \hat f_{\underline{\nu}} e^{i\underline{\nu}\cdot\underline{\varphi}} \cdot \Big\{N^{-1}\sum_{j=0}^{N-1} e^{i\underline{\nu}\cdot\underline{\omega}j}\Big\} =$$
$$= f_{\underline{0}} + \sum_{\substack{\underline{\nu}\in\mathbb{Z}^r\\ \underline{\nu}\neq\underline{0}}} \hat f_{\underline{\nu}} e^{i\underline{\nu}\cdot\underline{\varphi}} \cdot \Big\{N^{-1}\frac{1-e^{i\underline{\omega}\cdot\underline{\nu}N}}{1-e^{i\underline{\omega}\cdot\underline{\nu}}}\Big\} \xrightarrow[N\to\infty]{} \qquad (2.1.5)$$
$$\xrightarrow[N\to\infty]{} f_{\underline{0}} \equiv \int_{\mathbb{T}^r} f(\underline{\psi})\frac{d\underline{\psi}}{(2\pi)^r}.$$

since $\widehat{f}_{\underline{\nu}}$ decreases more rapidly than any power in $|\underline{\nu}|$ for $|\underline{\nu}| \to \infty$ allow us to exchange the limit on N with the sum on $\underline{\nu}$. The absolute value of the term in curly brackets in the first line of (2.1.5) is bounded above by 1 for every N (being an average of N numbers each of absolute value 1), and it tends to zero as $N \to \infty$ for each $\underline{\nu} \neq \underline{0}$ because $\underline{\omega}\cdot\underline{\nu}$ can never be multiple of 2π by the assumed rational independence of $(\omega_1, \ldots, \omega_r, 2\pi)$.

More generally let $\mathcal{M}$ denote the set of the vectors $\underline{\nu} \in \mathbb{Z}^r$ such that there exists an integer m for which

e2.1.6

$$\underline{\omega}\cdot\underline{\nu} + 2\pi m = 0; \qquad (2.1.6)$$

then the above argument implies

e2.1.7

$$N^{-1}\sum_{j=0}^{N-1} f(\underline{\varphi}+j\underline{\omega}) \xrightarrow[N\to\infty]{} \sum_{\underline{\nu}\in\mathcal{M}} \hat f_{\underline{\nu}}\, e^{i\underline{\nu}\cdot\underline{\varphi}}, \qquad (2.1.7)$$

thus completing the proof of the proposition. ■

In the course of the proof we have also obtained, see (2.1.2) and (2.1.3), the following corollary on functions f on $\mathbb{T}^r$ which are Riemann integrable in the sense that given any $\varepsilon > 0$ there exist $f_-, f_+ \in C^\infty(\mathbb{T}^r)$ such that $f_-(\underline{\varphi}) \le f(\underline{\varphi}) \le f_+(\underline{\varphi})$ for all $\underline{\varphi}$ and $\int (f_+(\underline{\varphi}) - f_-(\underline{\varphi}))d\underline{\varphi}/(2\pi)^r < \varepsilon$.

C2.1.1 **(2.1.1) Corollary:** (Average of Riemann integrable functions)
If S is the rotation map of $\mathbb{T}^r$ with rotation vector $\underline{\omega}$ such that $(\omega_1, \ldots, \omega_r, 2\pi)$ are rationally independent and f is a Riemann integrable function on $\mathbb{T}^r$ then for all choices of $\underline{\varphi}$ one has

e2.1.8

$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} f(S^j\underline{\varphi}) = \int_{\mathbb{T}^r} f(\underline{\psi})\frac{d\underline{\psi}}{(2\pi)^r}. \qquad (2.1.8)$$

Remark: Note that a Riemann integrable function in the above sense, see (2.1.2), is *necessarily* bounded.

The preceding proof can be refined to obtain the following proposition.

P2.1.2 **(2.1.2) Proposition:** (Rationally dependent rotations and foliations of the torus)
Let S be a rotation of $\mathbb{T}^r$ with rotation numbers $\underline{\omega} = (\omega_1, \ldots, \omega_r)$. Then there is an integer $0 \le g \le r$ such that $\mathbb{T}^r$ can be thought of as a union of a $(r-g)$–parameters family $\mathcal{T}_{\underline{\varphi}'}$, with $\underline{\varphi}' \in \mathbb{T}^{r-g}$, of sets homeomorphic to the torus $\mathbb{T}^g$ on which the map S acts as a rotation by a vector $(\omega'_1, \ldots, \omega'_g)$ with rationally independent components and such that either $(\omega'_1, \ldots, \omega'_g, 2\pi)$ are rationally independent too or $\omega'_1 = 2\pi/n$ for some integer n.
If $(\omega'_1, \ldots, \omega'_g, 2\pi)$ are rationally independent and E is a analytically regular set, one has

e2.1.9
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \chi_E(S^j \underline{\varphi}) = \int_{E \cap \mathcal{T}_{\underline{\varphi}'}} \frac{d\underline{\psi}}{(2\pi)^g}, \tag{2.1.9}$$

where $\underline{\varphi} \in \mathcal{T}_{\underline{\varphi}'}$.

Remark: The following proof relies only on the property that the intersection between an analytically regular set and an analytically regular surface is a Riemann measurable set: which allows us to reduce the proof to the case treated in proposition (2.1.1): see Appendix 1.4 and problem [1.4.13]. Rationally dependent rotations are also called *resonant rotations*.

Proof: Let $\{\widetilde{\omega}_1, \ldots, \widetilde{\omega}_g\}$, $g \le r$, be a subset of $\{\omega_1, \ldots, \omega_r\}$ consisting of g rationally independent numbers and suppose that it is a maximal subset among those sharing this property. Then there exist g numbers $(\overline{\omega}_1, \ldots, \overline{\omega}_g)$ with $\overline{\omega}_j = \widetilde{\omega}_j / M$, M integer, $j = 1, \ldots, g$, and a $r \times g$ matrix J with *integer entries* such that

e2.1.10
$$\omega_j = \sum_{k=1}^{g} J_{jk}\, \overline{\omega}_k, \qquad j = 1, \ldots, r. \tag{2.1.10}$$

Furthermore the matrix J contains a $g \times g$ diagonal submatrix, that we can suppose to be the first (we just imagine that $\omega_1, \ldots, \omega_r$ are ordered so that $\widetilde{\omega}_1 = \omega_1, \ldots, \widetilde{\omega}_g = \omega_g$), given by $J_{jk} = M\delta_{jk}$, $j, k = 1, \ldots g$.
It is therefore possible to complete the matrix J into a square $r \times r$ matrix with integer coefficients so that, still calling J this new matrix, one has $\det J \ne 0$. This is possible in infinitely many ways and we select one arbitrarily (for instance, in the case that $J_{jk} = M\delta_{jk}$, $j, k = 1, \ldots, g$, we can set $J_{jk} = 0$ if $k > g$ and $k \ne j$, $J_{jj} = 1$ if $j > g$).
The map

e2.1.11
$$\underline{\varphi} = J\underline{\varphi}' \mod 2\pi \tag{2.1.11}$$

transforms $\mathbb{T}^r$ into itself and it is of class C^∞: it is not one–to–one but to each $\underline{\varphi}$ there correspond $|\det J|$ values of $\underline{\varphi}'$ such that $J\underline{\varphi}' = \underline{\varphi} \mod 2\pi$.
If $E \subset \mathbb{T}^r$ is an analytically regular set, the function

e2.1.12
$$\chi(\underline{\varphi}') = \chi_E(J\underline{\varphi}') \qquad \underline{\varphi}' \in \mathbb{T}^r \tag{2.1.12}$$

is the characteristic function of a set $E' = J^{-1}E$ which is analytically regular in $\mathbb{T}^r$ and, if we set $S'\underline{\varphi}' = \underline{\varphi}' + \underline{\overline{\omega}}$ mod 2π, with $\underline{\overline{\omega}} = (\overline{\omega}_1, \ldots, \overline{\omega}_g, 0, \ldots 0)$ we get, by construction,

e2.1.13
$$J(S'\underline{\varphi}') = J(\underline{\varphi}' + \underline{\overline{\omega}}) = (J\underline{\varphi}') + \underline{\omega} = S(J\underline{\varphi}'). \tag{2.1.13}$$

This shows that the set $\mathcal{T}_{\varphi'_{g+1},\ldots,\varphi'_r}$ of the points $\underline{\varphi} \in \mathbb{T}^r$ of the form of (2.1.11) with $\underline{\varphi}' = (\varphi'_1, \ldots, \varphi'_g, \varphi'_{g+1}, \ldots, \varphi'_r)$ and $\varphi'_{g+1}, \ldots, \varphi'_r$ fixed is an invariant torus of dimension g. By the analytical regularity of E' also the
N2.1.2 intersection $E' \cap \mathcal{T}_{\varphi'_{g+1},\ldots,\varphi'_r}$ is analytically regular,[2] so that the restriction of S' on $\mathcal{T}_{\varphi'_{g+1},\ldots,\varphi'_r}$ acts as a rotation with rationally independent rotation numbers $\overline{\omega}_1, \ldots, \overline{\omega}_g$. If $\underline{\overline{\omega}} = (\overline{\omega}_1, \ldots, \overline{\omega}_g, 2\pi)$ are rationally independent (2.1.9) follows from proposition (2.1.1). See problem [2.1.8] for a more detailed anaysis of what happens when $\underline{\overline{\omega}} = (\overline{\omega}_1, \ldots, \overline{\omega}_g, 2\pi)$ are not rationally independent. ■

Problems for §2.1

Q2.1.1 **[2.1.1]:** (*One-dimensional Hamiltonian motions*)
Consider a Hamiltonian system with Hamiltonian $H(p,q) = p^2/2 + V(q)$, with $V \in C^\infty(\mathbb{R})$, $V(0) = 0$, $V'(q) > 0$ for $q > 0$, $V(q) = V(-q)$, $\lim_{q\to\infty} V(q) = +\infty$. Show that its motions of energy $E > 0$ are periodic with period $T(E) = 2\pi/\omega(E)$ where

$$\omega(E) = \pi\Big(\int_{-x(E)}^{x(E)} \frac{dq}{\sqrt{2(E - V(q))}}\Big)^{-1}, \quad V(x(E)) = E, \quad x(E) > 0.$$

Q2.1.2 **[2.1.2]:** (*Energy–frequency dispersion relation analyticity*)
Consider, in the context of problem [2.1.1], the case $V(q) = aq^4$, $a > 0$, and show that

$$\omega(E) = c(4aE)^{1/4} \qquad c = \Big(\frac{1}{\pi}\int_{-1}^{1} \frac{d\xi}{\sqrt{1 - \xi^4}}\Big)^{-1}$$

Q2.1.3 **[2.1.3]:** (*Analyticity of the energy–frequency dispersion relation*)
In the context of problem [2.1.1] show that if V is an even polynomial in q or, more generally, an even analytic function in q, the function $E \to \omega(E)$ is analytic for $E > 0$.

Q2.1.4 **[2.1.4]:** If $V(q) = aq^4$, $a > 0$, the set of the data (p,q) that generate, in the case of problem [2.1.1], a periodic motion with period rational with respect to 2π has measure zero but it is dense.

Q2.1.5 **[2.1.5]:** (*Anisochrony of one-dimensional motions*)
By making use of the result in problem [2.1.3] show that the same conclusion of the preceding problem holds in the general case contemplated in problem [2.1.3] unless $V(q)$ is proportional to q^2.

Q2.1.6 **[2.1.6]:** Show that in general, under the assumptions of problem [2.1.1], $E \to \omega(E)$ is of class C^∞ for $E > 0$ and, if $\omega'(E) \neq 0$ for $E \in (a,b)$, then the motions with energy in

[2] If E was assumed to be only C^∞–regular it would not be, in general, true that such intersection remains C^∞–regular or even Riemann measurable, see problems [1.4.12], [1.4.13].

(a,b) are such that $\omega(E)$ is incommensurable with 2π with the exception of a denumerable infinity of values of E dense in (a,b).

Q2.1.7 **[2.1.7]:** (*Analytic and canonical integrability in dimension one*)
Show that the system of problem [2.1.1] is canonically integrable in the region $W = \{p,q|H(p,q) > 0\}$. If V is an analytic function then the system is analytically and canonically integrable.

Q2.1.8 **[2.1.8]:** Refine the statement of proposition (2.1.2) by showing that if the g–dimensional vector $\underline{\overline{\omega}}$ of rationally independent components of $\underline{\omega}$ with maximal g is such that $(\underline{\overline{\omega}}, 2\pi)$ does not have rationally independent components then for some $n > 0$ the torus $\mathcal{T}_{\varphi'_{g+1},\ldots,\varphi'_r}$ can be further decomposed into a union of sets of dimension $g-1$ each consisting of n tori which are cyclically permuted by the rotation $\underline{\omega}$. (*Hint:* One can reduce the problem to the case where $\overline{\omega}_1 = 2\pi/n$ for some integer $n > 0$.)

Q2.1.9 **[2.1.9]:** (*Riemann integrabilty of sets and smooth approximations*)
Given a Riemann measurable set E show that there exist two C^∞ functions f^+ and f^- such that (2.1.2) holds. (*Hint:* Being E Riemann measurable one can find two set E^+ and E^- formed by rectangles such that $E^- \subset E \subset E^+$ and the areas of E^+ and E^- differs by less than $\varepsilon/2$. One can now enlarge E^+ obtaining $\tilde{E}^+$ and restrict E^- obtaining $\tilde{E}^-$ such that $\partial\tilde{E}^\pm \cap \partial E = \emptyset$ and the areas of $\tilde{E}^+$ and $\tilde{E}^-$ differs by less than ε. By regularazing the characteristic functions of $\tilde{E}^+$ and $\tilde{E}^-$ one can construct f^+ and f^-).

Bibliographical note to §2.1

Propositions (2.1.1) and (2.1.2) are classical results attributed to Jacobi (see [AA68]).

§2.2 Ergodic properties of quasi periodic motions. Ergodic sequences and measures.

The quasi-periodic motions examined in the preceding section have further remarkable properties.

For example if S is a rotation of $\mathbb{T}^r$ with rotation numbers $\underline{\omega} = (\omega_1,\ldots,\omega_r)$ such that $(\omega_1,\ldots,\omega_r,2\pi)$ are rationally independent parameters then, given an analytically regular or just Riemann–measurable set, the frequency of visit $\nu_{\underline{\varphi}}(E)$ by the trajectory of $\underline{\varphi} \in \mathbb{T}^r$ to E is, cf. equations (1.4.13),(2.1.4),

e2.2.1
$$\nu_{\underline{\varphi}}(E) \stackrel{def}{=} \lim_{N\to\infty} N^{-1}\sum_{k=0}^{N-1}\chi_E(\underline{\varphi}+k\underline{\omega}) = \int_{\mathbb{T}^r}\chi_E(\underline{\psi})\frac{d\underline{\psi}}{(2\pi)^r} \tag{2.2.1}$$

for all $\underline{\varphi} \in \mathbb{T}^r$, hence, by dominated convergence,

e2.2.2
$$\begin{aligned}&\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1}\nu_{\underline{\varphi}}(E\cap S^{-j}F) =\\ &= \lim_{N\to\infty}\int_{\mathbb{T}^r} N^{-1}\sum_{j=0}^{N-1}\chi_E(\underline{\psi})\chi_F(S^j\underline{\psi})\frac{d\underline{\psi}}{(2\pi)^r} = \nu_{\underline{\varphi}}(E)\,\nu_{\underline{\varphi}}(F)\end{aligned} \tag{2.2.2}$$

for every pair of analytically regular sets.

The latter relation can be read in colorful language by saying "*the visit to E followed, after a long time j, by the visit to F are events that in average happen independently in the history of $\underline{\varphi}$*", or "*knowing that the trajectory of $\underline{\varphi}$ has visited E does not allows us to derive informations on the sequence of times j in which the trajectory of $\underline{\varphi}$ will visit F with the exception, at most, of a sequence of times with zero density in time*". The reader should spend some time meditating on why such an interpretation of (2.2.2) is reasonable.

Property (2.2.2) is general within the class of motions studied in §2.1 and it is not just characteristic of the case in which $(\omega_1, \ldots, \omega_r, 2\pi)$ are rationally independent.

In general, under the hypothesis in which proposition (2.1.2) holds, *i.e.* for the system $(\mathbb{T}^r, S)$ with S an arbitrary rotation with rotation numbers $\underline{\omega} \in \mathbb{R}^r$ one has

e2.2.3
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \nu_{\underline{\varphi}}(E \cap S^{-j} F) = \nu_{\underline{\varphi}}(E) \nu_{\underline{\varphi}}(F) \tag{2.2.3}$$

for all analytically regular $E, F \subset \mathbb{T}^r$. In fact the function $\nu_{\underline{\varphi}}$ verifies (2.2.3) although *the last equality in (2.2.1) is not valid* when $(\underline{\omega}, 2\pi)$ are not rationally independent.

This follows from proposition (2.1.2). If $(\underline{\omega}, 2\pi)$ are rationally independent the relation (2.2.3) follows directly from (2.2.1) and (2.2.2). Otherwise the property (2.2.3) reduces again to the relations in (2.2.1) and (2.2.2) after remarking that the case with $(\underline{\omega}, 2\pi)$ not rationally independent can be reduced to the case of rational independence via the arguments seen in the proof of (2.1.2), see also problem [2.1.8].

A different way of expressing the above considerations is in the following proposition.

P2.2.1 **(2.2.1) Proposition:** (Quasi-periodic frequencies)
Let S be the rotation on $\mathbb{T}^r$ with rotation numbers $\underline{\omega}$. Let $\mathcal{P} = \{P_0, \ldots, P_n\}$ be an analytically regular partition of $\mathbb{T}^r$ and let $p\Big(\begin{smallmatrix} j_1 \ldots j_p i_1+k \ldots i_q+k \\ \sigma_1 \ldots \sigma_p \ \sigma'_1 \ \ldots \ \ \sigma'_q \end{smallmatrix} \Big| \underline{\sigma}(\underline{\varphi})\Big)$ be the frequency of visit of the points in the trajectory of $\underline{\varphi}$ to the set

e2.2.4
$$P^{j_1\ldots j_p\, i_1+k\ldots i_q+k}_{\sigma_1\ldots\sigma_p\ \sigma'_1\ \ldots\ \sigma'_q} = \bigcap_{m=1}^{p} S^{-j_m} P_{\sigma_m} \cap \bigcap_{n=1}^{q} S^{-(i_n+k)} P_{\sigma'_n}. \tag{2.2.4}$$

Then the histories $\underline{\sigma}(\underline{\varphi})$ of $\underline{\varphi} \in \mathbb{T}^r$, with respect to $\mathcal{P}$, have the following property (cf. *(1.4.16), (1.4.17)):*

e2.2.5
$$\begin{aligned} &\lim_{N\to\infty} N^{-1} \sum_{k=0}^{N-1} p\Big(\begin{matrix} j_1 \ldots j_p & i_1+k \ldots i_q+k \\ \sigma_1 \ldots \sigma_p & \sigma'_1 \ldots \sigma'_q \end{matrix} \Big| \underline{\sigma}(\underline{\varphi})\Big) = \\ &\qquad = p\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix} \Big| \underline{\sigma}(\underline{\varphi})\Big) p\Big(\begin{matrix} i_1 \ldots i_q \\ \sigma_1 \ldots \sigma'_q \end{matrix} \Big| \underline{\sigma}(\underline{\varphi})\Big) \end{aligned} \tag{2.2.5}$$

for all $p,q = 1,2,\ldots,$ *and for all* $\sigma_1,\ldots,\sigma_p,\ \sigma'_1,\ldots,\sigma'_p \in \{0,1,\ldots,n\}$, $\{j_1,\ldots,j_p\},\ \{i_1,\ldots,i_q\} \subset \mathbb{Z}$.

Proof: This proposition is just the statement (2.2.3) in which $E = P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}$ and $F = P^{i_1\ldots i_q}_{\sigma'_1\ldots\sigma'_q}$ because $S^{-k}P^{i_1\ldots i_q}_{\sigma'_1\ldots\sigma'_q} = P^{i_1+k\ldots i_q+k}_{\sigma'_1\ \ldots\ \sigma'_q}$. ■

Property (2.2.5) is, as we shall see, really remarkable: it tells us that the frequency of appearance in $\underline{\sigma}(\varphi)$ of two strings of history homologue to $\begin{pmatrix} j_1\ldots j_p \\ \sigma_1\ldots\sigma_p\end{pmatrix}$ and $\begin{pmatrix} i_1\ldots i_q \\ \sigma'_1\ldots\sigma'_q\end{pmatrix}$ is, "in average", equal to the product of the frequencies of appearance of each string. In other words strings of history homologue to two given strings of history appear, in the average, as independently distributed in $\underline{\sigma}(\varphi)$.

It is important to set up a definition for the purpose of formalizing and generalizing the above properties.

D2.2.1 **(2.2.1) Definition:** (Ergodic and mixing sequences)
A sequence $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$ *is said* ergodic *if it has defined frequencies and if*

e2.2.6
$$\lim_{N\to\infty} N^{-1}\sum_{k=0}^{N-1} p\begin{pmatrix} j_1\ldots j_p & i_1+k\ldots i_q+k \\ \sigma'_1\ldots\sigma'_p & \sigma''_q\ldots\sigma''_q \end{pmatrix}\!\Big|\underline{\sigma}\Big) = \tag{2.2.6}$$
$$= p\left(\begin{matrix} j_1\ldots j_p \\ \sigma'_1\ldots\sigma'_p\end{matrix}\Big|\underline{\sigma}\right) p\left(\begin{matrix} i_1\ldots i_q \\ \sigma''_1\ldots\sigma''_q\end{matrix}\Big|\underline{\sigma}\right)$$

for all $p,\ q = 1,2,\ldots,$ *and for all* $\sigma'_1,\ldots,\sigma'_p,\ \sigma''_1,\ldots,\sigma''_q \in \{0,\ldots,n\}$, $\{j_1,\ldots,j_p\},\ \{i_1,\ldots,i_q\} \subset \mathbb{Z}$.
We shall say that $\underline{\sigma}$ *is* mixing *if*

e2.2.7
$$\lim_{k\to\infty} p\Big(\begin{matrix} j_1\ldots j_p\ i_1+k\ldots i_q+k \\ \sigma'_1\ldots\sigma'_p\quad \sigma''_1\ldots\quad \sigma''_q\end{matrix}\,\Big|\underline{\sigma}\Big) = \tag{2.2.7}$$
$$= p\Big(\begin{matrix} j_p\ldots j_p \\ \sigma'_1\ldots\sigma'_p\end{matrix}\,\Big|\underline{\sigma}\Big)\, p\Big(\begin{matrix} i_1 \quad\quad i_q \\ \sigma''_1\ldots\sigma''_q\end{matrix}\,\Big|\underline{\sigma}\Big)$$

for all possible choices of the labels.

Remark: Obviously every mixing sequence $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$ is also ergodic.

For a deeper understanding of the meaning of the names used above one should keep in mind the statement of the next proposition (2.2.2) that, among other consequences, shows that the existence of the frequencies of visit of the $(\mathcal{P},S)$–histories of the points of a dynamical system is a general fact, much more so than what one could be led to expect at first sight.

P2.2.2 **(2.2.2) Proposition:** (Birkhoff theorem)
Let (Ω,S,μ) *be a discrete metric invertible dynamical system and let* $\mathcal{B}$ *be*

the σ–algebra on which μ is defined. If f is a μ–measurable function the limit

e2.2.8
$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} f(S^j x) = \overline{f}(x) \tag{2.2.8}$$

exists μ–almost everywhere and if $f \in L_1(\mu)$ the limit is also reached in $L_1(\mu)$.

Remarks: (1) This well known abstract theorem is proved in Appendix (2.2).
(2) The theorem implies that if E_1, $E_2, \ldots$ is a denumerable family of measurable sets the limits

e2.2.9
$$\lim_{N\to\infty} N^{-1}\sum_{k=0}^{N-1} \chi_{E_j}(S^k x) = \overline{\chi}_{E_j}(x), \qquad j = 1, 2, \ldots, \tag{2.2.9}$$

simultaneously exist μ–almost everywhere. Hence, if $\mathcal{P} = \{P_0, \ldots, P_n\}$ is a partition of Ω into measurable sets the limits

e2.2.10
$$\lim_{N\to\infty} N^{-1}\sum_{k=0}^{N-1} \chi_{P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}}(S^k x) = p\left(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\middle|\underline{\sigma}(\underline{x})\right) \tag{2.2.10}$$

exist for all possible choices of the labels and for μ–almost all x. This means that the $(\mathcal{P}, S)$–histories of μ–almost all points of Ω have defined frequencies.
(3) Let us, however, stress the deep difference between this statement and that of proposition (2.1.1) (or of proposition (2.1.2)): the latter shows, in a very particular case, that *all* points have histories with a defined frequency. Proposition (2.2.2) above, instead, shows us that in a very general case μ–almost all points have histories with defined frequencies.
In general, given a dynamical system, it is *very difficult* (and very interesting) to find points, if any exists, that have singular histories, *i.e.* that do not have defined frequencies with respect to a given partition $\mathcal{P}$: if the points are randomly chosen with distribution μ *and* μ *is* S*-invariant* proposition (2.2.2) says that this is impossible.
(4) The proof of proposition (2.2.2) is rather simple and abstract and this should help us to understand why it ends up by being not so useful in the analysis of concrete problems. It is nevertheless obviously important and very useful in theoretical questions of abstract ergodic theory and in the analysis of structural problems.
(5) From proposition (2.2.2) one can easily understand why sequences $\underline{\sigma}$ verifying (2.2.6) are called *ergodic*. Let indeed (Ω, S, μ) be a metric dynamical system for which the ergodic hypothesis of Section §1.1 holds. This means that the system is such that for all $E \in \mathcal{B}$

e2.2.11
$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} \chi_E(S^j x) = \mu(E) \qquad \mu - \text{almost everywhere}, \tag{2.2.11}$$

i.e. it is such that μ–almost all points $x \in \Omega$ visit a given measurable set for a time which is asymptotically proportional to its measure (in Section §1.1 such a hypothesis was formulated in the particular case in which Ω is the phase–space of given energy, S is a Hamiltonian evolution on Ω up to time 1, and μ is the Liouville measure on Ω).

Then, by the definition of $(\mathcal{P}, S)$–history with respect to a given partition $\mathcal{P} = \{P_0, \ldots, P_n\}$ of Ω into μ–measurable sets, we shall have that, μ–almost everywhere, $x \in \Omega$ has a history $\underline{\sigma}(x)$ with frequencies defined and independent of x, given by

e2.2.12
$$p\Big(\begin{matrix} j_1 \ldots j_p \\ \sigma_1 \ldots \sigma_p \end{matrix}\Big|\underline{\sigma}(x)\Big) = \mu\Big(P^{j_1\ldots j_p}_{\sigma_1\ldots\sigma_p}\Big) \tag{2.2.12}$$

for every possible choice of the labels.

Furthermore (2.2.11) implies also, for all $E, F \in \mathcal{B}$,

e2.2.13
$$N^{-1}\sum_{j=0}^{N-1} \mu(F \cap S^{-j}E) = \int_\Omega \chi_F(x) \cdot N^{-1} \sum_{j=0}^{N-1} \chi_E(S^j x)\mu(dx) \xrightarrow[N\to\infty]{} \xrightarrow[N\to\infty]{} \mu(E)\mu(F), \tag{2.2.13}$$

hence the $(\mathcal{P}, S)$–histories have frequencies (2.2.12) that verify (2.2.6), μ–almost everywhere: *i.e.* they are ergodic sequences in the sense of the definition (2.2.1).

(6) It is also interesting to remark that not only (2.2.13) is a consequence of (2.2.11) but, vice versa, it implies it. This follows from proposition (2.2.2) which says that the limit

e2.2.14
$$\overline{\chi}(x) = \lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \chi_E(S^j x) \tag{2.2.14}$$

exists μ–almost everywhere and defines an S–invariant function: *i.e.* $\overline{\chi}(x) = \overline{\chi}(Sx)$ holds μ–almost everywhere.

Suppose that $\overline{\chi}$ was not almost everywhere constant; then there would be an $\alpha \in \mathbb{R}$ such that the sets $E = \{x|\overline{\chi}(x) \le \alpha\}$ and $E^c = \{x|\overline{\chi}(x) > \alpha\}$ have respective measures $\mu(E)$ and $\mu(E^c)$ with $0 < \mu(E)$, $\mu(E^c) < 1$. The latter sets are such that $E = SE$, up to a set of zero measure; hence $E \cap S^j E = E \mod 0$ and

e2.2.15
$$\mu(E) \equiv N^{-1} \sum_{j=0}^{N-1} \mu(E \cap S^{-j}E) \xrightarrow[N\to\infty]{} \mu(E)^2, \tag{2.2.15}$$

if (2.2.13) holds. Therefore $\mu(E) = 0$ or $\mu(E) = 1$: the contradiction implies that $\overline{\chi}$ is constant μ–almost everywhere. Then integrating both members of (2.2.14) we deduce (2.2.11) from the invariance of $\overline{\chi}$ and μ.

(7) From the proof it will follow that the same theorem holds if instead of a metric invertible dynamical system we considered a metric invertible

dynamical system mod 0, *i.e.* if (Ω, S, μ) is replaced by $(\Omega \setminus N, S, \mu)$ with N a set with zero μ–measure.

It is useful to set one more definition, made natural by the previous observations, appending to it a comment to illustrate the adjective *mixing* introduced in definition (2.2.1).

*D*2.2.2 **(2.2.2) Definition:** (Ergodic and mixing dynamical system)
If (Ω, S, μ) is a metric invertible dynamical system we shall say that it is ergodic *if, for all E, $F \in \mathcal{B}$, one has*

*e*2.2.16
$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \mu(E \cap S^{-j}F) = \mu(E)\mu(F). \tag{2.2.16}$$

If, instead, for all E, $F \in \mathcal{B}$, one has

*e*2.2.17
$$\lim_{j\to\infty} \mu(E \cap S^{-j}F) = \mu(E)\mu(F), \tag{2.2.17}$$

we shall say that (Ω, S, μ) is mixing*.*
*If (Ω, S) is a topological dynamical system we shall denote by $\mathcal{M}_e(\Omega, S)$ (or by $\mathcal{M}_m(\Omega, S)$) the set of the S–invariant probability measures on the Borel sets $\mathcal{B}(\Omega)$ of Ω and such that (Ω, S, μ) is an ergodic (or mixing) system: its elements will be called S–*ergodic measures *(or S–*mixing measures*) on Ω. More generally $\mathcal{M}(\Omega, S)$ will be the set of all S–invariant Borel measures and $\mathcal{M}^0(\Omega)$ will be the set of all Borel measures (S–invariant or not).*

Remarks: (1) Remarks (5) and (6) to the (2.2.2) show that (2.2.16) is equivalent to (2.2.11) so that the present definition of ergodicity coincide with the definition given in Section §1.1.
(2) $\mathcal{M}_m(\Omega, S) \subset \mathcal{M}_e(\Omega, S)$.
(3) The mixing property owes its name to the fact that if j is large the set $S^{-j}F$ is, in mixing systems, very spread out in Ω and uniformly so if the size of sets E is measured by $\mu(E)$. This means that the fraction $\mu(E \cap S^{-j}F)/\mu(F)$ of its points that is found inside another set E (or
*N*2.2.1 inside a finite number of such sets) is proportional to the measure of E.[1] Every $F \in \mathcal{B}$ is mixed around in Ω in a uniform way after a time j large enough, of course depending on F, under the action of the map S.
(4) The results in proposition (2.1.1) and corollary (2.1.1) prove that the quasi-periodic rotations of the tori with rationally independent components of the vector $(\omega_1, \ldots, \omega_n, 2\pi)$ are ergodic.

In this book we mostly consider maps: however it is convenient to have in mind the definitions of ergodicity and mixing that can be given in the case of flows, *i.e.* dynamical systems with continuous time.

[1] Naturally the magnitude which j must reach in order that this happens within a prefixed approximation depends both on E and F.

D2.2.3 **(2.2.3) Definition:** (Ergodic and mixing flows)
Let (Ω, S_t, μ) *be a metric flow. Then we shall say that it is* ergodic *if, for all* $E,\ F \in \mathcal{B}(\Omega)$*, one has*

e2.2.18
$$\lim_{T\to\infty} T^{-1}\int_0^T \mu(E\cap S_{-t}F) = \mu(E)\mu(F). \tag{2.2.18}$$

If, moreover, for all $E,\ F \in \mathcal{B}(\Omega)$*, one has*

e2.2.19
$$\lim_{t\to\infty} \mu(E\cap S^{-t}F) = \mu(E)\mu(F), \tag{2.2.19}$$

we shall say that (Ω, S_t, μ) *is* mixing.

For flows one can also prove the analogue of Birkhoff's theorem.

P2.2.3 **(2.2.3) Proposition:** (Birkhoff's theorem for flows)
Let (Ω, S_t, μ) *be a metric flow. If* f *is a* μ*–measurable function the limit*

e2.2.20
$$\lim_{T\to\infty} T^{-1}\int_0^T f(S_t x) = \overline{f}(x) \tag{2.2.20}$$

exists μ*–almost everywhere and if* $f \in L_1(\mu)$ *the limit is also reached in* $L_1(\mu)$*.*

Remark: The proof of the latter proposition is very close to the corresponding one of proposition (2.2.2) given in the following appendix.

A general proposition providing an equivalent definition of ergodicity is

P2.2.4 **(2.2.4) Proposition:** *If* (Ω, S, μ) *and* (Ω, S_t, μ) *are, respectively, a metric dynamical system or a metric flow then a necessary and sufficient condition for ergodicity is that, for any given* $f \in L_1(\mu)$ *the following limit relations hold* μ*–almost everywhere*

$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} f(S^j x) = \int \mu(dy)\, f(y), \qquad \text{map case,}$$
$$\lim_{T\to\infty} T^{-1}\int_0^T f(S_t x)dt = \int \mu(dy)\, f(y), \qquad \text{flow case.}$$

Proof: Consider the case of a map S. Note that Birkhoff's theorem implies that $\int f(y)\mu(dy) \equiv \int \overline{f}(y)\mu(dy)$. If f is ergodic in the sense of definition (2.2.2) the average $\overline{f}(x)$ of a function must be constant almost everywhere otherwise there would be a set E with $0 < \mu(E) < 1$ and a number a such that E is the set of points where $\overline{f}(x) \le a$. The set E would be invariant and therefore by choosing $E = F$ in (2.2.16) one would get the contradiction $\mu(E) = \mu(E)^2$. Vice versa if the time average of any function is its μ–integral we can $f(x) = \chi_F(x)$ in (2.2.16) and note that its average is $\mu(F)$ so that (2.2.16) holds. ∎

Remark: Some simple, but classical and interesting, applications will be found among the problems below: among them the theory of average rotations in epicyclic and deferent motions.

Appendix 2.2: *A proof of Birkhoff's theorem*

Proposition (2.2.2) is easily proved by following the classical scheme of Garcia, see [Ja62], simpler than the original Birkhoff's proof, [Bi31]. We first prove the μ–almost everywhere convergence. Here (Ω, S, μ) is, quite generally, an invertible metric dynamical system $\mod 0$.

We first show (2.2.8) for $f \in L_\infty(\mu)$. Let

$$f^+(x) = \limsup_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^j x), \qquad f^-(x) = \liminf_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^j x).$$

Since $f^+(x) \equiv f^+(Sx) \equiv f^+(S^{-1}x)$ and $f^-(x) \equiv f^-(Sx) \equiv f^-(S^{-1}x)$ the sets of $\mathcal{B}$ defined by $D_{ab} = \{x | f^-(x) < a < b < f^+(x)\}$ are S–invariant.

If $f^+(x)$ is not equal to $f^-(x)$ μ–almost everywhere then $a < b$ must exist such that $\mu(D_{ab}) > 0$. We show instead that $\mu(D_{ab}) = 0$, for all $a < b$.

To simplify notations set $g(x) = f(x) - b$, and $h(x) = a - f(x)$. If $x \in D_{ab}$ it is clear that either $g(x) > 0$ or there exists $M \ge 2$ such that

$$\begin{aligned} &k^{-1}(g(x) + g(Sx) + \ldots + g(S^{k-1}x)) \le 0, \qquad k = 1, 2, \ldots, M-1, \\ &M^{-1}(g(x) + g(Sx) + \ldots + g(S^{M-1}x)) > 0. \end{aligned}$$

If we call $D^{(M)}$ the set of the $x \in D_{ab}$ for which the latter relations hold one has $D_{ab} = \cup_{M=1}^\infty D^{(M)}$. By induction one can check that the integral of g on $\cup_{M=1}^k D^{(M)}$ is not negative.

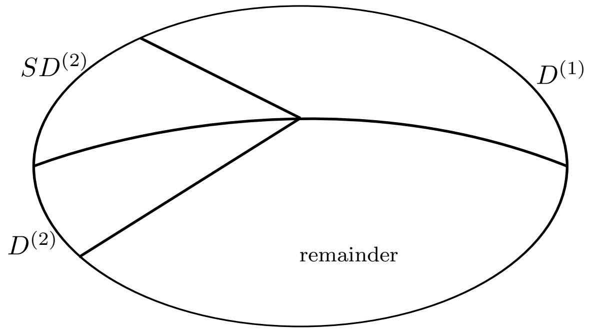
For $k = 1$ the statement is obvious. Note that, for $k = 2$, $D^{(2)}$ consists of points x such that $Sx \in D^{(1)}$ and, see Fig. (2.2.1), we see that

$$\begin{aligned} &\int_{D^{(1)}\cup D^{(2)}} g(x)\mu(dx) \\ &\qquad = \int_{D^{(1)}\setminus SD^{(2)}} g(x)\mu(dx) + \int_{D^{(2)}} (g(x) + g(Sx))\mu(dx) \ge 0, \end{aligned}$$

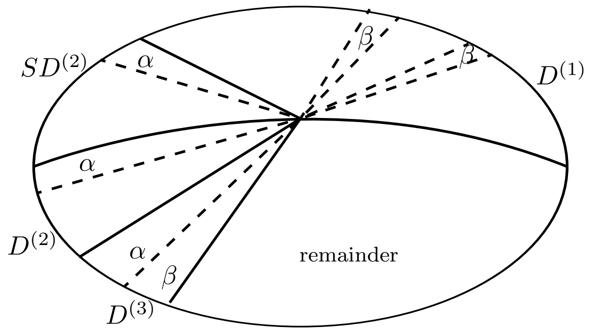
having used here the invariance of the measure μ.

Analogously $D^{(3)}$ is made of points x such that $S^2x \in D^{(1)}$ and $Sx \in D^{(1)} \cup D^{(2)}$. Then, see Fig. (2.2.2), the set $D^{(3)}$ can be divided in two parts $D^{(3)}_\alpha$ and $D^{(3)}_\beta$ such that $x \in D^{(3)}_\alpha \to Sx \in D^{(2)}$ and $x \in D^{(3)}_\beta \to Sx \in D^{(1)}$. Hence for $k = 3$, see Fig. (2.2.2), one has

$$\begin{aligned} &\int_{D^{(1)}\cup D^{(2)}\cup D^{(3)}} g(x)\mu(dx) = \int_{D^{(3)}} (g(x) + g(Sx) + g(S^2x))\mu(dx) + \\ &+ \int_{D^{(2)}\setminus SD^{(3)}_\alpha} (g(x) + g(Sx))\mu(dx) + \int_{D^{(1)}\setminus SD^{(2)}\cup SD^{(3)}_\beta \cup S^2 D^{(3)}_\beta} g(x)\mu(dx) \ge 0, \end{aligned}$$



F2.2.1 **Fig.(2.2.1)** *Illustration of the case $k = 2$: the negative contribution of the region $D^{(2)}$ is compensated by the (larger) contribution from its image $SD^{(2)}$.*



F2.2.2 **Fig.(2.2.2)** *Illustration of the case $k = 3$: the negative contribution of the region $D^{(3)}$ can be divided in the part α, with S–image and S^2–image respectively in $D^{(2)}$ and $D^{(1)}$, and in the part β with both S–image and S^2–image in $D^{(1)}$. Other cases are not possible and the negative contibutions are compensated by larger positive contributions.*

and obviously one can continue and formalize the above considerations
N2.2.2 building an inductive proof. [2]

We can proceed in an identical fashion with the function h thus arriving

[2] For instance: fix $N > 1$ (large) and let Δ_N be the set of points $x \in D_{ab}$ such that the sum $\sum_{j=0}^{N-1} g(S^j x) \geq 0$ but the sums $\sum_{j=0}^{k-1} g(S^j x) < 0$ for $k = 1, \ldots, N-1$. We define the set Δ_{N-1} as the set of points in $D_{ab} \backslash \cup_{j=0}^{N-1} S^j \Delta_N$ for which the sums $\sum_{j=0}^{k-1} g(S^j x)$ are negative if $k < N-1$ and non-negative for $k = N-1$ and we note that by construction $(\cup_{j=0}^{N-2} S^j \Delta_{N-1}) \cap (\cup_{j=0}^{N-1} S^j \Delta_N) = \emptyset$: likewise we can define Δ_j as a subset of $D_{ab}/(\cup_{i=j+1}^{N-1} \cup_{k=0}^{i-1} S^k \Delta_i)$ for $j = N, N-1, \ldots, 1$ and $\Delta_i \cap \Delta_j = \emptyset$. The integral of g over the set $D_N = \cup_{j=1}^{N} \cup_{i=0}^{j-1} S^i \Delta_j$ is therefore ≥ 0. Since $(\cup_{i=j+1}^{N-1} \cup_{k=0}^{i-1} S^k \Delta_i)$ is a family of sets which increases to D_{ab}, *i.e.* $D_{ab} = \cup_N (\cup_{i=j+1}^{N-1} \cup_{k=0}^{i-1} S^k \Delta_i)$ it follows $\int g(x) \mu(dx) \geq 0$.

to the two inequalities

$$\int_{D_{ab}} (f(x)-b)\mu(dx) \ge 0, \qquad \int_{D_{ab}} (a-f(x))\mu(dx) \ge 0,$$

that summed yield $\mu(D_{ab})\,(a-b) \ge 0$ that implies, since $a<b$, $\mu(D_{ab})=0$.

Hence we have shown that if f is in $L_\infty(\mu)$ then the limit (2.2.8) exists μ–almost everywhere. Since the sequence of averages of f is uniformly bounded if f is bounded, the limit (2.2.8) exists also in $L_1(\mu)$. However the operator $\mathcal{M}_N$, defined as $\mathcal{M}_N f(x) \stackrel{def}{=} N^{-1}\sum_{j=0}^{N-1} f(S^j x)$, is manifestly a uniformly (with respect to N) continuous operator in $L_1(\mu)$ (with norm 1) and therefore the limit (2.2.8) exists also in $L_1(\mu)$ for all $f \in L_1(\mu)$ because
N2.2.3 the space $L_\infty(\mu)$ is dense in $L_1(\mu)$.[3]

Upon a more mindful examination we see that the proof of the almost everywhere convergence could be repeated also for f just measurable with the only difference that now $f^\pm$ could assume the values $\pm\infty$ on a set of positive measure. So for all measurable functions the convergence takes place μ–almost everywhere.

Problems for §2.2

Q2.2.1 **[2.2.1]:** (*Continued fractions*)
Let $r>0$ be an irrational number represented by its continued fraction

$$r = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}} \stackrel{def}{=} [a_0, a_1, a_2, \ldots],$$

defined by setting, if $[x]$ denotes the integral part of x (to not be confused with the symbol $[a_0, a_1, a_2, \ldots]$ denoting the continued fraction), $a_0=[r]$, $r_1=(r-a_0)^{-1}$, $a_1=[r_1]$, $r_2=(r_1-a_1)^{-1}$, $a_2=[r_2]$, ... The numbers a_j are called the *entries* or the *partial quotients* of the continued fraction. Check that one has $a_j>0$ for all $j=0,1,2,\ldots$.

Q2.2.2 **[2.2.2]:** In the context of problem [2.2.1] let

$$R_k = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots \ldots + \frac{1}{a_k}}}} \stackrel{def}{=} [a_0, a_1, a_2, \ldots, a_k].$$

Show that $R_{2k} < r < R_{2k+1}$ for all $k \ge 0$.

Q2.2.3 **[2.2.3]:** (*Convergents*)
In the context of problems [2.2.1] and [2.2.2] note that if $[a_1,\ldots,a_k] = \frac{p'}{q'}$ then

[3] More explicitly this means that, since $\int_\Omega |\mathcal{M}_N f(x)|\mu(dx) \le \int_\Omega |f(x)|\mu(dx)$, if $f_n(x)$ is such that $\lim_{n\to\infty}\int_\Omega |f_n(x)-f(x)|\mu(dx) = 0$ then $\lim_{n\to\infty}\int_\Omega |\mathcal{M}_N f_n(x) - \mathcal{M}_N f(x)|\mu(dx) = 0$ and this limit is reached uniformly in N: hence $\lim_{n\to\infty}\lim_{N\to\infty}\int_\Omega |\mathcal{M}_N f_n(x) - \mathcal{M}_N f(x)|\mu(dx) = 0$, *i.e.* $\lim_{n\to\infty}\int_\Omega |\overline{f}_n(x) - \overline{f}(x)|\mu(dx) = 0$. It is now enough to remark that if $f \in L_1(\mu)$ then $f_n(x) = \min\{f(x), n\} \in L_\infty(\mu)$.

$[a_0, a_1, \dots, a_k] = \frac{a_0 p' + q'}{p'}$. Deduce from this that a vector $\underline{v}_k = (p_k, q_k) \in \mathbb{Z}_+^2$ such that $R_k = p_k/q_k$ can be taken to be

$$\underline{v}_k = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The quantities p_k/q_k are called the *convergents* of r.

Q2.2.4 **[2.2.4]:** (*Convergents recursion*)
Deduce from problem [2.2.3] that $\underline{v}_k = a_k \underline{v}_{k-1} + \underline{v}_{k-2}$, *i.e.*

$$p_k = a_k p_{k-1} + p_{k-2},$$
$$q_k = a_k q_{k-1} + q_{k-2},$$

for $k \geq 2$. (*Hint:* $\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a_k \\ 1 \end{pmatrix} = a_k \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} a_{k-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; then eliminate the last matrix in the product of matrices appearing in problem [2.2.3].)

Q2.2.5 **[2.2.5]:** From the recursion relation in problem [2.2.4] deduce that

$$q_k p_{k-1} - p_k q_{k-1} = -(q_{k-1} p_{k-2} - p_{k-1} q_{k-2}) = (-1)^k,$$
$$q_k p_{k-2} - p_k q_{k-2} = a_k (q_{k-1} p_{k-2} - p_{k-1} q_{k-2}) = (-1)^{k-1} a_k,$$

for $k \geq 2$, so that

$$\frac{p_{k-1}}{q_{k-1}} - \frac{p_k}{q_k} = \frac{(-1)^k}{q_k q_{k-1}}, \qquad \frac{p_{k-12}}{q_{k-2}} - \frac{p_k}{q_k} = \frac{(-1)^{k-1}}{q_k q_{k-2}} a_k.$$

(*Hint:* Multiply the first equation in the recursive formula in problem [2.2.4] by q_{k-1} and the second by p_{k-1} and subtract, *etc.*)

Q2.2.6 **[2.2.6]:** (*Convergents are relatively prime*)
Show that the numbers p_n, q_n are relatively prime for all n. (*Hint:* Obvious for p_0, q_0, p_1, q_1; for $k \geq 2$ this follows from the first relation in problem [2.2.5].)

Q2.2.7 **[2.2.7]:** (*Even–odd convergents*)
From problem [2.2.5] deduce that

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} \dots < r < \dots \frac{p_3}{q_3} < \frac{p_1}{q_1}, \qquad \text{and } \Big| r - \frac{p_k}{q_k} \Big| < \frac{1}{q_k q_{k+1}}.$$

Q2.2.8 **[2.2.8]:** Show that $q_k \geq 2^{(k-1)/2}$, $k \geq 0$ and $p_k \geq 2^{(k-2)/2}$, $k \geq 1$. (*Hint:* Note that $a_k \geq 1$ for all $k \geq 1$ and use the recursive relation in problem [2.2.4] and $p_j, q_j \geq 1$.)

Q2.2.9 **[2.2.9]:** (*Rational approximations by convergents*)
Show that

$$\frac{1}{q_k(q_k + q_{k+1})} < \Big| r - \frac{p_k}{q_k} \Big| < \frac{1}{q_k q_{k+1}}.$$

(*Hint:* If $\frac{a}{b} < \frac{c}{d}$ then $\frac{a + s\,c}{b + s\,d}$ increases with s for $s \geq 0$, while if $\frac{a}{b} > \frac{c}{d}$ it decreases. Hence if k is even $\frac{p_{k-2} + s\,p_{k-1}}{q_{k-2} + s\,q_{k-1}}$ increases with s and for $s = a_k$ it becomes $\frac{p_k}{q_k}$ which is such that $\frac{p_k}{q_k} < r < \frac{p_{k-1}}{q_{k-1}}$. Therefore

$$\frac{p_{k-2}}{q_{k-2}} \leq \frac{p_{k-2} + p_{k-1}}{q_{k-2} + q_{k-1}} < r,$$

hence

$$\Big|r-\frac{p_{k-2}}{q_{k-2}}\Big|>\Big|\frac{p_{k-2}+p_{k-1}}{q_{k-2}+q_{k-1}}-\frac{p_{k-2}}{q_{k-2}}\Big|\equiv\frac{1}{q_{k-2}(q_{k-2}+q_{k-1})},$$

while the other inequality follows from problem [2.2.7].

The following definition will be used below: a rational number p/q is a best approximation for r if for any pair p', q' with $q' < q$, one has $|q'r - p'| > |qr - p|$.

Q2.2.10 **[2.2.10]:** Assume r irrational, and set $\alpha_j = p_j/q_j$, where p_j/q_j are the convergents of the continued fraction of r. Let j be odd. Let p and q be positive integers such that $\alpha_{j-1} > \alpha > \alpha_{j+1}$, with $\alpha = p/q$. Then $q > q_j$. State and check the analogous result for j odd, by showing that the two results can be summarized by saying that if p/q is between two convergents of orders $j-1$ and $j+1$ then one has $q > q_j$. (*Hint:* $\alpha_{j-1} > \alpha > \alpha_{j+1} > r > \alpha_j$ so that $1/q_jq_{j-1} > |\alpha_{j-1} - r| > |\alpha_{j-1} - \alpha| = |p_{j-1}q - q_{j-1}p|/qq_{j-1} \ge 1/qq_{j-1}$, because $|p_{j-1}q - q_{j-1}p| \ge 1$).

Q2.2.11 **[2.2.11]:** In the context of problem [2.2.10] show that if j is odd and if $\alpha = p/q$, with p, q relatively prime integers and $\alpha_{j-1} < \alpha < \alpha_{j+1}$, is not a convergent, then $q_j\,|r - \alpha_j| < q\,|r - \alpha|$; a similar result holds for j even. (*Hint:* $q|\alpha - r| > q|\alpha - \alpha_{j+1}| = q|pq_{j+1} - qp_{j+1}|/qq_{j+1} \ge 1/q_{j+1} \ge q_j|\alpha_j - r|$).

Q2.2.12 **[2.2.12]:** (*Approximation by rationals*)
Show that problems [2.2.9],[2.2.10],[2.2.11] imply that if p/q is an approximation to r such that $|q'r - p'| > |qr - p|$ for all $q' < q$ then $q = q_j$, $p = p_j$ for some j. In other words every best approximant is a convergent.

Q2.2.13 **[2.2.13]:** (*Rational approximation and convergents*)
Show that if r is irrational every convergent is a best approximant. (*Hint:* If not then for some n there must exist $q < q_n$ with $|rq - p| < |rq_n - p_n| = \varepsilon_n$; let $\bar p, \bar q$ minimize the expression $|q'r - p'|$ for $q' < q_n$; if $\bar\varepsilon$ is the minimum value, one has $\bar\varepsilon < \varepsilon_n$; hence $\bar p/\bar q$ is a best approximation: so that $\bar p = p_s, \bar q = q_s$ for some $s < q_n$ and $1/(q_s + q_{s+1}) \le |q_sr - p_s| \le |q_nr - p_n| < 1/q_{n+1}$, *i.e.* $q_s + q_{s+1} > q_{n+1}$ which contradicts $q_{n+1} = a_{n+1}q_n + q_{n-1}$).

Q2.2.14 **[2.2.14]:** (*Best approximation and convergents*)
A necessary and sufficient condition in order that a rational approximation to an irrational number be a best approximation is that it is a convergent of the continued fraction of r. (*Hint:* Just a summary of problems [2.2.6] through [2.2.13]).

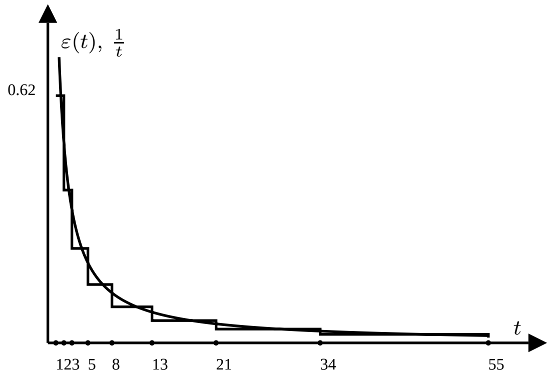
Q2.2.15 **[2.2.15]:** Show that if $q_{n-1} < q < q_n$ then $|qr - p| > |q_{n-1}r - p_{n-1}|$. Show that this can be interpreted as saying that the graph of the function $\eta(q) = \min_p |qr - p|$ is above that of the function $\eta_0(q) = \varepsilon_n = |q_nr - p_n|$ for $q_n \le q < q_{n+1}$. (*Hint:* If one had $|qr - p| \le |q_{n-1}r - p_{n-1}|$ and if $\bar\varepsilon = \min|qr - p|$ over $q_{n-1} < q < q_n$ and over p is reached at some $\bar q, \bar p$ then $\bar p/\bar q$ would be a best approximation).

Q2.2.16 **[2.2.16]:** (*Bounded entries continued fractions*)
Show that if the entries a_j of the irrational number r are uniformly bounded by N then the growth of q_n is bounded by an exponential (and one can estimate q_n by a constant times $[(N + (N^2 + 4)^{1/2})/2]^n$). Viceversa an exponential bound can hold if and only if the entries of the continued fraction are uniformly bounded. (*Hint:* The denominator q_n is bounded by that of the number with continued fraction with entries all equal to N).

Q2.2.17 **[2.2.17]:** Show that if the inequality: $|q_nr - p_n| > 1/Cq_n$ holds for all n and for a suitable C then q_n cannot grow faster than exponentially. (*Hint:* Problem [2.2.9] implies the inequality $1/Cq_n < 1/q_{n+1}$.)

Q2.2.18 **[2.2.18]:** (*Diophantine property and convergent growth*)
One says that a number r verifies a *Diophantine property* with constant C and exponent τ if $|qr - p| > C^{-1}q^{-\tau}$. Show that if the convergents growth is bounded by the law $q_{n+1} < Cq_n$, for some constant C, then the number r verifies a Diophantine property with exponent $\tau = 1$. More generally relate the growth of the convergents to the Diophantine property. (*Hint:* Refer to problem [2.2.9].)

Q2.2.19 **[2.2.19]:** (*Rotations and generation of the continued fractions entries*)
Suppose n even and think the interval $[0,1]$ as a circle of radius $1/2\pi$: the point $q_n r \mod 1$ can be represented as a point displaced by $\varepsilon_n = |q_n - p_n|$ to the right of 0, while $q_{n-1}r$ can be viewed as a point to the left of 0 by ε_{n-1}. Show that the points $qr \mod 1$, with $q_n < q < q_{n+1}$, are not in the interval $[-\varepsilon_{n-1}, 0]$ unless q is of the form $q = sq_n + q_{n-1}$, with $s \le a_{n+1}$ an integer. Furthermore show that the point $rq_{n+1} = r(a_{n+1}q_n + q_{n-1})$ is closer than ε_n to $q_{n-1}r$, and that it is the *next* position closest to 0 occurring after $q_n r$. Show that this provides a natural interpretation of the meaning of the numbers a_j in the continued fraction of r regarded as a rotation of the circle $[0,1]$, as well as a geometric interpretation of the relation $a_{n+1}q_n + q_{n-1} = q_{n+1}$. (*Hint:* Use the construction given for problem [2.2.9] to see that the points qr are between rq_{n-1} and rq_{n+1} for $q = sq_n + q_{n-1}$, with $1 \le s \le a_{n+1}$, while, for other values of $q \le q_{n+1}$, their distance from a point inside $[-\varepsilon_{n-1}, 0]$ is greater than ε_{n-1}.)



F2.2.3 **Fig.(2.2.3)** Graph of the best approximants for the golden mean. The horizontal lines end at the abscissa q_n, $n \ge 1$ and their height is the precision of the best rational approximation p/q with $q < q_n$. The continuous line is the graph of $1/t$.

Q2.2.20 **[2.2.20]:** (*Gaps in the rational approximations*)
Show that the function $\varepsilon(T)$ = *maximum gap between points of the form* $nr \mod 1, n = 1, 2, \ldots, T$ depends on T as

$$\begin{aligned} q_n \le T < q_n + q_{n-1} &\Rightarrow \varepsilon(T) = \varepsilon_{n-1} \\ q_n + q_{n-1} \le T < 2q_n + q_{n-1} &\Rightarrow \varepsilon(T) = \varepsilon_{n-1} - \varepsilon_n \\ \ldots &\Rightarrow \ldots \\ (a_{n+1}-1)q_n \le T < a_{n+1}q_n + q_{n-1} \equiv q_{n+1} &\Rightarrow \varepsilon(T) = \varepsilon_{n-1} - (a_{n+1}-1)\varepsilon_n \end{aligned}$$

and set $T_{n,k} = kq_n + q_{n-1}, k = 0, \ldots, a_{n+1} - 1$ and $\varepsilon(t) = \varepsilon(T_{n,k})$ for $T_{n,k} \le t < T_{n,k+1}$. Draw the graphs of the function equal to $\varepsilon(T)$ for $\varepsilon(T)$ and its inverse $T(\varepsilon)$ for the *golden mean*, *i.e.* the number with $a_0 = 0$ and $a_j \equiv 1$, $j \ge 1$. Explain why one should have expected that the plot of $(t, \varepsilon(t))$ for $t \ge 1$ and the plot of $(t, 1/t)$ should have the form in Fig.(2.2.3). Plot $-\log \varepsilon(T)$ in terms of $\log T$ and show that is is above a straight line with slope $\log r^{-1}$). (*Hint:* This is simply another interpretation of problem [2.2.19]).

Q2.2.21 **[2.2.21]:** (*Quadratic irrationals*)
Show that if a number has a continued fraction with entries which eventually are periodically repeated, then it is a number verifying a quadratic equation, *i.e.* it is a *quadratic irrational*. Viceversa, see problems [2.2.22] and [2.2.23], it can also be shown that all quadratic irrationals have a continued fraction with entries eventually periodic repeated.

Q2.2.22 **[2.2.22]:** Suppose that for some integers a, b, c one has $ar^2 + br + c = 0$. Remark that the argument in problem [2.2.3] shows that the number $r_n = [a_n, a_{n+1}, \ldots]$ verifies $r = (p_{n-1}r_n + p_{n-2})/(q_{n-1}r_n + q_{n-2})$. Substituting the latter expression in the equation for r one finds that r_n verifies an equation like $A_n r_n^2 + B_n r_n + C_n = 0$. Check, by direct calculation of A_n, B_n, C_n that

$$\begin{aligned} A_n &= a\, p_{n-1}^2 + b\, p_{n-1} q_{n-1} + c\, q_{n-1}^2,\\ C_n &= A_{n-1},\\ B_n^2 - 4A_n\, C_n &= b^2 - 4a\, c. \end{aligned}$$

Show that $|A_n|, |B_n|, |C_n|$ are uniformly bounded by $H = 2(2|a|r + |b| + |a)) + |b|$. Hence all quadratic irrationals are Diophantine with exponent $\tau = 1$. (*Hint:* It suffices to find a bound for $|A_n|$. Write $A_n = q_{n-1}^2(a(p_{n-1}/q_{n-1})^2 + b(p_{n-1}/q_{n-1}) + c)$ and use that $|r - p_{n-1}/q_{n-1}| < 1/q_{n-1}^2$ and $ar^2 + br + c = 0$).

Q2.2.23 **[2.2.23]:** (*Eventually periodic fractions and quadratic irrationals*)
Show that a quadratic irrational has an eventually periodic continued fraction because, as a consequence of the results of the previous problem, the numbers r_n can only take finitely many values. Show that, if H is the constant introduced in problem [2.2.22], the period length can be bounded by $2(2H+1)^3$ and that the periodic part has to start from the j-th entry with $j \le 2(2H+1)^3$.

Q2.2.24 **[2.2.24]:** (*Filling time bound*)
Show that the maximum time that one has to wait until the quasi-periodic motion $\underline{\beta} + \underline{\omega} t$ enters a ball of radius ε around $\underline{\beta}_1 \in \mathbb{T}^\ell$ can be estimated to be not larger than $const\, C\varepsilon^{-\ell-\tau}$ if $|\underline{\omega}\cdot\underline{\nu}| > C^{-1}|\underline{\nu}|^{-\tau}$ for all $\underline{0} \ne \underline{\nu} \in \mathbb{Z}^\ell$. (*Hint:* Suppose $\underline{\beta}_1 = \underline{0}$ (not restrictive). Let $\overline{\chi}(x) = (1-x^2)^{\tau+1}$ for $|x| \le 1$ and $\overline{\chi}(x) = 0$ otherwise. Define $\chi^\varepsilon(\underline{\alpha}) = \prod_{j=1}^\ell (\varepsilon\gamma)^{-1}\,\overline{\chi}(\alpha_j/\varepsilon)$, and $\gamma = (2\pi)^{-1}\int_{-1}^1 dx\,(1-x^2)^{\tau+1}$; then the Fourier transform of $\chi^\varepsilon(\underline{\alpha})$ is $\widehat{\chi}^\varepsilon_{\underline{\nu}} = \psi(\varepsilon\underline{\nu})$ with $\psi(\underline{0}) = 1, |\psi(\underline{\nu})| \le \text{const.}\prod_j \frac{1}{(1+|\nu_j|)^{2+\tau}}$. Note that if $T^{-1}\int_0^T \chi^\varepsilon(\underline{\beta}_0 + \underline{\omega}\, t)\, dt > 0$ then the motion will have visited the ball before time T. Evaluate the integral as $1 + \sum_{\underline{\nu}\ne\underline{0}} \widehat{\chi}^\varepsilon_{\underline{\nu}} e^{i\underline{\beta}_0\cdot\underline{\nu}} \frac{e^{i\underline{\omega}\cdot\underline{\nu}T}-1}{i\,\underline{\omega}\cdot\underline{\nu}\,T}$ and bound it below with

$$\Big(1 - \text{const.}\frac{C}{T}\sum_{\underline{\nu}\ne\underline{0}}\prod_{j=1}^{\ell}\frac{1}{(1+|\nu_j|\varepsilon)^{2+\tau}}|\underline{\nu}|^\tau\Big) \ge 1 - \frac{\Gamma C}{T\varepsilon^{\tau+\ell}},$$

where $\Gamma \sim \text{const.}\int_{-\infty}^\infty d^\ell \underline{x}\, |\underline{x}|^\tau \prod_j \frac{1}{(1+|x_j|)^{2+\tau}}$.)

Q2.2.25 **[2.2.25]:** (*Positive frequency time*)
In the context of problem [2.2.24] estimate the maximum time one has to wait so that while time varies between 0 and T the cylinder centered at $\underline{\beta}_0$ with basis a disk of radius ε and height 1 (say) has been visited a finite fraction of T. Show that it is not larger than $\overline{T} = const\, C\varepsilon^{-\tau-\ell+1}$: hence the estimate of the maximum filling time in problem [2.2.24] can be improved to become $\overline{T}$. (*Hint:* Just repeat the argument in problem [2.2.24] by replacing $\overline{\chi}(\underline{\alpha})$ by a function which differs from 0 only in a cylinder centered at $\underline{0}$ and having as basis a $(\ell-1)$–dimensional ball of radius ε and height 1 with axis parallel to $\underline{\omega}$.)

Q2.2.26 **[2.2.26]:** (*Minimum filling time*)
In the context of problem [2.2.24] note that the first visiting time of a ball of radius ε coincides with the first hitting time of a disk orthogonal to $\underline{\omega}$ through the center of the ball. This suggests that the estimate of $\overline{T}$ in problem [2.2.25] contains a factor $\varepsilon^{-(\ell-1)}$ which is a kind of "cross section" estimate and therefore the minimum time to enter the ball could, in fact, be simply $constC\,\varepsilon^{-\tau}$. Check that in the case $\ell = 2$ this is a consequence of the result of problems [2.2.14] through [2.2.20]. This is in fact true for all ℓ: see [BGW98].

Q2.2.27 **[2.2.27]:** Let $r = [a_0, a_1, \ldots, a_k]$, $a_i \ge 1, i > 0$, be a rational number and let $\underline{\omega} = (r, 1)$. Consider the periodic motion on $\mathbb{T}^2$ given by $\underline{\alpha}_0 + t\,\underline{\omega}$. Estimate (from below) the maximum distance that a point can have from the trajectory of $\underline{\alpha}_0$. (*Hint:* See problem [2.2.20].)

Q2.2.28 **[2.2.28]:** The dynamical system with $\Omega = \{0\}$, $S0 = 0$, $\mu(\{0\}) = 1$ is ergodic and mixing. The system $\Omega = \{0, 1\}$, $S0 = 1$ and $S1 = 0$, $\mu(\{0\}) = \mu(\{1\}) = 1/2$ is ergodic but not mixing.

Q2.2.29 **[2.2.29]:** (*S^2 ergodic implies S ergodic*)
If (Ω, S^2, μ) is ergodic also (Ω, S, μ) is such. If (Ω, S, μ) is mixing such is (Ω, S^2, μ) and vice versa. Find an example of an ergodic system (Ω, S, μ) such that (Ω, S^2, μ) is not ergodic. (*Hint:* Look into problem [2.2.28].)

Q2.2.30 **[2.2.30]:** (*Ergodicity of a non-resonant quasi-periodic flow*)
The continuous dynamical system $(\mathbb{T}^r, S_t, \lambda)$ where $S_t(\underline{\varphi}) = \underline{\varphi} + \underline{\omega}\, t \mod 2\pi$ and λ is the Lebesgue measure $\lambda(d\underline{\varphi}) = d\underline{\varphi}/(2\pi)^r$ is ergodic if and only if $(\omega_1, \ldots, \omega_r)$ are rationally independent.

Q2.2.31 **[2.2.31]:** (*Non-ergodicity of a resonant quasi-periodic flow*)
Consider the flow of problem [2.2.30] and, by suitably interpreting proposition (2.1.2), realize that it can be thought of as a "union" of ergodic systems.

Q2.2.32 **[2.2.32]:** (*Quasi-periodic flows are not mixing*)
The flow in problem [2.2.30] for all $r \geq 1$ is not mixing.

Q2.2.33 **[2.2.33]:** (*Epicyclic motion*)
Let $a_1, a_2, \ldots, a_n$ be $n > 2$ non-zero complex numbers, let $\omega_1, \ldots, \omega_n$ be n rationally independent numbers and let $\alpha_1, \ldots, \alpha_n$ be n angles. Consider the motion $z(\underline{\alpha} + \underline{\omega}t)$ of $z(\underline{\alpha}) = \sum_i a_i e^{i\alpha_i}$ (*epicyclic motion*): show that $z(\underline{\alpha}) = 0$ at most on a set of 0 volume of $\underline{\alpha}$'s in $\mathbb{T}^n$. Show that the function $\vartheta(\underline{\alpha}) = \arg(z(\underline{\alpha}))$ has derivatives $\partial_{\alpha_j}\vartheta(\underline{\alpha})$ which are measurable and the t–derivative of $\vartheta(\underline{\alpha} + \underline{\omega}t)$ is given by $f(\underline{\alpha} + \underline{\omega}t)$ with

$$f(\underline{\alpha}) = \frac{\sum_{i,j} a_i a_j \omega_i \cos(\alpha_i - \alpha_j)}{\sum_{i,j} a_i a_j \cos(\alpha_i - \alpha_j)} \equiv \sum_{i=1}^{n} \omega_i \partial_{\alpha_i}\vartheta(\underline{\alpha}),$$

in the case of real a's. Check that $f(\underline{\alpha})$ and $\partial_{\alpha_j}\vartheta(\underline{\alpha})$ are also Lebesgue integrable; if $z(\underline{\alpha})$ cannot vanish they are in fact even Riemann integrable. (*Hint:* Let $n = 3$, for instance; if $x + iy \stackrel{def}{=} z(\underline{\alpha} + \underline{\omega}t)$ the t–derivative of $\vartheta(\underline{\alpha} + \underline{\omega}t)$ is $(\dot{y}x - y\dot{x})/(x^2 + y^2)$, and the denominator can vanish when two suitable combinations of the angles α_j vanish; hence if the denominator vanishes it does so at second order at a point on a plane and at the same time the numerator vanishes to first order so that the divergence of $f(\underline{\alpha})$ is integrable, if at all present (*i.e.* if a_1, a_2, a_3 are the sides of a triangle).)

Q2.2.34 **[2.2.34]:** (*Epicyclic mean angular motion with deferent circle*)
In the context of problem [2.2.33] show that $z(\underline{\alpha} + \underline{\omega}t)$ is a point which has a well defined average rotation speed around the origin, given by $\Omega = \lim_{t\to\infty} \frac{1}{t} \arg z(\underline{\alpha} + \underline{\omega}t)$ for all $\underline{\alpha}$ if $|a_1| > \sum_{j\geq 2}^{n} |a_j|$, or more generally, if $z(\underline{\alpha})$ cannot be zero for suitably chosen $\underline{\alpha}$. Show also that one has $\Omega = \frac{1}{(2\pi)^n} \int f(\underline{\alpha}) d^n\underline{\alpha}$. The circle of radius $|a_1|$ is called *deferent* while the circles with radii a_j, $j > 1$ are called *epicycles.* (*Hint:* Since $f(\underline{\alpha})$ is Riemann integrable one can apply proposition (2.2.1).)

Q2.2.35 **[2.2.35]:** (*Epicyclic mean angular motion without deferent circle*)
In the context of problem [2.2.33] consider cases in which $z(\underline{\alpha})$ can be zero for suitably chosen $\underline{\alpha}$: *e.g.* the case $n = 3$ with a, b, c positive and equal to the sides of a triangle (we refer to such a case as a *singular epicyclic motion*). Show that the average rotation speed Ω exists for all but a set of zero measure of $\underline{\alpha}$'s. Check that in the latter cases the average is $\Omega = \frac{1}{(2\pi)^n} \int f(\underline{\alpha}) d^n\underline{\alpha}$ and that Ω is a linear combination of the components of $\underline{\omega}$ with coefficients p_j which add up to 1: $\Omega = \sum_{j=1}^{n} \omega_j p_j$. (*Hint:* Since $f(\underline{\alpha})$ is always Lebesgue measurable we can apply Birkhoff's theorem and use the ergodicity of the irrational rotations of the torus.)

Q2.2.36 **[2.2.36]:** (*No exceptions in the epicylic motion*)
In the context of problem [2.2.35] consider a singular epicyclic motion, *e.g.* the case in which $n = 3$ and a_1, a_2, a_3 are the sides of a triangle (see the previous problem). Show that the average angular motion actually does exist for *all* $\underline{\alpha} \in \mathbb{T}^3$ such that $z(\underline{\alpha} + \underline{\omega} t)$ never vanishes for any t's and is the *same* (*i.e.* Ω) for all: this is a problem posed by Lagrange and solved by Bohl. (*Hint:* Note that, by problem [2.2.35], as close as wished to every point $\overline{\underline{\alpha}}$ such that $z(\underline{\alpha} + \underline{\omega} t) \neq 0$ for all t there must be a point whose average rotation speed is Ω. The set of $\underline{\alpha}$ for which $z(\underline{\alpha})$ vanishes, "triangle configurations", is a line in $\mathbb{T}^3$: the set of points closer than ε to this line forms a tube of volume $< c\varepsilon$ for some $c > 0$. If $\underline{\alpha}_\varepsilon$ is a point with average speed of rotation Ω close to $\overline{\underline{\alpha}}$ within $\varepsilon/2$ then $\underline{\alpha}_\varepsilon + \underline{\omega} t$ and $\overline{\underline{\alpha}}_\varepsilon + \underline{\omega} t$ will stay within order of ε forever. But the fraction of time that $\underline{\alpha}_\varepsilon + \underline{\omega} t$ spends inside the tube around the singularity will be, by ergodicity, of order ε and therefore the fraction of time the argument of $z(\overline{\underline{\alpha}}_\varepsilon + \underline{\omega} t)$ may differ substantially from that of $z(\underline{\alpha}_\varepsilon + \underline{\omega} t)$ also has size of order ε; during such amounts of time the difference in the arguments can be at most 2π. Hence the average rotation of $z(\overline{\underline{\alpha}}_\varepsilon + \underline{\omega} t)$ is close within order ε to Ω and ε is arbitrary.)

Remark. Note that if $\underline{\alpha}$ is such that $z(\underline{\alpha} + \underline{\omega} t)$ vanishes one cannot define $\arg(z(\underline{\alpha} + \underline{\omega} t))$. Unless $z(\underline{\alpha} + \underline{\omega} t)$ vanishes for infinitely many t's the argument of $z(\underline{\alpha} + \underline{\omega} t)$ can be eventually meaningfully defined and by the argument in the latter problem [2.2.36] it equals Ω. However if there are two t's for which $z(\underline{\alpha} + \underline{\omega} t) = 0$ then the $\underline{\omega}$ cannot be rationally independent. Hence the result of the previous problem is that the average epicyclic motion exists aside from the cases in which it cannot be defined.

Q2.2.37 **[2.2.37]:** (*Epicylic motion in Lagrange's problem*, [AA68], p. 142)
In the context of problem [2.2.33] consider the function $\partial_{\alpha_j} \vartheta(\underline{\alpha})$ and show that the time average p_j of $\partial_{\alpha_j} \vartheta(\underline{\alpha} + \underline{\omega} t)$ is > 0 and equal to the fraction of the volume of $\mathbb{T}^n$ occupied by the $\underline{\alpha}$'s such that $|\sum_{k \neq j} a_k e^{i\alpha_k}| < |a_j|$. (*Hint:* Without loss of generality consider $j = 1$. Note that by ergodicity of the transformation $\underline{\alpha} \to \underline{\alpha} + \underline{\omega} t$ and of the transformation of the single angle $\alpha_1 \to \alpha_1 + t$ the above average is

$$p_1 = \int_{\mathbb{T}^n} \partial_{\alpha_1} \vartheta(\underline{\alpha}) \frac{d\underline{\alpha}}{(2\pi)^n} = \lim_{t \to \infty} \frac{1}{t} \int_0^t d\tau \int_{\mathbb{T}^{n-1}} \partial_{\alpha_1} \vartheta(\alpha_1 + \tau, \alpha_2, \ldots, \alpha_n) \frac{d\alpha_2 \ldots d\alpha_n}{(2\pi)^{n-1}}.$$

However $\int_0^{2\pi} d\tau \partial_{\alpha_1} \vartheta(\alpha_1 + \tau, \alpha_2, \ldots, \alpha_n) = 0$ if $|a_1| > |\sum_{j=2}^n a_j e^{i\alpha_j}|$, as a simple drawing shows if the point $z(\underline{\alpha})$ is represented as $\xi + a_1 e^{i\alpha_1}$ with $\xi = \sum_{j=2}^n a_j e^{i\alpha_j}$, and $\int_0^{2\pi} d\tau\, \partial_{\alpha_1} \vartheta(\alpha_1 + \tau, \alpha_2, \ldots, \alpha_n) = 2\pi$ otherwise.)

Q2.2.38 **[2.2.38]:** (*Bohl's formula for Lagrange's epicycles*)
In the context of problem [2.2.33] consider the case $n = 3$ with a_1, a_2, a_3 being sides of a triangle whose angles are A_1, A_2, A_3, with A_j opposite to a_j. Then the average motion is Ω given by

$$\Omega = \frac{1}{\pi}(A_1 \omega_1 + A_2 \omega_2 + A_3 \omega_3)$$

(*Hint:* This is a special case of the formula for p_j derived in problem [2.2.37].)

Q2.2.39 **[2.2.39]:** (*Existence of the rotation number*, [Le11])
Let $\vartheta' = S(\vartheta)$ be a monotonically increasing smooth function of the real variable ϑ. Suppose that $S(\vartheta + 2\pi) = S(\vartheta) + 2\pi$. Show the existence of Ω such that $S^n(\vartheta) = \vartheta + \Omega n + \varepsilon(n)$, with $\varepsilon(n)$ bounded uniformly in ϑ, n. (*Hint:* By the periodicity of S it suffices to consider $\vartheta \in [0, 2\pi)$. Assume that one has $S^n \vartheta \neq \vartheta + 2\pi p$ for all integers p and for all $\vartheta \in [0, 2\pi)$: otherwise the discussion is easier. Let $S^n \vartheta = \vartheta + 2\pi\, M_n(\vartheta) + r_n(\vartheta)$, with $M_n(\vartheta) \in \mathbb{Z}_+$ and $0 < r_n(\vartheta) < 2\pi$. But $M_n(\vartheta)$ can be discontinuous only if $r_n(\vartheta) = 0$, so that, by assumption, this can not happen. Hence $M_n(\vartheta)$ must be independent of ϑ: $M_n(\vartheta) = M_n$ for all $\vartheta \in [0, 2\pi]$. Thus $S^n(S^m \vartheta) = S^{n+m} \vartheta = \vartheta + 2\pi M_{n+m} + r_{n+m}(\vartheta)$ and

$$\begin{aligned} S^n(S^m \vartheta) &= S^n(\vartheta + 2\pi M_m + r_m(\vartheta)) = \\ &= S^n(\vartheta + r_m(\vartheta)) + 2\pi M_m = \vartheta + 2\pi(M_n + M_m) + r_n(\vartheta + r_m(\vartheta)). \end{aligned}$$

We deduce $M_n + M_m - 1 \le M_{n+m} \le M_n + M_m + 1$. By subadditivity this implies that $\lim_{n\to\infty} \frac{M_n}{n} = \Omega$ exists; the inequality also implies the existence of a constant C such that $|\frac{M_{np+m}}{np+m} - \frac{M_p}{p}| < \frac{C}{p}$ for all $p \ge 1, 0 \le m < p, n \ge 0$: hence $\frac{M_n}{n} = \Omega + \frac{\gamma_n}{n}$, with $|\gamma_n| \le C$.)

Q2.2.40 **[2.2.40]:** (*Existence of the mean motion*, from [Le11])
Let $\dot\vartheta = F(\vartheta, t)$ where $(\vartheta, t) \in \mathcal{R}^2$ are two angles and $F > 0$ is smooth on $\mathbb{R}^2$ and *periodic* with period 2π in both arguments. Show that for all initial data ϑ_0 the solution of the equation has the form $\vartheta(t) = \vartheta_0 + \Omega t + \varepsilon(t)$ with $\varepsilon(t)$ uniformly bounded for all t. (*Hint:* Define the circle map corresponding to the solutions of the differential equation considered at times $2\pi n$, $n = 0, 1, 2, \ldots, n$.)

Q2.2.41 **[2.2.41]:** Given a dynamical system (Ω, S, μ) let $\mathcal{E} = \{E_1, \ldots\}$ be a denumerable family
N2.2.4 of μ–measurable sets that generate [4] the σ–algebra $\mathcal{B}_\mu$ on which the complete S–invariant measure μ is defined. Then (Ω, S, μ) is ergodic if and only if

$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} \mu(E_k \cap S^{-j} E_h) = \mu(E_k)\mu(E_h) \text{ for all } h, k \ge 1,$$

and it is mixing if and only if

$$\lim_{j\to\infty} \mu(E_k \cap S^{-j} E_h) = \mu(E_k)\mu(E_h) \text{ for all } h, k \ge 1.$$

Q2.2.42 **[2.2.42]:** (*Bernoulli shift is mixing*)
The dynamical system (Ω, S, μ), with $\Omega = \{0,1\}^{\mathbb{Z}}$, S = the shift by one unit, μ = Bernoulli measure $B(1/2, 1/2)$ on Ω = the measure that attributes to the cylinders $C^{i_1\ldots i_p}_{\sigma_1\ldots\sigma_p}$ (cf. remark following definition (1.4.2) for the notion of cylinder) the measure 2^{-p}, is mixing.

Q2.2.43 **[2.2.43]:** (*The baker map*)
The *baker map* B is defined on the unit square $\mathcal{Q} = [0,1]^2$ of $\mathbb{R}^2$ by

$$x \to 2x \bmod 1 , \qquad y \to y/2 + [2x]/2,$$

(where $[\cdot]$ denotes the integer part function). Check that this map leaves invariant the Lebesgue measure λ on $[0,1]^2$. Show that the system (Ω, S, μ) of problem [2.2.42] is isomorphic mod 0 to $(\mathcal{Q}, B, \lambda)$. Find the measure zero set $N \subset [0,1]^2$ and $N' \subset \{0,1\}^{\mathbb{Z}}$ that have to be removed to define the isomorphism, cf. definition (1.2.3). Is B invertible? Is it invertible mod 0? (*Hint:* Write x and y in binary expansion and express B in terms of their binary expansions).

Q2.2.44 **[2.2.44]:** Let (Ω, S) be an invertible topological dynamical system and let Ω be compact metric; show that $\mu \in \mathcal{M}(\Omega, S)$ is ergodic or mixing if and only if for all $f, g \in C(\Omega)$ one has, respectively,

$$\lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^j x) = \int_\Omega f(x')\, \mu(dx')$$

or

$$\lim_{j\to\infty} \mu(f S^j g) \equiv \int_\Omega f(S^j x')\, g(x')\, \mu(dx') = \mu(f)\mu(g),$$

where $\mu(f) = \int_\Omega f(x')\mu(dx')$. Show also that in the above statements one can replace $C(\Omega)$ with any denumerable family $\{f_n\}_{n\in\mathbb{Z}}$ of functions of $C(\Omega)$ which is fundamental

[4] This means that $\mathcal{B}_\mu$ is the smallest μ–complete σ-algebra containing the sets in $\mathcal{E}$.

in $C(\Omega)$ (*i.e.* such that the linear space that it spans is dense in $C(\Omega)$ in the uniform
N2.2.5 norm).[5]

Q2.2.45 **[2.2.45]:** (*S^{-1} is ergodic or mixing if S is such*)
If (Ω, S, μ) is an invertible ergodic (or mixing) dynamical system also the "inverse" dynamical system (Ω, S^{-1}, μ) is such.

Q2.2.46 **[2.2.46]:** (*Equality of averages in the past and in the future*)
If (Ω, S, μ) is an invertible metric dynamical system, fixed $f \in L_1(\mu)$ the limits

$$\overline{f}^+(x) = \lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^j x) \quad \text{and} \quad \overline{f}^-(x) = \lim_{N\to\infty} N^{-1} \sum_{j=0}^{N-1} f(S^{-j} x)$$

exist μ–almost everywhere and are μ–almost everywhere equal. (*Hint:* Use proposition (2.2.2). Define $E = \{x | \overline{f}^+(x) > \overline{f}^-(x)\}$ and, if $\mu(E) > 0$, show that the average values in the future and in the past of $\chi_E f$ are respectively $\chi_E \overline{f}^+$ and $\chi_E \overline{f}^-$ and, hence, $(\overline{f}^+ - \overline{f}^-)\chi_E \ge 0$ and $\int \chi_E \overline{f}^+ d\mu = \int \chi_E \overline{f}^- d\mu = \int \chi_E f d\mu$ imply $(\overline{f}^+ - \overline{f}^-)\chi_E = 0$ μ–almost everywhere, *etc.*)

Q2.2.47 **[2.2.47]:** (*Frequency of digits*, see [AA68])
Let $\alpha = \log_b c$ with $b > c > 1$, b and c integers. Consider the map $Sx = x + \alpha \bmod 1$. Show that $[b^{S^n(0)}] \equiv [b^{n\alpha}]$, with $[\cdot]$ denoting the integer part, is the first digit of the development in base b of the number c^n. Making use of proposition (2.1.2), compute the frequency with which the digits 7 and 3 appear respectively as first digits of the decimal development of 3^n and show that 7 appears more often than 3 *assuming known that* $\log_3 10$ *is irrational*, (is it? check).

Q2.2.48 **[2.2.48]:** (*Arnold's cat map is mixing*)
Show that the map of $\mathbb{T}^2$: $S\underline{\varphi} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \bmod 2\pi$ is such that the dynamical system $(\mathbb{T}^2, S, \lambda)$ with $\lambda =$ Lebesgue measure $= (d\underline{\varphi})/(2\psi)^2$ is ergodic and mixing. (*Hint:* Apply the result of problem [2.2.44] and select the sequence $f_{\underline{\nu}} = e^{i\underline{\nu}\cdot\underline{\varphi}}$ as the basis for the Fourier series on $\mathbb{T}^2$.)

Q2.2.49 **[2.2.49]:** (*The product of mixing systems is mixing*)
Consider the metric dynamical system constructed starting from two metric invertible dynamical systems (Ω, S, μ) and (Ω', S', μ') and forming the new metric dynamical system, called the *tensor product* of the two dynamical systems, $(\Omega \times \Omega', S \times S', \mu \times \mu')$. Show that the tensor product is mixing if both (Ω, S, μ) and (Ω', S', μ') are mixing. (*Hint:* Make use of the result of problem [2.2.44] and of the density in $C(\Omega \times \Omega')$ of the functions having the form $\sum_{i=1}^p f_i(x) g_i(y)$ with $f_1, \ldots, f_p \in C(\Omega)$, $g_1, \ldots, g_p \in C(\Omega')$.)

Q2.2.50 **[2.2.50]:** (*Ergodicity and mixing of tensor products*)
The tensor product of two irrational rotations of the circle $\mathbb{T}^1$ with equal rotation numbers is not an ergodic system. Find esplicitly invariant sets of positive measure. However the tensor product of two mixing systems is mixing (cf. problem [2.2.49]).

Q2.2.51 **[2.2.51]:** Let (Ω, S, μ) be a metric invertible dynamical system; consider the unitary operator on $L_2(\mu)$: $(Uf)(x) = f(Sx)$. Show that (Ω, S, μ) is ergodic if and only if the equation $Uf = f$ in $L_2(\mu)$ admits only the solution $f =$ const.

Q2.2.52 **[2.2.52]:** If (Ω, S, μ) is as in problem [2.2.51] and if $f \in L_2(\Omega, \mu)$ such that $Uf = \lambda f$ with $\lambda \ne 1$ exist then (Ω, S, μ) is not mixing.

Q2.2.53 **[2.2.53]:** (*A non-Lebesgue-measurable set*)
Let $\omega/2\pi$ be irrational and consider the rotation of the unit circle $\mathbb{T}^1$ with rotation

[5] Given a set H of continuous functions $H \subset C(\Omega)$, the subspace spanned by H consists in the finite linear combinations of elements in H.

number ω. Select a point out of every trajectory. If E_0 is the set of points thus obtained check that $E_0 + j\omega$ is disjoint from $E_0 + i\omega$ for $i \neq j$. Conclude that E_0 is not Lebesgue measurable. (*Hint:* Since the $\cup_{j=-\infty}^{\infty} E_0 + j\omega = [0, 2\pi]$ and since the pairwise disjoint sets $E_0 + j\omega$, if measurable, would have all the same Lebesgue measure it follows that they could not have zero measure because the measure is completely additive and therefore the whole space would have zero measure. They cannot have positive measure because then the whole space would have infinite measure. An absurd situation.)

Q2.2.54 **[2.2.54]:** (*A non-Lebesgue-measurable function*)
Given E_0 as in problem [2.2.53] for every $x \in [0, 2\pi]$ the trajectory $T(x) = \{x'|x' = x + k\omega \bmod 2\pi$, with $k \in \mathbb{Z}\}$, intersects E_0 only in one point $\xi(x)$. Setting $f(x) = \xi(x)$ one defines a function such that $f(x) = f(x+\omega)$. Show that such function takes different values on different trajectories but it is not a first integral! hence it is not Lebesgue measurable. (*Hint:* Ergodic systems cannot have first integrals: hence the only way out is that the set E_0 cannot be measurable and the function $f(x)$ is also not measurable: but measurability is required in the definition of first integral).

Note that E_0 is the prototype of the sets which are *not Lebesgue measurable.* Making use of the axiom of the choice it becomes possible construct apparent nontrivial constants of motion for ergodic measures. However one obtains in this way examples of functions or sets which *are not measurable* with respect to the ergodic measures in question. The above set (described by Lebesgue) appears, therefore, as a non-measurable set associated with the ergodic action of a rotation of $\mathbb{T}^1$. Non measurability is the mathematically precise definition of a function that *cannot* be tabulated with any prefixed appximation. The axiom of choice is used here when the choice set E_0 is obtained, *i.e.* in Zermelo's formulation.

Bibliographical note to §2.2

The proof of Birkhoff's ergodic theorem goes back to 1931, [Bi31]. The one presented above is taken from [Ja62]. Use of continued fractions in Science seems to be documented to go back at least to Aristarchos (about 280 b.c.) (who gives as tropical year length $365 + (1/4 + 1/(20 + 2/60))$ mean solar days while the solar year was given as $365 + (1/4 - 1/(10 - 1/4))$ mean solar days) and it is likely to signal that he realized existence of equinox precession before Hypparchos did 150 years later. [Ra99] The continued fractions theory presentation is here inspired by [Ki64]; the case of Diophantine vectors of dimension higher than 2 cannot be reduced to the theory of continued fractions, nevertheless the above approach to continued fractions admits some extensions and in the problems for Section §(8.1) we provide a few examples inspired by [Ch99], see also [CM01]. The rotation number theory for circle maps is taken from [Le11]: it is due to Poincaré, [Po85]. The theory of epicyclic motions is essentially due to Lagrange in the cases in which a deferent is present. The case in absence of deferents has been solved by Bohl in the $n = 3$ case and extended by Weyl to the general case, [AA68].

§2.3 Ergodic points

The considerations and definitions of the previous sections lead to a natural classification of the motions associated with a continuous map S of a compact metric space Ω, observed on a partition $\mathcal{P}$. This classification divides motions in three classes: the *ergodic* motions, the motions *with defined frequencies* but which are not ergodic and the *exceptional* motions, *i.e.* motions with not defined frequencies.

Let $\widehat{\Omega}$ be the closure of the set of $(\mathcal{P},S)$–histories of points in Ω: this is, in general, larger than the set of histories of points in Ω and has been called in definition (1.4.2) the set of symbolic motions seen from $\mathcal{P}$. We begin by defining the exceptional points.

D2.3.1 **(2.3.1) Definition:** (Exceptional points)
Let (Ω,S) be an invertible dynamical system and let $\mathcal{P}=\{P_0,\ldots,P_n\}$ be a partition of Ω. If $x\in\Omega$ let $\underline{\sigma}(x)$ be its $(\mathcal{P},S)$–history.

*The set of the sequences $\underline{\sigma}\in\widehat{\Omega}$ with not defined frequencies will be called the set of $(\mathcal{P},S)$–*exceptional symbolic motions *and we shall denote it by $\widehat{\mathcal{E}}'$. The set of the points $x\in\Omega$ such that $\underline{\sigma}(x)\in\widehat{\mathcal{E}}'$ will be denoted $\mathcal{E}'(\mathcal{P},S)$ and it will be the set of the $(\mathcal{P},S)$–*exceptional points.

We need to investigate the classification of the remaining points of Ω and of $\widehat{\Omega}$ in order to show that *all statistical properties of motions are carried by the ergodic points*. This will be achieved in the next section. It is convenient to introduce, for this purpose, a few more definitions.

D2.3.2 **(2.3.2) Definition:** (Stationary distribution)
A sequence $\underline{p}$ of non negative numbers,

$$\underline{p}=\Big(p\Big(\begin{matrix}j_1\cdots j_q\\ \sigma_1\ldots\sigma_q\end{matrix}\Big)\Big),\quad q\ge 1,\quad \{j_1,\ldots,j_q\}\subset\mathbb{Z},\ \sigma_1,\ldots,\sigma_q\in\{0,\ldots,n\},$$

will be called a stationary distribution *on $\{0,\ldots,n\}^{\mathbb{Z}}$ if the following properties hold for every choice of the labels:*
(i) positivity*:*

e2.3.1
$$p\begin{pmatrix}J\\ \underline{\sigma}\end{pmatrix}=p\begin{pmatrix}j_1\cdots j_q\\ \sigma_1\ldots\sigma_q\end{pmatrix}\ge 0,\tag{2.3.1}$$

(ii) normalization*:*

e2.3.2
$$\sum_{\sigma_1,\ldots,\sigma_q=0}^{n}p\begin{pmatrix}j_1\cdots j_q\\ \sigma_1\ldots\sigma_q\end{pmatrix}=1,\tag{2.3.2}$$

(iii) compatibility*:*

$$\sum_{\sigma=0}^{n}p\Big(\begin{matrix}j_1\cdots j_{q-1}\ j_q\ j_{q+1}\cdots j_s\\ \sigma_1\ldots\sigma_{q-1}\ \sigma\ \sigma_{q+1}\ldots\sigma_s\end{matrix}\Big)=p\Big(\begin{matrix}j_1\cdots j_{q-1}\,j_{q+1}\cdots j_s\\ \sigma_1\ldots\sigma_{q-1}\,\sigma_{q+1}\ldots\sigma_s\end{matrix}\Big),$$

(iv) stationarity*: if $J+k=\{j_1+k,\ldots,j_q+k\}$ when $J=\{j_1,\ldots,j_q\}$ and for all q–ples $\underline{\sigma}=(\sigma_1,\ldots,\sigma_q)$ and all $k\in\mathbb{Z}$*

e2.3.3
$$p\Big(\begin{matrix}J+k\\ \underline{\sigma}\end{matrix}\Big)=p\Big(\begin{matrix}J\\ \underline{\sigma}\end{matrix}\Big)\qquad\textit{for all }k\in\mathbb{Z}.\tag{2.3.3}$$

Often a stationary distribution in the above sense is also called a *stochastic process* with space of states $\{0,1,\ldots,n\}$. Related notions are expressed by the following definition.

D2.3.3 **(2.3.3) Definition:** (Nonstationary, ergodic and mixing distributions) *If properties (i),(ii),(iii) of definition (2.3.2) hold then we shall simply say that* $\underline{p}$ *is a* distribution *on* $\{0,\ldots,n\}^{\mathbb{Z}}$*: if property (iv) does not hold we say that* $\underline{p}$ *is a* nonstationary distribution.
(i) The collection of all distributions, stationary and nonstationary, will be denoted $M^0(\{0,\ldots,n\}^{\mathbb{Z}})$*. It is a compact set in the topology inherited as a subset of*

e2.3.4
$$K=\prod_{q=1}^{\infty}\prod_{\{j_1,\ldots,j_q\}\subset\mathbb{Z}}\ \prod_{\sigma_1,\ldots,\sigma_q\in\{0,\ldots,n\}}[0,1], \tag{2.3.4}$$

endowed with the product topology.
(ii) The collection of the stationary distributions in $M^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *will be denoted by* $M(\{0,\ldots,n\}^{\mathbb{Z}})$*.*
(iii) The stationary distributions which are ergodic*, i.e. such that*

e2.3.5
$$\lim_{N\to\infty}N^{-1}\sum_{k=0}^{N-1}p\begin{pmatrix}J & J'+k\\ \underline{\sigma}_J & \underline{\sigma}'_{J'}\end{pmatrix}=p\begin{pmatrix}J\\ \underline{\sigma}\end{pmatrix}p\begin{pmatrix}J'\\ \underline{\sigma}'_{J'}\end{pmatrix} \tag{2.3.5}$$

for all finite J,J' *and for all strings* $\underline{\sigma}_J=(\sigma_j)_{j\in J}$*,* $\underline{\sigma}_{J'}=(\sigma'_{j'})_{j'\in J'}$*, will be denoted by* $M_e(\{0,\ldots,n\}^{\mathbb{Z}})$*.*
(iv) The stationary distributions which are mixing*, i.e. such that*

e2.3.6
$$\lim_{k\to\infty}p\begin{pmatrix}J & J'+k\\ \underline{\sigma}_J & \underline{\sigma}'_{J'}\end{pmatrix}=p\begin{pmatrix}J\\ \underline{\sigma}_J\end{pmatrix}p\begin{pmatrix}J'\\ \underline{\sigma}'_{J'}\end{pmatrix}, \tag{2.3.6}$$

will be denoted by $M_m(\{0,\ldots,n\}^{\mathbb{Z}})$*.*

Remarks: (1) One sees that M^0 and M are closed subsets in K, while M_e and M_m are Borel sets in K, being defined via limits of continuous functions.
(2) One also sees that if $\underline{\widehat{\sigma}}$ is a given sequence with defined frequencies then the set of the frequencies of $\underline{\widehat{\sigma}}$:

e2.3.7
$$p\Big(\begin{matrix}j_1\ldots j_p\\ \sigma_1\ldots\sigma_q\end{matrix}\,|\,\underline{\widehat{\sigma}}\Big), \tag{2.3.7}$$

$q\ge 1$, $\{j_1,\ldots,j_q\}\subset\mathbb{Z}$, $\sigma_1,\ldots,\sigma_q\in\{0,\ldots,n\}$, is an element of $M(\{0,\ldots,n\}^{\mathbb{Z}})$: it will be called the *distribution of* $\underline{\widehat{\sigma}}$.
(3) The notion of distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ can be identified with that of "probability measure" on the Borel sets of $\{0,\ldots,n\}^{\mathbb{Z}}$. If we denote $\mathcal{M}^0(K)$ the set of the probability measures on the Borel sets of a topological space K, the identification is described by the following proposition.

P2.3.1 **(2.3.1) Proposition:** (Distributions and probability measures)
(i) Let $\underline{p} \in M^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *and, for every cylinder* $C^J_{\underline{\sigma}} = C^{j_1\ldots j_q}_{\sigma_1\ldots\sigma_q} \subset \{0,\ldots,n\}^{\mathbb{Z}}$*, define*

e2.3.8
$$m_{\underline{p}}(C^J_{\underline{\sigma}}) = p\binom{J}{\underline{\sigma}}. \tag{2.3.8}$$

Then there is a unique probability measure $m_{\underline{p}} \in \mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *on* $\{0,\ldots,n\}^{\mathbb{Z}}$ *which attributes to the cylinder* $C^J_{\underline{\sigma}}$ *the measure* $m_{\underline{p}}(C^J_{\underline{\sigma}})$ *given by (2.3.8).*
(ii) Vice versa for each $m \in \mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *the sequence* $\underline{p} = (m(C^J_{\underline{\sigma}}))$ *is in* $M^0(\{0,\ldots,n\}^{\mathbb{Z}})$*.*
(iii) The correspondence $\underline{p} \longleftrightarrow m_{\underline{p}}$ *is one-to-one between* $M^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *and* $\mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}})$ *and it is continuous if the latter space is considered as a (compact) topological space with the weak topology induced by the continuous functions on* $\{0,\ldots,n\}^{\mathbb{Z}}$*.*[1] N2.3.1
(iv) Furthermore the correspondence transforms τ*–invariant measures (where* τ *is the translation* $(\tau\underline{\sigma})_i = \sigma_{i+1}$, $i \in \mathbb{Z}$*) into stationary distributions* $M(\{0,\ldots,n\}^{\mathbb{Z}})$*;* τ*–ergodic measures* $\mathcal{M}_e(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$ *are mapped into* $M_e(\{0,\ldots,n\}^{\mathbb{Z}})$*, and* τ*–mixing measures* $\mathcal{M}_m(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$ *into* $M_m(\{0,\ldots,n\}^{\mathbb{Z}})$*.*

Remark: Because of the above proposition we shall, in the following, identify the notions of distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ with that of Borel probability measure on $\{0,\ldots,n\}^{\mathbb{Z}}$. This means that we identify

$$\begin{aligned} M^0(\{0,\ldots,n\}^{\mathbb{Z}}) &\text{ with } \mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}}),\\ M(\{0,\ldots,n\}^{\mathbb{Z}}) &\text{ with } \mathcal{M}(\{0,\ldots,n\}^{\mathbb{Z}},\tau),\\ M_e(\{0,\ldots,n\}^{\mathbb{Z}}) &\text{ with } \mathcal{M}_e(\{0,\ldots,n\}^{\mathbb{Z}},\tau),\\ M_m(\{0,\ldots,n\}^{\mathbb{Z}}) &\text{ with } \mathcal{M}_m(\{0,\ldots,n\}^{\mathbb{Z}},\tau). \end{aligned}$$

Proof: Let $\widetilde{\mathcal{B}}$ be the algebra of subsets of $\{0,\ldots,n\}^{\mathbb{Z}}$ generated by the subsets E of $\{0,\ldots,n\}^{\mathbb{Z}}$ that can be represented as

e2.3.9
$$E = \bigcup_{\sigma_{-p}\ldots\sigma_p\in\Delta} C^{-p\;\cdots\;p}_{\sigma_{-p}\;\cdots\;\sigma_p}, \tag{2.3.9}$$

for some finite p and $\Delta \subset \{0,\ldots,n\}^{2p+1}$.
If $m_{\underline{p}}$ exists then it is uniquely determined on $\widetilde{\mathcal{B}}$ by (2.3.8); in fact by additivity one must have

e2.3.10
$$m_{\underline{p}}(E) = \sum_{\sigma_{-p}\ldots\sigma_p\in\Delta} p\begin{pmatrix} -p\ldots p\\ \sigma_{-p}\ldots\sigma_p\end{pmatrix}, \tag{2.3.10}$$

[1] Given a compact topological space Ω one says that $\mu_n \in \mathcal{M}^0(\Omega)$ converges weakly to $\mu \in \mathcal{M}^0(\Omega)$ if $\lim_{n\to\infty}\int_\Omega f(x)\mu_n(dx) = \int_\Omega f(x)\mu(dx)$ for every $f \in C(\Omega)$. The topology corresponding to this notion of convergence is called the weak topology on $\mathcal{M}^0(\Omega)$.

and, since $\widetilde{\mathcal{B}}$ generates the σ–algebra of Borel sets (the cylinders being a base for the topology of $\{0,\ldots,n\}^{\mathbb{Z}}$), $m_{\underline{p}}$ is fixed by (2.3.10), if it exists.

To show existence of $m_{\underline{p}}$ we first show that the (2.3.10) can be really used to define $m_{\underline{p}}(E)$, for all $E \in \widetilde{\mathcal{B}}$: the problem lies obviously in the fact that E can be represented in infinitely many ways in the form (2.3.9) and it is necessary to verify that the value of the sum (2.3.10) does not depend on the particular representation of E.

Suppose that E can also be represented as

e2.3.11 $$E = \bigcup_{\sigma'_{-p'}\ldots\sigma'_{p'}\in\Delta'} C^{-p'\,\ldots\,p'}_{\sigma'_{-p'}\ldots\sigma'_{p'}}, \tag{2.3.11}$$

and suppose that $p' \geq p$. Then the latter equality can be true if and only if

e2.3.12 $$\Delta' = \{\{\sigma'_{-p'}\ldots\sigma'_{p'}\} | \{\sigma'_{-p},\ldots,\sigma'_p\} \in \Delta\}; \tag{2.3.12}$$

in fact cylinders that have a different specification (cf. remark following definition (1.4.2)) are necessarily disjoint.

Then the measure of E computed with the representation (2.3.11) will be

e2.3.13 $$\sum_{\substack{\sigma'_{-p'}\ldots\sigma'_{-p-1}\\ \sigma'_{p+1}\ldots\sigma'_{p'}}} \sum_{\{\sigma'_{-p}\ldots\sigma'_p\}\in\Delta} p\begin{pmatrix} -p'\ldots -p-1 & -p\ \ldots\ p & p+1\ldots p' \\ \sigma'_{-p}\ldots\ \sigma'_{-p-1} & \sigma'_{-p}\ldots\sigma'_p & \sigma_{p+1}\ \ldots\sigma'_{p'} \end{pmatrix}, \tag{2.3.13}$$

that coincides with (2.3.10) by virtue of (iii) of definition (2.3.2).

Hence $m_{\underline{p}}(E)$ is well defined on $\widetilde{\mathcal{B}}$ and the function $E \to m_{\underline{p}}(E)$ is an additive non-negative function of the sets in $\widetilde{\mathcal{B}}$. Since the sets of $\widetilde{\mathcal{B}}$ are both open and closed the measure μ_0 is also a regular measure, *i.e.* every set of $\widetilde{\mathcal{B}}$ contains a closed set (itself) and is contained in an open set (itself) whose measures differ by less than an arbitrarily fixed amount, cf. [DS58], p.137.

To show that $m_{\underline{p}}$ can be extended to a completely additive measure on the
N2.3.2 smallest σ–algebra that contains $\widetilde{\mathcal{B}}$ (*i.e.* to the Borel sets) it will suffice[2] to show that $m_{\underline{p}}$ is completely additive on $\widetilde{\mathcal{B}}$. But this is so because all sets of $\widetilde{\mathcal{B}}$ are at the same time open and closed, hence compact (cf. (2.3.11)).

Then if $E \in \widetilde{\mathcal{B}}$ is a union $\cup_{i=1}^{\infty} E_i$, $E_i \in \widetilde{\mathcal{B}}$, and $E_i \cap E_j = \emptyset$, for all $i \neq j$, one must have $E_i = \emptyset$ except for a finite number of labels (since E, being compact, is covered by the open disjoint sets E_i). Hence $E = \cup_{i=1}^{N} E_i$ and therefore complete additivity on $\widetilde{\mathcal{B}}$ coincides with the finite additivity.

This shows that $m_{\underline{p}}$ can be uniquely extended to a probability measure on the Borel sets of $\{0,\ldots,n\}^{\mathbb{Z}}$. The remaining statements are left to the reader. ∎

It is now possible to complete the classification of the points of Ω and $\widehat{\Omega}$.

[2] This is a very general result of measure theory, *Alexandrov theorem*, see [DS58], p. 138.

D2.3.4 **(2.3.4) Definition:** (Ergodic motions, ergodic points)
With the notations of definitions (2.3.1), (2.3.2) and (2.3.3) one has the following.
(i) Given $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ *let* $\widehat{\mathcal{E}}_{\underline{p}}$ *be the set of sequences* $\underline{\widehat{\sigma}} \in \widehat{\Omega}$ *in which (*cf. *definition (1.4.3)) the string homologue to* $\begin{pmatrix} j_1 \cdots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix}$ *appears with defined frequency*

e2.3.14
$$p\begin{pmatrix} j_1 \cdots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix}\!\!\Big|\underline{\widehat{\sigma}}\Big) \stackrel{def}{=} p\begin{pmatrix} j_1 \cdots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix} \text{ for all } q,\ j_1,\ldots,\sigma_1\ldots \tag{2.3.14}$$

(ii) Let $\widehat{M}$ *be the set of distributions* $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ *that are the distributions of some* $\underline{\sigma} \in \widehat{\Omega}$*,*
(iii) Let $\widehat{M}_e \stackrel{def}{=} \widehat{M} \cap M_e(\{0,\ldots,n\}^{\mathbb{Z}})$*.*
(iv) If $\widehat{\mathcal{E}}'$ *denotes the set of exceptional sequences in* $\{0,\ldots,n\}^{\mathbb{Z}}$*, i.e. of the sequences in* $\widehat{\Omega}$ *which have not defined frequencies, see definition (2.3.1), we shall set*

e2.3.15
$$\widehat{\mathcal{E}} = \bigcup_{\underline{p}\in\widehat{M}_e} \widehat{\mathcal{E}}_{\underline{p}} \subset \widehat{\Omega}, \qquad \widehat{\mathcal{E}}'' = \widehat{\Omega}/(\widehat{\mathcal{E}} \cup \widehat{\mathcal{E}}') \tag{2.3.15}$$

and the points of $\widehat{\mathcal{E}}$ *will be called the* ergodic symbolic motions *of* $\widehat{\Omega}$*. Evidently* $\widehat{\Omega} = \widehat{\mathcal{E}} \cup \widehat{\mathcal{E}}' \cup \widehat{\mathcal{E}}''$*.*
(v) Likewise we shall denote $\mathcal{E}(\mathcal{P},S)$*,* $\mathcal{E}'(\mathcal{P},S)$*,* $\mathcal{E}''(\mathcal{P},S)$ *the points of* Ω *whose* $(\mathcal{P},S)$*–histories are in* $\widehat{\mathcal{E}}$*,* $\widehat{\mathcal{E}}'$*,* $\widehat{\mathcal{E}}$*" respectively.*

Remarks: (1) The sets $\widehat{\mathcal{E}}'$, $\widehat{\mathcal{E}}''$, $\widehat{\mathcal{E}}$, $\widehat{\mathcal{E}}_{\underline{p}}$ are Borel sets in $\widehat{\Omega}$. Note, in fact, that $\widehat{\Omega}$ is closed in $\{0,\ldots,n\}^{\mathbb{Z}}$ and the above sets can be defined in terms of properties of the following continuous functions on $\widehat{\Omega}$

e2.3.16
$$\underline{\widehat{\sigma}} \to p_N\Big(\begin{matrix} J \\ \underline{\sigma} \end{matrix}\,\Big|\,\underline{\widehat{\sigma}}\Big) = N^{-1}\{\text{number of strings homologue to } \tbinom{J}{\underline{\sigma}} \text{ appearing in } \underline{\widehat{\sigma}} \text{ between 0 and } N\}, \tag{2.3.16}$$

and of their limits as $N \to \infty$.
(2) It should be clear that usually $\widehat{\mathcal{E}}_{\underline{p}} = \emptyset$ because $\widehat{M}$ will be much smaller than $M(\{0,\ldots,n\}^{\mathbb{Z}})$.
(3) We shall call, by definition, *statistical properties* of the motions of a dynamical system (Ω,S) observed on $\mathcal{P}$ all properties that one can observe by randomly selecting a $\underline{\sigma} \in \widehat{\Omega}$ with a probability distribution $m \in \mathcal{M}(\widehat{\Omega},\tau)$, *i.e. with any probability distribution which is* τ*–invariant and concentrated on* $\widehat{\Omega}$.

After all the above definitions, the following propositions allow us to fix ideas in a somewhat more orderly fashion.

P2.3.2 **(2.3.2) Proposition:** (Characterization of distributions describing the statistical properties of motions)

Let (Ω, S) *be a dynamical system and let* $\mathcal{P} = \{P_0, \ldots, P_n\}$ *be a partition of* Ω.
(i) A necessary and sufficient condition in order that a distribution $\underline{p}$ *in* $M_e(\{0, \ldots, n\}^{\mathbb{Z}})$ *be the distribution of an ergodic point of* $\widehat{\Omega}$ *(i.e.* $\underline{p} \in \widehat{M_e}$*) is that* $p\binom{J}{\underline{\sigma}} = 0$ *if* $P^J_{\underline{\sigma}}$ *is empty. In such case* $m_{\underline{p}}(\widehat{\mathcal{E}}_{\underline{p}}) = 1$.
(ii) A necessary and sufficient condition in order that $\underline{p} \in M(\{0, \ldots, n\}^{\mathbb{Z}})$ *be such that* $m_{\underline{p}}(\widehat{\Omega}) = 1$ *is that* $p\binom{J}{\underline{\sigma}} = 0$ *if* $P^J_{\underline{\sigma}} = \emptyset$.

Remarks: (1) Hence all ergodic distributions of $\widehat{\Omega}$ can be built by computing the distribution of the frequencies of the ergodic sequences $\underline{\sigma} \in \widehat{\Omega}$: it is clear that the systems $(\widehat{\Omega}, \tau, m)$ which are ergodic are therefore constructed starting from the ergodic points of $\widehat{\Omega}$.
(2) If $m_{\underline{p}}(\widehat{\Omega}) = 1$ and $m_{\underline{p}}$ is an invariant measure it is not necessarily true that there exists $\underline{\sigma} \in \widehat{\Omega}$ whose distribution is precisely $\underline{p}$ (unless $\underline{p} \in M_e$). This "anomaly" will be "explained" by the ergodic decomposition theorem of the next section where it will be shown that, nevertheless, every such measure can be constructed from the ergodic measures.

Proof: Item (ii) will be discussed first. By definition (1.4.2) and (1.4.6), the set $\widehat{\Omega}$ is

e2.3.17
$$\widehat{\Omega} = \bigcap_{J \subset \mathcal{Z}, |J| < \infty} \{\underline{\sigma} \,|\, \underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}},\ P^J_{\underline{\sigma}_J} \neq \emptyset\} \stackrel{def}{=} \bigcap_{J \subset \mathbb{Z}, |J| < \infty} \Gamma_J, \qquad (2.3.17)$$

where $\underline{\sigma}_J = (\sigma_i)_{j \in J}$, $P^J_{\underline{\sigma}_J} = \cap_{j \in J} S^{-j} P_{\sigma_j}$ and Γ_J is implicitly defined.
The set Γ_J is a union of cylinders with base J and its complement is

e2.3.18
$$\{0, \ldots, n\}^{\mathbb{Z}} \backslash \Gamma_J = \bigcup_{\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}},\ P^J_{\underline{\sigma}_J} = \emptyset} C^J_{\underline{\sigma}_J}, \qquad (2.3.18)$$

which, if $p(C^J_{\underline{\sigma}}) = 0$ when $P^J_{\underline{\sigma}} = \emptyset$, has zero $m_{\underline{p}}$–measure. Hence in such case $m_{\underline{p}}(\Gamma_J) = 1$ and, therefore, $m_{\underline{p}}(\widehat{\Omega}) = 1$. The viceversa can be checked by running the argument backwards.
To show (i) let $\underline{p} \in M_e(\{0, \ldots, n\}^{\mathbb{Z}}$ and $p\binom{J}{\underline{\sigma}} = 0$ for every $\binom{J}{\underline{\sigma}}$ such that $P^J_{\underline{\sigma}} = \emptyset$. Then $m_{\underline{p}}(\widehat{\Omega}) = 1$ by (ii) and $m_{\underline{p}}$ is ergodic. Thus $m_{\underline{p}}$–almost everywhere one has

e2.3.19
$$\lim_{N \to \infty} N^{-1} \sum_{j=0}^{N-1} \chi_{C^{i_1 \ldots i_p}_{\sigma_1 \ldots \sigma_p}}(\tau^j \underline{\sigma}) = m_{\underline{p}}(C^{i_1 \ldots i_p}_{\sigma_1 \ldots \sigma_p}) = p \begin{pmatrix} i_1 \ldots i_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}, \qquad (2.3.19)$$

for every choice of the labels. Hence there exist (many) points $\underline{\sigma}$ whose distribution is $\underline{p}$: indeed $m_{\underline{p}}$–almost all points share this property so that $m_{\underline{p}}(\widehat{\mathcal{E}}_{\underline{p}}) = 1$. The viceversa is also not difficult. ■

It is useful to remark explicitly the following corollary.

C2.3.1 **(2.3.1) Corollary:** (Measurability of the ergodic distributions)
$\widehat{M}_e$ is a Borel set in $M(\{0,\ldots,n\}^{\mathbb{Z}})$.

Proof: $\widehat{M}_e = M_e(\{0,\ldots,n\}^{\mathbb{Z}}) \cap \{\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}}) \,|\, p\binom{J}{\underline{\sigma}} = 0$ for every J, $\underline{\sigma}$ such that $P^J_{\underline{\sigma}} = \emptyset\}$; hence M_e is a Borel set in M (cf. remark (1) to definition (2.3.4)) and the second set that appears in this intersection is obviously closed and $\widehat{M}_e$ is the intersection of two Borel sets. ∎

It is good, before proceeding to the announced discussion of the ergodic decomposition, to summarize and clarify further the relation between $\widehat{\Omega}$ and Ω in the case in which Ω is a compact metric space and S is expansive on $\mathcal{P}$. This is a case in which it is easy to establish a precise relation between the "real motions" in Ω and the "symbolic motions" in $\widehat{\Omega}$.

P2.3.3 **(2.3.3) Proposition:** (Coding real motions into symbolic motions and measurability properties of the codes)
Let (Ω, S) be an invertible topological dynamical system; let Ω be compact metric space, and let $\mathcal{P} = \{P_0,\ldots,P_n\}$ be a topological partition of Ω on which S is expansive.
Let $\Sigma : \Omega \to \widehat{\Omega}$ be the code that to every $x \in \Omega$ associates the $(\mathcal{P}, S)$–history $\underline{\sigma}(x)$ and $X : \widehat{\Omega} \to \Omega$ the map defined by (1.4.11). Then
(i) X is continuous;
(ii) $\Sigma^{-1}X^{-1}(x) = x$ for every x of Ω;
(iii) X and Σ are the inverse of each other if considered as maps between Ω and $\Sigma\Omega$ and, respectively, between $\Sigma\Omega$ and Ω;
(iv) $\Sigma\Omega$ is a Borel set in $\widehat{\Omega}$;
(v) the map $A : \mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}}) \to \mathcal{M}^0(\Omega)$:

e2.3.20
$$(A\,m)(E) = m(X^{-1}(E)) \qquad \text{for } E \in \mathcal{B}(\Omega) \tag{2.3.20}$$

is continuous;
(vi) the map $B : \mathcal{M}^0(\Omega) \to \mathcal{M}^0(\{0,\ldots,n\}^{\mathbb{Z}})$:

e2.3.21
$$(B\,\mu)(F) = \mu(\Sigma^{-1}(F)) \qquad \text{for } F \in \mathcal{B}(\{0,\ldots,n\}^{\mathbb{Z}}) \tag{2.3.21}$$

has image $M^0(\Sigma\Omega) = \{m | m \in M^0(\{0,\ldots,n\}^{\mathbb{Z}}),\ m(\Sigma\Omega) = 1\}$ and it is measurable (with respect to the Borel sets in $\mathcal{M}^0$ and M^0);
(vii) A and B are inverse of each other as maps between $M^0(\Sigma\Omega)$ and $\mathcal{M}^0(\Omega)$ and, respectively, between $\mathcal{M}^0(\Omega)$ and $M^0(\Sigma\Omega)$.

Proof: Points (i)-(iii) were proved in proposition (1.4.1).
The correspondence Σ is measurable because $\Sigma^{-1}C^J_{\underline{\sigma}} = P^J_{\underline{\sigma}}$ for all $J, \underline{\sigma}$; then (iv) follows from the one to one property of the correspondence Σ and from Kuratowsky's theorem (which implies that the image of a Borel set via a one-to-one Borel map between two complete separable metric spaces
N2.3.3 is still a Borel set).[3]

[3] A Borel map, cf. Appendix 1.2, is a map such that the inverse image of any Borel set is a Borel set; see [Pa67] Ch. I, Sec. 3, Theorem 3.9.

The remaining two items follow easily from the above remarks (and from Kuratowski's theorem, which is needed to solve the other measurability problems that arise). For instance measurability of B is reduced immediately to checking that the functions $\mu \to \mu(P^J_{\underline{\sigma}})$ are Borel functions as functions on $\mathcal{M}^0(\Omega)$; this follows from fact that $\mathcal{P}$ is a topological partition and from the Borel measurability of the functions $\mu \to \mu(E)$ on $M(\{0,\ldots,n\}^{\mathbb{Z}})$ for each closed E (due to the approximability of the characteristic functions χ_E by functions of $C(\Omega)$). ∎

Problems for §2.3

Q2.3.1 **[2.3.1]:** Let $\underline{p} \in M \equiv M(\{0,\ldots,n\}^{\mathbb{Z}})$ be such that $p\begin{pmatrix} j_1 \ldots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix} \neq 0$ if and only if $\sigma_1 = \ldots = \sigma_q = 0$. Which are the possible values of the measure of a generic set $E \in \mathcal{B}(M)$?

Q2.3.2 **[2.3.2]:** (*Bernoulli shift*)
Let $\underline{\pi} = (\pi_\sigma)_{\sigma=0,\ldots,n}$ be an $(n+1)$–ple of positive numbers such that $\sum_\sigma \pi_\sigma = 1$. If we set, for every choice of the labels,

$$p\begin{pmatrix} j_1 \ldots j_q \\ \sigma_1 \ldots \sigma_q \end{pmatrix} = \prod_{i=1}^{q} \pi_{\sigma_i},$$

check that $\underline{p}$ is a distribution in $M(\{0,\ldots,n\}^{\mathbb{Z}})$: it is a distribution often called the *Bernoulli shift (or Bernoulli scheme)* $B(\pi_0,\ldots,\pi_q)$. (*Hint:* See definition (2.3.2) and proposition (2.3.1).)

Q2.3.3 **[2.3.3]:** If for every $i \in \mathbb{Z}$ we assign $(n+1)$ numbers $\underline{\pi}^{(i)} = (\pi^{(i)}_\sigma)_{\sigma=0,\ldots,n}$, with $\sum_\sigma \pi^{(i)}_\sigma = 1$, and if we set $p\binom{j_1\ldots j_q}{\sigma_1\ldots\sigma_q} = \prod_{k=1}^q \pi^{(j_k)}_{\sigma_k}$, then: $\underline{p} \in M^0(\{0,\ldots,n\}^{\mathbb{Z}})$.

Q2.3.4 **[2.3.4]:** The distributions $\underline{p}$ defined in problems [2.3.1], [2.3.2] are in $M_e \cap M_m$.

Q2.3.5 **[2.3.5]:** Show that every $\underline{p} \in M^0(\{0,\ldots,n\}^{\mathbb{Z}})$ is uniquely determined by the sequence $\underline{p} = \left(p\binom{a\ldots b}{\sigma_a\ldots\sigma_b}\right)$ obtained prescribing the values of $p\binom{J}{\underline{\sigma}}$ only for the intervals $J = \{a, a+1, \ldots, b-1, b\}$, for any $a, b \in \mathbb{Z}$ with $a < b$.

Q2.3.6 **[2.3.6]:** Suppose given non-negative numbers $p\begin{pmatrix} a & a+1\ldots b \\ \sigma_a & \sigma_{a+1}\ldots\sigma_b \end{pmatrix}$ for $a,\ b \in \mathbb{Z}$, for every $a < b$, and $(\sigma_a, \sigma_{a+1}, \ldots, \sigma_b) \in \{0,\ldots,n\}^{(b-a+1)}$ so that

$$\sum_{\sigma_a\ldots\sigma_b} p\begin{pmatrix} a\ldots b \\ \sigma_a\ldots\sigma_b \end{pmatrix} = 1,$$

$$\sum_{\sigma_a} p\begin{pmatrix} a & a+1\ldots b \\ \sigma_a & \sigma_{a+1}\ldots\sigma_b \end{pmatrix} = p\begin{pmatrix} a+1\ldots b \\ \sigma_{a+1}\ldots\sigma_b \end{pmatrix},$$

$$\sum_{\sigma_b} p\begin{pmatrix} a\ldots b-1 b \\ \sigma_a\ldots\sigma_{b-1}\sigma_b \end{pmatrix} = p\begin{pmatrix} a\ldots b-1 \\ \sigma_a\ldots\sigma_{b-1} \end{pmatrix},$$

then there exists a unique measure m on $\{0,\ldots,n\}^{\mathbb{Z}}$ such that

$$m(C^{a\ldots b}_{\sigma_a\ldots\sigma_b}) = p\begin{pmatrix} a\ldots b \\ \sigma_a\ldots\sigma_b \end{pmatrix}.$$

(*Hint:* Note that in the proof of proposition (2.3.1) there appear only J having the form $(a, a+1, \ldots, b-1, b)$).

Q2.3.7 **[2.3.7]:** (*Perron–Frobenius theorem*)
Let $T_{\sigma\sigma'}$ be a $(n+1)\times(n+1)$ matrix with strictly positive entries. Show that T and its transpose T^* admits an eigenvalue $\lambda > 0$ with eigenvector $\underline{\pi}$ or $\underline{\pi}^*$ with positive components. Show that the construction suggested for this proof leads to two $(n+1)$–ples $\underline{\pi}$, $\underline{\pi}^*$ such that $T\underline{\pi} = \lambda\underline{\pi}$, and $T^*\underline{\pi}^* = \lambda\underline{\pi}^*$, $\Sigma_\sigma \pi_\sigma \pi^*_\sigma = 1$. (*Hint:* Apply the fixed point theorem, see [DS58], p. 453, to the map

$$(\pi_0, \ldots, \pi_n) \to \Big(\sum_{\sigma'} T_{\sigma\sigma'}\pi_{\sigma'} / \sum_{\sigma',\sigma''} T_{\sigma'\sigma''}\pi_{\sigma''}\Big)_{\sigma=0,1,\ldots,n}$$

defined on the $(n+1)$–ples $\underline{\pi}$ of non-negative numbers such that $\sum_{\sigma=0}^n \pi_\sigma = 1$: the fixed point π^* defines λ and the corresponding eigenvector π^* of T^*. Then apply it again to the map defined by

$$(\mu_0, \ldots, \mu_n) \to \big(\lambda^{-1}\sum_{\sigma'} T_{\sigma'\sigma}\,\mu_{\sigma'}\big)_{\sigma=0,\ldots,n}$$

on the $(n+1)$–ples $\underline{\mu}$ of non-negative numbers such that $\sum_{\sigma=0}^n \pi_\sigma\mu_\sigma = 1$ obtaining the eigenvector π.)

Q2.3.8 **[2.3.8]:** In the context of problem [2.3.7] let

$$p\begin{pmatrix} a & a+1\ldots b-1 & b \\ \sigma_a & \sigma_{a+1}\ldots\sigma_{b-1} & \sigma_b\end{pmatrix} = \pi^*_{\sigma_a}\lambda^{-1}T_{\sigma_a\sigma_{a+1}}\,\ldots\lambda^{-1}T_{\sigma_{b-1}\sigma_b}\,\pi_{\sigma_b},$$

and using the results of problem [2.3.6] show the existence of $m \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ such that $m(C^{a\ldots b}_{\sigma_b\ldots\sigma_b}) = p\begin{pmatrix} a\ldots b \\ \sigma_a\ldots\sigma_b\end{pmatrix}$.

Q2.3.9 **[2.3.9]:** Let $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ and let $\underline{p}_1$, $\underline{p}_2 \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ be two ergodic distributions; assume that $\underline{p} = a\underline{p}_1 + (1-a)\underline{p}_2$, *i.e.* that $p\binom{J}{\underline{\sigma}} = ap_1\binom{J}{\underline{\sigma}} + (1-a)p_2\binom{J}{\underline{\sigma}}$, with $0 < a < 1$. Let $\underline{\sigma}_1$ and $\underline{\sigma}_2$ be two sequences whose frequency distributions are precisely $\underline{p}_1$ and $\underline{p}_2$ respectively. Let $N_k = k$, and form the sequence obtained by writing the first $[a\,N_1]$ elements of $\underline{\sigma}_1$ followed by the first $[(1-a)N_1]$ elements of $\underline{\sigma}_2$ followed by the successive $[aN_2]$ of $\underline{\sigma}_1$, followed by the successive $[(1-a)N_2]$ elements of $\underline{\sigma}_2$ *etc.* Show that the sequence $\underline{\sigma}$ so obtained has distribution $\underline{p}$.

Q2.3.10 **[2.3.10]:** (*Perron–Frobenius spectral gap*)
Consider problems [2.3.7] and [2.3.8]. Show that if $T_{\sigma\sigma'} > 0$, then π and π^* are the only eigenvectors with components ≥ 0 and they have all components positive. The corresponding eigenvalues are maximal in the sense that every other eigenvalue is necessarily smaller in absolute value, and it is smaller at least by a factor $(1-e^{-2c})$, with $e^{-c} \stackrel{def}{=} \min_{\sigma,\sigma',\sigma''}\frac{T_{\sigma\sigma''}}{T_{\sigma'\sigma''}}$, that we call "gap". (*Hint:* Let π^*, π be as in problem [2.3.7]. Check that $e^{-c} \le \frac{((\lambda^{-1}T^k)f)_\sigma}{((\lambda^{-1}T^k)f)_{\sigma'}} \le e^c$ for all $k \ge 1$ and for all $f \ne 0$ with $f_\sigma \ge 0$. Hence if 1 is the vector with components all equal to 1 it will be $e^{-c} \le ((\lambda^{-1}T)^k 1)_\sigma \le e^c$ as it follows from $((\lambda^{-1}T)1)_\sigma \le e^c((\lambda^{-1}T)1)_{\sigma'}$ by multiplying both sides by $\pi^*_{\sigma'}$ or by π^*_σ and summing over σ' or over σ. Furthermore if $g_\sigma \ge 0$ it follow that $((\lambda^{-1}T)g)_\sigma \ge e^{-c}(\pi^*\cdot g) \ge e^{-2c}(\pi^*\cdot g)\pi_\sigma$ *componentwise.* Hence if $g_\pm \stackrel{def}{=} (|g| \pm g)/2$ componentwise and if $(\pi^*\cdot g) = 0$ (hence $(\pi^*\cdot g_+) = (\pi^*\cdot g_-)$ it is $(\lambda^{-1}T)g = (\lambda^{-1}T)(g_+ - e^{-2c}(\pi^*\cdot g_+)\,\pi) - (\lambda^{-1}T)(g_- - e^{-2c}(\pi^*\cdot g_-)\,\pi)$, componentwise, hence

$$|(\lambda^{-1}T)g| \le (\lambda^{-1}T)(g_+ - e^{-2c}(\pi^*\cdot g_+)\,\pi) + (\lambda^{-1}T)(g_- - e^{-2c}(\pi^*\cdot g_-)\,\pi),$$

so that, multiplying the components by those of π^* and summing, $(\pi^*\cdot|(\lambda^{-1}T)g|)\le (1-e^{-2c})\,(\pi^*\cdot|g|)$ which implies by recursion that $(\pi^*\cdot|(\lambda^{-1}T)^k g|)\le(1-e^{-2c})^k(\pi^*\cdot|g|)\le(1-e^{-2c})^k\max_\sigma|g_\sigma|$. Therefore given any f it is $f=(\pi^*\cdot f)\,h+(f-(\pi^*\cdot f)\,\pi)$ and $|(\lambda^{-1}T)^k f-(\pi^*\cdot f)\,\pi|\le 2(1-e^{-2c})^k\max_\sigma|f_\sigma|e^c$ which implies uniqueness of π,π^* as well as the statement about the spectral gap).

Q2.3.11 **[2.3.11]:** Show that under the hypotheses of problem [2.3.10] the measure m defined in problem [2.3.8] is mixing and find a quantitative estimate of the mixing rate in terms of the Perron–Frobenius gap in problem [2.3.10]. (*Hint:* Show that it is such on the cylinders of the form $C=C^{a\;a+1\ldots b-1\;b}_{\sigma_a\sigma_{a+1}\ldots\sigma_{b-1}\sigma_b}$, *i.e.* $\lim_{k\to\infty}m(C\cap S^kC')=m(C)m(C')$ if S is the translation on $\{0,\ldots,n\}^{\mathbb{Z}}$ and then use the hint to problem [2.3.6].)

Q2.3.12 **[2.3.12]:** (*Perron–Frobenius theorem for mixing matrices*)
The conclusions of [2.3.10] and [2.3.11] hold even if one only assumes that there exists $N\ge1$ such that $(T^N)_{\sigma\sigma'}>0$, in such case one says that the matrix T is *mixing.* (*Hint:* It suffices to note that every eigenvector of T is such also for T^N and then apply the result of [2.3.10] to T^N, *etc.*)

Q2.3.13 **[2.3.13]:** (*Markov processes*)
Show that the formula in problem [2.3.8] can also be written

$$p(\sigma_a)\prod_{i=a}^{b-1}p_{\sigma_i\sigma_{i+1}},$$

with $p(\sigma)=\pi^*_\sigma\pi_\sigma$ and $p_{\sigma\sigma'}=\pi_\sigma^{-1}\lambda^{-1}T_{\sigma\sigma'}\pi_{\sigma'}$. And check that $\sum_\sigma p(\sigma)=1$, $\sum_{\sigma'}p_{\sigma\sigma'}=1$. Because of the possibility of writing $p\binom{a\ldots b}{\sigma_a\ldots\sigma_b}$ as above the class of the measures thus obtained is the class of the *mixing Markov processes* with $\{0,\ldots,n\}$ as space of states.

Q2.3.14 **[2.3.14]:** Deduce, via a limiting procedure, that the statements of problem [2.3.7] also hold if T has non-negative matrix elements rather than strictly positive. Show (via examples) that, however, uniqueness of the eigenvalue with largest modulus does not hold.

Bibliographical note to §2.3

The fundamental notions of measure theory in the form used here can be derived from Ch. I, II, V of the book by Parthasarathy, [Pa67]. The quoted chapters also clarify various other notions that we shall introduce and use in the following.

§2.4 The ergodic decomposition

The following proposition shows that every invariant measure could be thought of as " resultant of a convex combination" of ergodic measures: the result is particularly clear when Ω is compact, S is a homeomorphism and the measure is a Borel measure.

We continue to use the symbols introduced in the previous sections. In particular, given a partition $\mathcal{P}$ we denote by $\widehat{\Omega}$ the closure of the set of sequences that are histories of points on the partition $\mathcal{P}$: we have called this set the set of the "$(\mathcal{P},S)$–symbolic motions", cf. definition (1.4.2); and we denote by τ the translation acting on the sequences of $\widehat{\Omega}$. If $J\subset\mathbb{Z}$ is a finite subset of times and $\underline{\sigma} = (\sigma_j)_{j\in J}$ is a string of labels in $\{0,\ldots,n\}$, we denote by $C^J_{\underline{\sigma}}$ the set of infinite sequences which at the times $j \in J$ agree with σ_j.

P2.4.1 **(2.4.1) Proposition:** (Ergodic decompositions)
(i) Let (Ω, S) be an invertible topological dynamical system and let $\mathcal{P} = \{P_0,\ldots,P_n\}$ be a partition of Ω. Let m be a translation invariant probability measure on the $(\mathcal{P}, S)$–symbolic motions $\widehat{\Omega}$.
Then there exists a unique probability measure ϑ_m defined on the Borel sets
N2.4.1 *of $\widehat{M}_e \subset M(\{0,\ldots,n\}^{\mathbb{Z}})$*[1]*, i.e. "concentrated on the τ–ergodic measures" such that m is the "center of mass" of a mass distributed over the ergodic measures ω or, analytically,*

e2.4.1
$$m(C^J_{\underline{\sigma}}) = \int_{\widehat{M}_e} \vartheta_m(d\omega)\,\omega(C^J_{\underline{\sigma}}) \tag{2.4.1}$$

for all $J \subset \mathbb{Z}$, for all $\underline{\sigma} \in \{0,\ldots,n\}^J$.
N2.4.2 *(ii) If Ω is a compact metric space then $\mathcal{M}_e(\Omega, S) \subset \mathcal{M}(\Omega)$ is a Borel set*[2] *in the weak topology induced by $C(\Omega)$ on $\mathcal{M}(\Omega)$, (which is compact).*
Furthermore if $\mu \in \mathcal{M}(\Omega, S)$ there is a Borel probability measure π_μ on $\mathcal{M}_e(\Omega, S)$, i.e. "concentrated on the S–ergodic measures ω on Ω", such that

e2.4.2
$$\mu(f) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\,\omega(f) \qquad \text{for all } f \in C(\Omega), \tag{2.4.2}$$

and π_μ is unique.

Remarks: (1) This is a theorem due to Ruelle, related to Choquet's theory of simplexes, see [Ru69].
(2) Proposition (2.4.1) implies that (2.4.1) holds also by replacing $C^J_{\underline{\sigma}}$ by any Borel set $E \subset \widehat{\Omega}$, see corollary (2.4.1). Hence (2.4.1) implies that if m is an arbitrary translation invariant probability distribution on $\widehat{\Omega}$ then $m(\widehat{\mathcal{E}}) = 1$ because by (i) of proposition (2.3.2) one has $\omega(\widehat{\mathcal{E}}) = 1$ for all $\omega \in \widehat{M}_e$. This shows that the problem of studying the statistical properties of motions as defined in remark (3) to definition (2.3.4) is "reduced" to studying the motions of the ergodic points of $\widehat{\mathcal{E}}$.
(3) From proposition (2.4.1) we see that the motions of the points in $\widehat{\mathcal{E}}''$ are not really different from those of $\widehat{\mathcal{E}}$. On the other hand the motions of $\widehat{\mathcal{E}}'$ are exceptional motions. Although they are often of remarkable interest, when existing, such motions do not have direct "statistical relevance" in the sense that if we choose randomly points of $\widehat{\Omega}$ with respect to an arbitrary invariant measure m we have m–probability zero to find a point of $\widehat{\mathcal{E}}' \cup \widehat{\mathcal{E}}''$: this is a corollary of proposition (2.4.1), see corollary (2.4.1).

---

[1] See definition (2.3.4).

[2] See definition (2.3.3).

(4) One says that the ergodic points of $\widehat{\mathcal{E}}$ contain all the possible statistical informations about the symbolic motions of the dynamical system.
(5) It is useful to remark explicitly that the extremal points of $\mathcal{M}$ are points
N2.4.3 in $\mathcal{M}_e$: this can be checked by contradiction.[3] And viceversa the points in
N2.4.4 $\mathcal{M}_e$ are extremal in $\mathcal{M}$.[4]

Proof: The set $\widehat{\Omega}$ is closed in $\{0,\ldots,n\}^{\mathbb{Z}}$ and it is τ–invariant: hence the dynamical system $(\widehat{\Omega},\tau)$ verifies the hypothesis of (ii); furthermore the characteristic functions of the cylinders are continuous on $\widehat{\Omega}$ so that (ii) implies (i). One sees that, in turn, (ii) is a consequence of (i) if S is expansive on $\mathcal{P}$, by using proposition (2.3.3).

We shall choose to prove (ii).

The set of the ergodic distributions $\mathcal{M}_e(\Omega,S)$ is a Borel set. In fact note that the S–invariant distribution m is in $\mathcal{M}_e(\Omega,S)$ if it belongs to the closed set $\mathcal{M}(\Omega,S)$ and if, furthermore, for a denumerable dense set $\{f_i\}_{i\in\mathbb{Z}_+}$ of
N2.4.5 continuous functions [5] one has

e2.4.3
$$N^{-1}\sum_{j=0}^{N-1}[m(f_h S^{-j} f_k)-m(f_h)m(f_k)] \xrightarrow[N\to\infty]{} 0 \quad \text{for all } h,k\in\mathbb{Z}_+. \tag{2.4.3}$$

The functions in square bracket, as functions of $m\in\mathcal{M}(\Omega,S)$, are continuous. This implies that the set $\mathcal{M}_e(\Omega,S)$ is a Borel set.

Therefore we only need to show the validity of (2.4.2). To define π_μ it will suffice to give the values of the integrals

e2.4.4
$$\int \pi_\mu(d\omega)\,\omega(f_1)\ldots\omega(f_q) \tag{2.4.4}$$

for each choice of $q\geq 1$ functions in $C(\Omega)$. Indeed the functions
N2.4.6 $\omega\to\omega(f_1)\ldots\omega(f_q)$ (as we vary $f_1,\ldots,f_q$) are a fundamental set in[6]
N2.4.7 $C(\mathcal{M}(\Omega,S))$.[7]

Then consider the quantities

e2.4.5
$$\lim_{N_2,\ldots,N_q\to\infty}(N_2N_3\ldots N_q)^{-1}\sum_{k_2=0}^{N_2-1}\ldots\sum_{k_q=0}^{N_q-1}\mu(f_1\cdot S^{-k_2}f_2\cdot S^{-k_3}f_3\cdot\ldots\cdot S^{-k_q}f_q). \tag{2.4.5}$$

---

[3] If $\overline{\mu}$ was extremal in $\mathcal{M}$ but not ergodic, one could find a τ–invariant Borel set E such that $0<\overline{\mu}(E)<1$ and one could define the two probability distributions $\mu_1(R)=\overline{\mu}(R\cap E)/\overline{\mu}(E)$ and $\mu_2(R)=\overline{\mu}(R\cap E^c)/(1-\overline{\mu}(E))$ and, hence, $\overline{\mu}=\overline{\mu}(E)\mu_1+(1-\overline{\mu}(E))\mu_2$ with $\mu_1\neq\mu_2$, against the extremality of $\overline{\mu}$.

[4] Assume that $\overline{\mu}$ is ergodic and there are two invariant measures μ_1 and μ_2 in $\mathcal{M}$ such that $\overline{\mu}=a\mu_1+(1-a)\mu_2$. Then both μ_1 and μ_2 are absolutely continuous with respect to $\overline{\mu}$ and therefore they must coincide with it by its ergodicity.

[5] cf. problem [2.2.44]. Here the function $S^{-j}f$ is the function $x\to f(S^jx)$.

[6] This means that they span a dense linear space.

[7] We use here Riesz' representation theorem for the dual of $C(A)$ as $\mathcal{M}^0(A)$, when A is a compact Hausdorff space. See [DS58], p. 265.

By Birkhoff's theorem the limits (2.4.5) always exist. If one admits that (2.4.2) holds then the expressions (2.4.4) can easily be computed as the limits (2.4.5). This follows immediately by inserting the expression under the limit sign in (2.4.2) and passing to the limits under the integral sign (by applying the criterion of dominated convergence and the ergodicity of the measures ω in the support of π_μ).

Therefore we proceed to show that (2.4.5) can be really used to define (2.4.4) and, hence, π_μ. For this purpose it is necessary to show that given an arbitrary "polynomial"

e2.4.6
$$\omega \to Q(\omega) = \sum_{i=1}^{R} c_i \prod_{k=1}^{q_i} \omega(f_{i,k}), \tag{2.4.6}$$

with $f_{i,k} \in C(\Omega)$ and with complex coefficients, its "integral" $I(Q)$, computed by means of the formula in (2.4.4) with the integral defined by (2.4.5), is well defined. This means that it has to be independent of the particular representation of Q in terms of the monomials, and furthermore that it satisfies

e2.4.7
$$|I(Q)| \le \sup_{\omega \in \mathcal{M}(\Omega,S)} |Q(\omega)|. \tag{2.4.7}$$

Establishing (2.4.7) will show, first of all, the independence of the $I(Q)$ from the particular representation (2.4.6) of Q. In fact if Q_1 and Q_2 are two different representations of the same polynomial then, if $\tilde{Q} = Q_1 - Q_2$, $|I(\tilde{Q})| \le \sup_{\omega\in\mathcal{M}(\Omega,S)} |\tilde{Q}(\omega)| = 0$. By applying Riesz' representation theorem of to $C(\mathcal{M}(\Omega,S))$, see [DS58], p. 265, the existence of π_μ and the validity of (2.4.2) will follow with $\mathcal{M}(\Omega,S)$ instead of $\mathcal{M}_e(\Omega,S)$. This will be a first step towards the proof of (2.4.2).

To check (2.4.7) it will be enough to consider the case $q_i = q$ for all $i = 1,\ldots,R$. Note that

e2.4.8
$$\begin{aligned}
I(Q) = &\lim_{N_2\to\infty} \ldots \lim_{N_q\to\infty} (N_2\ldots N_q)^{-1} \sum_{k_2,\ldots,k_q} \sum_{i\ge1} c_i \cdot \\
&\cdot \mu(f_{i,1} S^{-k_2} f_{i,2} \ldots S^{-k_q} f_{i,q}) = \lim_{N_2\to\infty} \ldots \lim_{N_q\to\infty} (N_2\ldots N_q)^{-1} \cdot \\
&\cdot \sum_{k_2,\ldots,k_q} \mu\big(\sum_{i\ge1} c_i f_{i,1} S^{-k_2} f_{i,2} \ldots S^{-k_q} f_{i,q}\big) = \\
&= \mu\big(\sum_{i\ge1} c_i f_{i,1}\, \overline{f}_{i,2} \ldots \overline{f}_{i,q}\big),
\end{aligned} \tag{2.4.8}$$

where Birkhoff's theorem is applied to the system (Ω, S, μ) denoting

e2.4.9
$$\overline{f}_{i,h}(x) = \lim_{N\to\infty} N^{-1} \sum_{k=0}^{N-1} f_{i,h}(S^k x). \tag{2.4.9}$$

Since μ and all functions appearing in every addend in the last member of (2.4.8) are S–invariant except the first, we can replace it too with its average, deducing by (2.4.8)

e2.4.10
$$I(Q)=\mu\big(\sum_{i\ge1} c_i\,\overline{f}_{i,1}\,\overline{f}_{i,2}\dots\overline{f}_{i,q}\big). \tag{2.4.10}$$

Let us now remark that there exists a set V, which is S–invariant and of μ–measure zero, outside which the limits

e2.4.11
$$\overline{f}(x)=\lim_{N\to\infty}N^{-1}\sum_{k=0}^{N-1}f(S^kx)\qquad\text{for all } f\in C(\Omega),\ x\notin V, \tag{2.4.11}$$

exist. Indeed if $f_1,\ f_2,\dots$ is a denumerable dense set in $C(\Omega)$ (that exists because Ω is metric compact) there must exist an S–invariant set outside of which the limits (2.4.11) exist for all functions $f_1,f_2,\dots$ simultaneously and $\mu(V)=0$. The density in norm of the set $f_1,f_2,\dots$ in $C(\Omega)$ implies then that, outside V, the limits (2.4.11) exist for every $f\in C(\Omega)$.

For each $x\in\Omega\setminus V$ the limit (2.4.11) defines a functional on $C(\Omega)$ that is obviously positive, normalized to 1 and S–invariant: $(S^{-1}\,\overline{f})(x)=\overline{f}(x)$, for all $f\in C(\Omega)$. Then such functional is generated (by Riesz' representation theorem) by an S–invariant probability measure on Ω, *i.e.* by a element ω of $\mathcal{M}(\Omega,S)$.

Hence (2.4.10) immediately implies (2.4.7) since μ is a probability measure.

From (2.4.7) one deduces, then, existence of a probability measure π_μ on the Borel sets of $\mathcal{M}(\Omega,S)$ such that

e2.4.12
$$I(Q)=\int_{\mathcal{M}(\Omega,S)}Q(\omega)\pi_\mu(d\omega). \tag{2.4.12}$$

Hence selecting $Q(\omega)=\omega(f)$, with $f\in C(\Omega)$, from (2.4.10) we deduce then

e2.4.13
$$\mu(f)=\int_{\mathcal{M}(\Omega,S)}\omega(f)\,\pi_\mu(d\omega). \tag{2.4.13}$$

Selecting, instead, $Q(\omega)=\omega(f_1S^{-k_2}f_2)\,\omega(f_3)\dots\omega(f_q)$, with $f_1,\,f_2,\dots,f_q\in C(\Omega)$, one deduces from (2.4.8), (2.4.10)

e2.4.14
$$\begin{aligned}N_2^{-1}&\sum_{k_2=0}^{N_2-1}\int\omega(f_1S^{-k_2}f_2)\,\omega(f_3)\dots\omega(f_q)\,\pi_\mu(d\omega)=\\ &=\lim_{N_3\to\infty}\dots\lim_{N_q\to\infty}N_2^{-1}\sum_{k_2=0}^{N_2-1}\cdot\\ &\cdot(N_3\dots N_q)^{-1}\sum_{k_3,\dots,k_q}\mu(f_1S^{-k_2}f_2S^{-k_3}f_3\dots S^{-k_q}f_q),\end{aligned}\tag{2.4.14}$$

and, in the limit as $N_2 \to \infty$, the first member of (2.4.14) converges to

e2.4.15
$$\int \omega(f_1\,\overline{f}_2)\omega(f_3)\ldots\omega(f_q)\,\pi_\mu(d\omega), \tag{2.4.15}$$

because, by Birkhoff's theorem applied to (Ω, S, ω), we can apply the dominated convergence criterion.

The second member of (2.4.14) converges, by (2.4.8) and (2.4.10), when $N_2 \to \infty$ to

e2.4.16
$$\int \omega(f_1)\omega(f_2)\ldots\omega(f_q)\,\pi_\mu(d\omega), \tag{2.4.16}$$

hence for all $f_3, \ldots, f_q,\ f_1,\ f_2 \in C(\Omega)$

e2.4.17
$$\int_{\mathcal{M}(\Omega,S)} [\omega(f_1\,\overline{f}_2) - \omega(f_1)\omega(f_2)]\,\omega(f_3)\ldots\omega(f_q)\,\pi_\mu(d\omega) = 0. \tag{2.4.17}$$

By the density of the polynomials in ω in the space $C(\mathcal{M}(\Omega, S))$ the relation (2.4.17) implies that $\omega(f_1\,\overline{f}_2) = \omega(f_1)\omega(f_2)$, for all $f_1,\ f_2$ and π_μ–almost everywhere. Again, since there exists a denumerable set of continuous functions dense in $C(\Omega)$, the relation $\omega(\,\overline{f}_1 f_2) = \omega(f_1)\omega(f_2)$ can be assumed valid for $\omega \notin W$, with the set W independent of f_1 and f_2 and of zero π_μ measure: $\pi_\mu(W) = 0$.

It is however clear that if $\omega(\,\overline{f}_1 f_2) = \omega(f_1)\omega(f_2)$, for all $f_1,\ f_2 \in C(\Omega)$, the measure ω is ergodic and hence $\pi_\mu(\mathcal{M}_e(\Omega, S)) = 1$.

The uniqueness of π_μ follows from the above analysis and from the uniqueness of the measures built by means of the above metioned Riesz' representation theorem. ■

Noting that (2.4.1) implies its validity also for all Borel sets of $\widehat{\Omega}$ one can ask if, in the hypotheses of (ii) of proposition (2.4.1), the relation

e2.4.18
$$\mu(E) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\omega(E) \tag{2.4.18}$$

holds at least for suitably chosen $E \subset \Omega$. The following proposition gives an answer to this question.

P2.4.2 **(2.4.2) Proposition:** *If (Ω, S) is an invertible topological dynamical system and if $\mathcal{P} = \{P_0, \ldots, P_n\}$ is a topological partition of Ω on which S is expansive then (2.4.18) holds for every $E \in \mathcal{B}_0 = \{$algebra of sets generated by all closed sets$\}$.*

Proof: Starting with a given measure $\mu \in \mathcal{M}(\Omega, S)$ and by applying proposition (2.3.3) we can construct the measure $B\,\mu \in M(\widehat{\Omega})$ that has support on $\Sigma\Omega$ (we use here the notations and proposition (2.3.3) as well as corollary (2.3.1), to insure measurability of the set $\widehat{M_e}$).

By applying (i) of proposition (2.4.1) to $B\,\mu$ one gets

e2.4.19
$$(B\,\mu)(F) = \int_{\widehat{M}_e} \overline{\vartheta}_\mu(d\omega')\,\omega'(F), \tag{2.4.19}$$

if $\overline{\vartheta}_\mu$ is suitably chosen and if F is a cylinder $C^J_{\underline{\sigma}}$ or if it is an element of the algebra $\widehat{\mathcal{B}}_0$ generated by the cylinders (*i.e.* if F is a finite union of disjoint cylinders).

We first show that $\overline{\vartheta}_\mu(\widehat{M_e}\cap M(\Sigma\Omega))=1$.

If, indeed, the support of $\overline{\vartheta}_\mu$ was not $\widehat{M_e}\cap M(\Sigma\Omega)$, there should exist $\eta>0$ and a subset $\mathcal{E}_\eta\subset\widehat{M_e}\backslash M(\Sigma\Omega)$ consisting of measures ω' such that:

e2.4.20
$$\omega'(\widehat{\Omega}/\Sigma\Omega)=1, \tag{2.4.20}$$

and $\overline{\vartheta}_\mu(\mathcal{E}_\eta)>\eta>0$ (note that $\omega(\widehat{\Omega}/\Sigma\Omega)=0,1$ if $\omega\in\widehat{M_e}$, because ω is ergodic and $\widehat{\Omega}\backslash\Sigma\Omega$ is a τ–invariant Borel set).

Let now $\Gamma\subset\Sigma\Omega$ be a closed subset of $\Sigma\Omega$ such that $(B\mu)(\Gamma)>1-\delta>1-\eta$ and let G_n, $n=1,2,\ldots$ be a sequence of open sets such that $G_{n+1}\subset G_n$, for all n and $\cap_n G_n=\Gamma$ (the sequence exists because π_μ is regular and $\widehat{\Omega}$ is a metric space). We can also suppose that G_n is in $\mathcal{B}_0$: indeed G_n is a union of cylinders that covers the compact set Γ and the cylinders are open. In this case we have, by (2.4.19), (2.4.20),

e2.4.21
$$\begin{aligned}(B\,\mu)(\Gamma)=\lim_{n\to\infty}(B\,\mu)(G_n)=\lim_{n\to\infty}\int_{\widehat{M}_e}\overline{\vartheta}_\mu(d\omega')\omega'(G_n)=\\ =\int_{\widehat{M}_e}\overline{\vartheta}_\mu(d\omega')\omega'(\Gamma)=\int_{\widehat{M}_e\backslash\mathcal{E}_\eta}\overline{\vartheta}_\mu(d\omega')\omega'(\Gamma),\end{aligned} \tag{2.4.21}$$

which implies

e2.4.22
$$1-\delta<(B\mu)(\Gamma)\le 1-\eta, \tag{2.4.22}$$

that contradicts the choice of Γ.

N2.4.8 Then the (2.4.19) can be replaced by [8]

e2.4.23
$$B\,\mu(F)=\int_{B\mathcal{M}_e(\Omega,S)}\overline{\vartheta}_\mu(d\omega')\,\omega'(F), \tag{2.4.23}$$

because, cf. (vii) in proposition (2.3.3), $B\mathcal{M}_e(\Omega,S)=\widehat{M_e}\cap M(\Sigma\Omega)$.

Noting that the first two equalities in (2.4.21) were obtained under the only hypothesis that Γ was closed, (the relation $B\mu(\Gamma)>1-\delta$ has been used only in (2.4.22)), we see that instead of (2.4.23) we can write

e2.4.24
$$B\,\mu(\Gamma)=\int_{B\mathcal{M}_e(\Omega,S)}\overline{\vartheta}_\mu(d\omega')\,\omega'(\Gamma), \tag{2.4.24}$$

for all Γ closed and hence for all $\Gamma\in\mathcal{B}_0$, if $\mathcal{B}_0$ is the algebra of the sets generated by the closed sets.

---

[8] Note that by proposition (2.3.3), by items (vi) and (vii) following it, and by Kuratowsky's theorem $B\mathcal{M}_e(\Omega,S)$ is a Borel set and such is $B\mathcal{E}$, if $\mathcal{E}\subset\mathcal{M}_e(\Omega,S)$ is a Borel set.

Let us set now, for $\varepsilon \subset \mathcal{M}_e(\Omega, S)$ and Borel set, see footnote 8,

e2.4.25
$$\pi_\mu(\varepsilon) = \overline{\vartheta}_\mu(B\varepsilon) \tag{2.4.25}$$

and recall that $AB =$ identity (cf. proposition (2.3.3), (vii)) and $\Sigma^{-1}X^{-1} =$ identity (cf. proposition (2.3.3), (ii)). We find that, if Δ is closed,

e2.4.26
$$\begin{aligned}\mu(\Delta) = \mu(\Sigma^{-1}X^{-1}\Delta) = (B\mu)(X^{-1}\Delta) = \int_{B\mathcal{M}_e(\Omega,S)} \overline{\vartheta}_\mu(d\omega')\omega'(X^{-1}\Delta) = \\ = \int_{B\mathcal{M}_e(\Omega,S)} \overline{\vartheta}_\mu(d\omega')(A\omega')(\Delta) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\,\omega(\Delta),\end{aligned} \tag{2.4.26}$$

having used the continuity statement (i) of proposition (2.3.3) to infer that $X^{-1}(\Delta)$ is closed and hence to apply (2.4.24).

One then deduces from (2.4.26) the identity between the first and the last member in the case in which Δ is open; hence

e2.4.27
$$\mu(E) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\omega(E) \tag{2.4.27}$$

for all $E \in \mathcal{B}_0$.

From (2.4.27) and from the observation that every continuous function on Ω can be obtained as a uniform limit of a sequence of functions which are constant on a finite number of closed or open sets it follows that:

e2.4.28
$$\mu(f) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\,\omega(f) \qquad \text{for all } f \in C(\Omega), \tag{2.4.28}$$

that shows the coincidence of the measure π_μ with the measure that realizes the ergodic decomposition of μ according to (ii) of proposition (2.4.1) (which, precisely, is unique). ■

The above leads us to an important conclusion.

C2.4.1 **(2.4.1) Corollary:** (Typical motions are ergodic)
Let (Ω, S) be an invertible topological system satisfying the hypotheses of proposition (2.4.2). For all stationary distributions $\mu \in \mathcal{M}(\Omega, S)$ one has

e2.4.29
$$\mu(E) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega)\,\omega(E) \tag{2.4.29}$$

for all Borel sets E. In particular $\mu(\mathcal{E}) = 1$.
Specializing the property to the case $(\widehat{\Omega}, \tau)$ where $\widehat{\Omega}$ are the symbolic motions of (Ω, S) observed om a partition $\mathcal{P}$ it follows that $m(\widehat{\mathcal{E}}) = 1$ for all stationary distributions $m \in M(\widehat{\Omega})$.

Remarks: (1) Hence by picking up randomly with a stationary distribution a point $x \in \Omega$ one gets (with probability 1) an ergodic point and by picking

up a symbolic motion in $\widehat{\Omega}$ with any stationary probability distribution on $\widehat{\Omega}$ one gets (with probability 1) an ergodic sequence.
(2) Therefore the probability of finding, by a random choice with a stationary distribution, an exceptional point in $\mathcal{E}'$ or a nonergodic point in $\mathcal{E}''$ is zero.

Proof: One notices that (2.4.29) has been established in proposition (2.4.2) for all sets in the algebra containing the closed sets. However both sides of (2.4.29) define countably additive measures on the Borel sets and the extension of a countably additive measure on an algebra of sets to the smallest σ–algebra containing it is unique. ∎

Problems for §2.4

Q2.4.1 **[2.4.1]:** If π is a permutation of $\{1,\ldots,n\}$ consider the dynamical system $(\{1,\ldots,n\}, S_\pi,\mu)$ where $S_\pi(\{i\})=\pi(\{i\})$, $\mu(\{i\})=1/n$ and find the ergodic decomposition of μ.

Q2.4.2 **[2.4.2]:** (*Simple ergodic decompositions*)
Find the ergodic decomposition of the (normalized) Lebesgue measure in the following dynamical systems:
(a) $\Omega=T^1$, $S\varphi=\varphi+\frac{2\pi}{n}$, n integer
(b) $\Omega=T^2$, $S\underline{\varphi}=\underline{\varphi}+\underline{\omega}$, $\frac{\omega_1}{\omega_2}=\frac{m}{n}$ e $\frac{\omega_1}{2\pi}$ irrational.
(c) $\Omega=T^2$, $S\underline{\varphi}=\underline{\varphi}+\underline{\omega}$, $\omega_1=\frac{2\pi}{n}$, $\frac{\omega_1}{\omega_2}$ irrational and $\frac{\omega_2}{2\pi}$ = irrational.

Q2.4.3 **[2.4.3]:** (*Unique ergodicity of irrational rotations*)
Let $(\omega_1,\ldots,\omega_n,2\pi)$ be rationally independent. Show that $\mathcal{M}(\mathbb{T}^n,S)$, with $S\underline{\varphi}=\underline{\varphi}+\underline{\omega} \bmod 2\pi$ is a set consisting in the single element given by the Lebesgue measure (one say that "the irrational rotations are uniquely ergodic").

Q2.4.4 **[2.4.4]:** (*Ergodic decomposition of a decomposable Markov process*)
Consider problem [2.3.7]. Suppose that $T_{\sigma\sigma'}=0$ if $\sigma\in\{0,\ldots,n_L\}$ and $\sigma'\in\{n_L+1,\ldots,n\}$ or viceversa (*i.e.* T is a "block matrix"). Consider, in the dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$, the ergodic decomposition of the measure μ built by varying the choice of the arbitrary vector $\underline{\pi}$ among the eigenvectors of T and of $\underline{\pi}^*$ among those of T^*. Show that if $T_{\sigma\sigma'}>0$ for the remaining values of the labels σ, σ', then μ can be expressed in terms of just two ergodic measures.

Q2.4.5 **[2.4.5]:** Consider problem [2.3.7] for a generic matrix T: show that as $\underline{\pi}$ and $\underline{\pi}^*$ vary the measures $\mu_{\underline{\pi},\underline{\pi}^*}$ can be decomposed into a finite number of ergodic measures. How many?

Q2.4.6 **[2.4.6]:** (*Consistency between ergodicity of points and of distributions*)
Under the hypotheses of corollary (2.4.1) let $\Delta\subset M_e(\widehat{\Omega})$ be a Borel set and $\pi_m(\Delta)=1$. Let $\mathcal{E}(\Delta)=\{$set of the ergodic points of $\widehat{\Omega}$ whose distribution generates a measure in $\Delta\}$. Then $\mathcal{E}(\Delta)$ is a Borel set and, furthermore, it is $m(\mathcal{E}(\Delta))=1$.

Bibliographical note to §2.4

The theory of the ergodic decomposition exposed here is substantially taken from Ruelle, [Ru66]; see also the book of Ruelle, [Ru69].

CHAPTER III

Entropy and complexity

§3.1 Complexity of motions and entropy

Continuing the analysis of general structural properties of motions of an invertible discrete dynamical system (Ω, S) we shall now discuss the foundations of the notion of complexity of motions on Ω and of its theory.

Let $\mathcal{P}$ be a partition of Ω: the complexity of a motion on Ω as observed on $\mathcal{P}$ will be defined in terms of the complexity of its $(\mathcal{P}, S)$–history. Therefore we begin by discussing the notion of complexity of a sequence $\underline{\sigma}$.

*N*3.1.1 If $\widehat{\underline{\sigma}}$ is a sequence $\widehat{\underline{\sigma}} \in \{0, \ldots, n\}^{\mathbb{Z}}$ of symbols with defined frequencies[1] and if $N > 0$ we consider, as $(\sigma_0 \ldots \sigma_{N-1}) \in \{0, \ldots, n\}^N$ varies, the strings of history homologue to $\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix}$ that "appear" in $\widehat{\underline{\sigma}}$, *i.e.* which have strictly positive frequency of appearance. We shall set

*e*3.1.1
$$\eta_{\rm abs}(\widehat{\underline{\sigma}}\,|\,N) = \{\text{number of distinct } N\text{–length strings that appear in } \widehat{\underline{\sigma}}\}. \tag{3.1.1}$$

It would at first seem natural to identify the size of the *complexity* of $\widehat{\underline{\sigma}}$ with the number

*e*3.1.2
$$\begin{aligned} s_{\rm abs}(\widehat{\underline{\sigma}}) &= \limsup_{N\to+\infty} N^{-1} \log \eta_{\rm abs}(\widehat{\underline{\sigma}}|N) = \\ &= \lim_{N\to\infty} N^{-1} \log \eta_{\rm abs}(\widehat{\underline{\sigma}}|N), \end{aligned} \tag{3.1.2}$$

[1] Cfr. definition (1.4.3).

where the logarithm and the factor N^{-1} are suggested by the expectation that $\eta_{\rm abs}(\underline{\widehat\sigma}|N)$ grows exponentially with N. The fact that

e3.1.3
$$\eta_{\rm abs}(\underline{\widehat\sigma}|N+M)\le\eta_{\rm abs}(\underline{\widehat\sigma}|N)\,\eta_{\rm abs}(\underline{\widehat\sigma}|M)\tag{3.1.3}$$

and $\eta_{\rm abs}(\underline{\widehat\sigma}|N)\le(n+1)^N$ imply the existence of the limit in (3.1.2). Indeed one has

$$f_k\equiv 2^{-k}\log\eta_{\rm abs}(\underline{\widehat\sigma}|2^k),\ \ f_k\le\log(n+1),\ \ f_{k+1}\le f_k,\ \text{ for all}\,k>0,$$

so that $\lim_{k\to\infty}f_k$ exists. Moreover, given k and taking N large one can write $N=m\,2^k+r$ and obtain that

$$\frac1N\log\eta_{\rm abs}(\underline{\widehat\sigma}|N)\le f_k+\frac1N\log\eta_{\rm abs}(\underline{\widehat\sigma}|r),$$

so that eventually $N^{-1}\log\eta_{\rm abs}(\underline{\widehat\sigma}|N)\le f_k+\varepsilon$ for every ε. With a similar reasoning, writing $2^k=m\,N+r$ with k large, one shows that also the inequality

$$\frac1N\log\eta_{\rm abs}(\underline{\widehat\sigma}|N)\ge f_k-\frac1{2^k}\log\eta_{\rm abs}(\underline{\widehat\sigma}|r)$$

holds. This proves that the limit exists.

However, obviously, this is not the only possible definition of complexity of $\underline{\widehat\sigma}$: for example here we put on the same level strings of history $\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ with very different frequency of appearance in $\underline{\widehat\sigma}$, $p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\,\Big|\,\underline{\widehat\sigma}\Big)$.

The following definition takes into some account the possible existence of very many history strings with low frequency of appearance in $\underline{\widehat\sigma}$.

Given $\varepsilon>0$ we shall consider all possible partitions of the strings of length N in two classes $\mathcal C_1$ and $\mathcal C_2$ such that

e3.1.4
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal C_2}p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\Big|\underline{\widehat\sigma}\Big)\le\varepsilon\tag{3.1.4}$$

setting

e3.1.5
$$\eta_\varepsilon(\underline{\widehat\sigma}|N)=\inf_{\mathcal C_1,\mathcal C_2}\{\text{number of elements of }\mathcal C_1\},\tag{3.1.5}$$

and noting that $\eta_\varepsilon(\underline{\widehat\sigma}|N)$ does not decrease as ε decreases, at N fixed, we shall define

e3.1.6
$$s(\underline{\widehat\sigma})=\lim_{\varepsilon\to0}\limsup_{N\to\infty}N^{-1}\log\eta_\varepsilon(\underline{\widehat\sigma}|N).\tag{3.1.6}$$

Note the relation between (3.1.2) and (3.1.6) by writing $s_{\rm abs}(\underline{\widehat\sigma})$ as

e3.1.7
$$s_{\rm abs}(\underline{\widehat\sigma})=\limsup_{N\to\infty}\lim_{\varepsilon\to0}N^{-1}\log\eta_\varepsilon(\underline{\widehat\sigma}|N).\tag{3.1.7}$$

when $\underline{\sigma}$ has defined frequencies.

Having seen the above two possible notions of complexity several others come to mind, depending on whether one wishes to judge of "minor importance" certain strings of history relative to the "importance" of others.

Here is a rather general definition. Let $\underline{V} = \{V_N\}_{N=1}^{\infty}$ be a sequence of functions $(\sigma_0 \ldots \sigma_{N-1}) \to V_N(\sigma_0 \ldots \sigma_{N-1})$ defined respectively on $\{0,\ldots,n\}^N$, $N = 1, 2, \ldots$. Given an infinite sequence $\widehat{\underline{\sigma}} \in \{0,\ldots,n\}^{\mathbb{Z}}$ with defined frequencies, consider the subdivisions $\mathcal{C}_1$, $\mathcal{C}_2$ of $\{0,\ldots,n\}^N$ in two classes such that

e3.1.8
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2} p\Big(\begin{matrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{matrix}\Big|\widehat{\underline{\sigma}}\Big) \le \varepsilon \tag{3.1.8}$$

and set

e3.1.9
$$\eta_\varepsilon(\widehat{\underline{\sigma}}\,|\,N;\underline{V}) = \inf_{\mathcal{C}_1,\mathcal{C}_2} \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_1} e^{-V_N(\sigma_0\ldots\sigma_{N-1})}. \tag{3.1.9}$$

Clearly $\eta_\varepsilon(\widehat{\underline{\sigma}}\,|\,N;\underline{0}) = \eta_\varepsilon(\widehat{\underline{\sigma}}\,|\,N)$. We shall set

e3.1.10
$$s(\widehat{\underline{\sigma}}\,|\,\underline{V}) = \lim_{\varepsilon\to 0}\limsup_{N\to\infty} N^{-1}\log\eta_\varepsilon(\widehat{\underline{\sigma}}\,|\,N;\underline{V}) \tag{3.1.10}$$

N3.1.2 and call this quantity the *complexity with weight* $e^{-\underline{V}}$.[2]

We summarize the above discussion in the following definition, where we shall make use of the symbols introduced so far without discussing them again.

D3.1.1 **(3.1.1) Definition:** (Complexity and entropy of a sequence)
If $\widehat{\underline{\sigma}}$ *is a sequence in* $\{0,\ldots,n\}^{\mathbb{Z}}$ *with defined frequencies and with distri-*
N3.1.3 *bution* $\underline{p}$ [3] *and if* $\underline{V} = \{V_N\}_{N\ge 1}$ *is a sequence of functions on* $\{0,\ldots,n\}^N$, $N = 1\ldots$*, respectively, one defines the* complexity of $\widehat{\underline{\sigma}}$ with weight $\underline{V}$ *the quantity (3.1.10). If* $V_N = 0$ *such a quantity takes the name of* entropy *and it is given by (3.1.6).*

As a first application we evaluate the complexity of the motions of analytically integrable systems, *i.e.* of motions associated with the rotations of a torus $\mathbb{T}^r$ observed on an analytically regular partition $\mathcal{P} = \{P_0,\ldots,P_n\}$, see definition (1.4.1).

P3.1.1 **(3.1.1) Proposition:** *If* $S : \underline{\varphi} \to \underline{\varphi} + \underline{\omega}$ *mod* 2π, $\underline{\omega} \in \mathbb{R}^r$, *is a rotation of* $\mathbb{T}^r$ *and if* $\mathcal{P}$ *is an analytically regular partition of* $\mathbb{T}^r$ *one has*

e3.1.11
$$s(\underline{\sigma}(\underline{\varphi})) = 0 \qquad \forall\,\underline{\varphi} \in \mathbb{T}^r. \tag{3.1.11}$$

One says, in a colorful language, that "the entropy of quasi-periodic motions is zero".

---

[2] There is nothing "mysterious" in the exponential: it is just a way to define the weight so that it is automatically non negative.

[3] cf. remark (2) to definition (2.3.2)

Remarks: (1) As in the case of proposition (2.1.1) if $(\underline{\omega}, 2\pi)$ are rationally independent one can take the atoms of $\mathcal{P}$ as Riemann measurable sets. The higher regularity is necessary to cover the general situation, see (2.1.2). See also problems [3.1.10] and [3.1.11] below to understand the essential role of the regularity hypothesis on the partition $\mathcal{P}$.

Proof: For simplicity we shall consider only the case in which $(\omega_1, \ldots, \omega_r, 2\pi)$ are $r+1$ rationally independent numbers.

In this case the distribution $\underline{p}$ of the history $\underline{\sigma}(\underline{\varphi})$ of $\underline{\varphi} \in \mathbb{T}^r$ is given by

e3.1.12
$$p\begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} = \int_{P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}} \frac{d\underline{\varphi}}{(2\pi)^r}; \tag{3.1.12}$$

cf. proposition (2.1.2) and equation (2.1.9).

We start with some geometric considerations. By the regularity hypothesis on $\mathcal{P}$ the surface area $|\,\partial P_\sigma\,|$ of the elements $P_\sigma \in \mathcal{P}$ is finite and we can set

e3.1.13
$$2L \stackrel{def}{=} \sum_{\sigma=0}^{n} |\,\partial P_\sigma\,| < +\infty. \tag{3.1.13}$$

The sum of the areas of the boundaries of the sets $P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ is, obviously, such that

e3.1.14
$$\begin{aligned}\sum_{\sigma_0\ldots\sigma_{N-1}} |\,\partial P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}\,| &\le \sum_{k=0}^{N-1}\sum_{\sigma=0}^{n} |\,S^{-k}\partial P_\sigma\,| \\ &= \sum_{k=0}^{N-1}\sum_{\sigma=0}^{n} |\,\partial P_\sigma\,| = 2LN,\end{aligned} \tag{3.1.14}$$

because the rotation S is rigid: therefore it does not alter lengths, areas or volumes.

The volume of $P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ defined by (3.1.12) can be bounded, if the diameter of $P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ is small enough with respect to 2π (*e.g.* $< \pi$), by [4]

N3.1.4

e3.1.15
$$p\begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \Gamma_r\, |\,\partial P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}\,|^{\,r/(r-1)}, \tag{3.1.15}$$

where Γ_r is a suitable constant easily expressible in terms of the volume of the unit r–dimensional sphere.

In general it is not possible to construct an analytically regular partition (see definition (1.4.1)) of the torus $\mathbb{T}^r$ composed by sets with small diameter. To make use of (3.1.15) we shall suppose that all the sets of $\mathcal{P}$ but one are, since the beginning, so small that $\mathrm{diam}(P_\sigma) < \pi$. We will call the set with

[4] This is the isoperimetric inequality: in $\mathbb{R}^r$ it is valid for every set, independently on the size of its diameter; on $\mathbb{T}^r$ it is necessary that the set "does not wrap around the torus". For sets with small enough diameter, it follows from the isoperimetric inequality in $\mathbb{R}^r$.

large diameter P_0. The general case can be treated analogously. In this situation we have that diam$\,(P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}) \geq \pi$ only if $\sigma_0 = \ldots = \sigma_N = 0$. In the following argument we can include the set $P^{0\ldots N-1}_{0\ldots 0}$ in the set $\mathcal{C}_1(N)$ for every N, without affecting the validity of the argument.

Let us define, given $\eta > 0$,

e3.1.16
$$\begin{aligned}\mathcal{C}_1(N) &= \Big\{\sigma_0,\ldots,\sigma_{N-1}\,\big|\, p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} > e^{-N\eta}\Big\},\\ \mathcal{C}_2(N) &= \Big\{\sigma_0,\ldots,\sigma_{N-1}\,\big|\, p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \leq e^{-N\eta}\Big\}.\end{aligned} \tag{3.1.16}$$

We get, for all $\gamma > 0$ and taking advantage of the well known idea behind Chebishev's inequality and of (3.1.15),

e3.1.17
$$\begin{aligned}&\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \leq\\ &\leq \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\left(\frac{e^{-N\eta}}{p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}}\right)^{\gamma} \leq\\ &\leq e^{-N\eta\gamma}\,\Gamma_r^{1-\gamma}\sum_{\sigma_0\ldots\sigma_{N-1}} |\,\partial P^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}|^{\,(1-\gamma)r/(r-1)},\end{aligned} \tag{3.1.17}$$

and selecting $\gamma = r^{-1}$, so that $(1-\gamma)r/(r-1) \equiv 1$, equations (3.1.17) and (3.1.14) imply

e3.1.18
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \leq 2LN\,\Gamma_r^{1-\gamma}\,e^{-N\eta\gamma}. \tag{3.1.18}$$

N3.1.5 On the other hand {number of elements in $\mathcal{C}_1(N)$} $\leq e^{+N\eta} + 1$:[5] hence $s(\underline{\sigma}(\underline{\varphi})) \leq \eta$ for all $\eta > 0$ (note that the $s(\underline{\sigma}(\underline{\varphi}))$, as it is defined in (3.1.6), is less than the quantity computed by setting $\varepsilon = e^{-N\eta}$, as we are doing). This means that $s(\underline{\sigma}(\underline{\varphi})) = 0$. ■

The preceding proof does not provide us with an optimal result, nor it is the simplest conceivable: under the same hypothesis we could easily deduce that even $s_{\rm abs}(\underline{\sigma}(\underline{\varphi})) = 0$. However the proof just given is more interesting because it can be easily extended to much more general situations: for instance to those contemplated in the following definition and proposition.

D3.1.2 **(3.1.2) Definition:** *Let (Ω, S) be an invertible dynamical system. Let μ_0 be a probability measure (not necessarily S–invariant) on Ω and let $\mathcal{P}$ be a μ_0–measurable partition of Ω.*

[5] The +1 is here to take into account the set $P^{0\ldots N-1}_{0\ldots N-1}$ which is included in $\mathcal{C}_1$.

(i) Denote by $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ *the symbolic motions* $\widehat{\underline{\sigma}}\in\widehat{\Omega}$*, observed on* $\mathcal{P}$ *with well defined frequencies and with distribution that can be expressed by integrals with respect to* μ_0 *as*

e3.1.19
$$p\Big(\begin{matrix} j_1\ldots j_q \\ \sigma_1\ldots\sigma_q\end{matrix}\mid\widehat{\underline{\sigma}}\Big)=\int_{P^{j_1\ldots j_q}_{\sigma_1\ldots\sigma_q}}\rho(x)\mu_0(dx), \tag{3.1.19}$$

where ρ *and also* $\rho^{-1}\in L_1(\mu_0)$*.*

We shall call the points of $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ symbolic motions which are absolutely continuous with respect to μ_0.

(ii) The points of Ω *whose* $(\mathcal{P},S)$*–histories are in* $\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$ *will be denoted* $\mathcal{E}(S,\mathcal{P},\mu_0)$ *and we shall call them the* points which are $(\mathcal{P},S)$–absolutely continuous with respect to μ_0.

(iii) If Ω *is a compact Riemannian manifold we call the* largest coefficient of expansion of a line element of Ω under the action of S *the quantity*

$$\lambda(S)=\sup_x\frac{\|dSv\|_{V_{Sx}}}{\|v\|_{V_x}},$$

where V_x *is the tangent space to* Ω *at* x *and* $v\in V_x$*.*

Then the following proposition holds.

P3.1.2 **(3.1.2) Proposition:** (Kouchnirenko's theorem)

Let (Ω,S) *be a dynamical system with* Ω *a* C^∞ r*–dimensional compact Riemannian manifold and with* S *a* C^∞ *diffeomorphism of* Ω*. Let* μ_0 *be the volume measure on* Ω *and let* $\mathcal{P}$ *be a* C^∞*–regular partition of* Ω *(see definition (1.4.1)). Let* $\widehat{\underline{\sigma}}\in\widehat{\mathcal{E}}(S,\mathcal{P},\mu_0)$*.*

Then

e3.1.20
$$s(\widehat{\underline{\sigma}})\le r\log\lambda, \tag{3.1.20}$$

where $\lambda=\max(\lambda(S),\lambda(S^{-1}))$*, see definition (3.1.2).*

Remarks: (1) The theorem is remarkable because it shows in a general enough context what at first sight might be surprising: namely the entropy of a motion whose initial datum is randomly selected with a probability distribution which is ergodic and equivalent to the volume measure on Ω *cannot exceed* $r\log\lambda$, no matter how fine the partition $\mathcal{P}$ of Ω is taken, provided $\mathcal{P}$ is a regular partition (of course).

(2) Proposition (3.1.2) and certain modifications of it have remarkable applications to the theory of Hamiltonian systems: in studying the latter in connection with Statistical Mechanics one is often interested in motions whose initial data x are randomly chosen, on the energy surface in phase space, with respect to the volume measure.

Consider for instance a Hamiltonian system and select the initial data with the Liouville distribution on an energy surface, or with a distribution equivalent to it. If the system is ergodic (3.1.19) holds (with $\rho=1$) and therefore the complexity of motions observed, say, at unit time intervals and on regular partitions is, with probability 1, bounded by a geometric constant which is independent on the particular motion considered.

Proof: Proceeding as in the proof of proposition (3.1.1) and assuming for the time being, for simplicity, that the sets $P_0, \ldots, P_n$ have diameter small with respect to the diameter of Ω (if not see footnote 6) and of a size such that for all sets of smaller diameter a generalization of the isoperimetric inequality holds, *i.e.*

e3.1.21
$$\mu_0(P) \le \Gamma \mid \partial P \mid^{r/(r-1)}, \tag{3.1.21}$$

we see that (3.1.14) becomes

e3.1.22
$$\begin{aligned} \sum_{\sigma_0 \ldots \sigma_{N-1}} \mid \partial P^{0 \ldots N-1}_{\sigma_0 \ldots \sigma_{N-1}} \mid &\le \sum_{k=0}^{N-1} \sum_{\sigma=0}^{n} \mid S^{-k} \partial P_\sigma \mid \\ &\le 2L \sum_{k=0}^{N-1} \lambda^k = 2L \frac{\lambda^N - 1}{\lambda - 1}. \end{aligned} \tag{3.1.22}$$

Note that in (3.1.22) one has λ^k rather than $\lambda^{k\,(r-1)}$, as perhaps one might expect, because the conservation of the measure $\rho\,\mu_0$ and the boundedness of ρ and of ρ^{-1} imply that a surface element can at most expand its area by a factor λ.

Hence, by proceeding as in (3.1.17) and (3.1.18), with (3.1.22) instead of (3.1.14), one obtains

e3.1.23
$$\begin{aligned} &\sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_2(N)} p\Big({0 \ldots N-1 \atop \sigma_0 \ldots \sigma_{N-1}} \mid \underline{\sigma} \Big) \le \\ &\quad \le e^{-N\eta/r} \lambda^{-N} \Gamma_r^{1-1/r} \, \|\rho\|_\infty^{1-1/r} 2L(\lambda^N - 1)/(\lambda - 1) \le G e^{-\eta N/r}, \end{aligned} \tag{3.1.23}$$

having set

$$\begin{aligned} \mathcal{C}_1(N) &= \Big\{ \sigma_0 \ldots \sigma_{N-1} \,\big|\, p\Big({0 \ldots N-1 \atop \sigma_0 \ldots \sigma_{N-1}} \mid \underline{\sigma} \Big) > e^{-\eta N} \lambda^{-Nr} \Big\}, \\ \mathcal{C}_2(N) &= \Big\{ \sigma_0 \ldots \sigma_{N-1} \,\big|\, p\Big({0 \ldots N-1 \atop \sigma_0 \ldots \sigma_{N-1}} \mid \underline{\sigma} \Big) \le e^{-\eta N} \lambda^{-Nr} \Big\}, \end{aligned}$$

and having chosen G to be a suitable constant.

Since it is clear that the number of elements of $\mathcal{C}_1(N)$ is smaller than
N3.1.6 $e^{N\eta} \lambda^{rN}$ we deduce $s(\widehat{\underline{\sigma}}) \le r \log \lambda$.[6] ∎

[6] If the sets have large diameter we can always divide them into smaller sets reducing to the case in which there is at most one set $P^{0,\ldots,N-1}_{0,\ldots,0}$ with a large diameter: and we shall add the latter set to $\mathcal{C}_1$ which will have a number of elements bounded by $e^{N\eta} \lambda^{rN} + 1$ and reaching the same conclusions. One makes use of the remark that finer partitions canoot have lower entropy: if $\mathcal{P}'$ is fined than $\mathcal{P}$ then $s(\mathcal{P}, S) \le s(\mathcal{P}', S)$.

After introducing the notion of complexity of motions we shall try of analyze the problem of the actual computation of $s(\underline{\sigma}(x))$.

A first remarkable result is the *theorem of Shannon–McMillan* that is fundamental for the major conceptual clarification that it introduces about the meaning of the entropy notion and for its applications to the theory of codes: we shall touch this subject also at the end of Chapter 10 and in some problems in the coming sections.

Problems for §3.1

Q3.1.1 **[3.1.1]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ such that all strings of length $\le N_0$ have frequencies well defined and positive.

Q3.1.2 **[3.1.2]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ with well defined frequencies.

Q3.1.3 **[3.1.3]:** Construct a sequence in $\{0,1\}^{\mathbb{Z}}$ in which all possible strings of finite length appear at least once but have frequency zero except those with specification $\sigma_0 = \sigma_1 = \ldots = 0$.

Q3.1.4 **[3.1.4]:** (*Finite algorithms are simple*) Let $k > 0$ and $f : \{0, \ldots, n\}^k \to \{0, \ldots, n\}$ and define arbitrarily $\sigma_{-1}, \sigma_{-2}, \ldots$; set, inductively for $i \ge 0$, $\sigma_i = f(\sigma_{i-1}, \ldots, \sigma_{i-k})$. Show that $s(\underline{\sigma}) = 0$: *i.e.* "is not possible construct complex sequences with finite algorithms".

Q3.1.5 **[3.1.5]:** If $\mathcal{P}$ is an analytically regular partition of $\mathbb{T}^r$, $\underline{\omega} \in \mathbb{R}^r$ and $S\underline{\varphi} = \underline{\varphi} + \underline{\omega} \bmod 2\pi$ show that $s_{\rm abs}(\underline{\varphi}) = 0$ (*Hint:* Consider first the case of the irrational rotations).

Q3.1.6 **[3.1.6]:** (*Typical entropy in a Bernoulli shift*)
A randomly selected sequence in $\{0,1\}^{\mathbb{Z}}$ with an equal weights Bernoulli distribution $B(1/2, 1/2)$ has well defined frequencies and entropy $\log 2$: show this by applying (3.1.6) and a combinatorial argument.

Q3.1.7 **[3.1.7]:** Study the problem analogous to problem [3.1.6] for the Bernoulli scheme $B(1/3, 2/3)$. The result is $s(\underline{\sigma}) = -(1/3)\log(1/3) - (2/3)\log(2/3))$. (*Hint:* Let $m(\underline{\sigma}) = N^{-1}\sum_{i-0}^{N} \sigma_i$. For any ε one can set $\mathcal{C}_2 = \{\underline{\sigma} \,|\, |m(\underline{\sigma}) - 2/3| > \delta\}$ chosing δ in a suitable way.)

Q3.1.8 **[3.1.8]:** Generalize problem [3.1.8] to the case of the Bernouilli scheme $B(\pi_0, \ldots, \pi_n)$ on $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$.

Q3.1.9 **[3.1.9]:** Consider the sequence obtained by writing n_0 zeroes followed by n_0 ones followed by n_1 zeroes followed by n_1 ones, *etc*where $(n_0 + n_1 + \ldots + n_k)^{-1} n_{k+1} \xrightarrow[k\to\infty]{} 0$. Show that the sequence has well defined frequencies and compute the associated distribution $\underline{p}$ and the entropy.

Q3.1.10 **[3.1.10]:** (*A complex sequence generated by a circle rotation*)
Let $\underline{\sigma}$ be a sequence in $\{0,1\}^{\mathbb{Z}}$ with well defined frequencies distributed as the Bernoulli scheme $B(1/2, 1/2)$ (see problem [3.1.6]). Let S an irrational rotation of $\mathbb{T}^1 : S\varphi = \varphi + \omega \bmod 2\pi$. Consider the trajectory of the origin $(S^k 0)_{k\in\mathbb{Z}}$. Associate with the point $S^k 0$ the symbol σ_k and construct two Borel sets $P_0 = \{x \,|\, x = S^k 0 \text{ for } k \text{ with } \sigma_k = 0\}$ and $P_1 = \mathbb{T}^1 \backslash P_0$. Show that the entropy of the motion of 0 observed on the partition $\mathcal{P}$ is $\log 2$ and that this does not contradict proposition (3.1.1). (*Hint:* The partition is not analytically regular).

Q3.1.11 **[3.1.11]:** (*Arbitrarily complex sequence generated by a quasi periodic motion*)
If S is an irrational rotation of the circle $\mathbb{T}^1$, given $\varphi \in \mathbb{T}^1$ and $M > 0$, there exists a Borel partition $\mathcal{P}$ of $\mathbb{T}^1$ on which the motion of φ appears with an entropy larger than M. Why this is not in contradiction with proposition (3.1.2)? (*Hint:* At the light of problem [3.1.10] consider a Bernoulli scheme with $\exp M$ symbols.)

Bibliographical note to §3.1

The notion of complexity goes back to Shannon, [Sh49]. Proposition (3.1.2) is inspired from Kouchnirenko's theorem in the version that one finds in [AA68], p. 46.

§3.2 The Shannon–McMillan theorem

It is the following proposition.

P3.2.1 **(3.2.1) Proposition:** (Shannon–McMillan theorem)
Let $\underline{\widehat{\sigma}} \in \{0,\ldots,n\}^{\mathbb{Z}}$ *be a sequence with defined frequencies and with distribution* $\underline{p} = \left(p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\right)$.
(i) The limit

e3.2.1
$$s = \lim_{N\to\infty} -N^{-1}\sum_{\sigma_0\ldots\sigma_{N-1}} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \tag{3.2.1}$$

exists and will be called average entropy *of* $\underline{\widehat{\sigma}}$,
(ii) If $\underline{\widehat{\sigma}}$ *is ergodic then*

e3.2.2
$$s(\underline{\widehat{\sigma}}) = s \tag{3.2.2}$$

and (3.2.1) often yields a rather convenient way of computing $s(\underline{\widehat{\sigma}})$.
(iii) If $\underline{\widehat{\sigma}}$ *is ergodic, given* $\varepsilon > 0$, *there exists* N_ε *such that for all* $N \ge N_\varepsilon$ *the elements of* $\{0,\ldots,n\}^N$ *can be split into two classes* $\mathcal{C}_{1,\varepsilon}(N)$ *and* $\mathcal{C}_{2,\varepsilon}(N)$ *with the properties*

e3.2.3
$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{2,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} < \varepsilon, \tag{3.2.3}$$

e3.2.4
$$\exp((s-\varepsilon)N) \le |\mathcal{C}_{1,\varepsilon}(N)| \le \exp((s+\varepsilon)N), \tag{3.2.4}$$

e3.2.5
$$\exp(-(s+\varepsilon)N) \le p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \exp(-(s-\varepsilon)N), \tag{3.2.5}$$

for every choice of the string $\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)$. *Here* $|\mathcal{C}_{1,\varepsilon}(N)|$ *denotes the number of elements in* $\mathcal{C}_{1,\varepsilon}(N)$.

Remarks: (1) Hence if $\underline{\widehat{\sigma}}$ is ergodic the strings of symbols of large length can be divided into two groups, one with small total probability and another consisting of "few" elements of approximately equal frequency of appearance in $\underline{\widehat{\sigma}}$ (in the rather weak sense of (3.2.5)).
(2) For a better understanding of the meaning of the entropy notion one can remark that, if $\underline{\widehat{\sigma}}$ is ergodic, a quantity s enjoying the properties described in (iii) is necessarily equal to the entropy of $\underline{\widehat{\sigma}}$.
Indeed if $s(\underline{\widehat{\sigma}}) = s'$ it is clear that (3.2.3) and (3.2.4) imply that $s' \le s$.

Then suppose that one has $s' < s$. Then there is an ε with $0 < \varepsilon < \min(1/4, (s - s')/4)$ and one can find ε' and a set $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ in $\{0, \ldots, n\}^N$ such that

e3.2.6
$$|\,\overline{\mathcal{C}}_{1,\varepsilon'}(N)\,| < \exp((s' + \varepsilon)N) \qquad\qquad \text{and}$$
$$\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_{1,\varepsilon'}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} < \varepsilon', \tag{3.2.6}$$

for infinitely many values of $N > N_{\varepsilon'}$, by the definition of entropy (cf. definition (3.1.1) and equation (3.2.5)), at least if ε is small enough.

The set $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ will contain ν elements of $\mathcal{C}_{1,\varepsilon}(N)$, $\nu \ge 0$, together with some others of $\mathcal{C}_{2,\varepsilon}(N)$. Hence $\overline{\mathcal{C}}_{1,\varepsilon'}(N)$ will be obtained from $\mathcal{C}_{1,\varepsilon}(N)$ subtracting from it $(|\,\mathcal{C}_{1,\varepsilon}(N)\,| - \nu)$ elements and adding to it a suitable number of other elements of $\mathcal{C}_{2,\varepsilon}(N)$.

But if N is such that (3.2.5) holds for it and if we note the inclusion $\mathcal{C}_{1,\varepsilon}(N)\backslash(\mathcal{C}_{1,\varepsilon}(N) \cap \overline{\mathcal{C}}_{1,\varepsilon'}(N)) \subset \overline{\mathcal{C}}_{2,\varepsilon'}(N)$, we get

e3.2.7
$$\begin{aligned}\varepsilon \ge &\sum_{(\sigma_0\ldots\sigma_{N-1})\in\mathcal{C}_{1,\varepsilon}(N)\backslash(\mathcal{C}_{1,\varepsilon}(N)\cap\overline{\mathcal{C}}_{1,\varepsilon'}(N))} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} =\\ &= \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} -\\ &- \sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)\cap\overline{\mathcal{C}}_{1,\varepsilon'}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\end{aligned} \tag{3.2.7}$$

and by (3.2.3), (3.2.5) the last difference can be bounded below by

e3.2.8
$$\begin{aligned}\varepsilon \ge 1 - \varepsilon - \nu &\max_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \ge\\ &\ge 1 - \varepsilon - \nu e^{-(s-\varepsilon)N} \ge 1 - \varepsilon - e^{(N(s'+\varepsilon)-N(s-\varepsilon))} \ge\\ &\ge 1 - \varepsilon - e^{(s'-s+2\varepsilon)N} \ge 1 - \varepsilon - e^{-2N\varepsilon},\end{aligned} \tag{3.2.8}$$

being $\nu \le |\,\overline{\mathcal{C}}_{1,\varepsilon}(N)\,| \le e^{(s'+\varepsilon)N}$ and $2\varepsilon < (s - s')/2$, $\varepsilon < 1/4$, at least if N is large enough. Hence the contradiction in (3.2.8) implies that the number s with the property (3.2.3), (3.2.4) and (3.2.5), if it exists, is necessarily the entropy of $\underline{\widehat{\sigma}}$.

(3) It is convenient to break the proof into a few lemmas.

L3.2.1 **(3.2.1) Lemma:** *Let $\underline{p}$ be the distribution of $\underline{\widehat{\sigma}}$ and let $m_{\underline{p}}$ be the probability distribution associated with it. If the function of $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$*

e3.2.9
$$\underline{\sigma} \;\to\; f_N(\underline{\sigma}) = -N^{-1} \log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \tag{3.2.9}$$

is such that the limit

e3.2.10
$$\widetilde{s}(\underline{\sigma}) \stackrel{def}{=} \lim_{N\to\infty} f_N(\underline{\sigma}) \quad in\ L_1(m_{\underline{p}}) \tag{3.2.10}$$

exists, then proposition (3.2.1) follows.

N3.2.1 *Proof of lemma (3.2.1):* By integrating[1] both sides of (3.2.10) we obtain (3.2.1) so that existence of the limit implies (i) of proposition (3.2.1).

The function $\widetilde{s}(\underline{\sigma})$, if the limit defining it exists almost everywhere, is translation invariant, *i.e.* $\widetilde{s}(\underline{\sigma}) = \widetilde{s}(\tau\underline{\sigma})$ holds $m_{\underline{p}}$–almost everywhere. In fact the monotonicity of the logarithm and the compatibility property fulfilled by the p's imply that $\widetilde{s}(\underline{\sigma}) \le \widetilde{s}(\tau\underline{\sigma})$ since

e3.2.11
$$\begin{aligned}\widetilde{s}(\underline{\sigma}) &= \lim_{N\to\infty} -\frac{1}{N}\log p\begin{pmatrix}0\dots N-1\\ \sigma_0\dots\sigma_{N-1}\end{pmatrix} \le \\ &\le \lim_{N\to\infty} -\frac{1}{N}\log p\begin{pmatrix}1\dots N-1\\ \sigma_1\dots\sigma_{N-1}\end{pmatrix} = \widetilde{s}(\tau\underline{\sigma}),\end{aligned} \tag{3.2.11}$$

hence, integrating this inequality and using the invariance of $m_{\underline{p}}$ (which implies the opposite inequality $\widetilde{s}(\tau\sigma) \le \widetilde{s}(\sigma)$), one deduces that $m_{\underline{p}}$–almost everywhere one has $\widetilde{s}(\underline{\sigma}) = \widetilde{s}(\tau\underline{\sigma})$.

Then if $\widehat{\underline{\sigma}}$ is ergodic also $m_{\underline{p}}$ is such, and therefore $\widetilde{s}(\underline{\sigma}) = \text{constant} = s$, $m_{\underline{p}}$–almost everywhere. In this case the convergence in (3.2.10), which implies convergence in $m_{\underline{p}}$-measure, implies also that for all $\varepsilon > 0$, there exist N_ε and, for all $N \ge N_\varepsilon$, a set $E_{\varepsilon,N} \subset \{0,\dots,n\}^{\mathbb{Z}}$ such that

e3.2.12
$$\begin{aligned} &m_{\underline{p}}(E_{\varepsilon,N}) < \varepsilon, \\ &|f_N(\underline{\sigma}) - s| < \varepsilon \qquad \text{for all } \underline{\sigma} \notin E_{\varepsilon,N},\end{aligned} \tag{3.2.12}$$

and it is clear that, since f_N is measurable on the algebra of the cylinders with base $[0, N-1]$, (*i.e.* it only depends on $\sigma_0,\dots,\sigma_{N-1}$), then also $E_{\varepsilon,N}$ can be chosen measurable on this algebra and, therefore, it is a union of cylinders with base $[0, N-1]$. We shall then set

e3.2.13
$$\begin{aligned}\mathcal{C}_{2,\varepsilon}(N) &= \{\sigma_0\dots\sigma_{N-1} | C^{0\dots N-1}_{\sigma_0\dots\sigma_{N-1}} \subset E_{\varepsilon,N}\}, \\ \mathcal{C}_{1,\varepsilon}(N) &= \{\sigma_0\dots\sigma_{N-1} | C^{0\dots N-1}_{\sigma_0\dots\sigma_{N-1}} \cap E_{\varepsilon,N} = \emptyset\} = \\ &= \{0,\dots,n\}^N \backslash \mathcal{C}_{2,\varepsilon}(N),\end{aligned} \tag{3.2.13}$$

and, therefore, (3.2.12) implies (3.2.3) and (3.2.5). Then (3.2.4) follows from (3.2.5) and from $\sum_{\sigma_0\dots\sigma_{N-1}} p\begin{pmatrix}0\dots N-1\\ \sigma_0\dots\sigma_{N-1}\end{pmatrix} = 1$, and (3.2.2) follows from

[1] *i.e.* my multiplying both sides times $p\begin{pmatrix}0\dots N-1\\ \sigma_0\dots\sigma_{N-1}\end{pmatrix}$ and summing over $\sigma_0\dots\sigma_{N-1}$.

(3.2.10) by integrating both sides (indeed the limit (3.2.10) takes place in $L_1(m_{\underline{p}})$ as well), and using the remark (2), above. ■

Hence proposition (3.2.1) follows from the $L_1(m_{\underline{p}})$–convergence of the limit (3.2.10). To prove convergence we consider, for $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$ and $j \ge 1$, the functions

e3.2.14
$$\varphi_j(\underline{\sigma}) = -\log \frac{p\begin{pmatrix} -j\ldots-1 & 0 \\ \sigma_{-j}\ldots\sigma_{-1} & \sigma_0 \end{pmatrix}}{p\begin{pmatrix} -j\ldots-1 \\ \sigma_{-j}\ldots\sigma_{-1} \end{pmatrix}}, \tag{3.2.14}$$

which are non-negative (possibly $+\infty$) and $m_{\underline{p}}$–measurable.

L3.2.2 **(3.2.2) Lemma:** *If the limit*

e3.2.15
$$\varphi(\underline{\sigma}) = \lim_{j\to\infty} \varphi_j(\underline{\sigma}) \qquad \textit{in} \quad L_1(m_{\underline{p}}) \tag{3.2.15}$$

exists, then proposition (3.2.1) follows.

Remark: In fact we shall also show that such a limit is also reached $m_{\underline{p}}$–almost everywhere.

Proof of lemma (3.2.2): Consider the following identity, for $N \ge 2$,

e3.2.16
$$\begin{aligned} &-N^{-1}\log p\begin{pmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1} \end{pmatrix} = \\ &= -N^{-1}\sum_{j=0}^{N-2} \log \frac{p\begin{pmatrix} 0\ldots j+1 \\ \sigma_0\ldots\sigma_{j+1} \end{pmatrix}}{p\begin{pmatrix} 0\ldots j \\ \sigma_0\ldots\sigma_j \end{pmatrix}} - N^{-1}\log p\begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix}, \end{aligned} \tag{3.2.16}$$

and note that $-N^{-1}\log p\begin{pmatrix} 0 \\ \sigma_0 \end{pmatrix} \xrightarrow[N\to\infty]{} 0$ both in $L_1(m_{\underline{p}})$ and $m_{\underline{p}}$–almost

N3.2.2 everywhere[2].

Furthermore the sum in (3.2.16) can be written, when $N \to \infty$,

e3.2.17
$$N^{-1}\sum_{j=1}^{N-1} \varphi_j(\tau^j\underline{\sigma}), \tag{3.2.17}$$

because, by the translation invariance of $\underline{p}$,

e3.2.18
$$\varphi_j(\tau^j\underline{\sigma}) = -\log \frac{p\begin{pmatrix} -j\ldots-1 & 0 \\ \sigma_0\ldots\sigma_{j-1} & \sigma_j \end{pmatrix}}{p\begin{pmatrix} -j\ldots-1 \\ \sigma_0\ldots\sigma_{j-1} \end{pmatrix}} = -\log \frac{p\begin{pmatrix} 0\ldots j \\ \sigma_0\ldots\sigma_j \end{pmatrix}}{p\begin{pmatrix} 0\ldots j-1 \\ \sigma_0\ldots\sigma_{j-1} \end{pmatrix}}. \tag{3.2.18}$$

[2] Remark that if $p\begin{pmatrix} 0 \\ \bar\sigma_0 \end{pmatrix} = 0$ for some $\bar\sigma_0$ the statement is still valid because the $\underline{\sigma}$ with $\sigma_0 = \bar\sigma_0$ have 0 measure.

Existence of the limit (3.2.15) implies, by the τ–invariance of $m_{\underline{p}}$,

e3.2.19
$$N^{-1}\sum_{j=1}^{N-1}(\varphi_j(\tau^j\underline{\sigma})-\varphi(\tau^j\underline{\sigma}))\xrightarrow[N\to\infty]{}0 \quad\text{in}\quad L_1(m_{\underline{p}}). \tag{3.2.19}$$

Hence, by Birkhoff's theorem, the limit as $N\to\infty$ exists in $L_1(m_{\underline{p}})$:

e3.2.20
$$\overline{\varphi}(\underline{\sigma})=\lim_{N\to\infty}N^{-1}\sum_{j=1}^{N-1}\varphi(\tau^j\underline{\sigma}) \tag{3.2.20}$$

and, therefore, it also follows $f_N(\underline{\sigma})\xrightarrow[N\to\infty]{}\overline{\varphi}(\underline{\sigma})$ in $L_1(m_{\underline{p}})$, and the function $\widetilde{s}(\underline{\sigma})$ will be just $\overline{\varphi}(\underline{\sigma})$. ∎

We can proceed to conclude the proof of the proposition.

Proof of proposition (3.2.1): The functions φ_j are a set equibounded in $L_1(m_{\underline{p}})$ and equisummable[3] with respect to $m_{\underline{p}}$. Indeed let $E_{j,k}$ be the set of the sequences $\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}$ such that
N3.2.3

e3.2.21
$$k\le\varphi_j(\underline{\sigma})<k+1. \tag{3.2.21}$$

For the sequences $\underline{\sigma}$ contained in such a set one has

e3.2.22
$$p\begin{pmatrix}-j\ldots 0\\ \sigma_{-j}\ldots\sigma_0\end{pmatrix}\le e^{-k}\,p\begin{pmatrix}-j\ldots -1\\ \sigma_{-j}\ldots\sigma_{-1}\end{pmatrix}. \tag{3.2.22}$$

Adding up these relations, summing over the choices of $\sigma_{-j},\ldots,\sigma_0\in\{0,\ldots,n\}^{j+1}$ such that $C^{-j\ldots 0}_{\sigma_{-j}\ldots\sigma_0}\subset E_{j,k}$, one finds

e3.2.23
$$m_{\underline{p}}(E_{j,k})\le(n+1)e^{-k}, \tag{3.2.23}$$

hence, if E is a $m_{\underline{p}}$–measurable set,

$$\int_E\varphi_j(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})=\sum_{k=0}^{\infty}\int\varphi_j(\underline{\sigma})\chi_E(\underline{\sigma})\chi_{E_{j,k}}(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})\le$$

$$\le\sum_{k=0}^{\infty}(1+k)\int\chi_E(\underline{\sigma})\chi_{E_{j,k}}(\underline{\sigma})m_{\underline{p}}(d\underline{\sigma})\le\sqrt{m_{\underline{p}}(E)}\sum_{k=0}^{\infty}(1+k)\sqrt{m_{\underline{p}}(E_{j,k})}\le$$

e3.2.24
$$\le\sqrt{m_{\underline{p}}(E)}\Big[\sqrt{(n+1)}\sum_{k=0}^{\infty}(1+k)e^{-k/2}\Big], \tag{3.2.24}$$

[3] A set of functions $F\subset L_1(\mu)$ is called *equisummable* if for every ε there is a δ such that if $\mu(E)\le\delta$ then $\int_E f(x)d\mu(x)\le\varepsilon$ for every $f\in F$.

that shows simultaneously (and "miracolously") the equiboundedness in $L_1(m_{\underline{p}})$ and the equisummability of φ_j, $j = 1, \ldots$.

Simple considerations of measure theory based on Vitali's convergence theorem and on Fatou's lemma[4] show that, to verify the convergence, in $L_1(m_{\underline{p}})$
N3.2.4 and almost everywhere, of the sequence $\{\varphi_j\}$, as $j \to \infty$, it suffices to verify convergence in $m_{\underline{p}}$–measure and, respectively, almost everywhere of the sequence of functions

e3.2.25
$$\exp(-\varphi_j(\underline{\sigma})) = \frac{p(C^{-j\ldots 0}_{-j\ldots\sigma_0})}{p(C^{-j\ldots -1}_{-j\ldots\sigma_{-1}})} \quad \text{for} \quad j \to \infty, \tag{3.2.25}$$

N3.2.5 as one can check.[5]

The problem of convergence in $m_{\underline{p}}$–measure and $m_{\underline{p}}$–almost everywhere, of $\exp(-\varphi_j)$ is very similar to the problem of proving the *Vitali–Lebesgue theorem* which states the existence of the limit as $I \to y$ of $|I|^{-1} \int_I F(x)dx$ if I are intervals containing y. In the present context the result is called *Doob's theorem.*

This result is a particular case of a general theorem of measure theory, and it is worth to make a small notational effort to reduce it to this general theorem.

All the functions $\exp(-\varphi_j)$, $j = 1, 2, \ldots$ are measurable with respect to the σ–algebra $\mathcal{B}$–generated by the cylinders with negative base; it will be useful
N3.2.6 to identify, in the obvious way, such cylinders with those of [6] $\{0, \ldots, n\}^{\mathbb{Z}_-}$ so that

e3.2.26
$$m(C^J_{\underline{\sigma}}) = m_{\underline{p}}(C^J_{\underline{\sigma}}) = p\begin{pmatrix} J \\ \underline{\sigma} \end{pmatrix} \qquad \forall\, J \subset \mathbb{Z}_-, \,\forall\, \underline{\sigma} \in \{0, \ldots, n\}^{|J|}; \tag{3.2.26}$$

in other words m coincides with $m_{\underline{p}}$ restricted to the cylinders witn negative
N3.2.7 base. By the Radon–Nykodim theorem,[7] given $\sigma_0 \in \{0, \ldots, n\}$, there exists a function in $L_1(m)$, that we shall denote g_{σ_0}, such that

e3.2.27
$$p\begin{pmatrix} -j\ldots -10 \\ \sigma_{-j}\ldots\sigma_{-1}\sigma_0 \end{pmatrix} = \int_{C^{-j\ldots -1}_{\sigma_{-j}\ldots\sigma_{-1}}} g_{\sigma_0}(\underline{\sigma}')\, m(d\underline{\sigma}'). \tag{3.2.27}$$

---

[4] See [DS58], p.150 and 152.

[5] If $\exp(-\varphi_j)$ converges in $m_{\underline{p}}$–measure and almost everywhere to a limit that we denote $\exp(-\varphi)$ one has $0 \le \varphi_j \xrightarrow[j\to\infty]{} \varphi$ almost everywhere. By Fatou's lemma φ is summable and hence $< +\infty$ almost everywhere; then convergence in $m_{\underline{p}}$–measure of $\exp(-\varphi_j)$ to $\exp(-\varphi)$ implies convergence in measure of φ_j to φ and therefore, given the equisummability of the functions φ_j, Vitali's criterion of convergence implies the convergence in $L_1(m_{\underline{p}})$ of φ_j to φ.

[6] $\mathbb{Z}_+$ denotes the integers ≥ 0 and $\mathbb{Z}_-$ denotes the integers < 0. Hence $\{0, \ldots, n\}^{\mathbb{Z}_-}$ are unilateral sequences with negative labels.

[7] See [DS58], p.176.

Hence

e3.2.28
$$\frac{p\begin{pmatrix} -j \ldots -1 & 0 \\ \sigma_{-j} \ldots \sigma_{-1} & \sigma_0 \end{pmatrix}}{p\begin{pmatrix} -j \ldots -1 \\ \sigma_{-j} \ldots \sigma_{-1} \end{pmatrix}} = \frac{\int_{C^{-1 \ldots -1}_{\sigma_{-j} \ldots \sigma_{-1}}} g_{\sigma_0}(\underline{\sigma}')\, m(d\underline{\sigma}')}{\int_{C^{-j \ldots -1}_{\sigma_{-j} \ldots \sigma_{-1}}} m(d\underline{\sigma}')}, \tag{3.2.28}$$

and we realize that our purpose is to show the convergence, in m–measure and m–almost everywhere, of this quantity regarded as a function of $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}_-}$, for every fixed σ_0. And the limit should be precisely g_{σ_0}.

It is now possible to reduce the analysis to some simple and classical constructions.

For simplicity we shall suppose that the measures $m_{\underline{p}}$ and m are not atomic.

Let us consider the map of $\{0,\ldots,n\}^{\mathbb{Z}_-}$ in $[0, n/n{+}1]$ defined by $\underline{\sigma} \to X(\underline{\sigma})$:

e3.2.29
$$X(\underline{\sigma}) = \sum_{j=1}^{\infty} \frac{\sigma_{-j}}{(2+n)^j}, \tag{3.2.29}$$

which is a homeomorphism between the sequences in $\{0,\ldots,n\}^{\mathbb{Z}_-}$ and the subset $X(\{0,\ldots,n\}^{\mathbb{Z}_-})$ of $[0,1]$, which is the Cantor set consisting of the numbers of $[0,1]$ whose development in base $(n+2)$ never contains the digit $n+1$.

Then via the map X we can transform the measure m into a measure $\overline{m}$ on $[0, n/(n+1)]$ and the function g_{σ_0} into $\widetilde{g} \in L_1(\overline{m})$ by setting

e3.2.30
$$\begin{aligned} \overline{m}(E) &= m(X^{-1}E), \qquad E \subset \mathcal{B}([0, n/n+1]), \\ \widetilde{g}(x) &= g_{\sigma_0}(X^{-1}(x)). \end{aligned} \tag{3.2.30}$$

Define then the map $Y : [0, n/n+1] \to [0,1]$ as

e3.2.31
$$y = Y(x) = \overline{m}([0,x]). \tag{3.2.31}$$

The function $x \to Y(x)$ is non-decreasing, more precisely it is strictly increasing except, possibly, in the union of a denumerable family of closed disjoint intervals. Hence Y is continuous and is invertible as a map between $[0, n/n+1]$ deprived of a denumerable infinity of closed disjoint sets and its image in $[0,1]$ (which consists in the same $[0,1]$ deprived, at most, of a denumerable infinity of points).

Therefore Y establishes an isomorphism mod 0 between $([0, n/n+1]), \overline{m})$ and $([0,1], \mu)$ where μ is the Lebesgue measure (because of the relation (3.2.31)). Via this isomorphism $\widetilde{g}$ becomes a function $\overline{g} \in L_1(\mu)$.

Another remarkable fact is that the set

e3.2.32
$$D(\sigma_{-j} \ldots \sigma_{-1}) = YX(C^{-j \ldots -1}_{\sigma_{-j} \ldots \sigma_{-1}}) \tag{3.2.32}$$

either is a connected interval or is empty; the latter possibility arises if and only if $m(C^{-j \ldots -1}_{\sigma_{-j} \ldots \sigma_{-1}}) = 0$. Indeed it suffices to remark that the set of the numbers that have the first j digits of the their development in base $(n+1)$ equal is an interval.

The nonempty intervals having the form (3.2.32) form a covering $\mathcal{D}_j$ of $[0,1]$ with intervals which, pairwise, have no internal points in common. Furthermore the covering $\mathcal{D}_{j+1}$ refines $\mathcal{D}_j$ because every interval of $\mathcal{D}_{j+1}$ is a union of intervals of $\mathcal{D}_j$.

If $x \in D(\sigma_{-j},\ldots,\sigma_{-1})$ and if we define the function

e3.2.33
$$x \to h_j(x) = \frac{\int_{D(\sigma_{-j}\ldots\sigma_{-1})} \overline{g}(x')dx'}{|D(\sigma_{-j}\ldots\sigma_{-1})|} \tag{3.2.33}$$

(which is a relation that has μ–almost everywhere meaning) we obtain the image of φ_j via YX.

The Vitali–Lebesgue theorem concerns exactly sequences of functions having the form

e3.2.34
$$x \to q_j(x) = \frac{\int_D g(x')dx'}{|D|} \quad \text{with} \quad g \in L_1(\mu), \tag{3.2.34}$$

where D is the interval that contains x extracted out of a pavement $\mathcal{D}_j$ of $[0,1]$ with intervals with no common internal points and refined by $\mathcal{D}_{j+1}$. The theorem says that $q_j \xrightarrow[j\to\infty]{} g$ almost everywhere with respect to the
N3.2.8 Lebesgue measure μ in $[0,1]$ and in $L_1(\mu)$.[8]

Applying the latter statement to (3.2.33) and translating it back to the original variables via the isomorphism YX we see that it means

e3.2.35
$$\lim_{j\to\infty} \varphi_j(\underline{\sigma}) = g_{\sigma_0}(\underline{\sigma}) \tag{3.2.35}$$

$m_{\underline{p}}$–almost everywhere and in $L_1(m_{\underline{p}})$. This yields the proof of Doob's theorem in the present special case and completes the proof of proposition (3.2.1).

Problems for §3.2

Q3.2.1 **[3.2.1]:** (*Approximability in entropy and distribution*)
Under the hypothesis of proposition (3.2.1) assume $\underline{\widehat{\sigma}}$ ergodic and set $S = s(\widehat{\sigma})$. Show that given an integer $u > 0$ and $\varepsilon > 0$, an integer $N(\varepsilon,u)$ exists such that for all $N > N(\varepsilon,u)$ it is possible to divide $\{0,\ldots,n\}^N$ into two classes $\mathcal{C}_{1,\varepsilon,u}(N)$ and $\mathcal{C}_{2,\varepsilon,u}(N)$ such that

$$\sum_{\sigma_0\ldots\sigma_{N-1}\in\mathcal{C}_{2,\varepsilon,u}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \varepsilon,$$
$$\exp((S-\varepsilon)N) \le |\mathcal{C}_{1,\varepsilon,u}(N)| \le \exp[(S+\varepsilon)N],$$
$$\exp[-(S+\varepsilon)N] \le p\begin{pmatrix}0\ldots N-1\\ \sigma'_0\ldots\sigma'_{N-1}\end{pmatrix} \le \exp[-(S-\varepsilon)N],$$

[8] See, for instance, [DS58] Vol. I, Ch. III, p. 214. One should note the "analogy" between the proof discussed here and that of Birkhoff's theorem in Appendix 2.2.

for all $(\sigma'_0 \dots \sigma'_{N-1}) \in \mathcal{C}_{1,\varepsilon,u}(N)$ and

$$\sum_{\sigma_0\dots\sigma_{u-1}\in\{0,\dots,n\}^u} \Big| p\begin{pmatrix} 0\dots u-1 \\ \sigma_0\dots\sigma_{u-1}\end{pmatrix} - \text{(frequency of appearance of } \begin{pmatrix} 0\dots u-1 \\ \sigma_0\dots\sigma_{u-1}\end{pmatrix} \text{ in } (\sigma'_0,\dots,\sigma'_{N-1})\Big| < \varepsilon$$

for all $(\sigma'_0 \dots \sigma'_{N-1}) \in \mathcal{C}_{1,\varepsilon,u}(N)$. (*Hint:* Proceed as in the above derivation of (3.2.3), (3.2.4), (3.2.5) from (3.2.12) observing that, by Birkhoff's theorem,

$$g_N(\underline{\sigma}) = \sum_{\substack{\sigma'_0\dots\sigma'_{u-1}\\ \in\{0,\dots,n\}^u}} p\begin{pmatrix} 0\dots u-1 \\ \sigma'_0\dots\sigma'_{u-1}\end{pmatrix} - \text{(frequency of appearance of } \begin{pmatrix} 0\dots u-1 \\ \sigma'_0\dots\sigma'_{u-1}\end{pmatrix} \text{ between 0 and } N-1 \text{ in } \underline{\sigma})$$

converges, for $N \to \infty$, to zero in $L_1(m_{\underline{p}})$ and $m_{\underline{p}}$–almost everywhere. Then choose $E_{\varepsilon,u,N}$ in analogy with the choice (3.2.12) but such that $|g_N(\underline{\sigma})| < \varepsilon$, *etc.*)

Q3.2.2 **[3.2.2]:** (*Entropy of a distribution on symbolic sequences*)
If $\underline{p} \in M(\{0,\dots,n\}^{\mathbb{Z}})$ we can define the entropy of $\underline{p}$ naturally as

$$s(\underline{p}) = \lim_{\varepsilon\to 0} \limsup_{N\to\infty} N^{-1} \log \mathcal{N}_\varepsilon(\underline{p}|N),$$

where $\mathcal{N}_\varepsilon(\underline{p}\,|\,N)$ is the minimum number of elements of $\{0,\dots,n\}^N$ that remain if we take out from $\{0,\dots,n\}^N$ a family $\mathcal{C}_2(N)$ of elements such that $\sum_{\sigma_0\dots\sigma_{N-1}\in\mathcal{C}_2(N)} p\begin{pmatrix} 0\dots N-1 \\ \sigma_0\dots\sigma_{N-1}\end{pmatrix} \le \varepsilon$.

Then proposition (3.2.1) formulated in terms of $\underline{p}$ has a meaning (in an obvious way, even if there is no $\widehat{\underline{\sigma}} \in \{0,\dots,n\}^{\mathbb{Z}}$ that generates $\underline{p}$) and is true (note that this statement does not strengthen proposition (3.2.1) except for what concerns its statement (i), cf. proposition (3.2.1), and the relative remarks).

Q3.2.3 **[3.2.3]:** (*Lebesgue measure on* $[0,1]^2$ *and the* $(\frac12,\frac12)$ *Bernoulli shift*)
Show that the Bernoulli distribution $B(1/2,1/2)$ on $\{0,1\}^{\mathbb{Z}}$ is isomorphic mod 0 to the Lebesgue measure on the square $[0,1]\times[0,1]$. The isomorphism is established by $(x,y) \in [0,1]^2 \to \underline{\sigma} \in \{0,1\}^{\mathbb{Z}}$ if $x = \sum_{j=0}^\infty \frac{\sigma_j}{2^{j+1}}$, $y = \sum_{j=1}^\infty \frac{\sigma_{-j}}{2^j}$.

Q3.2.4 **[3.2.4]:** (*The baker map and the* $(\frac12,\frac12)$ *Bernoulli shift*)
The isomorphism of problem [3.2.3] establishes an isomorphism mod 0 between the dynamical systems $(\{0,1\}^{\mathbb{Z}},\tau,B)$ and $([0,1]^2,S,\lambda)$ where $\lambda(dx) = dxdy$ and $S(x,y) = (2x,y/2)$ if $x < 1/2$, and $S(x,y) = (2x-1,(y+1)/2)$ if $x \ge 1/2$. The latter dynamical system is called *the baker map* (see also problem [2.2.43]).

Q3.2.5 **[3.2.5]:** (*Generalization of the binary and decimal expansions*)
Consider n positive numbers $p_1,\dots,p_n$ such that $\sum_{i=1}^n p_i = 1$. Consider n intervals I_1, $I_2,\dots,I_n$ that decompose $[0,1)$. Define $Sx = (x-a_i)/p_i$, if $x \in I_i = [a_i,a_{i+1})$, having set $a_0 = 0$ and $a_{i+1} = p_1+\dots+p_i$, $i = 1,\dots,n$. Draw the map S as a map of $[0,1]$ into itself and show that S conserves the Lebesgue measure on $[0,1)$. The code that associates with $x \in [0,1]$ its history on $(I_1,\dots,I_n) : x \to (\sigma_0,\sigma_1,\dots,) \in \{1,\dots,n\}^{\mathbb{Z}_+}$ transforms the Lebesgue measure into the unilateral Bernoulli measure $B(p_1,\dots,p_n)$ on $\{0,\dots,n\}^{\mathbb{Z}_+}$. This code generalizes the binary representation (which corresponds to the case $n = 2$ and $p_1 = p_2 = 1/2$).

Q3.2.6 **[3.2.6]:** (*A generalization of the baker map isomorphism with a Bernoulli shift*)
Generalize the result of problem [3.2.5] to show that the Bernoulli scheme with n symbols and probabilities $(p_1, \ldots, p_n)$ is isomorphic mod 0 to the Lebesgue measure on $[0,1]^2$ on which acts a suitable map S.

Q3.2.7 **[3.2.7]:** Consider the Bernoulli scheme $B(p_1, \ldots, p_n)$ on $\{1, \ldots, n\}^{\mathbb{Z}}$. Given k positive numbers $a_1, \ldots, a_k$ such that $a_1 + \ldots + a_k = 1$, show the existence of a Borel partition of $\{1, \ldots, n\}^{\mathbb{Z}}$ into k sets of measures, respectively, $a_1, a_2, \ldots, a_k$. (*Hint:* Use the result in problem [3.2.5] and the fact that such partitions trivially exist on $[0,1]^2$ considered with the Lebesgue measure).

Q3.2.8 **[3.2.8]:** Given a Borel partition $\mathcal{Q} = \{Q_1, \ldots, Q_k\}$ of $\{1, \ldots, n\}^{\mathbb{Z}}$, and given the Bernoulli measure μ on $\{1, \ldots, n\}^{\mathbb{Z}}$ with probabilities $B(p_1, \ldots, p_n)$ show, by making use of the results of problems [3.2.6] and [3.2.7], that there exists a family of partitions $t \to \mathcal{Q}(t)$ parameterized bt $t \in [0,1]$, such that $\mathcal{Q}(0) = \{\emptyset, \emptyset, \ldots, \emptyset, \{0, \ldots, n\}^{\mathbb{Z}}\}$, $\mathcal{Q}(1) = \{Q_1, \ldots, Q_k\}$ with a fixed number, k, of atoms and which is continuous in the sense that

$$\lim_{t \to t_0} |\mathcal{Q}(t), \mathcal{Q}(t_0)| \equiv \lim_{t \to t_0} \sum_{i=1}^{k} \mu(Q_i(t) \Delta Q_i(t_0)) = 0, \ \forall\, t_0 \in [0,1] \ ,$$

where $A \Delta B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference between A and B. (*Hint:* Consider the isomorphism discussed in problem [3.2.6] and use that such a property is easy to show in the case of Borel partitions of $[0,1]^2$).

Q3.2.9 **[3.2.9]:** (*Non-atomic Borel measures on* $\{0,1\}^{\mathbb{Z}}$ *are isomorphic* mod 0 *to the Lebesgue measure on* $[0,1]$)
Let μ be a non-atomic Borel measure on $\{0,1\}^{\mathbb{Z}}$ (*i.e.* a measure such that no positive measure set E exists which has no subsets of smaller but positive measure). Show that it is isomorphic mod 0 to the Lebesgue measure on $[0,1]$. (*Hint:* Use the idea and the map YX that appear at the end of the proof of proposition (3.2.1).)

Q3.2.10 **[3.2.10]:** Show that if μ is an S–invariant measure on a σ–algebra $\mathcal{B}$ of Ω and it is S–mixing then μ is non-atomic if $\mathcal{B}$ is not trivial.

Q3.2.11 **[3.2.11]:** If μ is a non-atomic Borel measure on a complete and separable metric space then μ is isomorphic mod 0 to a Borel measure on $\{0,1\}^{\mathbb{Z}}$. (*Hint:* Use Alexandrov and
N3.2.9 Urhyson theorems[9] stating that each separable and complete metric space is homeomorphic to a Borel subset of $[0,1]^{\mathbb{Z}}$ which, in turn, is in a one-to-one and bimeasurable correspondence with a set Borel of $\{0,1\}^{\mathbb{Z}}$. Hence by using the conclusions of problem [3.2.9] show that μ is isomorphic mod 0 to the Lebesgue measure on $[0,1]$.)

Q3.2.12 **[3.2.12]:** Show that every dynamical system (Ω, S, μ) with Ω complete metric separable and with μ non-atomic is isomorphic mod 0 to a dynamical system having the form $([0,1], \widetilde{S}, \mu_0)$ where μ_0 is the Lebesgue measure and $\widetilde{S}$ is a suitable map.

Q3.2.13 **[3.2.13]:** Interpret proposition (2.3.3) as a proof that every invertible topological dynamical system (Ω, S, μ) on a complete metric space admitting a topological separating partition is isomorphic mod 0 to a system of the type $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, m)$, where τ is the translation of the sequences of symbols.

Q3.2.14 **[3.2.14]:** If (Ω, S, μ) is an invertible dynamical system on Ω that is assumed be a complete metric separable space and if μ is a complete Borel measure (cf. Appendix 1.4), then (Ω, S, μ) is isomorphic mod 0 to a dynamical system of the type $(V^{\mathbb{Z}}, \tau, \widetilde{\mu})$ where V is a compact metric space, τ is the translation on $V^{\mathbb{Z}}$, and $\widetilde{\mu}$ is a τ–invariant complete Borel measure. (*Hint:* Let $x_1, x_2, \ldots$ be a denumerable dense set and consider the function $\varphi : \Omega \to [0,1]^{\mathbb{Z}_+}$ defined by $\varphi(x) = (d(x,x_i)/1 + d(x,x_i))_{i \in \mathbb{Z}}$, if $d(\ldots)$ is the metric on Ω. Then φ is an isomorphism between Ω and its image $\varphi(\Omega) \subset [0,1]^{\mathbb{Z}_+}$ that turns out

[9] See [DS58] pp. 24,138, and problem [3.2.14].

to be a Borel set, and in fact a G_δ–set (*i.e.* a countable intersection of open dense sets),
N3.2.10 in the product topology (theorem, of Alexandrov and Urhyson[10]). Associate then with x the sequence in $V^{\mathbb{Z}} = ([0,1]^{\mathbb{Z}_+})^{\mathbb{Z}}$ defined by $\Phi(x) = (\varphi(S^i x))_{i\in\mathbb{Z}}$; it is clear that Φ is a continuous and one-to-one map between Ω and $\varphi(\Omega)$; hence it is bimeasurable (by
N3.2.11 Kuratowsky's theorem[11] This implies that if we set $\mu(E) = \mu(\Phi^{-1}(E))$ for E Borel in $V^{\mathbb{Z}}$, then $(V^{\mathbb{Z}}, \tau, \widetilde{\mu})$ is isomorphic mod0 to (Ω, S, μ).)

One says, therefore, that "every metric invertible dynamical system constructed on a complete separable metric space by means of a map and of a Borel measure is isomorphic mod (0) to a topological dynamical system on a compact metric space".

Bibliographical note to §3.2

The proof of proposition (3.2.1) ("Shannon–McMillan theorem") is taken from [Ki57], p. 44–89.

The relation between Doob's theorem and Vitali–Lebesgue's is well known. The proof of Doob's theorem can be found in [Ki57]. A proof of the Vitali–Lebesgue theorem can be found, for instance, in [DS58], p.214.

We remark that for a proof of proposition (3.2.1) the L_1–convergence of (3.2.15) would be sufficient. Instead, we have also obtained (implicitly) the almost everywhere convergence: this provides a strenghtening of the Shannon–McMillan theorem (due to Breiman, cf. [Br57]).

A generalization of the notion of entropy to measures on $\{0,\ldots,n\}^{\mathbb{Z}}$ that are not invariant under translation can be found in [Ja59], where an extension of the theorem of Shannon–McMillan to "S-quasi–periodic" distributions (rather than S-invariant) is discussed.

§3.3 Elementary properties of the average entropy

In this section some consequences of the proof of proposition (3.2.1) are collected together with some elementary properties of entropy and with various interesting definitions.

C3.3.1 **(3.3.1) Corollary:** (Average entropy of a sequence)
If $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ is a stationary distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ the function defined by the limit

e3.3.1
$$\widetilde{s}(\underline{\sigma}) = \lim_{N\to\infty} -N^{-1} \log p \begin{pmatrix} O \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \tag{3.3.1}$$

exists in $L_1(m_{\underline{p}})$ and (therefore) the limit

e3.3.2
$$\widetilde{s}(\underline{p}) = \lim_{N\to\infty} -N^{-1} \sum_{\sigma_0\ldots\sigma_{N-1}} p \begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \log p \begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \tag{3.3.2}$$

---

[10] It is a useful exercise to look for a proof of this theorem without having recourse to the literature.

[11] See footnote 3, Section §2.3.

exists.

Proof: In the proof of statement (i) of proposition (3.2.1) (*i.e.* in lemma (3.2.1) needed to prove it) the sequence $\underline{\widehat{\sigma}}$ generating the distribution of frequencies $\underline{p}$ enters only through $m_{\underline{p}}$. And we only made use of the translation invariance of $m_{\underline{p}}$: the latter property is also true for the distribution $\underline{p}$ that appears in (3.3.1) even if $\underline{p}$ is not the distribution of a sequence. ∎

D3.3.1 **(3.3.1) Definition:** *If $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ the quantity $\widetilde{s}(\underline{p})$ in (3.3.2) will be called* average entropy *of $\underline{p}$. The notion is a priori different from that of* entropy *of $\underline{p}$, that we shall denote by $s(\underline{p})$, which can be defined by generalizing definition (3.1.1) to the case in which $\underline{p}$ is not generated by a*
N3.3.1 *sequence $\underline{\sigma}$.*[1]

Remarks: (1) If $\underline{p}$ is ergodic $m_{\underline{p}}$ is also ergodic and, therefore, $\widetilde{s}(\underline{\sigma})$ is almost everywhere constant (with respect to $m_{\underline{p}}$, because we have seen in the proof of lemma (3.2.1) that it is translation invariant) and one can repeat the first part of the proof of proposition (3.2.1) to conclude that for all $\varepsilon > 0$, there is N_ε such that for $N \geq N_\varepsilon$ the elements of $\{0,\ldots,1\}^N$ can be divided in two sets $\mathcal{C}_{1,\varepsilon}(N)$ and $\mathcal{C}_{2,\varepsilon}(N)$ satisfying (3.2.3), (3.2.4) and (3.2.5).

We can then repeat the arguments of remark (1) to proposition (3.2.1) and deduce that $\widetilde{s}(\underline{\sigma}) = s = s(\underline{p})$. Hence if $\underline{p}$ is ergodic we have

e3.3.3
$$\widetilde{s}(\underline{p}) = s(\underline{p}). \tag{3.3.3}$$

(2) The relation (3.3.3) does not hold in general if $\underline{p}$ is not ergodic.

The following proposition clarifies the remark (2) above.

P3.3.1 **(3.3.1) Proposition:** (Entropy and average entropy)
If $\underline{p}_1$, $\underline{p}_2 \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ and are ergodic and, if $\underline{p} = a\underline{p}_1 + (1-a)\underline{p}_2$ with $0 < a < 1$, in the sense that

e3.3.4
$$p\begin{pmatrix} j_1 \cdots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix} = a p_1 \begin{pmatrix} j_1 \cdots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix} + (1-a) p_2 \begin{pmatrix} j_1 \cdots j_p \\ \sigma_1 \ldots \sigma_p \end{pmatrix}, \tag{3.3.4}$$

one has

e3.3.5
$$\begin{aligned} \widetilde{s}(\underline{p}) &= a s(\underline{p}_1) + (1-a) s(\underline{p}_2), \\ s(\underline{p}) &= \max\{s(\underline{p}_1), s(\underline{p}_2)\}. \end{aligned} \tag{3.3.5}$$

The first property is called affinity *of the entropy.*

---

[1] In other words we can define the *complexity* with weight $e^{-\underline{V}}$ of a shift invariant distribution $\underline{p}$ the quantity $s(\underline{p},\underline{V})$ as in (3.1.8), (3.1.9) and (3.1.10), where $\mathcal{C}_1$ and $\mathcal{C}_2$ are now defined by replacing the frequencies $p\big(\begin{smallmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{smallmatrix} \,|\, \underline{\widehat{\sigma}}\big)$ by $p\big(\begin{smallmatrix} 0\ldots N-1 \\ \sigma_0\ldots\sigma_{N-1}\end{smallmatrix}\big)$ and by replacing $\eta_\varepsilon(\underline{\widehat{\sigma}}|N;\underline{V})$ with $\eta_\varepsilon(\underline{p}|N;\underline{V})$, where one defines $\eta_\varepsilon(\underline{p}|N;\underline{V})$ through the second member of (3.1.9). Then $s(\underline{p})$ is naturally defined as $s(\underline{p}\,|\,\underline{0})$, using the notations of definition (3.1.1) with $V_N \equiv 0$.

Proof: The first of (3.3.5) is based on the identity (3.3.3) for ergodic distributions and on the following simple inequalities.

The function $\xi \to -\xi \log \xi$ is convex for $\xi \in [0,1]$ and, therefore,

e3.3.6
$$-(ax+(1-a)y)\log(ax+(1-a)y) \geq -ax\log x-(1-a)y\log y \tag{3.3.6}$$

for every a, x, $y \in [0,1]$. Furthermore if $0 \leq x_1, \ldots, x_p \leq 1$, $0 \leq y_1, \ldots, y_p \leq 1$ and $\sum_i x_i = \sum_i y_i = 1$ the monotonicity of $\xi \to -\log \xi$ gives

e3.3.7
$$\begin{aligned}
&\sum_i -(ax_i+(1-a)y_i)\log(ax_i+(1-a)y_i) \equiv\\
&\quad\equiv \sum_i -ax_i\log(ax_i+(1-a)y_i)+\\
&\qquad+\sum_i -(1-a)y_i\log(ax_i+(1-a)y_i) \leq\\
&\quad\leq \sum_i -ax_i\log ax_i + \sum_i -(1-a)y_i\log(1-a)y_i =\\
&\quad= -a\log a-(1-a)\log(1-a)+\\
&\qquad+a\sum_i -x_i\log x_i+(1-a)\sum_i -y_i\log y_i.
\end{aligned} \tag{3.3.7}$$

Hence (3.3.6) and (3.3.7) imply

e3.3.8
$$\begin{aligned}
&a\sum_i -x_i\log x_i+(1-a)\sum_i y_i\log y_i \leq\\
&\quad\leq \sum_i -(ax_i+(1-a)y_i)\log(ax_i+(1-a)y_i) \leq\\
&\quad\leq -a\log a-(1-a)\log(1-a)+a\sum_i -x_i\log x_i+(1-a)\sum_i -y_i\log y_i,
\end{aligned} \tag{3.3.8}$$

so that, by selecting $i=(\sigma_0,\ldots,\sigma_{N-1})$, $x_i = p_1\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, $y_i = p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, then (3.3.8) and (3.3.2) imply the first of the (3.3.5).

The second of (3.3.5) is based on proposition (3.2.1): if $s(\underline{p}_1)=s_1 \leq s(\underline{p}_2)=s_2$ and if $\mathcal{C}^1_{1,\varepsilon}(N)$ and $\mathcal{C}^2_{1,\varepsilon}(N)$ denote the sets of specifications of large probability with respect to $m_{\underline{p}_1}$ and $m_{\underline{p}_2}$ we see that

e3.3.9
$$\begin{aligned}
|\,\mathcal{C}^1_{1,\varepsilon}(N)\cup\mathcal{C}^2_{1,\varepsilon}(N)\,| &\leq e^{N(s_1+\varepsilon)}+e^{N(s_2+\varepsilon)} \leq\\
&\leq e^{N(s_2+\varepsilon)}(1+e^{-N(s_2-s_1)}),
\end{aligned} \tag{3.3.9}$$

that shows, since

e3.3.10
$$\sum_{\sigma_0\ldots\sigma_{N-1}\notin\mathcal{C}^1_{1,\varepsilon}(N)\cup\mathcal{C}^2_{1,\varepsilon}(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \leq 2\varepsilon, \tag{3.3.10}$$

that $s = s(\underline{p}) \le s_2$.

But one cannot have $s < s_2$: if indeed $\overline{\mathcal{C}}_\varepsilon(N)$ was a set of large probability $(> 1-\varepsilon)$ and $|\,\overline{\mathcal{C}}_\varepsilon(N)\,| < e^{N(s+\varepsilon)}$, for all $N \ge N_\varepsilon$, we should have

e3.3.11
$$\begin{aligned}&\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} =\\ &= (1-a)^{-1}(1-a)\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p_2\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le\\ &\le (1-a)^{-1}\sum_{\sigma_0\ldots\sigma_{N-1}\notin\overline{\mathcal{C}}_\varepsilon(N)} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \le \varepsilon(1-a)^{-1},\end{aligned} \tag{3.3.11}$$

but this would contradict the fact that s_2 is the entropy of $\underline{p}_2$. ∎

Another simple but important property of $s(\underline{p})$ is the following one.

P3.3.2 **(3.3.2) Proposition:** (Average entropy as an infimum, semicontinuity) *Let $\underline{p} \in M(\{0,\ldots,n\}^{\mathbb{Z}})$ be an invariant distribution; one has*

e3.3.12
$$\begin{aligned}&\widetilde{s}(\underline{p}) \le \log(1+n),\\ &\widetilde{s}(\underline{p}) = \inf_{N=2^k} -N^{-1}\sum_{\sigma_0\ldots\sigma_{N-1}} p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix},\end{aligned} \tag{3.3.12}$$

where the infimum is taken over the integers k, by setting $N = 2^k$. Hence the average entropy is upper-semicontinuous, i.e. if $\underline{p}_n$ is a sequence of distributions which converges to a limit $\underline{p}_\infty$ in the topology of $M(\{0,\ldots,n\}^{\mathbb{Z}})$, i.e. in the sense that $p_n\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix} \xrightarrow[n\to\infty]{} p_\infty\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$ for all $N, \sigma_0, \sigma_1, \ldots, \sigma_N$, then

e3.3.13
$$\widetilde{s}(\underline{p}_\infty) \ge \limsup_{n\to\infty} \widetilde{s}(\underline{p}) \tag{3.3.13}$$

and $s(\mu)$ has a maximum on any compact set in $M(\{0,\ldots,n\}^{\mathbb{Z}})$.

Proof: The function $(x_1,\ldots,x_p) \to -\sum_i x_i \log x_i$, $0 \le x_1,\ldots,x_p \le 1$, $\sum_i x_i = 1$, has its maximum in $x_i = 1/p$ where its value is $\log p$. The sum (3.3.2) has precisely this form: this shows the first relation of (3.3.12).

To show the second relation in (3.3.12) let I and J be two sets of labels and let $(p_i)_{i\in I}$, $(p'_j)_{j\in J}$, $(p_{ij})_{ij\in I\times J}$ be three families of not negative numbers

N3.3.2 such that [2]

e3.3.14
$$\begin{aligned}&\sum_i p_i = \sum_j p'_j = \sum_{ij} p_{ij} = 1,\\ &p_i = \sum_j p_{ij}, \qquad p'_j = \sum_i p_{ij}.\end{aligned} \tag{3.3.14}$$

---

[2] We think here to $i = (\sigma_0\ldots\sigma_{N-1})$, $j = (\sigma'_0\ldots\sigma'_{N-1})$, $p_i = p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}$, $p'_j = p\begin{pmatrix}0\ldots M-1\\ \sigma'_0\ldots\sigma'_{M-1}\end{pmatrix}$, $p_{ij} = p\begin{pmatrix}0\ldots N-1 & N\ldots N+M-1\\ \sigma_0\ldots\sigma_{N-1} & \sigma'_0\ldots\sigma'_{M-1}\end{pmatrix}$.

Then

e3.3.15

$$\begin{aligned}&\sum_{ij} -p_{ij}\log p_{ij} \le \sum_i -p_i \log p_i + \sum_j -p'_j \log p'_j,\\ &\sum_{ij} -p_{ij}\log p_{ij} \ge \sum_i -p_i \log p_i,\end{aligned} \tag{3.3.15}$$

because

e3.3.16

$$\begin{aligned}\sum_{ij} - p_{ij}\log p_{ij} &= \sum_i p_i \sum_j -(p_{ij}/p_i)\log p_{ij} =\\ &= \sum_i p_i \sum_j -(p_{ij}/p_i)(\log p_{ij}/p_i + \log p_i) =\\ &= \sum_i p_i \sum_j -(p_{ij}/p_i)\log(p_{ij}/p_i) + \sum_i -p_i \log p_i =\\ &= \sum_j \Big\{ \sum_i p_i(-p_{ij}/p_i)\log p_{ij}/p_i)\Big\} + \sum_i -p_i\log p_i\end{aligned} \tag{3.3.16}$$

and the term in curly bracket is ≥ 0 because $(p_{ij}/p_i) \le 1$; moreover by the convexity of $\xi \to -\log \xi$, it is bounded above by

e3.3.17

$$-\sum_j \big(\sum_i p_i(p_{ij}/p_i)\big)\log\big(\sum_i p_i(p_{ij}/p_i)\big) = -\sum_j p'_j \log p'_j. \tag{3.3.17}$$

Choosing p_i, p_{ij} as in footnote 2, and setting

e3.3.18

$$H_N(\underline{p}) = \sum_{\sigma_0\ldots\sigma_{N-1}} -p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}\log p\begin{pmatrix}0\ldots N-1\\ \sigma_0\ldots\sigma_{N-1}\end{pmatrix}, \tag{3.3.18}$$

we see that the first of the (3.3.15) means

e3.3.19

$$H_{N+M}(\underline{p}) \le H_N(\underline{p}) + H_M(\underline{p}), \tag{3.3.19}$$

which implies that $N^{-1}H_N(\underline{p})$ is monotonic non-increasing on the sequence $N = 2^k$, $k = 0, 1, \ldots$ and, hence, the second of (3.3.12) follows. ∎

The second of the (3.3.15) will be useful in the following and sometimes we shall refer directly to it without formulating it as a separate proposition.

The following definition is remarkable and natural.

D3.3.2 **(3.3.2) Definition:** (Average entropy of a dynamical system)
Let (Ω, S, μ) be an invertible dynamical system and let $\mathcal{P} = \{P_0, \ldots, P_n\}$ be a partition of Ω into μ–measurable sets. We define the average entropy of S with respect to $\mathcal{P}$ and μ *the following quantity:*

e3.3.20

$$\widetilde{s}(\mathcal{P}, S, \mu) = \widetilde{s}(\underline{p}_\mu), \tag{3.3.20}$$

where $\underline{p}_\mu \in M(\{0,\dots,n\}^{\mathbb{Z}})$ *is defined by*

e3.3.21
$$p_\mu \begin{pmatrix} 0 \dots N-1 \\ \sigma_0 \dots \sigma_{N-1} \end{pmatrix} = \mu\big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k} \big). \tag{3.3.21}$$

We define the average entropy of S with respect to μ *the quantity*

e3.3.22
$$\widetilde{s}(S,\mu) = \sup_{\mathcal{P}} \widetilde{s}(\mathcal{P}, S, \mu), \tag{3.3.22}$$

where the supremum is considered over all the finite partitions of Ω *into* μ*–measurable sets.*

Remarks: (1) Noting that $p_\mu \begin{pmatrix} 0 \dots N-1 \\ \sigma_0 \dots \sigma_{N-1} \end{pmatrix} = p_\mu \begin{pmatrix} -N+1 \dots 0 \\ \sigma_0 \dots \sigma_{N-1} \end{pmatrix}$ we deduce from (3.3.2) that

e3.3.23
$$\widetilde{s}(\mathcal{P}, S, \mu) = \widetilde{s}(\mathcal{P}, S^{-1}, \mu). \tag{3.3.23}$$

(2) Furthermore it is clear that $\widetilde{s}(S,\mu) = \widetilde{s}(S',\mu')$ if (Ω, S, μ) is isomorphic mod 0 to (Ω', S', μ'): *the average entropy of a metric dynamical system is an invariant under isomorphisms mod* 0. For this reason the average entropy is also called the *Kolmogorov–Sinai invariant.*

Problems for §3.3

Q3.3.1 **[3.3.1]:** Making use of proposition (3.3.2) and of problem [3.2.8] show that, given $\varepsilon > 0$, one can find a partition $\mathcal{P}_\varepsilon$ of $\{0,1\}^{\mathbb{Z}}$ that has average entropy $< \varepsilon$ with respect to the action of the translation on the Bernoulli measure $B(1/2, 1/2)$. (*Hint:* Construct a partition that interpolates between the partition $\mathcal{P}_0 = \{\emptyset, \{0,1\}^{\mathbb{Z}}\}$ and $\mathcal{P}_1 = \{C^0_0, C^0_1\}$ and estimate $\widetilde{s}(\mathcal{P}_\varepsilon, S, \mu)$ by means of (3.3.12) with $N = 1$).

Q3.3.2 **[3.3.2]:** Consider the Bernoulli scheme $B(1/2, 1/2)$ and compute the entropy of the partition $\{C^{01}_{00}, C^{01}_{01}, C^{01}_{10}, C^{01}_{11}\}$.

Q3.3.3 **[3.3.3]:** By using the results of problem [2.4.6] show that if $\Delta = \{\omega \,|\, \omega \in \mathcal{M}_e(\{0,\dots,n\}^{\mathbb{Z}}, \tau),\ s(\omega) \in [\alpha, \beta]\}$, and if π is a Borel measure on $\mathcal{M}_e$ such that $\pi(\Delta) = 1$, then the measure $m = \int_\Delta \omega \pi(d\omega)$ has entropy $\widetilde{s}(m) \in [\alpha, \beta]$. The same happens if $[\alpha, \beta]$ is replaced, in the definition of Δ, by $[\alpha, \beta), (\alpha, \beta], (\alpha, \beta)$. (*Hint:* Note that $\varepsilon(\Delta)$ is a Borel set and make use of problem [2.4.6].)

Q3.3.4 **[3.3.4]:** (*Affinity of average entropy for finite mixtures*)
By using problem [3.3.3] show that the average entropy s is affine with respect to the ergodic decompositions of measures $\underline{p} \in M(\{0,\dots,n\}^{\mathbb{Z}})$, *i.e.* show that the affinity in proposition (3.3.1), for finite mixtures, implies via problem [3.3.3] and the Shannon–McMillan theorem, that if $\underline{p} \in M(\{0,\dots,n\}^{\mathbb{Z}})$ and π is its ergodic decomposition, then

$$\widetilde{s}(\underline{p}) = \int_{M_e} \pi(d\omega) \widetilde{s}(\omega).$$

Q3.3.5 **[3.3.5]:** (*Affinity of the average entropy for arbitrary mixtures*)
If (Ω, S) is an invertible topological dynamical system with Ω metric and compact and if $\mu \in \mathcal{M}(\Omega, S)$ and π_μ is the ergodic decomposition on $\mathcal{M}_e(\Omega, S)$ of μ then

$$\widetilde{s}(S,\mu) = \int_{\mathcal{M}_e(\Omega,S)} \pi_\mu(d\omega) \widetilde{s}(\omega),$$

i.e. the average entropy is *affine* also for ergodic decompositions which are not finite.

Q3.3.6 **[3.3.6]:** Let ρ be a probability distribution on $\{0,\ldots n\}^N$. Construct the measure μ_0 on $\{0,\ldots,n\}^{\mathbb{Z}}$ by assigning independent probabilities q to the blocks of variables $(\sigma_i)_{i\in[kN,(k+1)N)}, k\in\mathbb{Z}$. Show that μ_0 is invariant with respect to the action of translations which are multiples of the N steps translation τ^N. Show that, for $N>1$,

$$\mu(E)=N^{-1}\sum_{j=0}^{N-1}\mu_0(\tau^j E)\qquad \text{for all } E\in\mathcal{B}(\{0,\ldots,n\}^{\mathbb{Z}})$$

defines a τ–ergodic and τ^N–mixing measure which is not τ–mixing.

Q3.3.7 **[3.3.7]:** (*Average entropy of periodic distributions*)
Compute the average entropy of the measure μ defined in problem [3.3.6] regarded as an invariant measure for the dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}},\tau^N)$.

Q3.3.8 **[3.3.8]:** (*Approximability of measures by ergodic measures*)
If m is a shift invariant distribution on $\{0,\ldots,n\}^{\mathbb{Z}}$ and if $\varepsilon>0$ and M are given, there exists a probability distribution μ ergodic on $\{0,\ldots,n\}^{\mathbb{Z}}$ and such that

$$\sum_{\sigma_1\ldots\sigma_M}|\,m(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})-\mu(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})\,|\,<\varepsilon$$

(*Hint:* Consider $N\gg M$ and the distribution q on the sequences $\{0,\ldots,n\}^N$ defined by: $q(\sigma_1,\sigma_2,\ldots,\sigma_N)=m(C^{1,\ldots,N}_{\sigma_1,\ldots,\sigma_N})$; construct μ, starting from q, m as in problem [3.3.6]).

Q3.3.9 **[3.3.9]:** (*Approximability in distribution and entropy*)
Show that, as consequence of Shannon-McMillan theorem, every ergodic $m\in M_e(\{0,\ldots,n\}^{\mathbb{Z}})$ can be approximated in distribution and entropy by a measure μ which is ergodic and "of finite type", *i.e.* built as in problem [3.3.6]. By approximation in distribution and entropy one means that given $\varepsilon>0$ and $M>0$ there exists a μ of finite type for which

$$\sum_{\sigma_1\ldots\sigma_M}|\,m(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})-\mu(C^{1\ldots M}_{\sigma_1\ldots\sigma_M})\,|<\varepsilon,\qquad |\,s(\mu)-s(m)\,|<\varepsilon.$$

(*Hint:* Make use of problems [3.3.6], [3.3.7] and [3.3.8], and of (i) in proposition (3.2.1)).

Bibliographical note to §3.3

The properties discussed in this section are well known, see for instance p. 178 in [Ru69]. There are various other properties of entropy and mainly its extensions to "non-commutative cases" that are less simple and at times quite deep; see [We79], for a review.

§3.4 Further properties of the average entropy. Generator theorem

In this section we mainly discuss definition (3.3.2) and certain simplifications in the evaluation of the extremum in (3.3.22).

D3.4.1 **(3.4.1) Definition:** (Generating partition)
Let $\mathcal{B}$ be a σ–algebra in a space Ω and $\mathcal{P} = (\{P_0,\ldots,P_n\}$, $\mathcal{Q} = \{Q_0,\ldots,Q_m\}$ be two $\mathcal{B}$-measurable partitions of Ω. Define the partition $\mathcal{P}\vee\mathcal{Q}$, generated by $\mathcal{P}$ and $\mathcal{Q}$, *to be the partition whose atoms are*

e3.4.1
$$R_{\sigma\sigma'} = P_\sigma \cap Q_{\sigma'} \quad \sigma\in\{0,\ldots,n\},\ \sigma'\in\{0,\ldots,m\}. \tag{3.4.1}$$

If μ is a probability measure on $\mathcal{B}$ define

e3.4.2
$$H(\mathcal{P},\mu) = \sum_{\sigma=0}^{n} -\mu(P_\sigma)\log\mu(P_\sigma). \tag{3.4.2}$$

*If (Ω,S,μ) is an invertible metric system with μ defined on $\mathcal{B}$ we say that $\mathcal{P}$ is μ-S–*generating *if the smallest σ–algebra that contains the sets of the partitions $S^k\mathcal{P}$, for all $k\in\mathbb{Z}$, coincides μ-mod 0 with $\mathcal{B}$. When confusion does not arise we shall simply say that $\mathcal{P}$ is S–generating.*

Remarks: (1) By using the definition given in (3.3.2), the quantity $\widetilde{s}(\mathcal{P},S,\mu)$ can also be rewritten as (cf. also (3.3.18))

e3.4.3
$$\widetilde{s}(\mathcal{P},S,\mu) = \lim_{N\to\infty} N^{-1} H(\mathcal{P}\vee S^{-1}\mathcal{P}\vee\ldots\vee S^{-(N-1)}\mathcal{P},\mu). \tag{3.4.3}$$

(2) The identity (3.4.3) implies that for all h,k integer (positive, zero or negative), with $h\le k$, one has

e3.4.4
$$\widetilde{s}(S^h\mathcal{P}\vee\ldots\vee S^k\mathcal{P},S,\mu) = \widetilde{s}(\mathcal{P},S,\mu). \tag{3.4.4}$$

(3) From general measure theory it follows that if μ is isomorphic mod 0 to the Lebesgue measure on $[0,1]$ then a necessary and sufficient condition in order that $\mathcal{P}$ be S–generating is that $\mathcal{P}$ is *S–separating mod* 0, *i.e.* that there exists $N\in\mathcal{B}$, $\mu(N)=0$, such that if $x,y\notin N$ and the $(\mathcal{P},S)$–histories of x and y coincide then x and y coincide too.

Hence in dynamical systems (Ω,S,μ) in which (Ω,S) is a topological dynamical system and μ a Borel measure every S–separating partition is generating.[1] N3.4.1

(4) From general measure theory it follows that in order that $\mathcal{P}$ be S–generating it must happen that, given $\varepsilon>0$ and $E\in\mathcal{B}$, there exists N_ε such that the partition $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ is "fine enough" so that it is possible, by taking suitable unions of its atoms, to construct a set E_ε whose symmetric difference from E, $(E\Delta E_\varepsilon = (E\backslash E_\varepsilon)\cup(E_\varepsilon\backslash E)$, is small

e3.4.5
$$\mu(E\Delta E_\varepsilon) < \varepsilon. \tag{3.4.5}$$

(5) From remark (4) it follows that if $\mathcal{P}$ is S–generating and if $\mathcal{Q} = \{Q_0,\ldots,Q_m\}$ is an arbitrary μ–measurable partition, given $\varepsilon>0$ there

[1] Every Borel measure on a complete and separable metric space is isomorphic mod 0 to the sum of the Lebesgue measure on an interval and a denumerable sum of Dirac measures, cf. problem [3.2.9] and [Pa67].

exists N_ε such that by suitably collecting the atoms of $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ into $(m+1)$ groups one can form a partition $\mathcal{Q}^\varepsilon = \{Q_0^\varepsilon, \ldots, Q_m^\varepsilon\}$ such that

e3.4.6
$$d(\mathcal{Q}, \mathcal{Q}^\varepsilon) = \sum_{i=0}^{n} \mu(Q_i \Delta Q_i^\varepsilon) < \varepsilon, \tag{3.4.6}$$

where, given in general two partitions $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ and $\mathcal{Q}' = \{Q'_0, \ldots, Q'_m\}$ with an equal number of atoms, we define

e3.4.7
$$d(\mathcal{Q}, \mathcal{Q}') = \sum_{j=0}^{n} \mu(Q_i \Delta Q'_i). \tag{3.4.7}$$

The interest of the above observations and their relevance for the problem of the actual computation of the extreme in (3.3.22) lies in the corollary to the following proposition.

P3.4.1 **(3.4.1) Proposition:** (Continuity of the average entropy of a partition) *Let (Ω, S, μ) be an invertible metric dynamical system.*
(i) If $\mathcal{P} = \{P_0, \ldots, P_n\}$, $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ are two μ–measurable partitions of Ω, then

e3.4.8
$$\widetilde{s}(\mathcal{P}, S, \mu) \le \widetilde{s}(\mathcal{P} \vee \mathcal{Q}, S, \mu) \le \widetilde{s}(\mathcal{P}, S, \mu) + \widetilde{s}(\mathcal{Q}, S, \mu). \tag{3.4.8}$$

(ii) If $\mathcal{P} = \{P_0, \ldots, P_n\}$ and $\mathcal{Q} = \{Q_0, \ldots, Q_n\}$ are two μ–measurable partitions and if $d(\mathcal{P}, \mathcal{Q}) = \sum_{i=0}^{n} (P_i \Delta Q_i) = \varepsilon < (n+1)/(n+2)$, one has

e3.4.9
$$|\widetilde{s}(\mathcal{P}, S, \mu) - \widetilde{s}(\mathcal{Q}, S, \mu)| \le \varepsilon \log(n+1) - \varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon). \tag{3.4.9}$$

Remark: The statement (ii), *Sinai's theorem*, gives continuity in $\mathcal{P}$ at fixed number, n, of atoms.

Proof: Let $i = (\sigma_0 \ldots \sigma_{N-1})$, $j = (\lambda_0 \ldots \lambda_{N-1})$, $\sigma_k \in \{0, \ldots, n\}$, $\lambda_k \in \{0, \ldots, m\}$ and

e3.4.10
$$\begin{aligned} p_i &= \mu\Big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\Big), \qquad p_{ij} = \mu\Big(\bigcap_{k=0}^{N-1} S^{-k}(P_{\sigma_k} \cap Q_{\lambda_k})\Big), \\ p'_j &= \mu\Big(\bigcap_{k=0}^{N-1} S^{-k} Q_{\lambda_k}\Big); \end{aligned} \tag{3.4.10}$$

one has $\sum_i p_i = \sum_{ij} p_{ij} = 1$, $\sum_j p_{ij} = p_i$, $\sum_i p_{ij} = p'_j$. Then (3.4.8) is derived from the relation between the approximants $\sum_i(-p_i \log p_i)$, $\sum_{ij}(-p_{ij} \log p_{ij})$, $\sum_i(-p_i \log p_i + \sum_j -p'_j \log p'_j)$ obtained in (3.3.15) from (3.3.14).

To show (ii) let $\mathcal{R} = \{R_0, \ldots, R_n, R_{n+1}\}$ be the partition of Ω into $n+2$ sets defined by

e3.4.11
$$\begin{aligned} R_0 &= P_0 \Delta Q_0, \\ R_i &= (P_i \Delta Q_i) / \mathop{\cup}_{j=0}^{i-1} (P_j \Delta Q_j), \qquad i = 1, \ldots, n, \\ R_{n+1} &= \Omega / \mathop{\cup}_{i=0}^{n} R_i . \end{aligned} \tag{3.4.11}$$

We have $X \stackrel{def}{=} \mu(R_{n+1}) \ge 1 - \varepsilon$ and $\sum_{i=0}^{n} \mu(R_i) = 1 - X \le \varepsilon$. Furthermore (3.4.8) implies

e3.4.12
$$\widetilde{s}(\mathcal{P} \vee \mathcal{R}, S, \mu) \le \widetilde{s}(\mathcal{P}, S, \mu) + \widetilde{s}(\mathcal{R}, S, \mu), \tag{3.4.12}$$

and the similar relation with $\mathcal{Q}$ instead of $\mathcal{P}$. The relation $(\mathcal{P} \vee \mathcal{R}) = (\mathcal{Q} \vee \mathcal{R})$ (that follows immediately from the definitions) implies together with (3.4.12)

e3.4.13
$$|\widetilde{s}(\mathcal{P}, S, \mu) - \widetilde{s}(\mathcal{Q}, S, \mu)| \le \widetilde{s}(\mathcal{R}, S, \mu). \tag{3.4.13}$$

But from the second of (3.3.12) with $k = 0$ it follows

e3.4.14
$$\widetilde{s}(\mathcal{R}, S, \mu) \le \sum_{\sigma=0}^{n+1} -\mu(R_\sigma) \log \mu(R_\sigma), \tag{3.4.14}$$

which implies

e3.4.15
$$\begin{aligned} \widetilde{s}(\mathcal{R}, S, \mu) &\le -X \log X + \sum_{\sigma=0}^{n} -\mu(R_\sigma) \log \mu(R_\sigma) \equiv \\ &\equiv -X \log X - (1-X) \log(1-X) + (1-X) \sum_{\sigma=0}^{n} -\frac{\mu(R_\sigma)}{(1-X)} \log \frac{\mu(R_\sigma)}{(1-X)}; \end{aligned} \tag{3.4.15}$$

therefore the sum can be bounded above by $\log(n+1)$ because of the identity $\sum_{\sigma=0}^{n} \frac{\mu(R_\sigma)}{1-X} = 1$. Hence

e3.4.16
$$\widetilde{s}(\mathcal{R}, S, \mu) \le -X \log X - (1-X) \log(1-X) + (1-X) \log(n+1), \tag{3.4.16}$$

from which (3.4.9) follows because the function in the r.h.s. of (3.4.16) is monotonic decreasing between $(1 - (n+1)/(n+2))$ and 1, and one has $X \ge 1 - \varepsilon$. ■

C3.4.1 **(3.4.1) Corollary:** (Generator theorem)
Let (Ω, S, μ) be an invertible metric dynamical system. If $\mathcal{P} = \{P_0, \ldots, P_n\}$, $\mathcal{Q} = \{Q_0, \ldots, Q_m\}$ are two μ–measurable partitions and $\mathcal{P}$ is S–generating, then

e3.4.17
$$\widetilde{s}(\mathcal{Q}, S, \mu) \le \widetilde{s}(\mathcal{P}, S, \mu) = \widetilde{s}(S, \mu) \tag{3.4.17}$$

Remark: The above theorem is due to Sinai.

Proof: Let $\varepsilon > 0$ and let $\mathcal{Q}^\varepsilon$ be a partition obtained by forming unions of atoms of $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ and such that $d(\mathcal{Q},\mathcal{Q}^\varepsilon) < \varepsilon$, cf. remark (5) to definition (3.4.1). Then, by proposition (3.4.1)

e3.4.18
$$|\widetilde{s}(\mathcal{Q},S,\mu) - \widetilde{s}(\mathcal{Q}^\varepsilon,S,\mu)| \xrightarrow[\varepsilon\to 0]{} 0. \tag{3.4.18}$$

But it is clear that $\mathcal{Q}^\varepsilon$ is less fine than $\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$ (and therefore there is a partition $\mathcal{R}$ of Ω such that $\mathcal{Q}^\varepsilon \vee \mathcal{R} = \vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P}$); hence by (3.4.8) and (3.4.4)

e3.4.19
$$\widetilde{s}(\mathcal{Q}^\varepsilon,S,\mu) \le \widetilde{s}\Big(\vee_{-N_\varepsilon}^{N_\varepsilon} S^{-k}\mathcal{P},S,\mu\Big) = \widetilde{s}(\mathcal{P},S,\mu), \tag{3.4.19}$$

so that the first of (3.4.17) is proved. The second follows by the arbitrariness of $\mathcal{Q}$ and the definition in (3.3.22). ∎

C3.4.2 **(3.4.2) Corollary:** *Let (Ω,S,μ) be an invertible metric system mod 0 with μ defined on a σ–algebra $\mathcal{B}$. Let $\mathcal{P}_1$, $\mathcal{P}_2,\ldots$ be a sequence of μ–measurable partitions such that $\mathcal{P}_{n+1}$ refines $\mathcal{P}_n$ (i.e. such that the atoms of $\mathcal{P}_n$ are unions of atoms of $\mathcal{P}_{n+1}$) and such that the smallest σ–algebra that contains $\mathcal{P}_n$ and all its images under S, for all $n = 1,2\ldots$, is $\mathcal{B}$ (a* generating sequence of partitions*). One has*

e3.4.20
$$\widetilde{s}(S,\mu) = \lim_{n\to\infty} \widetilde{s}(\mathcal{P}_n,S,\mu). \tag{3.4.20}$$

The proof will be left to the reader. Finally an application.

P3.4.2 **(3.4.2) Proposition:** (Entropy bound for smooth measures)
Let Ω be a compact Riemannian manifold of class C^∞ and dimension r, S be a C^∞ diffeomorphism of Ω, and μ be an S–invariant Borel measure equivalent to the volume measure μ_0 on Ω (i.e. let $\mu = \rho\mu_0$ with $\rho^{\pm 1} \in L_1(\mu_0)$). The average entropy of S with respcet to μ can be bounded as

e3.4.21
$$\widetilde{s}(S,\mu) < r\,\log\lambda, \tag{3.4.21}$$

in terms of the largest expansion coefficient λ of the line elements of Ω under the action of $S^{\pm 1}$.

Remark: This is, essentially, again Kouchnirenko's theorem, cf. proposition (3.1.2): this time it is formulated on the average entropy and without the hypothesis of ergodicity of μ (so that the average entropy cannot be identified with the entropy).

Proof: Since Ω is locally diffeomorphic to $\mathbb{R}^r$ it is clear that there exists a sequence of partitions with sets with a piecewise C^∞ boundary and which verify the hypothesis of corollary (3.4.2). We can in fact suppose that the atoms of such partitions have always diameter less than a prefixed $\delta > 0$. We shall fix δ so that the isoperimetric inequality holds for every C^∞–regular set P with diameter diam $(P) < \delta$

e3.4.22
$$|\mu(P)| \le \Gamma |\partial P|^{r/(r-1)}, \tag{3.4.22}$$

where $|\partial P| =$ (area of the surface ∂P), and Γ is a suitable P–independent constant.

By corollary (3.4.2) it will suffice to show that for an arbitrary partition $\mathcal{P}$, C^∞–regular, with atoms of diameter $\le \delta$ one has $s(\mathcal{P}, S, \mu) \le r \log \lambda$.

Fixed $\mathcal{P}$ and $\eta > 0$ and proceeding as in proposition (3.1.2) we split the specifications $\sigma_0, \ldots, \sigma_{N-1}$ of length N, $\sigma_i \in \{0, \ldots, n\}$ into two classes:

e3.4.23
$$\begin{aligned} &\mathcal{C}_1(N) = \big\{\sigma_0, \ldots, \sigma_{N-1} \big| \mu\big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\big) > e^{-N\eta} \lambda^{-rN}\big\}, \\ &\mathcal{C}_2(N) = \big\{0, \ldots, n\big\}^N \backslash \mathcal{C}_1(N), \end{aligned} \tag{3.4.23}$$

and, as in the case of the mentioned proposition, we deduce that if

e3.4.24
$$p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} = \mu\Big(\bigcap_{k=0}^{N-1} S^{-k} P_{\sigma_k}\Big) \tag{3.4.24}$$

one has, for a suitable $C > 0$,

e3.4.25
$$X \stackrel{def}{=} \sum_{\sigma_0 \ldots \sigma_{N-1} \in \mathcal{C}_2(N)} p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix} \le C e^{-N\eta}. \tag{3.4.25}$$

Hence. having set $j = (\sigma_0, \ldots, \sigma_{N-1})$, $p_j = p\begin{pmatrix} 0 \ldots N-1 \\ \sigma_0 \ldots \sigma_{N-1} \end{pmatrix}$,

e3.4.26
$$\begin{aligned} &\sum_j -p_j \log p_j = \sum_{j \in \mathcal{C}_1(N)} -p_j \log p_j + \sum_{j \in \mathcal{C}_2(N)} -p_j \log p_j \le \\ &\le (r \log \lambda + \eta) N + X \sum_{j \in \mathcal{C}_2(N)} -(p_j/X) \log p_j = \\ &= (r \log \lambda + \eta) N - X \log X + X \sum_{j \in \mathcal{C}_2(N)} (-p_j/X) \log(p_j/X) = \\ &= (r \log \lambda + \eta) N - X \log X + X N \log(n+1), \end{aligned} \tag{3.4.26}$$

because {number of elements in $\mathcal{C}_2(N)$} $\le (n+1)^N$. Dividing the (3.4.26) by N and passing to the limit as $N \to \infty$ the terms containing X tend to zero and one finds $s(\mathcal{P}, S, \mu) \le \eta + r \log \lambda$, for every $\eta > 0$. ∎

Problems for §**3.4** (*Complements to Shannon–McMillan's theorem*)

Q3.4.1 **[3.4.1]:** (*Average entropy of a Bernoulli scheme*)
Consider the Bernoulli scheme on $\Omega = \{0, \ldots, n\}^{\mathbb{Z}}$ that associates with the symbols

the probabilities $\pi_0, \ldots, \pi_n$, $\sum_i \pi_i = 1$. Denoting by m the corresponding probability measure on Ω consider the system (Ω, τ, m) and show that its entropy $s = s(\tau, m)$ and its average entropy $\widetilde{s} = \widetilde{s}(\tau, m)$ satisfy $s = \widetilde{s} = -\sum_i \pi_i \log \pi_i$. (*Hint:* Recall the definition in (3.3.22), and use problem [3.1.8] and the Shannon–McMillan theorem.)

Q3.4.2 **[3.4.2]:** Show the existence of Bernoulli schemes (with infinitely many different symbols) with infinite average entropy. (*Hint:* Choose a sequence $\{a_n\}_{n=0}^{\infty}$ such that $\sum_{n=1}^{\infty} a_n < \infty$ and $\sum_{n=1}^{\infty} a_n \log a_n < \infty$, and normalize to 1 the first series; take for instance $a_n = 1/n \log^2 n$ for $n \ge 2$.)

Q3.4.3 **[3.4.3]:** (*Average entropy of a Markov chain*)
Compute the average entropy of the dynamical system $(\{0,1\}^{\mathbb{Z}}, \tau, m)$ where m is the distribution built in [2.3.8]. Show that $\widetilde{s}(\tau, m) = -\sum_{\sigma,\sigma'} \pi^*_\sigma \pi_{\sigma'} \frac{T_{\sigma\sigma'}}{\lambda} \log \frac{T_{\sigma\sigma'}}{\lambda}$.

Q3.4.4 **[3.4.4]:** Estimate the average entropy of the system of the example (1.2.5).

Q3.4.5 **[3.4.5]:** Show that two Bernoulli schemes with different entropy cannot be isomorphic mod 0.

Q3.4.6 **[3.4.6]:** (*A non–generating partition with maximal entropy*)
Consider the Markov process (cf. problem [2.3.8]) with transition matrix $T_{\sigma\sigma'} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$ and show that it admits a *non-generating* partition $\mathcal{Q}$ with largest entropy. (*Hint:* $Q_0 = \{\underline{\sigma} | \sigma_0 = \sigma_1\}$, $Q_1 = \{\underline{\sigma} | \sigma_0 \ne \sigma_1\}$, $s(\mathcal{Q}, \tau) = \log 2$.)

Q3.4.7 **[3.4.7]:** Let μ_n be a sequence of invariant distributions for a topological dynamical system (Ω, S) and suppose that there is a partition $\mathcal{P}$ which is S–generating for all (Ω, S, μ_n). Then if μ_n converges weakly to μ, *i.e.* $\mu_n(f) \xrightarrow[n\to\infty]{} \mu(f)$ for all continuous functions f, the average entropies of μ_n verify $\limsup_{n\to\infty} \widetilde{s}(S, \mu_n) \le \widetilde{s}(S, \mu)$. (*Hint:* See (3.3.13).)

Q3.4.8 **[3.4.8]:** (*Factors of arbitrary entropy*)
Show that given a dynamical system (Ω, S, μ) with a given partition $\mathcal{P}$, such that the measure μ on Ω is isomorphic mod 0 to the Lebesgue measure on $[0,1]$, there exists a partition $\mathcal{P}'$ of Ω such that $s(\mathcal{P}', S) = a\, s(\mathcal{P}, S)$ with a arbitrarily fixed in $[0,1]$. The dynamical system (Ω, S, μ') where μ' is the restriction of μ to the σ–algebra generated by $\mathcal{P}'$ is called a *factor* of the original dynamical system. (*Hint:* Use Sinai's theorem and the existence of a "continuous" family interpolating between the trivial partition and $\mathcal{P}$; proceed as in problem [3.2.8]). To appreciate the generality of this property note its relation with problems [3.2.1], [3.2.11] and [3.2.12].

Q3.4.9 **[3.4.9]:** (*Stacks for mixing systems*)
Consider a system $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ with μ mixing and non-atomic (hence such that $\sup \mu(C^{0\ldots k}_{\sigma_0\ldots\sigma_k}) \xrightarrow[k\to\infty]{} 0$) and, for simplicity, assume $\mu(C^{0\ldots k}_{\sigma_0\ldots\sigma_k}) > 0$ for all $\sigma_0, \ldots, \sigma_k$. Show that, given $\varepsilon > 0$ and $N > 0$ integer, there exists a Borel set F such that the sets F, $\tau F, \ldots, \tau^{N-1} F$ are pairwise disjoint and $\mu\Big(\cup_{i=0}^{N-1} \tau^i F\Big) \ge 1 - \varepsilon$. (*Hint:* : Let $M \gg M' \gg N$ and choose M' such that sup $\mu(C^{0\ldots M'}_{\sigma_0\ldots\sigma_{M'}}) < \varepsilon/4N$. Consider the string with M' elements $00000\ldots001$ and call $\widetilde{F}$ the union of the cylinders with base $(kN, \ldots, kN + M')$ and specification $00000\ldots001$ for $k = 0, 1, 2, \ldots, [M/N]$ and $F = \widetilde{F} / \cup_{i=1}^{N-1} \tau^i \widetilde{F}$ (*i.e.* F is the set of sequences for $k = 0, 1, 2, \ldots, [M/N]$ (ie F is the set of sequences containing the string $00000\ldots001$ with the first zero in position 0 or N or $2N$, *etc.*, for the first time between 0 and $M - 1$). It is then clear that $\tau^i F \cap \tau^j F = \emptyset$ $\forall 0 \le i \ne j \le N-1$, and $\{0, \ldots, n\}^{\mathbb{Z}} / \cup \tau^i F$ contain two types of points: those sequences $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ that never contain between 0 and $M - 1$ the string $00000\ldots001$ and those that do contain it between 0 and $N - 1$ The set of the points of the second type has measure lower than $(\varepsilon/4N)N = \varepsilon/4$ and the set of the first type has infinitesimal measure as $M \to \infty$.)

Q3.4.10 **[3.4.10]:** (*Rohlin's stack*)
Under the hypothesis of problem [3.4.9] let $\mathcal{Q} = (Q_0, \ldots, Q_k)$ be a Borel partition of

$\{0,\ldots,n\}^{\mathbb{Z}}$. It is possible, given N and $\varepsilon > 0$, to find F so that $\mu(Q_i \cap F)/\mu(F) = \mu(Q_p)$ for all $i = 0,\ldots,k$, and $\tau^i F \cap \tau^j F = \emptyset\ \forall 0 \le i \ne j \le N$. (*Hint:* Let F_0 be the set whose existence, in correspondence of the given N and ε, is assured by the result in problem [3.4.9] and verifying the properties described therein. Set $F_t = \tau^t F_0$, $t \in \mathbb{Z}$, and use the mixing property (assumed to hold for μ) to infer that

$$\lim_{t\to\infty} \sum_{i=1}^{k} \left| \frac{\mu(Q_i \cap F_t)}{\mu(F_t)} - \mu(Q_i) \right| = \lim_{t\to\infty} \eta_t = 0$$

If one chooses $\eta_t \ll \mu(F_t) = \mu(F_0)$ (by the τ-invarience of μ) it is clear that, being $(\{0,\ldots,n\}^{\mathbb{Z}}, \mu)$ isomorphic mod 0 to the Lebesgue measure on $[0,1]$ (cf. problem [3.2.12]) and regarding in this way $Q_0,\ldots,Q_k, F_t$ as sets of $[0,1]$ it is possible to take out of F_t a set $\Delta \subset F_t$ of points having small measure with respect to $\eta_t \ll \mu(F_0)$ to obtain that $\mu(Q_i \cap (F_t \setminus \Delta)) = \mu(Q_i)\mu(F_t \setminus \Delta)$ without deteriorating the bound on the measure of $\cup_{i=0}^{N-1} \tau^i (F_t \setminus \Delta))$, *i.e.* keeping it larger than $1-\varepsilon$).

Remark: This statement (*Rohlin's stack theorem*) does not require the hypothesis of mixing: ergodicity of μ suffices (the same can be said also for the result of the preceding problem [3.4.9]); however the proof is, in the latter cases, somewhat more elaborate.

The following problems provide a guided proof of the statement that two mixing shifts of equal entropy contain copies of each other (Sinai).

Q3.4.11 **[3.4.11]:** Let $(\{0,\ldots,n\}^{\mathbb{Z}}, \tau, \mu)$ and $(\{0,\ldots,n'\}^{\mathbb{Z}}, \tau, \mu')$ be two mixing shifts with $s(\mu',\tau) > s(\mu,\tau)$. Consider the partitions $\mathcal{P}$ and $\mathcal{P}'$ of $\{0,\ldots,n\}^{\mathbb{Z}}$ and of $\{0,\ldots,n'\}^{\mathbb{Z}}$ into the cylinders with base 0 (*i.e.* $\mathcal{P} = \{C^0_0, C^0_1, \ldots, C^0_n\}$ and $\mathcal{P}' = \{C'^0_0, \ldots, C'^0_{n'}\}$ respectively).

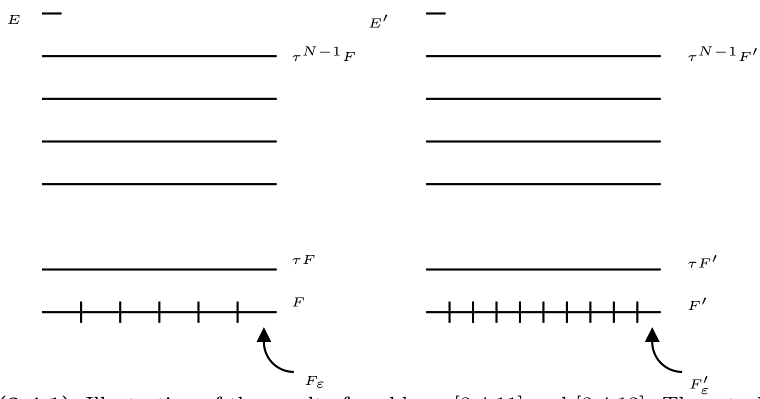
Given $\varepsilon > 0$ and u integer > 0 consider the partitions $\mathcal{P}_N = \vee_0^{N-1} \tau^i \mathcal{P}$ and $\mathcal{P}'_N = \vee_0^{N-1} \tau^i \mathcal{P}'$ and choose $N > N(u,\varepsilon)$, where $N(u,\varepsilon)$ is such that for $N > N(u,\varepsilon)$ the properties stated in problem [3.2.1] hold. Let $F \subset \{0,\ldots,n\}^{\mathbb{Z}}$ and $F' \subset \{0,\ldots,n'\}^{\mathbb{Z}}$ be two sets for which (cf. problem [3.4.10])

$$\frac{\mu(Q \cap F)}{\mu(F)} = \mu(Q), \text{ for all } Q \in \mathcal{P}_N\,; \quad \frac{\mu'(Q' \cap F')}{\mu'(F')} = \mu'(Q') \text{ for all } Q' \in \mathcal{P}'_N$$

and simultaneously $\tau^i F \cap \tau^j F = \emptyset = \tau^i F' \cap \tau^j F'$, for all $i \ne j$, $i,j = 0,\ldots,N-1$. Represent F and F' as two intervals (see Fig. (3.4.1)), and represent as intervals also the sets

$$\tau F, \ldots, \tau^{N-1} F, \qquad \tau F', \ldots, \tau^{N-1} F'$$

$$E = \{0,\ldots,n\}^{\mathbb{Z}} / \bigcup_{i=0}^{N-1} \tau^i F, \qquad E' = \{0,\ldots,n'\}^{\mathbb{Z}} / \bigcup_{i=0}^{N-1} \tau^i F'.$$

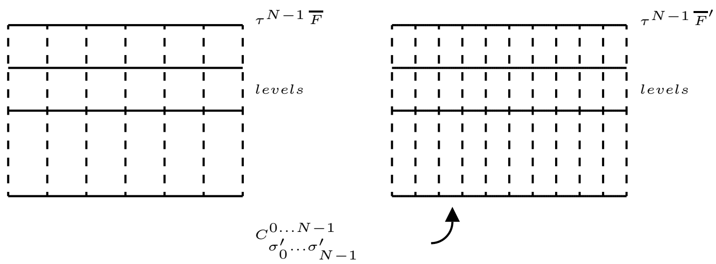


F3.4.1 **Fig.(3.4.1)** Illustration of the result of problems [3.4.11] and [3.4.12]. The sets F_ε and F'_ε are defined in problem [3.4.12]. The interval F, base of the stack, is divided into smaller intervals representing the sets $F\cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ with $\sigma_0\ldots,\sigma_{N-1}$ chosen in the large frequency collection $\mathcal{C}_{1,\varepsilon,u}(N)$, except the rigthmost interval, denoted F_ε, which represents the intersection of F with $\cup C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the cylinders with $\sigma_0,\ldots,\sigma_{N-1}$ in the collection $\mathcal{C}_{2,\varepsilon,u}$ of rare strings.

Check that the action of τ is naturally represented as an upward translation except for its action on E and E' and on $\tau^{N-1}F$ and $\tau^{N-1}F'$ (where it acts differently and in a way which, in general, is not simply representable graphically): in this representation the measures μ and μ' are represented by the Lebesgue measure on the several intervals whose lengths, in every stack, add up to 1.

Q3.4.12 **[3.4.12]:** In the situation of problem [3.4.11] draw as segments the several elements of the partition induced by $\mathcal{Q}\equiv\mathcal{P}_N$ on F (cf. Fig. (3.4.1)): $(Q\cap F)_{Q\in\mathcal{Q}}$; and represent as an interval also the set $F_\varepsilon=\cup F\cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ where the union is over the choices of $(\sigma_0,\ldots,\sigma_{N-1})$ in the collection $\mathcal{C}_{2,\varepsilon,u}(N)$ introduced in [3.2.1]. Perform the same construction over the stack relative to F'. Check that $\mu(F_\varepsilon)\le\varepsilon\mu(F)$, $\mu'(F'_\varepsilon)\le\varepsilon\mu'(F')$.

Q3.4.13 **[3.4.13]:** In the situation of problems [3.4.11] and [3.4.12] set $\overline{F}=F/F_\varepsilon$ and $\overline{F}'=F'/F'_\varepsilon$: such sets are split into disjoint parts by the partitions $\overline{F}\cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ and $\overline{F}'\cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}}$ with $\sigma_0,\ldots,\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon,u}(N)$ and $\sigma'_0,\ldots,\sigma'_{N-1}\in\mathcal{C}'_{1,\varepsilon,u}(N)$ (cf. problems [3.4.12] and [3.2.1]). Fix also $N\gg u$. We shall call "level" of the stack every image $\tau^j(\overline{F}\cap C^{0\ldots N-1}_{\sigma_0\ldots\sigma_{N-1}})$, with $0\le j\le N-1$ and $\sigma_0,\ldots,\sigma_{N-1}\in\mathcal{C}_{1,\varepsilon,u}(N)$. Likewise we define the levels for the stack with base $\overline{F}'$. See Fig. (3.4.2).



F3.4.2 **Fig.(3.4.2)** Illustration of the stack levels of the construction of problem [3.4.13].

Remark that by the Shannon–McMillan theorem, if $2\varepsilon<s(\mu')-s(\mu)$ (where ε is the same as in proposition (3.2.1)) and N is large enough, the "columns of levels" with base F' are much more numerous of those with base F (the number of columns is essentially given by (3.2.4)).

Put arbitrarily into correspondence every column with base on F with a different column with base on F' by assigning to the j–th level of a column with base on F the symbol σ'_j of the column with base on F' associated with it.

Collecting then the levels that, in this construction, come to have labels equal to $0,1,\ldots,n'$ respectively, form a partition of $\cup_{j=0}^{N-1}\tau^j\,\overline{F}$ in $n'+1$ sets $\tilde P'_0,\ldots,\tilde P'_{n'}$: imagine extending such a partition to a partition $\tilde{\mathcal{P}}'=\{\tilde P'_0,\ldots,\tilde P'_n\}$ of the whole space $\Omega=\{0,\ldots,n\}^{\mathbb{Z}}$, arbitrarily.

Show that if $\mathcal{F}$ is the partition of $\{0,\ldots,n\}^{\mathbb{Z}}$ into $\overline{F}$ and $\{0,\ldots,n\}^{\mathbb{Z}}/\overline{F}$ we get that $\vee_{-N}^{N}\tau^i(\widetilde{\mathcal{P}}'\vee\mathcal{F})$ contains a partition $\widetilde{\mathcal{P}}$ with n elements formed by unions of its atoms and such that: $d(\mathcal{P},\widetilde{\mathcal{P}})<2\varepsilon$. (*Hint:* The partition $\vee_{-N}^{N}\tau^i(\widetilde{\mathcal{P}}'\vee\mathcal{F})$ reduces on $\cup_{i=0}^{N-1}\tau^i\,\overline{F}$ to the partition into levels and, therefore, we reconstruct from it, by means of operations of unions of atoms, the partition $\mathcal{P}$ on $\cup_{i=0}^{N-1}\tau^i\,\overline{F}$, *etc.*)

Q3.4.14 **[3.4.14]:** Show that the partition $\widetilde{\mathcal{P}}$ built via the procedure illustrated in problems

[3.4.9] to [3.4.13], is such that

$$\sum_{\sigma'_0\ldots\sigma'_{n-1}} |\mu'(C^{0\quad u-1}_{\sigma'_0\ldots\sigma'_{n-1}}) - \mu(P^{0\ldots u-1}_{\sigma'_0\ldots\sigma'_{n-1}})| \xrightarrow[N\to\infty,\varepsilon\to 0]{} 0$$

$$|s(\widetilde{\mathcal{P}}',\tau) - s(\mathcal{P},\tau)| \xrightarrow[\varepsilon\to 0]{} 0$$

It is therefore possible to "represent a process with larger entropy into one of lower entropy and within a prefixed approximation" without losing in entropy more than a prefixed quantity beyond the obviously necessary loss $(s(\mu',\tau') - s(\mu,\tau))$. (*Hint:* Use the strengthened form of the Shannon–McMillan theorem in problems [3.2.1], [3.2.3] and the fact that if N is very large the strings (short because the length u is fixed) $\sigma'_0,\ldots,\sigma'_{u-1}$ appear with frequency almost equal to $\mu'(C^{0\ldots u-1}_{\sigma'_0\ldots\sigma'_{n-1}})$ in the sequences of $\mathcal{C}_{1,\varepsilon,u}(N)$.)

Q3.4.15 **[3.4.15]:** (*Copying a dynamical system into another*)
Deduce from the results of problem [3.4.14] that, under the same hypotheses of [3.4.11] but with $s(\mu',\tau) = s(\mu,\tau) = s$, then, given a positive integer u and given $\varepsilon > 0$, it is possible to "copy" the first dynamical system into the second in the sense that it is possible to construct a partition $\overline{\mathcal{P}}$ of $\{0,\ldots,n'\}^{\mathbb{Z}}$ so that

$$|s(\overline{\mathcal{P}},\tau) - s| < \varepsilon$$

$$\sum_{\sigma_0\ldots\sigma_{n-1}\in\{0,\ldots,n\}^u} |\mu(C^{0\ldots u-1}_{\sigma_0\ldots\sigma_{n-1}}) - \mu'(\overline{P}^{0\ldots u-1}_{\sigma_0\ldots\sigma_{u-1}})| < \varepsilon$$

(*Hint:* Find in $\{0,\ldots,n'\}^{\mathbb{Z}}$ a partition $\widetilde{\widetilde{\mathcal{P}}}'$ such that $s(\widetilde{\widetilde{\mathcal{P}}}',\tau) = s - 2\varepsilon$. Apply then again the construction of the [3.4.11], [3.4.12], [3.4.13] replacing $\mathcal{P}'$ with $\widetilde{\widetilde{\mathcal{P}}}'$ *etc*).

Bibliographical note §3.4

The generator theorem of Sinai is discussed here following Appendix 19, p. 163, of [AA69]. The problems on Rohlin's stack and its applications are drawn from Ornstein's theory of Bernoulli shifts isomorphisms, [Or74].

CHAPTER IV

Markovian pavements

§4.1 Histories compatibility. Markovian pavements

In the previous sections the problem of studying the statistics of motions of a dynamical system (Ω, S), as seen from a partition $\mathcal{P}$, has been shown to be equivalent to studying probability distributions on the space of the sequences of symbols associated with a partition $\mathcal{P}$ (cf. proposition (2.3.2)).

Upon further thought it is, however, clear that the analysis presented so far can be of little help in concrete problems. It is true that on $\{0,\ldots,n\}^{\mathbb{Z}}$ it is possible to study vast classes of ergodic distributions, and such a study can be also developed in a rather detailed and concrete way, but it is hard to give criteria that select those probability distributions (or measures) m that are relevant for the statistical study of motions of (Ω, S). Or, in other words, it is hard to give criteria that guarantee that $m(\widehat{\Omega}) = 1$, cf. proposition (2.3.2), and allow us to identify the possible symbolic motions, *i.e.* to recognize whether $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$ is the $(\mathcal{P}, S)$–history of some point $x \in \Omega$.

In general the values $\sigma_i(x)$, $i \in \mathbb{Z}$, are linked by very intricate relations and understanding them means a very detailed understanding of the motions structure. On the other hand it is necessarily so: indeed the action of S regarded as an action on the symbolic histories is trivial, being reduced to a mere shift (*i.e.* to a translation). Complexity of a given dynamical system must necessarily be hidden in the map, called *code* in Section §1.4, which associates with every $x \in \Omega$ its $(\mathcal{P}, S)$–history on $\mathcal{P}$; at least in the cases in which $\mathcal{P}$ is generating, *i.e.* it is fine enough to provide a faithful description of the motion.

The simplest compatibility condition between elements of $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$

is, perhaps, what can be called a *local condition of compatibility*.

D4.1.1 **(4.1.1) Definition:** (Compatibility matrix)
A $(n+1)\times(n+1)$ *matrix* T *with entries* $T_{\sigma\sigma'}$ *equal to* 0 *or to* 1 *will be called a* compatibility matrix*. Such a matrix will be called* transitive *if for every pair* σ,σ' *there is a suitable integer* $a_{\sigma\sigma'}$ *such that* $T^{1+a_{\sigma\sigma'}}_{\sigma\sigma'}>0$*. It will be called* mixing *if there is* $a\ge 0$ *such that* $T^{1+a}_{\sigma\sigma'}>0$ *for all* σ,σ' *and* a *is called the* mixing time *of* T*.*
A sequence $\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}$ *will be called* T–compatible *or* admissible *if and only if every pair of adjacent symbols that appear in* $\underline{\sigma}$ *is admissible, i.e.* $T_{\sigma_i\sigma_{i+1}}=1\,\forall i\in\mathbb{Z}$ *or, equivalently, if and only if*

e4.1.1
$$\prod_{i=-\infty}^{+\infty} T_{\sigma_i\sigma_{i+1}}=1. \tag{4.1.1}$$

We shall call $\{0,\ldots,n\}^{\mathbb{Z}}_T\subset\{0,\ldots,n\}^{\mathbb{Z}}$ *the (closed and translation invariant) subset of the sequences* $\underline{\sigma}$ *that verify (4.1.1).*
The dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$*, with* $(\tau\underline{\sigma})_i=\sigma_{i+1}$*, is often called a* translation *or* shift*, and the dynamical system* $(\{0,\ldots,n\}^{\mathbb{Z}}_T,\tau)$ *is called a* subshift of finite type*.*

One could naively think that by suitably selecting $\mathcal{P}$ it should be possible to obtain that the (Ω,S)–histories are all and only those sequences of $\{0,\ldots,n\}^{\mathbb{Z}}$ which verify (4.1.1) for some suitable T. It is however clear that the "totally disconnected" topological structure of $\{0,\ldots,n\}^{\mathbb{Z}}$ can be topologically incompatible with the structure of Ω that, very often, is a Riemannian manifold. In such cases a code that transforms the dynamics into a symbolic dynamics that is a subshift of finite type, even if existent, could not fail to show some pathology (like points not well coded, discontinuities, *etc.*), which in turn we must necessarily expect to produce, sooner or later, difficulties in the theory.
Nevertheless the simplicity of the condition (4.1.1) is so captivating that it is worth looking after systems that admit such a simple symbolic representations.

We shall therefore analyze topological dynamical systems (Ω,S) and we shall try to isolate some further conditions on (Ω,S) that will allow us to describe through simple symbolic dynamics large classes of probability distributions (or measures) and, more precisely, classes of probability distributions like the following ones.

D4.1.2 **(4.1.2) Definition:** (Topological probability distributions, topological pavements)
Given a compact topological space Ω *we call* topological probability distributions *the probability distributions* $\mu\in\mathcal{M}^0(\Omega)$ *defined on a* σ*–algebra of sets containing the Borel sets of* Ω *which satisfy the condition*

e4.1.2
$$\mu(G)>0 \qquad \textit{for all open sets } G. \tag{4.1.2}$$

If (Ω, S) is a topological dynamical system we denote by $\mathcal{M}_t(\Omega, S)$ the topological probability distributions that are S–invariant.
We say that, $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ is a pavement of Ω if it is a covering of Ω by closed sets, which are the closures of their internal points and are such that $Q_i \cap Q_j = \partial Q_i \cap \partial Q_j$ for all $i \neq j$.

The key notion that, as we shall see, allows us to use effectively symbolic dynamics for the dynamical systems for which it has a meaning, is that of *Markovian pavement* (also called *Markovian partition*; see remark (4) after the following definition).

D4.1.3 **(4.1.3) Definition:** (Markovian pavements)
Let (Ω, S) be an invertible topological dynamical system. Given a pavement $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ of Ω set $T_{\sigma\sigma'} = 1$ if $\mathrm{int}(Q_\sigma) \cap \mathrm{int}(S^{-1}Q_{\sigma'}) \neq \emptyset$ and $T_{\sigma\sigma'} = 0$ otherwise, $\sigma, \sigma' \in \{1, \ldots, q\}$, $q \geq 2$. We shall say that $\mathcal{Q}$ is Markovian *if the following holds.*
(i) The set

e4.1.3
$$\mathcal{X}(\underline{\sigma}) = \bigcap_{k=-\infty}^{+\infty} S^{-k} Q_{\sigma_k} \tag{4.1.3}$$

is not empty and consists of a single point $X(\underline{\sigma})$ for all $\underline{\sigma}$ such that

e4.1.4
$$\prod_{i=-\infty}^{+\infty} T_{\sigma_i \sigma_{i+1}} = 1. \tag{4.1.4}$$

Furthermore, for such $\underline{\sigma}$ the sets $\cap_{-N}^{N} S^{-k} Q_{\sigma_k}$ contain internal points for all N. We shall call the matrix T the compatibility matrix for $\mathcal{Q}$, *and the space $\underline{\sigma} \in \{1, \ldots, q\}_T^{\mathbb{Z}}$ of the sequences $\underline{\sigma}$ satisfying (4.1.4) will be called the space of the T–*compatible *sequences, or simply* compatible *sequences if no confusion is possible.*
(ii) The correspondence $\underline{\sigma} \to X(\underline{\sigma})$ between $\{1, \ldots, q\}_T^{\mathbb{Z}}$ and Ω is Hölder continuous, i.e. there exist $C, a > 0$ such that

e4.1.5
$$d(X(\underline{\sigma}), X(\underline{\sigma}')) \leq C\, d(\underline{\sigma}, \underline{\sigma}')^a, \tag{4.1.5}$$

where we define the distance between $\underline{\sigma}$ and $\underline{\sigma}'$ on $\{1, \ldots, q\}^{\mathbb{Z}}$ by

e4.1.6
$$d(\underline{\sigma}, \underline{\sigma}') = \exp\big(-\nu(\underline{\sigma}, \underline{\sigma}')\big), \tag{4.1.6}$$

where $\nu(\underline{\sigma}, \underline{\sigma}')$ is the largest integer j such that $\sigma_i = \sigma'_i$ for $|i| \leq j$.
(iii) There exists an upper bound $M < \infty$ on the number of compatible sequences mapped into a given point x, i.e. the inverse map X^{-1} verifies $\big| X^{-1}(x) \big| \leq M$, for all $x \in \Omega$. The number M will be called multiplicity of the code.
(iv) Setting $\partial_i = \partial Q_i$, $i = 1, \ldots, q$, and $\partial = \cup_{i=1}^{q} \partial_i$ there exist two closed sets ∂^+ and ∂^- such that

e4.1.7
$$\partial = \partial^+ \cup \partial^-, \qquad S\partial^- \subset \partial^-, \qquad S^{-1}\partial^+ \subset \partial^+, \tag{4.1.7}$$

i.e. the boundary ∂ can be decomposed into two parts, the first of which (denoted ∂^-) "contracts" under the action of S while the other (denoted ∂^+) "expands".

Remarks: (1) The continuity of $\underline{\sigma} \to X(\underline{\sigma})$ insures that X is a Borel map (*i.e.* the inverse images of the Borel sets are Borel sets).
(2) One has $X(\tau\underline{\sigma}) = SX(\underline{\sigma})$ so that S is coded into the symbolic dynamics.
(3) If $x \in \Omega \setminus \cup_{i\in\mathbb{Z}} S^{-i}\partial$ the $(\mathcal{Q}, S)$–history of x is naturally and unambiguously defined and the correspondence between $\Omega \setminus \cup_i S^{-i}\partial$ and $X^{-1}(\Omega \setminus \cup_i S^{-i}\partial)$ is one-to-one and maps, together with its inverse, Borel sets into Borel sets (by Kuratowsky's theorem, cf. footnote 1, Section §2.3).
(4) Although $\mathcal{Q}$ is not a partition of Ω it is convenient to adopt conventions and notations similar to those used for partitions. We shall denote

e4.1.8
$$Q^J_{\underline{\sigma}} = \mathop{\cap}_{j\in J} S^{-j} Q_{\sigma_i}, \qquad J \subset \mathbb{Z}, \quad \underline{\sigma} \in \{1, \ldots, q\}^J, \tag{4.1.8}$$

and a cylinder $C^J_{\underline{\sigma}}$ will be called T–compatible if $C^J_{\underline{\sigma}} \cap \{1, \ldots, q\}^{\mathbb{Z}}_T \neq \emptyset$. Note, however, that while the T–compatibility of $C^J_{\underline{\sigma}}$ implies $Q^J_{\underline{\sigma}} \neq \emptyset$ the *vice versa* in general is not true, because a point might belong to the boundary of several Q's.
(5) In analogy to what happens in the case of partitions, sometimes it can be convenient to consider also *non-generating Markovian pavements*: they are defined exactly as in definition (4.1.3), by eliminating only the condition that the set (4.1.3) consists of a single point. Then, if we want to stress the difference with respect to the ones defined in definition (4.1.3), we can refer to the latter as *generating Markovian pavements* (see problems [4.3.6] and [4.3.7] for some examples).
(6) The set $\Omega \setminus \cup_i S^{-i}\partial = \cap_i(\Omega \setminus S^{-i}\partial)$ is an intersection of a countable
N4.1.1 family of dense open sets, therefore it is not empty and, in fact, dense.[1] Therefore T "cannot have too many zeroes"; see problem [4.1.3].
(7) It is easy to realize that every point of Ω is the image of some sequence $\underline{\sigma} \in \{1, \ldots, q\}^{\mathbb{Z}}_T$. If $x \in \Omega \setminus \cup_i S^{-i}\partial$ this is obvious. If $x \in \cup_i S^{-i}\partial$ there exists σ such that $x \in Q_\sigma$; then we set $\sigma_0 = \sigma$ and there must exist σ' such that int $(Q_{\sigma_0}) \cap$ int$(S^{-1}Q_{\sigma'}) \neq \emptyset$, and $Sx \in Q_{\sigma'}$ (because $Q_\sigma = \overline{\text{int}\,(Q_\sigma)}$ and this property is also true for $S^{\pm 1}Q_\sigma$, since S is a homeomorphism). We shall set then $\sigma_1 = \sigma'$ and, by construction, $T_{\sigma_0\sigma_1} = 1$, *etc.*: in this way one constructs a sequence $\underline{\sigma} \in \{1, \ldots, q\}^{\mathbb{Z}}_T$ such that $x \in S^{-k}Q_{\sigma_k}$ for all $k \in \mathbb{Z}$. Therefore $x = X(\underline{\sigma})$.
(8) If $\mu \in \mathcal{M}^0(\Omega)$ is a Borel probability measure on Ω such that $\mu(S^k\partial) = 0$, for all $k \in \mathbb{Z}$, it is clear that the partition of $\Omega \setminus \cup_i S^{-i}\partial$ generated by $\mathcal{Q}$ is S–separating and in fact S–expansive, cf. (4.1.7), on $\mathcal{Q} \cap (\Omega \setminus \cup_i S^{-i}\partial)$. Hence $\mathcal{Q}$ restricted to $\Omega \setminus \cup_i S^{-i}\partial$ is a generating partition for every topological measure which is invariant and such that $\mu(\partial) = 0$.

[1] This is a consequence of a general Baire's theorem, see [DS58], p. 20.

The following proposition is, essentially, a tautology because definition (4.1.3) originated precisely by an effort to collect hypotheses sufficient to make true the properties that it states.

*P*4.1.1 **(4.1.1) Proposition:** (Codings of topological dynamical systems into symbolic ones via a Markovian pavement)
Let (Ω, S) be an invertible topological dynamical system which admits a Markovian pavement $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ with a compatibility matrix T.
(i) If the distribution m is in $M_0(\{1,\ldots,q\}^{\mathbb{Z}}_T)$ (i.e. $m \in M_0(\{1,\ldots,q\}^{\mathbb{Z}})$ and $m(\{1,\ldots,q\}^{\mathbb{Z}}_T) = 1$) the relation

*e*4.1.9
$$Am(E) = m(X^{-1}E) \qquad \textit{for all } E \in \mathcal{B}(\Omega) \tag{4.1.9}$$

defines a probability measure $Am \in \mathcal{M}^0(\Omega)$. The map A transforms τ–invariant measures in S–invariant measures, τ–ergodic measures in S–ergodic measures, etc. Since Ω is compact A is continuous.
(ii) If $m \in M_e(\{1,\ldots,q\}^{\mathbb{Z}}_T)$ is a topological measure, then $Am \in \mathcal{M}_e(\Omega, S)$ and Am is a topological measure.
(iii) The correspondence A between ergodic topological measures on the space of compatible sequences $\{1,\ldots,q\}^{\mathbb{Z}}_T$ and on Ω is, imagining the measures to
*N*4.1.2 *be completed,*[2] *a correspondence between measures isomorphic mod 0. The dynamical systems $(\{1,\ldots,q\}^{\mathbb{Z}}_T, \tau, m)$ and (Ω, S, Am) are, for such measures m, isomorphic mod 0.*

Remarks: (1) Because of property (iii) it is possible to "reduce" the analysis of the S–invariant ergodic topological measures on Ω to the analysis of the analogous τ–ergodic topological measures on $\{1,\ldots,q\}^{\mathbb{Z}}_T$. Since the latter, as we shall see, can sometimes be studied in detail, this possibility is of great interest. For instance if T is mixing, *i.e.* there exists $N > 0$ such that $(T^N)_{\sigma\sigma'} > 0$ for all σ, σ', cf. definition (4.1.1), it is easy to see that there exist S–ergodic topological measures and this is, by itself, already a nontrivial fact: we shall come back to this point in more detail in the forthcoming sections.
(2) The validity of proposition (4.1.1) rests mainly on the remark that, if $\mu \in \mathcal{M}_e(\Omega, S)$ is a topological ergodic measure, then $\mu(\partial^-) = 0$. In fact
*N*4.1.3 $\partial^- \supset S\partial^-$ and, hence, $\mu(\partial^-)$ is 0 or 1 by ergodicity:[3] however $\Omega \setminus \partial^-$ is open and hence $\mu(\Omega \setminus \partial^-) > 0$ and $\mu(\partial^-) = 0$. Likewise the ergodicity of μ with respect to S^{-1} implies $\mu(\partial^+) = 0$, hence $\mu(\partial) = 0$; by invariance it follows then that $\mu(\cup_i S^{-i}\partial) = 0$.

Proof: Continuity of the code X implies that the inverse image of a Borel set $E \in \mathcal{B}(\Omega)$ is again a Borel set, so that Am is well defined and one has only to check that it is a measure: this is a general property because the

---

[2] *i.e.* we imagine extending the σ–algebra of the measurable sets by adding the sets which are not Borel sets but which are contained in zero measure Borel sets.

[3] Since $SE \subset E$ and $\mu(SE) = \mu(E)$ the sets E and SE differ mod 0, *i.e.* E is invariant mod 0 and therefore $\mu(E) = 0, 1$.

X^{-1}–images of the elements of a countable family of pairwise disjoint sets is a countable family of pairwise disjoint sets.

To prove (ii) note that if m is ergodic then also Am is ergodic. In fact if E is an invariant set for S then $X^{-1}E = X^{-1}SE = \tau X^{-1}E$, so if E has Am–measure different from 0 or 1 also $X^{-1}E$ has the same m–measure. By the remark (2) one has $Am(\cup_i S^i\partial) = 0$; this means that $m(X^{-1}(\cup_i S^i\partial)) = 0$ and, hence, X (which is a one-to-one correspondence between $\Omega\setminus\cup_i S^i\partial$ and $\{1,\ldots,q\}^{\mathbb{Z}}_T \setminus X^{-1}(\cup_i S^i\partial)$) is an isomorphism mod 0 between $A\,m$ and m.

Furthermore if G is open in Ω and $x \in G\setminus\cup_i S^{-i}\partial$ let $\underline{\sigma}(x)$ be a compatible sequence coded into x. There must exist N such that $\cap_{-N}^{N} S^{-i}Q_{\sigma_i(x)} \subset G$; indeed by (4.1.5) the diameter of this set is infinitesimal as $N \to \infty$. But

$$X^{-1}\big(\cap_{-N}^{N} S^{-i}Q_{\sigma_i(x)}\big) \supset C^{-N\ \ldots\ N}_{\sigma_{-N}(x)\ldots\sigma_N(x)},$$

and the measure of the latter cylinder is positive because it is T–compatible and m is a topological measure: hence $Am(G) > 0$.

To prove (iii) note that from the ergodicity of $\mu \in \mathcal{M}_e(\Omega, S)$ and from the preceding remark (2) it follows that $\mu(\cup_i S^{-i}\partial) = 0$; then the measure m on $\{1,\ldots,q\}^{\mathbb{Z}}_T$ defined by

e4.1.10 $$m(E) = \mu(X(E \setminus X^{-1}(\cup_i S^{-i}\partial)) \qquad \text{for all } E \in \mathcal{B}(\{1,\ldots,q\}^{\mathbb{Z}}_T) \tag{4.1.10}$$

is isomorphic mod 0 to μ and $Am = \mu$.

If the cylinder $C^{-N\ldots N}_{\sigma_{-N}\ldots\sigma_N}$ is T–compatible we have that

e4.1.11 $$\begin{aligned} &X(C^{-N\ldots N}_{\sigma_{-N}\ldots\sigma_N} \setminus X^{-1}(\cup_i S^{-i}\partial)) =\\ &= \cap_{-N}^{N} S^{-i}Q_{\sigma_i} \setminus \cup_i S^{-i}\partial = \cap_{-N}^{N} S^{-i}Q_{\sigma_i} \qquad \text{mod } 0, \end{aligned} \tag{4.1.11}$$

and, therefore, $m(C^{-N\ldots N}_{\sigma_N\ldots\sigma_N}) > 0$ because μ is a topological measure and, by (i), int $\big(\cap_{-N}^{N} S^{-i}Q_{\sigma_i}\big) \neq \emptyset$. Combining this with property (ii) property (iii) follows. ∎

Remarks: (1) In fact the above arguments also prove that A establishes a one-to-one correspondence between measures μ on Ω such that $\mu(\cup_i S^i\partial) = 0$ and measures m on $\{1,\ldots,q\}^{\mathbb{Z}}_T$ such that $m(X^{-1}(\cup_i S^{-i}\partial)) = 0$; furthermore corresponding measures are isomorphic mod 0.

(2) The preceding proposition would remain valid if the condition (4.1.5), in the definition of Markovian pavement, was modified into "$d(X(\underline{\sigma}), X(\underline{\sigma}')) \to 0$ if $d(\underline{\sigma}, \sigma') \to 0$". We chose to state the result in a less general form because, as we shall see, Hölder continuity is a very natural regularity property of codes X and it will play an essential role in various applications.

Problems for §**4.1** (*On Perron–Frobenius' theorem and on the structure of Markovian chains)*

Q4.1.1 **[4.1.1]:** Consider the dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}}, \tau)$ and show that the pavement $\mathcal{Q} = \{Q_0,\ldots,Q_n\}$ of $\{0,\ldots,n\}^{\mathbb{Z}}$ with $Q_\sigma = \{\underline{\sigma}' \,|\, \underline{\sigma}'_0 = \sigma\}$ is Markovian. (*Hint:* In this case ∂ is empty.)

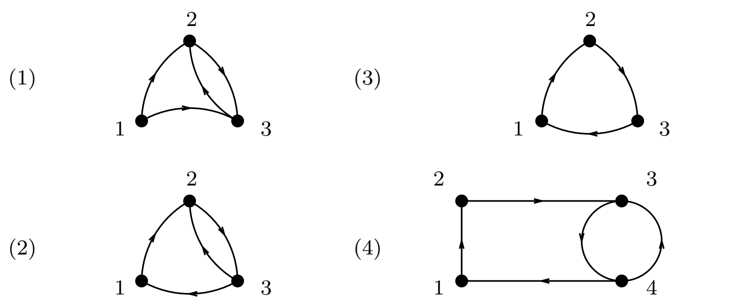
Q4.1.2 **[4.1.2]:** Find examples of dynamical systems with Markovian pavements and matrices of compatibility with some zero entry. (*Hint:* The space of sequences of 0, 1's with $T = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$.)

Q4.1.3 **[4.1.3]:** Check that the matrix $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ cannot be a compatibility matrix for a Markovian pavement. (*Hint:* No point would have a compatible symbolic history).

Q4.1.4 **[4.1.4]:** (*Compatibility graphs*)
Let $\sigma, \sigma' = 0, 1, \ldots, n$ and let $T_{\sigma\sigma'} \ge 0$ be a matrix. Let G_T be the graph obtained by connecting all pairs σ, σ' verifying $T_{\sigma,\sigma'} > 0$ by an arrow pointing from σ to σ': we say that σ' follows σ. A symbol σ can follow itself (*i.e.* $T_{\sigma\sigma} > 0$) or it can follow and be followed by a symbol σ' (*i.e.* $T_{\sigma\sigma'} > 0$ and $T_{\sigma'\sigma} > 0$). Two labels σ, σ' will be called *equivalent* if there is a closed loop of coherently oriented arrows in G_T that start at σ and return, proceeding always in the direction of the arrows, to σ after passing through σ'. Show that the set of labels can be divided into the set $I_0 = \mathcal{I}$ of labels that are inequivalent to any other label, that we call *inessential labels*, and into sets $I_1, \ldots, I_a$ (called "classes") such that each I_j contains labels equivalent to any of the other labels in the same I_j but inequivalent to any label in I_k if $k \ne j$. (*Hint:* Try to draw some special cases first, like the ones corresponding to the matrices

$$T_1 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\; T_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},\; T_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},\; T_4 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix},$$

and illustrated in Fig. (4.1.1).)



F4.1.1 **Fig.(4.1.1)** The graphs G_T corresponding to the three matrices $T = T_1, T_2, T_3, T_4$ of problem [4.1.4].

Q4.1.5 **[4.1.5]:** In the context of problem [4.1.4] call a semi-infinite sequence $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}_+}$ *compatible* if $T_{\sigma_i \sigma_{i+1}} > 0$ for all $i \ge 0$. Likewise call an infinite sequence $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ *compatible* if $T_{\sigma_i \sigma_{i+1}} > 0$ for all $i \in \mathbb{Z}$. The spaces of compatible sequences will be denoted $\{0, \ldots, n\}_T^{\mathbb{Z}_+}$ or $\{0, \ldots, n\}_T^{\mathbb{Z}}$: they are *subshifts of finite type* in the sense introduced in this section (see definition (4.1.1). Show that any semi-infinite or infinite compatible sequence contains at most a finite number of inessential labels. Show also that a semi-infinite compatible sequence consists eventually only of labels $\sigma_i \in I_{k_+}$ for some k_+. Likewise an infinite compatible sequence consists, to the right of 0, eventually of labels in some I_{k_+} and, to the left of 0, eventually of labels in some I_{k_-}, for some k_+, k_-. However if $k_+ = k_- = k$ then $\sigma_i \in I_k$ for all i. Furthermore no infinite compatible sequence with $k_- = k_+$ contains inessential labels.

Q4.1.6 **[4.1.6]:** In the context of problems [4.1.4], [4.1.5] let $\{0, \ldots, n\}_T^{\mathbb{Z}}$ be the space of the compatible sequences. Show that if E is a translation invariant Borel set in $\{0, \ldots, n\}_T^{\mathbb{Z}}$ and μ is an ergodic measure such that $\mu(E) = 1$ then μ–almost all elements $\underline{\sigma} \in E$ have symbols σ_i in a *single* I_j. (*Hint:* By problem [4.1.5] all compatible sequences $\underline{\sigma}$ are such that the frequency $\ell_j^+(\underline{\sigma})$ of appearance of symbols in I_j in the "future part $\sigma_0, \sigma_1, \ldots$ of $\underline{\sigma}$" is well defined and it is either 0 or 1. By ergodicity almost all sequences must have the same value of $\ell_j^+(\underline{\sigma})$, so that there will be a single j_0 such that $\ell_{j_0}^+(\underline{\sigma}) = 1$ while for all $j \neq j_0$ one will have $\ell_j^+(\underline{\sigma}) = 0$. But the frequency in the future is equal to the frequency in the past (*i.e.* in the "past part $\ldots, \sigma_{-1}, \sigma_0$ of $\underline{\sigma}$", see problem [2.2.46]), therefore one has $\ell_j^-(\underline{\sigma}) \equiv \ell_j^+(\underline{\sigma})$ μ–almost everywhere: hence the "typical" sequences will have symbols that eventually in the past and the future all lie in I_{j_0}: in the language of problem [4.1.5] one has $k_- = k_+ = j_0$, hence $\sigma_i \in I_{j_0}$ for all i.)

Q4.1.7 **[4.1.7]:** (*Classes and periods of a compatibility matrix*)
If $\sigma \in I_j$ set $d(\sigma)$ = greatest common divisor of the integers $n_\sigma > 0$ such that $(T^{n_\sigma})_{\sigma\sigma} > 0$. Show that $d(\sigma)$ is constant for $\sigma \in I_j$. This allows us to define d_j for the class I_j as $d_j = \{d(\sigma) : \sigma \in I_j\}$; we shall say, for reasons that will be clarified later, that d_j is the *period* of the class I_j. (*Hint:* If s, m, n are such that $T^s_{\sigma\sigma} > 0$, $T^m_{\sigma\sigma'} > 0$, $T^n_{\sigma'\sigma} > 0$, one has $T^{n+m+ks}_{\sigma'\sigma'} \geq T^m_{\sigma'\sigma} T^{ks}_{\sigma\sigma} T^n_{\sigma\sigma'} > 0$, so that $d(\sigma')$ divides $n + m + ks$ for all k, hence it divides s. This implies $d(\sigma') \leq d(\sigma)$; by changing the roles of σ and σ' one finds $d(\sigma) = d(\sigma')$.)

Q4.1.8 **[4.1.8]:** (*Transitivity and mixing of compatibility matrices*)
In the context of the previous problems suppose that the matrix T is such that all labels σ are equivalent (*i.e.* there are no inessential labels and all the labels form a single class I_1). Show that if the period is $d = 1$ then there is p such that $T^p_{\sigma\sigma'} > 0$ for all σ, σ'. If the period is $d \geq 2$ then the set I_1 can be divided into d disjoint subsets $I_{1,1}, \ldots, I_{1,d}$, with $I_{1,d+1} \equiv I_{1,1}$, such that $T_{\sigma\sigma'} > 0$ only if $\sigma \in I_{1,i}$ and $\sigma' \in I_{1,i+1}$ for some i. Furthermore if $p > 0$ is large enough $T^{pd}_{\sigma\sigma'} > 0$ for all $\sigma, \sigma' \in I_{1,i}$ for some i and 0 otherwise, *i.e.* the block of the matrix T^d corresponding to the labels $\sigma, \sigma' \in I_{1,i}$ is *mixing*. (*Hint:* Let $n_{\sigma\sigma'} > 0$ be such that $T^{n_{\sigma\sigma'}}_{\sigma\sigma'} > 0$. Suppose first that there is an element σ_0 such that $T_{\sigma_0\sigma_0} > 0$, hence $n_{\sigma_0\sigma_0} = 1$ and $d_1 = 1$. Then we take $p = \sum_{\sigma\sigma'} n_{\sigma\sigma'}$, $\overline{n} = p - n_{\sigma\sigma_0} - n_{\sigma_0\sigma'}$ and we see that $T^p_{\sigma\sigma'} \geq T^{n_{\sigma\sigma_0}}_{\sigma\sigma_0} T^{\overline{n}}_{\sigma_0\sigma_0} T^{n_{\sigma_0\sigma'}}_{\sigma_0\sigma'} > 0$. Consider next the case in which $d = 1$ and $\sigma = 0, 1$, and let n_{00}, n_{11} be relatively prime integers such that $T^{n_{00}}_{00} > 0, T^{n_{11}}_{11} > 0$. Then any integer k large enough can be *simultaneously* written in the following forms

$$\begin{aligned}
k &= m n_{00} + n_{01} + m' n_{11}, \\
k &= \widetilde{m}' n_{11} + n_{10} + \widetilde{m} n_{00}, \\
k &= \widehat{m} n_{00} + n_{01} + \widehat{m}' n_{11} + n_{10}, \\
k &= \widehat{m}' n_{11} + n_{10} + \widehat{m} n_{00} + n_{01},
\end{aligned}$$

for suitably chosen integers $m, m', \widetilde{m}, \ldots$ because if k is large also $k - \sum n_{\sigma\sigma'}$ is large. Therefore we can write $T^k = (T^{n_{00}})^m T^{n_{01}} (T^{n_{11}})^{m'}$, using the first expression and, in this way, we realize that $T^k_{01} > 0$, *etc.* This clearly implies that $T^k_{\sigma\sigma'} > 0$ for all pairs σ, σ' and for all k large enough. Analogously one discusses the case $d = 1$ with $n > 1$. The general case, $d \geq 2$ and $n \geq 1$, is similar.)

Q4.1.9 **[4.1.9]:** Particularize the results of problem [4.1.7] to the case of the matrices of problem [4.1.4].

Q4.1.10 **[4.1.10]:** (*Iterates of a compatibility matrix*)
Check that the result of the previous problems means that the space $\widetilde{\Omega} = \{0, \ldots, n\}^{\mathbb{Z}}_T$ of compatible infinite sequences can be divided into d disjoint spaces $\Omega_1, \Omega_2, \ldots, \Omega_d$, and the shift maps Ω_j onto Ω_{j+1}, with $\Omega_{d+1} \equiv \Omega_1$. (*Hint:* Note that, by the previous problems, d will turn out to be the period of the class I_1.)

Q4.1.11 **[4.1.11]:** (*Spectral decomposition of subshifts*)
Under the hypotheses of the previous problems and assuming that all labels are essential and belong to the same class I_1 show that the space $\widetilde{\Omega} = \{0, \ldots, n\}^{\mathbb{Z}}_T$ can be split as a union of d disjoint closed sets $\Omega_1, \ldots, \Omega_d$, and the shift τ maps Ω_i into Ω_{i+1}, with

$\Omega_{d+1} \equiv \Omega_1$, so that τ^d maps Ω_i into itself for each i, and given any pair F, G of relatively open sets in Ω_i there is a large enough $q > p$ such that $\tau^{q'\,d} F \cap G \neq \emptyset$ for all $q' > q$. A matrix T with only essential labels and only one class of them is called *transitive matrix*. (*Hint:* One has to note, see problem [4.1.5], that a sequence is in Ω_i if and only if $\sigma_0 \in I_{1,i}$; furthermore any open set can be obtained as a union of cylinders with a long enough *finite* base. The fact that $T^{dp}_{\sigma\sigma'} > 0$ if $\sigma, \sigma' \in I_{1,i}$ means that fixed i we can obtain compatible sequences which have at sites multiple of dp arbitrary labels in Ω_i.)

Q4.1.12 **[4.1.12]:** (*Perron-Frobenius theorem for transitive matrices*)
Under the hypotheses of problems [4.1.11] show that there exist d eigenvectors of the matrix T^d, to be denoted $e^{(0)}$, $e^{(1)}, \ldots, e^{(d-1)}$, with zero components except those in correspondence of the labels of $I_{1,1}, \ldots, I_{1,d}$, respectively, relative to the matrix T^d. The components of $e^{(i)}$ with labels in $I_{1,i}$ are strictly positive. (*Hint:* Use that T^d is a block matrix which acts in a mixing way in every block, cf. problem [4.1.8], and apply Perron–Frobenius' theorem in its elementary form discussed in problems [2.3.7]÷[2.3.12]).

Q4.1.13 **[4.1.13]:** In the context of problem [4.1.12] show that $Te^{(i)} = \lambda e^{(i+1)}$, for all $i = 0, 1, \ldots, d-1$, if we set $e^{(d)} = e^{(0)}$, and if the eigenvectors are suitably rescaled and $\lambda > 0$ is suitably chosen. (*Hint:* Note that T transforms a vector with nonzero components on the group of labels $I_{1,i}$ into one with components nonzero on the successive group $I_{1,i+1}$, *etc.*)

Q4.1.14 **[4.1.14]:** Deduce from problem [4.1.13] that the eigenvalue of T^d relative to $e^{(i)}$ is $\lambda^d > 0$, independently from i.

Q4.1.15 **[4.1.15]:** Always in the context of problem [4.1.12], let e be an eigenvector for T with eigenvalue $\mu e^{i\varphi}$, $\mu > 0$. Setting $\widehat{e}^{(i)}_j = 0$ if $j \notin I_{1,i}$, and $\widehat{e}^{(i)}_j = e_j$ if $j \in I_{1,i}$, one has $T^d \widehat{e}^{(i)} = (\mu e^{i\varphi})^d \widehat{e}^{(i)}$. If the eigenvalue $\mu e^{i\varphi}$ is, among those of T, one with the largest absolute value, show that $\widehat{e}^{(i)}$ is proportional to $e^{(i)}$ and, furthermore, $(e^{i\varphi})^d = 1$, $\mu = \lambda$.

Q4.1.16 **[4.1.16]:** By making use of the result of problem [4.1.15] show that if the eigenvalue corresponding to e is one with largest absolute value, then it has the form $e = \sum_{j=0}^{d-1} e^{-\frac{2\pi i}{d}\,pj} e^{(j)}$ and this corresponds to the eigenvalue $\lambda e^{\frac{2\pi i}{d}\,p}$.

Q4.1.17 **[4.1.17]:** (*Perron–Frobenius theorem for general matrices*)
Deduce from problems [4.1.10]÷[4.1.16] that if T is a matrix with non-negative entries the eigenvalues of largest absolute value of T are arranged proportionally to the d–th roots of unity on a circle of radius $\lambda \geq 0$. The number d varies, if $\lambda > 0$, in a subset of the set of the periods of the blocks of equivalent labels. In fact, more generally, to each of these blocks I of period d_I correspond d_I eigenvectors of the type $\lambda_I e^{2\pi i p/d_I}$, $p = 0, \ldots, d_I - 1$, with simple multiplicity: between them one finds those of largest absolute value (that are precisely those that maximize λ_I).

Q4.1.18 **[4.1.18]:** (*Errant and non wandering points*)
If (Ω, S) is a topological dynamical system we say that a point $x \in \Omega$ is *errant* or *wandering* if one of its neighborhoods U is such that eventually no point initially in U evolves into point again in U, *i.e.* if there exists an open $U \ni x$ and an integer N_U such that $S^n U \cap U = \emptyset$, for all $n \geq N_U$. Interpret the results of the previous problems [4.1.4] and [4.1.5] as the statement that the set of the nonwandering points of $\{1, \ldots, q\}^{\mathbb{Z}}_T$ is $(I_1)^{\mathbb{Z}}_T \cup \ldots \cup (I_a)^{\mathbb{Z}}_T$, where the subscript T means that one considers only compatible sequences.

Q4.1.19 **[4.1.19]:** (*Topological transitivity and mixing*)
If (Ω, S) is a topological dynamical system with Ω compact we say that (Ω, S) is *topologically transitive* if there exists $x \in \Omega$ such that the set $\cup_{n\geq 0}\{S^n x\}$ is dense in Ω; we say that (Ω, S) is *topologically mixing* if given F, G open there exists N^0 such that $F \cap S^N G \neq \emptyset \;\forall N \geq N^0$. Show that if (Ω, S) is topologically transitive or mixing and if it admits a Markovian pavement with compatibility matrix T, then T is transitive or,

N4.1.4 respectively, mixing.[4]

Q4.1.20 **[4.1.20]:** (*Smale spectral theorem*)
Let (Ω, S) be a topological invertible dynamical system endowed with a Markovian pavement $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$. Show that the set $\Omega_{\rm nw}$ of the nonwandering points of Ω, apart from a set of zero measure for all the ergodic topological measures on Ω, is representable as

$$\Omega_{\rm nw} = \bigcup_{i=1}^{a} \bigcup_{j=1}^{d_i} \Omega_{i,j},$$

where $\Omega_{i,j}$ are closed sets such that $S\Omega_{i,j} = \Omega_{i,j+1}$, with $\Omega_{i,d_i+1} = \Omega_{i,1}$, and S^{d_i} is topologically mixing on $\Omega_{i,j}$. (*Hint:* Use the results of problem [4.1.19] and set $\Omega_{i,j} = X(\widetilde{\Omega}_{i,j})$; see problem [4.1.19] for the notion of topological mixing).

Q4.1.21 **[4.1.21]:** Find the decomposition into equivalence classes of communicating labels of the matrices that follow, compute the periods and determine the subclasses $\Omega_{i,j}$ for every indecomposable block:

$$T_3 = \begin{pmatrix} 0&1&1&0&0\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&0&0&1\\ 1&0&0&0&0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1&0&0&0&1\\ 0&0&1&0&0\\ 0&0&0&1&0\\ 0&1&0&0&0\\ 0&0&0&0&1 \end{pmatrix}, \quad T_1 = \begin{pmatrix} 1&1\\ 0&1 \end{pmatrix}.$$

Bibliographical note to §4.1

The abstract notion of Markovian pavement, given here, is inspired by the some of its concrete applications. Other directions in which one can think for an abstract interpretation of the results that lead to the notion of Markovian pavement are possible. See for instance, [Ca76].

The idea of using Markovian pavements to study certain classes of topological measures goes back to the work of Adler and Weiss, [AW68] and to the works of Sinai, [Si68a], [Si68b], with important extensions in [Si72], who proposed a very original method to treat the existence problem and the analysis of the ergodic properties of invariant topological measures associated with a hyperbolic system.

§4.2 Markovian pavements for hyperbolic systems

An interesting class of dynamical systems for which it is possible to construct Markovian pavements $\mathcal{Q}$ consisting of sets of arbitrarily small diameter is the class of the *smooth hyperbolic systems*, also called *Anosov systems*.

The prototype of such systems is among those discussed in Section §1.2: it is the dynamical system (Ω, S), where $\Omega = \mathbb{T}^2$ = bidimensional torus regarded as a Riemannian manifold with the flat metric $ds^2 = d\varphi_1^2 + d\varphi_2^2$ and

e4.2.1
$$S \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \mod 2\pi. \tag{4.2.1}$$

[4] See definition (4.1.1).

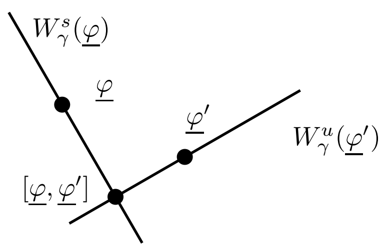
This system is hyperbolic in the sense that for every point $\underline{\varphi} \in \Omega$ there exist two manifolds $W^u(\underline{\varphi})$ and $W^s(\underline{\varphi})$ which enjoy the properties that we now list.
$W^s(\underline{\varphi})$ is the straight line directed as the eigenvector $\underline{v}_1$ relative to the eigenvalue $\lambda < 1$ of the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ while $W^u(\underline{\varphi})$ is the straight line directed as the eigenvector $\underline{v}_2$ relative to the eigenvalue $\lambda^{-1} > 1$ of the same matrix:

e4.2.2
$$\underline{v}_1 = \begin{pmatrix} 1 \\ \lambda - 1 \end{pmatrix}, \quad \underline{v}_2 = \begin{pmatrix} 1 \\ \lambda^{-1} - 1 \end{pmatrix}, \quad \lambda = (3 - \sqrt{5})/2, \tag{4.2.2}$$

and, since their slope is irrational, the two straight lines fill densely Ω.
Furthermore the lines $W^a(\underline{\varphi})$ are *covariant*, *i.e.* $SW^a(\underline{\varphi}) = W^a(S\underline{\varphi})$ for $a = u, s$, and any two points $\underline{\psi}$ and $\underline{\psi}'$ on $W^s(\underline{\varphi})$ become close at exponential rate under the action of S, while any two points on $W^u(\underline{\varphi})$ become separated with exponential rate, *i.e.* , for all $n > 0$,

e4.2.3
$$\begin{aligned} d(S^n\underline{\psi}, S^n\underline{\psi}') &\le \lambda^n d(\underline{\psi}, \underline{\psi}') \qquad \text{for all } \underline{\psi}, \underline{\psi}' \in W^s(\underline{\varphi}), \\ d(S^{-n}\underline{\psi}, S^{-n}\underline{\psi}') &\le \lambda^n d(\underline{\psi}, \underline{\psi}') \qquad \text{for all } \underline{\psi}, \underline{\psi}' \in W^u(\underline{\varphi}), \end{aligned} \tag{4.2.3}$$

where d is the distance measured along $W^s(\underline{\varphi})$ or $W^u(\underline{\varphi})$ respectively (and it coincides with the geodesic distance if $\underline{\psi}$, $\underline{\psi}'$ are close enough).
From Fig. (4.2.1) it is also clear that, if $W^u_\gamma(\underline{\varphi})$ and $W^s_\gamma(\underline{\varphi})$ denote the connected parts of $W^u(\underline{\varphi})$ and $W^s(\underline{\varphi})$ containing $\underline{\varphi}$ and contained in a circle
N4.2.1 of small enough radius γ,[1] then there exists $\varepsilon > 0$ such that if $d(\underline{\varphi}, \underline{\varphi}') < \varepsilon$
N4.2.2 it follows that $[\underline{\varphi}', \underline{\varphi}] \stackrel{def}{=} W^u_\gamma(\underline{\varphi}') \cap W^s_\gamma(\underline{\varphi})$ consists of a unique point.[2]
Furthermore if ε is small enough $d(\underline{\varphi}, \underline{\varphi}') < \varepsilon$ implies that $[\underline{\varphi}', \underline{\varphi}]$ depends continuously on $\underline{\varphi}, \underline{\varphi}'$.



F4.2.1 **Fig.(4.2.1)** Representation of the operation that associates $[\underline{\varphi}', \underline{\varphi}]$ with the two points $\underline{\varphi}$ and $\underline{\varphi}'$ as the intersection of a short connected part $W^u_\gamma(\underline{\varphi}')$ of the unstable manifold of $\underline{\varphi}'$ and of a short connected part $W^s_\gamma(\underline{\varphi})$ of the stable manifold of $\underline{\varphi}$. The size γ is short “enough”, compared to the diameter of Ω, and it is represented by the segments to the right and left of $\underline{\varphi}$ and $\underline{\varphi}'$.

---

[1] The circle is defined in terms of the geodesic distance.

[2] Why is it necessary to require smallness of γ to have, say, uniqueness here?

The manifolds (straight lines, in this case) $W^u(\underline{\varphi})$ and $W^s(\underline{\varphi})$ are called, respectively, the *unstable manifold* and the *stable manifold* of the point $\underline{\varphi}$. Such manifolds are covariant with respect to the action of W, *i.e.*

e4.2.4
$$SW^s(\underline{\varphi}) = W^s(S\underline{\varphi}), \qquad S^{-1}W^u(\underline{\varphi}) = W^u(S^{-1}\underline{\varphi}). \tag{4.2.4}$$

It is natural to call "hyperbolic" the map S: in every point the action of S is strongly unstable analogously to what happens near the unstable equilibrium points called in stability theory "hyperbolic".

The example just discussed is very special and simple because of the local linearity of the map S. However the above described situation is sufficiently simple to admit natural generalizations to the case of more complex maps.

D4.2.1 **(4.2.1) Definition:** (Smooth hyperbolic system or Anosov system)
Let (Ω, S) be a dynamical system on a compact connected Riemannian manifold Ω of class C^∞ with S a diffeomorphism of class C^∞.[3] *Suppose that the system is topologically transitive, i.e. there is a dense orbit.*[4] *Suppose furthermore that there exists a smooth Riemannian metric d (possibly different from the one given on Ω, yet equivalent to it) such that measuring lengths with d the following properties hold.*

N4.2.3 N4.2.4

(i) (Splitting property) There exist two manifolds $\widetilde{W}^u(x)$ and $\widetilde{W}^s(x)$, that we shall suppose of class C^k, with $k > 2$, and with tangent plane at x depending on x with Hölder regularity. Furthermore the manifolds $\widetilde{W}^u(x)$ and $\widetilde{W}^s(x)$ are transversal in x and have complementary positive dimensions.[5]

N4.2.5

(ii) (Covariance property) Calling $\Sigma_\gamma(x)$ the sphere of radius γ centered at x and setting $W^u_\gamma(x) = \{$connected part of $\widetilde{W}^u(x) \cap \Sigma_\gamma(x)$ containing $x\}$ and $W^s_\gamma(x) = \{$connected part of $\widetilde{W}^s(x) \cap \Sigma_\gamma(x)$ containing $x\}$, there exists $\gamma > 0$ such that

e4.2.5
$$SW^s_\gamma(x) \subset W^s_\gamma(Sx), \qquad S^{-1}W^u_\gamma(x) \subset W^u_\gamma(S^{-1}x). \tag{4.2.5}$$

(iii) (Hyperbolicity property) There exists $\lambda < 1$ such that, for all $n \geq 0$,

e4.2.6
$$\begin{aligned} d(S^n y, S^n z) &\leq \lambda^n d(y,z) \qquad \text{for all } y,z \in W^s_\gamma(x),\\ d(S^{-n} y, S^{-n} z) &\leq \lambda^n d(y,z) \qquad \text{for all } y,z \in W^u_\gamma(x). \end{aligned} \tag{4.2.6}$$

(iv) There exists $\varepsilon > 0$, $\varepsilon < \gamma$ such that, if $x, y \in \Omega$ and $d(x,y) < \varepsilon$, the set $W^u_\gamma(x) \cap W^s_\gamma(y)$ consists in just a single point $[x,y]$ which depends continuously on x and y.
In the above circumstances we say that (Ω, S) is a smooth hyperbolic system *or an* Anosov system *and that the Riemannian metric d is* adapted *to the map S.*

---

[3] One often requires just class C^2.

[4] Or, equivalently, for all nonempty open $U, V \subset \Omega$ one has $U \cap S^n V \neq \emptyset$ for some n.

[5] The Hölder continuous dependence on x means there is $\delta > 0$ such that in each chart of an atlas for Ω we can find a base of vectors in the plane V^α_x tangent to $\widetilde{W}^a(x)$, $a = u, s$, whose *corresponding* components ξ_x verify $|\xi_x - \xi_y| < C d(x,y)^\alpha$ if $d(x,y) < \delta$.

The latter definition is formulated so that it will turn out to be useful for the applications but it contains *several elements of redundancy* and it involves conditions and assumptions that appear rather difficult to verify: the following proposition could be used to replace definition (4.2.1) with a "purer" one. One can show that the following holds.

P4.2.1 **(4.2.1) Proposition:** (Anosov)
Let (Ω, S) be a dynamical system on a compact connected C^∞ Riemannian manifold Ω and let S be a C^∞ diffeomorphism. Suppose the following property for the linearization[6] *of S^n hold at the every point $x \in \Omega$.* (N4.2.6)
(i) (Splitting) It is possible to decompose (or "split") the tangent space V_x at $x \in \Omega$ into two complementary (non-zero) spaces:

e4.2.7
$$V_x = V_x^s \oplus V_x^u, \tag{4.2.7}$$

*such that $dS^n V_x^s = V^s_{S^n x}$ and $dS^{-n} V_x^u = V^u_{S^{-n} x}$ (*covariance*), and V_x^s and V_x^u depend with continuity on x (*continuity*).*
(ii) (Hyperbolicity) There exist $C > 0$ and $\lambda < 1$ for which, for all $n \geq 0$,

e4.2.8
$$\begin{aligned} \|(dS^n)v\|_{V_{S^n x}} &\leq C\lambda^n \|v\|_{V_x} \qquad \textit{for all } v \in V_x^s, \\ \|(dS^{-n})v\|_{V_{S^{-n} x}} &\leq C\lambda^n \|v\|_{V_x} \qquad \textit{for all } v \in V_x^u, \end{aligned} \tag{4.2.8}$$

where the subscripts to the modulus signs indicate that the lengths are measured in the metric used for the tangent vectors at the appropriate points.
(iii) (Transitivity) There is a point with a dense orbit.
Then (Ω, S) is a smooth hyperbolic system. The decomposition of the tangent space into the two spaces V_x^s, V_x^u is called the hyperbolic splitting.

Remarks: (1) *Vice versa*, if (Ω, S) is hyperbolic in the sense of the definition (4.2.1) then it verifies the hypothesis of proposition (4.2.1) and therefore proposition (4.2.1) can be considered as a differential definition of hyperbolicity. Sometimes one defines a hyperbolic dynamical system as a system that verifies the hypothesis of proposition (4.2.1), possibly replacing the regularity requirement in class C^∞ by that of regularity in class C^2.
(2) Definition (4.2.1) brings directly into light the properties which are useful for proving existence of Markovian pavements and it is better suited for further generalizations.
(3) Therefore one can take as definition either the statements of definition (4.2.1) or of proposition (4.2.1). *However* an even weaker and more satisfactory definition, implying all the above statements, can be given. It is discussed in problem [4.2.1] below.
(4) One could define Anosov systems without requiring the existence of a dense trajectory. It is not known whether a system verifying the properties (i) and (ii) of definition (4.2.1) is necessarily transitive.

[6] Usually denoted dS^n and thought of as an operator between the tangent space V_x at x and the tangent space $V_{S^n x}$ at $S^n x$.

(5) A proof of a stronger result is presented in the problems at the end of the section.

P4.2.2 **(4.2.2) Proposition:** (Anosov)
Given any $0 < \alpha < 1$ and under the assumptions of proposition (4.2.1) the fields of spaces V^s_x, V^u_x are Hölder continuous in x with exponent α.

Remarks: (1) See footnote 5 for a definition of Hölder continuity of a field of spaces.
(2) This theorem is particularly important because in a sense it is optimal: there exist analytically smooth Anosov maps whose stable and unstable spaces split the tangent plane in a nonsmooth way (still Hölder continuous, of course). For the proof see problem [4.2.8].

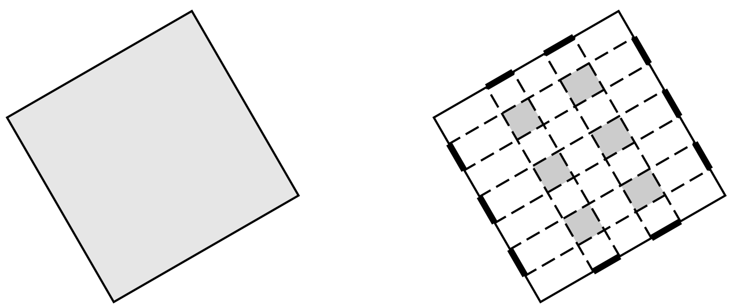
The construction of Markovian pavements for hyperbolic systems is based on the notion of *S–rectangle* or *hyperbolic rectangle.*

D4.2.2 **(4.2.2) Definition:** (S–rectangle)
Using the notations of definition (4.2.1) a set $R \subset \Omega$ is an S–rectangle for the hyperbolic system S, cf. *definition (4.2.1), if*

$$(a)\ R = \overline{\mathrm{int}(R)}, \qquad (b)\ \mathrm{diam}(R) < \varepsilon, \qquad (c)\ x, y \in R \Rightarrow [x, y] \in R,$$

with ε small enough, so that property (iv) in definition (4.2.1) (where the operation $[\cdot,\cdot]$ is defined) holds.

In the example at the beginning of this section simple S–rectangles are (small) parallelograms with sides parallel to the eigenvectors $\underline{v}_1$ and $\underline{v}_2$ of the matrix $\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}$.



F4.2.2 **Fig.(4.2.2)** Two examples of S–rectangles, shaded in the figure, for the map of the torus $\mathbb{T}^2$ generated by the matrix $\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}$. In the second case the rectangle is not a connected set and is obtained by letting x run on the disconnected intervals marked on one of the sides of the previous rectangle and y run over a neighbouring side and forming $[x, y]$ defined in definition (4.2.1), (iv), and illustrated in Fig.(4.2.1).

However a more general S–rectangle is, in this case, a union of such parallelograms constructed by assigning a point $\underline{\varphi}$ and drawing from it $W^s(\underline{\varphi})$

and $W^u(\underline{\varphi})$ and, on them, a finite number of closed disconnected intervals and, then, performing the construction in Fig. (4.2.2).

Hence an S–rectangle can be disconnected (nevertheless, in this section, only connected S–rectangles will be considered).

$P4.2.3$ **(4.2.3) Proposition:** (Existence of Markovian pavements)
If (Ω, S) is an Anosov system and $\delta > 0$, there exists a Markovian pavement $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ of Ω consisting of S–rectangles of diameter less than δ.

Remarks: (1) In the general case the proof that follows becomes somewhat involved because one cannot rely on simple drawings; however with some imagination the two-dimensional analysis carries over essentially unchanged to the higher dimensional cases. Therefore we restrict ourselves, here, to the two–dimensional case, but we do not make use of several simplifications that would make the analysis not immediately extendible to the general case. See problem [4.3.10] for a simpler construction in the two-dimensional case. Our present analysis therefore follows quite closely the work of Bowen, see [Bo75] p.78–83, in which existence of a Markovian pavement for much more general dynamical systems is derived (cf. definition (4.2.3) and proposition (4.2.5) at the end of this section).
(2) For a first reading it may be useful to follow the arguments by imagining that we are dealing with the very special case of Arnold's cat map, *i.e.* of the diffeomorphism of $\mathbb{T}^2$ generated by the matrix $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. In this case a simpler proof could be devised but, again, it would not be simply extendible to the higher dimensional cases.
(3) Most of what follows is devoted to illustrate a few very simple geometrical constructions. The wording may look intricate but the actual constructions are very simple and easily automatized for use on a digital computer. Therefore drawing what is described in words make the various statements easily intelligible.

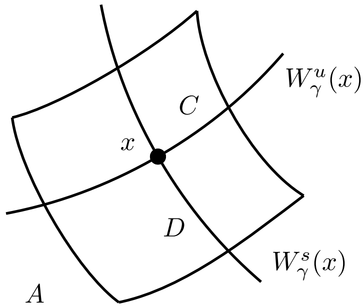
Proof: (case $d = 2$).

(A) *Geometric description of a generic rectangle.*
If A is an S–rectangle and $x \in \mathrm{int}(A)$, set $C = W^u_\gamma(x) \cap A$, $D = W^s_\gamma(x) \cap A$; hence we have

$e4.2.9$
$$A = [C, D] = \bigcup_{y \in C,\, z \in D} [y, z] \tag{4.2.9}$$

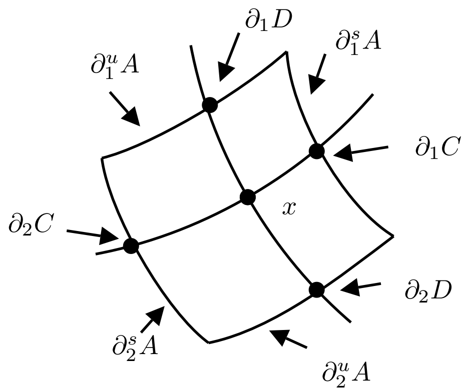
and we call x the *center* of A with respect to the *cross*, or *pair of axes*, C and D (note that any point of A is a center with respect to suitable pair of axes). This is illustrated in Fig.(4.2.3). We shall say that C is an unstable axis and D a stable one; if C, D and C', D' are two pairs of axes for the same rectangle we say that C and C', or D and D', are "parallel"; one has either $C \equiv C'$ or $C \cap C' = \emptyset$.

The boundary of A is composed by four connected sides $\partial^s_\beta A$, $\beta = 1, 2$, $\partial^u_\beta A$, $\beta = 1, 2$; the first two are parallel to the stable axis C and the other two to the unstable axis D, and they can be defined in terms of the boundaries



F4.2.3 **Fig.(4.2.3)** A rectangle A with a pair of axes C, D crossing at the corresponding center x.

∂C and ∂D of C and D considered as subsets of the unstable and stable lines that contain them. Each such boundary consists of two points $\partial_1 C$ and $\partial_2 C$ or $\partial_1 D$ and $\partial_2 D$, see Fig. (4.2.4).



F4.2.4 **Fig.(4.2.4)** The stable and unstable boundaries of a rectangle A and the two pairs of points $\partial_1 C, \partial_2 C$ and $\partial_1 D, \partial_2 D$ that generate them.

We define the stable and unstable parts of the boundary as

e4.2.10
$$\partial^s_\beta A = [D, \partial_\beta C], \qquad \partial^u_\beta A = [\partial_\beta D, C], \qquad \beta = 1, 2. \tag{4.2.10}$$

We shall prove the existence of a pavement $\mathcal{Q}$ of Ω with S–rectangles of diameter $< 2\alpha$, where $\alpha = \delta/2$ is half the preassigned δ, such that for all $\beta = 1, 2$ and for all $Q \in \mathcal{Q}$ one has

e4.2.11
$$S\partial^s_\beta Q \subset \partial^s_{\beta'} Q', \qquad S^{-1}\partial^u_\beta Q \subset \partial^u_{\beta''} Q'', \tag{4.2.11}$$

where β', β'', Q', Q'' depend on β and Q.

It is important to realize immediately that a pavement $\mathcal{Q}$ built with S–rectangles verifying (4.2.11) is Markovian. In fact equation (4.2.11) obviously implies (setting $\partial^s Q = \partial^s_1 Q \cup \partial^s_2 Q$ and $\partial^u Q = \partial^u_1 Q \cup \partial^u_2 Q$)

e4.2.12
$$S\,(\bigcup_{Q\in\mathcal{Q}} \partial^s Q) \subset \bigcup_{Q\in\mathcal{Q}} \partial^s Q, \qquad S^{-1}(\bigcup_{Q\in\mathcal{Q}} \partial^u Q) \subset \bigcup_{Q\in\mathcal{Q}} \partial^u Q, \tag{4.2.12}$$

N4.2.7 and it is also implied by (4.2.12).[7]

(B) *An equivalent problem.*
It is convenient to suppose that the map S has very large expansion and contraction rates. This is not a loss of generality because the problem can be further reduced to showing existence of a pavement $\mathcal{B}$ of Ω consisting of S–rectangles of diameter $< \alpha$ and such that for some $m > 0$

e4.2.13
$$S^m \mathop{\cup}_{B\in\mathcal{B}} \partial^s B \subset \mathop{\cup}_{B\in\mathcal{B}} \partial^s B\ , \qquad S^{-m} \mathop{\cup}_{B\in\mathcal{B}} \partial^u B \subset \mathop{\cup}_{B\in\mathcal{B}} \partial^u B. \tag{4.2.13}$$

Given such a pavement $\mathcal{B}$, a pavement $\mathcal{Q}$ satisfying the previous property (4.2.12) can be constructed as follows. We consider, if $m > 1$, the pavement $\mathcal{Q}$ whose elements are the S–rectangles Q having the form

e4.2.14
$$Q = \mathop{\cap}_{k=0}^{m-1} S^k B_k, \tag{4.2.14}$$

with $B_1, \ldots, B_{m-1} \in \mathcal{B}$. Here we use the fact that the intersection between two small S–rectangles *is still an S–rectangle*; it is easy to check that such S–rectangles are a pavement of Ω verifying (4.2.12).
The proof of the existence of $\mathcal{B}$, which is what remains to do, is in fact a description of a nice explicit construction of $\mathcal{B}$.

(C) *Construction of a covering by rectangles.*
We begin by considering a *covering* $\mathcal{A}^0 = \{A_1^0, \ldots, A_r^0\}$ of Ω by means of connected S–rectangles whose internal points cover Ω; this is possible because Ω is a compact connected Riemannian manifold. Furthermore we suppose that such S–rectangles have small diameter $\alpha < \min\{\delta/2, \gamma/2\}$, where δ is the length prefixed in the statement of the proposition and γ is the size of the local portions of stable and unstable manifolds (whose existence is part of the definition of Anosov system).
We imagine that the rectangles A_j^0 are constructed, as discussed in item (A) above, as products of two axes C_j^0 and D_j^0 through a center point x_j:

e4.2.15
$$A_j^0 = [C_j^0, D_j^0]. \tag{4.2.15}$$

(D) *Lebesgue length of the covering.*
A key role will be plaid by a certain length $a > 0$ that is a natural extension of the notion of *Lebesgue length* associated with a covering of a compact space by finitely many open sets. A Lebesgue length of a covering is defined as a length a such that every point x is "well inside" *some* element of the

[7] One checks directly the properties of definition (4.1.3), after having visualized (with the help of a drawing) the geometric meaning of (4.2.11). Use that the two relations in (4.2.12) must be simultaneously valid and that S is a map that transforms points relatively internal to $\partial^s_\beta C$ into points relatively internal to $S\partial^s_\beta C$: assuming that (4.2.11) does not follow from (4.2.12) attempt to represent the situation by a drawing in order to see the arising absurdity.

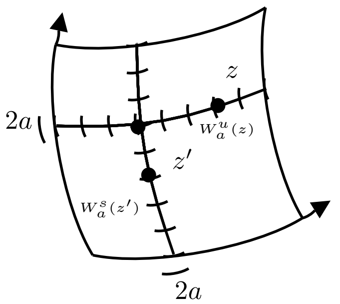
covering in the sense that the entire sphere of radius a around any point x is contained in the interior of some element of the covering.

The extension that we need here is that out of a covering by S–rectangles like $\mathcal{A}^0$ we can always find for each x an element $A_{F(x)} \in \mathcal{A}^0$, with a suitable label $F(x)$, whose boundary parallel to the s–direction stays at a distance $> a$ from $W^s_\gamma(x) \cap A_{F(x)}$, *i.e.* from the stable axis of $A_{F(x)}$ through x; and another one $A_{F'(x)} \in \mathcal{A}^0$, with a suitable label $F'(x)$, whose boundary parallel to the u–direction stays at a distance $> a$ from $W^u_\gamma(x) \cap A_{F'(x)}$, *i.e.* from the unstable axis of $A_{F'(x)}$ through x; actually we can even suppose that $F(x) = F'(x)$. To be precise we should require that, for every $y \in W^s_x \cap A^0_{F(x)}$, the distance between y and the intersection of $W^s_\gamma(y)$ with $\partial^u A^0_{F(x)}$ is greater than a if it is measured along $W^s_\gamma(y)$. Here and in what follows we will always assume that the length α is so small that one can think of $W^s_a(x)$ as a straight segment of length $2a$ centered in x. This simplifies the presentation following argument without any loss of generality.

Given $\mathcal{A}^0$ we shall fix a length $a > 0$ enjoying the above properties and such that $a < \alpha/2$ where $\alpha < \delta/2$ is the above defined maximum diameter of the elements of $\mathcal{A}\mathcal{A}^0$ (the reason for this choice will become clear in following). We call a a Lebesgue length for $\mathcal{A}^0$.

(1) Imagine to have drawn through x the manifolds $W^u_\gamma(x) \cap A^0_{F(x)}$ and $W^s_\gamma(x) \cap A^0_{F(x)}$.

(2) Through each of the points z in $W^u_\gamma(x) \cap A^0_{F(x)}$ and z' in $W^s_\gamma(x) \cap A^0_{F(x)}$ we can draw, respectively, the segments $W^s_a(z)$ and $W^u_a(z')$ of unstable and stable manifold, respectively, with $a < \alpha/2$: by our definitions and our choice of the Lebesgue length a such segments *lie entirely in* $A^0_{F(x)}$, as illustrated in Fig.(4.2.5).



F4.2.5 **Fig.(4.2.5)** Illustration of the check that each point of Ω is far at least as a in the direction of both stable and unstable manifolds from the boundaries of "some" element of the covering $\mathcal{A}^0$. The short segments are portions of stable or unstable manifold of length $2a$, where a is (much) smaller than the diameter α of the rectangle.

(E) *Replacing S by $g = S^m$ with m large.*

Let $m > 0$ be an integer such that the following sum is small enough (*i.e.* as small as indicated)

e4.2.16
$$2\sum_{k=1}^{\infty} \lambda^{mk}\alpha = 2\lambda^m(1-\lambda^m)^{-1}\alpha < a/2. \qquad (4.2.16)$$

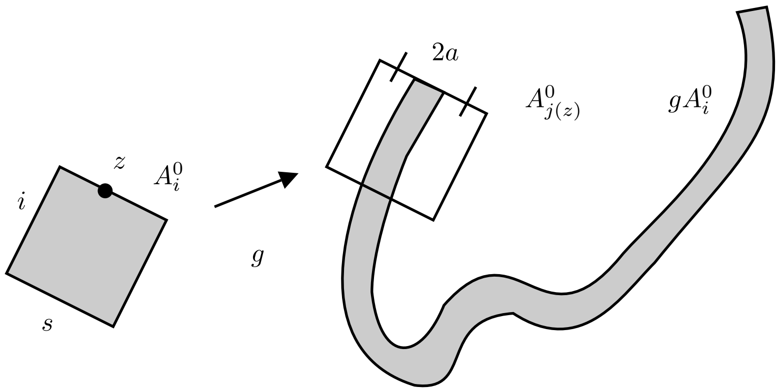
Calling $g \stackrel{def}{=} S^m$ we show that a Markovian pavement $\mathcal{B}$ for g can be constructed starting from a covering $\mathcal{A}$ made of S–rectangles with diameter less than $2\alpha < \delta$ for which $a < \alpha/2$ is a Lebesgue distance and verifying the following properties.

(I) For every $\beta = 1, 2$ and $A \in \mathcal{A}$ there exist β', $\beta'' \in \{1, 2\}$, A', $A'' \in \mathcal{A}$ for which

e4.2.17
$$g\,\partial^s_\beta A \subset \partial^s_{\beta'} A', \qquad g^{-1}\,\partial^u_\beta A \subset \partial^u_{\beta''} A''. \tag{4.2.17}$$

(II) Every $z \in \partial^s A_i$ is such that gz is "well inside" a boundary of the same type (*i.e.* stable) of some other rectangle $A_{j(z)}$, *i.e.* $W^s_{a/2}(gz) \subset \partial^s A_{j(z)}$ for a suitable $j(z)$; and, symmetrically, every $z \in \partial^u A_i$ is such that $W^u_{a/2}(g^{-1}z) \subset \partial^u A_{k(z)}$ for a suitable $k(z)$ (see Fig. (4.2.6)).

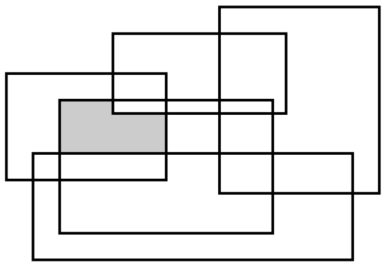
To construct $\mathcal{B}$ starting from $\mathcal{A}$ is the main difficulty of the analysis because realizing properties (I) and (II) will be rather straightforward. We shall see later how to construct the covering $\mathcal{A}$ starting form the initial covering $\mathcal{A}^0$.



F4.2.6 **Fig.(4.2.6)** Geometrical meaning of the properties (I) and (II). The g image of the part of the boundary of the (shaded) rectangle A^0_i on the left containing z is mapped, greatly shortened to a size $< a/2$ at least by (4.2.16), on the upper boundary of another rectangle of the covering $A^0_{j(z)}$; and in fact well in the middle so that even by widening it on either side by $(1-\frac{1}{2})a = a/2$ it remains in the upper boundary of $A^0_{j(z)}$.

Indeed suppose that the covering $\mathcal{A}$ satisfies the above described properties (I) and (II). We can obtain a finer pavement of Ω by intersecting some of the set of $\mathcal{A}$ with the complements of the others. More precisely for every subset $\mathbf{r} \subset \{1, \ldots, r\}$ we can set $P_\mathbf{r} = (\cap_{i\in\mathbf{r}} A_i) \cap (\cup_{i\not\in\mathbf{r}} \Omega\backslash A_i)$. Clearly the
N4.2.8 collection $\mathcal{P}$ of the non-empty $P_\mathbf{r}$ forms a pavement of Ω [8] such that for all $P \in \mathcal{P}$ one has $g\partial^s P \subset \cup_{P'\in\mathcal{P}}\,\partial^s P'$, $g^{-1}\partial^u P \subset \cup_{P'\in\mathcal{P}}\,\partial^u P'$, *i.e.* $\mathcal{P}$ is close to satisfy (4.2.13). Moreover remark that the sets $P_\mathbf{r}$ are the connected

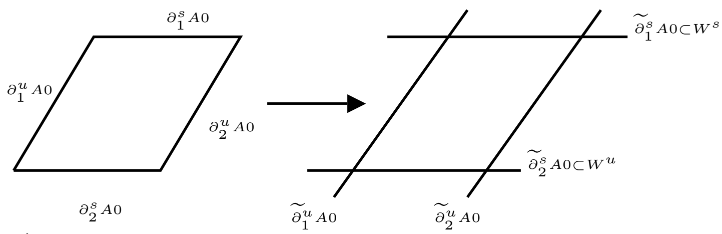
[8] The elements P of $\mathcal{P}$ have boundaries constituted by stable parts ($\partial^s P$) and by unstable parts ($\partial^u P$).



F4.2.7 **Fig.(4.2.7)** The sets obtained by intersecting in all possible ways the elements of the covering $\mathcal{A}$ have interiors which are pairwise disjoint but are not necessarily rectangles, *e.g.* see the gray set. One can call such sets "incomplete" rectangles.

components of $\Omega \backslash \cup_{A\in\mathcal{A}}\, \partial A$. *However* $\mathcal{P}$ has the defect of not necessarily consisting of S–rectangles (see Fig. (4.2.7)).

Nonetheless it is possible to build from $\mathcal{P}$ a pavement consisting of S–rectangles and verifying (4.2.13). One just "continues a little" the sides of every set $A \in \mathcal{A}$ along the stable or unstable manifold that contains them, see Fig. (4.2.8), *until an encounter with a boundary* (of unstable or stable type, respectively, *i.e.* of "opposite" type) is obtained. The boundary where we stop the continuation is the first one meets (this construction depends on the order in which the the sides to be continued are examined, hence it contains some arbitrariness). Since the sets in $\mathcal{A}$ are very small the length of the added parts of lines will be small (*i.e.* not exceeding 2α).



F4.2.8 **Fig.(4.2.8)** Prolonging the sides of the rectangles of the covering $\mathcal{A}$ in order to "complete" the incomplete rectangles like the one shaded in Fig.(4.2.7). The continuation is performed until a line reaches a boundary of a set in $\mathcal{A}$ (see Fig.(4.2.9)) and it means prolonging by a length at most 2α.

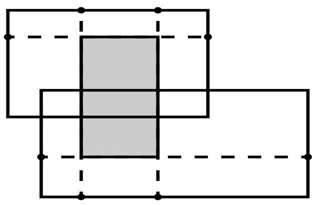
More precisely the continuation can be done by replacing $\partial^s_\beta A, \partial^u_\beta A$ by

e4.2.18
$$\widetilde{\partial}^s_\beta A = \bigcup_{x\in\partial^s_\beta A} W^s_\gamma(x), \qquad \widetilde{\partial}^u_\beta A = \bigcup_{x\in\partial^u_\beta A} W^u_\gamma(x), \tag{4.2.18}$$

i.e. we continue $\partial^s_\beta A, \partial^u_\beta A$ on either side by adding a piece of manifold of the same type (*i.e.* stable or unstable) of size γ.

Of course the sets $\widetilde{\partial}^s_\beta A, \widetilde{\partial}^u_\beta A$ will go beyond the point where they first meet the boundary of the elements of $\mathcal{A}$ which intersect A. This is so because diam $(A) < 2\alpha$ for all $A \in \mathcal{A}$ and $\gamma > 2\alpha$.

However we can just delete the parts of such lines that go outside the elements of $\mathcal{A}$ which intersect A. What is left is represented in Fig.(4.2.9) where A_0 is the shaded rectangle and the dashed lines correspond to the part of the continuations of the boundaries of A_0 which have not been deleted: we denote the lines so constructed by $\widehat{\partial}^s_\beta A \subset \widetilde{\partial}^s_\beta A$ and $\widehat{\partial}^u_\beta A \subset \widetilde{\partial}^u_\beta A$. This construction is repeated for each of the rectangles of $\mathcal{A}$ and at the end we shall have a pavement consisting only of S–rectangles, *i.e.* there will be no more "incomplete rectangles" like those appearing in Fig.(4.2.7).



F4.2.9 **Fig.(4.2.9)** Rectangles obtained by the geometric operations illustrated in Fig. (4.2.7), (4.2.8) applied to the shaded rectangle: we see that the number of "incomplete" rectangles becomes smaller (at the expense of an increase in the number of rectangles). The size of the dashed lines is $\le 2\alpha$. The construction has to be repeated for each rectangle, successively.

If $z \notin \cup_j(\widehat{\partial}^s A_j \cup \widehat{\partial}^u A_j)$, *i.e.* if z is not on the boundary of any of the just constructed rectangles, there exists a unique connected component of $\Omega \setminus \cup_j(\widehat{\partial}^s A_j \cup \widehat{\partial}^u A_j)$ that contains z. This component is clearly an open set and its closure is a S-rectangle B. As z varies such S–rectangles describe a finite family $\mathcal{B}$ that is a pavement of Ω with S–rectangles of diameter $< 2\alpha$, being intersections of rectangles with diameter already $< 2\alpha$.

Let $z \in \partial^s_\beta B$, $B \in \mathcal{B}$; then z will not be, in general, on the boundary of some elements of $\mathcal{A}$ but, by construction, there will exist $x \in \{$boundary of a suitable rectangle $A \in \mathcal{A}\}$ such that $z \in W^s_{2\alpha}(x)$. Since $\mathcal{A}$ enjoys, by hypothesis, properties (I) and (II) (cf. (4.2.17)) and $gW^s_{2\alpha}(x) \subset W^s_{a/2}(gx)$ by (4.2.16) we get

e4.2.19 $$g\,x \in \partial^s A_{j(x)} \qquad \text{and} \qquad W^s_{a/2}(g\,x) \subset \partial^s A_{j(x)}. \tag{4.2.19}$$

Noting that the diameter of $g\,W^s_{2\alpha}(x)$ is smaller than $2\lambda^m\alpha < a/2$ by the choice (4.2.16) of m, we see that $gW^s_{2\alpha}(x) \subset W^s_{a/2}(gx)$ so that (4.2.19) implies

e4.2.20 $$g\,\partial^s_\beta B \subset \bigcup_{B' \in \mathcal{B}} \partial^s B', \tag{4.2.20}$$

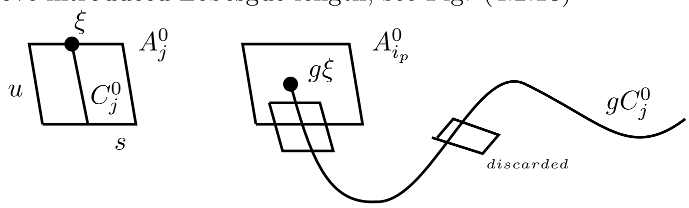
and, likewise, we would evince that

e4.2.21 $$g^{-1}\partial^u_\beta B \subset \bigcup_{B' \in \mathcal{B}} \partial^u B'. \tag{4.2.21}$$

This means that $\mathcal{B}$ satisfies (4.2.13), *i.e.* it is a Markovian pavement for the map g. As we noticed in point (B) it is easy to construct a Markovian pavement $\mathcal{B}'$ for S starting from $\mathcal{B}$, see (4.2.14). Moreover in (4.2.14) the element B_0 of $\mathcal{B}$ appearing in the intersection has diameter less than 2α so that the diameter of the elements of $\mathcal{B}'$ is also smaller that 2α.

Therefore we are reduced to show that if the covering $\mathcal{A}^0$ does not already enjoy properties (I) and (II) stated in connection with (4.2.17) above, it is still possible to modify it slightly so that it becomes a covering $\mathcal{A}$ of the same type which, however, satisfies properties (I) and (II) above. This can be done through an inductive procedure of successive approximations.

(F) *Successive refinements to build a covering verifying properties (I) and (II)*
Consider $g\,C^0_j$, $j=1,\ldots,r$, and note that it is a "very long curve" (because of our choice of large m and because $g\equiv S^m$). Select a covering $A^0_{i_1},\ldots,A^0_{i_{k_j}}$ of $g\,C^0_j$ made of elements of $\mathcal{A}^0$ such that gC^0_j *"passes"* through each of them and *well inside*, *i.e.* at a distance $\ge a$ from the parallel boundary where a is the above introduced Lebesgue length, see Fig. (4.2.10)



F4.2.10 **Fig.(4.2.10)** Here the g-image of C^0_i is represented as a long curve. The covering of the g-image is realized by discarding the elements of $\mathcal{A}^0$ whose expanding boundaries pass too close to $g\,C^0_i$, *i.e.* closer than the Lebesgue length of $\mathcal{A}^0$, see item (D).

In general some $A^0_{i_p}$, in this covering, are not completely crossed by gC^0_j; see the square $A^0_{i_p}$ in Fig. (4.2.10). Then continue a little C^0_j so that gC^0_j crosses *all* the sets $A^0_{i_1},\ldots,A^0_{i_k}$ of the just considered covering of $g\,A^0_j$: this means replacing A^0_j with $A^1_j=[C^1_j,D^0_j]$, where

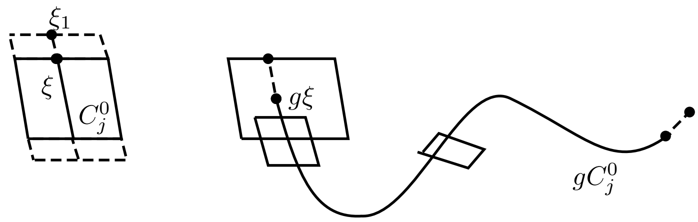
$$C^1_j=\mathop{\cup}_{h=1}^{k_j} g^{-1}\Big(\{\text{continuation of } gC^0_j \text{ until it crosses all of } A^0_{i_h}\}\cap A^0_{i_h}\Big)=\mathop{\cup}_{h=1}^{k_j} g^{-1}([C^0_{i_h},g\,C^0_j\cap A^0_{i_h}]);$$

see Fig. (4.2.11).

Inductively we shall set

e4.2.22 $$C^n_j=\mathop{\cup}_{h=1}^{k_j} g^{-1}([C^{n-1}_{i_h},g\,C^{n-1}_j\cap A^{n-1}_{i_h}]). \tag{4.2.22}$$

At each step the lengthening of C^n_j with respect to C^{n-1}_j is by about a factor λ^m shorter so that it is $\le 2(\lambda^m)^n\alpha$: this is seen by induction, recalling that diam $(A^0_j)\le\alpha$.



F4.2.11 **Fig.(4.2.11)** Enlarging the manifold C_j^0 up to the point ξ_1 so that the g–image of the enlarged manifold ends on the appropriate boundary of the rectangle $A^0_{i_p}$, with the notations of the previous figure.

Hence we can pass to the limit $n \to \infty$ to define $C_j = \bigcup_{n=0}^{\infty} C_j^n$ and this makes sense if α and λ^m are so small that (see (4.2.16)) $\sum_{n\ge 1}^{\infty}(\lambda^m)^n 2\alpha < a < \alpha/2 < \gamma/2$.

Likewise one proceeds to lengthen D_j^0 to construct D_j^n and to define D_j (using g^{-1} instead of g).

Define then the rectangles A_j, with $j = 1, \ldots, r$, as $[C_j, D_j]$, that form a covering $\mathcal{A}$ of Ω with sets of diameter smaller than 2α.

By construction, if $z \in A_j \in \mathcal{A}$, there exist $A_{j'} \in \mathcal{A}$ such that $g\, z \in A_{j'}$ and $g^{-1}z \in A_{j''}$ and

e4.2.23
$$g\,[C_j, z] \supset [C_{j'}, g\, z], \qquad g^{-1}[z, D_j] \supset [g^{-1}z, D_{j''}]. \tag{4.2.23}$$

Furthermore: $A_{j'} \supset W^s_{\frac{a}{2}}(y)$ for all $y \in [C_{j'}, gz]$, and $A_{j''} \supset W^u_{\frac{a}{2}}(y)$ for all $y \in [g^{-1}z,\, D_{j''}]$, provided that, as we can suppose, α and, therefore, the rectangles are small enough so that their sides can be treated as straight in the continuation operations. Drawings can be very useful to visualize the above statements; furthermore diam $(A_j) < 2\alpha \le \delta$.

Hence $\mathcal{A}$ satisfies the properties (I) and (II) stated in connection with the (4.2.17). The correspondence that associates a point $x \in \Omega$ with a compatible sequence $\underline{\sigma}$ is Hölder continuous, (4.1.6), and its inverse has finite multiplicity if the value of δ is small enough so that the intersection between any element of $\mathcal{P}$ and the S–image of any other is a connected set, which certainly happens if α is small enough. Hence the proof is complete. ■

A consequence of the definition (4.2.1) is the following result.

P4.2.4 **(4.2.4) Proposition:** (Transitivity and mixing of compatibility matrices for Markovian pavements for Anosov maps)
If $\mathcal{P}$ is a Markovian pavement for an Anosov map, cf. *definition (4.2.1), then its compatibility matrix is transitive and mixing (*cf. *definition (4.1.1)).*

Proof: By definition (4.2.1) and problem [4.1.19] transitivity follows. The mixing property is tightly related to the assumption that Anosov systems are maps of a *connected* phase space Ω. Even dropping connectedness it is still possible to construct a Markovian pavement by the construction of proposition (4.2.3). The transition matrix will have the properties described in the problems [4.1.4] through [4.1.11]. Hence the system can be decomposed into a union of disjoint closed sets $\cup_{i=1}^{a-1} \cup_{j=1}^{d_i} \Omega_{i,j}$ and a "remainder set" Ω' open and invariant (consisting of the wandering points, cf. problem [4.1.16]) by problem [4.1.17] which must be empty by the transitivity assumption: connectedness then implies $a = 1, d = 1$ hence $\Omega = \Omega_{1,1}$ and S acts in a topologically mixing way on Ω so that also T is mixing. ■

The just described construction of a Markovian pavement is generalizable to a situation that, in some applications, turns out to be quite useful; we shall describe it in the following definition (due to Ruelle, who called the systems that will be defined below an *abstract hyperbolic systems* or *Smale systems*, see p. 125–130 in [Ru78]).

D4.2.3 **(4.2.3) Definition:** (Abstract hyperbolic systems)
Let (Ω, S) be a topological dynamical system. Suppose that (Ω, S) is topologically transitive (i.e. there exists a point x of Ω with a dense orbit $(S^n x)_{n\in\mathbb{Z}^+}$).
The system (Ω, S) is said to be an abstract hyperbolic (transitive) system *if there exist a metric d for the topology of Ω and three constants $\lambda < 1$, $\gamma > 0$, $\varepsilon > 0$, with $\lambda > 0$ and $\gamma > \varepsilon$, and a function $x,\, y \to [x, y]$ defined on the pairs $\{x, y \,|\, x, y \in \Omega, d(x, y) < \varepsilon\}$ such that*
(1) $[x, x] = x$
(2) $[[x, y], z] = [x, z]$ and $[x, [y, z]] = [x, z]$, whenever the last expressions have a meaning.
(3) $S[x, y] = [Sx, Sy]$, when both sides have a meaning.
(4) Setting

e4.2.24
$$\begin{aligned} W^s_\gamma(x) &= \{u | u = [u, x] \ , \qquad d(u, x) < \gamma\},\\ W^u_\gamma(x) &= \{u | u = [x, u] \ , \qquad d(u, x) < \gamma\},\end{aligned} \tag{4.2.24}$$

it follows that for all $n > 0$

e4.2.25
$$\begin{aligned} d(S^n y, S^n z) &< \lambda^n d(y, z) \quad \textit{if} \quad y, z \in W^s_\gamma(x),\\ d(S^{-n} y, S^{-n} z) &< \lambda^n d(y, z) \quad \textit{if} \quad y, z \in W^u_\gamma(x).\end{aligned} \tag{4.2.25}$$

Remark: The simplest examples of such systems are naturally the subshifts of finite type with a transitive compatibility matrix. The set $W^s_\gamma(\underline{\sigma})$ consists of all the strings that agree with $\underline{\sigma}$ on the sites with label $i \geq -p$ if $\gamma = e^{-p}$.

Then the following proposition holds.

P4.2.5 **(4.2.5) Proposition:** (Bowen)
Every abstract hyperbolic system admits a Markovian pavement with sets of diameter $\leq \delta$ where $\delta > 0$ is arbitrarily prefixed.

The proof of this proposition follows, *grosso modo*, the ideas in the proof of proposition (4.2.3). It presents difficulties of topological character due to the lack of hypotheses of connectedness and differentiability on Ω and

S: this forces us to be careful in stating as obvious certain properties that are such in the case studied in proposition (4.2.3). See, for details, [Bo75], [Ru76].

A remarkable property enjoyed by such systems is the so called *shadowing* or "existence of a shadow motion".

P4.2.6 **(4.2.6) Proposition:** (Shadow symbols)
Let T be a transitive $n \times n$ compatibility matrix. Let $\{\underline{\sigma}_j\}_{j=-\infty}^{\infty}$ be a sequence of elements (sequences) in $\{0,\ldots,n-1\}_T^{\mathbb{Z}}$ such that $d(\tau\underline{\sigma}_j, \underline{\sigma}_{j+1}) < \delta$ where τ is the shift; the sequence $\{\underline{\sigma}_j\}$ is called a δ–approximate motion for the dynamical system $(\{0,\ldots,n-1\}_T^{\mathbb{Z}}, \tau)$. Then there exists $C > 0$ independent of the sequence $\{\underline{\sigma}_j\}_{j=-\infty}^{\infty}$ and $\underline{\sigma} \in \{0,\ldots,n-1\}_T^{\mathbb{Z}}$ such that $d(\tau^j\underline{\sigma}, \underline{\sigma}_j) < C\delta$.

Remark: Therefore given a symbolic motion that is at each time perturbed by δ one can find a true motion which remains close to the perturbed motion within $C\delta$ for *all times* $j \ge 0$ where C is a universal constant (*e.g.* $C = e$). In colorful words "the perturbed motion is a shadow of a true motion" (under a slightly trembling light). This is particularly interesting as it implies that the "same" property holds for motions in a more general Smale system, hence for Anosov maps: see proposition (4.2.7) below.

Proof: The distance between elements $\underline{\sigma}, \underline{\sigma}'$ of $\{0,\ldots,n-1\}_T^{\mathbb{Z}}$ is the exponential of minus the maximal k such that $\sigma_i = \sigma_i'$ for all $|i| \le k$; see (4.1.6). Let $\delta = e^{-p}$. We define σ_j for $|j| \le p$ by $\sigma_j = (\underline{\sigma}_0)_j$. Then we set $\sigma_{p+1} = (\underline{\sigma}_1)_{p+1}$ and more generally $\sigma_{p+j} = (\underline{\sigma}_j)_{p+j}$ for $j > 0$ while $\sigma_{-p+j} = (\underline{\sigma}_j)_{-p+j}$ for $j < 0$. Then the result is true with $C = 1$. If δ does not have the form e^{-p} with p integer then the same argument can be repeated but one gets that C can be larger than 1 but always $< e$. Hence $C = e$. ∎

P4.2.7 **(4.2.7) Proposition:** (Shadowing)
Let (Ω, S) be an abstract hyperbolic system, in the sense of definition (4.2.3) above. There is a constant C such that if $\ldots, x_{-1}, x_0, x_1, \ldots$ is a sequence with $d(Sx_j, x_{j+1}) < \delta$ and δ is small enough then there exists a point x such that $d(S^j x, x_j) < C\delta$ for all j.

Proof: The ambiguity in the code X that associates a symbolic motion with a point forbids saying that the proof of the previous proposition implies this result. However we can simply mimic the proof of proposition (4.2.6). We start with x_0 and construct the image Sx_0 which will be very close to x_1 so that a point $\xi_1 = [Sx_0, x_1]$ on the unstable manifold of Sx_0 and on the stable of x_1 exists. We have that $S^{-1}\xi_1$ is on the unstable manifold of x_0. Then we construct $S\xi_1$ and define $\xi_2 = [S\xi_1, x_2]$ which is a point such that $S^{-2}\xi_2$ is on the unstable manifold of x_0. More generally $\xi_n = [S\xi_{n-1}, x_n]$: all points $S^{-i}\xi_i$ are on the unstable manifold of x_0 and all points ξ_i are at distance $< \delta$ from x_i and from the $S\xi_{i-1}$. The limit $\xi = \lim_{n\to\infty} S^{-n}\xi_n$ exists because by construction the variation in position of $S^{-n}\xi_n$ with respect to $S^{-(n-1)}\xi_{n-1}$

has size of order $\lambda^n\delta$ because the points ξ_n lie systematically on the unstable manifold of x_0: hence ξ exists and is on the unstable manifold of x_0. This also implies that there is C such that $d(S^j\xi, x_j) < C\delta$ for $j > 0$. We can repeat this construction using S^{-1} and the points $x_{-1}, x_{-2}, \ldots$ and find a point η on the stable manifold of x_0 such that $d(S^j\eta, x_j) < C\delta$ for $j < 0$. It is now enough to set $x = [\xi, \eta]$. ■

Remark: It should be noted that the proof of proposition (4.2.6) immediately suggests the proof of proposition (4.2.7) if one takes into account that in symbolic dynamics the unstable "manifold" (quotes used here because it is not a manifold but just a set) consists of the compatible sequences which agree over the labels to the left of a suitable label (*i.e.* for times preceding a certain time).

Problems for the §4.2 (*Existence, regularity, smoothness, uniqueness of the stable and unstable foliations for Anosov maps*)

Q4.2.1 **[4.2.1]:** (*Continuity and invariance of the foliations*, from [AS67])
In the context of proposition (4.2.1) do not assume continuity nor covariance of the fields of spaces V^s_x, V^u_x as functions of x and replace the assumption in (4.2.8) by

$$\begin{aligned}
\|(dS^n)v\|_{V_{S^n x}} &\le C\lambda^n\|v\|_{V_x} && \text{for all } v\in V^s_x,\\
\|(dS^{-n})v\|_{V_{S^{-n} x}} &\le C\lambda^n\|v\|_{V_x} && \text{for all } v\in V^u_x,\\
\|(dS^{-n})v\|_{V_{S^{-n} x}} &\ge C\lambda^{-n}\|v\|_{V_x} && \text{for all } v\in V^s_x,\\
\|(dS^n)v\|_{V_{S^n x}} &\ge C\lambda^{-n}\|v\|_{V_x} && \text{for all } v\in V^u_x,
\end{aligned} \tag{$*$}$$

for all positive n. Show that this implies (a) covariance of the fields, and (b) continuity of the fields. (*Hint:* Continuity at fixed n of $dS^{\pm n}$ implies that for $a = s, u$ if $x_j \to x_0$ and $v_j \in V^a_{x_j} \to v_0$ then $v_0 \in V^a_{x_0}$. This proves "lower semicontinuity" of the function $x \to d^a(x) \stackrel{def}{=}$ dimension of V^a_x. However the dimension is an integer and $d^s(x)+d^u(x)$ is identically equal to the dimension of the manifold. Hence the dimensions of V^a_x are both upper and lower semicontinuous, *i.e.* they are continuous, and being integer valued they are constant. Therefore the above inclusion $\lim V^a_{x_n} \subseteq V^a_{x_0}$ implies continuity of V^a_x.)

Q4.2.2 **[4.2.2]:** Let J be a matrix on $\mathbb{R}^d$ and assume that $\mathbb{R}^d = V^s \oplus V^u$ and that there is $\lambda < 1$ such that for all $n \ge 0$

$$\begin{aligned}
|J^n v| &< \lambda^n|v|, \quad v\in V^s, \qquad |J^{-n}v| < \lambda^n|v| \quad v\in V^u,\\
|J^{-n} v| &> \lambda^{-n}|v|, \quad v\in V^s, \qquad |J^{n}v| > \lambda^{-n}|v| \quad v\in V^u,
\end{aligned} \tag{$**$}$$

then V^s, V^u are invariant under the action of $J^{\pm1}$. Furthermore there exist $\alpha, \varepsilon > 0$ such that the cones $\Gamma^{u,\alpha}$ and $\Gamma^{s,\alpha}$ consisting of all the lines forming an angle $\le \alpha$ with the plane V^s or, respectively, V^u are such that

$$J\Gamma^{u,\alpha} \subseteq \Gamma^{u,(1-\varepsilon)\alpha}, \qquad J^{-1}\Gamma^{s,\alpha} \subseteq \Gamma^{s,(1-\varepsilon)\alpha}$$

i.e. the cones $\Gamma^{u,\alpha}$ and $\Gamma^{s,\alpha}$ "shrink by a factor $(1-\varepsilon)$" under the action of J or, respectively, J^{-1}.

Q4.2.3 **[4.2.3]:** If in the right hand side of $(**)$, in problem [4.2.3], one inserts a constant C then the same conclusions hold if J is replaced by J^{n_0} with n_0 large enough.

Q4.2.4 **[4.2.4]:** (*Hadamard–Perron theorem at a hyperbolic fixed point*)
Let S be a C^∞ map of $\mathbb{R}^d$ into itself with a fixed point at the origin and with a Jacobian matrix J at the origin which verifies the property in $(**)$, in problem [4.2.4], for a given

$\lambda < 1$ and for all $n \ge 0$. Given $k > 0$, inside a sphere of radius $\delta_k > 0$, with δ_k small enough, around the origin there exist two surfaces W^s and W^u of class C^k tangent to V^s and V^u at the origin, respectively, and such that $SW^s \subset W^s$ and $S^{-1}W^u \subset W^u$. Furthermore there exists $C > 0$ and $\lambda' < 1$ such that $|S^n x| < C\lambda'^n|x|$ for $x \in W^s$ and $|S^{-n}x| < C\lambda'^n|x|$ for $x \in W^u$. (*Hint:* Let $k = 0$ and let Σ_δ be the sphere in V^s of radius δ and center at the equilibrium point. The results of problems [4.2.2] and [4.2.3] imply that if δ is small enough then a smooth surface whose points have the form $(s, \gamma(s))$, with $s \in V^s \cap \Sigma_\delta$ and $\gamma(s)$ a smooth function, and whose tangent plane forms everywhere an angle with respect to V^s not exceeding α is mapped by S^{-1} into a surface of the same type inside Σ_δ but extends beyond the boundary of Σ_δ. Then one modifies S^{-1} *outside* the sphere Σ_δ. Let (s, u) be a point near the origin and write $S^{-1}(s, u) = (f(s, u), J^{-1}u + g(s, u))$, where, with a slight abuse of notation, we are denoting $J^{-1}(0, u) = (0, J^{-1}u)$: modify the map S^{-1} outside Σ_δ into a new auxiliary map

$$X(s, u) = \Big(\sigma(s, u)f(s, u)\,,\; J^{-1}u + \sigma(s, u)g(s, u)\Big),$$

with $\sigma(s, u) = \tilde\sigma(\rho)$ where $\rho = sqrtu^2 + s^2$ and $\tilde\sigma(\rho)$ is a C^∞ non-negative decreasing function that is identically 1 for $\rho \le \delta$ and is 0 for $\rho > 2\delta$. Check that if δ is small enough then the map X still maps the cone $\Gamma^{s,\alpha}$ into itself *as well as any surface γ tangent to V^s at the origin and forming with V^s and angle $< \alpha$ into a surface of the same type.* Furthermore if $(s, \gamma(s))$ and $(s, \eta(s))$ are two such surfaces and $(s, \gamma'(s))$, $(s, \eta'(s))$ are their X–images, one has $\max_{|s|<2\delta} |\gamma'(s) - \eta'(s)| \le \lambda' \max_{|s|<2\delta} |\gamma(s) - \eta(s)|$, where λ' can be chosen as close as wished to λ provided δ is taken sufficiently small. Therefore if $(s, \gamma_0(s))$ with $\gamma_0(s) \equiv 0$ is a special initial surface Δ_0 the iterates $X^n\Delta_0$ converge exponentially fast to a limit Δ_∞ which has the property of being tangent at the origin to V^s (because such are all $S^n\Delta_0$) and of being invariant (because it is a limit of iterates); it has also the property of contracting exponentially to the origin (because every point in the cone $\Gamma^{s,\alpha} \cap \Sigma_\delta$ gets closer to the origin); furthermore the function $\gamma_\infty(s)$ defining Δ_∞ verifies a Lipshitz property (being the limit of functions with derivatives bounded by α). Hence Δ_∞ is a continuous surface and, in fact, it is even Lipshitz continuous. More generally if we fix $k > 0$ one proceeds in a similar way by controlling also the derivatives of the function $\gamma_\infty(s)$, and one can prove that Δ_∞ is a C^{k-1} regular surface with $(k-1)$–th derivatives verifying a Lipshitz property.)

Q4.2.5 **[4.2.5]:** (*Hadamard–Perron theorem at a hyperbolic point*)
Let S be a system that satisfies point (i) and (ii) of proposition (4.2.1) with C replaced by 1 in (4.2.8). Moreover let x_0 be a point whose trajectory under S is $x_j = S^j x_0$, $j \in \mathbb{Z}$. We imagine to consider a coordinate system that "follows" x_0 in its motion: this means that we consider a sequence of spheres $\Sigma_\delta(S^j x)$ on which a coordinate system with origin at x_j is defined introducing coordinates (s, u) such that $s = 0$ is a surface tangent at the origin to $V^u_{S^j x_0}$ and $u = 0$ is a surface tangent at the origin to $V^s_{S^j x_0}$. One can choose the coordinate systems to be quite similar to each other, *i.e.* one can construct all of them by choosing the origin and a neighborhood of it out of the charts of a finite atlas for Ω (thus the "units of measure" of the lengths cannot vary wildly as j varies). Show that if δ is small enough the cones $\Gamma^{s,\alpha}_{S^j x_0}$ and $\Gamma^{u,\alpha}_{S^j x_0}$ regarded as sets of points of $\Sigma_\delta(S^j x_0)$ with coordinates (s, u) such that $|u| < \alpha|s|$ and, respectively, $|s| < \alpha|u|$, for α small enough, are mapped into $\Gamma^{s,(1-\varepsilon')\alpha}_{S^{j-1}x_0}$ and $\Gamma^{u,(1-\varepsilon')\alpha}_{S^{j+1}x_0}$ respectively by S^{-1} and S, for some $\varepsilon' > 0$. Likewise a surface in $\Sigma_{S^j x_0}$ whose tangent plane forms everywhere angles $< \alpha$ to the axis $V^s_{S^j x_0}$ is mapped into a surface with the same property in $\Sigma_\delta(S^{j-1}x_0)$ by S^{-1}. Furthermore two sequences of surfaces γ^j, η^j in $\Sigma_\delta(S^j x_0)$ with the property of being contained in the cone $\Gamma^{s,\alpha}_{S^j x_0}$ or $\Gamma^{u,\alpha}_{S^j x_0}$ will be mapped by S^{-1} or, respectively, by S into surfaces which are closer by a factor λ' to each other at every point. Conclude that there exist manifolds $W^s(S^j x_0), W^u(S^j x_0)$ in each $\Sigma_\delta(S^j x_0)$ which are tangent to $V^s_{S^j x_0}$ and $V^u_{S^j x_0}$ which verify the properties (4.2.5) and (4.2.6). (*Hint:* Repeat the analysis of problem [4.2.4] carrying along the label j).

Q4.2.6 **[4.2.6]:** Making use of problem [4.2.3] show that one can repeat the analysis in problems [4.2.4] and [4.2.5] with C not equal to one but suitably modifying (4.2.6), *i.e.* substituting

λ^n with $C\lambda^n$. (*Hint:* Check the existence of surfaces W^s, W^u near every point that satisfies (4.2.5) with S replaced by S^{n_0} and n_0 large enough.)

Q4.2.7 **[4.2.7]:** (*Existence of an adapted metric*)
Show that it is always possible to change the Riemannian metric on Ω in such a way that if (4.2.8) holds for the old metric then it holds also for the new metric but with $C = 1$ and a different λ, *i.e.* show the existence of an adapted metric.(*Hint:* Let λ_1 be such that $\lambda < \lambda_1 < 1$. Setting $v = v^s + v^u$ with $v^s \in V^s_x, v^u \in V^u_x$ let $||v|| = \sum_{n=0}^\infty \lambda_1^{-n}||\partial S^n v^s||_{V^s_{S^n x}} + \sum_{n=0}^\infty \lambda_1^{-n}||\partial S^{-n} v^u||_{V^u_{S^n x}}$ The new metric satisfies (4.2.8) with $C = 1$ but with λ_1 replacing λ. However it can fail to be smooth. To avoid this show that it is enough to limit the sum in the definition to a large N, suitably chosen (*i.e.* so that for a suitably large C' one has $\lambda_1(1 + (\lambda/\lambda_1)^N C') < \lambda_2 < 1$), to obtain (4.2.8) with λ_2 replacing λ.)

Q4.2.8 **[4.2.8]:** (*Anosov theorem*)
Show that the spaces V^s_x, V^u_x are not only continuous (as discussed in problem [4.2.1]) but they are also Hölder continuous with an exponent < 1 (arbitrarily prefixed), *i.e.* prove proposition (4.2.2). (*Hint:* Let for simplicity $d = 2$, *i.e.* consider the two dimensional case first. Since we have seen that the surfaces W^u_x, W^s_x are smooth in C^k (for any prefixed k) we only have to see how close is V^s_x to V^s_y when x, y are in the same manifold W^u_x and close enough. We study the various geometrical objects that we introduce in a coordinate system (s, u) in a sphere $\Sigma_\delta(x)$ in which $u = 0$ and $s = 0$ are surfaces tangent to V^s_x, V^u_x. Suppose that there is a sequence $y_n \to x$ with $y_n \in W^u_x$ and $d(y_n, x) = \delta_n$, *but suppose also* that the angle $\alpha(y_n)$ between the stable and unstable manifolds at y_n is such that $\alpha(y_n) - \alpha(x) > \sqrt{\delta_n}$. Then the differential of S at x and the one at y_n differ by $O(\delta_n)$: $dS_{y_n} = dS_x + O(\delta_n)$, and the vector v_n tangent to $W^s_{y_n}$ has a component along V^u_x which by assumption has size $\sqrt{\delta_n}$ at least. Therefore if we apply dS^m to v_n with $m = -\log(\sqrt{\delta_n})$ we get a vector which has lenght $O(\lambda^{-m}\sqrt{\delta_n}) \simeq 1$ times the initial length instead of a vector of length of order λ^m times the initial one: the point is that the difference between dS_{y_n} and dS_x is of order δ_n and therefore dS^m cannot rotate the vector v_n enough to overcome the expansion along the unstable direction at least not if m is not too large. A similar argument works if $\sqrt{\delta_n}$ is replaced by δ_n^α with any $\alpha < 1$: but *not* with $\alpha = 1$. Therefore the V^s_x are Hölder continuous but not necessarily Lipshitz (or smoother): and in fact counterexamples to smoothness can be constructed. This shows that V^s_x is Hölder continuous with exponent $\alpha < 1$ at x. Examining more carefully the argument one concludes that the Hölder continuity constant can be chosen $C = 1$ independently of x for any prefixed $\alpha < 1$: the price that one pays is that the smaller is C or the closer is α to 1 the closer y has to the to x so that $|\xi_y - \xi_x| < Cd(x, y)^\alpha$ holds for a component ξ_y of the tensor that determines the plane V^s_y in a chart on Ω.)

Q4.2.9 **[4.2.9]:** (*Uniqueness of the stable and unstable manifolds*)
Show that if (Ω, S) is a hyperbolic system in the sense of definition (4.2.1) there cannot exist two manifolds for $x \in \Omega$, $\widetilde W^s(x)$ and $\widetilde{\widetilde W}^s(x)$ of class C^∞ and mutually transversal, such that if $y,\ z \in \widetilde W^s(x)$ or $y,\ z \in \widetilde{\widetilde W}^s(x)$ one has $d(S^n y, S^n z) \le C\lambda^n d(y, z)$, for all $n \ge 0$. (*Hint:* Consider the following figure which describes an impossible situation if $\widetilde W \ne \widetilde{\widetilde W}$ and if one imagines to apply to it iterates of the map S.)

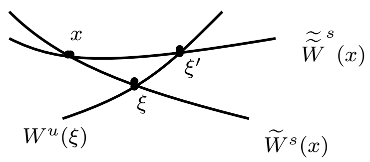


Fig.4.2.12: Illustration of the contradiction in which one would incur if the unstable manifold of x was not unique and two stable manifolds would emerge from x.

Q4.2.10 **[4.2.10]:** (*Mixing and non-transitive hyperbolic systems*)

If the connectedness condition in the definition of Anosov system is dropped then one obtains a hyperbolic system whose phase space Ω can be decomposed into a union of disjoint closed sets $\cup_{i=1}^{a}\cup_{j=1}^{d_i}\Omega_{i,j}$ and a "remainder set" Ω'. The sets $\Omega_{i,j}$ are permuted by the action of S and each of them is invariant under a suitable iterate of S which then acts in a topologically mixing way on it; the points $x\in\Omega'$ admit small neighborhoods U_x whose points evolve under the iterations of S without ever returning to visit U_x and they are called *wandering points*. If (Ω, S) is a connected Anosov system then it is topologically mixing. (*Hint:* It follows from the fact that even dropping connectedness it is still possible to construct a Markovian pavement by the construction of this section. The transition matrix will have the properties described in the problems of Section §4.1 and the result of problems [4.1.19] and [4.1.20] implies the first part. That a connected Anosov system has to be topologically mixing can be seen by noting that if $\Omega'\neq\emptyset$ then there cannot be a dense orbit; while if Ω' is empty then Ω will be the union of a finite number of disjoint sets and therefore it cannot be connected unless there is only one set $\Omega_{i,j}\equiv\Omega$ and in this case Ω will be topologically mixing.)

Q4.2.11 **[4.2.11]:** (*Splitting and hyperbolicity alone imply the Anosov property*)
Let (Ω, S) be a smooth dynamical system which verifies the properties of Anosov systems with the possible exception of the existence of a dense orbit. Show that it is necessarily an Anosov system. (*Hint:* The construction of a Markovian pavement discussed in the proof of proposition (4.2.1) can be performed and it leads to a representation of the points of Ω by symbolic strings with a certain compatibility matrix T: the result of problem [4.1.13] shows that the system can fail to be an Anosov system only if the matrix T is not mixing. However in this case the space Ω is disconnected *or* the matrix T contains inessential labels and correspondingly Ω contains wandering points.)

Q4.2.12 **[4.2.12]:** Find an explicit bound for C and a of (4.1.5) in the pavement constructed in proposition (4.2.3).

General properties of smooth hyperbolic systems discussed without the use of Markovian pavements.

Q4.2.13 **[4.2.13]:** (*Existence of a dense set of periodic orbits*)
Let (Ω, S) be an Anosov system (cf. definition (4.2.1)). The system (Ω, S) admits a dense set of periodic points. (*Hint:* Consider the case $d=2$. Let x_0 generate a dense orbit. Then choose $\overline{x}\in\Omega$ and let P be the rectangle with axes $W^s_\delta(\overline{x})\times W^u_\delta(\overline{x})$ with δ very small. Suppose, without loss of generality that $x_0\in P$. Then there will exist k_0 large such that $S^{k_0}x_0\in P$ again. The image $S^{k_0}P$ will be, if k_0 is large enough long and thin and it will cross P from "left to right" (*i.e.* from one of the two stable boundaries to the other) in a narrow connected band that contains $S^{k_0}x_0$. This band that we can call P_1 will be the image of a very narrow band "from top to bottom" of P (*i.e.* from one of the unstable boundaries of P to the other) containing x_0 and which we shall call P_0. The band P_1 is the connected part of $S^{k_0}P$ that contains $S^{k_0}x_0$ and $P_1=S^{k_0}P_0$. A much narrower band P_{-1} inside P_0 will become under iteration by S^{2k_0} a much narrower band $P_2\subset P_1$, *etc.* In this way we determine a unique point $\xi\subset P_j$, for all j. And one has $S^{k_0}\xi=\xi$: hence there is a periodic point of period k_0 inside P. The latter set was a rectangle of prefixed size of $O(\delta)$ around a prefixed point $\overline{x}$: hence periodic points are dense.)

Q4.2.14 **[4.2.14]:** (*Connectedness and Anosov systems*)
Let (Ω, S) be an Anosov system. Suppose that there is a fixed point $\overline{x}$. Show that $\cup_{i=0}^{\infty}S^iW^u_\delta(\overline{x})$ is dense in Ω. Assuming that (Ω, S) verifies all the properties in definition (4.2.1) *except* the connectedness of Ω show that connectedness follows. (*Hint:* Do not suppose that Ω is connected. Let x_0 generate a dense orbit. Suppose $\Omega\neq A\stackrel{def}{=}\overline{\cup_{i=0}^{\infty}S^iW^{u,\delta}_{\overline{x}}}$. Then there will be a point $y\notin A$ at distance $d(y,A)\stackrel{def}{=}\varepsilon>0$ from A. We can find a point $S^{k_+}x_0$ at distance $<\varepsilon/4$ from y and $k_-<k_+$ such that $S^{k_-}x_0$ is closer than $\varepsilon/4$ to the set A. And ε can be prefixed to be $\varepsilon\ll\delta$. Then the (part of) stable manifold $W^s_\delta(S^{k_-}x_0)$ will intersect the unstable manifold of a point $z\in A$. This will imply that $d(S^{k_+-k_-}z, S^{k_+}x_0)<\varepsilon$ which contradicts $d(y,A)\stackrel{def}{=}\varepsilon$. So far connectedness of Ω has

not been used. Furthermore if $\overline{x}$ is a fixed point then $\cup_{i=0}^{\infty} S^i W_\delta^u(\overline{x})$ is connected and therefore Ω is connected and any two points $z, w \in \Omega$ will be very close to points of A and therefore their stable manifolds will intersect $\cup_{i=0}^{\infty} S^i W_\delta^u(\overline{x})$ so that we can connect by a continuous curve z, w because $\cup_{i=0}^{\infty} S^i W_\delta^u(\overline{x})$ is connected.)

Q4.2.15 **[4.2.15]:** (*Density of stable and unstable manifolds for fixed points in Anosov maps*) Let (Ω, S) be an Anosov system. Let $\overline{x}$ be a fixed point. Then for all $y \in \Omega$ one has $\Omega = \overline{\cup_{i=0}^{\infty} S^i\, W_\delta^u(y)}$. (*Hint:* Let the point x_0 generate a dense orbit. The points of $W_\delta^u(y)$ will be very close to the points that are obtained by iterating a small segment of $W_\delta^u(\overline{x})$. Therefore it suffices to show that the result is true if $y \in W_{\delta'}^u(\overline{x})$ no matter how small $\delta' > 0$ is. But if $z \notin \overline{\cup_{i=0}^{\infty} S^i\, W_{\delta'}^u(y)}$ we can find an iterate of x_0 very close to y and a successive iterate very close to z and we are in a situation like the one met in the hint to problem [4.2.14]).

Q4.2.16 **[4.2.16]:** (*Density of stable and unstable manifolds for periodic points in Anosov maps*) Let (Ω, S) be an Anosov system and assume that the point x_0 generates a dense orbit. Let $\overline{x}$ be a periodic point of period $\overline{k}$. Then $\Omega = \overline{\cup_{i=0}^{\infty} S^i W_\delta^u(\overline{x})}$. (*Hint:* Same proof as for problem [4.2.14]: however we could not conclude that Ω is connected if we dropped the connectedness assumption as in problem [4.2.14].)

Q4.2.17 **[4.2.17]:** Let (Ω, S) be an Anosov system. Let $\overline{x}$ be a periodic point of minimal period $\overline{k}$. Then let $\Omega_0 = \overline{\cup_{i=0}^{\infty} S^{i\,\overline{k}}\, W_\delta^u(\overline{x})}$, $\Omega_1 = S\Omega_0, \ldots, \Omega_{\overline{k}-1} = S^{\,\overline{k}-1}\Omega_0$. Show that the sets Ω_j must coincide. Show also that if x_0 generates a dense orbit, then the iterates multiples of $\overline{k}$ of x_0 are dense as well. (*Hint:* Suppose $\overline{k} = 2$ for simplicity. Then if $S^{2i}x_0$ is dense in Ω_0 by repeating the argument in problem [4.2.15] we get that $\Omega_0 = \overline{\cup_{j=-\infty}^{\infty} S^{\,\overline{k}j}\, W_\delta^u(y)}$ if $y \in \Omega_0$. Thus if $y \in \Omega_0 \cap \Omega_1 \neq \emptyset$ one should have $\Omega_0 = \Omega_1$. Therefore we have to check that the even iterates of x_0 are dense in Ω_0 or, if not, then the odd iterates of Sx_0 are dense in Ω_0 and then replace x_0 by Sx_0. Suppose, for instance, that $\overline{x}$ can be approximated by a sequence of even iterates of x_0. Then a sequence of even iterates of x_0 can approximate any point $y \in \cup_{i=0}^{\infty} S^{i\,\overline{k}}\, W_\delta^u(\overline{x})$ or in the closure of the latter set, by the last hint to problem [4.2.15].)

Q4.2.18 **[4.2.18]:** (*Density of the stable and unstable manifolds of all points in Anosov maps*) Let (Ω, S) be an Anosov system. Show that the stable and unstable manifolds of any point, defined as $\cup_{-\infty}^{\infty} S^j W_\gamma^a(x)$, $a = u, s$, are dense in Ω.

Q4.2.19 **[4.2.19]:** (*Non-wandering points in systems with an absolutely continuous invariant measure*) Let (Ω, S) be a smooth dynamical system admitting an invariant measure μ which is absolutely continuous with respect the the volume measure: show that the set of non-wandering points, see problem [4.2.10], is the whole Ω. (*Hint:* Poincaré's recurrence theorem holds.)

Q4.2.20 **[4.2.20]:** (*Non-wandering points and density of periodic points*) Let (Ω, S) be a smooth dynamical system which verifies the properties of Anosov systems with the possible exception of the existence of a dense orbit. Show that if the set of non-wandering points is the whole Ω then the periodic points are dense. (*Hint:* The absence of wandering points allows us to imitate the argument in problem [4.2.12].)

Bibliographical note to §4.2

The proof of the existence theorem of Markovian pavements is taken from the paper of Bowen, see [Bo70], by particularizing its proof to case investigated here. The proof of Bowen widely generalizes and at the same time simplifies the original proof of Sinai, [Si68a], [Si69b]. The first idea and construction of a Markovian pavement appeared in connection with the theory of the automorphism $\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}$ of $\mathbb{T}^2$ in $\mathbb{T}^2$ by Adler and Weiss (see [AW68]): in this work one explicitly and analytically constructs some pavements. This preceded the work of Sinai where the general notion of Markovian pavement for hyperbolic maps is introduced and its existence is shown in full generality.

It is useful to consult also the monograph [Bo75]. The notion of hyperbolic system goes back to Anosov: for elementary as well as for some not elementary properties of such systems see [AS67], (where they are called U–systems), [AA68] (where they are called C–systems), and [Av76]. The abstract version of hyperbolic system is due to Ruelle, see [Ru68].

The problems are classical results, see [AS67], [KH95], and summarize Anosov's extension of the stability theory for fixed points.

§4.3 Coding of the volume measure of smooth hyperbolic systems

In this section we illustrate via an example the method to perform the operations that we have shown to be possible in Section §4.1 for systems that admit Markovian pavements, considering the case of smooth hyperbolic systems.

We shall see that the volume measure on Ω can be described by means of the symbolic code X associated with a Markovian pavement $\mathcal{Q}$.

The fundamental notions for this study are described in the following two definitions: the first of them, setting the notion of conditional probability, will also be fundamental in the coming sections.

D4.3.1 **(4.3.1) Definition:** (Conditional probability)
Let m be a Borel probability measure on the space $\{1,\ldots,q\}^{\mathbb{Z}}$ of the sequences with q symbols and let $J=\{j_1,\ldots,j_s\}\subset\mathbb{Z}$ and $\underline{\sigma}_J=(\sigma_1,\ldots,\sigma_s)\in\{1,\ldots,q\}$.

Consider the σ–algebras $\mathcal{B}(J^c)$ generated by the cylinders $C^{J'}_{\underline{\sigma}'}$, with base $J'\subset J^c=\mathbb{Z}\setminus J$. Given $\underline{\sigma}_J$ we can define on $\mathcal{B}(J^c)$ the measures

e4.3.1
$$E\to m'(E)\overset{def}{=}m(E\cap C^J_{\underline{\sigma}_J}),\qquad E\to\overline{m}(E)\overset{def}{=}m(E),\tag{4.3.1}$$

for $E\in\mathcal{B}(J^c)$. Therefore the measure m' is absolutely continuous with respect to $\overline{m}$, i.e. m' is proportional to $\overline{m}$ via a suitable function (the Radon–Nykodim derivative $dm'/d\overline{m}$ of m' with respect to $\overline{m}$). Set

e4.3.2
$$m(\underline{\sigma}_J|\underline{\sigma}'_{J^c})=\frac{dm'}{d\overline{m}}(\underline{\sigma}'),\tag{4.3.2}$$

where the notation is correct because $\underline{\sigma}_J$ is fixed and $(dm'/d\overline{m})(\underline{\sigma}')$ is $\mathcal{B}(J^c)$–measurable: hence the latter depends, m–almost everywhere, only on $\underline{\sigma}'_{J^c}$, i.e. $(dm'/d\overline{m})(\underline{\sigma}')=(dm'/d\overline{m})(\underline{\sigma}'_{J^c})$.

The expression (4.3.2) is called probability of $\underline{\sigma}_J$ conditional to $\underline{\sigma}'_{J^c}$*: if we think of it as a function of* $\underline{\sigma}'$ *it is m–measurable and* $\overline{m}$*–measurable as well.*

Remarks: (1) Note that $\overline{m}$ is the restriction of m to the sets in $\mathcal{B}(J^c)$.
(2) The notion of conditional probability is important because, as we shall see, in terms of it one can describe in a simple way large classes of nontrivial measures on $\{1,\ldots,q\}^{\mathbb{Z}}$.
(3) The functions $\underline{\sigma}_J$, $\underline{\sigma}'_{J^c} \to m(\underline{\sigma}_J|\underline{\sigma}'_{J^c})$ must verify suitable compatibility conditions implicit in their definition; for instance $\sum_{\underline{\sigma}_J} m(\underline{\sigma}_J|\underline{\sigma}'_{J^c}) = 1$.
(4) The just mentioned compatibility condition shows that in order to define the function in (4.3.2) it is sufficient, if J is fixed, to assign all ratios

e4.3.3
$$m(\underline{\sigma}'_J|\underline{\sigma}_{J^c})/m(\underline{\sigma}''_J|\underline{\sigma}_{J^c}) \quad \text{for all } \underline{\sigma}'_J, \underline{\sigma}''_J, \underline{\sigma}_{J^c}. \tag{4.3.3}$$

Coming back to Anosov systems we note that the stable and unstable manifolds are smooth (see proposition (4.2.2) and problem [4.2.8]). Therefore it makes sense to consider, at x, the map S as a map between a neighborhood of x on W^a_x and a neighborhood of Sx on W^a_{Sx}, $a = u, s$. The Jacobian matrices of these maps are linear maps, *i.e.* matrices, between V^a_x and V^a_{Sx}. We shall call the latter matrices the expanding and the contracting Jacobian matrices at x (their dimensions will be $d^a \times d^a$ if d^a is the dimension of the manifolds W^a, $a = u, s$).

Note *however* that such matrices will be smooth functions of x *along the (smooth) manifolds* $W^a_{x_0}$ of any given point x_0, but they will only be Hölder continuous as functions of x as x varies in Ω. The Hölder continuity exponent can be taken to be $\alpha < 1$ and the Hölder continuity modulus C as well can be taken to be independent of the point x if the comparison involves close enough points (how close may depend on the choice of α).

We consider for simplicity only the two-dimensional case: but most of what we say carries over to the higher dimensional cases simply by replacing the contraction and expansion coefficients by the absolute values of the determinant of the expanding and contracting Jacobian matrices. The latter are Hölder continuous functions on Ω.

D4.3.2 **(4.3.2) Definition:** (Local expansion and contraction coefficients and exponents)
Let (Ω, S) *be a bidimensional Anosov system with a Markovian pavement* $\mathcal{Q} = \{Q_1, \ldots, Q_q\}$ *built with S–rectangles. Let T be its compatibility matrix (see definition (4.1.3)) and call X the corresponding code* $X : \{1,\ldots,q\}^{\mathbb{Z}}_T \to \Omega$*. We set*

e4.3.4
$$\lambda_u^{-1}(x) = \Big|\frac{dS^{-1}}{d\xi}(x)\Big|_u, \qquad \lambda_s(x) = \Big|\frac{dS}{d\xi}(x)\Big|_s, \tag{4.3.4}$$

where the subscripts indicate the derivative of S considered as a map from the stable or unstable manifold at x into the corresponding manifold at S(x) and ξ represent the arc length parameterization of these manifolds. They *will be called the local* expansion *and* contraction coefficients*, while (the opposite of) their logarithms*

e4.3.5
$$A_u(\underline{\sigma}) = -\log \lambda_u^{-1}(X(\underline{\sigma})), \qquad A_s(\underline{\sigma}) = -\log \lambda_s(X(\underline{\sigma})), \tag{4.3.5}$$

defined on the space of the compatible sequences $\underline{\sigma} \in \{1,\ldots,q\}_T^{\mathbb{Z}}$*, will be called the local* expansion *and* contraction exponents *or* rates*.*

Remarks: (1) For all $x \in \Omega$ one has

e4.3.6
$$\Big|\frac{dS^{-M}}{d\xi}(x)\Big|_u = \prod_{k=0}^{M-1} \Big|\lambda_u(S^{-k}x)\Big|^{-1}, \qquad \Big|\frac{dS^{M}}{d\xi}(x)\Big|_s = \prod_{k=0}^{M-1} \Big|\lambda_s(S^{k}x)\Big|, \tag{4.3.6}$$

by the differentiation rules and by the covariance of the manifolds W_x^s and W_x^u, cf. (4.2.5).
(2) Note that in an adapted metric, see definition (4.2.1), $\lambda_u^{-1}, \lambda_s$ will be everywhere $< \lambda < 1$, and of course > 0. However in general it might well be otherwise: the values of the local expansion and contraction coefficients can be quite arbitrarily changed by changing the metric on Ω. Of course *eventually* for large M both quantities in (4.3.6) will become < 1.
(3) The definition of smooth hyperbolic system implies that the functions (4.3.4) and (4.3.5) are Hölder continuous in $x \in \Omega$. Since the code X is Hölder continuous as well (cf. definition (4.1.3)) we deduce

e4.3.7
$$|A_a(\underline{\sigma}) - A_a(\underline{\sigma}')| \le C\, e^{-\kappa\nu(\underline{\sigma},\underline{\sigma}')}, \qquad a = u, s, \tag{4.3.7}$$

with suitable positive C and κ, if $\nu(\underline{\sigma},\underline{\sigma}') = \max\{j | \sigma_k = \sigma_k', \text{for all } |k| \le j\}$. In other words $\underline{\sigma} \to A_a(\underline{\sigma})$ is "Hölder continuous" on the space $\{1,\ldots,q\}_T^{\mathbb{Z}}$.
It is convenient to find a representation of A_a in terms of *potentials*.

P4.3.1 **(4.3.1) Proposition:** (Contraction and expansion potentials)
Let (Ω, S) *be an Anosov system and let* $\mathcal{Q} = \{Q_1,\ldots,Q_q\}$ *be a Markovian pavement of* Ω *with compatibility matrix* $T_{\sigma\sigma'}$*. Consider the local expansion and contraction exponents* $A_a(\underline{\sigma})$ *defined by (4.3.5). Then the function*
N4.3.1 $A_a(\underline{\sigma})$ *can be written as a sum of cylindrical functions:*[1]

e4.3.8
$$A_a(\underline{\sigma}) = constant + \sum_{n=0}^{\infty} \Phi^a_{2n+1}(\sigma_{-n},\ldots,\sigma_n), \quad a = u, s, \tag{4.3.8}$$

where the functions Φ^u, Φ^s*, which will be called the* expansion *and* contraction potentials*, are defined on the* T*–compatible strings of* $2n+1$ *symbols, i.e. on* $\{1,\ldots,q\}_T^{2n+1}$*, and decay exponentially in the sense that*

e4.3.9
$$|\Phi^a_{2n+1}(\sigma_{-n},\ldots,\sigma_n)| \le \overline{C}\, e^{-\overline{\kappa} n} \tag{4.3.9}$$

[1] A function on $\{1,\ldots,q\}^{\mathbb{Z}}$ is called *cylindrical* if it depends only on the values of σ_j corresponding to a finite number of labels j.

for suitable positive constants $\overline{C}, \overline{\kappa}$.

Proof: To check (4.3.8) and (4.3.9) let $\underline{\sigma}^0 \in \{1,\ldots,q\}_T^{\mathbb{Z}}$ and let $z > 0$ be such that $T^z_{\sigma\sigma'} > 0$, for all pairs σ, σ': such z exists because of the topological transitivity and mixing properties of Anosov systems (see problems [4.1.19], [4.2.10]). We construct $\underline{\sigma}^0$, $\underline{\sigma}^1$, $\underline{\sigma}^2, \ldots \in \{1,\ldots,q\}_T^{\mathbb{Z}}$ by "chopping" the sequence $\underline{\sigma}$ at sites $\pm j$ and "attaching" to the right and to the left of it the semi-infinite strings read out of the parts of $\underline{\sigma}^0$ with positive or negative labels. This will be done so that the sequence $\underline{\sigma}^1$ has the entry with the label in the site 0 coinciding with the corresponding entry of $\underline{\sigma}$, while it coincides with $\underline{\sigma}^0$ in the sites j with $|j| > z$, $\underline{\sigma}^2$ coincides with $\underline{\sigma}$ in the sites between -1 and 1 and with $\underline{\sigma}^0$ in those with $|j| > 1 + z$, $\underline{\sigma}^3$ coincides with $\underline{\sigma}$ between -2 and 2 and with $\underline{\sigma}^0$ in the sites with $|j| > 2 + z$, *etc.*: the $2z$ insertions "between" $\underline{\sigma}$ and $\underline{\sigma}^0$ are each time made arbitrarily (and they are possible because $T^z_{\sigma\sigma'} > 0$, by our choice of z). Set *constant* $= A_a(\underline{\sigma}^0)$ and

e4.3.10
$$\Phi^a_{2n+1}(\sigma_{-n},\ldots,\sigma_n) = A_a(\underline{\sigma}^{n+1}) - A_a(\underline{\sigma}^n); \tag{4.3.10}$$

then equations (4.3.8) and (4.3.9) follow immediately from (4.3.7), *i.e.* from the Hölder continuity of the symbolic code and of the stable and unstable manifolds. ■

We can now describe the action of the code X on the (normalized) volume measure μ^0 on Ω (Sinai).

P4.3.2 **(4.3.2) Proposition:** (Symbolic code for the volume measure)
Let (Ω, S) *be a bidimensional smooth topologically mixing hyperbolic system and let* $\mathcal{Q} = \{Q_1,\ldots,Q_q\}$ *be a Markovian pavement with compatibility matrix* T *(necessarily mixing,* cf. *problem [4.1.19]); let* $X : \{1,\ldots,q\}_T^{\mathbb{Z}} \to \Omega$ *be the corresponding code. Let* μ^0 *be the volume measure on* Ω *associated with the metric on* Ω *and let* m_0 *be the measure on* $\{1,\ldots,q\}_T^{\mathbb{Z}}$ *isomorphic to it via* X *(*cf. *remark (1) following (4.1.11)).*
(i) The measure m_0 *has the following conditional probabilities:*

e4.3.11
$$\begin{aligned}\frac{m_0(\sigma'_{-n}\ldots\sigma'_n|\sigma_j, |j|>n)}{m_0(\sigma''_{-n}\ldots\sigma''_n|\sigma_j, |j|>n)} &= \frac{\sin\varphi(X(\underline{\sigma}'))}{\sin\varphi(X(\underline{\sigma}''))}\cdot \\ &\cdot \exp\Big(-\sum_{k=1}^{\infty}[A_u(\tau^k\underline{\sigma}') - A_u(\tau^k\underline{\sigma}'') + A_s(\tau^{-k}\underline{\sigma}') - A_s(\tau^{-k}\underline{\sigma}'')]\Big),\end{aligned} \tag{4.3.11}$$

where A_s, A_u *are defined in (4.3.5) and satisfy (4.3.7),* $x \to \varphi(x)$ *is the angle between the stable and unstable manifolds in* x*, and* $\underline{\sigma}'$, $\underline{\sigma}''$ *are sequences in* $\{1,\ldots,q\}_T^{\mathbb{Z}}$ *whose values on the labels* j *with* $|j| > n$ *coincide, i.e.* $\sigma'_j = \sigma''_j = \sigma_j$ *for* $|j| > n$.
(ii) The restriction m_0^+ *of* m_0 *to the* σ*–algebra* $\mathcal{B}(\mathbb{Z}^+)$*, generated by the*

cylinders with base in $\mathbb{Z}^+$, *has conditional probabilities of the form*

e4.3.12

$$\begin{aligned}\frac{m_0^+(\sigma'_0\ldots\sigma'_n|\sigma_i,j>n)}{m_0^+(\sigma''_0\ldots\sigma''_n|\sigma_j,j>n)} &= \frac{f(\sigma')}{f(\underline{\sigma}'')}\cdot\\ &\cdot\exp\Big(-\sum_{k=0}^{\infty}[\widehat{A}_u(\tau^k\underline{\sigma}')-\widehat{A}_u(\tau^k\underline{\sigma}'')]\Big),\end{aligned}\tag{4.3.12}$$

where $f>0$ *and* $\widehat{A}_u$ *are Hölder continuous functions on* $\{1,\ldots,q\}^{\mathbb{Z}}_T$.
(iii) Finally there exist $\Phi^a_{2n+1}:\{1,\ldots,q\}^{2n+1}\to\mathbb{R}$ *such that*

e4.3.13

$$A_a(\underline{\sigma})=\sum_{n=0}^{\infty}\Phi^a_{2n+1}(\sigma_{-n}\ldots\sigma_n),\qquad \widehat{A}_a(\underline{\sigma})=\sum_{n=0}^{\infty}\Phi^a_{2n+1}(\sigma_0\ldots\sigma_{2n+1}),\tag{4.3.13}$$

for $a=u,s$ *and there exist constants* $\widetilde{C},\,\widetilde{\kappa}>0$ *for which*

e4.3.14

$$|\Phi^a_{2n+1}(\sigma_{-n}\ldots\sigma_n)|\le\widetilde{C}e^{-\widetilde{\kappa}n}\qquad \textit{for all } n\ge 0.\tag{4.3.14}$$

Remarks: (1) Even though, at first sight, this may appear strange, it is convenient to think of (4.3.13) in the following form. Let $(\Phi^a_X)_{X\subset\mathbb{Z}}$ be a family of functions parameterized by the finite subsets $X\subset\mathbb{Z}$, $\Phi^a_X:\{1,\ldots,q\}^{\mathbb{Z}}\to\mathbb{R}$, such that

e4.3.15

$$\begin{aligned}&\Phi^a_X(\underline{\sigma}_X)=0 \qquad \text{unless } X=(z,z+1,\ldots,z+2n),\\ &\Phi^a_X(\underline{\sigma}_X)=\Phi^a_{2n+1}(\underline{\sigma}_X)\quad \text{if } X=(z,z+1,\ldots,z+2n),\end{aligned}\tag{4.3.15}$$

for $\underline{\sigma}_X\in\{1,\ldots,q\}^X=\{1,\ldots,q\}^{2n+1}$ and some $z\in\mathbb{Z}$. Then one has, or one defines

e4.3.16

$$\begin{aligned}A_a(\underline{\sigma})&\equiv\sum_{X\text{ centered on }0}\Phi^a_X(\underline{\sigma}_X),\qquad a=u,s,\\ \widehat{A}_a(\underline{\sigma})&\equiv\sum_{X\ni 0,X\subset\mathbb{Z}^+}\Phi^a_X(\underline{\sigma}_X),\qquad a=u,s,\\ \overline{A}_a(\underline{\sigma})&\stackrel{def}{=}\sum_{X\ni 0}\frac{\Phi^a_X(\underline{\sigma}_X)}{|X|},\qquad a=u,s.\end{aligned}\tag{4.3.16}$$

The above notations are useful for later use and for a quick check of the following algebraic identity that will be used below:

e4.3.17

$$\begin{aligned}\sum_{k=1}^{\infty}[A_a(\tau^k\underline{\sigma}')-A_a(\tau^k\underline{\sigma}'')]&=\sum_{k=0}^{\infty}[\widehat{A}_a(\tau^k\underline{\sigma}')-\widehat{A}_a(\tau^k\underline{\sigma}'')]+\\ &\qquad+W(\underline{\sigma}')-W(\underline{\sigma}''),\\ \sum_{k=-\infty}^{+\infty}[A_a(\tau^k\underline{\sigma}')-A_a(\tau^k\underline{\sigma}'')]&=\sum_{k=-\infty}^{+\infty}[\widehat{A}_a(\tau^k\underline{\sigma}')-\widehat{A}_a(\tau^k\underline{\sigma}'')]=\\ &=\sum_{k=-\infty}^{+\infty}[\,\overline{A}_a(\tau^k\underline{\sigma}')-\overline{A}_a(\tau^k\underline{\sigma}'')],\end{aligned}\tag{4.3.17}$$

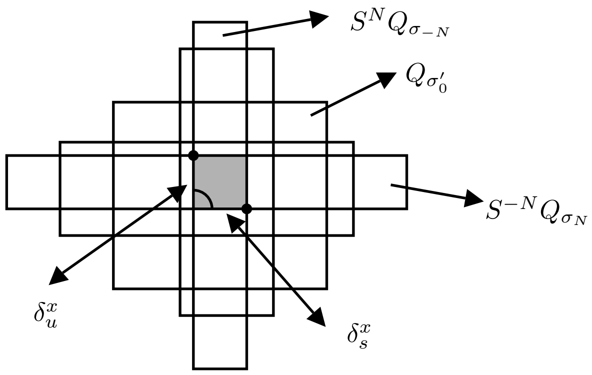
where W is a suitable Hölder continuous function on $\{1,\ldots,q\}_T^{\mathbb{Z}}$ and $\underline{\sigma}'$, $\underline{\sigma}''$ are defined after (4.3.11) (the check is left to the reader).
(2) If one follows the steps of the proof by considering the particular case of the paradigm of the Anosov systems, *i.e.* Arnold's cat map,

e4.3.18 $$\Omega=\mathbb{T}^2,\ S\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}=\begin{pmatrix}1&1\\1&2\end{pmatrix}\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix},\ \mu_0(d\varphi_1 d\varphi_2)=\frac{d\varphi_1 d\varphi_2}{(2\pi)^2},\qquad(4.3.18)$$

with $\mathcal{Q}$ an arbitrary Markovian pavement, one should realize that in this case the proof is very simple because the necessary computation of several limits becomes trivial.

Proof: We prove (4.3.11) in the case $n=0$. The proof however can be immediately adapted to cover the arbitrary $n\ge 0$ case. Let $\underline{\sigma}'$ and $\underline{\sigma}''\in\{1,\ldots,q\}_T^{\mathbb{Z}}$ with $\sigma'_j=\sigma''_j=\sigma_j$ for $|j|>0$, and set $X(\underline{\sigma}')=x$, $X(\underline{\sigma}'')=y$, assuming, furthermore, that $x,y\notin\cup_{i=1}^q\partial Q_i$. By Doob's theorem (cf. the part of the proof of proposition (3.2.1) that follows equation (3.2.25) and, in particular, equation (3.2.28)) one has, m–almost everywhere,

e4.3.19 $$\frac{m_0(\sigma'_0|\sigma_j,|j|>0)}{m_0(\sigma''_0|\sigma_j,|j|>0)}=\lim_{N\to\infty}\frac{m_0(C^{-N\ldots-1\ \ 0\ \ 1\ldots N}_{\sigma_{-N}\ldots\sigma_{-1}\sigma'_0\sigma_1\ldots\sigma_N})}{m_0(C^{-N\ldots-1\ \ 0\ \ 1\ldots N}_{\sigma_{-N}\ldots\sigma_{-1}\sigma''_0\sigma_1\ldots\sigma_N})}.\qquad(4.3.19)$$



F4.3.1 **Fig.(4.3.1)** The angle $\varphi(x')$ is marked in the dashed region (in general it is not 90^o) around the corner denoted x' in the text; the marked corners represent x'' and x''' (with the first up and the latter on the left of x'); the shadowed region represents the intersection $\cap_{-N}^N S^{-j}Q_{\sigma_j}$, with $\sigma_0=\sigma'_0$.

A schematic geometric representation of the numerator of the second member of (4.3.19) is in the drawing in Fig. (4.3.1): the shadowed rectangle is precisely $C^{-N\ldots0\ldots N}_{\sigma_{-N}\ldots\sigma'_0\ldots\sigma_N}$. Note also that such a small rectangle has sides $\le\alpha\lambda^{+N}$, if α is the largest diameter of the elements of $\mathcal{Q}$ and λ is the hyperbolicity parameter entering in the definition of smooth hyperbolic system
N4.3.2 (cf. definition (4.2.1) in §4.2.) The area of the rectangle is [2]

e4.3.20 $$|\delta^x_u|\cdot|\delta^x_s|\,\sin\varphi(x)+\ \text{higher order infinitesimals},\qquad(4.3.20)$$

[2] By higher order infinitesimal one means $o(\lambda^{2N})$, whereas the first term in (4.3.20) is $O(\lambda^{2N})$.

where, see caption to Fig. (4.3.1), δ_u^x and δ_s^x are the segments $x'x''$ and $x'x'''$, $|\delta_u^x|$ and $|\delta_s^x|$ are the corresponding lengths (in the S–adapted metric considered here), and $\varphi(x)$ is, up to higher order infinitesimals, equal to the angle $\varphi(x')$ between them (always in the considered metric). By definition of m_0 (4.3.20) is also the m_0–measure of the cylinder $C^{-N\,\ldots 0\,\ldots\,N}_{\sigma_{-N}\ldots\sigma'_0\ldots\sigma_N}$.

One repeats an analogous argument for the cylinder $C^{-N\ldots\,0\ldots\,N}_{\sigma_{-N}\ldots\sigma''_0\ldots\sigma_N}$ and, therefore, the limit (4.3.19) coincides with

e4.3.21
$$\frac{\sin\varphi(x)}{\sin\varphi(y)}\lim_{N\to\infty}\frac{|\delta_u^x|\cdot|\delta_s^x|}{|\delta_u^y|\cdot|\delta_s^y|}. \tag{4.3.21}$$

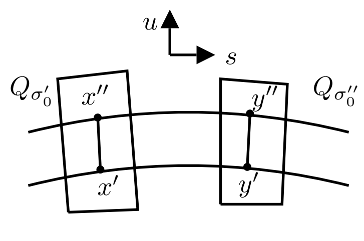
If we call x',x'' the extremes of δ_u^x and y',y'' the extremes of δ_u^y we see that x',x'' are on the same unstable boundary of $S^NQ_{\sigma_{-N}}$ and, likewise, y',y'' are on the same unstable boundary of $S^NQ_{\sigma_{-N}}$, see Fig. (4.3.2), although of course the segments δ_u^x and δ_u^x may be very far from each other as the first is inside $Q_{\sigma'_0}$ and the second inside $Q_{\sigma''_0}$ with $\sigma'_0\neq\sigma''_0$.

We can therefore write, for all k with $0\le k<N$,

e4.3.22
$$\frac{|\delta_u^x|}{|\delta_u^y|}=\frac{|S^{-k}S^k\delta_u^x|}{|S^{-k}S^k\delta_u^y|}=\frac{(\prod_{j=1}^k\lambda_u^{-1}(S^jx'))|S^k\delta_u^x|}{(\prod_{j=1}^k\lambda_u^{-1}(S^jy'))|S^k\delta_u^y|}, \tag{4.3.22}$$

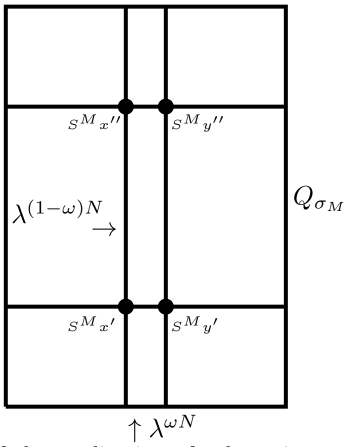
where x' and y' tend, respectively, to x and y for $N\to\infty$. We can remark that k is arbitrary, so that we conveniently choose it to be $k=\omega N$ with $\omega\in(0,1)$ to be fixed below.

Since the points S^jx' and S^jy' are on the same stable manifold they get close exponentially fast and their distance is $\le\lambda^j$, see Fig.(4.3.2).



F4.3.2 **Fig.(4.3.2)** The points x',y' and x'',y'' are, respectively, on the same stable manifolds because their symbols coincide at all sites different from 0 so that they are on the same "vertical" boundary of $S^NQ_{\sigma_{-N}}$. However they lie in different elements of the Markovian pavement (namely $Q_{\sigma'_0}$ and $Q_{\sigma''_0}$).

The Hölder continuity of $\lambda_u(x)$ therefore implies that the ratio of the products in (4.3.22) converges as $k=\omega N\to\infty$. The two arcs $S^{\omega N}\delta_u^x$ and $S^{\omega N}\delta_u^y$ become very close to each other and their distance is of the order of $\le O(\lambda^{\omega N})$ because they are on the same stable manifolds and initially they are far away at a distance of order 1. Considering that the initial length of δ_u^x and δ_u^y is $O(\lambda^N)$ we see that although their extreme points (x',x'' or y',y'') grow quite far from each other these segments stay short,



F4.3.3 **Fig.(4.3.3)** The result of the application of a large iterate $M = \omega N$ of S to the region x', x'', y', y'' in Fig. (4.3.2) if the segments $(x', x'') = \delta_u^x$ and $(y', y'') = \delta_u^y$ are short enough. The stable manifold arc x', y' becomes very short while the arcs x', x'' and y', y'' remain short because they were extremely close and, for large N, $\omega N = M \ll N$. The result is that the S^M image of the rectangle x', x'', y', y'' is small in all directions and if ω is chosen close enough to 1 (but < 1) it will look like a rectangle which is very elongated vertically, *i.e.* in the unstable direction: $d(S^M x', S^M y'), d(S^M x'', S^M y'') \ll d(S^M x', S^M x''), d(S^M y', S^M y'')$. In the figure the two dimensions of the rectangle are drawn of the same order of magnitude and the rectangle is drawn so big for illustration purposes but it should be much smaller than the enclosing rectangle Q_{σ_M}.

i.e. $O(\lambda^{(1-\omega)N})$. Therefore we can find $0 < \omega < 1$ such that for $M = \omega N$ the picture is as in Fig.(4.3.3).

The angles between the four sides of the rectangle vary quite smoothly (*i.e.* Hölder continuously), therefore the ratio between the lengths $|S^M \delta_u^x|$ and $|S^M \delta_u^y|$ gets as close to 1 as wished for $N \to \infty$: this, together with a symmetric argument to study the ration $|\delta_s^x|/|\delta_s^y|$, yields the conclusion

e4.3.23
$$\lim_{N\to\infty} \frac{\delta_u^x}{\delta_u^y} = \prod_{k=1}^{\infty} \frac{\lambda_u^{-1}(S^k x)}{\lambda_u^{-1}(S^k y)}, \qquad \lim_{N\to\infty} \frac{\delta_s^x}{\delta_s^y} = \prod_{k=1}^{\infty} \frac{\lambda_s(S^{-k} x)}{\lambda_s(S^{-k} y)}. \tag{4.3.23}$$

This proves property (i).

To show property (ii) one proceeds analogously and the details are omitted.

The statement (iii) is a consequence of (4.3.8) and (4.3.9) and of the explicit equations that are derived to obtain (4.3.12): one thus deduces naturally an expression for the f in (4.3.12) and its Hölder continuity. Details are again omitted: essentially identical arguments will be exposed in detail in Section §(5.3) in connection with a similar problem and the reader can refer to it. ∎

A corollary of the above proof, in particular of the argument leading to (4.3.23), is the following version of Fubini's theorem

C4.3.1 **(4.3.1) Corollary:** (Adapted Fubini's theorem)
Let (M, S) be a two-dimensional Anosov map. Given $x_0 \in M$ consider

the rectangle $R = W^s_\delta(x_0) \times W^u_\delta(x_0)$ *where* δ *is small enough so that the point* $[x, y]$ *is uniquely defined (*cf. *Fig.(4.2.1)). Then the volume of a subset* $E \subset R$ *is given by*

e4.3.24
$$\begin{aligned}\mu_0(E) = &\int_{W^u_\delta(x_0)} d\sigma_y \int_{W^s_\delta(x_0)} d\sigma_x \cdot \\ &\cdot \sin\alpha([y,x]) \prod_{i=1}^{\infty} \frac{\lambda_u^{-1}(S^{-i}[y,x])}{\lambda_u^{-1}(S^{-i}x)} \frac{\lambda_s(S^i[y,x])}{\lambda_s(S^i x)} \chi_E([y,x]),\end{aligned} \tag{4.3.24}$$

where $d\sigma_y, d\sigma_x$ *are the area measures (i.e. arc length) on* $W^u_\delta(x_0), W^s_\delta(x_0)$ *and* $\alpha([y,x])$ *is the angle between the stable and unstable manifolds at* $[y, x]$.

This result in fact is general and it holds in any dimension. An interesting similar result will be mentioned in Section §(6.2).

Problems for §4.3

Q4.3.1 **[4.3.1]:** (*Markovian interval maps*)
Consider a continuous map S of $[0, 1]$ into itself. Suppose that S is of class $C^{1+\varepsilon}$, $\varepsilon > 0$, in $[a_i, a_{i+1}] = Q_i$, $i = 0, \ldots, n-1$, where $a_0 = 0 < a_1 < a_2 < \ldots < a_{n-1} < a_n = 1$. Suppose that $|S'(x)| \ge \lambda > 1$, for all $x \in [a_i, a_{i+1}]$ (at the extremes one should interpret S' as right or left derivative). Finally assume that S is "Markovian": for each $i = 0, \ldots, n$ there is $j(i)$ such that $Sa_i = a_{j(i)}$. Set $T_{\sigma\sigma'} = 0$ if $S(a_0, a_{\sigma+1}) \cap (a_{\sigma'}, a_{\sigma'+1}) = \emptyset$ and $T_{\sigma\sigma'} = 1$ otherwise. Show that the code that associates with $\underline{\sigma} \in \{0, \ldots, n-1\}^{\mathbb{Z}_+}_T$ the point

$$X(\underline{\sigma}) = \mathop{\cap}_{k=0}^{\infty} S^{-k} Q_{\sigma_k}$$

is Hölder continuous in the sense that $|X(\underline{\sigma}) - X(\underline{\sigma}')| < C\ d(\underline{\sigma}, \underline{\sigma}')^\alpha$, with $C > 0$ and $\alpha = \log\lambda$.

Q4.3.2 **[4.3.2]:** Denote by $\varphi_\sigma : [Sa_\sigma, Sa_{\sigma+1}] \to [a_\sigma, a_{\sigma+1}]$ the inverse function of S on $[Sa_\sigma, Sa_{\sigma+1}]$ ("σ–th branch of the inverse S^{-1} of S"). Show that the function $\widehat{A}(\underline{\sigma}) = -\log|\varphi'_{\sigma_0}(X(\sigma_1\sigma_2\ldots))|$ is Hölder continuous on $\{0, \ldots, n-1\}^{\mathbb{Z}}_T$.

Q4.3.3 **[4.3.3]:** (*Coding of Lebesgue measure via a Markovian interval map*)
Show that the Lebesgue measure μ_0 on $[0, 1]$ is coded by X into a measure m_0 on $\{0, \ldots, n-1\}^{\mathbb{Z}_+}_T$ such that

$$\frac{m_0(\sigma'_0 \ldots \sigma'_n | \sigma_{n+1} \ldots)}{m_0(\sigma''_0 \ldots \sigma''_n | \sigma_{n+1} \ldots)} = \exp\Big(-\sum_{k=0}^{\infty} [\widehat{A}(\tau^k \underline{\sigma}') - \widehat{A}(\tau^k \underline{\sigma}'')] \Big),$$

where $\tau(\sigma_0\sigma_1\ldots) = (\sigma_1\ldots)$. Note that the sum is, in fact, finite. (*Hint:* One repeats the argument leading to (4.3.23): however in this case it becomes much simpler.)

Q4.3.4 **[4.3.4]:** (*Expansive maps and Markovian pavements*)
Give a definition of Markovian pavement for an expansive map of a compact topological metric space inspired by the previous problems and by the definition (4.1.3). Here we say that S is expansive on Ω (a metric compact space) if there exist $\lambda < 1$ and $\varepsilon > 0$ such that $d(x, y) \le \lambda \cdot d(Sx, Sy)$, for every pair x, y such that $d(x, y) < \varepsilon$; and furthermore the equation $Sy = y'$ has only one solution y such that $d(x, y) < \varepsilon/2$ for every x such that $d(Sx, y') < \varepsilon/2$.

Q4.3.5 **[4.3.5]:** (*Interval maps and invariant absolutely continuous measures*)
Show, in the context of problem [4.3.3], that the condition for the existence of an S–invariant measure $\overline{\mu} = \overline{h}\mu_0$ absolutely continuous with respect to μ_0 is that there exists

a solution to the equation in $L_1(m_0)$

$$h(\sigma_0\sigma_1\sigma_2\ldots) = \sum_{\sigma'=0}^{n-1} e^{-\widehat{A}(\sigma'\sigma_0\sigma_1\ldots)} h(\sigma'\sigma_0\sigma_1\ldots)$$

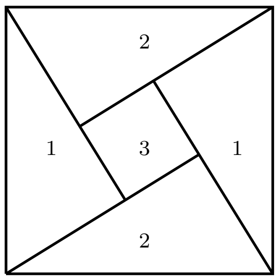
m_0–a.e, if we set $h(\underline{\sigma}) = \overline{h}(X(\underline{\sigma}))$. (*Hint:* Write the condition $\overline{\mu}(E) = \overline{\mu}(S^{-1}E)$, that is

$$\overline{h}(x) = \sum_{\sigma'=0}^{n-1} |\varphi'_{\sigma'}(x)|\,\overline{h}(\varphi_{\sigma'}(x)) \qquad x \in L_1(\mu_0)$$

in the symbolic variables $\underline{\sigma}$.)

Q4.3.6 **[4.3.6]:** (*A non-generating Markovian pavement for the square root of Arnold's cat map*)

Consider the map of $\mathbb{T}^2$ into itself defined by the matrix $S = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ and let $\underline{v}^{\pm}$ be the two eigenvectors of S with components $(1, \lambda_+ - 1)$ and $(1, \lambda_- - 1)$ where $\lambda_\pm = (3 \pm \sqrt{5})/2$ are the eigenvalues of S. Construct the pavement formed by three rectangles with sides parallel to $\underline{v}^{\pm}$ as in Fig. (4.3.4).



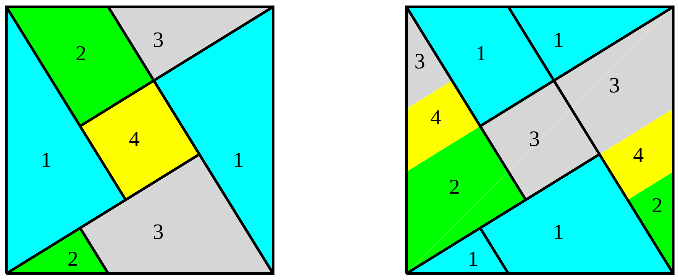
F4.3.4 **Fig.(4.3.4)** The pavement of problem [4.3.6] with three rectangles whose sides lie on two *connected* portions of stable and unstable manifold of the fixed point at the origin.

Check that it verifies the property (4.1.7). Show that nevertheless *it is not* a Markovian pavement according to definition (4.1.3) (see remark (5) after that definition). Show that the same pavement is also not Markovian for the map of $\mathbb{T}^2$ generated by the matrix $S_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and note that $S_0^2 \equiv S$. Compute the transition matrix for the latter map showing that it is

$$T = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

and that $T^5_{ij} > 0$ for all $i, j = 1, 2, 3$. (*Hint:* The property (4.1.7) follows simply because the unions of the boundaries consisting of stable or of unstable manifolds of the origin are connected and therefore invariant under S or S^{-1} respectively. It is not Markovian because the sets in (4.1.3) do not consist of a single point, so that the property (i) in definition (4.1.3) is not fulfilled. The image of the rectangle labeled 1 (for instance) crosses the rectangle labeled 2 in two sets with disconnected interiors.)
grestore

Q4.3.7 **[4.3.7]:** (*A generating Markovian pavement of the square root of Arnold's cat map*)
Consider the partition in Fig.(4.3.5) and show that it is a Markovian partition for the



F4.3.5 **Fig.(4.3.5)** A Markovian pavement (left) for the square root S_0 of Arnold's cat map. It is obtained from the partition in Fig. (4.3.4) by continuing a little further the stable manifold of the origin breaking into two parts the rectangle labeled 2 (whose large size was responsible for the non-Markovian nature of the pavement in Fig. (4.3.4)). The images under S_0 of the pavement rectangles is shown in the right figure: corresponding rectangle are marked by the same colors.

square root S_0 of Arnold's cat map with transition matrix

$$T = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{matrix},$$

and check that $T^5_{\sigma\sigma'} > 0$ for all $\sigma, \sigma' = 1, 2, 3, 4$.

Q4.3.8 **[4.3.8]:** (*An example of a hyperbolic algebraic map on* $\mathbb{T}^3$)
The compatibility matrix of a Markovian pavement may define a hyperbolic algebraic map on a torus of dimension equal to the number of elements of the pavement. Check that this is the case for the matrix T in problem [4.3.6]. Check also that this is not the case for the matrix T, which is the compatibility matrix of a generating pavement, in problem [4.3.7]. (*Hint:* The characteristic equation for the eigenvalues of the matrix T of problem [4.3.6] is $\lambda^3 = \lambda + 1$ and the eigenvector corresponding to the largest eigenvalue λ (*spiral mean*) is $(1, \lambda, \lambda^{-1})$. This is a vector with rationally independent components, see problems [8.1.2], [8.1.4] and [8.1.4] below.)

Q4.3.9 **[4.3.9]:** (*A simple construction of Markovian pavements for two-dimensional Anosov systems*)
Show that the construction of Markovian pavements in two dimensional Anosov systems admitting a fixed point can be easily obtained by generalizing the construction in problem [4.3.6], *i.e.* by drawing a connected part of the stable and unstable manifolds of the fixed point and letting them "go around" until they form a net whose elements have a diameter smaller than a prefixed δ (using the density of the stable and unstable manifolds, see problem [4.2.18]) and stopping the drawing of the stable manifold when its extremes cross the unstable manifold and viceversa, as done in the illustration in Fig.(4.3.4).

Q4.3.10 **[4.3.10]:** By problem [4.2.18]) also the stable and unstable manifolds of a periodic point are dense. Furthermore Anosov systems admit a dense set of periodic points, as problem [4.2.13] shows. Show that this implies that the construction in problem [4.3.9] can be estended to Anosov systems which have no fixed point. (*Hint:* Let x_0 be a periodic point for S with period p. Then x_0 is a fixed point for the Anosov map S^p and the stable and unstable manifolds for S^p and for S are the same. Then we construct a Markovian pavement $\mathcal{P}_0$ with the method of problem [4.3.9] and, by the argument discussed in the proof of proposition (4.2.1), at item (C), $\cap_{j=0}^{p-1} S^j \mathcal{P}_0$ is a Markovian pavement for S.)

Bibliographical note to §4.3

The idea of using Markovian pavements to code symbolically topological measures associated with an algebraic hyperbolic map of the torus is due to Adler and Weiss, [AW68], and Sinai, [Si68a], [Si68b], [Si72], who proved and applied various versions of proposition (4.3.2). Another interesting application to the theory of systems endowed of an attractor that verifies the axiom A of Smale is in [Ru76]. Markovian maps of the interval have been studied by several authors, see for early contributions [Ru77], [PY79] and [CE80a].

CHAPTER V

Gibbs distributions

§5.1 Gibbs distributions

The study of the general structure of dynamical systems, begun in the previous sections, could continue and constitutes one of the directions in which ergodic theory can be developed. We shall, however, look in a somewhat different direction dedicating attention to a few concrete problems that do not belong to the general theory. The more concrete studies involve analytic work of "classical" type and are more directly related to the applications.

In the previous sections, for instance in proposition (4.3.2), we faced for the first time, in implicit form, the following problem: given a probability distribution on $\{1,\ldots,q\}_T^{\mathbb{Z}}$ described by its family of conditional probabilities, under which conditions do the latter determine the probability distribution uniquely?

If the conditional probabilities are assigned in a form similar to (4.3.11) or (4.3.12) with the functions A_a having the form (4.3.13) or (4.3.16) the problem is known as the "determination of a Gibbs distribution from its potential".

To proceed orderly it is convenient to set up a general definition and an appropriately suggestive nomenclature, in spite of a few repetitions of notions and definitions already discussed previously in different contexts.

Let T be a $(n+1)\times(n+1)$ compatibility matrix, with entries equal to 0 or 1, and let $\Omega=\{0,\ldots,n\}_T^{\mathbb{Z}}=\{\underline{\sigma}\,|\,\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}},\ \prod_{-\infty}^{+\infty}T_{\sigma_i\sigma_{i+1}}=1\}$ be the space of the T–compatible sequences, cf. definition (4.1.1).

We shall say that the cylinder $C^J_{\underline{\sigma}}$ with base J and specification $\underline{\sigma}$ ($J=$

$\{j_1,\ldots,j_q\} \subset \mathbb{Z}$, $\underline{\sigma} \in \{0,\ldots,n\}^J$) is a T–compatible cylinder if $C^J_{\underline{\sigma}} \cap \{0,\ldots,n\}^{\mathbb{Z}}_T \neq \emptyset$.

If $\Lambda \subset \mathbb{Z}$ is a set, finite or not, $\mathcal{B}(\Lambda)$ will be the σ–algebra generated by the cylinders with base in Λ. If Λ is finite with $|\Lambda|$ elements then $\mathcal{B}(\Lambda)$ is a finite σ–algebra with at most $|\Lambda|^{n+1}$ atoms.

The matrix T is said *mixing*, cf. (4.1.1), if there exists an integer $z \geq 0$ such that $T^{z+1}_{\sigma\sigma'} > 0$ for all $\sigma,\sigma' \in \{0,\ldots,n\}$: the minimum value of such z will be denoted $a(T)$ and it will be called *mixing time* (or *mixing length*) of T.

If T is mixing the dynamical system $(\{0,\ldots,n\}^{\mathbb{Z}}_T, \tau)$ will be topologically
N5.1.1 mixing.[1]
N5.1.2 If m is a probability distribution on Ω and if $\Lambda \subset \mathbb{Z}$ is a finite set,[2] we can define, for every T–compatible cylinder $C^\Lambda_{\underline{\sigma}_\Lambda}$ with base Λ, the probability distribution on $\mathcal{B}(\Lambda^c)$

e5.1.1
$$m'(E) = m(C^\Lambda_{\underline{\sigma}_\Lambda} \cap E), \qquad E \in \mathcal{B}(\Lambda^c). \tag{5.1.1}$$

Such a probability distribution is obviously absolutely continuous with respect to the restriction of m to $\mathcal{B}(\Lambda^c)$, which we shall still denote with m, and the Radon–Nykodim derivative of m' with respect to m will be a $(m,\mathcal{B}(\Lambda^c))$–measurable function:

e5.1.2
$$\underline{\sigma}' \to m(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) = \frac{dm'}{dm}(\underline{\sigma}'), \tag{5.1.2}$$

where the notation is admissible because $\frac{dm'}{dm}(\underline{\sigma}')$, being $(m,\mathcal{B}(\Lambda^c))$–measurable, depends on $\underline{\sigma}'$ only via the restriction to Λ^c of $\underline{\sigma}'$: $\underline{\sigma}'_{\Lambda^c} = (\sigma'_j)_{j\in\Lambda^c}$.

The (5.1.2) is the *probability of the event* $\underline{\sigma}_\Lambda$ *in* Λ *conditional to the event* $\underline{\sigma}'_{\Lambda^c}$ *in* Λ^c.

Extending a convention, employed so far, we shall consider pairs J_1 and $J_2 \subset \mathbb{Z}$, $J_1 \cap J_2 = \emptyset$, and if $\underline{\sigma}_{J_1} \in \{0,\ldots,n\}^{J_1}$ and $\underline{\sigma}_{J_2} \in \{0,\ldots,n\}^{J_2}$ then $\underline{\sigma}_{J_1}\underline{\sigma}_{J_2}$ will denote the element $\underline{\sigma}'_{J_1\cup J_2} \in \{0,\ldots,n\}^{J_1\cup J_2}$ such that $\sigma'_i = \sigma_i$, for all $i \in J_1 \cup J_2$.

The following general definition of *Gibbs distribution* or *Gibbs measure* or *Gibbs state* is obviously inspired by (4.3.11), (4.3.12) and (4.3.16); making use of the notions and notations previously introduced we set the following definition.

D5.1.1 **(5.1.1) Definition:** (Potentials for symbolic dynamics)
Let T be a mixing compatibility matrix for the sequences in $\{0,\ldots,n\}^{\mathbb{Z}}$ and let $a(T)$ be the mixing time of T, cf. *definition (4.1.1). Let B the space*

---

[1] Given two open sets F and G there exists N_0 such that $S^N F \cap G \neq \emptyset$ for all $N \geq N_0$.

[2] We shall say that Λ is an interval or, equivalently, that it is connected if Λ is given by the intersection of $\mathbb{Z}$ with an interval of $\mathbb{R}$.

of the sequences $\Phi = \{\Phi_X\}_{X\subset\mathbb{Z}}$ *parameterized by the finite subsets of* $\mathbb{Z}$, *consisting in the functions*

e5.1.3
$$\Phi_X : \{0,\ldots,n\}^X \to \mathbb{R} \tag{5.1.3}$$

such that, having set $\|\Phi_X\| = \max_{\underline{\sigma}\in\{0,\ldots,n\}^X} |\Phi_X(\underline{\sigma})|$, *one has*
(i) (Shift invariance) Φ *is invariant under translations of* $\mathbb{Z}$, *i.e. for all* $\sigma_1,\ldots,\sigma_p$

e5.1.4
$$\Phi_X(\sigma_1,\ldots,\sigma_p) = \Phi_{\tau X}(\sigma_1,\ldots,\sigma_p), \tag{5.1.4}$$

if $X = (\xi_1,\ldots,\xi_p)$ *and* $\tau X = (\xi_1+1,\ldots,\xi_p+1)$.
(ii) (Stability) Φ *is "summable":*

e5.1.5
$$\|\Phi\| \equiv \sum_{X\ni 0} \frac{\|\Phi_X\|}{|X|} < +\infty. \tag{5.1.5}$$

We shall say that B *is the space of the* potentials *on* $\{0,\ldots,n\}^{\mathbb{Z}}_T = \Omega$. *The function on* $\{0,\ldots,n\}^{\mathbb{Z}}_T$ (cf. *the third of (4.3.16)) defined by*

e5.1.6
$$A_\Phi(\underline{\sigma}) = \sum_{X\ni 0} \frac{\Phi_X(\underline{\sigma}_X)}{|X|} \tag{5.1.6}$$

will be called potential energy per site *or* energy function *associated with* Φ.

Remark: In classical potential theory the energy of a configuration of points on $\mathbb{Z}$ interacting via a potential $\Phi(x,y)$ is written as

e5.1.7
$$\sum_{(x,y)\in\mathbb{Z}^2} \sigma_x\sigma_y\Phi(x,y) = \sum_{x\in\mathbb{Z}}\Big(\sum_{\substack{y\in\mathbb{Z}\\ y\neq x}} \frac{\sigma_x\sigma_y\Phi(x,y)}{2}\Big), \tag{5.1.7}$$

where $\sigma_x = 0$ if the site x is empty and $\sigma_x = 1$ if it is occupied. This explains the name given to (5.1.6) and why Φ is called a *many-body potential* or, more properly, a collection of many-body potentials. In the case of (5.1.7) one can imagine that $\Phi_X = 0$ unless $X = \{x,y\}$ and $\Phi_{\{x,y\}}(\sigma,\sigma') = \sigma\sigma'\Phi(x,y)$: this is the case in which the potential is a *two-body potential*.

In terms of potentials it is possible to give a general enough notion of Gibbs distribution.

D5.1.2 **(5.1.2) Definition:** (Gibbs distributions, DLR equations)
In the context of definition (5.1.1) we shall say that $m \in M^0(\{0,\ldots,n\}^{\mathbb{Z}}_T)$ *is a Gibbs distribution on* $\{0,\ldots,n\}^{\mathbb{Z}}_T$ *with potential* $\Phi\in B$ *if it is a probability distribution on the Borel sets of* $\{0,\ldots,n\}^{\mathbb{Z}}_T$ *whose conditional probabilities*
N5.1.3 $m(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda_c})$ *are,* m*–almost everywhere, such that* [3]

e5.1.8
$$\frac{m(\underline{\sigma}'_\Lambda|\underline{\sigma}_{\Lambda^c})}{m(\underline{\sigma}''_\Lambda|\underline{\sigma}_{\Lambda^c})} = \exp\Big(-\sum_{k=-\infty}^{+\infty}\{A_\Phi(\tau^k\underline{\sigma}') - A_\Phi(\tau^k\underline{\sigma}'')\}\Big) \tag{5.1.8}$$

[3] One refers to (5.1.8) by calling it *Dobrushin–Lanford–Ruelle relations* or simply *DLR relations*; see also remark (5) after this definition.

for every interval $\Lambda\subset\mathbb{Z}$ *longer than* $a(T)$ *and for* $\underline{\sigma}'=(\underline{\sigma}'_\Lambda\underline{\sigma}_{\Lambda^c})$, $\underline{\sigma}''=$
N5.1.4 $(\underline{\sigma}''_\Lambda\underline{\sigma}_{\Lambda^c})\in\{0,\ldots,n\}^{\mathbb{Z}}_T$.[4]

Equivalently: m is a Gibbs distribution with potential Φ *if for every long enough interval* Λ *one has, m–almost everywhere,*

e5.1.9
$$m(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})=\frac{\exp\big(-\sum_{R\cap\Lambda\ne\emptyset}\Phi_R(\underline{\sigma}_R)\big)}{\sum_{\underline{\sigma}'_\Lambda}\exp\big(-\sum_{R\cap\Lambda\ne\emptyset}\Phi_R(\underline{\sigma}'_R)\big)},\tag{5.1.9}$$

where (5.1.9) are interpreted as zero if $\underline{\sigma}\notin\{0,\ldots,n\}^{\mathbb{Z}}_T$.

The set of the Gibbs distributions with potential Φ *will be denoted* $G^0(\Phi)$, $G(\Phi)\subset G^0(\Phi)$ *will be the set of the Gibbs distributions on* $\{0,\ldots,n\}^{\mathbb{Z}}_T$
N5.1.5 *which are invariant under translations,*[5] $G_e(\Phi)\subset G(\Phi)$ *will be the set of the ergodic Gibbs distributions with potential* Φ*, etc.*

Remarks: (1) It is necessary to associate with this definition an existence theorem. Indeed it is by no means clear that Gibbs distributions exist, as it is not obvious that translation invariance of the potential implies translation invariance of the relative Gibbs distributions. In fact shift invariance is not in general a consequence of the shift invariance of the potential. We shall not meet such "pathologies" in what follows because the potentials we consider will have further properties which allow to exclude them.

(2) Equivalence between (5.1.8) and (5.1.9) follows from (5.1.6) and from the observation that

e5.1.10
$$\sum_{k=-\infty}^{\infty}\{A_\Phi(\tau^k\underline{\sigma}')-A_\Phi(\tau^k\underline{\sigma}'')\}=\sum_{R\cap\Lambda\ne\emptyset}\{\Phi_R(\underline{\sigma}'_R)-\Phi_R(\underline{\sigma}''_R)\},\tag{5.1.10}$$

that is obtained from the definitions. Furthermore in (5.1.10) the sum can be decomposed, as shown by the right hand side, into 2 absolutely convergent sums. This shows immediately that $m(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})$ must be proportional to $\exp(-\sum_{R\cap\Lambda\ne\emptyset}\Phi_R(\underline{\sigma}_R))$. The denominator in (5.1.9) is precisely the normalization coefficient determined by the condition $\sum_{\sigma_\Lambda}m(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})=1$, which holds because $m(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})$ is a conditional probability.

What said so far is correct for every connected $\Lambda\subset\mathbb{Z}$ provided all the elements of T are positive. In the general case, if T is only mixing, it is necessary to consider in the previous argument only T–compatible sequences. This does not present particular difficulties provided Λ is a large enough interval (at least longer than $a(T)$). If Λ is too short and T has many zeroes it could happen that the denominator of (5.1.9) is zero for some $\underline{\sigma}$ because

[4] More precisely: chosen $\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}_T$ m–almost everywhere and chosen $\underline{\sigma}'_\Lambda$, $\underline{\sigma}''_\Lambda\in\{0,\ldots,n\}^{\Lambda}_T$ then (5.1.8) holds provided Λ is connected and of length larger than $a(T)$, where $a(T)$ is the mixing time of T. The latter condition is needed to make sure that the denominator in (5.1.8) or (5.1.9) does not vanish.

[5] We call a distribution m on $\{0,\ldots,n\}^{\mathbb{Z}}_T$ *invariant under translation* if $\tau m=m$ where the action of τ on the probability distributions is obvious: we set $(\tau m)(E)=m(\tau^{-1}E)$, for all $E\in\mathcal{B}$.

there might be no possibility to match the sequence $\underline{\sigma}_{\Lambda^c}$ outside Λ with at least one sequence $\underline{\sigma}_\Lambda$ internal to Λ so that $\underline{\sigma}_\Lambda\underline{\sigma}_{\Lambda^c}$ is T–compatible. In the latter case the definition becomes more involved as one has to appeal to the fact that sequences $\underline{\sigma}$ with probability 1 will be compatible: we simply state as a part of the definition that Λ be connected and long enough to avoid having to deal with the problem, as it is not necessary to do so. Of course once the conditional probabilities relative to an interval Λ are known the ones relative to shorter intervals $\Lambda' \subset \Lambda$ are also determined so that requiring property (5.1.9) for long intervals L is not restrictive.
(3) The reader familiar with the theory of Markov processes will recognize without difficulty that the case $\Phi_X = 0$ if $|X| > 2$ or if $X = \{x, y\}$ is not a pair of nearest neighbors corresponds to the case of a mixing Markov process: hence *Gibbs processes*, *i.e.* Gibbs distributions, constitute a (non-trivial) generalization of Markov processes.
(4) We shall denote the expression (5.1.9) with the symbol $p_\Phi(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})$ and its denominator will be called simply the "normalization".
(5) The properties (5.1.8) and the equivalent (5.1.9) are called *DLR equations*. Here they are taken as defining properties of Gibbs distributions. However in other approaches the Gibbs distributions are defined in a different way and the DLR relations become theorems. See also Section §(6.1).

Before discussing some properties of Gibbs distributions it is convenient to set the following definition.

D5.1.3 **(5.1.3) Definition:** (Bulk and surface energies)
In the context of the definitions (5.1.1) and (5.1.2), let $\Phi \in B$, $\Omega \in \{0,\ldots,n\}^{\mathbb{Z}}_T$, $\Lambda \subset \mathbb{Z}$, $|\Lambda| < +\infty$, *and set*

e5.1.11
$$\begin{aligned} &\textit{(i)}\quad U^0_\Lambda(\underline{\sigma}) = \sum_{R\subset\Lambda} \Phi_R(\underline{\sigma}_R),\\ &\textit{(ii)}\quad U_\Lambda(\underline{\sigma}) = \sum_{R\cap\Lambda\neq\emptyset} \Phi_R(\underline{\sigma}_R),\\ &\textit{(iii)}\quad E_\Lambda(\underline{\sigma}) = \sum_{j\in\Lambda} A_\Phi(\tau^j\underline{\sigma}),\end{aligned} \tag{5.1.11}$$

which we call, respectively, the energy of $\underline{\sigma}_\Lambda$ in Λ, *the* energy of $\underline{\sigma}_\Lambda$ in Λ with boundary condition $\underline{\sigma}_{\Lambda^c}$ outside Λ, *and the* contribution of Λ to the energy of the configuration $\underline{\sigma}$. *Furthermore we set*

e5.1.12
$$U_\Lambda(\underline{\sigma}) = E_\Lambda(\underline{\sigma}) + D_{\Lambda,1}(\underline{\sigma}), \qquad U^0_\Lambda(\underline{\sigma}) = E_\Lambda(\underline{\sigma}) + D_{\Lambda,2}(\underline{\sigma}), \tag{5.1.12}$$

and, finally,

e5.1.13
$$\varepsilon(\Lambda) = \sup_{\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}_T} \Big(|D_{\Lambda,1}(\underline{\sigma})| + |D_{\Lambda,2}(\underline{\sigma})|\Big), \tag{5.1.13}$$

e5.1.14
$$\varepsilon_{N,a} = \varepsilon([a, a+1, a+2, \ldots, a+N-1]), \tag{5.1.14}$$

and $D_{\Lambda,1}(\underline{\sigma}), D_{\Lambda,2}(\underline{\sigma})$ *will be called* surface corrections *or* surface terms.

Remarks: (1) The names come from the form and the interpretation of equations (5.1.11)÷(5.1.14) in the case of the theory of the potential discussed in the remark to definition (5.1.1): the reader should write the various quantities just defined in that case.
(2) Note that $|U^0_\Lambda(\underline{\sigma})|$, $|U_\Lambda(\underline{\sigma})|$, $|E_\Lambda(\underline{\sigma})| \le ||\Phi||\,|\Lambda|$.
(3) With the new notations the r.h.s. of (5.1.9) is written as

e5.1.15
$$p_\Phi(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c}) = \frac{\exp\big(-U_\Lambda(\underline{\sigma})\big)}{\text{normalization}}, \tag{5.1.15}$$

and it is an expression that *makes sense without referring to any distribution* m. It will be called the "probability of the configuration $\underline{\sigma}_\Lambda$ in presence of the external configuration $\underline{\sigma}_{\Lambda^c}$" and we shall say that it is "proportional to the exponential of the opposite of its energy". In Statistical Mechanics the probability distributions, on the space of the configurations of a system, which associate with a configuration a probability proportional to the exponential of $-U$, if U is the energy of the configuration, are called *Boltzmann–Gibbs distributions* and have great importance for Physics.
(4) An immediate consequence of the finiteness of the norm $\|\Phi\|$ is that

e5.1.16
$$\lim_{N\to\infty} \varepsilon_{N,a}/N = 0, \tag{5.1.16}$$

that explains the name of "surface corrections" used for $D_{\Lambda,1}(\underline{\sigma}), D_{\Lambda,2}(\underline{\sigma})$. The reader should check that if $B^0 \subset B$ is the space of the *finite range potentials*, *i.e.* of the potentials $\Phi \in B$ with $\Phi_X = 0$ except for a finite number of sets X, then

e5.1.17
$$\varepsilon_{N,a} \le \|\Phi\| R_\Phi, \tag{5.1.17}$$

where R_Φ is defined by

e5.1.18
$$R_\Phi = \max_{X,\Phi_X\ne 0} \text{diam}(X), \tag{5.1.18}$$

and is called the *range* of Φ.

We conclude this section by proving the following proposition.

P5.1.1 **(5.1.1) Proposition:** (Existence of Gibbs states)
If T *is a mixing compatibility matrix with labels* $0,\ldots,n$ *and if* Φ *is a potential on* $\{0,\ldots,n\}^{\mathbb{Z}}$ *then the sets* $G^0(\Phi)$ *and* $G(\Phi)$ *are not empty: therefore there exists at least one translation invariant Gibbs distribution with potential* $\Phi \in B$, cf. *definition (5.1.1).*

Proof: To check the proposition we begin with a general remark on probability distributions on spaces of sequences: as implied by the definition of conditional probability a function $\underline{\sigma}' \to m(\underline{\sigma}_\Lambda|\underline{\sigma}'_{\Lambda^c})$ is the conditional probability of the event $C^\Lambda_{\underline{\sigma}_\Lambda}$ with respect to the σ–algebra $\mathcal{B}(\Lambda^c)$ and to the probability distribution m if and only if, given arbitrarily two functions $\underline{\sigma} \to f(\underline{\sigma}_\Lambda)$ and $\underline{\sigma} \to g(\underline{\sigma}_{\Lambda^c})$ continuous and, respectively, $\mathcal{B}(\Lambda)$ and $\mathcal{B}(\Lambda^c)$–measurable, the following *Fubini's conditional integration* holds

e5.1.19
$$\int f g dm \equiv \sum_{\underline{\sigma}_\Lambda} \int f(\underline{\sigma}_\Lambda) g(\underline{\sigma}'_{\Lambda^c}) m(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) m(d\underline{\sigma}'_{\Lambda^c}), \tag{5.1.19}$$

where, for ease of notations, we do not use a different symbol for the probability distribution m on $\mathcal{B}$ and for its restriction to the σ–algebra $\mathcal{B}(\Lambda^c)$.

Note, furthermore, that $\Phi \in B$ (so that $||\Phi|| < \infty$, see definition (5.1.1)) implies that $\underline{\sigma} \to p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c})$, defined by the (5.1.15) with Λ large enough, is a continuous function of the variable $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}_T$.

We first show that $G^0(\Phi)$ is a convex and compact (possibly empty) set if it is regarded as a subset of the space of the probability distributions $\mathcal{M}^0(\{0, \ldots, n\}^{\mathbb{Z}}_T)$ thought of, as usual, as a topological space with the weak topology induced on it by the continuous functions on $\{0, \ldots, n\}^{\mathbb{Z}}_T$.

Indeed (5.1.19) written for m_1 and m_2 in $G^0(\Phi)$, *i.e.* with $m(\underline{\sigma}_\Lambda \,|\, \underline{\sigma}_{\Lambda_c}) = p_\Phi(\underline{\sigma}_\Lambda \,|\, \underline{\sigma}_{\Lambda_c})$ in both cases, implies that also $\alpha m_1 + (1-\alpha) m_2$, $\alpha \in (0,1)$, is in $G^0(\Phi)$ (as we see by performing the suitable linear combination of (5.1.19) written for m_1 and m_2). Furthermore, if we consider a sequence of probability distributions $m_k \in G^0(\Phi)$ weakly converging to m and we write (5.1.19) for m_k we see that, due to the continuity of f, g and p_Φ as functions of $\underline{\sigma}$, also $m \in G^0(\Phi)$, *i.e.* that $G^0(\Phi)$ is closed (hence compact).[6]
N5.1.6

Another preliminary observation is that if $\Phi^k \in B$, $k = 1, 2, \ldots$ is a sequence of potentials in B which converges in the norm (5.1.5) to $\Phi \in B$ and if $m_k \in G^0(\Phi^k)$, $k = 1, \ldots$, is a corresponding sequence of Gibbs distributions converging to a probability distribution m, then $m \in G^0(\Phi)$ (*continuity of Gibbs states as functions of their potential*). In fact the convergence of Φ^k to a Φ in B implies, as it is immediate to see, that for every $\Lambda = \{-N, \ldots, N\}$, $N > a(T)$, the limit

e5.1.20
$$\lim_{k\to\infty} p_{\Phi^k}(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c}) = p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c}) \tag{5.1.20}$$

takes place uniformly in $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}_T$. Therefore, due to the uniformity in $\underline{\sigma}$ of the limit (5.1.20), equation (5.1.19) written for m_k implies that $m \in G^0(\Phi)$ (by taking the limit $k \to \infty$).

From the latter considerations and from the metric compactness of the space $M^0(\{0, \ldots, n\}^{\mathbb{Z}}_T)$ of the probability distributions on $\{0, \ldots, n\}^{\mathbb{Z}}_T$ it follows that it will suffice to show that $G^0(\Phi) \neq \emptyset$, for all $\Phi \in B^0$, where $B^0 \subset B$ is an arbitrary subset of B which is dense in B in the norm (5.1.5).

---

[6] We always consider the weak topology on the space of probability distributions, *i.e.* two probability distributions are close to each other if the integrals that they attribute to a finite large enough family of continuous functions are close to each other, see footnote 1, Section §2.3. In this topology the space of the probability distributions on a compact separable space is compact.

This will also show that $G(\Phi)\neq\emptyset$: indeed if $G^0(\Phi)\neq\emptyset$ we shall be able to consider, given $m_0\in G^0(\Phi)$, the average over translations

e5.1.21
$$m_0^{(N)}(E)=N^{-1}\sum_{k=0}^{N-1} m_0(\tau^k E),\qquad E\in\mathcal{B}(\{0,\ldots,n\}_T^{\mathbb{Z}}),\tag{5.1.21}$$

which is a convex combination of elements in $G^0(\Phi)$ because if $m\in G^0(\Phi)$ also $E\to m_0(\tau E)$ is a probability distribution in $G^0(\Phi)$ (as follows from the translation invariance of Φ and p_Φ and from (5.1.19)).

Every limit point of the sequence $m_0^{(N)}$ defined in (5.1.21) is in $G^0(\Phi)$, since $G^0(\Phi)$ is closed, and furthermore it is obviously τ–invariant, *i.e.* it is in $G(\Phi)$, which, therefore, is not empty.

It remains to show that $G^0(\Phi)\neq\emptyset$ for Φ in a dense set $B^0\subset B$: it is natural to select the set B^0 to be the set of the finite range potentials (cf. remark (4) to definition (5.1.3)). If $\Phi\in B^0$ we denote by R_Φ its range, as in (5.1.18), and define a sequence of "approximate" Gibbs distributions via the integrals that they assign to continuous functions.

Let $\widehat{\underline{\sigma}}$ be a configuration arbitrarily chosen in $\{0,\ldots,n\}_T^{\mathbb{Z}}$ (which will be used to define a boundary condition) and let $\Lambda_k=[-k,k]$, $k>a(T)$. We define

e5.1.22
$$\int f(\underline{\sigma})m_{k,\widehat{\underline{\sigma}}}(d\underline{\sigma})\stackrel{def}{=}\sum_{\underline{\sigma}'_{\Lambda_k}} f(\underline{\sigma}'_{\Lambda_k}\widehat{\underline{\sigma}}_{\Lambda_k^c})\frac{e^{-U_{\Lambda_k}(\underline{\sigma}'_{\Lambda_k}\widehat{\underline{\sigma}}_{\Lambda_k^c})}}{\sum_{\underline{\sigma}''_{\Lambda_k}} e^{-U_{\Lambda_k}(\underline{\sigma}''_{\Lambda_k}\widehat{\sigma}_{\Lambda_k^c})}},\tag{5.1.22}$$

where the two sums are over the configurations $\underline{\sigma}'_{\Lambda_k}$ and $\underline{\sigma}''_{\Lambda_k}\in\{0,\ldots,n\}^{\Lambda_k}$ compatible with $\widehat{\underline{\sigma}}$, *i.e.* such that $\underline{\sigma}'_{\Lambda_k}\widehat{\underline{\sigma}}_{\Lambda_k^c}$ as well as $\underline{\sigma}''_{\Lambda_k}\widehat{\sigma}_{\Lambda_k^c}$ are sequences in $\{0,\ldots,n\}_T^{\mathbb{Z}}$. The sums in the denominator involve at least one positive element because $k>a(T)$. It should be noted that $m_{k,\widehat{\underline{\sigma}}}$ only depends on $\widehat{\underline{\sigma}}$ via $\widehat{\underline{\sigma}}_{\Lambda_k^c}$: in other words fixing $\widehat{\underline{\sigma}}$ actually is a convenient way to fix a sequence of boundary conditions on the intervals Λ_k.

We can compute the conditional probability $m_k(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})$ of the probability distributions m_k just defined, finding

e5.1.23
$$m_k(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})=p_\Phi(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})\qquad\text{for all }\underline{\sigma}\in\{0,\ldots,n\}_T^{\mathbb{Z}}\tag{5.1.23}$$

for every interval $\Lambda=[a,b]$ such that $|\Lambda|>a(T)$ and Λ is "well inside" Λ_k, *i.e.* $\Lambda^{R_\Phi}=[a-R_\Phi,b+R_\Phi]\subset\Lambda_k$.

Therefore m_k fails to verify (5.1.9), *i.e.* fails to be in $G^0(\Phi)$, "only" because its validity is restricted to Λ well inside Λ_k. Hence it is natural to take the limit $k\to\infty$: let $m_{k_i}\to m$ be a convergent subsequence extracted from the sequence $\{m_k\}_{k\in\mathbb{N}}$.

Imagine to write (5.1.19) for cylindrical functions f and g (*i.e.* for functions dependent on $\underline{\sigma}$ via the values of σ_j corresponding to a finite number of $j\in\mathbb{Z}$); then the k–independence of the r.h.s. of (5.1.23) (for $\Lambda_k\supset\Lambda$)

implies that the limit m, as $k \to \infty$ of the selected subsequence verifies (5.1.19) for all pairs of cylindrical functions f, g.

By the density of the cylindrical functions in the continuous functions and by (5.1.23), (5.1.19) extends to all pairs of continuous functions so that $m \in G^0(\Phi)$ and $G^0(\Phi)$, hence $G(\Phi)$ as well, is not empty. ∎

In fact the method of the above proof gives us also an approximation procedure for the actual construction of an element of $G^0(\Phi)$.

C5.1.1 **(5.1.1) Corollary:** (Continuity of Gibbs states with respect to the potential)
Under the hypotheses of proposition (5.1.1) one has:
(i) $G^0(\Phi)$ and $G(\Phi)$ are not empty, convex, compact and τ–invariant.
(ii) If $\Phi^{(k)} \xrightarrow[k\to\infty]{} \Phi$ is a convergent sequence of potentials in B (convergence in the norm (5.1.5) of B) and if $m_k \in G^0(\Phi^{(k)})$ and $m_k \xrightarrow[k\to\infty]{} m$, then $m \in G^0(\Phi)$; symbolically we shall write

e5.1.24
$$\lim_{\Phi\to\overline{\Phi}} G^0(\Phi) \subset G^0(\overline{\Phi}). \tag{5.1.24}$$

(iii) If $\Phi^{(k)} \xrightarrow[k\to\infty]{} \Phi$ in B and if $m_k \in G(\Phi^{(k)})$ and $m_k \xrightarrow[k\to\infty]{} m$, then $m \in G(\Phi)$; symbolically

e5.1.25
$$\lim_{\Phi\to\overline{\Phi}} G(\Phi) \subset G(\overline{\Phi}). \tag{5.1.25}$$

Proof: (i) and (ii) have been proved within the proof of proposition (5.1.1) and (iii) is a consequence of (ii). ∎

Problems for §5.1

Q5.1.1 **[5.1.1]:** (*Bernoulli shifts and Gibbs states*)
Study the Gibbs state on $\{0,1\}^{\mathbb{Z}}$ with potential $\Phi_X = 0$ if $|X| \neq 1$ and $\Phi_{\{x\}}(\sigma_x) = h(\sigma_x)$ and prove that it is a Bernoulli scheme.

Q5.1.2 **[5.1.2]:** (*A Markov process*)
Study the Gibbs state on $\{0,1,2\}^{\mathbb{Z}}_T$ with $\Phi = 0$ and $T = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ and prove that is a Markov process. Compute its transition probabilities in terms of the spectral properties of the matrix T.

Q5.1.3 **[5.1.3]:** (*Gibbs distributions on higher dimensional lattices*)
Generalize the notion of Gibbs distribution to the case of the space of d-dimensional sequences $\{0,1\}^{\mathbb{Z}^d}$, without compatibility conditions, replacing the translation τ with the group of the translations of $\mathbb{Z}^d$ so that the statements analogous to proposition (5.1.1) and to corollary (5.1.1) remain valid. (*Hint:* Just suppose that Λ is a finite cubic region and require that (5.1.15) gives the conditional probabilities of configurations in Λ when outside of Λ there is a fixed configuration.)

Q5.1.4 **[5.1.4]:** (*Gibbs states and Markov processes in $\mathbb{Z}^d$*)
Generalize the notion of Markov process to the space of the sequences $\{0,1\}^{\mathbb{Z}^d}$. (*Hint:* See problem [5.1.3] and suppose that the range of Φ is 1, *i.e.* identify Markov processes with nearest neighbour interaction Gibbs states.).

Q5.1.5 **[5.1.5]:** (*Averintzev–Spitzer theorem*)

Consider the most general probability distribution on $\{0,1\}^{\mathbb{Z}^2}$ with conditional probability $m(\sigma_0|\underline{\sigma}') = p(\sigma_0|\sigma'_1\sigma'_2\sigma'_3\sigma'_4)$ where $1,2,3,4$ are the 4 nearest neighbour sites of 0 and $\sigma_0\underline{\sigma}' \in \{0,1\}^{\mathbb{Z}^2}$. Suppose, furthermore, that m is invariant under rotations (of $\pi/2$) and that $p(.|\ldots) > 0$. Show that m is a Gibbs state with a potential Φ which can be chosen so that $\Phi_X = 0$ if $|X| > 2$, $\Phi_{\{i,j\}}(\sigma_i,\sigma_j) = 0$ if $|i-j| > 1$, $\Phi_{\{i,j\}}(\sigma_i,\sigma_j) = J\sigma_i\sigma_j$ if $|i-j| = 1$ and $\Phi_{\{0\}}(\sigma_0) = h\sigma_0$.

Q5.1.6 **[5.1.6]:** (*An equivalence condition for potentials*)

If B is the space of potentials on $\{0,\ldots,n\}^{\mathbb{Z}}_T$, show that if Φ and $\Psi \in B$ and if there exists $C \in \mathbb{R}$ and a continuous function F over $\{0,\ldots,n\}^{\mathbb{Z}}_T$ such that $A_\Phi(\underline{\sigma}) = C + A_\Psi(\underline{\sigma}) + F(\underline{\sigma}) - F(\tau\underline{\sigma})$, then $G(\Phi) = G(\Psi)$. (*Hint:* Use (5.1.8)).

Q5.1.7 **[5.1.7]:** (*Potentials and Hölder continuous functions*)

If $A \in C(\{0,\ldots,n\}^{\mathbb{Z}}_T)$ is a Hölder continuous function and if T is a mixing compatibility matrix then there exist $\Phi \in B$ such that $A = A_\Phi$. (*Hint:* See proposition (4.3.2) and equation (4.3.13).)

Q5.1.8 **[5.1.8]:** (*Particle potentials*)

Let B^0 be the space of finite range potentials. Under the hypotheses of problem [5.1.6] with $n = 1$, $T_{\sigma\sigma'} > 0$, show that if $\Phi \in B^0$ there always exists $\Psi \in B^0$ such that $G(\Phi) = G(\Psi)$ and $\Psi_X(\sigma_X) = 0$ unless $\sigma_\xi \neq 0$, for some $\xi \in X$. We say that Ψ is a potential equivalent to Φ and with $\underline{\sigma} = \underline{0}$ as *vacuum configuration*. (*Hint:* Define Ψ recursively keeping in mind the equivalence condition in problem [5.1.6].)

Q5.1.9 **[5.1.9]:** (*Existence of particle potentials*)

In the context of the problem [5.1.8] find some sufficient conditions in order that $\Phi \notin B^0$ admits an equivalent potential Ψ (*i.e.* such that $G(\Phi) = G(\Psi)$) with the property of the Ψ in problem [5.1.8]. (*Hint:* The recursion suggested in the hint to problem [5.1.8] leads to an expression for Ψ in terms of sums over values of Φ over various configurations. One just makes sure that the sums involved are absolutely convergent, including the sum that provides an estimate for the norm $||\Psi||$.)

Q5.1.10 **[5.1.10]:** (*Open boundary conditions*)

Assume that the compatibility matrix T has no vanishing entries. Replace in the r.h.s. of (5.1.22) $U_{\Lambda_k}(\underline{\sigma}_{\Lambda_k}\widehat{\underline{\sigma}}_{\Lambda_k^c})$ with $U^0_{\Lambda_k}(\underline{\sigma}_{\Lambda_k})$ and $f(\underline{\sigma}_{\Lambda_k}\underline{\sigma}_{\Lambda_k^c})$ with $f(\underline{\sigma}_{\Lambda_k}\underline{0})$ where $\underline{0}$ denotes the sequence of symbols identically 0. This is equivalent to replacing m_k with

$$\widetilde{m}_k = \sum_{\underline{\sigma}_{\Lambda_k}} \frac{\exp\Big(-U^0_{\Lambda_k}(\underline{\sigma}_{\Lambda_k})\Big)}{\sum_{\underline{\sigma}'_{\Lambda_k}} \exp\Big(-U^0_{\Lambda_k}(\underline{\sigma}'_{\Lambda_k})\Big)}\,\delta_{(\underline{\sigma}_{\Lambda_k}\underline{0})},$$

where $\delta(\underline{\sigma}_{\Lambda_k}\underline{0})$ is the Dirac probability distribution concentrated on the configuration that inside Λ_k coincides with $\underline{\sigma}_{\Lambda_k}$ and outside of Λ_k is identically zero. The distribution $\widetilde{m}_k$ is called a *finite volume Gibbs distribution with open boundary conditions* when $\underline{0}$ is the "vacuum" for the potential Φ in the sense of problem [5.1.8]. Show that every limit point of the sequence $\widetilde{m}_k$ is in $G^0(\Phi)$ if $\Phi \in B^0$, where B^0 is space of finite range potentials. (*Hint:* Repeat word by word the proof of proposition (5.1.1).)

Q5.1.11 **[5.1.11]:** Same as problem [5.1.10] but replacing B^0 with the set B (in which B^0 is dense).

Q5.1.12 **[5.1.12]:** Adapt definitions and results of Section §5.1 and the corresponding problems to the case in which the matrix T is just transitive (rather than mixing).

Bibliographical note to §5.1 and §(5.2)

The notion of Gibbs distribution represents an interesting product of the close interaction that took place between Mathematical Physics and Theoretical Physics in the 1960's. It was due to two main reasons. On the one hand a large number of scientists with a basic formation and research activity experience in High Energy Physics became interested in the mathematical problems connected with their previous works. This was done within the framework of a general rethinking on the foundations of a theory that seemed to undergo a deep methodology crisis (field theory, *i.e.* relativistic quantum mechanics, after the failure of its naive application to the theory of strong interactions). On the other hand the need by several condensed matter physicists to refine the theoretical prediction instruments of statistical mechanics in order to interpret the experimental results on phase transitions (that were produced in great abundance thanks to the substantial progress of the experimental techniques). The interest into rigorous results developed mainly for the purposes of having reliable terms of comparison to check the reliability of hitherto uncontrolled approximations needed to solve delicate theoretical problems like the theoretical computation of the critical exponents (made possible, to a previously unimaginable extent, by the progress of electronic computational machines)

The notion of Gibbs distribution was developed independently in the West (thanks mainly to the works of Ruelle, Fisher, Griffiths, Lanford, *etc.*) and in the East (thanks mainly to the works of Dobrushin, Minlos, Sinai, etc.).

The more or less definitive formulation of the notion and of the basic properties of a Gibbs distribution can be found in the classical papers [Do68a], [Do68b], [Do68c], [Do69], [Ru69] and [LR69].

§5.2 Properties of Gibbs distributions

For a better understanding of the nature and properties of Gibbs distributions we shall discuss an important uniqueness criterion and some of its simple consequences.

P5.2.1 **(5.2.1) Proposition:** (Uniqueness of Gibbs distributions)
Let T a mixing compatibility matrix with labels $0,\ldots,n$ and with mixing time $a(T)$. Let $\Phi\in B$ be a potential (see definition (5.1.1)) on $\{0,\ldots,n\}_T^{\mathbb{Z}}$ and suppose

e5.2.1
$$\|\Phi\|_1=\sum_{X\ni 0}\frac{(1+\mathrm{diam}(X))}{|X|}\,\|\Phi_X\|<+\infty, \tag{5.2.1}$$

(i) Every T–compatible cylinder has positive m–measure for all probability distributions $m\in G(\Phi)$ (hence every set of m–probability 1 is dense in $\{0,\ldots,n\}_T^{\mathbb{Z}}$).
(ii) $G^0(\Phi)$ contains a unique element m.

Remarks: (1) Condition (5.2.1) says that "the interaction energy between the left half of a configuration $\underline{\sigma}$ and the other half is finite, uniformly in

$\underline{\sigma}$": indeed the interaction energy of the left half of the configuration $\underline{\sigma}$ with the right half is naturally defined by

e5.2.2
$$\sum_{R\cap\mathbb{Z}_+\neq\emptyset,\, R\cap\mathbb{Z}_-\neq\emptyset}\Phi_R(\underline{\sigma}_R)=W(\underline{\sigma}^-,\underline{\sigma}^+),\tag{5.2.2}$$

having performed the division of $\underline{\sigma}$ in the left half $\underline{\sigma}^-\equiv(\sigma_j)_{j<0}$ and the right half $\underline{\sigma}^+=(\sigma_j)_{j\geq 0}$. Then (5.2.2) is bounded above by (5.2.1) as

e5.2.3
$$|W(\underline{\sigma}^-,\underline{\sigma}^+)|\leq\|\Phi\|_1;\tag{5.2.3}$$

furthermore there exists potentials $\Phi\in B$ and sequences $\underline{\sigma}\in\{0,\ldots,n\}^{\mathbb{Z}}_T$ for which equality sign holds.
(2) Technically (5.2.1) or (5.2.3) will be employed to compare conditional probabilities. For example if $\underline{\sigma}'=(\ldots\widetilde{\sigma}_{a-1}\sigma'_a\ldots\sigma'_b\widetilde{\sigma}_{b+1}\ldots)$ and $\underline{\sigma}''=(\ldots\widetilde{\sigma}_{a-1}\sigma''_a\ldots\sigma''_b\widetilde{\sigma}_{b+1}\ldots)$ are in $\{0,\ldots,n\}^{\mathbb{Z}}_T$ and if we set $\Lambda=[a,b]$, one has

e5.2.4
$$\begin{aligned}\frac{p_\Phi(\sigma'_a\ldots\sigma'_b|\ldots\widetilde{\sigma}_{a-1}\widetilde{\sigma}_{b+1}\ldots)}{p_\Phi(\sigma''_a\ldots\sigma''_b|\ldots\widetilde{\sigma}_{a-1}\widetilde{\sigma}_{b+1}\ldots)}&=e^{-\sum_{R\cap\Lambda\neq\emptyset}(\Phi_R(\underline{\sigma}'_R)-\Phi_R(\underline{\sigma}''_R))}\leq\\&\leq e^{4\|\Phi\|_1}e^{-\sum_{R\subset\Lambda}(\Phi_R(\underline{\sigma}'_R)-\Phi_R(\underline{\sigma}''_R))},\end{aligned}\tag{5.2.4}$$

as it follows immediately from (5.2.1) and from the translation invariance of Φ.
Likewise if $(\ldots\widetilde{\sigma}_{a-1}\sigma_a\ldots\sigma_b\widetilde{\sigma}_{b+1})$ and $(\ldots\widehat{\sigma}_{a-1}\sigma_a\ldots\sigma_b\widehat{\sigma}_{b+1}\ldots)$ are in $\{0,\ldots,n\}^{\mathbb{Z}}_T$ one has, if $b-a>a(T)$,

e5.2.5
$$\frac{p_\Phi(\sigma_a\ldots\sigma_b|\ldots\widetilde{\sigma}_{a-1}\widetilde{\sigma}_{b+1}\ldots)}{p_\Phi(\sigma_a\ldots\sigma_b|\ldots\widehat{\sigma}_{a-1}\widehat{\sigma}_{b+1}\ldots)}\leq e^{8\|\Phi\|_1}(n+1)^{a(T)},\tag{5.2.5}$$

as we see starting from the explicit expression for p_Φ, see (5.1.9) and (5.1.15), and paying attention to the normalization factor.
(3) Note that (5.2.4) and (5.2.5) also imply lower bounds with $\|\Phi\|_1$ replaced by $-\|\Phi\|_1$ and $(n+1)^{a(T)}$ by $(n+1)^{-a(T)}$, because of the arbitrariness of $\underline{\sigma}'$, $\underline{\sigma}''$, $\underline{\widetilde{\sigma}}$, $\underline{\widehat{\sigma}}$.

Proof: In the proof of the second statement we shall suppose, for simplicity, $a(T)=0$, *i.e.* $T_{\sigma\sigma'}\equiv 1$.
Note that if m, $m_1\in G(\Phi)$ then m is absolutely continuous with respect to m_1, and viceversa, with Radon–Nykodim derivative between $\exp(-8\|\Phi\|_1)$ and $\exp(8\|\Phi\|_1)$. Indeed one has

e5.2.6
$$\begin{aligned}m(C^{a\ldots b}_{\sigma_a\ldots\sigma_b})&\equiv\int m(C^{a\ldots b}_{\sigma_a\ldots\sigma_b})\,m_1(d\underline{\widetilde{\sigma}})=\\&=\int m_1(d\underline{\widetilde{\sigma}})\int m(d\underline{\widehat{\sigma}})p_\Phi(\sigma_a\ldots\sigma_b|\ldots\widehat{\sigma}_{a-1}\widehat{\sigma}_{b+1}\ldots)\leq\\&\leq e^{8\|\Phi\|_1}\int m_1(d\underline{\widetilde{\sigma}})\int m(d\underline{\widehat{\sigma}})p_\Phi(\sigma_a\ldots\sigma_b|\ldots\widetilde{\sigma}_{a-1}\widetilde{\sigma}_{b+1}\ldots)\equiv\\&\equiv e^{8\|\Phi\|_1}\int m_1(d\underline{\widetilde{\sigma}})p_\Phi(\sigma_a\ldots\sigma_b|\ldots\widetilde{\sigma}_{a-1}\widetilde{\sigma}_{b+1}\ldots)\equiv e^{8\|\Phi\|_1}m_1(C^{a\ldots b}_{\sigma_a\ldots\sigma_b}),\end{aligned}\tag{5.2.6}$$

having used inequality (5.2.5) in the third step.

The arbitrariness of a, b and $\sigma_a \ldots \sigma_b$ allow us to deduce from (5.2.6) that for all $E \in \mathcal{B}$

e5.2.7 $$m(E) \le e^{8\|\Phi\|_1} m_1(E), \tag{5.2.7}$$

that shows that m is absolutely continuous with respect to any probability distribution m_1 with Radon–Nykodim derivative f that, by the symmetric role of m and m_1, is in $L_\infty(m_1)$ and

e5.2.8 $$e^{-8\|\Phi\|_1} \le f(\underline{\sigma}) \le e^{8\|\Phi\|_1} \qquad m_1 - \text{almost everywhere.} \tag{5.2.8}$$

One has $m = f\, m_1$ and, at the same time, m and m_1 have the same conditional probability $p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c})$, for all $\Lambda = [a, b]$: this implies that for every $\Lambda = [a, b]$ and every $\underline{\sigma}^1$, $\underline{\sigma}^2 \in \{0, \ldots, n\}^\Lambda$ one has

e5.2.9 $$\frac{f(\underline{\sigma}^1_\Lambda \underline{\sigma}_{\Lambda^c})}{f(\underline{\sigma}^2_\Lambda \underline{\sigma}_{\Lambda^c})} \equiv 1 \qquad m - \text{almost everywhere in } \underline{\sigma}_{\Lambda^c}, \tag{5.2.9}$$

N5.2.1 which from a formal point of view is an obvious relation.[1]

If we set for $k \ge 0$

e5.2.10 $$r(k) \equiv \int f(\underline{\sigma})^k\, m_1(d\underline{\sigma}), \tag{5.2.10}$$

it follows, also, that $r(k)^{-1} f^k m_1$ is (for every $k \in \mathbb{Z}$) a probability distribution of $G^0(\Phi)$, since (5.2.9) remains valid if we replace f with the powers of f itself. By (5.2.8) applied by selecting $m' = r(k)^{-1} f^k m_1$ instead of m one has

e5.2.11 $$e^{-8\|\Phi\|_1} \le r(k)^{-1} f^k(\underline{\sigma}) \le e^{8\|\Phi\|_1} \qquad m_1 - \text{almost everywhere,} \tag{5.2.11}$$

for all $k \in \mathbb{Z}$.

The relation (5.2.11) and the arbitrariness of k imply that f is constant m_1–almost everywhere, *i.e.* $m = m_1$. Hence $G^0(\Phi)$ consists of a single point and, therefore, $G^0(\Phi) = G(\Phi)$. This proves the statement (ii).

To show that $m(C^{a\,\cdots\,b}_{\sigma_a\ldots\sigma_b}) > 0$ if $C^{a\,\cdots\,b}_{\sigma_a\ldots\sigma_b}$ is T–compatible we treat the general case, since the case $T_{\sigma\sigma'} = 1$ is easy but too special. Having set $J = [a - a(T), b + a(T)]$, let $\overline{\underline{\sigma}}_J$ be a string such that $m(C^J_{\overline{\underline{\sigma}}_J}) > 0$ (which certainly exists because otherwise m itself would vanish). Then

e5.2.12 $$m(C^J_{\overline{\underline{\sigma}}_J}) = \int p_\Phi(\overline{\underline{\sigma}}_J | \underline{\sigma}'_{J^c}) m(d\underline{\sigma}') > 0, \tag{5.2.12}$$

and, hence, there exists a set D of configurations $\underline{\sigma}'$ such that $m(D) > \varepsilon$ and $p_\Phi(\overline{\underline{\sigma}}_J | \underline{\sigma}'_{J^c}) > \varepsilon'$, for $\underline{\sigma}' \in D$, for some ε, $\varepsilon' > 0$. One can therefore construct a configuration $\widehat{\underline{\sigma}}_J$, T–compatible, that coincides with $\overline{\underline{\sigma}}_J$ on the

[1] A rigorous check of (5.2.9) passes through Doob's theorem already cited in Section §3.2 after (3.2.25): we leave this to the reader although it is somewhat subtle.

extreme sites $a-a(T)$ and $b+b(T)$ and with $\sigma_a,\ldots,\sigma_b$ in the sites in $[a,b]$: it is clear that $\widehat{\underline\sigma}_J$ is T–compatible with $\underline\sigma'_{J^c}$ for all $\underline\sigma'$ in D. Hence (5.2.4), the mentioned properties of $m(D)$ and the strict positivity of $p_\Phi(\underline\sigma_J|\underline\sigma'_{J^c})$ whenever $\underline\sigma_J$ and $\underline\sigma'_{J^c}$ are T–compatible immediately imply that also $m(C^J_{\widehat{\underline\sigma}_J})>0$. Since the cylinder $C^{a\ \cdots\ b}_{\sigma_a\ldots\sigma_b}$ that we are considering contains, by construction, $C^J_{\underline\sigma_J}$ the result follows.

Note the great simplification of the above proof in the case $a(T)=0$, *i.e.* in the case in which all sequences are compatible. ■

C5.2.1 **(5.2.1) Corollary:** (Ergodicity and mixing of Gibbs states)
Under the hypotheses of proposition (5.2.1) one has $G^0(\Phi)=G(\Phi)=G_e(\Phi)=G_m(\Phi)$.

Remark: Hence if $\|\Phi\|_1<+\infty$ the corresponding Gibbs distribution m is ergodic, and in fact mixing.

Proof: Let m be an arbitrary probability distribution on $\{0,\ldots,n\}^{\mathbb Z}_T$, possibly not in $G(\Phi)$. Let $\mathcal B(\infty)_m$ be the m–complete σ–algebra generated by the functions of $L_1(m)$ that are $\mathcal B([-N,N]^c)$–measurable for all $N>0$.

The latter functions are often called *functions measurable at infinity* because their values do not change by changing any finite number of labels in their argument $\underline\sigma$. They form an algebra that is called the *algebra at infinity of the probability distribution m*.

N5.2.2 If $\mathcal B([-N,N]^c)_m$ is the completion[2] with respect to m of $\mathcal B([-N,N]^c)$ one has, by definition

e5.2.13
$$\mathcal B(\infty)_m=\bigcap_{N>0}\mathcal B([-N,N]^c)_m. \tag{5.2.13}$$

If $f>0$ is a $\mathcal B(\infty)_m$–measurable function and if it is m–summable with integral 1, then the probability distribution $m_1=fm$ has the same conditional probabilities, *i.e.* $m(\underline\sigma_\Lambda|\underline\sigma_{\Lambda^c})$, of the probability distribution m. Indeed such probabilities would be in general given by the left hand side expression in the relation

e5.2.14
$$\frac{f(\underline\sigma'_\Lambda\underline\sigma_{\Lambda^c})m(\underline\sigma'_\Lambda|\underline\sigma_{\Lambda^c})}{f(\underline\sigma''_\Lambda\underline\sigma_{\Lambda^c})m(\underline\sigma''_\Lambda|\underline\sigma_{\Lambda^c})}=\frac{m(\underline\sigma'_\Lambda|\underline\sigma_{\Lambda^c})}{m(\underline\sigma''_\Lambda|\underline\sigma_{\Lambda^c})}, \tag{5.2.14}$$

where the equality takes place because, if f is $\mathcal B(\infty)$–measurable, $f(\underline\sigma'_\Lambda\underline\sigma_{\Lambda^c})=f(\underline\sigma''\underline\sigma_{\Lambda^c})$ since f must assume the same value on configurations that differ only in a finite number of sites.

It follows that if $m\in G(\Phi)$ one must have that $\mathcal B(\infty)_m$ is a trivial σ–algebra: if indeed nonconstant $\mathcal B(\infty)_m$–measurable functions f existed then there would exist positive ones among them and bounded away from 0 and $+\infty$ (*e.g.* if f was such a function one could take $(1+|f(\cdot)|)/(2+|f(\cdot)|)$). It could therefore be possible, by multiplying m by any of the latter, to construct a Gibbs distribution fm different from m itself, against the uniqueness shown in proposition (5.2.1)

[2] See appendix 1.2.

This implies that $m \in G_m(\Phi)$, *i.e.* that m is mixing. Indeed let $f,\, g \in C(\{0,\ldots,\}^{\mathbb{Z}}_{T})$; consider the sequence of functions $\underline{\sigma} \to g(\tau^k\underline{\sigma})$, $k = 0,1,\ldots$: by the compactness of the spheres of $L_\infty(m)$ with respect to the weak topology induced by $L_1(m)$ one can extract from the sequence a subsequence $g(\tau^{k_i}\underline{\sigma})$ converging, always in the weak topology, as $i \to \infty$ to a limit $\widetilde{g}(\underline{\sigma}) \in L_\infty(m)$ such that $\|\widetilde{g}\|_{L_\infty(m)} \le \max_{\underline{\sigma}} |g(\underline{\sigma})|$.
It is however clear that $\widetilde{g}$ is $\mathcal{B}(\infty)$–measurable because

e5.2.15
$$\lim_{k\to\infty} |g(\tau^k\underline{\sigma}') - g(\tau^k\underline{\sigma}'')| = 0, \tag{5.2.15}$$

if $\underline{\sigma}'$ and $\underline{\sigma}''$ differ only on a finite number of labels (recall that g is assumed continuous). It follows that $\widetilde{g}$ is constant m–almost everywhere, hence

e5.2.16
$$\widetilde{g} = \lim_{i\to\infty} m(1\cdot\tau^{k_i} g) \equiv m(g), \tag{5.2.16}$$

i.e. $\widetilde{g}$ does not depend on the choice of the subsequence. Therefore we deduce that $g(\tau^k\cdot) \xrightarrow[k\to+\infty]{} m(g)$ in the $L_1(m)$–topology of $L_\infty(m)$. In other words, and in particular,

e5.2.17
$$\lim_{k\to\infty} m(f\tau^k\, g) = m(f)\, m(g) \tag{5.2.17}$$

for every pair of continuous functions f and g and, hence, by density it follows that m is mixing. ■

Remarks: (1) We have also shown that $\mathcal{B}(\infty)_m$ (and hence also $\mathcal{B}(-\infty)_m = \cap_{N>0}\mathcal{B}([-\infty,-N])_m \subseteq \mathcal{B}(\infty)_m$) is a trivial σ–algebra in the sense that all functions measurable with respect to it are necessarily constant. The latter property is, in general, stronger than mixing: the invertible dynamical systems $(\{0,\ldots,n\}^{\mathbb{Z}},\tau,m)$ such that $\mathcal{B}(\infty)_m$ is trivial are called *systems with trivial algebra at infinity*, while those for which "already" $\mathcal{B}(-\infty)_m$ is trivial are called *K–systems* or *systems with trivial remote past.*
(2) The notion of K–system is naturally generalizable to invertible metric dynamical systems (Ω,S,μ): let $\mathcal{P} = \{P_0,\ldots,P_n\}$ be a nontrivial partition of Ω into μ–measurable sets (*i.e.* $n \ge 1$ and $\mu(P_i) > 0$ for all i) and let $(\{0,\ldots,n\}^{\mathbb{Z}},\tau,m)$ be the symbolic dynamical system associated with the $(S,\mathcal{P})$–histories by (2.3.19) and proposition (2.3.1); *we shall say that* (Ω,S,μ) *is a K–system* if $\mathcal{B}(-\infty)_m$ is trivial for all choices of $\mathcal{P}$.
(3) An interesting property, among many, of K–systems is that one can show (Sinai) that the triviality of $\mathcal{B}(-\infty)_m$ is equivalent to the requirement that, for all nontrivial μ–measurable partitions, one has $s(\mathcal{P},S,\mu) > 0$, cf. definition (3.3.2). We shall not enter into the discussion of the properties of K–systems.

Problems for §5.2

Q5.2.1 **[5.2.1]:** (*Fisher potential*)
Consider the potential on $\{0,1\}^{\mathbb{Z}}$ defined as follows:

(i) $\Phi_X(\underline{\sigma}_X) = 0$ unless $X = [a, b]$, $b \ge a$, and $\underline{\sigma}_X = (1, 1, \ldots, 1)$,
(ii) if $X = \{a, a+1, a+2, \ldots, a+n\}$ and $\sigma_a = \sigma_{a+1} = \ldots = \sigma_{a+n} = 1$

$$\Phi_X(\underline{\sigma}_X) = -\Phi_n.$$

Show that $\|\Phi\| = \sum_{n=0}^{\infty} |\Phi_n|$, $\|\Phi\|_1 = \sum_{n=0}^{\infty}(n+1)|\Phi_n|$, and check directly the inequality $|W(\underline{\sigma}^-, \underline{\sigma}^+)| \le \|\Phi\|_1$.

§5.3 Gibbs distributions on $\mathbb{Z}^+$

Let T be a $(n+1) \times (n+1)$ mixing compatibility matrix with mixing time $a(T)$ (*i.e.* $T^{a(T)+1}_{\sigma\sigma'} > 0$) and let $\Phi \in B$ be a potential on $\{0, \ldots, n\}^{\mathbb{Z}}_T$. Let $\{0, \ldots, n\}^{\mathbb{Z}^+}_T$ be the space of sequences $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}^+}_T$ such that $T_{\sigma_i\sigma_{i+1}} = 1$, $i = 0, 1, \ldots$, *i.e.* the space of the "unilateral" sequences with labels in $\mathbb{Z}^+$.

If $m \in \mathcal{M}^0(\{0, \ldots, n\}^{\mathbb{Z}^+}_T)$ and if $\Lambda = [a, a+1, \ldots, b-1, b] \subset \mathbb{Z}^+$, $\underline{\sigma}_\Lambda \in \{0, \ldots, n\}^{\Lambda}_T$, we called the function $\underline{\sigma}' \to m(\underline{\sigma}_\Lambda|\underline{\sigma}'_{\Lambda^c})$, defined by $\frac{dm'}{dm}(\underline{\sigma}')$, where $m'(E) = m(E \cap C^{\Lambda}_{\underline{\sigma}_\Lambda})$, the "probability of $\underline{\sigma}_\Lambda$ conditional to $\underline{\sigma}'$ with respect to the probability distribution m" (cf. definition (5.1.2)).

D5.3.1 **(5.3.1) Definition:** (Semiinfinite Gibbs distributions)
Let $\Phi \in B$ be a potential for $\{0, \ldots, n\}^{\mathbb{Z}}_T$ (see definition (5.1.1)), where T is mixing with mixing time $a(T) \ge 0$; we shall say that $m \in \mathcal{M}^0(\{0, \ldots, n\}^{\mathbb{Z}^+}_T)$ is a semiinfinite Gibbs distribution *or a* Gibbs distribution on $\mathbb{Z}^+$ *with potential Φ if the conditional probabilities of m verify*

e5.3.1
$$\frac{m(\sigma'_0, \ldots, \sigma'_a|\sigma_{a+1}, \ldots)}{m(\sigma''_0, \ldots, \sigma''_a|\sigma_{a+1}, \ldots)} = \exp\Big(-\sum_{\substack{X \cap [0,a] \ne \emptyset \\ X \subset \mathbb{Z}^+}} [\Phi_X(\underline{\sigma}'_X) - \Phi_X(\underline{\sigma}''_X)]\Big) \quad (5.3.1)$$

for every choice of $a \ge a(T)$ and of the sequences $\underline{\sigma}' = (\sigma'_0 \ldots \sigma'_a \sigma_{a+1} \ldots)$, $\underline{\sigma}'' = (\sigma''_0 \ldots \sigma''_a \sigma_{a+1}, \ldots) \in \{0, \ldots, n\}^{\mathbb{Z}^+}_T$.

We define the shift τ *on $\{0, \ldots, n\}^{\mathbb{Z}^+}_T$ as*

e5.3.2
$$\tau(\sigma_0, \sigma_1, \ldots) = (\sigma_1, \sigma_2, \ldots), \qquad (5.3.2)$$

and the semiinfinite energy per site *of Φ as*

e5.3.3
$$\widehat{A}(\underline{\sigma}) = \sum_{X \ni 0,\, X \subset \mathbb{Z}^+} \Phi_X(\underline{\sigma}_X) \qquad \underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}_T. \qquad (5.3.3)$$

In terms of $\widehat{A}$ we can equivalently say that m is a Gibbs distribution on $\mathbb{Z}^+$ with potential $\Phi \in B$ if

e5.3.4
$$\frac{m(\sigma_0' \ldots \sigma_a' | \sigma_{a+1} \ldots)}{m(\sigma_0'' \ldots \sigma_a'' | \sigma_{a+1} \ldots)} = \exp\Big(- \sum_{k=0}^{\infty} (\widehat{A}(\tau^k \underline{\sigma}') - \widehat{A}(\tau^k \underline{\sigma}''))\Big), \qquad (5.3.4)$$

with the notations in (5.3.1), (5.3.2) and (5.3.3), (note that the series in (5.3.4) is, in fact, a finite sum of at most $a+1$ terms). We call $G_+^0(\Phi)$ the set of Gibbs distributions on $\mathbb{Z}^+$ with potential $\Phi \in B$, and $G_+(\Phi)$ the set of Gibbs distributions on $\mathbb{Z}^+$ with potential $\Phi \in B$ which are invariant under translation.

Proceeding exactly as in Sections §5.1 and §5.2 one shows the following proposition, analogous to proposition (5.2.1).

P5.3.1 **(5.3.1) Proposition:** (Uniqueness of semiinfinite Gibbs distributions) *Under the hypotheses of proposition (5.1.1) one has*
(i) $G_+(\Phi)$ contains at least one element for all $\Phi \in B$, and it is convex and compact;
(ii) if $\|\Phi\|_1 < +\infty$, cf. *(5.2.1), $G_+(\Phi)$ contains a unique element $\widetilde{m}_+$;*
(iii) the distribution $\widetilde{m}_+$ attributes positive probability to all T–compatible cylinders (hence every set of $\widetilde{m}_+$–probability 1 is dense in $\{0, \ldots, n\}_T^{\mathbb{Z}^+}$);
(iv) if $\|\Phi\|_1 < +\infty$ and if $\mathcal{B}(+\infty)_{\widetilde{m}^+} = \cap_{N>0} \mathcal{B}([N, +\infty])_{\widetilde{m}^+}$ then $\mathcal{B}(+\infty)_{\widetilde{m}^+}$ is a trivial σ–algebra in the sense that if f is $\mathcal{B}(+\infty)_{\widetilde{m}^+}$–measurable then f is constant $\widetilde{m}_+$–almost everywhere;
v) if $\Phi^{(k)} \to \Phi$ in B then $G_+(\Phi^{(k)}) \to G_+(\Phi)$ in the sense analogous to that discussed in item (ii) of corollary (5.1.1).

Analogous definitions can be given for *Gibbs distributions on $\mathbb{Z}^-$*: formulations are left to the reader.

The interest of the semiinfinite Gibbs distributions lies in the following remark. Let $\Lambda_k = [-k, k]$ and, for simplicity, let $T_{\sigma\sigma'} \equiv 1$. Then, if $\underline{\sigma} \in \{0, \ldots, n\}^{\mathbb{Z}}$ we can regard $\underline{\sigma}$ as composed by its "left and right parts" $\underline{\sigma}^- = (\sigma_i)_{i \in \mathbb{Z}^-}$ and $\underline{\sigma}^+ = (\sigma_i)_{i \in \mathbb{Z}^+}$, $\underline{\sigma} = (\underline{\sigma}^-, \underline{\sigma}^+)$; then (cf. second equation in (5.1.11))

e5.3.5
$$\begin{aligned} U_{\Lambda_k}(\underline{\sigma}_{\Lambda_k}, \underline{\sigma}_{\Lambda_k^c}) = & \sum_{\substack{X \subset \mathbb{Z}^- \\ X \cap \Lambda_k \neq \emptyset}} \Phi_X(\underline{\sigma}_X^-) + \sum_{\substack{X \subset \mathbb{Z}^+ \\ X \cap \Lambda_k \neq \emptyset}} \Phi_X(\underline{\sigma}_X^+) + W_\Phi(\underline{\sigma}^-, \underline{\sigma}^+) = \\ & \stackrel{def}{=} U_k^-(\underline{\sigma}^-) + U_k^+(\underline{\sigma}^+) + W_\Phi(\underline{\sigma}^-, \underline{\sigma}^+), \end{aligned} \qquad (5.3.5)$$

where W_Φ is defined in (5.2.2) with the further constraint $R \cap \Lambda_k \neq \emptyset$ (and it verifies (5.2.3)), and $U_k^\pm$ are implicitly defined in (5.3.5).

Equation (5.3.5), inserted in (5.1.23), shows that the "finite volume" approximation m_k, defined there, to the Gibbs distribution with potential Φ can be expressed as

e5.3.6
$$m_k(d\underline{\sigma}) = \frac{\widetilde{m}_k^+(d\underline{\sigma}^+)\, \widetilde{m}_k^-(d\underline{\sigma}^-)\, e^{-W_\Phi(\underline{\sigma}^+, \underline{\sigma}^-)}}{c(k)}, \qquad (5.3.6)$$

where $c(k) > 0$ is a normalization constant and $\widetilde{m}_k^+$, $\widetilde{m}_k^-$ are two suitable distributions on $\{0,\ldots,n\}^{\mathbb{Z}^+}$ and on $\{0,\ldots,n\}^{\mathbb{Z}^-}$. For example $\widetilde{m}_k^+$ is defined by

e5.3.7
$$\int f(\underline{\sigma})\,\widetilde{m}_k^+(d\underline{\sigma}) = \sum_{\underline{\sigma}\in\{0,\ldots,n\}^{k+1}} \frac{f(\underline{\sigma}\,\underline{\widetilde{\sigma}}_{\Lambda_k^{+,c}})\,e^{-U_k^+(\underline{\sigma}\,\underline{\widetilde{\sigma}}_{\Lambda_k^{+,c}})}}{\sum\limits_{\underline{\sigma}'\in\{0,\ldots,n\}^{k+1}} e^{-U_k^+(\underline{\sigma}'\,\underline{\widetilde{\sigma}}_{\Lambda_k^{+,c}})}} \tag{5.3.7}$$

for every $f \in C(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$, if $\Lambda_k^+ = [0,k]$, hence $\Lambda_k^{+,c} = [k+1,\ldots,+\infty)$ and if $\underline{\widetilde{\sigma}} \in \{0,\ldots,n\}_T^{\mathbb{Z}_+}$ is an arbitrarily fixed reference configuration (a device already repeatedly used, *e.g.* in (5.1.22)): for instance one could take $\widetilde{\sigma}_j \equiv 0$ for all $j \geq 0$, in the case $T_{\sigma\sigma'} \equiv 1$.

The following proposition appears, therefore, quite natural.

P5.3.2 **(5.3.2) Proposition:** (Relation between infinite and semiinfinite Gibbs distributions)
Under the hypotheses of proposition (5.2.1) there exist two distributions $\widetilde{m}^+$, on $\{0,\ldots,n\}_T^{\mathbb{Z}^+}$, and $\widetilde{m}^-$, on $\{0,\ldots,n\}_T^{\mathbb{Z}^-}$, such that

e5.3.8
$$m(d\underline{\sigma}) = C^{-1}\widetilde{m}^+(d\underline{\sigma}^+)\widetilde{m}^-(d\underline{\sigma}^-)\chi(\underline{\sigma}^-,\underline{\sigma}^+)e^{-W_\Phi(\underline{\sigma}^-,\underline{\sigma}^+)}, \tag{5.3.8}$$

where $C > 0$ is a normalization constant, $\chi(\underline{\sigma}^-,\underline{\sigma}^+) = 1$ if the bilateral sequence $\underline{\sigma}^-\underline{\sigma}^+$, obtained by appending $\underline{\sigma}^+$ to the right of $\underline{\sigma}^-$, is in $\{0,\ldots,n\}_T^{\mathbb{Z}}$ and $\chi(\underline{\sigma}^-,\underline{\sigma}^+) = 0$ otherwise, and m is the Gibbs distribution (unique by proposition (5.2.1)) with potential Φ.

Furthermore $\widetilde{m}^+$ is the Gibbs distribution on $\mathbb{Z}^+$ with potential Φ and $\widetilde{m}^-$ is the analogous Gibbs distribution on $\mathbb{Z}^-$, cf. *definition (5.3.1).*

Proof: For simplicity we consider only the case $T_{\sigma\sigma'} = 1$. A computation identical to that leading from (5.1.22) to (5.1.23) shows that if Φ has finite range R then m_k^+ has conditional probability such that

e5.3.9
$$\frac{m_k^+(\sigma_0'\ldots\sigma_a'|\sigma_{a+1}\ldots)}{m_k^+(\sigma_0''\ldots\sigma_a''|\sigma_{a+1}\ldots)} = \exp\Big(-\sum_{h=0}^{\infty}[\widehat{A}(\tau^h\underline{\sigma}') - \widehat{A}(\tau^h\underline{\sigma}'')]\Big), \tag{5.3.9}$$

if $a + R < k$, and m_k^+ is defined in (5.3.7).

Assuming that Φ is in the space B_0 of the finite range potentials and proceeding as in the proof of proposition (5.1.1), we see that m_k^+ converges to the semiinfinite Gibbs state $\widetilde{m}^+$. A similar argument holds for m_k^-; therefore (5.3.6) and the continuity of the function $(\underline{\sigma}^-,\underline{\sigma}^+) \to W_\Phi(\underline{\sigma}^-,\underline{\sigma}^+)$ imply (5.3.8) in the limit as $k \to \infty$, with $\chi = 1$ (because under our hypotheses $T_{\sigma\sigma'} = 1$).

If $\Phi \notin B_0$ but $\|\Phi\|_1 < +\infty$ there exists a sequence $\Phi^{(k)} \in B_0$, $k = 1,2,\ldots$, such that $\|\Phi^{(k)} - \Phi\|_1 \xrightarrow[k\to\infty]{} 0$. Then from (5.3.8) written for $\Phi^{(k)}$ and from the uniform continuity, with respect to Φ, of the function $(\underline{\sigma}^-,\underline{\sigma}^+)\to W_\Phi(\underline{\sigma}^-,\underline{\sigma}^+)$ the validity of (5.3.8) for Φ follows in the limit $k\to\infty$. ■

The relation (5.3.8) and propositions (5.2.1) and (5.3.1) imply the following corollary; if $a(T)>0$ one must also make use of (5.2.5).

C5.3.1 **(5.3.1) Corollary:** (Absolute continuity of the restriction to $\mathcal{B}(\mathbb{Z}_+)$ of a Gibbs state with respect to the corresponding semiinfinite Gibbs state)
If $\Phi\in B$ and $\|\Phi\|_1<+\infty$, if m^+ denotes the restriction of the Gibbs distribution $m\in G(\Phi)$ to the σ–algebra $\mathcal{B}(\mathbb{Z}^+)$ and if $\widetilde{m}^+$ denotes the element of $G_+(\Phi)$, one has

e5.3.10
$$\left(\frac{dm^+}{d\widetilde{m}^+}\right)^{\pm1}\le e^{4\|\Phi\|_1}((1+n)e^{2\|\Phi\|})^{a(T)}. \tag{5.3.10}$$

Furthermore the function $h\overset{def}{=}\frac{dm^+}{d\widetilde{m}^+}$ is continuous: $h\in C(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$.

Proof: This follows immediately from the integration of (5.3.6) with respect to $\underline{\sigma}^-$ and from proposition (5.3.2): by expressing $\widetilde{m}_k^-$ with the formula analogous to (5.3.7) one has to use that, for any string $\ldots,\sigma_{-a-2},\sigma_{-a-1}$ there is at least one string $\underline{\sigma}_0\in\{0,\ldots,n\}^a$ such that $\ldots,\sigma_{-a-2},\sigma_{-a-1},\underline{\sigma}_0,\underline{\sigma}^+$ is a compatible string, while the number of such connection strings is at most
N5.3.1 $(n+1)^a$.[1] ■

It is interesting to remark that the semiinfinite Gibbs states solve certain eigenvalue problems.

P5.3.3 **(5.3.3) Proposition:** (Gibbs distributions as solutions of eigenvalue problems)
Let Φ be a potential for $\{0,\ldots,n\}_T^{\mathbb{Z}}$ with $\|\Phi\|_1<\infty$ and T a mixing compatibility matrix; using the notations of corollary (5.3.1), proposition (5.3.2) and definition (5.3.1), set

e5.3.11
$$(Lf)(\sigma_0\sigma_1\ldots)=\sum_{\sigma=0}^{n}e^{-\widehat{A}(\sigma\sigma_0\sigma_1\ldots)}f(\sigma\sigma_0\sigma_1\ldots), \tag{5.3.11}$$

where $(\sigma_0\sigma_1\ldots)\in\{0,\ldots,n\}_T^{\mathbb{Z}^+}$, $f\in C(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$, $\widehat{A}$ is defined in (5.3.3) and, finally, the sum runs over $\sigma\in\{0,\ldots,n\}$ such that $(\sigma\sigma_0\sigma_1,\ldots)\in\{0,\ldots,n\}_T^{\mathbb{Z}^+}$.

Then L is a continuous operator on $C(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$ and its norm $||L||$ is bounded by $(n+1)e^{\|\Phi\|}$. Denote L^* the adjoint operator of L acting on $M^0(\{0,\ldots,n\}_T^{\mathbb{Z}})$. There exists $\lambda>0$ such that

e5.3.12
$$L\,h=\lambda h,\qquad L^*\widetilde{m}^+=\lambda\widetilde{m}^+,\qquad \lambda>0. \tag{5.3.12}$$

[1] To obtain the upper bound in (5.3.10) the factor $(n+1)$ could be replaced with 1: the expression in (5.3.10) has the advantage of being symmetric for the upper and lower bounds.

Furthermore the equations (5.3.12), regarded as eigenvalue problems in the unknowns $h > 0$, $h \in C(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$, $\lambda > 0$, *and in the unknowns* $\widetilde{m}^+ \in M^0(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$, $\lambda > 0$, *admit as unique solution* h *and* $\widetilde{m}^+$ *defined in proposition (5.3.2) and corollary (5.3.1).*

Remark: The operator L is called the *transfer operator*. The proof below derives the results from the previously studied existence and uniqueness theorems for infinite and semiinfinite Gibbs states. A proof entirely based on the spectral theory of the transfer operator is also possible, and classical since [Ru67]: see problems below.

Proof: Consider for simplicity only the case $a(T) = 0$ (*i.e.* $T_{\sigma\sigma'} \equiv 1$). Note that if $k > a(T)$ the equation $(L^*)^k \,\overline{m} = \overline{\lambda}^k \,\overline{m}$, $\overline{\lambda} > 0$, $\overline{m} \in M^0(\{0,\ldots,n\}_T^{\mathbb{Z}^+})$ is a different way of writing the condition that $\overline{m}$ has conditional probability verifying (5.3.4): hence $\widetilde{m}^+$ verifies equation (5.3.12) and it is its only solution because any other solution would be an element of $G_+(\Phi)$ that, instead, contains only one element.
Evidently $\lambda = (L^*\widetilde{m}^+)(1)$.
By proposition (5.3.2) one has $h\widetilde{m}^+ = m^+$ for a suitable continuous h: in fact $h(\underline{\sigma}_+)$ is proportional to $\int \widetilde{m}^-(d\underline{\sigma}_-)e^{-W(\underline{\sigma}_-,\underline{\sigma}_+)}$ for $\underline{\sigma}_+ \in \{0,\ldots,n\}^{\mathbb{Z}_+}$ by integrating (5.3.8) over $\underline{\sigma}_-$.
Furthermore m^+ is the restriction of the probability distribution m to $\mathcal{B}(\mathbb{Z}^+)$ and m is τ–invariant. Hence for all $f \in C(\{0,\ldots,\}_T^{\mathbb{Z}^+})$

e5.3.13
$$\begin{aligned}\int f\, dm^+ = &\int f(\sigma_0\sigma_1\ldots)h(\sigma_0\sigma_1\ldots)\widetilde{m}^+(d\sigma_0\sigma_1\ldots) = \\ &= \int f(\sigma_1\sigma_2\ldots)h(\sigma_0\sigma_1\ldots)\widetilde{m}^+(d\sigma_0\sigma_1\ldots).\end{aligned} \tag{5.3.13}$$

Using in the right hand side the relation $\widetilde{m}^+ = \lambda^{-1}L^*\widetilde{m}^+$, we see that

e5.3.14
$$\begin{aligned}\int fhd\widetilde{m}^+ = &\sum_{\sigma_0=0}^{n}\int f(\sigma_1\sigma_2\ldots)\lambda^{-1}h(\sigma_0\sigma_1\ldots)e^{-\widehat{A}(\sigma_0\sigma_1\ldots)}\widetilde{m}^+(d\sigma_1\sigma_2\ldots) \\ &= \int f(\sigma_1\sigma_2\ldots)(\lambda^{-1}Lh)(\sigma_1\sigma_2\ldots)\widetilde{m}^+(d\sigma_1\sigma_2\ldots) = \\ &= \int f\cdot\lambda^{-1}Lh\, d\widetilde{m}^+,\end{aligned} \tag{5.3.14}$$

hence by the arbitrariness of f we deduce that $\lambda^{-1}Lh = h$, $\widetilde{m}^+$–almost everywhere. However, as already observed in proposition (5.3.1), (iii), every set with m^+–measure 1 is dense so that the equality $\lambda^{-1}Lh = h$, holding on a dense set and involving continuous functions, must hold everywhere.
To prove uniqueness of h suppose that there exists another $\overline{h} > 0$, $\overline{h} \in C(\{0,\ldots,n\}_T^{\mathbb{Z}})$, such that $L\,\overline{h} = \overline{\lambda}\,\overline{h}$, $\overline{\lambda} > 0$. Then we begin by remarking that by integrating the latter relation with respect to $\widetilde{m}^+$ and using $L^*\widetilde{m}^+ =$

$\lambda\widetilde{m}^+$ we see that $\lambda\widetilde{m}^+(\overline{h}) = \overline{\lambda}\widetilde{m}^+(\overline{h})$, *i.e.* $\lambda = \overline{\lambda}$ (because $\overline{h} > 0$ and is continuous). We can and shall suppose $\widetilde{m}^+(\overline{h}) = 1$.

Following backwards (5.3.14), (5.3.13) we deduce that the probability distribution $(\lambda^{-1}L\,\overline{h})\widetilde{m}^+$ is identical to $\overline{h}\widetilde{m}^+$ and it is τ–invariant in the sense that

e5.3.15
$$\begin{aligned}&\int f(\sigma_0\sigma_1\ldots)\,\overline{h}(\sigma_0\sigma_1\ldots)\widetilde{m}^+(d\sigma_0\sigma_1\ldots) =\\ &= \int f(\sigma_1\sigma_2\ldots)\,\overline{h}(\sigma_0\sigma_1\ldots)\widetilde{m}^+(d\sigma_0\sigma_1\ldots).\end{aligned} \tag{5.3.15}$$

It is then possible to define a τ–invariant distribution $\overline{m}$ on $\{0,\ldots,n\}^{\mathbb{Z}}$ such that its restriction to $\mathcal{B}(\mathbb{Z}^+)$ is precisely $\overline{h}\widetilde{m}^+$: indeed we shall set, for $f \in \mathcal{B}([-N,+\infty))$–measurable,

e5.3.16
$$\int f(\sigma)d\,\overline{m} = \int f(\sigma_{-N}\sigma_{-N+1}\ldots)\,\overline{h}(\sigma_{-N}\sigma_{-N+1}\ldots)\widetilde{m}(d\sigma_{-N}\ldots), \tag{5.3.16}$$

where $\underline{\sigma} = (\sigma_{-N}\sigma_{-N+1}\ldots) \in \{0,\ldots,n\}^{[-N,+\infty)}$ is regarded as an element $\underline{\sigma}'$ of $\{0,\ldots,\}^{\mathbb{Z}^+}$ by setting $\sigma'_j = \sigma_{j-N}$, for all $j \ge 0$.

The compatibility problems that must be faced to check that this is a definition of a linear, continuous and positive functional on $C(\{0,\ldots,n\}^{\mathbb{Z}^+})$ can be solved via (5.3.15).

We proceed to compute the conditional probability of the probability distribution $\overline{m}$ in order to show that $\overline{m} = m$.

By Doob's theorem one has, $\overline{m}$–almost everywhere in $\underline{\sigma}$,

e5.3.17
$$\begin{aligned}&\frac{\overline{m}(\sigma'_{-a}\ldots\sigma'_a|\sigma_j,|j|>a)}{\overline{m}(\sigma''_{-a}\ldots\sigma''_a|\sigma_j,|j|>a)} =\\ &= \lim_{N\to\infty}\frac{\overline{m}(\sigma'_{-a}\ldots\sigma'_a|\sigma_{-N}\ldots\sigma_{-a-1}\sigma_{a+1}\ldots\sigma_N)}{\overline{m}(\sigma''_{-a}\ldots\sigma''_a|\sigma_{-N}\ldots\sigma_{-a-1}\sigma_{a+1}\ldots\sigma_N)} =\\ &= \lim_{N\to\infty}\frac{\overline{m}(\sigma_{-N}\ldots\sigma_{-a-1}\sigma'_{-a}\ldots\sigma'_a|\sigma_{a+1}\ldots\sigma_N)}{\overline{m}(\sigma_{-N}\ldots\sigma_{-a-1}\sigma''_{-a}\ldots\sigma''_a|\sigma_{a+1}\ldots\sigma_N)} =\\ &= \lim_{N\to\infty}\frac{\overline{h}(\sigma_{-N}\ldots\sigma_{-a-1}\sigma'_{-a}\ldots\sigma'_s\sigma_{a+1}\ldots\sigma_N)}{\overline{h}(\sigma_{-N}\ldots\sigma_{-a-1}\sigma''_{-a}\ldots\sigma''_a\sigma_{a+1}\ldots\sigma_N)}\cdot\\ &\quad\cdot\frac{\widetilde{m}^+(\sigma_{-N}\ldots\sigma_{-a-1}\sigma'_{-a}\ldots\sigma'_a|\sigma_{a+1}\ldots\sigma_N)}{\widetilde{m}^+(\sigma_{-N}\ldots\sigma_{-a-1}\sigma''_{-a}\ldots\sigma''_a|\sigma_{a+1}\ldots\sigma_N)},\end{aligned} \tag{5.3.17}$$

where, in the last expression, the sequences $\underline{\sigma}' = (\sigma_{-N}\ldots\sigma_{-a-1}\sigma'_{-a}\ldots\sigma'_a\,\sigma_{a+1}\ldots)$ and $\underline{\sigma}'' = (\sigma_{-N}\ldots\sigma_{-a-1}\sigma''_{-a}\ldots\sigma''_a\sigma_{a+1}\ldots)$ are thought of as elements $\widehat{\underline{\sigma}}'$ and $\widehat{\underline{\sigma}}''$ of $\{0,\ldots,\}^{\mathbb{Z}^+}$ setting, as above, $\widehat{\sigma}'_j = \sigma'_{j-N}$, $\widehat{\sigma}''_j = \sigma''_{j-N}$, $j \ge 0$.

Since the distance between $\widehat{\sigma}'$ and $\widehat{\sigma}''$ as elements of $\{0,\ldots,n\}^{\mathbb{Z}^+}$ is $\le e^{-(N-a)}$ and since $\overline{h}$ is uniformly continuous on $\{0,\ldots,n\}^{\mathbb{Z}^+}$ the first ratio in the limit in (5.3.17) tends to 1.

The second ratio is, by (5.3.1) (recall that $\widetilde m^+\in G^0_+(\Phi)$),

e5.3.18
$$\begin{aligned}&\exp\Big(-\sum_{\substack{X\subset[-N,+\infty)\\ X\cap[-N,a]\ne\emptyset}}[\Phi_X(\widehat{\underline\sigma}'_X)-\Phi_X(\widehat{\underline\sigma}''_X)]\Big)=\\ &=\exp\Big(-\sum_{\substack{X\subset[-N,+\infty)\\ X\cap[-a,a]\ne\emptyset}}[\Phi_X(\underline\sigma'_X)-\Phi_X(\underline\sigma''_X)]\Big)\xrightarrow[N\to\infty]{}\\ &\xrightarrow[N\to\infty]{}\exp\Big(-\sum_{X\cap[-a,a]\ne\emptyset}[\Phi_X(\underline\sigma'_X)-\ \Phi_X(\underline\sigma''_X)]\Big)=\\ &=\exp\Big(-\sum_{k=-\infty}^{\infty}[A_\Phi(\tau^k\underline\sigma')-A_\Phi(\tau^k\underline\sigma'')]\Big),\end{aligned}\tag{5.3.18}$$

showing that $\overline m$ has the correct conditional probabilities to say that it is the Gibbs distribution $m\in G(\Phi)$. It follows that $\overline h\widetilde m^+=h\widetilde m^+$, *i.e.* $\overline h=h$, $\widetilde m^+$–almost everywhere. This means that $\overline h=h$ everywhere because $\overline h$ and h are continuous and furthermore, as already observed, the sets of m^+–probability 1 are dense.

Problems for §5.3 (*Spectral theory of the transfer operator and Gibbs states*)

Remark: The spectral theory of the operator L on the space of the continuous functions $C(\{0,\ldots,n\}^{\mathbb Z})$ is made easy by the results of proposition (5.3.3) proving existence of positive eigenfunctions $h,\widetilde m_+$ in (5.3.12) relative to a positive eigenvalue λ. However one can prove the existence of a positive solution to the eigenvalue problems posed by (5.3.12) independently of proposition (5.3.3). Therefore we present problems which can be solved by *assuming* that (5.3.12) has positive solutions $h,\widetilde m_+$ with a positive eigenvalue λ. The independent proof of their existence (and therefore an alternative proof of proposition (5.3.3)) will be discussed in problems [5.3.17], [5.3.18] and [5.3.19].

Q5.3.1 **[5.3.1]:** (*Positivity of the transfer operator*)
Let L be the operator defined in (5.3.11), where Φ is a potential for $\{0,\ldots,n\}^{\mathbb Z}$ and $\|\Phi\|_1<+\infty$ and let $h,\widetilde m_+,\lambda$ be as in (5.3.12) and assume the normalization $\widetilde m^+(h)=1$. Show that the functions $(\lambda^{-1}L)^kf$ are equicontinuous and equibounded with respect to k, if f is positive and continuous. (*Hint:* Consider first the case $f=1$, *i.e.* $f(\underline\sigma)\stackrel{def}{=}1(\underline\sigma)\equiv1$ and proceed by comparing with 1 the ratios $(\lambda^{-1}L)^k1(\underline\sigma)/(\lambda^{-1}L)^k1(\underline\sigma')$; for this purpose write explicitly the ratios and use the uniform boundedness to deduce the result from the constancy of $\widetilde m^+$–integral of the functions $(\lambda^{-1}L)^k1(\underline\sigma)$ as k varies.)

Q5.3.2 **[5.3.2]:** In the context and under the hypotheses of problem [5.3.1], given a continuous non-negative function f show that there exists N_f such that for all $k\ge N_f$ one has $(\lambda^{-1}L)^kf>0$. (*Hint:* First note that this is obvious for cylindrical functions. Then just take a very good cylindrical approximation of f and then k larger than the size of the base of the cylinder on which f is cylindrical.)

Q5.3.3 **[5.3.3]:** In the context of problem [5.3.1], if $f\ge0$ is continuous and $\mathcal B([0,N])$–measurable there exists N_f such that for every $k\ge N_f$ one has $(\lambda^{-1}L)^kf\ge e^{-2\|\Phi\|_1}\widetilde m^+(f)$, where $\widetilde m_+$ is defined as in (5.3.12). (*Hint:* Same argument as the one suggested in problem [5.3.1] making use of $\widetilde m^+((\lambda L)^kf)\equiv\widetilde m^+(f)$; $N_f=N$.)

Q5.3.4 **[5.3.4]:** (*Contractivity of iterates of $\lambda^{-1}L$*; see [Ru67])
In the context of problem [5.3.1], let f be continuous, $\mathcal B([0,N])$–measurable and such

that $\widetilde{m}^+(f) = 0$. There exists N_f such that for all $k \ge N_f$ one has

$$\widetilde{m}^+(|(\lambda^{-1}L)^k f|) \le (1 - e^{-2\|\Phi\|_1})\, \widetilde{m}^+(|f|).$$

(*Hint:* Let $f^+ = (|f|+f)/2$ and $f^- = (|f|-f)/2$, then $\widetilde{m}^+(f^+) = \widetilde{m}^+(f^-)$. Furthermore by using the result of problem [5.3.3] for k large enough one has

$$\begin{aligned}&|(\lambda^{-1}L)^k f| = |(\lambda^{-1}L)^k f^+ - (\lambda^{-1}L)^k f^-| =\\ &|(\lambda^{-1}L)^k f^+ - e^{-2\|\Phi\|_1}\widetilde{m}^+(f^+) - (\lambda^{-1}L)^k f^- + e^{-2\|\Phi\|_1}\widetilde{m}^+(f^-)| \le\\ &\le (\lambda^{-1}L)^k f^+ - e^{-2\|\Phi\|_1}\widetilde{m}^+(f^+) + (\lambda^{-1}L)^k f^- - e^{-2\|\Phi\|_1}\widetilde{m}^+(f^-).\end{aligned}$$

Then integrate both sides with respect to $\widetilde{m}^+$ and note that $\widetilde{m}^+$ is an eigenvector for $(\lambda^{-1}L^*)$ with eigenvalue 1, and that $\widetilde{m}^+(|f|) = \widetilde{m}^+(f^+) + \widetilde{m}^+(f^-)$.)

Q5.3.5 **[5.3.5]:** From the result in problem [5.3.4] show that if f is continuous and $\widetilde{m}^+(f) = 0$ then the limit as $k \to \infty$ of $(\lambda^{-1}L)^k f$ is 0 (in the the uniform convergence topology).

Q5.3.6 **[5.3.6]:** (*Convergence of* $(\lambda^{-1}L)^k f$)
Deduce from the result of problem [5.3.5] that if f is continuous the limit for $k \to \infty$ of $(\lambda^{-1}L)^k f$ is $h\,\widetilde{m}^+(f)$, uniformly if $h > 0$ is a normalized eigenfunction $Lh = \lambda h$. Check that such an eigenfunction exists because of (5.3.12). *Existence of h can be obtained independently of proposition (5.3.3): see problems [5.3.18] and [5.3.19].* (*Hint:* : Write $f = h\widetilde{m}^+(f) + (f - h\widetilde{m}^+(f))$).

Q5.3.7 **[5.3.7]:** (*Exponential convergence*; see [Ru67] and [GL70])
Deduce from the results of problems [5.3.3], [5.3.4] and [5.3.5] that if for some $\kappa > 0$ the sum $\sum_{X \ni 0} e^{\kappa\,(\mathrm{diam}X)}\|\Phi_X\| < \infty$ then $|(\lambda^{-1}L)^n 1(\underline{\sigma}) - h(\underline{\sigma})| \le Ce^{-\kappa' n}$, for $0 < \kappa' < \kappa$.

Q5.3.8 **[5.3.8]:** (*Case of subshifts*; see [GM70])
Solve the problems analogous to the previous ones with $\{0, \ldots, n\}^{\mathbb{Z}}$ replaced by $\{0, \ldots, n\}^{\mathbb{Z}}_T$ under the only assumption that T is mixing.

Q5.3.9 **[5.3.9]:** (*Decimation of finite range Gibbs states*)
Let $\Phi \in B$ be a potential on $\{0, \ldots, n\}^{\mathbb{Z}}_T$ with T a mixing compatibility matrix and with finite range R, see definition (5.1.1); denote by $\alpha = 0, 1, \ldots, M$ the elements of $\{0, \ldots, n\}^{a(T)+R}_T$. Consider the restriction of $m \in G(\Phi)$ to the σ–algebra generated by the cylinders with base given by $\cup_i \Lambda_{k_i}$ where $k_i \in \mathbb{Z}$ and $\Lambda_k = \{0 + 2k(a(T) + R),$ $1 + 2k((a(T) + R), \ldots, a(T) + R - 1 + 2k(a(T) + R)\}$ *i.e.* the cylinders "measurable on the blocks of $a(T) + R$ sites spaced by $(a(T) + R)$". Show that it is interpretable as a Gibbs state on $\{0, \ldots, M\}^{\mathbb{Z}}$ with a "nearest neighbours" potential $\widetilde{\Phi}$ (*i.e.* $\widetilde{\Phi}_X = 0$ unless X consists of only one or two points and, in the latter case, it is also zero unless that the two points are adjacent). Find a possible expression for $\widetilde{\Phi}$.

Q5.3.10 **[5.3.10]:** (*Decimation of a Markov process*)
Let m be a Markov process on $\{0, \ldots, n\}^{\mathbb{Z}}$. Show that the restriction of m to the σ–algebra generated by the cylinders with base on the even sites is still a Markov process and that its potential $\Phi^{(1)}$ can be chosen by defining suitably a map $\Theta : \Phi \to \Theta\Phi = \Phi^{(1)}$. Such map can be chosen so that one has (having set $\Phi^{(n)} = \Theta^n \Phi$)

$$\Phi^{(n)}_{\{x\}}(\sigma_x) \xrightarrow[n\to\infty]{} w(\sigma_x), \qquad |\Phi^{(n)}_{\{x,x+1\}}(\sigma, \sigma')| \le C\varepsilon^{2^n},$$

with w suitable, $C > 0$ and $\varepsilon < 1$.

Q5.3.11 **[5.3.11]:** (*Transfer matrix*)
Check that in the finite range potential case the operator L becomes "in a certain sense" a finite matrix (called *transfer matrix*) and the problems of existence of $\widetilde{m}_+, h$ reduce to the Perron–Frobenius theorem for matrices.. (*Hint:* Note that the equations for h,

$\lambda^{-1}Lh = h$ can be interpreted, if Φ has finite range, as an equation for a positive eigenfunction of a finite mixing matrix with non-negative entries.)

Q5.3.12 **[5.3.12]:** (*Finite range potentials and algebraic spectral problem for the transfer operator*)
Check that from the proof of the result of problem [5.3.11] it also follows that the determination of $m(C^{0\ldots p}_{\sigma_0\ldots\sigma_p})$ is, if $R<\infty$, an "algebraic" problem (*i.e.* it is reduced to the study of the eigenvalues and eigenvectors of a suitable finite matrix).

Q5.3.13 **[5.3.13]:** (*Decimation theorem*)
If Φ is a potential on $\{0,\ldots,n\}^{\mathbb{Z}}_T$, with T mixing, such that $\|\Phi\|_1<\infty$, the restriction of $m\in G(\Phi)$ to the algebra generated by the cylinders measurable on the sites $k\,n$, $n\in\mathbb{Z}$ is, for k large, a Gibbs state (in a natural sense) with potential $\Phi^{(k)}$ that can be chosen such that $\sum_{X\ni 0,|X|\ge 2}\|\Phi^{(k)}_X\|\le\varepsilon_\Phi(k)$, where $\varepsilon_\Phi(k)\xrightarrow[k\to\infty]{}0$, while $\Phi^{(k)}_{\{0\}}(\sigma)$ is a convergent sequence as $k\to\infty$. Finally the $\varepsilon_\Phi(k)$ is continuous, in Φ at fixed k, with respect to the norm $\|\Phi\|_1$. Try to prove this result in the case proposed in the following problem and refer to [CO81], where it is proved.

Q5.3.14 **[5.3.14]:** (*Transfer operator for Fisher potentials*)
Check the statement in problem [5.3.13] in the case of the potential of problem [5.2.1].

Q5.3.15 **[5.3.15]:** Via the arguments necessary to solve problem [5.3.7] and under the same assumptions deduce the existence of $C_k,\alpha>0$ for which

$$|h(\sigma'_0\ldots\sigma'_k\;\sigma_{k+1}\ldots)-h(\sigma''_0\ldots\sigma''_k\sigma_{k+1}\ldots)|\le C_k e^{-\alpha k}.$$

Q5.3.16 **[5.3.16]:** (*Exponential mixing rates*)
By the arguments necessary to solve problems [5.3.7] and [5.3.15] and under the same assumptions deduce that the distribution $\widetilde m^+$ has the property of "exponential mixing on the cylinders":

$$\begin{aligned}&|\widetilde m^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N}\cap C^{j\ldots j+M}_{\sigma'_0\ldots\sigma'_M})-\widetilde m^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})(h\widetilde m^+)(C^{0\ldots M}_{\sigma'_0\ldots\sigma'_M})|\le\\&\le K\,\widetilde m^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})(h\widetilde m^+)(C^{0\ldots M}_{\sigma'_0\ldots\sigma'_M})\;\min\{1,\,e^{-\kappa\,(j-N)}\}\end{aligned}$$

for suitably chosen K, $\kappa>0$. An identical property holds if in this relation we replace $\widetilde m^+$ with $h\widetilde m^+$ everywhere it appears alone (*i.e.* except where already appears $h\widetilde m^+$).

Q5.3.17 **[5.3.17]:** (*Alternative proof of existence of eigenfunction of L^**)
Let L^* be the adjoint of the operator L on $C(\{0,\ldots,n\}^{\mathbb{Z}})$: write it explicitly as an operator on the measures on $\{0,\ldots,n\}^{\mathbb{Z}}$. Define $m\to\frac{L^*m}{L^*m(1)}$ on the space of the probability distributions m on $\{0,\ldots,n\}^{\mathbb{Z}}$ where 1 denotes the function identically 1 on $\{0,\ldots,n\}^{\mathbb{Z}}$. This is a continuous map of $M(\{0,\ldots,n\}^{\mathbb{Z}})$ into itself in the natural weak topology induced by $C(\{0,\ldots,n\}^{\mathbb{Z}})$. Show that there is a fixed point $\widetilde m_+$ such that $\widetilde m_+=\frac{L^*\widetilde m_+}{L^*\widetilde m_+(1)}\equiv\lambda^{-1}L^*\widetilde m_+$ where $\lambda=L^*\widetilde m_+(1)>0$. (*Hint:* The space $M(\{0,\ldots,n\}^{\mathbb{Z}})$ is compact and convex so that by the fixed point theorem there is a fixed point.)

Q5.3.18 **[5.3.18]:** (*Equiboundedness and equicontinuity of $(\lambda^{-1}L)^k1$*)
Show that $(\lambda^{-1}L)^k1(\sigma_0,\sigma_1,\ldots)$ is uniformly bounded and uniformly continuous in $\underline\sigma$. (*Hint:* If $U(\eta_0,\ldots,\eta_{k-1}\,|\,\sigma_0,\sigma_1,\ldots)\stackrel{def}{=}\sum_{X\cap 0,\ldots,k-1\ne\emptyset}\Phi_X((\underline{\eta\sigma})_X)$ it is, comparing ratios of terms corresponding to the same $\underline\eta$ –labels,

$$e^{-2||\Phi||_1}\frac{\sum_{\eta_1,\ldots,\eta_k}e^{-U(\eta_1,\ldots,\eta_k\,|\,\sigma_0,\sigma_1,\ldots)}}{\sum_{\eta_1,\ldots,\eta_k}e^{-U(\eta_1,\ldots,\eta_k\,|\,\sigma'_0,\sigma'_1,\ldots)}}\le e^{2||\Phi||_1}$$

However the ratio is just $\frac{L^k 1(\underline{\sigma})}{L^k 1(\underline{\sigma}')} \equiv \frac{\lambda^{-k} L^k 1(\underline{\sigma})}{\lambda^{-k} L^k 1(\underline{\sigma}')}$ and there must exist a point $\underline{\sigma}$ where $\lambda^{-k} L^k 1(\underline{\sigma}) \le 1$ and one point where $\lambda^{-k} L^k 1(\underline{\sigma}) \ge 1$ because $\widetilde{m}_+(\lambda^{-k} L^k 1) = \widetilde{m}_+(1) = 1$. Hence equiboundedness follows. Likewise one sees equicontinuity if the sequences $(\sigma_0, \sigma_1, \ldots)$ are close *i.e.* $(\sigma'_0, \sigma'_1, \ldots)$ are such that $\sigma_i = \sigma'_i$ for many $i's$.)

Q5.3.19 **[5.3.19]:** (*Alternative proof of existence of an eigenfunction h for* $\lambda^{-1}L$)
Consider the sequence $N^{-1} \sum_{k=0}^{N-1} (\lambda^{-1} L)^k 1(\underline{\sigma})$: show that the problem [5.3.18] implies that this sequence is equicontinuous and equibounded . Therefore it admits a uniformly convergent subsequence (by Ascoli–Arzelà theorem): show that its limit is a positive function h with $\lambda^{-1} L h = h$.

Bibliographical note to §5.3

The results of this section and of Section §5.2 are due to Ruelle,[Ru67], who substantially extended previous results and conjectures (by Van Hove, [VH50]). The choice of the arguments of this section and of the previous, §5.1 and §5.2, is inspired by the book [Ru78] although the exposition of this section is somewhat different from Ruelle's original.

§5.4 An application: expansive maps of [0,1]

As an application of the results of Section §5.3 we shall discuss a theory of invariant probability distributions for the simplest expansive maps of $[0,1]$, cf. example (1.2.7).

Let $f_0, f_1, \ldots, f_n$ be $(n+1)$ functions defined on the intervals $[a_\sigma, a_{\sigma+1}]$, respectively, with

e5.4.1
$$0 = a_0 < a_1 < \ldots < a_n < a_{n+1} = 1, \tag{5.4.1}$$

and with values in $[0,1]$, such that each f_σ is of class $C^{1+\varepsilon}([a_\sigma, a_{\sigma+1}])$, $\varepsilon > 0$. Let S be the map of $[0,1]$ into itself defined by

e5.4.2
$$x \to Sx = f_\sigma(x) \quad \text{if} \quad x \in (a_\sigma, a_{\sigma+1}), \tag{5.4.2}$$

and defined in 0 as $f_0(0)$, in 1 as $f_{n+1}(1)$, and, arbitrarily, in a_σ as $f_\sigma(a_\sigma)$ or $f_{\sigma-1}(a_\sigma)$, for $\sigma = 1, \ldots, n$.

The (noninvertible) dynamical system $([0,1], S)$ is well defined but in general the Lebesgue measure $\mu_0(dx) = dx$ is *not S–invariant* .

If one chooses randomly a point $x \in [0,1]$ with distribution μ_0 the asymptotic behavior of the sequence $n \to F(S^n x)$, where $F \in C^\infty([0,1])$ can, sometimes, be described by means of a Borel probability distribution μ on $[0,1]$, such that

e5.4.3
$$N^{-1} \sum_{j=0}^{N-1} F(S^j x) \xrightarrow[N \to \infty]{} \int_0^1 \mu(dx') F(x') \tag{5.4.3}$$

for μ_0–almost all points $x \in [0,1]$.

This case, obviously, occurs when there exists a probability distribution μ absolutely continuous with respect to μ_0, with a positive density h, and which is S–invariant and S–ergodic:

e5.4.4
$$\mu(dx) = h(x)dx = h(x)\mu_0(dx), \qquad \mu(E) = \mu(S^{-1}E). \tag{5.4.4}$$

The first question that can be asked is, thus, to find conditions for the existence of a measure μ which is S–invariant, S–ergodic and absolutely
N5.4.1 equivalent[1] or, at least, absolutely continuous with respect to μ_0.

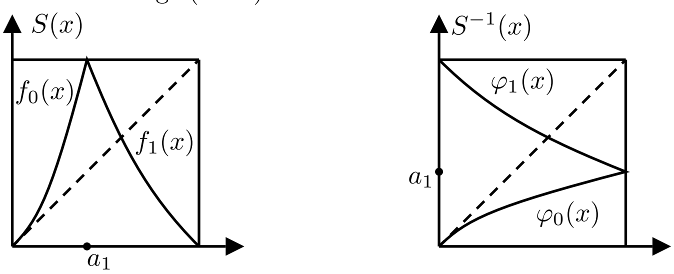
We shall not discuss the motivations of such questions which, at least for what concerns their interest for Physics, rest on the necessity of simple mathematical models to classify the phenomenology of the asymptotic behaviour of the solutions of nonlinear ordinary differential equations or of sequences of points obtained by iterating a map, starting from a given randomly chosen initial point.

A certain success has been achieved in the interpretation of experimental results in terms of simple objects like, in particular, maps of the interval $[0,1]$: above all famous is the interpretation of certain turbulence phenomena by Lorenz, [Lo63]. A kind of "experimental mathematics" has been (and is being) developed with the aim of achieving an orderly classification of the wealth of results that computer simulations or actual experiments produce.

We shall consider here the simplest case in which the map S is *strictly expansive and surjective*; this corresponds to the following conditions:

(1) $|f'_\sigma(x)| \ge \lambda > 1$ for all $x \in [a_\sigma, a_{\sigma+1}]$,
(2) $f_\sigma : [a_\sigma, a_{\sigma+1}] \leftrightarrow [0,1]$ for all $\sigma = 0, 1, \ldots, n$.

An illustration is in Fig. (5.4.1).



F5.4.1 **Fig.(5.4.1)** Graphs of an expansive and surjective map S and of its inverse.

The following proposition holds.

P5.4.1 **(5.4.1) Proposition:** (Mixing maps of the interval)
Let S be a strictly expansive surjective map of class $C^{1+\varepsilon}$, $\varepsilon > 0$, of the
N5.4.2 *interval* $[0,1]$ *into itself.*[2]

---

[1] μ is said to be absolutely equivalent to μ_0 if $\mu = h\mu_0$ and $h, h^{-1} \in L_1(\mu_0)$, *i.e.* μ and μ_0 are absolutely equivalent if μ is absolutely continuous with respect to μ_0 and μ_0 is absolutely continuous with respect to μ.

[2] This means that $f_\sigma \in C^{1+\varepsilon}([a_\sigma, a_{\sigma+1}])$, for all $\sigma = 0, 1, \ldots, n$.

(i) There exists an S–invariant probability distribution μ which is absolutely equivalent to the Lebesgue measure.
(ii) Such a probability distribution μ is mixing and one has

e5.4.5
$$\int_0^1 F(S^m x)G(x)\,\mu(dx) \xrightarrow[m\to\infty]{} \Big(\int_0^1 F(x)\mu(dx)\Big)\Big(\int_0^1 G(x)\mu(dx)\Big) \tag{5.4.5}$$

exponentially fast if $F, G \in C^\alpha([0,1])$, $\alpha > 0$. This means that $([0,1], S, \mu)$ is a mixing dynamical system that "mixes at an exponential rate the Hölder continuous functions".

Remark: This theorem is a particular case of a general theorem of Lasota–Yorke, [LY73]. We discuss it here by using a different technique of proof based on the theory of Gibbs distributions developed in the previous sections.

Proof: Consider $\{0,\ldots,n\}^{\mathbb{Z}^+}$ and the symbolic code $X : \underline{\sigma} \to X(\underline{\sigma}) = \cap_{j=0}^{+\infty} S^{-j} P_{\sigma_j}$, if we denote by P_σ the interval $P_\sigma = [a_\sigma, a_{\sigma+1}]$ and $\underline{\sigma} = (\sigma_j)_{j\in\mathbb{Z}^+}$, $\sigma_j = 0,\ldots,n$.

One deduces, by induction, calling $\varphi_\sigma : [0,1] \leftrightarrow [a_\sigma, a_{\sigma+1}]$ the inverse function of f_σ (cf. Fig. (5.4.1)) that

e5.4.6
$$\begin{aligned} &P_{\sigma_0} = P^0_{\sigma_0} = \varphi_{\sigma_0}([0,1])\ , \quad P^{0\ 1}_{\sigma_0\sigma_1} = P_{\sigma_0} \cap S^{-1} P_{\sigma_1} = \varphi_{\sigma_0}\varphi_{\sigma_1}([0,1]),\\ &P^{0\ldots N}_{\sigma_0\ldots\sigma_N} = P_{\sigma_0} \cap S^{-1}P_{\sigma_1} \cap \ldots \cap S^{-N} P_{\sigma_N} = \varphi_{\sigma_0}\varphi_{\sigma_1}\ldots\varphi_{\sigma_N}([0,1]). \end{aligned} \tag{5.4.6}$$

Furthermore φ_σ transforms intervals with interior points into intervals with interior points, so that $P^{0\ldots N}_{\sigma_0\ldots\sigma_N}$ is an interval with interior points. Hence $X(\underline{\sigma}) \neq \emptyset$.

However φ_σ "contracts" (see Fig.(5.4.1)) by a scale factor which is at least λ^{-1}, because f_σ expands by at least λ, hence the length $|\varphi_{\sigma_0}\ldots\varphi_{\sigma_N}([0,1])|$ of the interval $\varphi_{\sigma_0}\varphi_{\sigma_1}\ldots\varphi_{\sigma_N}([0,1])$ is

e5.4.7
$$|\varphi_{\sigma_0}\ldots\varphi_{\sigma_N}([0,1])| \le \lambda^{-(N+1)}. \tag{5.4.7}$$

so that $X(\underline{\sigma})$ consists of a single point. Therefore the code X is Hölder continuous and by (4.1.6)

e5.4.8
$$d(X(\underline{\sigma}), X(\underline{\sigma}')) \le d(\underline{\sigma}, \underline{\sigma}')^{\log\lambda}. \tag{5.4.8}$$

The probability distribution μ_0 is coded by X into a probability distribution $\overline{\mu}_0$ on $\{0,\ldots,n\}^{\mathbb{Z}^+}$ by setting

e5.4.9
$$\overline{\mu}_0(E) \stackrel{def}{=} \mu_0(X(E)) \qquad \text{for all } E \in \mathcal{B}(\{0,\ldots,n\}^{\mathbb{Z}^+}) \tag{5.4.9}$$

N5.4.3 isomorphic to it mod 0.[3]

[3] Indeed X is continuous and invertible outside a set of measure 0,

$$\cup_{j=-\infty}^{+\infty} S^{-j}\{a_0,\ldots,a_{n+1}\} = N \subset [0,1],$$

and it is therefore bimeasurable, see Kuratowsky's theorem quoted in the proof of proposition (2.3.3), as a map between $[0,1]\setminus N$ and $X([0,1]\setminus N)$.

The probability distribution $\overline{\mu}_0$ is now easily identified as the Gibbs distribution on $\mathbb{Z}^+$ with potential Φ such that

$$\begin{aligned}&\Phi_X(\underline{\sigma}_X)=0 \quad \text{if} \quad X\neq\{a,\ldots,a+p\} \text{ and for all } a, \text{ for all } p\ge 0,\\ &\Phi_{\{a,\ldots,a+p\}}(\sigma_0\ldots\sigma_p)=\hat{A}(\sigma_0\ldots\sigma_p 000\ldots)-\hat{A}(\sigma_0\ldots\sigma_{p-1}000\ldots),\end{aligned} \tag{5.4.10}$$

e5.4.10 where

e5.4.11

$$\hat{A}(\underline{\sigma})=-\log|\varphi'_{\sigma_0}(X(\sigma_1\sigma_2\ldots))|. \tag{5.4.11}$$

Indeed, if $\underline{\sigma}'=(\sigma'_0\ldots\sigma'_N\;\sigma_{N+1}\ldots)$, $\underline{\sigma}''=(\sigma''_0\ldots\sigma''_N\sigma_{N+1}\ldots)$ and $x'=X(\underline{\sigma}')$, $x''=X(\underline{\sigma}'')$, one has

e5.4.12

$$\begin{aligned}&\frac{\overline{\mu}_0(\sigma'_0\ldots\sigma'_N|\sigma_{N+1}\ldots)}{\overline{\mu}_0(\sigma''_0\ldots\sigma''_N|\sigma_{N+1}\ldots)}=\lim_{M\to\infty}\frac{\overline{\mu}_0(\sigma'_0\ldots\sigma'_N|\sigma_{N+1}\ldots\sigma_M)}{\overline{\mu}_0(\sigma''_0\ldots\sigma''_N|\sigma_{N+1}\ldots\sigma_M)}=\\ &=\lim_{M\to\infty}\frac{\overline{\mu}_0\Big(C^{0\ldots N\;N+1\ldots M}_{\sigma'_0\ldots\sigma'_N\;\sigma_{N+1}\ldots\sigma_M}\Big)}{\overline{\mu}_0\Big(C^{0\ldots N\;N+1\ldots M}_{\sigma''_0\ldots\sigma''_N\;\sigma_{N+1}\ldots\sigma_M}\Big)}=\\ &=\lim_{M\to\infty}\frac{|\varphi_{\sigma'_0}\ldots\varphi_{\sigma'_N}\varphi_{\sigma_{N+1}}\ldots\varphi_{\sigma_M}([0,1])|}{|\varphi_{\sigma''_0}\ldots\varphi_{\sigma''_N}\varphi_{\sigma_{N+1}}\ldots\varphi_{\sigma_M}([0,1])|}=\\ &=\lim_{M\to\infty}\frac{|\varphi_{\sigma'_0}\ldots\varphi_{\sigma'_N}(\beta)-\varphi_{\sigma'_0}\ldots\varphi_{\sigma'_N}(\alpha)|}{|\varphi_{\sigma''_0}\ldots\varphi_{\sigma''_N}(\beta)-\varphi_{\sigma''_0}\ldots\varphi_{\sigma''_N}(\alpha)|},\end{aligned} \tag{5.4.12}$$

where $[\alpha,\beta]=\varphi_{\sigma_{N+1}}\ldots\varphi_{\sigma_M}([0,1])$ is an interval of size at most $\lambda^{-(M-N)}$ around the point $X(\sigma_{N+1}\ldots)=\xi$. Hence

e5.4.13

$$\frac{\overline{\mu}_0(\sigma'_0\ldots\sigma'_N|\sigma_{N+1}\ldots)}{\overline{\mu}_0(\sigma''_0\ldots\sigma''_N|\sigma_{N+1}\ldots)}=\frac{\Big|\Big[\frac{d}{dx}\,\varphi_{\sigma'_0}\ldots\varphi_{\sigma'_N}(x)\Big]_{x=\xi}\Big|}{\Big|\Big[\frac{d}{dx}\,\varphi_{\sigma''_0}\ldots\varphi_{\sigma''_N}(x)\Big]_{x=\xi}\Big|}, \tag{5.4.13}$$

and, by the composition of derivatives, this ratio is

e5.4.14

$$\prod_{j=0}^{N}\frac{|\varphi'_{\sigma'_j}(\varphi_{\sigma'_{j+1}}\ldots\varphi_{\sigma'_N}(\xi))|}{|\varphi'_{\sigma''_j}(\varphi_{\sigma''_{j+1}}\ldots\varphi_{\sigma''_N}(\xi))|}. \tag{5.4.14}$$

Noting that $\varphi_{\sigma'_{j+1}}\ldots\varphi_{\sigma'_N}(\xi)\;\equiv\;\varphi_{\sigma'_{j+1}}\ldots\varphi_{\sigma'_N}(X(\sigma_{N+1}\ldots))$ so that $\varphi_{\sigma'_{j+1}}\ldots\varphi_{\sigma'_N}(\xi)$ is just $X(\sigma'_{j+1}\ldots\sigma'_N\;\sigma_{N+1}\ldots)$ we find

e5.4.15

$$\begin{aligned}&\frac{\overline{\mu}_0(\sigma'_0\ldots\sigma'_N|\sigma_{N+1}\ldots)}{\overline{\mu}_0(\sigma''_0\ldots\sigma''_N|\sigma_{N+1}\ldots)}=\prod_{j=0}^{N}\frac{|\varphi'_{\sigma'_j}(X(\sigma'_{j+1}\ldots))|}{|\varphi'_{\sigma''_j}(X(\sigma''_{j+1}\ldots))|}=\\ &=\exp\Big[-\sum_{j=0}^{N}\Big(-\log|\varphi'_{\sigma'_j}(X(\sigma'_{j+1}\ldots))|+\log|\varphi'_{\sigma''_j}(X(\sigma''_{j+1}\ldots))|\Big)\Big]=\\ &=\exp\Big[-\sum_{j=0}^{N}\Big(\hat{A}(\tau^j\underline{\sigma}')-\hat{A}(\tau^j\underline{\sigma}'')\Big)\Big]=\exp\Big[-\sum_{j=0}^{\infty}\Big(\hat{A}(\tau^j\underline{\sigma}')-\hat{A}(\tau^j\underline{\sigma}'')\Big)\Big],\end{aligned} \tag{5.4.15}$$

from (5.4.11) and because $\tau^M \underline{\sigma}' = \tau^M \underline{\sigma}''$, if $M > N$. Therefore if we note that

e5.4.16
$$\begin{aligned} &\hat{A}(\underline{\sigma}) = \hat{A}(000\ldots) + \sum_{k=0}^{\infty} \Phi_k(\sigma_0 \ldots \sigma_k), \\ &\Phi_k(\sigma_0 \ldots \sigma_k) = \hat{A}(\sigma_0 \ldots \sigma_k 000\ldots) - \hat{A}(\sigma_0 \ldots \sigma_{k-1} 000\ldots), \end{aligned} \tag{5.4.16}$$

we deduce that

e5.4.17
$$\frac{\overline{\mu}_0(\sigma'_0 \ldots \sigma'_N | \sigma_{N+1} \ldots)}{\overline{\mu}_0(\sigma''_0 \ldots \sigma''_N | \sigma_{N+1} \ldots)} = e^{-\sum_{j,k=0}^{\infty} \{\Phi_k(\sigma'_j \ldots \sigma'_{j+k}) - \Phi_k(\sigma''_j \ldots \sigma''_{j+k})\}}. \tag{5.4.17}$$

Hence by comparing (5.4.17) with (5.3.1) (or by comparing (5.4.15) with (5.3.4)), we see that $\overline{\mu}_0$ is the Gibbs distribution with potential Φ given by (5.4.10), provided no convergence problems arise in summing the series in (5.4.16).

If $k > 2$

e5.4.18
$$\begin{aligned} &|\Phi_k(\sigma_0 \ldots \sigma_k)| \equiv |\hat{A}(\sigma_0 \ldots \sigma_k 000\ldots) - \hat{A}(\sigma_0 \ldots \sigma_{k-1} 000\ldots)| \equiv \\ &\equiv \Big| \log \frac{|\varphi'_{\sigma_0}(X(\sigma_1\sigma_2 \ldots \sigma_k 000\ldots))|}{|\varphi'_{\sigma_0}(X(\sigma_1\sigma_2 \ldots \sigma_{k-1} 000\ldots))|} \Big| \le \\ &\le (\inf_{x\in[0,1],\sigma} |\varphi'_\sigma(x)|)^{-1} \\ &\qquad \Big| |\varphi'_{\sigma_0}(X(\sigma_1\sigma_2 \ldots \sigma_k 000\ldots))| - |\varphi'_{\sigma_0}(X(\sigma_1\sigma_2 \ldots \sigma_{k-1} 000\ldots))| \Big| \le \\ &\le \frac{\sup_\sigma C_{\varphi_\sigma}}{\inf |\varphi'_\sigma(x)|} \Big| X(\sigma_1\sigma_2 \ldots \sigma_k 000\ldots) - X(\sigma_1\sigma_2 \ldots \sigma_{k-1} 000\ldots) \Big|^{\varepsilon} \le \\ &\le \frac{\sup_\sigma C_{\varphi_\sigma}}{\inf |\varphi'_\sigma(x)|} \quad \lambda^{-\varepsilon(k-1)} \le C \lambda^{-\varepsilon k}, \end{aligned} \tag{5.4.18}$$

if C_{φ_σ} is the Hölder continuity modulus in $C^\varepsilon([0,1])$ of $x \to |\varphi'_{\sigma_0}(X)|$ (we recall that $f_\sigma \in C^{1+\varepsilon}([a_\sigma, a_{\sigma+1}])$ and $|f'_\sigma| \ge \lambda > 1$ by assumption) and C is a suitable constant. Hence not only the series in (5.4.16) is totally convergent with respect to the variations of $\underline{\sigma}$, but also one has

e5.4.19
$$\begin{aligned} &\|\Phi\|_1 < +\infty, \qquad \text{and} \\ &\sum_{X \ni 0} e^{\kappa\,(\text{diam } X)} \|\Phi_X\| < +\infty \text{ for all } \kappa < \varepsilon \log \lambda. \end{aligned} \tag{5.4.19}$$

From item (ii) in proposition (5.3.1) it follows that $\overline{\mu}_0$ is uniquely determined by its conditional probability (5.4.17).

Furthermore from corollary (5.3.1) it follows that the Gibbs distribution $\overline{\mu}$ with potential Φ, on $\{0,\ldots,n\}^{\mathbb{Z}}$ (*i.e.* on the *bilateral* sequences), restricted to the σ–algebra $\mathcal{B}(\mathbb{Z}^+)$ is absolutely continuous (in fact equivalent) with respect to $\overline{\mu}_0$ and it is τ–invariant:

e5.4.20
$$\overline{\mu}(E) = \overline{\mu}(\tau^{-1}E) \qquad \text{for all } E \in \mathcal{B}(\mathbb{Z}^+), \tag{5.4.20}$$

as a consequence of the τ–invariance of $\overline{\mu}$.
The triad $([0,1], S, \mu)$ with μ defined by

e5.4.21
$$\mu(E) = \overline{\mu}(X^{-1}(E)) \qquad \text{for all } E \in \mathcal{B}([0,1]) \tag{5.4.21}$$

is a dynamical system in which μ is S–invariant and S–mixing: this follows from the corresponding property of $\overline{\mu}$ that, in turn, follows from the results of Section §5.2 on Gibbs distributions on $\{0,\ldots,n\}^{\mathbb{Z}}$.
The bound on the exponential mixing rate follows from the property of $\overline{\mu}$

e5.4.22
$$\begin{aligned}|\,\overline{\mu}(C^{0\ldots N}_{\sigma_0\ldots\sigma_N}\cap C^{j\ldots j+M}_{\sigma'_0\ldots\sigma'_M}) - \overline{\mu}(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})\,\overline{\mu}(C^{0\ldots M}_{\sigma'_0\ldots\sigma'_N})| \le\\ \le K\,\overline{\mu}(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})\,\overline{\mu}(C^{0\ldots M}_{\sigma'_0\ldots\sigma'_M})\min\{1, e^{-\kappa\,(j-N)}\},\end{aligned} \tag{5.4.22}$$

where $\kappa > 0$, $K > 0$; the latter property follows from the results discussed in the problems [5.3.7], [5.3.15] and [5.3.16] and the second of (5.4.20).
Every $F \in C^\alpha([0,1])$, $\alpha > 0$, with a Hölder continuity modulus C_F can be expressed in the coordinates $\underline{\sigma}$ by setting

e5.4.23
$$\overline{F}(\underline{\sigma}) = F(X(\underline{\sigma})) = F(X(000\ldots)) + \sum_{k=0}^{\infty}\psi_k(\sigma_0\ldots\sigma_k), \tag{5.4.23}$$

where

e5.4.24
$$\begin{aligned}&\psi_k(\sigma_0\ldots\sigma_k) = F(X(\sigma_0\ldots\sigma_k 000\ldots)) - F(X(\sigma_0\ldots\sigma_{k-1}000\ldots)),\\ &|\psi_k(\sigma_0\ldots\sigma_k)| \le C_F|X(\sigma_0\ldots\sigma_k 000\ldots) - X(\sigma_0\ldots\sigma_{k-1}000)|^\alpha \le C'\lambda^{-\alpha k},\end{aligned} \tag{5.4.24}$$

for a suitable $C' > 0$. Therefore we can remark that, having set $\psi_{-1}(\sigma_{-1}) = F(X(000\ldots))$,

e5.4.25
$$\begin{aligned}\int F(S^j x)\,F(x)\mu(dx) = \int \overline{F}(\tau^j\underline{\sigma})\;\overline{F}(\underline{\sigma})\,\overline{\mu}(d\underline{\sigma}) =\\ \sum_{h,k=-1}^{\infty}\sum_{\sigma}\psi_k(\sigma_0\ldots\sigma_k)\psi_h(\sigma_j\ldots\sigma_{j+h})\,\overline{\mu}(C^{0\ldots k}_{\sigma_0\ldots\sigma_k}\cap C^{j\ldots j+h}_{\sigma_j\ldots\sigma_{j+h}}),\end{aligned} \tag{5.4.25}$$

where $\sum_\sigma$ denotes summation over $\sigma_0\ldots\sigma_k, \sigma_j\ldots\sigma_{j+h}$. From (5.4.22) and (5.4.23) we see that the (5.4.25) converges to the square of

$$\sum_{k=-1}^{\infty}\sum_{\sigma_0\ldots\sigma_k}\psi_k(\sigma_0\ldots\sigma_k)\,\overline{\mu}(C^{0\ldots k}_{\sigma_0\ldots\sigma_k})$$

with exponential rate, as $j \to \infty$, and the constant κ' of the exponential convergence can be chosen arbitrarily provided $\kappa' < (\min\{\alpha,\varepsilon\})\log\lambda$. ■

Problems for §5.4

Q5.4.1 **[5.4.1]:** (*Markovian maps of the interval*)
If S is strictly expansive but not surjective (cf. proposition (5.4.1) and Fig. (5.4.1))

check that the same arguments of the proof of proposition (5.4.1) can be repeated under the hypothesis that S is *Markovian i.e.* $\cup_{\sigma=0}^{n+1} S^{\pm} a_\sigma \subseteq \{0, a_1, \ldots, a_{n+1}\}$, where $S^{\pm}(a_\sigma)$ denotes the right and left limits of $S(x)$ as $x \to a_\sigma$. (*Hint:* We must replace $\{0,\ldots,n\}^{\mathbb{Z}^\pm}$ with $\{0,\ldots,n\}_T^{\mathbb{Z}^\pm}$, where T is a suitable compatibility matrix (which?), see problems of §5.3.)

Q5.4.2 **[5.4.2]:** (*Symbolic theory of invariant distributions for Markovian maps of the interval*)
Show that the equation for the density h of the probability distribution $\overline{\mu}$ with respect to $\overline{\mu}_0$ is, by proposition (5.3.3), the positive solution of the equation $Lf = f$ on $C(\{0,\ldots,n\}^{\mathbb{Z}})$ with

$$(Lf)(\underline{\sigma}) = \sum_{\sigma=0}^{n} e^{-\widehat{A}(\sigma\sigma_0\sigma_1\ldots)} f(\sigma\sigma_0\sigma_1\ldots).$$

Q5.4.3 **[5.4.3]:** (*Density for the invariant distribution for expansive Markovian maps of the interval*)
Show that the equation of problem [5.4.2] is the symbolic version of the equation, on $L_\infty([0,1])$, $Lg = g$ with:

$$(Lg)(x) = \sum_{\sigma=0}^{n} |\varphi_0'(x)| g(\varphi_\sigma(x)).$$

Give a geometric interpretation of this equation.

Q5.4.4 **[5.4.4]:** (*Existence of coordinates in which the Lebesgue measure is invariant for an expansive Markovian map of the interval*)
If the interval map S admits μ as invariant probability distribution $\mu(dx) = h(x)dx$ with h continuous and positive and if we set $y = \pi(x) = \int_0^x h(\xi)d\xi$ we define a change of coordinates $x \leftrightarrow y$ such that $\overline{S} = \pi S \pi^{-1}$, image of S in the new coordinates, is an expansive map (not necessarily strictly such) that admits Lebesgue measure as invariant measure.

Q5.4.5 **[5.4.5]:** Interpret the result of problem [5.4.4] as the statement that "a necessary condition in order that S admits an invariant measure absolutely continuous with respect to Lebesgue measure, with continuous and positive density h, is that there exists a system of coordinates on $[0,1]$ in which S appears as expansive", [Pi79].

Q5.4.6 **[5.4.6]:** (*About Ulam–Von Neumann map*)
Strict expansivity is not a necessary condition in order that a sufficiently regular surjective map admits an invariant probability distribution which is absolutely continuous with respect to Lebesgue measure. For an example check that the map $S(x) = 4x(1-x)$ admits the measure with density $(\pi^2 x(1-x))^{-1/2}$ as an invariant probability distribution.

Q5.4.7 **[5.4.7]:** (*Markovian quadratic maps of the interval and Ruelle's points*)
The map $S(x) = Ax(1-x)$, $A < 4$ is Markovian in the sense of problem [5.4.1], with respect to suitable decompositions $a_0 = 0 < a_1 < \ldots < a_{n+1} = 1$ of $[0,1]$, for infinitely many values of A (however the decomposition can depend on A): find one such value of A. The latter values of A are called "Ruelle's points". (*Hint:* If A is such that $S^3(1/2) = x_A$, with $x_A =$ non-zero fixed point of S, then the decomposition 0, 1/2, x_A, 1 is Markovian, [Ru77], [Pi79], [Pi80] and [Pi81].)

Bibliographical note to §5.4

The theory of the maps of $[0,1]$ into itself has a long history. Its origin is independent from the theory of Gibbs distributions and it often deals with cases that are not easily reduced to problems relative to Gibbs distributions: see for instance the works of Renij, [Re57], Ulam and von Neumann, [UV47], and Lasota and Yorke, [LY73]. The method of symbolic dynamics, is easily applicable only in the very special "Markovian" cases (*i.e.* when there exists a finite decomposition of $[0, 1]$ into intervals such that the points on the boundaries of the intervals are transformed by the map into a subset of themselves, and, furthermore, in every interval of the decomposition the map is monotonic), has been employed in [La78], [Ru77] and [Bo75]. Other more recent applications often using other methods are in [Mi79], [CE80a], [Pi79], [Pi80], [Pi??], [PY79], [Fe78], [Fe79] and [CEL80].

CHAPTER VI

General properties of Gibbs and SRB distributions

§6.1 Variational properties of Gibbs distributions

The study of Gibbs distributions can be performed to remarkable depth as we shall hint in this section and in the forthcoming ones. We begin with a structure theorem which can be articulated into various propositions.

P6.1.1 **(6.1.1) Proposition:** (Thermodynamic limit)
Let T be a mixing $(n+1)\times(n+1)$ compatibility matrix and Φ a potential in the space B of the potentials for $\{0,\ldots,n\}_T^{\mathbb{Z}}$, cf. *definition (5.1.1). Define the* partition function *of Φ in the box $\Lambda_N=[1,N]$ as*

e6.1.1
$$Z_N(\Phi)=\sum_{\underline{\sigma}\in\{0,\ldots,n\}_T^{\Lambda_N}}\exp\Big[-\Big(\sum_{R\subset\Lambda_N}\Phi_R(\underline{\sigma}_R)\Big)\Big]. \tag{6.1.1}$$

Then the limit

e6.1.2
$$P(\Phi)=\lim_{N\to\infty}N^{-1}\log Z_N(\Phi) \tag{6.1.2}$$

exists and it defines a function $\Phi\to P(\Phi)$ on B which is convex and Lipshitz continuous:

e6.1.3
$$|P(\Phi)-P(\Psi)|\le\|\Phi-\Psi\|. \tag{6.1.3}$$

The function $P(\Phi)$ is called the pressure *of the potential Φ.*

N6.1.1 *Proof:* Setting[1] $U_\Lambda^\Phi(\underline{\sigma})=\sum_{R\subset\Lambda}\Phi_R(\underline{\sigma}_R)$, cf. (5.1.11) in definition (5.1.3),

[1] With the notations used in definition (5.1.3) we should write $U_\Lambda^{0,\Phi}(\sigma)$ instead of $U_\Lambda^\Phi(\sigma)$: we prefer to drop the label 0 throughout all this chapter in order not to use an overwhelming notation.

one has

e6.1.4

$$\begin{aligned}&|N^{-1}\log Z_N(\Phi)-N^{-1}\log Z_N(\Psi)|=\\&=\Big|N^{-1}\int_0^1 dt\,\frac{d}{d\,t}\,\log Z_N(\Psi+t(\Phi-\Psi))\Big|=\qquad(6.1.4)\\&=\Big|N^{-1}\int_0^1 dt\frac{\sum_{\underline\sigma\in\{0,\ldots,n\}_T^{\Lambda_N}}U^{\Phi-\Psi}_{\Lambda_N}(\underline\sigma)\,e^{-U^{\Psi+t(\Phi-\Psi)}_{\Lambda_N}(\underline\sigma)}}{\sum_{\underline\sigma\in\{0,\ldots,n\}_T^{\Lambda_N}}e^{-U^{\Psi+t(\Phi-\Psi)}_{\Lambda_N}(\underline\sigma)}}\Big|\le\|\Phi-\Psi\|,\end{aligned}$$

having used remark (2) to definition (5.1.3), *i.e.*

e6.1.5

$$|U^\Phi_{\Lambda_N}(\underline\sigma)|\le\|\Phi\|N\qquad\text{for all }\Phi\in B.\tag{6.1.5}$$

Equation (6.1.1) and (6.1.5) immediately imply

e6.1.6

$$-\|\Phi\|+\log(n+1)\le N^{-1}\log Z_N(\Phi)\le\|\Phi\|+\log(n+1),\tag{6.1.6}$$

and therefore (6.1.4) and (6.1.6) imply that in order to show existence of the limit $P(\Phi)$ it suffices to consider $\Phi\in B^0$ where B^0 is an arbitrary set of potentials dense in B. Naturally, we shall select the set B^0 of the potentials Φ with finite range $R_\Phi<\infty$, cf. remark (4) to definition (5.1.3).

For simplicity we consider only the case $T_{\sigma\sigma'}=1\ \forall\sigma,\sigma'$.

If the finite range of Φ is $R_\Phi\equiv R$ and if $\underline\sigma=(\sigma'_1\ldots\sigma'_N\sigma''_1\ldots\sigma''_M)=(\underline\sigma'\underline\sigma'')\in\{0,\ldots,n\}^{N+M}$, we get

e6.1.7

$$|U^\Phi_{\Lambda_{N+M}}(\underline\sigma'\underline\sigma'')-U^\Phi_{\Lambda_N}(\underline\sigma')-U^\Phi_{\Lambda_M}(\underline\sigma'')|\le R\|\Phi\|,\tag{6.1.7}$$

by using (5.1.18), (5.2.1) and (5.2.3). So that, by (6.1.7) and (6.1.1),

e6.1.8

$$Z_{N+M}(\Phi)=Z_N(\Phi)Z_M(\Phi)e^{oR\|\Phi\|},\tag{6.1.8}$$

with $o\in[-1,1]$. If $N=M=2^k$ and $P_k=2^{-k}\log Z_{2^k}(\Phi)$, (6.1.8) says that

e6.1.9

$$P_k-2^{-(k+1)}R\|\Phi\|\le P_{k+1}\le P_k+2^{-(k+1)}R\|\Phi\|,\tag{6.1.9}$$

hence (6.1.9) imply that the limit as $k\to\infty$ of P_k exists.

Let $N=m2^k+d$, $0\le d<2^k$, $0\le m$; equation (6.1.8), repeatedly applied ($m+1$ times) yields

e6.1.10

$$Z_N(\Phi)=(Z_{2^k}(\Phi))^m Z_d(\Phi)\,e^{o\,(m+1)R\|\Phi\|},\tag{6.1.10}$$

with $o\in[-1,1]$. Then, at fixed k,

e6.1.11

$$\begin{aligned}N^{-1}\log Z_N(\Phi)=&\frac{m2^k}{m2^k+d}\,P_k+\frac{m+1}{m2^k+d}\,oR\|\Phi\|\quad+\\&+\frac{1}{m2^k+d}\,\log Z_d(\Phi)\xrightarrow[m\to\infty]{}P_k+2^{-k}oR\|\Phi\|.\end{aligned}\tag{6.1.11}$$

This shows, by the arbitrariness of k, the existence of the limit (6.1.2) and that it coincides with $\lim_{k\to\infty} P_k$.

Convexity is evident for $N^{-1}\log Z_N(\Phi)$ and hence for $P(\Phi)$. Lipschitz-continuity follows from (6.1.4) and from the existence of the limit (6.1.2). ■

P6.1.2 **(6.1.2) Proposition:** (Tangent plane to the graph of the pressure) *Under the hypotheses of the preceding proposition and if $\mu \in G(\Phi)$, the functional*

e6.1.12
$$\Psi \to \alpha(\Psi) = -\mu(A_\Psi) \tag{6.1.12}$$

is "tangent" to the graph of $\Phi \to P(\Phi)$. This means

e6.1.13
$$P(\Phi+\Psi) \geq P(\Phi) + \alpha(\Psi) \qquad \textit{for all } \Psi \in B. \tag{6.1.13}$$

Vice versa every functional α, tangent to the graph of the function $P(\Phi)$, has the form (6.1.12) with $\mu \in G(\Phi)$ and μ is uniquely determined by α.

Remark: The Gibbs states $\mu \in G(\Phi)$ have therefore the geometric interpretation of "tangent planes" to the graph of $\Phi \to P(\Phi)$. The function $P(\Phi)$ takes the role of *generating function* of the Gibbs states, via the relation

e6.1.14
$$\mu(A_\Psi) = -\frac{d}{d\varepsilon} P(\Phi+\varepsilon\Psi)\Big|_{\varepsilon=0}, \tag{6.1.14}$$

that is well defined, by the convexity of the function $\varepsilon \to P(\Phi+\varepsilon\Psi)$, at least where $G(\Phi)$ consists of a single point (and, hence, $\Phi \to P(\Phi)$ has a unique tangent plane).

This explains the importance attributed to the functional P which, because of its meaning in various problems of Statistical Mechanics, is called *pressure*.

Proof: As usual we shall suppose, for simplicity, that $T_{\sigma\sigma'} = 1\ \forall \sigma,\sigma'$.

If $\Lambda_N = [1,N]$ and if we recall definitions (5.1.2) and (5.1.3) (see in particular (5.1.9), (5.1.11) and (5.1.12)) the conditional probabilities of the distributions $\mu \in G(\Phi)$ can be expressed as

e6.1.15
$$p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c}) = \frac{\exp\Big(-\sum_{j=1}^N A_\Phi(\tau^j\underline{\sigma}) - D_{\Lambda_N,1}(\underline{\sigma})\Big)}{\sum_{\underline{\sigma}'}\exp\Big(-\sum_{j=1}^N A_\Phi(\tau^j\underline{\sigma}') - D_{\Lambda_N,1}(\underline{\sigma}')\Big)}, \tag{6.1.15}$$

with $\underline{\sigma} = (\underline{\sigma}_{\Lambda_N}\underline{\sigma}_{\Lambda_N^c})$, $\underline{\sigma}' = (\underline{\sigma}'_{\Lambda_N},\underline{\sigma}_{\Lambda_N^c}) \in \{0,\ldots,n\}^{\mathbb{Z}}$, see also (5.1.8). Here $D_{\Lambda_N,1}(\underline{\sigma})$ is a *boundary term* or, as we called it in definition (5.1.3), a *surface correction*. Define

e6.1.16
$$Z_{\Lambda_N}(\Phi,\underline{\sigma}_{\Lambda_N^c}) = \sum_{\underline{\sigma}'_{\Lambda_N}} \exp\Big(-\sum_{j=1}^N A_\Phi(\tau^j\underline{\sigma}') - D_{\Lambda_N,2}(\underline{\sigma}')\Big), \tag{6.1.16}$$

where $D_{\Lambda_N,2}(\underline{\sigma})$ is another surface correction (see again definition (5.1.3)). Given a sequence $u_i \in \mathbb{R}$, let $\overline{p}_i = e^{-u_i}/\sum_j e^{-u_j}$. Then from the inequality

e6.1.17
$$\begin{aligned}\log \sum_i e^{-u_i} = & \max_{\sum_i p_i=1,\, p_i\ge 0} \sum_i p_i(-\log p_i - u_i) = \\ & = \sum_i \overline{p}_i(-\log \overline{p}_i - u_i),\end{aligned} \tag{6.1.17}$$

we deduce

e6.1.18
$$\begin{aligned}\log Z_{\Lambda_N}(\Phi,\underline{\sigma}_{\Lambda_N^c}) = & \sum_{\underline{\sigma}'_{\Lambda_N}} p_\Phi(\underline{\sigma}'_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c})\Big(-\log p_\Phi(\underline{\sigma}'_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c}) - \\ & - \sum_{j=1}^N A_\Phi(\tau^j\underline{\sigma}') - D_{\Lambda_N,2}(\underline{\sigma}')\Big).\end{aligned} \tag{6.1.18}$$

To bound this expression in terms of $P(\Phi)$, we note that (5.1.12), (5.1.13) (5.1.14) and (5.1.16) imply the existence of a simple relation between $Z_N(\Phi)$ defined in (6.1.1) and $Z_{\Lambda_N}(\Phi,\underline{\sigma}_{\Lambda_N^c})$ in (6.1.16); namely

e6.1.19
$$Z_N(\Phi)/Z_{\Lambda_N}(\Phi,\underline{\sigma}_{\Lambda_N^c}) \le \exp\vartheta'\varepsilon_N, \tag{6.1.19}$$

with $\vartheta' \in [-1,1]$ and $\varepsilon_N = \varepsilon_{N,1}$ (see (5.1.14)) such that $\varepsilon_N/N \xrightarrow[N\to\infty]{} 0$.

Furthermore, if $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}}$, it is tautological that

e6.1.20
$$p_\Phi(\underline{\sigma}_{\Lambda_N}\,|\,\underline{\sigma}_{\Lambda_N^c}) = \int \mu(d\underline{\sigma}')\, p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c}) \tag{6.1.20}$$

(because the integrand is constant) but, by (5.2.5) (or, better, by the same computation that leads to (5.2.5) and without making use of the hypothesis $\|\Phi\|_1 < +\infty$) we see that

e6.1.21
$$\frac{p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}'_{\Lambda_N^c})}{p_\Phi(\underline{\sigma}_{\Lambda_N}\,|\,\underline{\sigma}_{\Lambda_N^c})} \le \exp\vartheta''\varepsilon'_N, \tag{6.1.21}$$

N6.1.2 with $\varepsilon'_N/N \xrightarrow[N\to\infty]{} 0$ and $\vartheta'' \in [-1,1]$.[2] Hence combining the tautology (6.1.20) with the bound (6.1.21) we get a bound for $p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}'_{\Lambda_N^c})$

e6.1.22
$$\begin{aligned}p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c}) \ge & \exp(-\varepsilon'_N)\int p_\Phi(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}'_{\Lambda_N^c})\mu(d\underline{\sigma}') \equiv \\ & \equiv \exp(-\varepsilon'_N)\mu(C^{\Lambda_N}_{\underline{\sigma}_{\Lambda_N}}).\end{aligned} \tag{6.1.22}$$

[2] ε'_N can be taken to depend only on N.

Then (6.1.19) and (6.1.18) say that

e6.1.23
$$\begin{aligned} N^{-1}\log Z_N(\Phi) \le N^{-1}\sum_{\underline{\sigma}'_{\Lambda_N}} p_\Phi(\underline{\sigma}'_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c})\Big(-\log\mu(C^{\Lambda_N}_{\underline{\sigma}'_{\Lambda_N}})-\\ -\sum_{j=1}^{N} A(\tau^j\underline{\sigma}')\Big)+\eta_N, \end{aligned} \tag{6.1.23}$$

with $\eta_N \xrightarrow[N\to\infty]{} 0$, uniformly in $\underline{\sigma}_{\Lambda_N^c}$.

Equation (6.1.23) can be thought of as an inequality between functions of
N6.1.3 $\underline{\sigma}_{\Lambda_N^c}$ and it can be integrated with respect to μ. We find [3]

e6.1.24
$$\begin{aligned} N^{-1}\log Z_N(\Phi) \le N^{-1}\sum_{\underline{\sigma}'_{\Lambda_N}} -\mu(C^{\Lambda_N}_{\underline{\sigma}'_{\Lambda_N}})\log\mu(C^{\Lambda_N}_{\underline{\sigma}'_{\Lambda_N}})-\\ -N^{-1}\sum_{j=1}^{N}\mu(A_\Phi)+\overline{\eta}_N, \end{aligned} \tag{6.1.24}$$

and, taking the limit as $N\to\infty$,

e6.1.25
$$P(\Phi)\le s(\mu)-\mu(A_\Phi), \tag{6.1.25}$$

where $s(\mu)$ is the average entropy of the probability distribution μ with respect to the translation, (cf. definition (3.3.2) and the theorem of the generator given in corollary (3.4.1)).

On the other hand, in general, by (6.1.17) and for an arbitrary τ–invariant probability distribution $m\in\mathcal{M}(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$ one has

e6.1.26
$$N^{-1}\log Z_N(\Phi) > N^{-1}\sum_{\underline{\sigma}_{\Lambda_N}} m(C^{\Lambda_N}_{\underline{\sigma}_{\Lambda_N}})(-\log m(C^{\Lambda_N}_{\underline{\sigma}_{\Lambda_N}})-U^\Phi_{\Lambda_N}(\underline{\sigma}_{\Lambda_N})), \tag{6.1.26}$$

and since

e6.1.27
$$\begin{aligned} \sum_{\underline{\sigma}_{\Lambda_N}} m(C^N_{\underline{\sigma}_{\Lambda_N}})U^\Phi_{\Lambda_N}(\underline{\sigma}_{\Lambda_N}) = \int m(d\underline{\sigma})U^\Phi_{\Lambda_N}(\underline{\sigma}_{\Lambda_N}) =\\ = \int m(d\underline{\sigma})\Big(\sum_{j=1}^{N} A_\Phi(\tau^j\underline{\sigma})+\Delta_N(\underline{\sigma})\Big) = N(m(A_\Phi)+\eta'_N) \end{aligned} \tag{6.1.27}$$

with $\eta'_N \xrightarrow[N\to\infty]{} 0$, one finds for all $m\in\mathcal{M}(\{0,\ldots,n\}^{\mathbb{Z}},\tau)$

e6.1.28
$$P(\Phi)\ge s(m)-m(A_\Phi). \tag{6.1.28}$$

[3] $\overline{\eta}_N$ is the integral of η_N.

By (6.1.25) and (6.1.28) we therefore obtain

e6.1.29
$$P(\Phi) = s(\mu) - \mu(A_\Phi) \qquad \text{for all } \mu \in G(\Phi). \tag{6.1.29}$$

Hence by (6.1.28) and (6.1.29)

e6.1.30
$$\begin{aligned} P(\Phi+\Psi) \ge s(\mu) - \mu(A_{\Phi+\Psi}) =& \\ = s(\mu) - \mu(A_\Phi) - \mu(A_\Psi) = P(\Phi) - \mu(A_\Psi)& \end{aligned} \tag{6.1.30}$$

which means that $\Psi \to \alpha(\Psi) = -\mu(A_\Psi)$ is a tangent functional.

To show the converse part of the proposition we make use of a general property of tangent planes to graphs of continuous and convex functions defined on a separable Banach space: the tangent plane is unique on a set which is at least dense and, furthermore, given $\Phi \in B$ and a functional α_Φ tangent to the graph of $\Phi \to P(\Phi)$ in the point Φ, one can find, for every $k \ge 0$, s_k points $\Phi_1^{(k)}, \Phi_2^{(k)}, \ldots, \Phi_{s_k}^{(k)}$, with $\|\Phi_j^{(k)} - \Phi\| < 1/k$, in which the graph of P has a unique tangent plane $\alpha_{\Phi^{(k)}}$ and s_k positive numbers $a_1^{(k)}$, $a_2^{(k)}, \ldots, a_{s_k}^{(k)}$ such that

e6.1.31
$$\begin{aligned} &\sum_{j=1}^{s_k} a_j^{(k)} = 1, \\ &\sum_{j=1}^{s_k} a_j^{(k)} \alpha_{\Phi_j^{(k)}}(\Psi) \xrightarrow[k\to\infty]{} \alpha(\Psi) \qquad \text{for all } \Psi \in B. \end{aligned} \tag{6.1.31}$$

N6.1.4 This is a general property of convex functionals [4] which can be applied to our case in the following way.

We begin by remarking that given a functional α and *assuming* that there is a probability distribution $\mu \in \mathcal{M}(\{0,\ldots,n\}^{\mathbb{Z}}, \tau)$ such that $\alpha(\Psi) = -\mu(A_\Psi)$, for all $\Psi \in B$, then μ is unique. The reason is that by choosing a suitable Ψ we can compute the μ–probability of a given cylinder $C^J_{\widehat{\underline{\sigma}}_J}$ with base on the set J.

The choice of the potential Ψ is such that $\Psi_X = 0$ unless the set X is a translate $\tau^k J$ of J for some $k \in \mathbb{Z}$ and

e6.1.32
$$\Psi_{\tau^k J}(\underline{\sigma}_J) = \delta_{\underline{\sigma}_J \widehat{\underline{\sigma}}_J}. \tag{6.1.32}$$

Then

e6.1.33
$$\mu(A_\Psi) = \sum_{X \ni 0} \frac{1}{|X|} \mu(\Psi_X(\underline{\sigma}_X)) \equiv \mu(C^J_{\widehat{\underline{\sigma}}_J}), \tag{6.1.33}$$

[4] In other words every tangent plane in Φ can be obtained as a limit of a sequence of planes that are (finite) convex linear combinations of planes tangent in points where the tangent plane is unique and whose largest distance to Φ is infinitesimal. For discussion and proof cf. p. 450 of [DS58] and p. 329 of [LR68].

that proves the statement, because μ is in it turn determined by its values on the cylinders.

Therefore given a tangent plane α to the graph of P in Φ one may be tempted to define the probability distribution μ, that should generate the plane α, by setting $\mu(A_\Psi) = -\alpha(\Psi)$.

The problem is that we do not know whether the numbers $\mu(C^J_{\underline{\sigma}_J})$, defined in this way in terms of α, and which should be μ–measures of cylinders, satisfy the properties (i), (ii), (iii) of definition (2.3.2), needed to reconstruct from them a probability distribution (cf. proposition (2.3.1)): note that condition (iv), *i.e.* the translation invariance, of definition (2.3.2) is in this case automatically satisfied.

However equation (6.1.31) says that, if $\mu(C^J_{\underline{\sigma}_J})$ has to be defined in terms of α by (6.1.33). Note that the already proved direct part of the proposition and the remark that α determines μ, if μ exists, imply that if there is a unique tangent plane in Φ then $G(\Phi)$ contains a unique point. From this follows that in the points $\Phi^{(k)}_j$ where the tangent plane is unique, one must necessarily have $\alpha_{\Phi^{(k)}_j}(\Psi) = \mu_{\Phi^{(k)}_j}(A_\Psi)$, for all $\Psi \in B$, where $\mu_{\Phi^{(k)}_j}$ is the unique element of $G(\Phi^{(k)}_j)$ This implies that

e6.1.34 $$\mu(C^J_{\underline{\sigma}_J}) = \lim_{k\to\infty} \sum_{j=1}^{s_k} a^{(k)}_j \mu_{\Phi^{(k)}_j}(C^J_{\underline{\sigma}_J}). \tag{6.1.34}$$

Since $\sum_j a^{(k)}_j = 1$ and the quantities $\mu_{\Phi^{(k)}_j}(C^J_{\underline{\sigma}_j})$ are values of measures of the cylinders $C^J_{\underline{\sigma}_J}$ with respect to some probability distribution $\mu_{\Phi^{(k)}_j}$ the r.h.s. of (6.1.34) verifies the (i), (ii), (iii) of definition (2.3.2) so that also $\mu(C^J_{\underline{\sigma}_J})$ verifies the same properties. Finally, by proposition (2.3.1), there exists $\mu \in \mathcal{M}(\{0,\ldots,\}^{\mathbb{Z}}, \tau)$ such that

e6.1.35 $$\alpha(\Psi) = -\mu(A_\Psi) \qquad \text{for all } \Psi \in B. \tag{6.1.35}$$

It remains to check that the distribution μ is in $G(\Phi)$.

For this purpose we remark that the first of (6.1.31) and the explicit definition of conditional probability for a Gibbs distribution (cf. Section §5.1, (5.1.9)), immediately imply

e6.1.36 $$\sup_j \sup_{\underline{\sigma}} |p_\Phi(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c}) - p_{\Phi^{(k)}_j}(\underline{\sigma}_\Lambda|\underline{\sigma}_{\Lambda^c})| \xrightarrow[k\to\infty]{} 0 \tag{6.1.36}$$

for all $\Lambda \subset \mathbb{Z}$. Furthermore (6.1.34) means that μ is the weak limit $\lim_{N\to\infty} \sum_{j=1}^k a^{(k)}_j \mu_{\Phi^{(k)}_j}$. Hence if $\underline{\sigma}_\Lambda \to f(\underline{\sigma}_\Lambda)$ is a $\mathcal{B}(\Lambda)$–measurable function we have

$$\mu(f) = \lim_{k\to\infty} \sum_{j=1}^{s_k} a^{(k)}_j \int \mu_{\Phi^{(k)}_j}(d\underline{\sigma}) f(\underline{\sigma}_\Lambda) =$$

$$= \lim_{k\to\infty} \sum_{j=1}^{s_k} a_j^{(k)} \sum_{\underline{\sigma}_\Lambda} \int \mu_{\Phi_j^{(k)}}(d\underline{\sigma}') p_{\Phi_j^{(k)}}(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) f(\underline{\sigma}_\Lambda) =$$

e6.1.37

$$= \lim_{k\to\infty} \sum_{\underline{\sigma}_\Lambda} \int \Big(\sum_{j=1}^{s_k} a_j^{(k)} \mu_{\Phi_j^{(k)}}(d\underline{\sigma}') \Big) p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) f(\underline{\sigma}_\Lambda) + \tag{6.1.37}$$

$$+ \lim_{k\to\infty} \sum_{\underline{\sigma}_\Lambda} \int \Big(\sum_{j=1}^{s_k} a_j^{(k)} \mu_{\Phi_j^{(k)}}(d\underline{\sigma}') \Big) \big(p_{\Phi_j^{(k)}}(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) - p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) \big) f(\underline{\sigma}_\Lambda) =$$

$$= \sum_{\underline{\sigma}_\Lambda} \int \mu(d\underline{\sigma}') p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}'_{\Lambda^c}) f(\underline{\sigma}_\Lambda),$$

by (6.1.36) and because $\underline{\sigma} \to p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c})$ is a continuous function (hence uniformly continuous). Equality between first and last term in (6.1.37) means that $p_\Phi(\underline{\sigma}_\Lambda | \underline{\sigma}_{\Lambda^c})$ is the conditional probability relative to the region Λ of the distribution μ: thus, by the arbitrariness of Λ, we see that $\mu \in G(\Phi)$. ∎

We implicitly obtained also the following results.

C6.1.1 **(6.1.1) Corollary:** (Variational principle for the pressure)
Under the hypotheses of proposition (6.1.1) the set $G(\Phi)$ consists of all probability distributions $\mu \in \mathcal{M}(\{0,\ldots,n\}_T^{\mathbb{Z}}, \tau)$ for which

e6.1.38

$$P(\Phi) = \max_{m\in\mathcal{M}(\{0,\ldots,n\}_T^{\mathbb{Z}},\tau)} (s(m) - m(A_\Phi)) = s(\mu) - \mu(A_\Phi). \tag{6.1.38}$$

Remark: This is the *variational principle* for the determination of the Gibbs distributions of given potential (Ruelle).

C6.1.2 **(6.1.2) Corollary:** (Decomposition into pure phases)
Under the hypotheses of proposition (6.1.1) the following properties hold.
*(i) If $\mu \in G(\Phi)$ then its ergodic decomposition π_μ (*cf. *Section §2.4) has support in $G_e(\Phi) = G(\Phi) \cap \mathcal{M}_e$.*
*(ii) The set $\mathcal{E}_e(\Phi) = \{$set of the ergodic points of $\{0,\ldots,n\}_T^{\mathbb{Z}}$ which generate a probability distribution in $G_e(\Phi)\}$ (*cf. *Section §2.3) is a Borel set and $\mu(\mathcal{E}_e(\Phi)) = 1$, for all $\mu \in G(\Phi)$.*
N6.1.5 *(iii) The extremal points of $G(\Phi)$ are the points of $G_e(\Phi)$.*[5]

Remark: A consequence of this corollary is an interesting interpretation of the pressure $P(\Phi)$ as complexity (cf. definition (3.1.1)). Let in fact D_N be a sequence of functions on $\{0,\ldots,n\}^N$, $N = 1, 2, \ldots$, such that

e6.1.39

$$\lim_{N\to\infty} \sup_{\sigma_1\ldots\sigma_N} N^{-1} |D_N(\sigma_1 \ldots \sigma_N)| = 0, \tag{6.1.39}$$

[5] If G is a closed convex set in a linear topological space we say that $\mu \in G$ is extremal if the relation $\mu = \alpha'\mu' + \alpha''\mu''$, with $\mu', \mu'' \in G$ and $\alpha' + \alpha'' = 1$, $\alpha', \alpha'' \in (0,1)$, implies $\mu' = \mu''$.

and set

e6.1.40
$$V_N(\sigma_1 \ldots \sigma_N) = U^{\Phi}_{\Lambda_N}(\sigma_1 \ldots \sigma_N) + D_N(\sigma_1 \ldots \sigma_N). \tag{6.1.40}$$

Then, if $\mu \in G(\Phi)$, μ–almost all points $\underline{\sigma}$ of $\{0, \ldots, n\}^{\mathbb{Z}}_T$ have complexity with weight $\underline{V} = \{V_N\}_{N=0}^{\infty}$ given by

e6.1.41
$$s(\underline{\sigma}, \{V_N\}_{N=0}^{\infty}) = P(\Phi) = s(\mu) - \mu(A_\Phi). \tag{6.1.41}$$

This is a consequence of the above corollaries and of the Shannon–McMillan theorem, see also problem [6.1.2].

Proof: The set $G(\Phi)$ is convex and compact, see corollary (5.1.1), and item (i) follows from the ergodic decomposition (proposition (2.4.1)) which implies that $\pi_\mu(\mathcal{M}_e \cap G(\Phi)) = 1$, if $\mu \in G(\Phi)$.
Item (ii) follows immediately from the results of problem [2.4.6].
Item (iii) can be proved by noting that the extremal points of $\mathcal{M}$ are the points in $\mathcal{M}_e$: see remark (5) to proposition (2.4.1).
Therefore assuming existence of a distribution μ, extremal in $G(\Phi)$ but not in $\mathcal{M}$ (*i.e.* not ergodic), one would find $\alpha \in (0,1)$ and two probability distributions in μ_1 and μ_2 with $\mu_1 \neq \mu_2$ and $\mu_2 \in \mathcal{M} \setminus G(\Phi)$ such that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$. Then $P(\Phi) = s(\mu) - \mu(A_\Phi) = \alpha(s(\mu_1)) - \mu_1(A_\Phi)) + (1-\alpha)(s(\mu_2) - \mu_2(A_\Phi)) < P(\Phi)$ by the convexity of the average entropy and by the variational principle in (6.1.38). Hence the extremal points of $G(\Phi)$ would be actually extremal also in $\mathcal{M}$ and therefore ergodic. The proof is therefore complete. ■

Problems for §6.1

Q6.1.1 **[6.1.1]:** Prove (6.1.41). (*Hint:* Consider the function

$$\widehat{f}_N(\underline{\sigma}) = -N^{-1}\log(\mu(C^N_{\underline{\sigma}_{\Lambda_N}})e^{-V_N(\underline{\sigma}_{\Lambda_N})})$$

and show that as $N \to \infty$ it converges in measure to $s(\mu) - \mu(A_\Phi)$ if $\mu \in G_e(\Phi)$; then make use of the argument given in the remark (2) to proposition (3.2.1))

Q6.1.2 **[6.1.2]:** (*Pressure of the Ising potential*)
Compute the pressure of the potential on $\{0,1\}^{\mathbb{Z}}$ such that $\Phi_X = 0$ unless X is a set which is a translation of a single point or of two nearest neighbors and

$$\Phi_{\{0\}}(\sigma) = \begin{cases} h & \sigma = 1 \\ -h & \sigma = 0 \end{cases}, \qquad \Phi_{\{0,1\}}(\sigma, \sigma') = J(2\sigma - 1)(2\sigma' - 1),$$

which is called "Ising potential in external field h." (*Hint:* : Remark that Z_N can be written as the scalar product $v \cdot \Xi^N v$, where Ξ is a suitable 2×2 matrix (*transfer matrix*) and $v = (1,1)$. $P(\Phi)$ is the linked to the eigenvalues of Ξ.)

Q6.1.3 **[6.1.3]:** Compute by means of the variational principle the probability $\mu(C^{00}_{11})$ with respect to the probability distribution of parameters J and h considered in problem [6.1.2].

Q6.1.4 **[6.1.4]:** (*Fisher potential*)
Like problem [6.1.3] for the *Fisher potential* introduced in problem [5.2.1].

Q6.1.5 **[6.1.5]:** (*Exact solubility of Fisher potentials*)
Compute the measure of the cylinders $C^{0\cdots k}_{1\cdots 1}$ for the Fisher potential of the preceding problem, using equation (6.1.33). The fact that an exact computation of the measure of *all* cylinders and of the pressure, entropy and other thermodynamic functions, is possible (and not difficult) makes the Fisher potentials an important tool for the analysis of examples and counterexamples.

Q6.1.6 **[6.1.6]:** (*Pressure singularities and non-uniqueness*)
Check that if $P(\Phi)$ is analytic in Φ for Φ in an open subset Δ in the space of the potentials, $\Delta \subset B$, then the tangent plane to Φ is unique in every point of Δ. (*Hint:* One says that a functional P defined on a Banach space B is analytic at a point Φ if given arbitrarily $k > 0$ and $\Psi_1, \ldots, \Psi_k \in B$ the function $\varepsilon_1 \ldots \varepsilon_k \to P(\Phi + \varepsilon_1 \Psi_1 + \ldots + \varepsilon_k \Psi_k)$ is analytic in $\varepsilon_1, \ldots, \varepsilon_k$ in a neighborhood of the origin.)

Q6.1.7 **[6.1.7]:** (*A Fisher potential with phase transition*) Show that if $P(\Phi + \varepsilon\Psi)$ has right derivative different from the left derivative, as function of ε, at $\varepsilon = 0$ then $G(\Phi)$ contains at least two different points (one says that a "*phase transition*" takes place as ε varies).

Q6.1.8 **[6.1.8]:** Show the existence of Fisher potentials with $\|\Phi\|_1 = +\infty$ (cf. problem [6.1.4]) for which P is not an analytic function of the "one-body" component of Φ (*i.e.* of $\Phi_{\{0\}}(1) \equiv -\Phi_0$) and that $P(\Phi)$ has right derivative different from the left derivative at a suitable values of Φ_0. Check that this cannot happen if $\|\Phi\|_1 < +\infty$. (*Hint:* Consider the case $\Phi_n = (1+n)^{-1-\varepsilon}$, with $0 < \varepsilon < 1$, for $n \geq 1$ and Φ_0 arbitrary; then P is not analytic as a function of Φ_0, by explicit computation. This requires having solved problem [6.1.5].)

Bibliographical note to §6.1

The contents of this section are taken from various sources, for instance from Ch. 7, §3,4 of [Ru69], p. 450 of [DS58], [LR68]. See also the bibliographical note to §5.1. The theory of the pressure regarded as a generating function of a Gibbs state is ancient, some examples of "rigorous" applications of this elegant technique are in [Ru72].

§6.2 Applications to Anosov systems. SRB distribution

A classical application of the theory of Gibbs distributions concerns the analysis of the invariant probability distributions associated with an Anosov system (Ω, S) on a compact Riemannian manifold (Sinai).

The definition that we adopted of Anosov system, cf. definition (4.2.1), assumes topological transitivity and implies topological mixing, cf. problem [4.2.10], *i.e.* that, given two open sets $G,\, F \subset \Omega$ there exists $k_0 > 0$ such that if $k \geq k_0$ then $S^k G \cap F \neq \emptyset$. We shall rely on the latter properties although several results could be suitably extended to cover much more general hyperbolic systems. In agreement with the notations introduced in definition (2.2.2) we shall recall or set the following definitions.

D6.2.1 **(6.2.1) Definition:** (Topological and ergodic distributions)
We denote
(i) $\mathcal{M}(\Omega, S)$ *the set of the* S*–invariant Borel distributions on* Ω*,*
(ii) $\mathcal{M}_e(\Omega, S) \subset \mathcal{M}(\Omega, S)$ *the set of the distributions on* Ω *which are* S*–ergodic,*

(iii) $\mathcal{M}_e^t(\Omega,S)\subset\mathcal{M}_e(\Omega,S)\subset\mathcal{M}(\Omega,S)$ *the set of the* S*–ergodic distributions which give a positive measure to all open sets.*
More generally $\mathcal{M}^{0,t}(\Omega)$ *will be the set of* topological distributions*,* cf. *definition (4.1.2), on* Ω*.*

The aim of the following analysis is to show that Anosov systems admit *"very many" probability distributions in* $\mathcal{M}_e^t(\Omega,S)$ and to show that, among them, one can identify several ones either on the basis of a criterion of "maximum complexity" or by attibuting a special role to the statistical properties of motions whose initial data are chosen randomly with a probability distribution which is absolutely continuous with respect to the volume.

*D*6.2.2 **(6.2.2) Definition:** (SRB distributions and f–complex distributions)
Given an Anosov map (Ω,S)
(i) A topological probability distribution $\mu\in\mathcal{M}(\Omega,S)$ *is said to have maximum complexity with respect to a continuous function* f *if it maximizes the quantity* $s(\mu)-\mu(f)$*.*
(ii) If μ_0 *is a probability distribution absolutely continuous with respect to the volume measure on* Ω *and if for all continuous functions* F *on* Ω *and for* μ_0*–almost all initial data* $x\in\Omega$ *the limit* $\lim_{N\to\infty}N^{-1}\sum_{j=0}^{N-1}F(S^jx)$ *esists and can be written as*

6.2.0
$$\lim_{N\to\infty}N^{-1}\sum_{j=0}^{N-1}F(S^jx)=\int_\Omega\mu(dy)F(y) \tag{6.2.1}$$

with $\mu\in\mathcal{M}(\Omega,S)$*, then* μ *is called the SRB distribution of* (Ω,S)*.*

Remark: (1) The results of this section are general results preparing the key results of this Chapter: namely that all Anosov systems admit a SRB distribution, proposition (6.3.3) and corollary (6.3.1), which furthermore is the invariant probability distribution which maximizes complexity with respect to the continuous function $f(x)=\lambda_u(x)$, proposition (6.3.4).
(2) Technically the above properties will be closely related to the results of this section which allow us to identify probability distributions that maximize complexity with suitable Gibbs distributions with short range potentials, see lemma (6.2.1), and with properties of periodic motions, see proposition (6.2.1).

The notion of complexity has been discussed in Section §3.1, where we have given the definitions of complexity of sequences of symbols $\underline{\sigma}$ with respect to families $\{U_N\}_{N\ge0}^\infty$ of weights, *i.e.* of functions defined on strings of length N in $\underline{\sigma}\in\{0,\ldots,n\}_T^N$.

Given a potential $\Phi\in B$, cf. definition (5.1.1), corollary (6.1.1) shows that a distribution μ on $\{0,\ldots,n\}_T^{\mathbb{Z}}$ that maximizes the complexity of μ–almost all the points relatively to a weight $U_N(\underline{\sigma})=\sum_{R\subset\{1,\ldots,N\}}\Phi_R(\underline{\sigma}_R)$ is a Gibbs distribution with potential Φ; such complexity is, then, equal to the pressure

*e*6.2.2
$$P(\Phi)=\max_{\mu'\in\mathcal{M}_e(\{0,\ldots,n\}_T^{\mathbb{Z}})}(s(\mu')-\mu'(A_\Phi)). \tag{6.2.2}$$

In Anosov systems, as seen in Section §4.2, Markovian pavements, which always exist, allow us to describe points of Ω by means of symbolic sequences and their evolution by the shift map.

One is thus led to the search of invariant and topological distributions on Ω via symbolic dynamics. At the same time one would like to characterize the distributions that in the symbolic language become Gibbs distributions in terms of quantities that do not depend explicitly on specific Markovian pavements.

Given the one-to-one correspondence between the S–ergodic topological distributions on Ω and the τ–ergodic topological distributions on $\{1,\ldots,q\}_T^{\mathbb{Z}}$ discussed in proposition (4.1.1) and in the relative remark (2) one is tempted, for instance, to associate with (6.2.2) the "intrinsic", *i.e.* Markovian pavement independent, problem of finding the distributions μ' that maximize

e6.2.3
$$s(\mu')-\mu'(f) \tag{6.2.3}$$

on $\mathcal{M}_e^t(\Omega,S)$ with f a continuous function on Ω.

As we shall see, if f is Hölder continuous, *i.e.* if it verifies a "modest" requirement of regularity beyond the mere continuity, it will be possible to give a quite satisfactory answer to the above questions.

Let (Ω,S) be an Anosov system. Denote $\mathcal{P}=\{P_1,\ldots,P_q\}$ a Markovian pavement constructed with S–rectangles as in Section §4.2: let T be its compatibility matrix (cf. Section §4.1). Then (Ω,S) is mixing and, hence, there exists $a\ge 0$ such that $(T^k)_{\sigma\sigma'}>0$ for all $k>a$.

Denote by $\underline{\sigma}\to X(\underline{\sigma})$ the Hölder continuous code of $\{1,\ldots,q\}_T^{\mathbb{Z}}$ into Ω discussed in Section §4.1 defined by

e6.2.4
$$x=X(\underline{\sigma})=\bigcap_{j=-\infty}^{+\infty} S^{-j}P_{\sigma_j}. \tag{6.2.4}$$

The following preliminary result holds.

L6.2.1 **(6.2.1) Lemma:** *Let (Ω,S) be an Anosov system and let $\mathcal{P}=\{P_0,\ldots,P_q\}$ be a Markovian pavement for it with compatibility matrix T.*
(i) If f is Hölder continuous the function $\mu\to s(\mu)-\mu(f)$ admits a maximum $p(f)$ on $\mathcal{M}_e^t(\Omega)$.
*(ii) Let $\nu_N\in\mathcal{M}^{0,t}(\Omega)$ be an arbitrary sequence of topological distributions (*cf.* definition (6.2.1)) on Ω and, given $\underline{\sigma}\in\{1,\ldots,q\}_T^{[0,N]}$, let*

e6.2.5
$$U_N(\sigma_0\ldots\sigma_N)=\frac{\int_{P^{0\ldots N}_{\sigma_0\ldots\sigma_N}}\sum_{j=0}^N f(S^jx)\nu_N(dx)}{\int_{P^{0\ldots N}_{\sigma_0\ldots\sigma_N}}\nu_N(dx)}, \tag{6.2.5}$$

where $P^{0\ldots N}_{\sigma_0\ldots\sigma_N}=\bigcap_{j=0}^{N}S^{-j}P_{\sigma_j}$. Then the image m of $\mu\in\mathcal{M}_e^t(\Omega,S)$ via the isomorphism mod 0 *defined by (6.2.4) maximizes the complexity with weight $\{U_N\}_{N=1}^{\infty}$ if and only if μ maximizes the function $\mu\to s(\mu)-\mu(f)$ on $\mathcal{M}_e^t(\Omega,S)$.*

Remark: Note that the sequence ν_N and the Markovian pavement $\mathcal{P}$ used to formulate this theorem are in a certain sense irrelevant: the image m of the probability distribution $\mu \in \mathcal{M}^t_e(\Omega, S)$, that maximizes $s(\mu) - \mu(f)$, makes largest the complexity associated with U_N for any choice of $\mathcal{P}$ or of the sequence ν_N that allows us to construct U_N. Hence it is natural to call $s(\mu) - \mu(f) = p(f)$ the *complexity with weight* f of μ on Ω: this is the interest of the statement (ii).

Proof: Let a be the mixing time of T. Let $\underline{\rho} \in \{1, \ldots, q\}^{\mathbb{Z}}_T$ be a T–compatible sequence, cf. definition (4.1.1), chosen once and for all.
We first need to define a way to merge together two compatible strings into a longer compatible string. This can be done in several "standard ways" provided the compatibility matrix T is mixing.

Suppose that $a \ge 1$ (*i.e.* we do not assume that the compatibility relation is trivial). Associate with each pair of symbols σ, σ' a T–compatible string $(\sigma\eta_0 \ldots \eta_{a-1}\sigma')$ and let $\underline{\vartheta}(\sigma, \sigma') = (\eta_0 \ldots \eta_{a-1})$.
Given such a correspondence ϑ, which we shall call an *interpolation string*, for any two finite T–compatible strings $\underline{\sigma} = (\sigma_1 \ldots \sigma_p)$ and $\underline{\sigma}' = (\sigma'_1 \ldots \sigma'_q)$ we can form the compatible strings

$$\underline{\sigma}\,\underline{\vartheta}(\sigma_p, \sigma'_1)\,\underline{\sigma}' \qquad \text{and} \qquad \underline{\sigma}'\,\underline{\vartheta}(\sigma'_q, \sigma_1)\,\underline{\sigma},$$

where, as usual, we indicate with $\underline{\sigma}\,\underline{\vartheta}(\sigma_p, \sigma'_1)\,\underline{\sigma}'$ the string starting with $\underline{\sigma}$ continuing with $\underline{\vartheta}(\sigma_p, \sigma'_1)$ and ending with $\underline{\sigma}'$. If $\underline{\sigma}$ is semiinfinite only one of the above "mergers" is possible and if both $\underline{\sigma}, \underline{\sigma}'$ are semiinfinite one merger is possible provided $\underline{\sigma}$ is infinite to the left and $\underline{\sigma}'$ is infinite to the right or *vice versa*. We shall simply denote $\underline{\sigma} \circ \underline{\vartheta} \circ \underline{\sigma}'$ the merger of $\underline{\sigma}$ and $\underline{\sigma}'$ with an interpolation string ϑ.
In the continuation of the proof just started we imagine that the set of q^2 interpolating strings $\underline{\vartheta}$ has been fixed once and for all. If $\underline{\sigma}, \underline{\rho} \in \{1, \ldots, q\}^{\mathbb{Z}}_T$ and $N \in \mathbb{N}$ we shall denote by $(\underline{\sigma}^{(N)}\underline{\rho})$ a sequence whose labels for $i \in [-N, N]$ coincide with σ_i and for $i \notin (-N-a, N+a)$ coincide with ρ_i, while for the other i's the sequence $(\underline{\sigma}^{(N)}, \underline{\rho})$ is interpolated by the appropriate interpolating strings $\underline{\vartheta}$.

If $f \in C^\alpha(\Omega)$, as already remarked elsewhere, (cf. remark (3) to definition (4.3.2)), the quantity $\underline{\sigma} \to f(X(\underline{\sigma}))$ can be well developed in terms of cylindrical functions starting from the identity

e6.2.6
$$f(X(\underline{\sigma})) = f(X(\underline{\rho})) + \sum_{N=0}^{\infty} [f(X(\underline{\sigma}^{(N)}\underline{\rho})) - f(X(\underline{\sigma}^{(N-1)}\underline{\rho}))], \tag{6.2.6}$$

where $f(X(\underline{\sigma}^{(-1)}\underline{\rho})) \stackrel{def}{=} f(X(\underline{\rho}))$. In fact set

e6.2.7
$$\begin{aligned} \Gamma_Y(\underline{\sigma}_Y) &= 0 \qquad Y \ne [-N, N] \text{ for some } N \ge 0 \text{ and} \\ \Gamma_Y(\underline{\sigma}_Y) &= \big(f(X(\underline{\sigma}^{(N)}\underline{\rho})) - f(X(\underline{\sigma}^{(N-1)}\underline{\rho}))\big), \end{aligned} \tag{6.2.7}$$

where in the second line we suppose that Y is the interval $Y = [-N, N]$ centered at 0. Then we rewrite (6.2.6) as

e6.2.8
$$f(X(\underline{\sigma})) = f(X(\underline{\rho})) + \sum_{Y \ni 0} \Gamma_Y(\underline{\sigma}_Y), \tag{6.2.8}$$

and we note that, if $C_{f,\alpha}$ is the Hölder continuity modulus of $f \in C^\alpha(\Omega)$ while α' is the Hölder continuity exponent of the code X (cf. definition (4.1.3) and (4.1.5)) and C' is the relative modulus, one has

e6.2.9
$$|\Gamma_Y(\underline{\sigma}_Y)| \le C_{f,\alpha} C'^{\alpha} \exp(-\alpha\alpha'(N-1)) = \widetilde{C} e^{-b\,\mathrm{diam}(Y)}, \tag{6.2.9}$$

with $\widetilde{C}, b > 0$. Therefore if we set

e6.2.10
$$\begin{aligned} &\Phi_X(\underline{\sigma}_X) = \Gamma_X(\tau^{-i}\underline{\sigma}_X) &&\text{if } X = \tau^i([-N,N]) \text{ for some } N \text{ and } i,\\ &\Phi_X(\underline{\sigma}_X) = 0 &&\text{otherwise,} \end{aligned} \tag{6.2.10}$$

and

e6.2.11
$$A_\Phi(\underline{\sigma}) = \sum_{X \ni 0} \frac{\Phi_X(\underline{\sigma}_X)}{|X|}, \tag{6.2.11}$$

we find that Φ is a potential in the space B introduced in definition (5.1.1), because Φ verifies the condition of the problem [5.3.7] of Section §5.3 (*exponential decay of the potential*), so that $||\Phi||_1 < \infty$. Hence one deduces

e6.2.12
$$\Big| \sum_{j=0}^{N-1} \Big(f(S^j X(\underline{\sigma})) - f(S^j X(\underline{\rho})) \Big) - \sum_{j=0}^{N-1} A_\Phi(\tau^j \underline{\sigma}) \Big| \le C'' \tag{6.2.12}$$

for every $\underline{\sigma} \in \{1, \ldots, q\}_T^{\mathbb{Z}}$ and for all $N > 0$, if C'' is suitably chosen. We will choose, if possible, $\underline{\rho}$ such that $\tau^j \underline{\rho} = \underline{\rho}$, *i.e.* $\rho = (\ldots i\, i\, i \ldots)$ for some label $i \in \{0, \ldots, q\}$. However the latter sequence may fail to be compatible: in such a case we choose $\underline{\rho}$ to be a periodic sequence of period a, which exists because of the mixing property of T. Below we suppose, for simplicity, that $\rho = (\ldots i\, i\, i \ldots)$ for some label $i \in \{0, \ldots, q\}$ is a compatible sequence.

From corollary (6.1.2) and (5.1.1), (5.1.17), it follows that there exists an ergodic distribution m on $\{1, \ldots, q\}_T^{\mathbb{Z}}$ which maximizes the complexity with weight $\{U_N\}_{N=1}^\infty$ because by (6.2.5) and (6.2.12)

e6.2.13
$$\Big| U_N(\underline{\sigma}) - \sum_{R \subset [0,\ldots,N]} \Phi_R(\underline{\sigma}_R) \Big| \le C''' \quad \text{for all } N, \tag{6.2.13}$$

cf. remark to lemma (6.2.1). Therefore from the general theory of Gibbs distributions, see proposition (5.2.1), it follows that $m \in \mathcal{M}_e^t(\{1, \ldots, q\}_T^{\mathbb{Z}}, \tau)$.

Let now μ' be the distribution isomorphic mod 0, via the isomorphism X, to a given distribution m' arbitrarily chosen in $\mathcal{M}_e^t(\{1, \ldots, q\}_T^{\mathbb{Z}}, \tau)$ (cf.

proposition (4.1.1) and (4.1.9)). Then $\mu' \in \mathcal{M}_e^t(\Omega, S)$ and furthermore $s(\mu') = s(m')$, since entropy is invariant under isomorphisms mod 0, and

e6.2.14

$$\begin{aligned}\mu'(f - f(X(\underline{\rho}))) &= \lim_{N\to\infty} \mu'\Big(N^{-1}\sum_{j=0}^{N-1}[f(S^j(\cdot)) - f(X(\underline{\rho}))]\Big) = \\ &= \lim_{N\to\infty} m'\Big(N^{-1}\sum_{j=0}^{N-1}[f(X(\tau^j\cdot)) - f(X(\underline{\rho}))]\Big) = \\ &= \lim_{N\to\infty} m'\Big(N^{-1}\sum_{j=0}^{N-1} A_\Phi(\tau^j\cdot)\Big) = m'(A_\Phi),\end{aligned} \tag{6.2.14}$$

and therefore, for all $\mu' \in \mathcal{M}_e^t(\{1, \ldots, q\}_T^{\mathbb{Z}}, \tau)$,

e6.2.15

$$s(\mu') - \mu'(f) = -f(X(\underline{\rho})) + s(m') - m'(A_\Phi), \tag{6.2.15}$$

so that the functional on $\mathcal{M}_e^t(\Omega, S)$ in the l.h.s. of (6.2.15) has a maximum which is reached, by the variational principle in corollary (6.1.1), on the $X-$ image of the Gibbs distribution $m \in G_e(\Phi)$. Item (ii) is thus also proved. ■

It is convenient to remark that lemma (6.2.1), or better its proof, also yields an explicit method to study the distributions μ that maximize $s(\mu) - \mu(f)$: indeed in deriving (6.2.10) we have shown that such distributions can be considered as Gibbs distributions with a suitable potential Φ_f which satisfies the hypotheses of problems [5.3.15] and [5.3.7], (*i.e. exponential decay*), if $f \in C^\alpha(\Omega)$. This implies the following result.

C6.2.1 **(6.2.1) Corollary:** (Variational principle for pressure of smooth functions)
If $f \in C^\alpha(\Omega)$ there exists a unique distribution that maximizes $s(\mu) - \mu(f)$ over $\mu \in \mathcal{M}_e^t(\Omega)$. Such a distribution μ_f is mixing and

e6.2.16

$$\mu_f(FS^kG) = \int \mu_f(dx)F(x)G(S^kx) \xrightarrow[k\to\infty]{} \mu_f(F)\mu_f(G), \tag{6.2.16}$$

with exponential rate, for $k \to \infty$, if F, G are Hölder continuous functions on Ω.

Proof: We only have to check the exponential decay, which, however, follows by the same argument seen in (5.4.22). ■

An interesting "intrinsic" method to compute $p(f)$ and μ from the stability properties of the periodic points is given by the following proposition.

P6.2.1 **(6.2.1) Proposition:** (Periodic orbits expansion)
(i) The quantity

e6.2.17

$$p(f) = \max_{\mu\in\mathcal{M}_e^t(\Omega,S)} (s(\mu) - \mu(f)), \tag{6.2.17}$$

with $f \in C^\alpha(\Omega)$, can be computed as

e6.2.18
$$p(f) = \lim_{N\to\infty} N^{-1} \log \sum_{x\in\Omega,\, S^N x = x} e^{-\sum_{j=0}^{N-1} f(S^j x)}. \tag{6.2.18}$$

(ii) If μ_f is the point where the maximum in (6.2.17) is reached, μ_f can be computed as a weak limit as $N \to \infty$ of the distributions on Ω

e6.2.19
$$\mu_N(dy) = \frac{\sum\limits_{S^N x = x} \Big(\exp\big(-\sum\limits_{j=0}^{N-1} f(S^j x)\big)\Big)\delta_x(dy)}{\sum_{S^N x = x} \Big(\exp\big(-\sum_{j=0}^{N-1} f(S^j x)\big)\Big)}, \tag{6.2.19}$$

where δ_x is the Dirac measure on x.

Proof: Let $\mathcal{Q}$ be a Markovian pavement: we shall denote by $\pi_N \equiv \pi_N(\Omega, S)$ the set of the points of Ω periodic with period N and by $\widehat{\pi}_N \equiv \widehat{\pi}_N(\Omega, S)$ the set of the points of $\{1, \ldots, q\}_T^{\mathbb{Z}}$ periodic with period N. Then

e6.2.20
$$X(\widehat{\pi}_N) \subseteq \pi_N, \tag{6.2.20}$$

and to every $x \in \pi_N$ there can correspond at most d elements $\underline{\sigma} \in \widehat{\pi}_N$ such that $X(\underline{\sigma}) = x$, if d is the multiplicity M of the code, cf. definition (4.1.3). But there could exist elements of $\widehat{\pi}_N$ that are not symbolic representations of elements of π_N because the compatible histories of a point of period N might have a period multiple of N (as a consequence of the ambiguity of the coding of points that are on the boundaries of the elements of the Markovian pavement). Hence we can write the inequality

e6.2.21
$$\sum_{x\in\pi_N} e^{-\sum_{j=0}^{N-1} f(S^j x)} \ge d^{-1} \sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-\sum_{j=0}^{N-1} f(X(\tau^j \underline{\sigma}))}, \tag{6.2.21}$$

which, by (6.2.12), implies

e6.2.22
$$\sum_{x\in\pi_N} e^{-\sum_{j=0}^{N_1} f(S^j x)} \ge d^{-1} \sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-C'' - N f(X(\underline{\rho})) - \sum_{j=0}^{N-1} A_\Phi(\tau^j \underline{\sigma})}. \tag{6.2.22}$$

This is a useful information because we can check that

e6.2.23
$$\lim_{N\to\infty} N^{-1} \log \sum_{\underline{\sigma}\in\widehat{\pi}_N} \exp\Big(-\sum_{j=0}^{N-1} A_\Phi(\tau^j \underline{\sigma})\Big) = P(\Phi); \tag{6.2.23}$$

indeed such a limit is certainly $\le P(\Phi)$ by virtue of (6.1.2) and of (5.1.12), (5.1.14) and (5.1.16), and because it is possible to establish, in an obvious way, a one-to-one correspondence between $\widehat{\pi}_N$ and a subset of $\{1, \ldots, q\}_T^N$.

On the other hand, since T is a mixing matrix with mixing time a, every configuration of $\{1,\ldots,q\}_T^{N-a}$ can be continued into a sequence in $\widehat{\pi}_N$, according to a prefixed rule to determine the interpolating strings $\underline{\vartheta}$. If $\underline{\sigma}'$ is any compatible string of length $N-a$, $\underline{\sigma}'=(\sigma'_0,\ldots,\sigma'_{N-a-1})$ we simply extend it to the right by the string $\underline{\vartheta}(\sigma'_{N-a-1},\sigma'_0)$ so that the string $\underline{\sigma}'\circ\underline{\vartheta}$ thus constructed has length N and can be indefinitely repeated to form a period N infinite sequence $\widetilde{\sigma}'$. Thus one sees that, with the notations of Section §6.1, see beginning of the proof of (6.1.2), there is a constant $\widehat{C}>0$ such that

e6.2.24
$$\Big|U^{\Phi}_{[0,1,\ldots,N-a-1]}(\underline{\sigma}')-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\widetilde{\sigma}}')\Big|\le\widehat{C},\tag{6.2.24}$$

so that

e6.2.25
$$\sum_{\underline{\sigma}'\in\{1,\ldots,n\}^{N-a}}e^{-U^{\Phi}_{[0,1,\ldots,N-a-1]}(\underline{\sigma}')}\le\sum_{\underline{\sigma}'\in\widehat{\pi}_N}e^{\widehat{C}-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma}')},\tag{6.2.25}$$

which gives, together with (6.1.1), the inverse inequality necessary to deduce (6.2.23).

Having established (6.2.23), we note that (6.2.22), (6.2.23) and (6.1.29) imply, if $m\in G(\Phi)$ and μ is its image via X,

e6.2.26
$$\begin{aligned}&\lim_{N\to\infty}N^{-1}\log\sum_{x\in\pi_N}e^{-\sum_{j=0}^{N-1}f(S^jx)}\ge\\&\qquad\ge-f(X(\underline{\rho}))+s(m)-m(A_\Phi)=s(\mu)-\mu(f),\end{aligned}\tag{6.2.26}$$

having also used the general relation (6.2.15) involving a distribution $m'\in\mathcal{M}^t_e\ (\{1,\ldots,q\}^{\mathbb{Z}}_T,\tau)$ and its X–image μ'.

To find the inequality opposite to (6.2.26) we proceed in an analogous way. Assuming for simplicity that N is even we associate with every $x\in\pi_N$ several elements $\underline{\sigma}\in\widehat{\pi}_{N+2a}$ such that

e6.2.27
$$d(X(\underline{\sigma}),x)\le\overline{C}e^{-\overline{\alpha}N},\tag{6.2.27}$$

for suitable $\overline{C},\ \overline{\alpha}>0$. The rule that allows us to construct this correspondence is the following. Let $\underline{\widehat{\sigma}}$ be one of the (up to d) elements in $\{1,\ldots,q\}^{\mathbb{Z}}_T$ such that $X(\underline{\sigma})=x$. Set

e6.2.28
$$\sigma_i=\widehat{\sigma}_i\qquad\forall|i|\le N/2,\tag{6.2.28}$$

and define $\sigma_{N/2+1},\ldots,\sigma_{N/2+a}$, $\sigma_{-N/2-1},\ldots,\sigma_{-N/2-a}$ so that

e6.2.29
$$\sigma_{\pm(N/2+a)}=0,\qquad\prod_{j=N/2}^{N/2+a-1}T_{\sigma_j\sigma_{j+1}}T_{\sigma_{-j}\sigma_{-j-1}}=1.\tag{6.2.29}$$

This is possible by the mixing property of T and in up to q^{2a} different ways.

The sequence $\underline{\sigma}$, defined in this way for $i \in [-N/2-a,\, N/2+a]$ is then periodically extended into a periodic sequence with period $N+2a$: note that this is now possible because $\sigma_{-N/2-a} = \sigma_{N/2+a}$ by (6.2.29).

The construction shows that with every $x \in \pi_N$ we can associate (up to q^{2a}) elements $\underline{\sigma} \in \widehat{\pi}_{2N/2+2a}$ and, furthermore, such classes of elements are pairwise disjoint (*i.e.* to different x necessarily there correspond different sequences $\underline{\sigma}$).

Furthermore (6.2.28) implies (6.2.27) by the Hölder continuity of the code, and (6.2.27) in turn means that if $\underline{\sigma}$ is associated with an x by the preceding construction there is a constant $\overline{C}' > 0$ such that

e6.2.30
$$\Big|\sum_{j=0}^{N-1} f(S^j x) - \sum_{j=0}^{N-1} f(X(\tau^j \underline{\sigma}))\Big| \le \overline{C}', \tag{6.2.30}$$

because f is Hölder continuous and, by (6.2.27), there exists $\overline{C}'' > 0$ such that

e6.2.31
$$\sum_{j=0}^{N-1} d(S^j x, S^j X(\underline{\sigma}))^\alpha \le \overline{C}'' \tag{6.2.31}$$

(note that the distance between $S^j x$ and $S^j X(\underline{\sigma})$ begins to grow substantially only when $|j|$ is close to $N/2$, by the construction of $\underline{\sigma}'$). Hence by (6.2.30)

e6.2.32
$$\begin{aligned}&\sum_{\underline{\sigma}\in\widehat{\pi}_{N+2a}} e^{-\sum_{j=0}^{N+2a-1} f(S^j X(\underline{\sigma}))} \ge \\ &\ge e^{-\overline{C}'-2a\|f\|_\infty} \cdot \sum_{x\in\pi_N} e^{-\sum_{j=0}^{N-1} f(S^j x)}.\end{aligned} \tag{6.2.32}$$

This inequality implies, by (6.2.23), the validity of the inequality opposite to (6.2.26), and (6.2.18) is proved. This shows that the limit (6.2.18) exists and equals $P(\Phi) - f(X(\underline{\rho}))$: in particular the latter expression does not depend on the Markovian pavement $\mathcal{Q}$ used, in spite of the fact that both Φ and the code X do depend on $\mathcal{Q}$.

To show the validity of item (ii) call $\overline{\mu}$ a weak limit of the sequence of distributions in (6.2.19). Since the distributions μ_N in (6.2.19) are invariant,[1] the limit of any convergent subsequence of μ_N will be S–invariant and hence $\overline{\mu}$ will be S–invariant. N6.2.1

To check that $\overline{\mu} = \mu$ it will suffice to show that $\overline{\mu}$ is absolutely continuous with respect to μ: then the ergodicity of μ will imply $\overline{\mu} = \mu$. The absolute continuity will follow (since μ and $\overline{\mu}$ are regular Borel measures) from an inequality of the type $\overline{\mu}(|F|) \le K\mu(|F|)$, for F that varies in a set which spans densely in $C(\Omega)$. For example for every Hölder continuous F.

---

[1] Observe that if $F(x) = \sum_{j=0}^{N-1} f(x)$ and $S^N x = x$ then $F(Sx) = F(x)$. Moreover the set of points of period N is clearly S-invariant.

Hence we shall fix $1>\beta>0$ and show the existence of $K>0$ such that $\overline{\mu}(F)\le K\mu(F)$ for all $F\ge 0$, $F\in C^\beta(\Omega)$. We have

e6.2.33
$$\begin{aligned}
\mu_N(F)=&\frac{\sum_{x\in\pi_N} e^{-\sum_{j=0}^{N-1} f(S^jx)}F(x)}{\sum_{x'\in\pi_N} e^{-\sum_{j=0}^{N-1} f(S^jx')}}\ \le\ \text{by (6.2.22)}\ \le\\
&\le d\,e^{C''}\,\frac{\sum_{x\in\pi_N} e^{-\sum_{j=0}^{N-1} f(S^jx)}F(x)}{\sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-Nf(X(\rho))-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma})}}\le\\
&\le\ \text{by (6.2.27), (6.2.30), (6.2.32)}\ \le de^{C''+\overline{C}'+2a\|f\|_\infty}\cdot\\
&\cdot\frac{\sum_{\underline{\sigma}\in\widehat{\pi}_{N+2a}} e^{-\sum_{j=0}^{N+2a-1} f(X(\tau^j\underline{\sigma}))}\big(F(X(\underline{\sigma}))+\,\overline{\overline{C}}e^{-\,\overline{\overline{\alpha}}N}\big)}{\sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{Nf(X(\underline{\rho}))-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma})}},
\end{aligned}\tag{6.2.33}$$

having also used the Hölder continuity of F to define naturally $\overline{\overline{C}}$ and $\overline{\overline{\alpha}}$. By (6.2.12), the (6.2.33) can be further bounded above by

e6.2.34
$$\begin{aligned}
\mu_N(F)\le de^{2C''+\overline{C}'+2a\|f\|_\infty}&\sum_{\underline{\sigma}\in\widehat{\pi}_{N+2a}} e^{-(N+2a)f(X(\underline{\rho}))}\cdot\\
&\cdot\frac{e^{-\sum_{j=0}^{N+2a-1}A_\Phi(\tau^j\underline{\sigma})}\big(F(X(\underline{\sigma}))+\,\overline{\overline{C}}e^{-\,\overline{\overline{\alpha}}N}\big)}{\sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-N\,f(X(\underline{\rho})-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma}))}},
\end{aligned}\tag{6.2.34}$$

and, in the limit as $N\to\infty$, we get

e6.2.35
$$\begin{aligned}
&\overline{\mu}(F)\le de^{2C''+\overline{C}'+4a\|f\|_\infty}\lim_{N\to\infty}\\
&\frac{\sum_{\underline{\sigma}\in\widehat{\pi}_{N+2a-1}} e^{-\sum_{j=0}^{N+2a-1}A_\Phi(\tau^j\underline{\sigma})}\big(F(X(\underline{\sigma}))+\,\overline{\overline{C}}e^{-\alpha N}\big)}{\sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma})}}.
\end{aligned}\tag{6.2.35}$$

However as a consequence of the finiteness of $\|\Phi\|_1$, for all $p\ge 0$ there exists a constant $B_p<\infty$ such that

e6.2.36
$$B_p^{-1}\le\frac{\sum_{\underline{\sigma}\in\widehat{\pi}_N} e^{-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma})}}{\sum_{\underline{\sigma}\in\widehat{\pi}_{N+p}} e^{-\sum_{j=0}^{N-1}A_\Phi(\tau^j\underline{\sigma})}}\le B_p,\tag{6.2.36}$$

because this ratio can be estimated by the same procedure followed to solve problem [5.3.1] (we leave this to the reader).

Then the limit in (6.2.35) is not larger than $B_{2a}\mu(F)$, for all $F\in C^\alpha(\Omega)$, $F\ge 0$, and this implies existence of $K>0$ such that $\overline{\mu}(F)\le K\mu(F)$, for all $F\in C^\alpha(\Omega)$, $F\ge 0$; and the proof is complete. ■

Bibliographical note to §6.2

This section illustrates some aspects of Sinai's theory, [Si72]; see the bibliographical note to §4.1.

§6.3 Periodic orbits, invariant probability distributions and entropy

An interesting and in a certain sense surprising corollary of the results of the preceding section is the following proposition.

P6.3.1 **(6.3.1) Proposition:** (Topological entropy)
Let (Ω, S) be an Anosov system and let $N_m(S) = \{$number of periodic points of period m contained in $\Omega\}$.
(i) The limit

e6.3.1
$$s_0 = \lim_{m\to\infty} m^{-1} \log N_m(S) \tag{6.3.1}$$

exists and has value $s_0 = s(\mu)$, where μ is a probability distribution which maximizes the entropy function $s(\cdot)$ in $\mathcal{M}_e^t(\Omega, S)$.
(ii) If $\mathcal{Q}$ is a Markovian pavement of Ω with compatibility matrix T and if λ_T is the largest eigenvalue of T (simple by Perron–Frobenius' theorem, cf. *problems [2.3.7], [2.3.10] and [2.3.12]), one has*

e6.3.2
$$s_0 = P(0) = \log \lambda_T > 0. \tag{6.3.2}$$

(iii) The closure of the set of periodic points coincides with Ω.

Proof: Property (iii) is true for the subshift $(\{1,\ldots,q\}_T^{\mathbb{Z}}, \tau)$ because T is transitive (see proposition (4.2.4)) and therefore it holds for (Ω, S) because the symbolic code X is continuous and surjective. A proof not based on symbolic dynamics is perhaps more instructive, see problem [4.2.13].

Property (i) holds because the probability distribution μ that maximizes $s(\mu')$ on $\mathcal{M}_e^t(\Omega)$ is such that $s_0 = s(\mu)$ by (6.2.18), while, by proposition (6.2.1), it is the image of the Gibbs distribution on $\{1,\ldots,q\}_T^{\mathbb{Z}}$ with potential $\Phi = 0$ (cf. proof of lemma (6.2.1), in particular (6.2.7) and (6.2.10)). Then

e6.3.3
$$\begin{aligned} P(0) &= \lim_{N\to\infty} N^{-1} \log \sum_{\underline{\sigma}\in\{1,\ldots,q\}_T^N} 1 = \\ &= \lim_{N\to\infty} N^{-1} \log \sum_{\sigma_1,\ldots,\sigma_N} T_{\sigma_1\sigma_2} T_{\sigma_2\sigma_3} \ldots T_{\sigma_{N-1}\sigma_N} = \\ &= \lim_{N\to\infty} N^{-1} \log \sum_{\sigma,\sigma'} (T^{N-1})_{\sigma\sigma'} = \log \lambda_T, \end{aligned} \tag{6.3.3}$$

by Perron–Frobenius' theorem, since T is mixing. Hence also (ii) follows. ■

The problem of the counting of periodic points can be further refined: for instance the following proposition holds (Manning–Ruelle).

P6.3.2 **(6.3.2) Proposition:** (Zeta function)
Let (Ω, S) be an Anosov system.
(i) The function

e6.3.4
$$\zeta(s,f) = \exp\Big[\sum_{n=1}^{\infty}\Big(\frac{e^{-ns}}{n}\sum_{T_x^n=x} e^{-\sum_{j=0}^{N-1} f(S^j x)}\Big)\Big] \tag{6.3.4}$$

is holomorphic in the complex plane in the variable s for $\mathrm{Re}\, s > p(f)$, if $f \in C^\alpha(\Omega)$, $\alpha > 0$. Such a function can furthermore be extended to a holomorphic function in the half plane Re $s > p(f) - \varepsilon(f)$, with $\varepsilon(f) > 0$ suitable, deprived of the point $s = p(f)$, where a simple pole is present .
(ii) The function $s \to \zeta(s, 0)$ extends to a meromorphic function in the complex plane s.

We shall not enter into the details of the proof of this proposition that requires an accurate analysis of the multiplicity of the code X associated with a Markovian pavement on Ω of the type of those built in Section §4.2, at least for what concerns the periodic points $x \in \Omega \cap \partial\mathcal{Q}$, [Ma71].

Another question, implicitly solved in the previous sections, is the existence of invariant probability distributions which are absolutely continuous with respect to the volume measure on Ω: we shall only formulate the results in the case in which Ω is two–dimensional, in order to make use of the analysis in Section §4.3. It is however true that both the results of Section §4.3 and those that follow can be extended to Anosov systems with dimension larger than 2.

P6.3.3 **(6.3.3) Proposition:** (SRB distribution)
Under the hypotheses of lemma (6.2.1) there exist two probability distributions μ^+, $\mu^- \in \mathcal{M}_e^t(\Omega, \tau)$ such that for every $F \in C^\alpha(\Omega)$, $\alpha > 0$,

e6.3.5
$$\lim_{N\to\infty} N^{-1}\sum_{j=0}^{N-1} F(S^{\pm j}x) = \int \mu^\pm(d\zeta)F(\zeta) \tag{6.3.5}$$

almost everywhere in x with respect to the volume measure μ_0 on Ω.
The dynamical system (Ω, S) admits an ergodic topological probability distribution μ absolutely continuous with respect to the volume measure if and only if $\mu^+ = \mu^-$. In such a case $\mu_0 = \mu^+ = \mu^-$.

Remarks: (1) The above theorem is due to Sinai, Ruelle, and Bowen, [Si68], [Si72], [Si77], [Ru76], [Ru78], [Ru95], [Ru99] and [Bo70]. The difference between this theorem and Birkhoff's theorem, see proposition (2.2.2), is striking. Here the distribution with which the initial data x are sampled *is not invariant*, *i.e.* x is chosen almost everywhere with respect to μ_0 rather than with respect to an invariant probability distribution.
(2) The above proposition says that the statistical properties of motions whose initial data are chosen randomly with a distribution proportional

to the volume measure on phase space *exist and are independent from the choice of the data*. Such a property is often taken for granted in Statistical Mechanics problems and it is formally called the 0-*th law of statistical mechanics* and $\mu^\pm$ are called the *statistics* of the motions generated by the dynamical system; this *attributes a philosophical primacy to the random choices based on distributions with density with respect to the volume of phase space*.
(3) Note that in general the statistics "toward the future" (μ^+) and the statistics "toward the past" (μ^-) are *different*.
(4) A system (Ω, S) is called *reversible* if there exists an isometry $I:\Omega\to\Omega$ such that

e6.3.6
$$IS = S^{-1}I, \qquad I^2 = 1. \tag{6.3.6}$$

This symmetry, which *ought not be confused with invertibility* which is just the existence of S^{-1}, exists in many concrete applications and it often coincides with the velocity reversal of the constituents of a mechanical system. What might be perhaps surprising is that even in systems in which time reversal symmetry holds one has, in general, $\mu^+\ne\mu^-$.

Proof: From proposition (4.3.2) it follows that the probability distribution m_0, image via the code X associated with the Markovian pavement $\mathcal{Q}$, of the volume measure μ_0 on Ω, is a probability distribution with conditional probabilities given by (4.3.11). The restriction m_0^+ of m_0 to the σ-algebra $\mathcal{B}(\mathbb{Z}^+)$ has instead conditional probabilities given by (4.3.12).
Let m_u be the Gibbs distribution on $\{1,\ldots,q\}^{\mathbb{Z}}_T$ with potential Φ^u described, according to the notations of (4.3.13), by $\Phi^u_X = 0$ unless X is a set translated of $\{0,\ldots,2n+1\}$ for some n, and define

e6.3.7
$$\Phi^u_{2n+1}(\sigma_0,\ldots,\sigma_{2n+1}) \equiv \Phi^u_{(0,\ldots,2n+1)}(\sigma_0,\ldots,\sigma_{2n+1}). \tag{6.3.7}$$

Such a probability distribution exists and is unique, ergodic, mixing with exponential rate on the cylinders (because Φ^u verifies (4.3.14) and, therefore, the bound (5.4.19) and its consequence (5.4.22), for the same reason discussed there).
We apply again the methods and ideas employed to study Gibbs distributions in Section §5.2, see in particular the proof of proposition (5.2.1).
The probability distribution m_u^+, restriction of m_u to the σ–algebra $\mathcal{B}(\mathbb{Z}^+)$, and the probability distribution m_0^+, restriction to $\mathcal{B}(\mathbb{Z}^+)$ of the probability distribution m_0 image via X of the volume measure μ_0, are absolutely equivalent. This can be verified by using (4.3.12), (4.3.13), (4.3.14) and (5.2.5) (which is, with the obvious modifications, applicable also to the conditional probability, (4.3.11), of m_0): in fact one has

e6.3.8
$$\begin{aligned}\frac{m_0^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})}{m_u^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})} &= \frac{\int m_0(\sigma_0\ldots\sigma_N|\widetilde\sigma_{N+1}\ldots,\widetilde\sigma_{-1}\ldots)m_0(d\underline{\widetilde\sigma})}{\int m_u(\sigma_0\ldots\sigma_N|\widehat\sigma_{N+1}\ldots,\widetilde\sigma_{-1}\ldots)m_u(d\underline{\widetilde\sigma})} \le \qquad (6.3.8)\\ &\le (q^a\exp(8\|\Phi^u\|_1))^2\Delta\frac{\int m_0(\sigma_0\ldots\sigma_N|\widehat\sigma_{N+1}\ldots,\widetilde\sigma_{-1}\ldots)m_0(d\underline{\widetilde\sigma})}{\int m_u(\sigma_0\ldots\sigma_N|\widehat\sigma_{N+1}\ldots,\widetilde\sigma_{-1}\ldots)m_u(d\underline{\widetilde\sigma})},\end{aligned}$$

N6.3.1 where a is the mixing time of T and[1] $\Delta = \max_{\underline{\sigma},\underline{\sigma}'} \left|\frac{\sin\varphi(X(\underline{\sigma}))}{\sin\varphi(X(\underline{\sigma}'))}\right| \le \frac{1}{\min\limits_{x\in\Omega}|\sin\varphi(x)|}$ and $\widehat{\underline{\sigma}}$ is a reference configuration arbitrarily chosen among those that continue $\ldots\widetilde{\underline{\sigma}}_{-1}\,\sigma_0\ldots\sigma_N$ into a configuration $\ldots\widetilde{\underline{\sigma}}_{-1}\,\sigma_0\ldots\sigma_N\,\widehat{\sigma}_{N+1}\ldots$: the $\widehat{\underline{\sigma}}$ is chosen so that it depends exclusively on σ_N and not, in particular, on $\widetilde{\underline{\sigma}}$, see the beginning of the proof of lemma (6.2.1). In (6.3.8) all integrals must be understood to be extended to regions such that the configurations that appear in the conditional probabilities are in $\{1,\ldots,q\}^{\mathbb{Z}}_T$.

Since the sequence $\widehat{\underline{\sigma}}$ is fixed the integrals in (6.3.8) can be immediately performed and (6.3.8) becomes

e6.3.9
$$\frac{m_0^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})}{m_u^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})} \le K_1 \frac{m_0^+(\sigma_0\ldots\sigma_N|\widehat{\sigma}_{N+1}\ldots)}{m_u^+(\sigma_0\ldots\sigma_N|\widehat{\sigma}_{N+1}\ldots)}, \tag{6.3.9}$$

for a suitable constant K_1. The probability distribution m_0 can be treated as the probability distribution m of Section §5.2 and we can prove for it propositions analogous to propositions (5.3.1), (5.3.2) and corollary (5.3.1) for which the probability distribution m_0^+ turns out to be equivalent (with a bound for the Radon–Nykodim derivative similar to (5.3.10)) to the probability distribution $\widetilde{m}_u^+$, a Gibbs distribution on $\mathbb{Z}^+$ with potential Φ^u. Since also m_u^+ is, by corollary (5.3.1), absolutely continuous with respect to $\widetilde{m}_u^+$ and (5.3.10) holds, the r.h.s. of (6.3.9) can be bounded by a constant K_2. By what said above one deduces that we can choose

e6.3.10
$$K_2 = e^{8\|\Phi^u\|_1 + 4a\|\Phi^u\|}\Delta\, q^{2a} K_1, \tag{6.3.10}$$

and, in conclusion,

e6.3.11
$$\frac{m_0^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})}{m_u^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})} \le K_2 \text{ and, likewise, } \ge K_2^{-1}. \tag{6.3.11}$$

Hence m_0^+ is absolutely continuous and equivalent to m_u^+ and $\pi_u \stackrel{def}{=} \frac{dm_0^+}{dm_u^+}$ is a $\mathcal{B}(\mathbb{Z}^+)$–measurable function such that $K_2^{-1} \le \pi_u(\underline{\sigma}) \le K_2$, m_0^+–almost everywhere. In fact with a little extra effort and by using the ideas of Section §5.2 and of the proof of proposition (5.3.2) in particular, it is possible to show that π_u can be chosen to be Hölder continuous on $\{1,\ldots,q\}^{\mathbb{Z}^+}_T$ (*i.e.* $|\pi_u(\underline{\sigma}) - \pi_u(\underline{\sigma}')| \le C\, d(\underline{\sigma},\underline{\sigma}')^\alpha$ for all $\underline{\sigma},\underline{\sigma}' \in \{1,\ldots,q\}^{\mathbb{Z}^+}_T$, , with $C, \alpha > 0$): we leave this derivation to the reader.

If $\underline{\sigma} \to \widehat{F}(\underline{\sigma})$ is a $\mathcal{B}(\mathbb{Z}^+)$–measurable function on $\{1,\ldots,q\}^{\mathbb{Z}}_T$ we shall have m_u^+–almost everywhere and, therefore, m_0^+–almost everywhere and

[1] Note that in the course of the proof of proposition (4.3.2) we deduced an explicit expression for the function f that appears in (4.3.12), *i.e.* $f(\underline{\sigma}) = \sin\varphi(X(\underline{\sigma}))$, where $\varphi(X(\underline{\sigma}))$ denotes the angle between the direction of the manifold stable and that of the unstable manifold at the point $x = X(\underline{\sigma})$. Such an expression is here used to get the bound (6.3.8).

m_0–almost everywhere

e6.3.12
$$N^{-1}\sum_{j=0}^{N-1}\widehat F(\tau^j\underline\sigma)\xrightarrow[N\to\infty]{}\int\widehat F(\underline\sigma')m_u^+(d\underline\sigma')=\int\widehat F(\underline\sigma')m_u(d\underline\sigma'),\qquad(6.3.12)$$

by Birkhoff theorem and by the ergodicity of m_u.

Hence, by a density argument, the limit (6.3.12) is valid for all $F\in C(\{1,\ldots,q\}_T^{\mathbb{Z}})$.[2]

N6.3.2 If $F\in C^\alpha(\Omega)$, $\alpha>0$, and μ^+ is the image (isomorphic mod 0 to m_u) via the code X, we shall have that the function $\widehat F(\underline\sigma)=F(X(\underline\sigma))$ is in $C(\{1,\ldots,q\}_T^{\mathbb{Z}})$ and, by (6.3.12), if $x=X(\underline\sigma)$

e6.3.13
$$\begin{aligned}&N^{-1}\sum_{j=0}^{N-1}F(S^jx)=N^{-1}\sum_{j=0}^{N-1}\widehat F(\tau^j\underline\sigma)\xrightarrow[N\to\infty]{}\\&\xrightarrow[N\to\infty]{}\int\widehat F(\underline\sigma')m_u(d\underline\sigma')=\int F(x')\mu^+(dx')\end{aligned}\qquad(6.3.13)$$

μ_0–almost everywhere.

Similar arguments can be given for the averages in the past and lead to the probability distribution μ^-.

The second part of the proposition is a consequence of the first part. ∎

Note that in the preceding proof we obtained something more.

Let indeed $\widehat F$, $\widehat G$ be a pair of bounded and $\mathcal{B}(\mathbb{Z}^+)$–measurable functions; then, for all $j\ge0$, obviously

e6.3.14
$$\begin{aligned}&\int\widehat F(\tau^j\underline\sigma)\widehat G(\underline\sigma)m_0(d\underline\sigma)=\int\widehat F(\tau^j\underline\sigma)\widehat G(\underline\sigma)m_0^+(d\underline\sigma)=\\&=\int\widehat F(\tau^j\underline\sigma)\widehat G(\underline\sigma)\pi_u(\underline\sigma)m_u^+(d\underline\sigma)=\int\widehat F(\tau^j\underline\sigma)\widehat G(\underline\sigma)\pi_u(\underline\sigma)m_u(d\underline\sigma),\end{aligned}\qquad(6.3.14)$$

where π_u is the function, Hölder continuous on $\{1,\ldots,q\}_T^{\mathbb{Z}^+}$, defined after the (6.3.11). Since m_u is mixing we obtain

e6.3.15
$$\lim_{j\to\infty}m_0(\widehat G\tau^j\widehat F)=m_u(\pi_u\widehat G)m_u(\widehat F)=m_0(\widehat G)m_u(\widehat F),\qquad(6.3.15)$$

and such a limit is reached at exponential rate if $\widehat F$ and $\widehat G$ are also Hölder continuous. All this follows easily from (5.4.22) which holds for m_u as noted after (6.3.7). By density this remains true for every $\widehat F$, $\widehat G$ which are Hölder continuous and $\mathcal{B}(\mathbb{Z})$–measurable.

---

[2] Note that the assumption for $\widehat F$ to be $\mathcal{B}(\mathbb{Z}^+)$-measurable can be obviously weakened into $\mathcal{B}([k,+\infty))$-measurable, for some $k\in\mathbb{Z}$, and such continuous functions are dense in $C(\{1,\ldots,q\}_T^{\mathbb{Z}})$.

Therefore if $F,G\in C^\alpha(\Omega)$ and if we set $\widehat F(\underline\sigma)=F(X(\underline\sigma))$, $\widehat G(\underline\sigma)=G(X(\underline\sigma))$ we see that this gives the following.

C6.3.1 **(6.3.1) Corollary:** (Mixing of the non invariant volume distributions)
If $F,\,G\in C^\alpha(\Omega)$, $\alpha>0$, one has

e6.3.16
$$\lim_{j\to\pm\infty}\mu_0(GS^jF)=\mu_0(G)\mu_\pm(F) \tag{6.3.16}$$

and the limits are attained with exponential rate.

Finally the following proposition is remarkable.

P6.3.4 **(6.3.4) Proposition:** (Pesin formula)
Let (Ω,S) be an Anosov system. With the usual notations of definition (4.3.1) we have

e6.3.17
$$s(\mu^+)-\mu^+(\log\lambda_u)=s(\mu^-)+\mu^-(\log\lambda_s)=0 \tag{6.3.17}$$

Furthermore μ^+ and μ^- maximize in $\mathcal M^t_e(\Omega,S)$, respectively, the functionals $s(\mu)-\mu(\log\lambda_u)$ and $s(\mu)+\mu(\log\lambda_s)$.

Proof (hint): The second statement follows immediately because μ^+ and μ^- are X–images of Gibbs distributions with potentials Φ^u and Φ^s defined via (4.3.5), as shown by proposition (4.3.2) (cf. (4.3.15)) and by the proof of proposition (6.3.3): this statement is then reduced to the variational principle for Gibbs distributions, discussed in the previous sections.

To check (6.3.17) note that, by (6.3.11), if $P^{0\ldots N}_{\sigma_0\ldots\sigma_N}=\bigcap_{j=0}^N S^{-j}Q_{\sigma_j}$,

e6.3.18
$$K_2^{-1}\le\frac{m_u^+(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})}{\mu_0(P^{0\ldots N}_{\sigma_0\ldots\sigma_N})}\le K_2, \tag{6.3.18}$$

and, therefore, to compute the value of $s(m_u)=s(\mu^+)$ we can make use of Shannon-McMillan's theorem and, by (6.3.18), we deduce that

e6.3.19
$$s(\mu^+)=s(m_u)=-\lim_{N\to\infty}N^{-1}\log\mu_0(C^{0\ldots N}_{\sigma_0\ldots\sigma_N}), \tag{6.3.19}$$

where the convergence of $f_N(\underline\sigma)=-N^{-1}\log\mu_0(C^{0\ldots N}_{\sigma_0\ldots\sigma_N})$ to $s(\mu^+)$ takes place in $L_1(m_u)$; see lemma (3.2.1).

Let $\underline\sigma\in\{1,\ldots,q\}^{\mathbb Z}_T$ and $x=X(\underline\sigma)$. It is possible to see, following the methods used to obtain the estimates of Section §4.3 that there exists K' such that

e6.3.20
$$(K')^{-1}\le\frac{\mu_0(P^{0\ldots N}_{\sigma_0\ldots\sigma_N})}{\prod_{j=0}^{N-1}\lambda_u^{-1}(S^jX(\underline\sigma))}\le K' \tag{6.3.20}$$

(this is first heuristically shown by interpreting it geometrically with the help of a drawing like Fig. (4.3.1) and Fig. (4.3.4)).

Equation (6.3.20) implies immediately, by Shannon–McMillan's and Birkhoff's theorems, by proposition (6.3.3) and by considering the logarithm divided by N of both sides of (6.3.20), that the entropy $s(\mu^+) = s(m_u)$ is

e6.3.21
$$s(\mu^+) = s(m_u) = \lim_{N\to\infty} -N^{-1} \log \mu_0(P^{0\ldots N}_{\sigma_0\ldots\sigma_N}) = \qquad (6.3.21)$$
$$= \lim_{N\to\infty} -N^{-1} \sum_{j=0}^{N-1} \log \lambda_u^{-1}(S^j x) = \int \mu^+(dx) \log \lambda_u^{-1}(x),$$

so that (6.3.17) follows. ■

We have therefore proved also the following corollary.

C6.3.2 **(6.3.2) Corollary:** *Let (Ω, S) be an Anosov system. Suppose the the map S preserves the volume measure μ_0 on Ω then*

e6.3.22
$$s(\mu_0) = -\int \mu_0(dx) \log \lambda_u(x), \qquad (6.3.22)$$

i.e. "the entropy of an absolutely continuous invariant distribution is the average value of the logarithm of the contraction coefficient".

Remarks: (1) Note the relation between this theorem and the theorem of Kouchnirenko of Section §3.4.

(2) Hence the dynamical system on $\mathbb{T}^2$ defined by the map $S\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix} = \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}\varphi_1\\ \varphi_2\end{pmatrix}$ mod 2π is such that the entropy of the Lebesgue measure μ_0 is

e6.3.23
$$s = \log(3+\sqrt{5})/2 = \log\lambda. \qquad (6.3.23)$$

(3) Furthermore μ_0 is the probability distribution that maximizes the functional $s(\mu') - \mu'(\log\lambda)$ in the space of the ergodic probability distributions on $\mathbb{T}^2$. This means that it makes maximal the entropy, since λ is constant.
(4) The number of the S–periodic points is (cf. proposition (6.3.1))

e6.3.24
$$N_m(s) = e^{s\,m+o(m)} = \Big(\frac{3+\sqrt{5}}{2}\Big)^{m+o(m)}. \qquad (6.3.24)$$

(5) This means that every Markovian pavement of $(\mathbb{T}^2, S)$ has a compatibility matrix T with the largest eigenvalue equal to $\frac{3+\sqrt{5}}{2} > 2$. See also the problems below.
(6) Hence a Markovian pavement for the system $(\mathbb{T}^2, S)$ consisting of only two S–rectangles cannot exist.

A representation of the SRB distribution similar to the one for the volume measure discussed in corollary (4.3.1) follows from the argument leading to (4.3.23) and from the analysis of Sections 6.2, 6.3 and the following proposition will be useful.

C6.3.3 **(6.3.3) Corollary:** (Fubini's theorem for the SRB distribution)
*Let (M,S) be a two-dimensional Anosov map. Given $x_0 \in M$ consider the rectangle $R = W^s_\delta(x_0) \times W^u_\delta(x_0)$ where δ is small enough so that the point $[x,y]$ is uniquely defined (*cf. *Fig.(4.2.1)). Then the SRB probability $\mu(E)$ of a subset $E \subset R$ is given by*

e6.3.25
$$\begin{aligned}\mu(E) = &\int_{W^u_\delta(x_0)} d\sigma_x \int_{W^s_\delta(x_0)} \nu(d\sigma_y)\cdot\\ &\cdot \prod_{i=1}^{\infty} \frac{\lambda_u^{-1}(S^{-i}[y,x])}{\lambda_u^{-1}(S^{-i}x)} \chi_E([y,x])\end{aligned} \tag{6.3.25}$$

where $d\sigma_x$ is the area measure on $W^u_\delta(x_0)$ and $\nu(d\sigma_y)$ is a measure on $W^s_\delta(x_0)$.

The measure ν will in general *not* be absolutely continuous with respect to the arc length $d\sigma_y$. This result in fact is general and it holds in any dimension.

Problems for §6.3

Q6.3.1 **[6.3.1]:** Let (Ω, S) be an Anosov system and let $\mathcal{Q}$ be a pavement of Ω with S–rectangles (cf. definition (4.2.2)) which have the Markov property (4.2.12). If $\mathcal{Q}$ is generating then we could conclude that the compatibility matrix, see Section §4.1, defined by $T_{\sigma\sigma'} = 1$ if $\text{int}(SQ_\sigma) \cap \text{int}(Q_{\sigma'}) \neq \emptyset$ and $T_{\sigma\sigma'} = 0$ otherwise, has the logarithm of the largest eigenvalue equal to the topological entropy by proposition (6.3.1). Suppose that $\mathcal{Q}$ is non-generating: show that the logarithm of the largest eigenvalue is less than (or equal to) the topological entropy. However it cannot differ from the topological entropy by more than the logarithm of the maximum number of connected components of the sets $\text{int}(Q_\sigma) \cap \text{int}(Q_{\sigma'})$. (*Hint:* If $\mathcal{Q}$ is non-generating but it is not trivial then the correspondence between compatible symbolic sequences and points whose trajectories never fall on the boundaries will not be one-to-one but it will have multiplicity.)

Q6.3.2 **[6.3.2]:** Show that corollary (6.3.2) implies that the matrix $S_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ and the matrix of problem [4.3.7] have the same largest eigenvalue, *i.e.* $(1+\sqrt{5})/2$, cf. Fig.(4.3.4).

Q6.3.3 **[6.3.3]:** Check that the matrix S_0 and the matrix of problem [4.3.7] have the same largest eigenvalue by directly computing it.

Bibliographical note to §6.3

This section illustrates some aspects of Sinai's theory, [Si72], [Si77]; see the bibliographical notes to §4.1.

§6.4 Equivalent potentials. Gibbs distributions with transitive vacuum

Let B be the space of potentials for $\{0,\ldots,n\}^{\mathbb{Z}}_T$, where T is a mixing compatibility matrix (cf. definitions (4.1.1)) and (5.1.1).

A natural question is whether it is possible that two potentials Φ and Ψ generate the same Gibbs states and, in the affirmative case, which is the relation between Φ and Ψ.

A fully satisfactory answer does not seem to exist, partly because the space B is not a "natural" space of potentials. It is just a family of conveniently restricted potentials large enough to give rise to a theory endowed of a remarkable variety of phenomena and singularities.

Nevertheless the following propositions will give an idea of what it means that Φ and Ψ produce the same Gibbs states: if this happens we say that the potential Φ and Ψ are *equivalent potentials.*

P6.4.1 **(6.4.1) Proposition:** (Equivalent potentials)
Let $B_s^h = \cup_{\kappa>s} B^{(\kappa)}$, where $s \geq 0$, and $B^{(\kappa)} \subset B$ is the space of the potentials $\Phi \in B$ such that

e6.4.1
$$\|\Phi\|^{(\kappa)} = \sum_{X \ni 0} e^{\kappa \,\mathrm{diam}(X)} \|\Phi_X\| < +\infty. \tag{6.4.1}$$

The B_s^h is called the space of the potentials that "decay exponentially" on a
N6.4.1 *scale shorter than s^{-1}.*[1] *Then*
(i) If, given Φ, $\Psi \in B_0^h$, there exists a constant C and a Hölder continuous function $F \in C(\{0,\ldots,n\}_T^{\mathbb{Z}})$ such that

e6.4.2
$$A_\Phi(\underline{\sigma}) = A_\Psi(\underline{\sigma}) + F(\tau\underline{\sigma}) - F(\underline{\sigma}) + C \tag{6.4.2}$$

holds then Φ and Ψ satisfy $G(\Phi) = G(\Psi)$.
(ii) Viceversa if Φ, $\Psi \in B_0^h$ and $G(\Phi) = G(\Psi)$ then there exists a constant C and a Hölder continuous function $F \in C(\{0,\ldots,n\}_T^{\mathbb{Z}})$ such that (6.4.2) holds.

Proof: Since (6.4.2) implies that for all invariant probability distributions
N6.4.2 $\mu \in M(\{0,\ldots,n\}_T^{\mathbb{Z}})$ [2] one has $s(\mu) - \mu(A_\Phi) = s(\mu) - \mu(A_\Psi) + C$, item (i) follows from the variational principle (corollary (6.1.1)).

To prove (ii) suppose that Φ, $\Psi \in B_0^h$ and that there exists $C \in \mathbb{R}$ for which

e6.4.3
$$\mu(A_\Phi) = \mu(A_\Psi) + C \qquad \text{for all}\, \mu \in M(\{0,\ldots,n\}_T^{\mathbb{Z}}), \tag{6.4.3}$$

We first prove that, under the assumption (6.4.3) there exists $F \in C(\{0,\ldots,n\}_T^{\mathbb{Z}})$, Hölder continuous, for which (6.4.2) holds.

In fact note that the function $g = A_{\Phi-\Psi} - C$ has zero integral with respect to an arbitrary $\mu \in M(\{0,\ldots,n\}_T^{\mathbb{Z}})$ and hence, in particular,

e6.4.4
$$\sum_{j=0}^{N-1} g(\tau^j \underline{\sigma}) = 0 \tag{6.4.4}$$

[1] Note that one has $B_0^h \subset B$; the superscript h refers to the fact that the function $A_\Phi(\underline{\sigma})$ corresponding to any $\Phi \in B_s^h$, with $s \geq 0$, is Hölder-continuous.

[2] See remark to proposition (2.3.1) for notations.

for every $\underline{\sigma}$ that belongs to the set of the periodic points with period N in
N6.4.3 $\{0,\ldots,n\}_T^{\mathbb{Z}}$.[3]

Furthermore the hypothesis of mixing on T guarantees the existence of $\underline{\sigma}_0$ such that $\cup_{j=0}^{\infty}\{\tau^j\underline{\sigma}_0\}$ is dense in $\{0,\ldots,n\}_T^{\mathbb{Z}}$ (cf. problem [4.1.19]).

We can then set

e6.4.5
$$\begin{aligned}&F(\underline{\sigma}_0)=0,\ F(\tau\underline{\sigma}_0)=g(\underline{\sigma}_0),\ F(\tau^2\underline{\sigma}_0)=g(\underline{\sigma}_0)+g(\tau\underline{\sigma}_0),\\ &\ldots,F(\tau^h\underline{\sigma}_0)=\sum_{j=0}^{h-1}g(\tau^j\underline{\sigma}_0),\ldots\end{aligned}\tag{6.4.5}$$

We now show that the function F thus defined on $\cup_{j=0}^{\infty}\{\tau^j\underline{\sigma}_0\}$ is in fact uniformly Hölder continuous on this set (*i.e.* the Hölder constant and the exponent can be chosen independently the point in $\cup_{j=0}^{\infty}\{\tau^j\underline{\sigma}_0\}$) and *therefore it extends to a continuous function* on all $\{0,\ldots,n\}_T^{\mathbb{Z}}$ that, by construction, verifies $g(\underline{\sigma})=F(\tau\underline{\sigma})-F(\underline{\sigma})$, and this implies (6.4.2).

To see the uniform Hölder continuity we remark that one has $A_{\Phi-\Psi}(\underline{\sigma})=\sum_{X\ni 0}\frac{\Phi_X(\underline{\sigma}_X)-\Psi_X(\underline{\sigma}_X)}{|X|}$ and recall that the definition of distance $d(\underline{\sigma},\underline{\sigma}')$ between two sequences $\underline{\sigma},\underline{\sigma}'$ is $d(\underline{\sigma},\underline{\sigma}')=e^{-\nu}$ if $\sigma_j=\sigma'_j$, for all $|j|\le\nu$ and $\sigma_{\pm(\nu+1)}\ne\sigma'_{\pm(\nu+1)}$ (see definition (4.1.3)). If $\underline{\sigma}$ and $\underline{\sigma}'$ are close then (6.4.1) implies existence of two constants $\kappa'>0$ and $C>0$ such that

e6.4.6
$$|g(\underline{\sigma})-g(\underline{\sigma}')|\le C\ d(\underline{\sigma},\underline{\sigma}')^{\kappa'}.\tag{6.4.6}$$

Then if $\tau^k\underline{\sigma}_0$ and $\tau^h\underline{\sigma}_0$ happen to be very close, *i.e.* if $(\tau^k\underline{\sigma}_0)_i=(\tau^h\underline{\sigma}_0)_i$ for all $|i|\le\nu$, and if $p=k-h\ge 0$ we set $\underline{\sigma}_1=\tau^h\underline{\sigma}_0$ and note that

e6.4.7
$$|F(\tau^k\underline{\sigma}_0)-F(\tau^h\underline{\sigma}_0)|=\Big|\sum_{j=0}^{p-1}g(\tau^j\underline{\sigma}_1)\Big|.\tag{6.4.7}$$

The definitions imply that $(\underline{\sigma}_1)_i=(\underline{\sigma}_1)_{i+p}$, for all $|i|\le\nu$, and hence there exists a configuration $\widehat{\underline{\sigma}}_1\in\{0,\ldots,n\}_T^{\mathbb{Z}}$ which is periodic with period p and such that

e6.4.8
$$(\widehat{\underline{\sigma}}_1)_j=(\underline{\sigma}_1)_j\qquad -\nu\le j\le\nu,\tag{6.4.8}$$

(this is $\widehat{\sigma}_{1j}=\sigma_{1j}$ for $-\nu\le j<-\nu+p$ and then $\widehat{\underline{\sigma}}$ is continued periodically). So that, by (6.4.4) and (6.4.7),

e6.4.9
$$|F(\tau^k\underline{\sigma}_0)-F(\tau^h\underline{\sigma}_0)|\le\sum_{j=0}^{p-1}|g(\tau^j\underline{\sigma}_1)-g(\tau^j\widehat{\underline{\sigma}}_1)|\le\frac{2C_{\kappa'}e^{-\kappa'\nu}}{1-e^{-\kappa'}}.\tag{6.4.9}$$

Therefore F is uniformly Hölder continuous on $\cup_{j=0}^{\infty}\{\tau^j\underline{\sigma}_0\}$ and (6.4.2) is proved under the assumption in (6.4.3).

---

[3] The measure $\mu=N^{-1}\sum_{j=0}^{N-1}\delta_{\tau^j\underline{\sigma}}$ is in $M(\{0,\ldots,n\}_T^{\mathbb{Z}})$!

To prove the statement in item (ii) we must show that (6.4.3) follows from the assumptions that $\Phi, \Psi \in B_0^h$ and that $G(\Phi) = G(\Psi)$. This means that the tangent planes to the graph of the pressure $P(\Phi)$ in Φ and in Ψ (which are unique, cf. propositions (5.2.1), (6.1.2)) coincide. Since $P(\Phi)$ is convex this means that they are also tangent to P at the points $P(\Phi+t(\Psi-\Phi))$ for $t \in [0,1]$: hence *for all t real*, because the DLR relations are real analytic in the potentials. Let $C = \mu(\Psi - \Phi)$: the variational principle (cf. corollary (6.1.1)) then tells us that

e6.4.10
$$\begin{aligned} &\max_{\nu\in M(\{0,\ldots,n\}^{\mathbb{Z}}_T)} \big(s(\nu) - \nu(A_{\Phi+t\,(\Psi-\Phi)})\big) = \\ &= s(\mu) - \mu(A_{\Phi+t\,(\Psi-\Phi)}) = P(\Phi) - Ct \end{aligned} \tag{6.4.10}$$

Noting that $s(\nu) - \nu(A_{\Phi+t\,(\Psi-\Phi)}) \equiv s(\nu) - \nu(A_\Phi) - t\nu(A_{\Psi-\Phi}) \le P(\Phi) - Ct$ for all t real and for all $\nu \in M(\{0,\ldots,n\}^{\mathbb{Z}}_T)$ can only be if $\nu(A_{\Psi-\Phi}) \equiv C$ we get $A_\Phi - A_\Psi = C + G$ with $\nu(G) \equiv 0$ *for all* $\nu \in M(\{0,\ldots,n\}^{\mathbb{Z}}_T)$. Hence we see that (6.4.3) holds. ■

Remark: We can define on B_0^h an equivalence relation setting $\Phi \approx \Psi$ if there exist $F \in C(\{0,\ldots,n\}^{\mathbb{Z}}_T)$ and a constant $C \in \mathbb{R}$ for which (6.4.2) holds. A norm on the equivalence classes defined by this relation is given by

e6.4.11
$$\|\Phi\| = \inf_{C,F} \sup_{\underline{\sigma}} |A_\Phi(\underline{\sigma}) + C + F(\tau\underline{\sigma}) - F(\underline{\sigma})|, \tag{6.4.11}$$

It would be natural to call "space of the potentials" the space obtained by taking the closure B' of the classes of B_0^h with respect to the norm (6.4.11).

But this new space B' of (equivalence classes of) potentials is difficult to study and it is not clear not only whether $\Phi \ne \Psi$ implies $G(\Phi) \ne G(\Psi)$ for Φ, $\Psi \in B'$, but even whether we could extend the notion of Gibbs distribution to the elements Φ of the completion of the classes of B_0^h.

Nevertheless the idea of choosing one representative for every class of equivalent potentials can, in several important cases, be realized. For example we consider the following case.

C6.4.1 **(6.4.1) Corollary:** (Potentials in systems with transitive vacuum) *If $T_{0\sigma} = T_{\sigma 0} = 1$ for all $\sigma = 0,\ldots,n$, ("transitivity of T on the vacuum") then every $\Psi \in B^h_{\log 2}$ is equivalent to a potential $\Phi \in B_0^h$ such that $\Phi_X(\underline{\sigma}_X) = 0$ if $\sigma_j = 0$ for some $j \in X$; such a potential is simply defined by*

e6.4.12
$$\begin{aligned} &\Phi_X(\underline{\sigma}_X) = 0 \qquad \text{if } \sigma_j = 0 \text{ for some } j \in X, \\ &\Phi_X(\underline{\sigma}_X) = \sum_{Y\supset X} \Psi_Y(\underline{\sigma}_X\underline{0}), \end{aligned} \tag{6.4.12}$$

N6.4.4 *where $(\underline{\sigma}_X\underline{0})_j = 0$ if $j \in Y/X$ and $(\underline{\sigma}_X\underline{0})_j = \sigma_j$ $(\ne 0)$ if $j \in X$.*[4]

[4] The notation we are using here is the same used in Section §5.1.

Remark: In this context the symbol 0 is called here "vacuum": however its only property is that transitions from $\sigma = 0$ to any σ' and vice versa are possible. Therefore if another symbol enjoys this property we could equally well take it as a reference symbol, call it "vacuum", and draw analogous conclusions.

Proof: It is immediate to check that if $\Psi \in B^h_{\log 2}$ then $\Phi \in B^h_0$, by using the definition in (6.4.12).

Taking into account proposition (5.1.6), (5.1.11) and (5.1.16) one finds for all $\mu \in M(\{0,\ldots,n\}^{\mathbb{Z}}_T)$

e6.4.13

$$\begin{aligned}
\mu(A_\Psi) &= \lim_{N\to\infty} N^{-1}\mu\Big(\sum_{j=0}^{N-1} A_\Psi(\tau^j\underline{\sigma})\Big) = \lim_{N\to\infty} N^{-1}\mu\Big(\sum_{X\subset[0,\ldots,N]} \Psi_X(\underline{\sigma}_X)\Big) = \\
&= \lim_{N\to\infty} N^{-1}\mu\Big(\sum_{X\subset[0,\ldots,N]} \Phi_X(\underline{\sigma}_X) + N\sum_{X\ni 0}\frac{\Psi_X(\underline{0})}{|X|}\Big) = \\
&= \lim_{N\to\infty} N^{-1}\mu\Big(\sum_{j=0}^{N-1} A_\Phi(\tau^j\underline{\sigma}) + N\sum_{X\ni 0}\frac{\Psi_X(\underline{0})}{|X|}\Big) = \qquad (6.4.13)\\
&= \mu(A_\Phi) + \sum_{X\ni 0}\frac{\Psi_X(\underline{0})}{|X|},
\end{aligned}$$

hence we are in the situation of item (i) of proposition (6.4.1). ■

It is then convenient to set the following definition.

D6.4.1 **(6.4.1) Definition:** (Gibbs states with transitive vacuum, particle potentials)

Let T be a mixing compatibility matrix with labels in $\{0,\ldots,n\}$ which is transitive on the vacuum, i.e. such that $T_{0\sigma} = T_{\sigma 0} = 1$ for all $\sigma = 0,\ldots,n$. Denote by $\widetilde{B} \subset B$ the space of the potentials for $\{0,\ldots,n\}^{\mathbb{Z}}_T$ such that

e6.4.14

$$\Phi_X(\underline{\sigma}_X) = 0 \qquad \textit{if} \quad \sigma_j = 0 \textit{ for some } j \in X. \qquad (6.4.14)$$

We shall say that $\{0,\ldots,n\}^{\mathbb{Z}}_T$ is the space of the configurations of n species of particles, denoted by the label $\sigma = 1,\ldots,n$, on $\mathbb{Z}$ subjected to a hard core *condition described by T (i.e. we interpret every 0 of the matrix T as a condition of incompatibility between two particles, that forbids the possibility of configurations in which they appear as nearest neighbours).*

The potentials $\Phi \in \widetilde{B}$ will be said particle potentials for n species of particles and a vacuum *and the Gibbs states on $\{0,\ldots,n\}^{\mathbb{Z}}_T$ associated with potentials $\Phi \in \widetilde{B}$ will be said* Gibbs states for n species of particles and a vacuum.

Remarks: The wording in the preceding definition is useful because it allows us to give a suggestive physical interpretation to the Gibbs states on $\{0,\ldots,n\}^{\mathbb{Z}}_T$ with T transitive on the vacuum. The configuration $\underline{\sigma} = \underline{0}$, with $\sigma_i \equiv 0$, is called, for obvious reasons, the *vacuum*.

The following proposition holds (Griffiths–Ruelle), [GR71].

P6.4.2 **(6.4.2) Proposition:** (Different particle potentials generate different Gibbs states)
If $\{0,\ldots,n\}_T^{\mathbb{Z}}$ is defined by a mixing matrix T which is transitive on the vacuum and if $\widetilde{B}$ is the set of the particles potentials *on $\{0,\ldots,n\}_T^{\mathbb{Z}}$, then Φ, $\Psi\in\widetilde{B}$ and $\Phi\neq\Psi$ imply $G(\Phi)\cap G(\Psi)=\emptyset$.*

Remark: If we consider the graph of $\Phi\to P(\Phi)$ for $\Phi\in\widetilde{B}$ we see that the above proposition, combined with the results of propositions (6.1.1) and (6.1.2), means that the graph of $P(\Phi)$ on $\widetilde{B}$ is *strictly convex*, *i.e.* it is convex and in every point it has a different tangent plane (one checks that the propositions (6.1.1) and (6.1.2) and the remark following proposition (6.1.2) remain valid if, under the present hypotheses on T, we replace B with $\widetilde{B}\subset B$).

Proof: Let $\Lambda_0\subset\mathbb{Z}$ be a fixed finite region and let $\Lambda_N=[-N,N]\supset\Lambda_0$.
The DLR equations, see (5.1.9), imply that if $\mu\in G(\Phi)$ one has

e6.4.15
$$\mu(C^{\Lambda_N}_{\underline{\sigma}_{\Lambda_0}\underline{0}})=\int\frac{e^{-\sum_{X\subset\Lambda_0}\Phi_X(\underline{\sigma}_X)-W(\underline{\sigma}_{\Lambda_0}\underline{0}|\underline{\sigma}_{\Lambda_N^c})}}{Z_{\Lambda_N}(\Phi;\underline{\sigma}_{\Lambda_N^c})}\,\mu(d\underline{\sigma}_{\Lambda_N^c}),\tag{6.4.15}$$

where $\mu(d\underline{\sigma}_{\Lambda_N^c})$ denotes the restriction to $\mathcal{B}(\Lambda_N^c)$ of the measure μ, the configuration $(\underline{\sigma}_{\Lambda_0}\underline{0})\in\{0,\ldots,n\}_T^{\Lambda_N}$ is defined to be $\sigma'_j=0$ if $j\notin\Lambda_0$ and $\sigma'_j=\sigma_j$ if $j\in\Lambda_0$,

e6.4.16
$$W(\underline{\sigma}'_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c})=\sum_{\substack{R\cap\Lambda_N\neq\emptyset\\ R\cap\Lambda_N^c\neq\emptyset}}\Phi_R(\underline{\sigma}'_R),\tag{6.4.16}$$

with $\underline{\sigma}'=((\underline{\sigma}_{\Lambda_0}\underline{0})\underline{\sigma}_{\Lambda_N^c})$ and

e6.4.17
$$Z_{\Lambda_N}(\Phi;\underline{\sigma}_{\Lambda_N^c})=\sum_{\underline{\sigma}_{\Lambda_N}}e^{-\sum\limits_{X\subset\Lambda_N}\Phi_X(\underline{\sigma}_X)-W(\underline{\sigma}_{\Lambda_N}|\underline{\sigma}_{\Lambda_N^c})},\tag{6.4.17}$$

according to the notations of Section §6.1.
The condition $\|\Phi\|<+\infty$ shows that if Λ_0 is fixed

e6.4.18
$$\lim_{N\to\infty}W(\underline{\sigma}_{\Lambda_0}\underline{0}|\underline{\sigma}_{\Lambda_N^c})=0,\tag{6.4.18}$$

uniformly in $\underline{\sigma}_{\Lambda_0}$, $\underline{\sigma}_{\Lambda^c}$ and hence

e6.4.19
$$\frac{\mu(C^{\Lambda_N}_{\underline{\sigma}_{\Lambda_0}\underline{0}})}{\mu(C^{\Lambda_N}_{\underline{0}})}\xrightarrow[N\to\infty]{}e^{-\sum_{X\subset\Lambda_0}\Phi_X(\underline{\sigma}_X)}.\tag{6.4.19}$$

This means that $\mu \in G(\Phi)$ determines uniquely $\overline{U}_{\Lambda_0}(\underline{\sigma}_{\Lambda_0}) = \sum_{X\subset\Lambda_0} \Phi_X(\underline{\sigma}_X)$
N6.4.5 for all $\sigma_{\Lambda_0} \in \{0,\ldots,n\}^{\Lambda_0}_T$.[5] In turn $\overline{U}_{\Lambda_0}$ determines Φ recursively as

e6.4.20
$$\Phi_X(\underline{\sigma}_X) = \sum_{R\subset X} (-1)^{|R|}\,\overline{U}_R(\underline{\sigma}_R), \tag{6.4.20}$$

therefore $\mu \in G(\Phi)$ determines Φ uniquely if $\Phi \in \widetilde{B}$. ■

We conclude this section by discussing an interesting property related to the exponential mixing properties of the SRB distributions. If (Ω, S) is an Anosov system with a SRB distribution μ and f is a Hölder continuous function the limit as $N \to \infty$ of the dispersion of the variable $N^{-\frac{1}{2}} \sum_{k=0}^{N-1} f(S^k x)$ will be finite if f has zero SRB average. It will be given by the exponentially rapidly converging series $D = \sum_{k=-\infty}^{\infty} \langle f(S^k\cdot) f(\cdot)\rangle_\mu$ where $\langle\cdot\rangle_\mu$ denotes average with respect to μ. Clearly $D \ge 0$ and a natural question is whether this can vanish. The following proposition gives a necessary and sufficient condition (Livsic–Sinai).

P6.4.3 **(6.4.3) Proposition:** (A vanishing dispersion condition)
Let (Ω, S) be an Anosov system and let f be a Hölder continuous function with vanishing SRB average, $\int f d\mu = 0$, and with vanishing dispersion, $D \stackrel{def}{=} \sum_{k=-\infty}^{\infty} \langle f(S^k\cdot) f(\cdot)\rangle_\mu = 0$. Let $A(x)$ be the energy function for the SRB distribution μ, i.e. the energy function A, see definition (5.1.1), associated with the potential Φ for the SRB distribution. Then
(i) A and $A+f$ are equivalent energy functions, i.e. the same Gibbs distribution μ maximizes $s(\mu') - \mu'(A)$ and $s(\mu') - \mu'(A+f)$ over $\mu' \in \mathcal{M}^t_e(\Omega, S)$, cf. *corollary (6.2.1).*
(ii) $f(x) = F(x) - F(Sx)$ for a suitable F which is Hölder continuous on Ω.
Property (ii) implies that a necessary and sufficient condition in order to have $D > 0$ is that the average of f along a periodic orbit does not vanish.

See Appendix (6.4).

Appendix 6.4: *Vanishing dispersion conditions*

The following lemma shows that $D = 0$ has interesting implications and prepares the proof of proposition (6.4.3)

L6.4.1 **(6.4.1) Lemma:** (Vanishing dispersion: a L_2 condition)
Let (Ω, S) be an Anosov system. A necessary and sufficient condition in order that a Hölder continuous function f with zero SRB average ($\int d\mu f = 0$) has a zero dispersion $D \stackrel{def}{=} \sum_{k=-\infty}^{\infty} \int f(S^k x) f(x) \mu(dx)$ in the SRB distribution μ for (Ω, S) is that there is a function $F(x)$ defined on a set $V_0 \subset \Omega$ such that

eA6.4.1
$$f(x) = F(x) - F(Sx), \qquad \text{for } x \in V_0 \text{ and } \mu(V_0) = 1 \tag{A6.4.1}$$

and there exist constants $C, a > 0$ and $V \subset \Omega \times \Omega, \mu \times \mu(V) = 1$ such that

[5] With $\{0,\ldots,n\}^{\Lambda_0}_T$ we mean (naturally) the set of the elements of $\{0,\ldots,n\}^{\Lambda_0}$ that can be extended in at least one way to an element of $\{0,\ldots,n\}^{\mathbb{Z}}_T$.

eA6.4.2
$$|F(x) - F(y)| < C|x-y|^\alpha \qquad \textit{for all } (x,y) \in V \tag{A6.4.2}$$

Hence F is bounded μ–almost everywhere.

Remark: From (A6.4.2) we *cannot* (yet) conclude that F is Hölder continuous because $V \neq \Omega \times \Omega$.

Proof: We discuss the case of 2–dimensional Anosov maps, for simplicity. Consider the sum $F_N(x) \stackrel{def}{=} \sum_{k=0}^{N-1} f(S^k x)$. If $D = 0$ one finds that $\int d\mu\,(\sum_{k=0}^{N-1} f(S^k x))^2$ is bounded uniformly in N. Therefore the sequence $\sum_{k=0}^{N-1} f(S^k x)$ is bounded in $L_2(\mu)$. Furthermore for all g smooth $\lim_{N\to\infty} \int \sum_{j=0}^{N-1} f(S^j x) g(x) \mu(dx)$ exists (by mixing) so that the limit $F(x) = \lim_{N\to\infty} \sum_{k=0}^{\infty} f(S^k x)$ exists weakly in $L_2(\mu)$ and (therefore) $f(x) = F(x) - F(Sx)$ μ–almost everywhere. The difficulty is in the proof of the weak kind of Hölder continuity of F claimed in (A6.4.2).

The same arguments as above can be made for time tending to $-\infty$: thus we can define a function $F_-(x) = \sum_{k=-1}^{-\infty} f(S^k x)$. The function $F_-(x) + F(x)$ is however invariant: and therefore it must be constant. The constant must be 0 because f has zero average.

Let $\xi_0 \in \Omega$ and let $R = W^u_\gamma(\xi_0) \times W^s_\gamma(\xi_0)$ be an S–rectangle, in the sense of definition (4.2.2), with center ξ_0 and pair of axes $C = W^u_\gamma(\xi_0)$ and $D = W^s_\gamma(\xi_0)$ with the notation of Section §4.2, cf. Fig.(4.2.1). The generic point in R has the form $[\varphi, \varphi']$ with $\varphi \in C, \varphi' \in D$, where $[\varphi, \varphi'] = W^s_\gamma(\varphi) \cap W^u_\gamma(\varphi')$. We can represent, by corollary (6.3.3), the integration of a function G on R as

eA6.4.3
$$\int_R G(x)\mu(dx) = \int_C d\sigma_\varphi \int_D \nu(d\varphi')\rho([\varphi,\varphi'])G([\varphi,\varphi']) \tag{A6.4.3}$$

where $d\sigma_\varphi$ is the arc length on C and ν is a suitably defined finite measure on the arc D and $\rho(\varphi, \varphi') = \prod_{k=0}^{\infty} \frac{\lambda_u^{-1}(S^k[\varphi,\varphi'])}{\lambda_u^{-1}(S^k\varphi)}$; note that $\rho(\varphi, \varphi')$ is Hölder continuous in $(\varphi, \varphi') \in C \times D$ and bounded away from $0, \infty$.

The key remark is that the function on $V \times V$ defined by $F_N([x,y])$ is weakly convergent
N6.4.6 to the limit $F([x,y])$ in $L_2(\mu)$. Indeed let $\Gamma(x,y)$ be a bounded test function. Then [6]

eA6.4.4
$$\begin{aligned}\int_R \Gamma(x,y)F_N([x,y])\mu(dx)\mu(dy) = &\int_{C\times C} d\sigma_{\varphi_0} d\sigma_{\varphi_1} \int_{D\times D} \nu(d\varphi'_0)\nu(d\varphi'_1)\cdot \\ &\cdot \rho(\varphi_0,\varphi'_0)\rho(\varphi_1,\varphi'_1)\Gamma([\varphi_0,\varphi'_0],[\varphi_1,\varphi'_1])F_N([\varphi_0,\varphi'_1]),\end{aligned} \tag{A6.4.4}$$

which, multiplying and dividing by $\rho(\varphi_0, \varphi'_1)$ can be written $\int_R F_N(z)\widetilde\Gamma(z)\,\mu(dz)$ where, if $z = [\varphi_0, \varphi'_1]$,

eA6.4.5
$$\widetilde\Gamma(z) = \int_{C\times D} d\sigma_{\varphi_1}\nu(d\varphi'_0)\Gamma([\varphi_0,\varphi'_0],[\varphi_1,\varphi'_1])\frac{\rho(\varphi_0,\varphi'_0)\rho(\varphi_1,\varphi'_1)}{\rho(\varphi_0,\varphi'_1)} \tag{A6.4.5}$$

is a bounded function.

Therefore $\int \Gamma(x,y)F_N([x,y])\mu(dx)\mu(dy) \xrightarrow[N\to\infty]{} \int_R \mu(dz)F(z)\widetilde\Gamma(z)$ because F_N converges weakly to F in $L_2(\mu)$. By going backwards this means that $F_N([x,y])$ is weakly convergent to $F([x,y])$ in $L_2(\mu \times \mu)$.

The function $\sum_{k=0}^{N-1}(f(S^k x) - f(S^k[x,y]))$ is bounded by

eA6.4.6
$$\Big|\sum_{k=0}^{N-1}(f(S^k x) - f(S^k[y,x]))\Big| \le C|x-y|^\alpha, \tag{A6.4.6}$$

[6] We use that $[x,y] = [\varphi_0, \varphi'_1]$ if $x = [\varphi_0, \varphi'_0]$ and $y = [\varphi_1, \varphi'_1]$.

if C is a suitable constant and α is the Holder continuity exponent of f times the logarithm of the constant λ bounding the minimum expansion rate. This is so because the points x and $[x,y]$ are on the same stable manifold and therefore S^kx and $S^k[x,y]$ get closer and closer exponentially fast with their distance bounded proportionally to λ^{-k}. Hence for a suitable $C>0$

eA6.4.7
$$|F(x)-F[x,y]|\le C|x-y|^\alpha \tag{A6.4.7}$$

holds almost everywhere with respect to $\mu\times\mu$. Likewise $|F_-(y)-F_-([x,y])|\le C|x-y|^\alpha$ almost everywhere with respect to $\mu\times\mu$. Since $F_-(x)=-F(x)$ holds μ–almost everywhere we get

eA6.4.8
$$|F(x)-F(y)|\le 2^{1-\alpha}C|x-y|^\alpha \qquad \text{for all } (x,y)\in V \tag{A6.4.8}$$

where V is such that $\mu\times\mu(V)=\mu\times\mu(R)=1$.

Hence for μ–almost all x one has $|F(x)-F(y)|\le 2^{1-\alpha}C\,\mathrm{diam}(R)^\alpha$ for μ-almost all y: and for μ-almost all y one has $|F(y)|\le 2^{1-\alpha}C\,\mathrm{diam}(R)^\alpha$ so that the function F is bounded uniformly μ-almost everywhere. ∎

If b is a μ-almost everywhere bound on $|F(x)|$ the above lemma implies that

eA6.4.9
$$e^{-2b}\mu(dx)\le\mu_N(dx)=\mu(dx)e^{\sum_{j=-N}^N f(S^jx)}\le e^{2b}\mu(dx) \tag{A6.4.9}$$

because $\sum_{j=-N}^N f(S^jx)=F(S^Nx)-F(S^{-N}x)$. On the other hand it follows also that the limit as $N\to\infty$ of $c_N\mu_N$ with c_N being a normalization factor ($e^{-2b}\le c_N\le e^{2b}$) is the Gibbs distribution with energy function $A(x)+f(x)$ if $A(x)$ is the potential function for μ, cf. (6.2.11). Indeed by proposition (5.2.1) we know that the Gibbs distributions associated with A and with $A+f$ are ergodic: they coincide with μ and with the limit as $N\to\infty$ of μ_N, respectively, hence (A6.4.9) implies that the two distributions are absolutely continuous with respect to each other hence they coincide. Therefore the following result holds.

C6.4.2 **(6.4.2) Corollary:** *Let (Ω,S) be an Anosov system and let f be a Hölder continuous function with vanishing SRB average, $\int f d\mu=0$, and vanishing dispersion $D=\sum_{j=-\infty}^\infty\mu\big(f(S^j\cdot)f(\cdot)\big)=0$. If $A(x)$ is the potential function for μ then $A(x)+f(x)$ is also a potential function for μ.*

The problem posed before Lemma (6.4.1) above is solved by the following proposition

Proof of proposition (6.4.3): The argument preceding corollary (6.4.2) shows that $|\sum_{j=-N}^N f(S^jx)|<2b$ everywhere (because f is continuous). Therefore by the same argument one sees that given any Gibbs distribution μ_0 with potential function A_0 the potentials A_0 and A_0+f are equivalent. Therefore we can apply (i) of proposition (6.4.1) to conclude that there exists F continuous and a constant C such that $f(x)=F(x)-F(Sx)+C$; in our case $C=0$ because $\int fd\mu=0$ and (i) and (ii) are checked. Item (iii) follows from (ii). ∎

Problems for §6.4

Q6.4.1 **[6.4.1]:** Consider the potential on $\{0,1\}^{\mathbb{Z}}$

$$\Phi_X(\underline{\sigma}_X)=0 \qquad |X|\ge 2\ , \qquad \Phi_{\{0\}}(\underline{\sigma})=h\sigma.$$

Find an element Φ' of B equivalent to it (*i.e.* such that $G(\Phi)=G(\Phi')$) but with potential with non vanishing many-body components and which: (a) decay exponentially ($\Phi'\in B_0^h$) (see proposition (6.4.1) and definition (6.4.1)), (b) is such that $\|\Phi'\|_1=+\infty$, (c) does not contain potentials with more than three bodies (*here a, b, c are meant as mutually excluding properties*).

Q6.4.2 **[6.4.2]:** Does it exist $\Phi\in B$ that is not equivalent to some $\Phi'\in\widetilde{B}$? (see proposition (6.4.1)).

Q6.4.3 **[6.4.3]:** Check that in the theory of Gibbs measures the function $A_\Phi = \sum_{X\ni 0} \frac{\Phi_X(\underline{\sigma}_X)}{|X|}$ can be replaced without substantial modifications by $A'_\Phi(\underline{\sigma}) = \sum_{X\ni 0,\, X\ge 0} \Phi_X(\underline{\sigma}_X)$, ($X \ge 0$ means $X \subset \mathbb{Z}^+$).

Q6.4.4 **[6.4.4]:** Making use of the results in problem [6.4.3] show that the potential $\Phi_X = 0$ unless $|X| = 3$ and $\Phi_{\{0,j,j+1\}}(\sigma_0, \sigma_j, \sigma_{j+1}) = (\sigma_j - \sigma_{j+1})\,\Phi_j$ and zero otherwise, is equivalent to zero if $\sum_j |\Phi_j| < +\infty$.

Q6.4.5 **[6.4.5]:** Consider the group of the $(n+1)^{\rm th}$ roots of unity: $r_0,\ r_1, \ldots, r_n$. Show that every potential for $\{0,\ldots,n\}^{\mathbb{Z}}$, $\Phi \in B^h_s$, with s large enough, is equivalent to a potential $\widetilde\Phi$ (*spin potential*) such that $\widetilde\Phi_X(\underline{\sigma}_X) = \Phi_X \cdot\ {\rm Re}\,\Big(\prod_{j\in X} r_{\sigma_j}\Big)$ with $\Phi_X \in \mathbb{R}$ and $\sum_{X\ni 0} e^{\kappa {\rm diam}(X)} |\Phi_X| < +\infty$, with $\kappa > 0$ suitable.

Bibliographical note to §6.4

The contents of this section are based on [GR71], and on the interesting proposition (6.4.3) (cf. p. 78, theorem 5.7 and the bibliographical note at p. 96 in [Ru78]). This theorem can be extended to potentials more general than those verifying (6.4.1) (it suffices to push the precision of the estimates, giving up proving properties as strong as (6.4.6) or (6.4.9) but looking for the minimal conditions sufficient to obtain continuity of F). The vanishing dispersion analysis is taken, at Ruelle's suggestion, from [Ru78].

CHAPTER VII

Analyticity, singularity and phase transitions

§7.1 Polymers

There are various instances in which the construction of the Gibbs distributions can be performed in great detail, almost completely explicitly, allowing us to answer satisfactorily to questions concerning, for instance, mixing rates of Gibbs states and smoothness of their dependence on the potential.

The methods are based on series expansions and on the use of recurrence relations to put bounds on their addends: they apply really quite generally for Gibbs states on one-dimensional lattices but they can be extended to Gibbs states on lattices of dimension higher than 1 only at the cost of severe restrictions. We devote the present chapter to a discussion of such methods and to an attempt to present various results from the unified viewpoint which is the *polymer theory* and the associated *cluster expansion*, a more modern term to indicate a method based on equations well known in Statistical Mechanics under the name of *Kirkwood-Salsburg* equations and *Mayer-Montroll* equations, see [Ga00] and appendix to the following Section §(7.2).

Let $\mathcal{B}$ be the space of the potentials Φ for $\{0,\ldots,n\}_T^{\mathbb{Z}}$, cf. definition (5.1.1) and remark the change in notation: in this chapter we denote the space of the potentials with script $\mathcal{B}$'s with appropriate labels when necessary to avoid confusion with the many constants B appearing in the bounds that we discuss. It is natural to ask for regularity properties of the function $P(\Phi)$; hence we set the following definition.

D7.1.1 **(7.1.1) Definition:** (Smoothness and analyticity of Gibbs distributions) *Let $\mathcal{B}$ be the space of the potentials for symbolic dynamics of definition (5.1.1) and let $\widetilde{\mathcal{B}}$ be a subspace in $\mathcal{B}$ large enough to contain the space $\mathcal{B}_0$ of potentials vanishing for all sets X except for the translates of a single set X_0,* cf. *Section §6.1.*

(i) We say that $\Phi \to P(\Phi)$ is C^k smooth near $\Phi \in \widetilde{\mathcal{B}}$ on the subspace $\widetilde{\mathcal{B}}$ if the pressure $P(\Phi)$ and the corresponding Gibbs distribution $\mu_\Phi \in G(\Phi)$ are smooth of class C^k on $\widetilde{\mathcal{B}}$. This means that for all $V \subset \mathbb{Z}$, $\underline{\sigma}_V \in \{0,\ldots,n\}^V_T$ and $\Psi_1,\ldots,\Psi_d \in \widetilde{\mathcal{B}}$ the functions of $\lambda_1,\ldots,\lambda_d$

e7.1.1
$$P(\Phi + \sum_{j=1}^{d} \lambda_i \Psi_i), \qquad \mu_{\Phi+\sum_{j=1}^{d}\lambda_i\Psi_i}(C^V_{\underline{\sigma}_V}) \tag{7.1.1}$$

are smooth of class C^k and, respectively, of class C^{k-1} in the variables $\lambda_1,\ldots,\lambda_d$ in a neighborhood of the origin of $\mathbb{R}^d$.

(ii) We say that $\Phi \to P(\Phi)$ is analytic near Φ on the subspace $\widetilde{\mathcal{B}}$ if the functions in (7.1.1) are analytic near the origin.

Remarks: (1) To clarify the above settings we note that in applications relevant for us $\widetilde{\mathcal{B}}$ will be a subspace consisting of potentials verifying some extra condition besides that of being in $\mathcal{B}$: for instance that some other norm is also finite or small (see the examples in Section §(7.2)) besides the stability condition expressed by the finiteness of the norm in (5.1.5).

(2) The space $\mathcal{B}_0$ was used in Section §6.1 in the course of the proof that knowledge of the λ–derivatives of $P(\Phi+\lambda\Psi)$ at $\lambda = 0$ yields the translation invariant Gibbs distribution μ_Φ at least when the latter is unique, cf. (6.1.14), (6.1.33) and proposition (6.1.2). Furthermore existence of the derivative for all $\Psi \in \mathcal{B}_0$ implies uniqueness of the translation invariant Gibbs distribution; and continuity in Φ of the derivatives implies continuity of the probabilities $\mu_\Phi(C^J_{\sigma_J})$, hence of the Gibbs distribution as a function of Φ.

We shall discuss here a few rather general propositions about smooth or analytic Φ-dependence of μ_Φ which will also be useful for applications to dynamical systems. We have in mind providing an example of the construction of an SRB distribution and of the analysis of its regularity (analyticity) in nontrivial case exhibiting what is called *spatio–temporal chaos*. For this purpose it will be eventually necessary to consider symbolic dynamics on lattices $\mathbb{Z}^{d+1}$ with $d \ge 0$ (where the first d coordinates represent *space* and the last *time*). We follow the custom of calling the symbols σ of symbolic dynamics *spins*: if the number of values that the symbols can assume is $n+1$, say $\sigma \in \{0,1,\ldots,n\}$, we shall call the space $\Omega = \{0,1,\ldots,n\}^{\mathbb{Z}^{d+1}}$ the *configurations space* of a $\frac{n}{2}$ spin system on a $(d+1)$–dimensional (square) lattice. Hence the following definition will be useful.

D7.1.2 **(7.1.2) Definition:** (Potentials for spin systems on $\mathbb{Z}^d$) *A compatibility matrix for a d–dimensional lattice $\frac{n}{2}$-spin system is a func-*

tion $T_{\sigma,\{\sigma_1,\ldots,\sigma_{2d}\}} = 0,1$ *of* $2d+1$ *symbols (spins) in* $\{0,\ldots,n\}$*: a sequence* $\underline{\sigma} \in \{0,\ldots,n\}^{\mathbb{Z}^d}$ *labeled by the points in* $\mathbb{Z}^d$ *is called* T*–compatible if* $\prod_{\xi\in\mathbb{Z}^d} T_{\sigma_\xi,\{\sigma_{\xi_i}\}_{i=1}^{2d}} = 1$*, where the* $2d$ *nearest neighbors of* ξ *are labeled by* $i=1,\ldots,2d$ *(so that one has* $|\xi_i-\xi| = 1$ *for* $i=1,\ldots,2d$*). The* T*–compatible sequences will be denoted* $\{0,\ldots,n\}_T^{\mathbb{Z}^d}$ *and* $\{0,\ldots,n\}_T^X$ *will denote the possible restrictions to* X *of* T*–compatible sequences* $\underline{\sigma}$*.*
We shall say that T *is* transitive on the vacuum *if* $T_{0,\{\sigma_i\}_{i=1}^{2d}} = 1$ *for any choice of the labels* $\sigma_1,\ldots,\sigma_{2d}$*.*
Let $\mathcal{B}$ *be the space of the sequences* $\Phi = \{\Phi_X\}_{X\subset\mathbb{Z}^d}$ *parameterized by the finite subsets of* $\mathbb{Z}^d$*, consisting in the functions* $\Phi_X : \{0,\ldots,n\}_T^X \to \mathbb{R}$ *such*
N7.1.1 *that, having set* $\|\Phi_X\| = \max_{\underline{\sigma}\in\{0,\ldots,n\}^X} |\Phi_X(\underline{\sigma})|$*, one has* [1]

e7.1.2
$$\|\Phi\| \equiv \sup_{\xi\in\mathbb{Z}^d} \sum_{X\ni\xi} \|\Phi_X\| < +\infty. \tag{7.1.2}$$

We shall say that $\mathcal{B}$ *is the space of the* potentials *on* $\{0,\ldots,n\}_T^{\mathbb{Z}^d} \equiv \Omega$ *and we can define* A_Φ *via (5.1.6), and* $U_\Lambda^0(\underline{\sigma}_\Lambda)$ *via (5.1.11), with the new meaning of* $X,\Lambda,\Phi,\underline{\sigma}$*. We say that the interaction is a* particle potential *if* T *is transitive on the vacuum and* $\Phi_X(\underline{\sigma}_X) = 0$ *if* $\sigma_j = 0$ *for some* $j\in X$*.*
We say that a probability distribution μ *on* Ω *is a Gibbs distribution for the potential* $\Phi\in\mathcal{B}$ *if for all finite* $\Lambda\subset\mathbb{Z}^d$ *the conditional probabilities for finding* $\underline{\sigma}_\Lambda$ *in* Λ *given* $\underline{\sigma}_{\Lambda^c}$ *in the complement of* Λ *verifies the DLR equations (5.1.9).*
If Φ *is invariant under translations the pressure* $P(\Phi)$ *can be defined by the limit in (6.1.1) and it can be checked to verify proposition (6.1.1) by adapting its proof to the new context.*

Remark: The values of the symbols σ have no *a priori* meaning and they will occasionally be different from $0,\ldots,n$ (*e.g.* $1,\ldots,n$ or other non-numerical symbols).

The analysis of the regularity of $P(\Phi)$ and of the Gibbs distributions μ_Φ for spin systems in $\mathbb{Z}^d$, when they are uniquely defined, will be done gradually. In the cases $d>1$, unlike what we saw for $d=1$, cf. proposition (5.2.1), we shall have to impose rather strong conditions on Φ. We begin by studying general representations of sums of the form

e7.1.3
$$Z_\Phi(\Lambda) = \sum_{\underline{\sigma}\in\{0,1,\ldots,n\}^\Lambda} e^{-\sum_{Y\subset\Lambda}\Phi_Y(\underline{\sigma}_Y)}, \tag{7.1.3}$$

where the potentials Φ are not even supposed to be translation invariant and Λ denotes a finite subset of arbitrary shape contained in a d–dimensional lattice $\mathbb{Z}^d$. The following definition will be used below.

---

[1] Note that the norm defined in (7.1.2) is not the same as in (5.1.5).

D7.1.3 **(7.1.3) Definition:** (Tree length of a set)
N7.1.2 *Given a finite subset $X \subset \mathbb{Z}^d$ consider the trees* [2] *that have all the points of X as nodes. We call such trees* spanning trees *for X. The length of a tree ϑ will be the sum $\delta_\vartheta(X)$ of the lengths of its branches. The minimum value of $\delta_\vartheta(X)$ will be called the* tree length *of X or the* tree distance *of the points of X, and it will be denoted by $\delta(X)$.*

Remarks: (1) If $d = 1$ one has $\delta(X) = \mathrm{diam}(X)$ and if X is an interval $\delta(X) = |X| - 1$.
(2) The number of different trees is of the order of $|X|!$ if $|X|$ is the number of points of X; see problems at the end of the section.

Consider a function $\gamma \to \zeta(\gamma)$ defined on the finite sets $\emptyset \neq \gamma \subset \Lambda$: we call the sets γ *polymers* and $\zeta(\gamma)$ their *activities.* The following key restriction will be assumed

e7.1.4
$$|\zeta(\gamma)| \le b_0^2 \nu_0^{2|\gamma|} e^{-2\kappa_0 \delta(\gamma)} \stackrel{def}{=} z(\gamma)^2, \tag{7.1.4}$$

where $\nu_0, b_0, \kappa_0 > 0$ and $\delta(\gamma)$ is the *tree length* of the set γ. The length scale $\kappa_0^{-1} > 0$ can be considered, as we shall see below, to be an estimate of the *size* of the polymers. The quantity $|\gamma|$ denotes the number of points in γ. The 2 in the exponents is there for esthetic reasons (*i.e.* to obtain (7.1.30) below) while the subscript 0 is for easier reference in what follows.

If $k(d, \kappa_0) \stackrel{def}{=} 2\, e \sum_{0 \neq \xi \in \mathbb{Z}^d} e^{-\kappa_0 |\xi|}$ and $B_0 \stackrel{def}{=} 2 b_0$ then (7.1.4) implies (see also problem [7.1.1])

e7.1.5
$$\sup_{\xi \in \mathbb{Z}^d} \sum_{\gamma \ni \xi, \delta(\gamma) \ge r} |\zeta(\gamma)|^{1/2} \le B_0 \nu_0 e^{-\frac{\kappa_0}{2} r} \quad \text{if} \quad \nu_0\, k(d, \tfrac{\kappa_0}{2}) < 1. \tag{7.1.5}$$

Note that $k(d, \kappa_0) = O(\kappa_0^{-d})$ for κ_0 small and $k(d, \kappa_0) = O(e^{-\kappa_0})$ for κ_0 large.

(A) **Remarkable algebraic identities.**

A natural adaptation, see [GK70], of an original method of proof of smoothness of the functions in (7.1.1), [Ru69], led to the following *algebraic approach*, which has become widely known as *cluster expansion* after having been rediscovered several times in the span of a decade or two.

Consider the set of *finite polymer configurations* $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ in which the polymers γ_i are allowed to overlap and even to coincide. Formally one such configuration is (therefore) a function $\gamma \to \Gamma(\gamma)$ with non negative integer values such that $\sum_\gamma \Gamma(\gamma) \equiv N(\Gamma) < \infty$. The number $\Gamma(\gamma)$ will be called the *multiplicity* of γ in Γ. The just introduced configurations can be

[2] Given a set of points $\xi_1, \ldots, \xi_n$ we call *tree on* $\xi_1, \ldots, \xi_n$ a graph without loops connecting all the points. The points are called the *nodes* (or vertices) of the tree, while the lines connecting the nodes are sometimes called *branches.* The number of lines is $n - 1$. A connected set of lines is called a *path*: given any two nodes of the tree there is only one path connecting them.

added by setting $(\Gamma_1+\Gamma_2)(\gamma)\stackrel{def}{=}\Gamma_1(\gamma)+\Gamma_2(\gamma)$ and, if $\Gamma_1(\gamma)\ge\Gamma_2(\gamma)$, they can also be *subtracted*.

Given a polymer configurations Γ we denote

e7.1.6
$$\widetilde\Gamma \text{ the set of polymers } \gamma \text{ such that } \Gamma(\gamma)>0, \tag{7.1.6}$$

hence $\widetilde\Gamma$ is the *support* of Γ and it consists of the polymers that are in Γ, each counted without taking multiplicity into account. If all polymers $\gamma\in\widetilde\Gamma$ are contained in a set Λ we say that Γ is contained in Λ and write $\Gamma\subset\Lambda$. If $\Gamma(\gamma)=0,1$ we say that Γ is a polymer configuration *without multiplicities* and we write also (improperly) $\Gamma=\widetilde\Gamma$. The case $\Gamma(\gamma)\equiv0$ for all γ will also be allowed and it will be called the *vacuum* or the *empty configuration* (and it will be denoted with $\emptyset$).

Let $\mathcal F$ be the space of real valued functions Ψ of the polymer configurations $\Gamma\to\Psi(\Gamma)$ such that $\sup_{N(\Gamma)=n}|\Psi(\Gamma)|<\infty$ for all $n\ge0$. Let $\mathcal F_0,\mathcal F_1$ be the subspaces of $\mathcal F$ consisting in the functions Ψ such that $\Psi(\emptyset)=0$ or $\Psi(\emptyset)=1$, respectively.

Given $\Psi_1,\Psi_2\in\mathcal F$ we define the *convolution product* $\Psi_1*\Psi_2$ by

e7.1.7
$$\Psi_1*\Psi_2(\Gamma)=\sum_{\Gamma_1+\Gamma_2=\Gamma}\Psi_1(\Gamma_1)\Psi_2(\Gamma_2), \tag{7.1.7}$$

which is always a sum over finitely many addends. Note that if we define $\mathbf 1(\Gamma)=1$ if $\Gamma=\emptyset$ and $\mathbf 1(\Gamma)=0$ otherwise the function $\mathbf 1$ so defined is in $\mathcal F_1$ and $\mathbf 1*\Psi\equiv\Psi*\mathbf 1\equiv\Psi$ for all $\Psi\in\mathcal F$, *i.e.* $\mathbf 1$ is the identity for the product in (7.1.7).

The functions in $\mathcal F_1$ have the form $\Psi(\Gamma)=\mathbf 1(\Gamma)+\Psi_0(\Gamma)$ with $\Psi_0\in\mathcal F_0$. If $\Psi\in\mathcal F_0$ (*i.e.* $\Psi(\emptyset)=0$) we define the *exponential*

e7.1.8
$$\begin{aligned}(\mathrm{Exp}\Psi)(\Gamma)&=\sum_{n\ge0}\frac{1}{n!}\Psi^n(\Gamma)=\\&=\mathbf 1(\Gamma)+\sum_{n\ge1}\frac{1}{n!}\sum_{\Gamma_1+\ldots+\Gamma_n=\Gamma}\Psi(\Gamma_1)\ldots\Psi(\Gamma_n),\end{aligned} \tag{7.1.8}$$

where the power Ψ^n is intended in the sense of the product in (7.1.7) and we set $\Psi^0(\Gamma)\stackrel{def}{=}\mathbf 1(\Gamma)$. For each Γ the sum in the r.h.s. of (7.1.8) involves only a finite sum.

The *logarithm* can also be defined on $\mathcal F_1$: if $\Psi\in\mathcal F_1$ and $\Psi=\mathbf 1+\Psi_0$ with $\Psi_0\in\mathcal F_0$ then we set

e7.1.9
$$\begin{aligned}(\mathrm{Log}\Psi)(\Gamma)&=\sum_{n\ge1}\frac{(-1)^{n+1}}{n}\Psi_0^n(\Gamma)=\\&=\sum_{n\ge1}\frac{(-1)^{n+1}}{n}\sum_{\Gamma_1+\ldots+\Gamma_n=\Gamma}\Psi_0(\Gamma_1)\ldots\Psi_0(\Gamma_n),\end{aligned} \tag{7.1.9}$$

where the sum is finite and $\operatorname{Log}\Psi \in \mathcal{F}_0$; one has also $\operatorname{Log}\operatorname{Exp}\Psi_0 \equiv \Psi_0$ for all $\Psi_0 \in \mathcal{F}_0$ and $\operatorname{Exp}\operatorname{Log}\Psi = \Psi$ for all $\Psi \in \mathcal{F}_1$. Furthermore $\operatorname{Exp}(\Psi_1 + \Psi_2) \equiv \operatorname{Exp}(\Psi_1) * \operatorname{Exp}(\Psi_2)$ if $\Psi_1, \Psi_2 \in \mathcal{F}_0$.

Let $\chi(\Gamma)$ be a *multiplicative function i.e.* a function such that $\chi(\Gamma_1+\Gamma_2) = \chi(\Gamma_1)\chi(\Gamma_2)$. A notable example is the function $\chi_V(\Gamma) = 1$ if all $\gamma \in \Gamma$ are contained in V (*i.e.* if $\Gamma \subset V$) and 0 otherwise. The following properties are, as we shall see, the main reason for introducing the above convolution property:

e7.1.10
$$\begin{aligned}&\sum_\Gamma (\Psi_1 * \Psi_2)(\Gamma)\chi(\Gamma) = \Big(\sum_\Gamma \Psi_1(\Gamma)\chi(\Gamma)\Big)\Big(\sum_\Gamma \Psi_2(\Gamma)\chi(\Gamma)\Big),\\ &\sum_\Gamma (\operatorname{Exp}\Psi)(\Gamma)\chi(\Gamma) = e^{\sum_\Gamma \Psi(\Gamma)\chi(\Gamma)} \qquad \text{for all } \Psi \in \mathcal{F}_0,\end{aligned} \tag{7.1.10}$$

provided $\sum_\Gamma |\Psi_i(\Gamma)| < \infty$ and χ is multiplicative and bounded.

(B) *An application: converting a sum of exponentials into an exponential of a sum.*

Given a polymer configuration Γ we define its total activity $\zeta(\Gamma)$ simply as the product of the activities of the polymers in Γ each counted according to its multiplicity $\Gamma(\gamma)$, *i.e.*

e7.1.11
$$\zeta(\Gamma) = \prod_{\gamma \in \widetilde{\Gamma}} \zeta(\gamma)^{\Gamma(\gamma)}. \tag{7.1.11}$$

The function $\zeta(\Gamma)$ is in $\mathcal{F}_1$ and

e7.1.12
$$\zeta(\Gamma_1 + \Gamma_2) \equiv \zeta(\Gamma_1)\zeta(\Gamma_2), \tag{7.1.12}$$

so that it is a multiplicative function.

Therefore we can define the partition function $Z(\Lambda)$ for the *configurations* $\Gamma = \{\gamma_1, \ldots, \gamma_q\}$ *of nonoverlapping polymers* in Λ as

e7.1.13
$$Z(\Lambda) \stackrel{def}{=} \sum_{\substack{\Gamma \subset \Lambda \\ \Gamma = \{\gamma_1,\ldots,\gamma_q\},\, \gamma_i \cap \gamma_j = \emptyset}} \zeta(\Gamma), \tag{7.1.13}$$

where the case $q = 0$ represents the contribution arising from the empty configuration and it is, by definition, a term in the sum equal to 1.

We shall say that a polymer configuration $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ is *compatible* if no overlapping occurs, *i.e.* if $\gamma_i \cap \gamma_j = \emptyset$ for all $i \neq j$. Hence (7.1.13) is a sum over compatible polymer configurations.

The *partition function* $Z(\Lambda)$ can be written by considering the function $\varphi(\Gamma) = 1$ if Γ contains no pairs of overlapping polymers or if $\Gamma = \emptyset$ and $\varphi(\Gamma) = 0$ otherwise, which is also a function in $\mathcal{F}_1$. Calling $\varphi^T(\Gamma) \stackrel{def}{=} (\operatorname{Log}\varphi)(\Gamma)$ we get

e7.1.14
$$Z(\Lambda) = \sum_\Gamma \varphi(\Gamma)\zeta(\Gamma)\chi_\Lambda(\Gamma) = e^{\sum_\Gamma \varphi^T(\Gamma)\zeta(\Gamma)\chi_\Lambda(\Gamma)}, \tag{7.1.14}$$

if $\sum_{\Gamma\subseteq\Lambda}|\varphi^T(\Gamma)||\zeta(\Gamma)|<\infty$, cf. (7.1.10), because the functions $\zeta(\Gamma)$ and $\chi(\Gamma)$ are multiplicative.

Remarks: (1) It is interesting to see what the above formulae mean in the simple case in which Λ consists of a single point. In this case there is only one polymer ξ and we can call its activity $\zeta(\xi)\stackrel{def}{=}z$. The polymer can be repeated several times in a polymer configuration Γ and we can denote by $n\xi$ the polymer Γ consisting of ξ repeated n times. Then (7.1.14) becomes the elementary relation $Z(\xi)=(1+z)=e^{\log(1+z)}$, hence

e7.1.15
$$\varphi^T(n\xi)=\frac{(-1)^{n+1}}{n}. \tag{7.1.15}$$

Therefore (7.1.14) is a natural generalization of this elementary formula and it achieves the feat of expressing a sum as the exponential of a sum (*i.e.* it is an algorithm to evaluate a logarithm).
(2) Note that φ^T is a purely *combinatorial* function (see equation (7.1.22) for a more explicit expression), while $\zeta(\gamma)$ verifies the bounds (7.1.5) if ν_0, see (7.1.4), is small enough.

(C) *An example of cluster expansion.*

The following proposition holds.

P7.1.1 **(7.1.1) Proposition:** (Polymers and cluster expansion)
Let $\Lambda\subset\mathbb{Z}^d$ be a finite set and let $\gamma\to\zeta(\gamma)$ be a polymer activity, for polymers $\gamma\subset\Lambda$, verifying, cf. *(7.1.4),*

e7.1.16
$$|\zeta(\gamma)|\le b_0^2\nu_0^{2|\gamma|}e^{-2\kappa_0\delta(\gamma)}\stackrel{def}{=}z(\gamma)^2, \tag{7.1.16}$$

*for some $b_0,\nu_0,\kappa_0>0$. Suppose that one has (*cf. *(7.1.5))*

e7.1.17
$$\big((1+B_0)e^{B_0\nu_0}+k(d,\kappa_0/2)\big)\nu_0<1, \tag{7.1.17}$$

with $B_0=2b_0$. Then the sum $Z(\Lambda)$ in (7.1.13) can be expressed as

e7.1.18
$$Z(\Lambda)=\exp\sum_{\Gamma\subset\Lambda}\varphi^T(\Gamma)\zeta(\Gamma), \tag{7.1.18}$$

where Γ is a polymer configuration (not necessarily such that $\widetilde\Gamma\equiv\Gamma$). And the functions $\varphi^T(\Gamma)$ satisfy

e7.1.19
$$\sup_{\xi\in\Lambda}\sum_{\Gamma\ni\xi,\delta(\Gamma)\ge r}|\varphi^T(\Gamma)||\zeta(\Gamma)|<B_0\,\nu_0\,e^{-\frac{\kappa_0}{2}r}. \tag{7.1.19}$$

Proof: Define the *differentiation operation* as

e7.1.20
$$(D_\Gamma\Psi)(H)\stackrel{def}{=}\Psi(\Gamma+H)\frac{(\Gamma+H)!}{H!} \tag{7.1.20}$$

with $\Gamma! = \prod_{\gamma\in\widetilde\Gamma}\Gamma(\gamma)!$. The name is attributed because of the validity of
N7.1.3 the following rules[3]

e7.1.21
$$\begin{aligned}
&D_\gamma(\Psi_1*\Psi_2)=(D_\gamma\Psi_1)*\Psi_2+\Psi_1*(D_\gamma\Psi_2),\\
&\frac{D_\Gamma(\Psi_1*\Psi_2)}{\Gamma!}=\sum_{\Gamma_1+\Gamma_2=\Gamma}\Big(\frac{D_{\Gamma_1}\Psi_1}{\Gamma_1!}\Big)*\Big(\frac{D_{\Gamma_2}\Psi_2}{\Gamma_2!}\Big),\\
&D_\gamma\mathrm{Exp}\Psi=D_\gamma\Psi*\mathrm{Exp}\Psi,\\
&\frac{D_\Gamma\,\mathrm{Exp}\Psi}{\Gamma!}=\sum_{n\ge1}\frac1{n!}\sum_{\sum_{i=1}^n\Gamma_i=\Gamma}\Big(\prod_i\frac{D_{\Gamma_i}\Psi}{\Gamma_i!}\Big)*(\mathrm{Exp}\Psi).
\end{aligned}\tag{7.1.21}$$

The second relation above, *Leibniz rule*, follows from the combinatorial identity $\sum_{p_1+p_2=n}\binom{q_1}{p_1}\binom{q_2}{p_2}=\binom{q_1+q_2}{n}$ for all n,q_1,q_2 with $n\le q_1+q_2$.

The above definitions allow us to derive an expression for $\varphi^T(\Gamma)$ which is quite explicit and which has the merit of implying immediately that $\varphi^T(\Gamma)$
N7.1.4 vanishes for nonconnected Γ's.[4]

Given $\Gamma=\{\gamma_1,\ldots,\gamma_n\}$ consider the graph G visiting all points $1,\ldots,n$ (*i.e.* with nodes $1,\ldots,n$) and with edges connecting all pairs i,j such that γ_i,γ_j overlap. Then (see problem [7.1.6]) one has $\varphi^T(\gamma)=1$ and for $n>1$:

e7.1.22
$$\varphi^T(\Gamma)=\frac1{\Gamma!}\sum_{C\subseteq G}(-1)^{\text{number of edges in } C},\tag{7.1.22}$$

where the sum runs over all connected subgraphs C of G *which visit all the n points* $1,\ldots,n$. This expression, *Mayer's coefficients* formula, immediately implies that $\varphi^T(\Gamma)=0$ unless Γ is connected. Note that if we define $\overline\varphi(\Gamma)\stackrel{def}{=}\zeta(\Gamma)\varphi(\Gamma)$ and $\overline\varphi^T=\mathrm{Log}\,\overline\varphi$ then $\varphi^T(\emptyset)=1$ and (of course)

e7.1.23
$$\overline\varphi^T(\Gamma)=\zeta(\Gamma)\varphi^T(\Gamma).\tag{7.1.23}$$

There are important cancellations between the various terms in (7.1.22): the cancellations are essential to show that the sum in (7.1.22) adds up to a quantity that is *much smaller* than the number of terms in the sum (which if of the order of $2^{n^2/2}$ if Γ consists of n polymers).

Existence of cancellations can be proved in several ways: for instance, following a method introduced by Ruelle who called it *algebraic method*, [Ru69], by taking advantage of recursive relations well known in the theory

[3] One should realize that the check of the above relations, which is left to the reader, can be reduced to the case in which $\Gamma=n\gamma$, *i.e.* to the case in which there is only one polymer species γ; in the case of the first relation it is useful to consider the generating function $F(z)=\sum_{n=0}^\infty z^nD_\gamma(\Psi_1*\Psi_2)(n\gamma)\equiv\sum_n(n+1)z^n(\Psi_1*\Psi_2)((n+1)\gamma)$ and verify that the relation becomes the ordinary differentiation rule for a product of functions of z.

[4] Here we say that Γ is connected if given any pair $\gamma,\gamma'\in\Gamma$ there is a sequence $\gamma_1=\gamma,\gamma_2,\ldots,\gamma_k=\gamma'$ such that $\Gamma(\gamma_i)>0$ and γ_i intersects γ_{i+1}.

of the Kirkwood-Salsburg equations in statistical mechanics, [Ga00]. Note that if $\Psi(\emptyset) \neq 0$ and $\Psi(\Gamma) = 0$ for $\Gamma! \neq 1$ (*i.e.* $\Gamma \neq \widetilde{\Gamma}$) then one can define a unique Ψ^{-1} such that $\Psi^{-1} * \Psi = \mathbf{1}$. Then we can consider the quantity

e7.1.24
$$\Delta_\Gamma(H) = (\overline{\varphi}^{-1} * D_\Gamma \overline{\varphi})(H) = \sum_{H_1+H_2=H} \overline{\varphi}^{-1}(H_1)\,\overline{\varphi}(\Gamma + H_2), \tag{7.1.24}$$

where $\overline{\varphi} = \zeta\varphi$ and φ is defined before (7.1.14). It is important to note that $D_\Gamma\,\overline{\varphi}(H_2) = \overline{\varphi}(\Gamma + H_2)$ because $\overline{\varphi}(\Gamma + H_2)$ vanishes if $(\Gamma + H_2)! > 1$, so that $(\Gamma + H_2)!/H_2! \equiv 1$ when $\overline{\varphi}(\Gamma + H_2) \neq 0$ Therefore $\Delta_\Gamma(H) = 0$ if Γ is not compatible (while H needs not be compatible because $\Gamma \cup H_2$ and H_1 could be compatible even though H is not).

Since $\varphi(\gamma_1, \ldots, \gamma_n) = \prod_{i<j}(1 + g(\gamma_i, \gamma_j))$ with $g(\gamma, \gamma') = -1$ if $\gamma \cap \gamma' \neq \emptyset$ and $g(\gamma, \gamma') = 0$ otherwise, we can write

e7.1.25
$$\begin{aligned}\overline{\varphi}(\gamma + \Gamma + H_2) &= \zeta(\gamma)\,\overline{\varphi}(\Gamma + H_2) \prod_{\gamma' \in H_2} (1 + g(\gamma, \gamma')) = \\ &= \zeta(\gamma)\,\overline{\varphi}(\Gamma + H_2) \sum_{S \subseteq H_2}^{*} (-1)^{N(S)},\end{aligned} \tag{7.1.25}$$

if H_2 is without multiplicities and γ is compatible with Γ; the $*$ on the last sum means that it is a sum over the sets S of polymers all of which are incompatible with γ. Setting $H_2 = S + H_3$ we get

e7.1.26
$$\begin{aligned}\Delta_{\gamma+\Gamma}(H) &= \zeta(\gamma) \sum_{S \subseteq H}^{*} (-1)^{N(S)} \sum_{H_1+H_3=H-S} \overline{\varphi}^{-1}(H_1)\,\overline{\varphi}(\Gamma + S + H_3) = \\ &= \zeta(\gamma) \sum_{S \subseteq H}^{*} (-1)^{N(S)} \Delta_{S+\Gamma}(H - S),\end{aligned} \tag{7.1.26}$$

if γ is compatible with Γ.

In the sum the term with $S = \emptyset$ must be included and $\Delta_\emptyset(H) = \mathbf{1}(H)$ is the correct interpretation of the symbols arising in this case.

Equation (7.1.26) determines $\Delta_\Gamma(H)$ with $N(\Gamma) + N(H) = m + 1$ in terms of $\Delta_\Gamma(H)$ with $N(\Gamma) + N(H) = m$ for $m = 1, 2, \ldots$, making the following useful estimate possible. Let

e7.1.27
$$I_m \equiv \sup_{\substack{\gamma_1,\ldots,\gamma_q \\ m \geq q \geq 1}} \sum_{H:\, N(H)=m-q} |\Delta_{\{\gamma_1,\ldots,\gamma_q\}}(H)| \prod_{i=1}^{q} z(\gamma_i)^{-1}. \tag{7.1.27}$$

Summing (7.1.26) over H we see, if p denotes a point in the lattice, that I_{m+1} is the supremum of

$$\sum_{H\,:\,N(\Gamma)+N(H)=m} |\Delta_{\gamma+\Gamma}(H)|\, z(\gamma + \Gamma)^{-1} \leq$$

e7.1.28
$$\leq \sum_{H\,:\,N(\Gamma)+N(H)=m} \sum_{S\subseteq H}^{*} |\Delta_{\Gamma+S}(H-S)| z(\gamma+\Gamma)^{-1}|\zeta(\gamma)| \leq \tag{7.1.28}$$
$$\leq z(\gamma) \sum_{S\,:\,N(S)\leq m-N(\Gamma)}^{*} I_m\, z(S) \leq z(\gamma)\, I_m \sum_{n\geq 0} \frac{1}{n!}\Big(\sum_{\gamma'\cap\gamma\neq\emptyset} z(\gamma')\Big)^n \leq$$
$$\leq z(\gamma)\, I_m \exp(\sum_{p\in\gamma}\sum_{\gamma'\ni p} z(\gamma')) \leq z(\gamma)\, I_m \exp\Big(|\gamma| \max_p \sum_{\gamma'\ni p} z(\gamma')\Big) \leq z_1(\gamma) I_m,$$

with (see (7.1.4) and (7.1.5))

e7.1.29
$$\begin{aligned} z_1(\gamma) =& b_0 \nu_0^{|\gamma|} e^{-\kappa_0\delta(\gamma)} e^{|\gamma| \max_p \sum_{\gamma'\ni p} z(\gamma')} \\ \leq& b_0 \nu_0^{|\gamma|} e^{-\kappa_0\delta(\gamma)} e^{B_0\nu_0|\gamma|} = z(\gamma) e^{B_0\nu_0|\gamma|}, \end{aligned} \tag{7.1.29}$$

if $k(d,\frac{\kappa_0}{2})\nu_0 < 1$ and $B_0 = 2b_0$.

If $q=1$ the polymer configuration H must contain at least one polymer so that $\Delta_\emptyset(H) = \mathbf{1}(H) = 0$: this allows us to estimate I_m inductively starting from I_1. Indeed if $\nu_0 e^{B_0\nu_0} < 1$ and $2b_0\nu_0 e^{B_0\nu} < 1$, as assumed in (7.1.17), one has (*since* $\Delta_\emptyset(H) = \mathbf{1}(H) = 0$ for $H \neq \emptyset$) $z_1(\gamma) \leq \frac{1}{2}$ and $I_1 \leq \frac{1}{2}$. Hence

e7.1.30
$$I_m \leq 2^{-m} \qquad \text{for all } m \geq 1. \tag{7.1.30}$$

The bound allows us to estimate $\overline{\varphi}^T$ because the third in (7.1.21) implies

e7.1.31
$$\Delta_\gamma(\Gamma) = (\overline{\varphi}^{-1} * D_\gamma\, \overline{\varphi})(\Gamma) = D_\gamma\, \overline{\varphi}^T(\Gamma) = \overline{\varphi}^T(\gamma+\Gamma)\frac{(\gamma+\Gamma)!}{\Gamma!}, \tag{7.1.31}$$

hence, making use of the above bound (7.1.30), we get

e7.1.32
$$\sum_\Gamma |\,\overline{\varphi}^T(\gamma+\Gamma)| \leq \sum_{\substack{m\geq 1\\ \Gamma\,:\,N(\Gamma)=m-1}} |\Delta_\gamma(\Gamma)| \leq z(\gamma) \sum_{m\geq 1} I_m. \tag{7.1.32}$$

Likewise we can bound, by (7.1.5),

e7.1.33
$$\sum_{\Gamma\ni p} |\,\overline{\varphi}^T(\Gamma)| \leq \sum_{\gamma\ni p}\sum_{\Gamma} |\,\overline{\varphi}^T(\gamma+\Gamma)| \leq B_0\nu_0. \tag{7.1.33}$$

Finally a bound that is useful in studying the mixing properties of the Gibbs distribution μ_Φ is a bound on the sum $\sum_{\Gamma\ni p\,:\,\Gamma\cap Q\neq\emptyset} |\,\overline{\varphi}^T(\Gamma)|$, where Q is any set in $\mathbb{Z}^d$. Let r be the distance between p and the set Q: proceeding in an analogous way we find, cf. (7.1.5),

e7.1.34
$$\sum_{\Gamma\ni p\,:\,\Gamma\cap Q\neq\emptyset} |\,\overline{\varphi}^T(\Gamma)| \leq B_0\nu_0\, e^{-\kappa_0 r/2}. \tag{7.1.34}$$

Taking as Q in (7.1.34) the set of point ξ such that $|\xi - p| > r$ we obtain (7.1.19). This completes the check of proposition (7.1.1). ∎

An immediate corollary to the above proposition is the following result.

C7.1.1 **(7.1.1) Corollary:** (Smoothness and polymer expansion)
Under the assumptions of proposition (7.1.1) suppose that the polymer activity depends on a parameter α and that its α-derivative $\partial_\alpha\zeta(\gamma)$ is continuous and verifies a bound

e7.1.35
$$|\partial_\alpha\zeta(\gamma)| < b_0'\nu_0^{a|\gamma|}e^{-a\kappa_0\delta(\gamma)}, \tag{7.1.35}$$

with $b_0', a > 0$ and $|\gamma|$ equal to the number of points in the polymer γ. Then for ν_0 small enough $|\Lambda|^{-1}\log Z(\Lambda)$ is continuously differentiable in α uniformly in Λ and if $\zeta(\gamma)$ is translation invariant the limit as $\Lambda\to\infty$ of $|\Lambda|^{-1}\log Z(\Lambda)$ is given by

e7.1.36
$$P = \lim_{\Lambda\to\infty}|\Lambda|^{-1}\log Z(\Lambda) = \sum_{\Gamma\ni 0}\frac{\varphi^T(\Gamma)\zeta(\Gamma)}{|n(\Gamma)|}. \tag{7.1.36}$$

where $n(\Gamma)$ is the number of distinct lattice points in the polymers of Γ.
Likewise if $\zeta(\gamma)$ is analytic in the parameter α and for α in a complex domain the inequality

e7.1.37
$$|\zeta(\gamma)| \le b_0\nu_0^{a|\gamma|}e^{-a\kappa_0\delta(\gamma)} \tag{7.1.37}$$

holds then $|\Lambda|^{-1}\log Z(\Lambda)$ is analytic in α in the same complex domain for all Λ and, in the translation invariant case, P is also analytic.

Remarks: (1) In the bounds (7.1.35) and (7.1.37) it is important that there is a factor $\nu_0^{a|\gamma|}$ for some $a>0$: a will in general be $a<2$ if ζ verifies the bound in (7.1.4). In fact one imagines that "part" of the constants in (7.1.4) is used up in order to check bounds like (7.1.35) and (7.1.37): the derivative $\partial_\alpha^k\zeta(\gamma)$ can be often bounded proportionally to $|\gamma|^k|\zeta(\gamma)|$, so that the smaller exponent a accounts for the necessity of bounding the extra $|\gamma|^k$ that may arise in attempting to differentiate with respect to parameters. For instance this happens if in the examples studied in Section §(7.2) one replaces the potential Φ with $\beta\Phi$ with β a parameter. A paradigmatic example, not discussed here, has been a case in which the polymers γ are the vertices of closed contours, joining points on the lattice $\mathbb{Z}^2$, of length $|\gamma|$ and $\zeta(\gamma)=e^{-\beta|\gamma|}$: the latter problem arises in the Ising model theory as developed in [MS67], see [Ga00] and [GMM73].
(2) Likewise in (7.1.37) the initial exponential decay $\nu_0^{2|\gamma|}$ may be partially used to bound the k-th derivative of ζ dimensionally (*i.e.* by Cauchy's theorem).

In the following section we apply the above propositions and corollary to derive some notable results on the theory of Gibbs states.

Problems for §7.1

Q7.1.1 **[7.1.1]:** (*Tree length properties*)
Let γ be a set in $\mathbb{Z}^d$ and let $\delta(\gamma)$ be the tree length of γ. Given $\kappa>0$ estimate the sum

$\sum_{\gamma\ni 0}\nu^{|\gamma|}e^{-\kappa\delta(\gamma)}\stackrel{def}{=}\overline{B}$ for ν small enough and determine the dependence of $\overline{B}$ on κ, d. (*Hint:* The sum over γ can be written as the sum over $q\ge 0$ of $q!^{-1}$ times the sum over the *ordered* q–ples $(\xi_1,\ldots,\xi_q)$ of points in $\mathbb{Z}^d$ and $e^{-\kappa\delta(\gamma)}$ can be bounded by the sum over the trees ϑ with $q+1$ nodes labeled $0,\xi_1,\ldots,\xi_q$ of $e^{-\kappa\delta_\vartheta(0,\xi_1,\ldots,\xi_q)}$, where $\delta_\vartheta(\gamma)$ is the tree length of γ measured on the particular tree ϑ, hence $\delta_\vartheta(\gamma)\ge\delta(\gamma)$. The sums over ξ_i can be performed one by one starting with the *external nodes of the tree.* Exploit the fact that the number of trees with n nodes is bounded by $e^n n!$: a more precise estimate is derived in the following problems.)

Q7.1.2 **[7.1.2]:** (*Taylor formula*)
For u real let us set $u^\Gamma\equiv\prod_{\gamma\in\widetilde{\Gamma}}u^{\Gamma(\gamma)}$. If $\Psi\in\mathcal{F}$, with $\sum_\Gamma|\Psi(\Gamma)|<\infty$, define, for v real, $\Psi_v=\sum_\Gamma\frac{v^\Gamma}{\Gamma!}D_\Gamma\Psi$; check that one has $\Psi_{v+u}(H)\equiv\sum_\Gamma\frac{u^\Gamma}{\Gamma!}D_\Gamma\Psi_v(H)$.

Q7.1.3 **[7.1.3]:** (*Counting rooted trees*)
Given n unit oriented segments labeled $1,2,\ldots,n$, call *origin* v of a segment the starting point of the segment (in the order established by its orientation); the other extreme will be called the *endpoint* of the segment. Fix *a priori* one of the segments (called *root branch*) and put it on the plane with its endpoint at the origin of the plane. Proceed by putting down on the plane one after the other in all possible ways the remaining $n-1$ segments attaching their endpoints to the origins of the segments that have already been put down. The resulting graph will be called a *rooted tree.* The starting points of the segments will be called nodes v of the tree, and d_v-1, $d_v\ge 1$, will denote the number of segments that enter the node v (d_v is the *branching number* of v). Show that the number of distinct rooted trees that can be formed with the given segments is bounded by $c^n n!$, for some constant c (*e.g.* $c=4$ would be a possible choice; more refined estimates are in the following problems).

Q7.1.4 **[7.1.4]:** (*Counting spanning trees*)
Given a set Γ of points on the plane, labeled $1,2,\ldots,n$, consider the set of spanning trees ϑ for Γ, *i.e.* the set of graphs without loops connecting the points. The points represent the nodes of the trees ϑ; for each node v of the tree we denote with d_v the number of branches that are incident with the node v; one has $d_v\ge 1$ and $\sum_v d_v=2(n-1)$. The numbers d_v will be called *branching numbers* of the tree. Show that the number of distinct trees that can be formed with fixed branching numbers $\{d_v\}$ is

$$\frac{(n-2)!}{\prod_v(d_v-1)!}.$$

(*Hint:* Each branch connecting two nodes v and v' can be considered as a contraction of a line coming out from the node v (exiting line) with a line entering into the node v' (entering line). There must be always a node, say v_0, with branching number $d_{v_0}=1$, so that we can orient all the lines in such a way that they point toward such a node v_0: then we can imagine a tree as obtained by contracting $n-1$ exiting lines with $n-1$ entering lines. If e_v and c_v denote the number of lines entering and, respectively, coming out from the node v, one has $e_{v_0}=1$ and $c_{v_0}=0$, while all the other nodes have at least one exiting line, so that $\forall v\ne v_0$ one has $c_v=1$ and $e_v\ge 0$; one has $d_v=e_v+c_v$. Call C the set of exiting lines and E the set of entering lines, and call E' the set obtained from E by neglecting the line entering the node v_0; one has $|E|=|E'|+1=\sum_v e_v=n-1$ and $|C|=\sum_v c_v=n-1$. Then a tree ϑ can be obtained as follows. Consider the subset C_1 of C formed by the lines coming out from the nodes v with $e_v=0$: choose a line in C_1 and contract it with any line in E', then choose another line in C_1 and contract it with another line in E', and so on until we exhaust all lines in C_1. Then consider the subset C_2 of C formed by the lines coming out from the nodes with $e_v\ge 1$ such that all the entering lines have been already contracted: choose a line in C_2 and contract it with any uncontracted line in E', and so on until all lines in C_2 have been contracted. Then define C_3 as the subset of C formed by the lines coming out from the nodes with $e_v\ge 1$ such that all the entering lines have been contracted along one of the previous steps, and so on: we iterate the construction until there will be left only one uncontracted line in C: such a line will necessarily have to be contracted with the line entering v_0. In this way we would have obtained $(n-2)!$ trees (as $n-2$ is the number of elements in E'), but we have still to divide such a number by $1/\prod_{v\neq v_0} e_v! = 1/\prod_v (d_v-1)!$ to avoid overcounting and obtain the right counting of distinct trees, as contraction of a line in C with any line in E entering the same node produce the same tree.)

Q7.1.5 **[7.1.5]:** (*Cayley's formula*)
Count the total number of rooted trees with n branches showing that it is given by n^{n-2}. (*Hint:* Use the multinomial formula

$$\sum_{\substack{n_1,\ldots,n_p\ge 0\\ n_1+\ldots+n_p=n}} \frac{n!}{n_1!\ldots n_p!} x_1^{n_1}\ldots x_{n_p}^{n_p} = (x_1+\ldots+x_p)^n\,,$$

and the result of the previous problem.)

Q7.1.6 **[7.1.6]:** (*Mayer coefficients*)
Let $\Gamma=\{\gamma_1,\ldots,\gamma_q\}$ be a polymer configuration. Regard the $\Gamma(\gamma)$ copies of each polymer contained in Γ as distinct by adding to each $\gamma\in\Gamma$ an extra label $1,2,\ldots,\Gamma(\gamma)$. Define $g_{ij}=-1$ if the polymer γ_i is incompatible with the polymer γ_j and $g_{ij}=0$ otherwise. Consider the function $\varphi(\Gamma)=\prod_{i<j}(1+g_{ij})$ and prove that $\varphi(\Gamma)\equiv\frac{\varphi(\Gamma)}{\Gamma!}=(\mathrm{Exp}\varphi^T)(\Gamma)$, where φ^T is defined by the r.h.s. of the first relation in (7.1.22), *i.e.* $\varphi^T=\mathrm{Log}\,\varphi$. (*Hint:* Note that by using (7.1.22) as a definition of φ^T

$$\varphi(\Gamma)=\prod_{i<j}(1+g_{ij})=\sum_k \sum_{\substack{C_1,\ldots,C_k \text{ unordered}\\ C_1+\ldots+C_k=\Gamma}} \prod_j (C_j!\,\varphi^T(C_j))=$$

$$=\sum_k\frac{1}{k!}\sum_{\substack{C_1,\ldots,C_k \text{ ordered}\\ C_1+\ldots+C_k=\Gamma}} \frac{\Gamma!}{\prod_j C_j!}\prod_j (C_j!\,\varphi^T(C_j))=\Gamma!\,(\mathrm{Exp}\,\varphi^T)(\Gamma),$$

if one collects together terms which are different because of the artificial distinction of the multiply repeated polymers.)

Bibliographical note for §7.1

The *algebraic* method followed in this section to obtain the considered estimates is due to Ruelle (see p. 86 in [Ru69]).

The notion of tree distance of the points of a set Γ has been introduced as a concept relevant for the theory of Gibbs distributions in the [DJS73].

The cluster expansion based on polymer expansions, introduced in [GK69], has been applied to several problems for instance the infinite volume limit (*infrared problem*) in the scalar field theories φ^4_d in dimension $d=2,3$, see [GJS73], [MS76], [OS73] and the remarkable lectures [Ec75]. Another example is the proof of the area law for the Wilson loop in a simple lattice gauge theory [GGM78]. One can also mention the microscopic theory of phase coexistence, see [MS67], [GMM73], [Ga72a] and developed in great detail later, see for instance [Mi95]. Problems on large deviations in the theory of Gibbs states can also be solved quite easily by the method, see [GMM78], [GLM02].

Here we have adopted the method of [GK78] following in detail the version used in [GMM73].

§7.2 Cluster expansions

A first answer to the question raised at the beginning of Section §7.1 is obtained in the case $d = 1$, *i.e.* of one-dimensional systems, in which T is a $(n+1) \times (n+1)$ compatibility matrix which is mixing and transitive on the vacuum (*i.e.* $T_{0\sigma} = T_{\sigma 0} = 1$ for all $\sigma = 0, \ldots, n$, see definition (6.4.1)) and $\Phi \in \widetilde{\mathcal{B}}$, where $\widetilde{\mathcal{B}}$ is the space of translation invariant particles potentials on $\{0, \ldots, n\}_T^{\mathbb{Z}}$ (cf. definition (6.4.1)). We recall that, in particular, this means that $\Phi_X(\sigma_X) \equiv 0$ if $\sigma_x = 0$ for some $x \in X$, thus allowing us to interpret $\sigma_x = 0$ as the vacuum at site x while $\sigma_x \neq 0$ will be interpreted as indicating that the site x is occupied by a particle of species σ_x.
More generally we shall consider Gibbs distributions on lattices of dimension $d > 1$, cf. definition (7.1.2).

Before proceeding it is convenient to adopt the notation $(h, \overline{\Phi})$ to denote a potential $\Phi \in \widetilde{\mathcal{B}}$ with

e7.2.1
$$\begin{aligned} & h(\sigma) = \Phi_\xi(\sigma) \text{ and } \overline{\Phi}_\xi(\sigma) = 0, \\ & \overline{\Phi}_X(\underline{\sigma}_X) \equiv \Phi_X(\underline{\sigma}_X) \text{ for all } |X| > 1, \end{aligned} \tag{7.2.1}$$

i.e. we wish to distinguish between the single-body component h of Φ and its many-body part $\overline{\Phi}$.
Furthermore, we shall denote by $\widetilde{\mathcal{B}}^{\text{complex}}$ the space of the translationally invariant particle potentials which are complex (*i.e.* such that the components of $\overline{\Phi}$ and h can assume complex values) with the natural norm

e7.2.2
$$\begin{aligned} & \|\Phi\| = \|h\| + \|\overline{\Phi}\|, \\ & \|h\| \equiv \max_\sigma |h_\xi(\sigma)|, \qquad \|\overline{\Phi}\| \equiv \sum_{X \ni \xi,\, |X| \geq 2} \max_{\underline{\sigma}_X} |\Phi_X(\underline{\sigma}_X)|, \end{aligned} \tag{7.2.2}$$

i.e. "$\widetilde{\mathcal{B}}$ is the real part of $\widetilde{\mathcal{B}}^{\text{complex}}$". Sometimes we shall write $\widetilde{\mathcal{B}}^{\text{complex}}(\mathbb{Z}^d)$ instead of $\widetilde{\mathcal{B}}^{\text{complex}}$ to stress that the potential Φ is an interaction potential for a d–dimensional spin system.
In the following discussion we shall often allow also single-body components $\Phi_\xi(\sigma) = h_\xi(\sigma)$ which are not translationally invariant (*i.e.* h_ξ may depend on ξ) and even not translationally invariant many spins potentials $\Phi_X(\underline{\sigma}) \neq \Phi_{X+\xi}(\underline{\sigma})$ for $\xi \in \mathbb{Z}^d$. In such cases $\|h\|$ and $\|\overline{\Phi}\|$ will be defined as in (7.2.2) with a $\sup_\xi$ added on the r.h.s.

The following proposition is remarkable and, historically, it has been the first to be proved among many similar ones. It is not really relevant for our intended applications to dynamical systems. We perform for completeness its analysis in Appendix (7.2).

P7.2.1 **(7.2.1) Proposition:** (Mayer expansion at small activity)
With the just introduced notations let $\widetilde{\mathcal{B}}^{\rm complex}(\mathbb{Z}^d)$ *be the space of the potentials* $\Phi = (h, \overline{\Phi})$ *for a* d*–dimensional* $\frac{n}{2}$*–spin system with a transitive*
N7.2.1 *vacuum (corresponding to* $\sigma = 0$*) and, possibly, a hard core interaction.*[1]
Define

e7.2.3
$$z \stackrel{def}{=} \sum_{\sigma>0} e^{-h(\sigma)}, \qquad \overline{z} \stackrel{def}{=} \sum_{\sigma>0} e^{-{\rm Re}h(\sigma)}, \tag{7.2.3}$$

where z *is such that* $|z| \le \overline{z}$ *and will be called* activity*. Given* $c > 0$ *let*

e7.2.4
$$\Sigma \stackrel{def}{=} \Big\{\Phi|\Phi = (h, \overline{\Phi}) \in \widetilde{\mathcal{B}}^{\rm complex},\ \overline{z}\, e^{\|\,\overline{\Phi}\|}(1 + (e^{\|\,\overline{\Phi}\|} - 1)\, c) < 1\Big\}. \tag{7.2.4}$$

(i) There is a constant c*, depending only on the matrix* T *describing the hard cores, such that the function* $\Phi \to P(\Phi)$*, defined for* $\Phi \in \widetilde{\mathcal{B}}$*, can be*
N7.2.2 *analytically extended from* $\Sigma \cap \widetilde{\mathcal{B}}$ *to* Σ*.*[2]
(ii) In absence of hard cores and for $\Phi \in \widetilde{\mathcal{B}}$ *real analyticity of* $\Phi \to P(\Phi)$ *holds under the condition*

e7.2.5
$$z\, e^{\|\,\overline{\Phi}\|}\left(\frac{1}{1 + z e^{\|\,\overline{\Phi}\|}} + c\, e^{\|\,\overline{\Phi}\|}(e^{\|\,\overline{\Phi}\|} - 1)\right) < 1, \tag{7.2.5}$$

where one can take $c = 2n$ *if the spin takes* $n+1$ *values. Note that the l.h.s. of (7.2.5) tend to* $z/(1+z) < 1$ *as* $\|\,\overline{\Phi}\| \to 0$*.*

Remarks: (1) On $\Sigma \cap \widetilde{\mathcal{B}}$ we shall have that the function $\Phi \to P(\Phi)$ is analytic in the sense of definition (7.1.1).
(2) The proof goes back to Groeneveld, Penrose and Ruelle who treated the *pair potential case*, *i.e.* the case $\Phi_X = 0$ if $|X| > 2$. This is a classical result among the deepest in the theory of Gibbs distributions: it bears many other consequences that we shall not need and which, therefore, we do not comment.
(3) The method of proof is generalizable or adaptable to a wide variety of cases and it is the foundation of most of the really constructive results that are known in the theory of Gibbs distributions, particularly in the case of lattice systems in more than one dimension. For this reason we have stated the above theorem *without restricting ourselves to the case of one dimensional spin systems.* However condition (7.2.4) is an extremely restrictive condition which physically means that a sequence $\underline{\sigma}$ which is typical for the Gibbs distribution will mostly consist of the symbol 0, *i.e.* it will be close to the vacuum. Therefore we shall also be interested in weakening the restriction, when possible.

---

[1] Cf. definition (7.1.1).

[2] This means that there exists a function $\Phi \to P(\Phi)$ that coincides on the set of real valued potentials $\Sigma \cap \widetilde{\mathcal{B}}$ with $P(\Phi) = \lim_{\Lambda\to\infty} |\Lambda|^{-1} \log Z_\Lambda(\Phi)$ and which is analytic in every point Φ of Σ, *i.e.* the function $(\lambda_1, \ldots, \lambda_d) \to P(\Phi + \lambda_1\Psi_1 + \ldots + \lambda_d\Psi_d)$ is analytic near the origin, for all $\Psi_1, \ldots, \Psi_d \in \widetilde{\mathcal{B}}^{\rm complex}$. See definition (7.1.1).

(4) If $\overline{\Phi}$ becomes small and if there are no hard cores in the interactions one expects that the system, *i.e.* the distribution μ_Φ, approaches the "free" one, in which the variables σ_x are independently distributed and $e^{-h(\sigma)}/(1+z)$ is the probability that σ_x has value σ. Whether this is true or not may depend on what we mean by "small" $\overline{\Phi}$. The last statement in the proposition shows that this is the case if the size of $\overline{\Phi}$ is measured by $\|\overline{\Phi}\|$ (cf. (7.2.2)): for this reason the definition of $\|\overline{\Phi}\|$ is natural and it repeatedly arises when one tries to control the properties of the distribution μ_Φ. Below we develop the theory in a different direction obtaining results of this type *but under more stringent conditions on the size of* $\overline{\Phi}$, *i.e.* we study cases which can be immediately reduced to the general analysis in Section §7.1. For completeness the main elements of the classical proof of (7.2.5) are in Appendix (7.2) below.

A proposition similar to proposition (7.2.1) but somewhat weaker is an immediate consequence of the polymer theory analysis of Section §7.1. Recalling that $\delta(X)$ denotes the tree length of the set X, *i.e.* the minimal value of the sum of the lengths of the branches of a tree that spans the set X (cf. definition (7.1.3)), it can be formulated as follows.

P7.2.2 **(7.2.2) Proposition:** (Cluster expansion in systems without hard cores and small activity)
*Let $\widetilde{B}^{\rm complex}$ be the space of the particle potentials $\Phi=(h,\overline{\Phi})$ for a lattice spin system with spin $\sigma=0,\ldots,n$ and without hard cores (*cf. *definition (6.4.1)). Given $\kappa>0$ define*

e7.2.6
$$\|\overline{\Phi}\|_\kappa \stackrel{def}{=} \sum_{X\ni 0}\max_{\underline{\sigma}_X}|\Phi_X(\underline{\sigma}_X)|\,e^{\kappa\delta(X)}, \tag{7.2.6}$$

and $\overline{z}=\sum_{\sigma>0}e^{-{\rm Re}\,h(\sigma)}$.
*If $\overline{z}$ is small enough, i.e. $\overline{z}\,e^{2(\|\overline{\Phi}\|_\kappa+b)+1}<\varepsilon$ for a suitably small ε the Gibbs distribution μ_Φ is analytic in Φ (*cf. *definition (7.1.1)).*

Proof: Consider, for $\Lambda=\{1,\ldots,N\}$ and $\Phi\in\widetilde{\mathcal{B}}^{\rm complex}$,

e7.2.7
$$\begin{aligned} Z_\Phi(\Lambda)=&\sum_{\underline{\sigma}\in\{0,1,\ldots,n\}^\Lambda} e^{-\sum_{Y\subset\Lambda}\Phi_Y(\underline{\sigma}_Y)}=\\ =&\sum_{E\subset\Lambda}\sum_{\underline{\sigma}_E\in\{1,\ldots,n\}^E}\Big(\prod_{\xi\in E}e^{-h(\sigma_\xi)}\Big)\Big(\prod_{Y\subset E}e^{-\overline{\Phi}_Y(\underline{\sigma}_Y)}\Big), \end{aligned}\tag{7.2.7}$$

because $\Phi\in\widetilde{\mathcal{B}}^{\rm complex}$ implies (by definition) $\Phi_E(\underline{\sigma}_E)=0$ if $\sigma_j=0$ for some $j\in E$. We must study the logarithm of $Z_\Phi(\Lambda)$ and precisely the limit $P(\Phi)=\lim_{\Lambda\to\infty}|\Lambda|^{-1}\log Z_\Phi(\Lambda)$: for this purpose we transform the sum (7.2.7) into an exponential form. This can be done by adding and subtracting 1 in each of the factors of the last product in (7.2.7) turning each factor into a binomial: then developing the binomial and collecting the "connected clusters". The union X of the clusters is contained in E and if we keep it fixed the same X may arise from any of the choices of E which contain X: hence each X is counted $2^{|\Lambda|-|X|}$ which is the source of various factors 2 below.

The partition function becomes $Z_\Phi(\Lambda) = 2^{|\Lambda|}\sum_\Gamma \varphi(\Gamma)\zeta(\Gamma)$, where $\Gamma = \{\gamma_1,\ldots,\gamma_p\}$ is a polymer configuration, $\varphi(\Gamma) = 1$ if the polymers in Γ do not overlap (so that in particular $\Gamma! = 1$), $\varphi(\Gamma) = 0$ otherwise, and $\zeta(\Gamma) = \prod_{\gamma\in\Gamma}\zeta(\gamma)$ with $\zeta(\emptyset) = 1$ and

$$\zeta(\gamma) = \frac{z}{2} \qquad \text{if } \gamma \text{ is a single point; otherwise:}$$

e7.2.8

$$\zeta(\gamma) = \langle \sum^{*}_{\substack{q,\,Y_1,\ldots,Y_q\\ \cup_j Y_j=\gamma,\,|Y_j|>1}} \prod_{\xi\in\gamma} e^{i\,\mathrm{Im}\,h(\sigma_\xi)} \prod_{j=1}^{q}(e^{-\overline{\Phi}_{Y_j}(\sigma_{Y_j})}-1)\rangle \,\big(\frac{\overline{z}}{2}\big)^{|\gamma|} \tag{7.2.8}$$

where $z = \sum_{\sigma>0} e^{-h(\sigma)}$, $\langle\bullet\rangle$ is the average $\overline{z}^{-|\gamma|}\sum_{\underline{\sigma}_\gamma>0}\prod_{\xi\in\gamma} e^{-\mathrm{Re}\,h(\sigma_\xi)}\bullet$, the $*$ over the sum recalls that the sets Y_i must contain chains connecting any two points of γ and $\underline{\sigma}_\gamma > 0$ means $\sigma_\xi > 0$ for all $\xi\in\gamma$.

N7.2.3 One obtains an expression for the partition function in terms of polymers whose activity can be immediately estimated, cf. (7.2.6), by [3]

e7.2.9

$$|\zeta(\gamma)| \le \big(\frac{\overline{z}}{2}\,e^{2\|\,\overline{\Phi}\|_\kappa+1}\big)^{|\gamma|}(1+\|\,\overline{\Phi}\|_\kappa)e^{-\kappa\delta(\gamma)}. \tag{7.2.9}$$

which implies (7.1.4) with $b_0^2 = 1+\|\,\overline{\Phi}\|_\kappa$, $\nu_0^2 = \frac{\overline{z}}{2}e^{2\|\,\overline{\Phi}\|+1}$ and $\kappa_0 = \kappa/2$: so that by taking z small enough we fall under the conditions of applicability

N7.2.4 of proposition (7.1.1).[4]

The probability $\mu_\Phi(C^V_{\underline{\sigma}^0_V})$ of the cylinder $C^V_{\underline{\sigma}^0_V}$ is the limit as $\Lambda\to\infty$ of

e7.2.10

$$\frac{\sum_{\underline{\sigma}_\Lambda} e^{-U_\Phi(\underline{\sigma}_\Lambda)}\chi_{V,\underline{\sigma}^0_V}(\underline{\sigma}_\Lambda)}{\sum_{\underline{\sigma}_\Lambda} e^{-U_\Phi(\underline{\sigma}_\Lambda)}}, \tag{7.2.10}$$

where $\chi_{V,\underline{\sigma}^0_V}(\underline{\sigma}_\Lambda) = \prod_{\xi\in\,V}\chi_{\xi,\sigma^0_\xi}(\sigma_\xi)$ is the characteristic function of $C^V_{\underline{\sigma}^0_V}$. The ratio can be written (after (7.1.18)) as

e7.2.11

$$\exp\Big(\sum_{\Gamma=\{\gamma_1,\ldots,\gamma_q\},\gamma_i\in\Lambda}\varphi^T(\Gamma)(\zeta_V(\Gamma)-\zeta(\Gamma))\Big), \tag{7.2.11}$$

---

[3] From $|e^x-1|\le e^{|x|}|x|$ we bound $\prod_{j=1}^q(e^{-\overline{\Phi}_{Y_j}(\underline{\sigma}_{Y_j})}-1)$ by $e^{|\gamma|\|\,\overline{\Phi}\|_0}\prod_j(|\,\overline{\Phi}_{Y_j}|e^{\kappa\delta(Y_j)})\cdot e^{-\kappa\delta(\gamma)}$ (because $\sum_j\delta(Y_j)\ge\delta(\gamma)$): we use that $\sum_{q\ge n}x^q/q!\le e^x x^n/n!$ and that the sum is over $q\ge 1$ which implies (7.2.9) after bounding $e^{\|\,\overline{\Phi}\|_\kappa|\gamma|}\|\,\overline{\Phi}\|_\kappa|\gamma|\le e^{(\|\,\overline{\Phi}\|_\kappa+1)|\gamma|}\|\,\overline{\Phi}\|_\kappa$ (the addend 1 bounds the contribution from the one point polymer).

[4] Note that (7.2.9) can be interpreted to hold even if $z=0$ because h is here real valued.

where $\zeta_V(\Gamma)=\prod_{\gamma\in\Gamma}\zeta_V(\gamma)$ with $\zeta_V(\gamma)$ defined in a way similar to (7.2.8) with the insertion inside the average of an extra factor $\prod_{\xi\in\gamma\cup V}\chi_{\xi,\sigma^0_\xi}(\sigma_\xi)$, when γ is not a single point; if $\gamma=\xi$ is a single point $\zeta_V(\gamma)=\langle\chi_{\xi,\sigma^0_\xi}(\sigma_\xi)\rangle\, z$ if $\xi\in V$ and $\zeta_V(\gamma)=z$ otherwise. This implies

e7.2.12 $$\zeta_V(\Gamma)\equiv\zeta(\Gamma)\quad\text{if } V\cap\widetilde{\Gamma}=\emptyset,\qquad\text{and}\quad|\zeta_V(\Gamma)|\le|\zeta(\Gamma)|.\tag{7.2.12}$$

Therefore we conclude that the limit as $\Lambda\to\infty$ of (7.2.11) exists and is simply given by

e7.2.13 $$\mu_\Phi(C^V_{\sigma^0_V})=\exp\Big(\sum_\Gamma\varphi^T(\Gamma)(\zeta_V(\Gamma)-\zeta(\Gamma))\Big),\tag{7.2.13}$$

which is convergent because of (7.2.12), (7.2.9), (7.1.4) and (7.1.19).

The translation invariance of $\zeta(\Gamma)$ implies, see (7.1.36), the existence of the "thermodynamic limit" $P(\Phi)=\lim_{\Lambda\to\infty}|\Lambda|^{-1}\log Z_\Phi(\Lambda)$ and its identity with the function $P(\Phi)=\sum_{\Gamma\ni 0}\frac{\varphi^T(\Gamma)}{n(\Gamma)}\zeta(\Gamma)$, where $n(\Gamma)$ is the number of lattice points contained in Γ, cf. (7.1.36). Analyticity in Φ follows from corollary (7.1.1) to proposition (7.1.1). ■

Remarks: (1) The above proof leads to a result weaker than (ii) of proposition (7.2.1) since it requires absence of hard cores, *i.e.* $T_{\sigma,\{\sigma'\}}\equiv 1$.
(2) In one dimension, however, the assumption of absence of hard core can be eliminated as we shall discuss in Section §(7.3).
(3) Note that *no condition* had to be imposed on the imaginary part of $h_\xi(\sigma)$.

We now turn to the question of regularity of $P(\Phi)$ without particular assumptions on the size of the activity z. In absence of hard cores the special role of $\sigma=0$ can be eliminated by means of another remarkable immediate corollary of the polymer cluster expansion which deals with a general lattice potential, cf. definition (5.1.1). The assumption that a spin value (*i.e.* $\sigma=0$) is "privileged" (*i.e.* $\Phi_\xi(\sigma),\sigma\ne 0$, is large so that the activity z of (7.2.3) is small) is replaced by the assumption that $\overline{\Phi}$ is small. To achieve this we can employ an expansion in which the trivial polymers, *i.e.* polymers γ with $|\gamma|=1$, are absent. The following proposition shows the possibility, at least for real valued potentials, of an expansion without trivial polymers, *i.e.* an expansion in terms of polymers γ with $|\gamma|\ge 2$ whose activity depends necessarily (unlike the trivial polymers cases) on the non-local part $\overline{\Phi}$ of the potential. And indeed corollaries will show how the results of the proposition can be implemented in a useful way for this purpose.

P7.2.3 **(7.2.3) Proposition:** (Cluster expansion at any activity)
Consider a $\frac{n}{2}$-spin system on a d–dimensional lattice $\mathbb{Z}^d$ and without hard core interactions. Let Φ be a complex potential not necessarily translation invariant and let $\|\Phi_X\|=\max_{\sigma_X}|\Phi_X(\underline{\sigma}_X)|$ and set, for $\kappa>0$,

e7.2.14 $$\|\overline{\Phi}\|_\kappa\stackrel{def}{=}\sup_\xi\sum_{X\ni\xi,|X|>1}\|\Phi_X\|\,e^{\kappa\delta(X)},\qquad b\stackrel{def}{=}\sup_\xi|\mathrm{Im}\,\Phi_\xi|,\tag{7.2.14}$$

where $\delta(X)$ denotes the tree length of X.
(i) Whether Φ is translation invariant or not, the partition function $Z_\Phi(\Lambda)$ defined as $\sum_{\underline{\sigma}_\Lambda} e^{-U_\Phi(\underline{\sigma}_\Lambda)}$ in (7.2.18) can be expressed formally as

e7.2.15
$$Z_\Phi(\Lambda) = \Big(\prod_{\xi\in\Lambda} \Xi_1(\xi)\Big)\, \exp\Big(\sum_{X\subset\Lambda} \varphi^T(X)\zeta(X)\Big), \tag{7.2.15}$$

where $\Xi_1(\xi) = \sum_\sigma e^{-\mathrm{Re}\,\Phi_\xi(\sigma)}$, $X = \{\gamma_1, \ldots, \gamma_n\}$ is a polymer configuration, $\varphi^T(X)$ are the combinatorial functions of X defined in (7.1.14), $\zeta(X) = \prod_{\gamma\in X} \zeta(\gamma)$ for suitably defined $\zeta(\gamma)$, which verify

e7.2.16
$$|\zeta(\gamma)| \le (e^{2\|\overline{\Phi}\|_\kappa+2b+1})^{|\gamma|}(\|\overline{\Phi}\|_\kappa + b)^{\nu(\gamma)} e^{-\kappa\delta(\gamma)}, \tag{7.2.16}$$

where $\nu(\gamma)$ is the minimum number of sets Y with $\Phi_Y \not\equiv 0$ and $|Y| \ge 2$ or with $|Y| = 1$, $Y = \xi$ and $|\mathrm{Im}\,\Phi_\xi| \ne 0$, necessary to cover γ. Note that $|\gamma| = 1$ implies $\nu(\gamma) = 1$, so that in this case $\nu(\gamma) \equiv |\gamma|$.
(ii) (7.2.15) becomes an identity if the series in the exponent is absolutely convergent.
(iii) If $\nu(\gamma) > c|\gamma|$ for $|\gamma| > 1$ and a suitable $0 < c \le 1$ there exist $B, \varepsilon > 0$, independent of the size n of the spins, such that if $\|\overline{\Phi}\|_\kappa + b < \varepsilon$

e7.2.17
$$\sup_{\xi\in\mathbb{Z}^d} \sum_{\widetilde{X}\ni\xi,\delta(X)\ge r} |\varphi^T(X)||\zeta(X)| < B\,\varepsilon\, e^{-\kappa r}. \tag{7.2.17}$$

(iv) Furthermore $\zeta(\gamma)$ are analytic as functions of Φ if Φ varies in the region $\|\overline{\Phi}\|_\kappa + b < \varepsilon$. In the translation invariant case the Gibbs distribution μ_Φ is analytic in Φ.

Remarks: (1) For later developments we stress that the result (ii) holds for potentials which are not necessarily translation invariant (but which must admit the translation invariant bound (7.2.14)) and which are not necessarily real. Furthermore the values of ε, B do not depend on the "spin size" n: the latter remarkable property will be used extensively in the following. The novelty in the expansion is that there is no special spin value corresponding to the vacuum and therefore no condition of the type $\overline{z}$ small is imposed. On the other hand the assumptions demand that all spin interactions should be small *with the possible exception of the real part of the single spin interaction $\Phi_\xi(\sigma)$.*
(2) The property $\nu(\gamma) > c|\gamma|$ for $|\gamma| > 1$ can fail if there are sets X with $\Phi_X \not\equiv 0$ so large to cover large sets γ, *i.e.* requiring the existence of c is in some sense a condition on the range of the potential. However in several cases of interest the parameter κ in (7.2.14) can be adjusted: in such cases the strong condition $\nu(\gamma) > c|\gamma|$ can be replaced by κ large enough and b small enough, see also the comments after (7.1.5). This will be exploited in one of the corollaries that follows.
(3) It is important to check, while looking at the coming proof, that the same theorem holds if the spins that are located at each site ξ can take a number

of values n_ξ which depends on ξ and is bounded by some $N>0$: this can be looked at as a *on site* hard core. This will be a necessary property in Section §(7.3): in fact it is the key to understand hard core interactions in one dimension.

Proof: We begin by restricting the sets that we shall call polymers (*i.e.* the polymers which have a non-zero activity in the coming polymer expansion associated with the potential considered): a polymer (with non-zero activity) will be a finite set γ with the following property.

("Connectivity") *For all pairs $\xi,\eta\in\gamma$ there is a connecting chain $Y_1,\ldots,Y_m$ of subsets of γ such that $\xi\in Y_1,\eta\in Y_m$ and $Y_i\cap Y_{i+1}\neq\emptyset$ and, furthermore, for each Y_i there is at least one configuration $\underline{\sigma}_{Y_i}$ with $\Phi_{Y_i}(\underline{\sigma}_{Y_i})\neq 0$.*

The polymer representation of partition functions stems from the identity

e7.2.18

$$\begin{aligned}Z_\Phi(\Lambda)=&\sum_{\underline{\sigma}\in\{0,1,\ldots,n\}^\Lambda} e^{-\sum_{Y\subset\Lambda}\Phi_Y(\underline{\sigma}_Y)}=\\ =&\sum_{\underline{\sigma}\in\{0,1,\ldots,n\}^\Lambda}\prod_{\xi\in\Lambda} e^{-\mathrm{Re}\,\Phi_\xi(\sigma_\xi)}\prod_{Y\subset\Lambda}\Big((e^{-\Phi_Y(\underline{\sigma}_Y)}-1)+1\Big),\end{aligned}\tag{7.2.18}$$

where in the second product the factors with $|Y|=1$ must be interpreted as $(e^{-i\mathrm{Im}\,\Phi_\xi(\sigma_\xi)}-1)$, *i.e.* we treat differently the real part and the imaginary parts of the single spin potential.

Developing the product as a binomial product we can collect the sets Y into groups forming polymers; thus representing the sum as a sum over collections of *non-overlapping polymers* $\gamma_1,\ldots,\gamma_p$:

e7.2.19

$$\begin{aligned}Z_\Phi(\Lambda)=&\Xi_1(\Lambda)\sum^{@}_{\gamma_1,\ldots,\gamma_p\subset\Lambda}\prod_{i=1}^p\Big(\Xi_1(\gamma_i)^{-1}\sum_{\underline{\sigma}_{\gamma_i}}\prod_{\xi\in\gamma_i}e^{-\mathrm{Re}\,\Phi_\xi(\sigma_\xi)}\cdot\\ &\cdot\Big(\sum_{q_i}\sum^{*}_{\substack{Y^{(i)}_1,\ldots,Y^{(i)}_{q_i}\subset\gamma_i\\ \bigcup_j Y^{(i)}_j\equiv\gamma_i}}\prod_{j=1}^{q_i}\big(e^{-\Phi_{Y^{(i)}_j}(\underline{\sigma}_{Y^{(i)}_j})}-1\big)\Big)\Big),\end{aligned}\tag{7.2.19}$$

where $\Xi_1(V)\stackrel{def}{=}\prod_{\xi\in V}(\sum_\sigma e^{-\mathrm{Re}\,\Phi_\xi(\sigma)})\equiv\prod_{\xi\in V}\Xi_1(\xi)$ for $V\subseteq\Lambda$; the symbol @ over the first sum recalls that the polymers γ_i are nonoverlapping and the $*$ over the fourth sum recalls the further restriction that the sets $Y^{(i)}_j$ must also contain chains connecting any two points of γ_i. Also in this case the $Y^{(i)}_j$ consisting of a single point ξ have to be interpreted as $(e^{-i\mathrm{Im}\,\Phi_\xi(\sigma_\xi)}-1)$.

Introduce the *activities* of a polymer γ and of a polymers configuration X

as

e7.2.20
$$\zeta(\gamma)=\langle\sum_q \sum^*_{\substack{Y_1,\ldots,Y_q\subset\gamma\\ \bigcup_j Y_j=\gamma}} \prod_{j=1}^q \big(e^{-\Phi_{Y_j}(\underline{\sigma}_{Y_j})}-1\big)\rangle,$$
$$\zeta(X)=\prod_{\gamma\in\tilde X}\zeta(\gamma)^{X(\gamma)},\qquad \zeta(\emptyset)\equiv 1, \tag{7.2.20}$$

where the $\langle\bullet\rangle$ denotes the average $\Xi_1(\gamma)^{-1}\sum_{\underline{\sigma}_\gamma}\prod_{\xi\in\gamma}e^{-\mathrm{Re}\,\Phi_\xi(\sigma_\xi)}\bullet$.

The key remark is that the function $\zeta(\gamma)$ verifies (7.1.4), from (7.2.20), by repeating the argument used to derive (7.2.9) (see footnote 3), noting that $\sum_{q\ge n}(xy)^q/q!\le e^y x^n$ if $x<1$: in our case $x=\|\overline{\Phi}\|_\kappa+b$ and $y=|\gamma|$ if $|\gamma|>1$ while if $|\gamma|=1$ we take, instead, $x=b$ (because $|e^{ix}-1|<|x|$ for real x).

Therefore, if $\nu(\gamma)>c|\gamma|$, setting $b_0=1,\nu_0^2=e^{2(\|\overline{\Phi}\|_\kappa+b)+1}(\|\overline{\Phi}\|_\kappa+b)^c,\kappa_0=\kappa/2$ the result is an immediate corollary of proposition (7.1.1) and of its corollary (7.1.1). ■

Remark: It is interesting to note that the reason why we write an expansion treating in a different way the real and imaginary parts of the single spin potential is that we want to allow for an arbitrary real part of the single spin potential. If we also require, rather than (7.2.15), that the single spin potential be small, *i.e.* that

e7.2.21
$$\|\Phi\|_\kappa=\sup_{\xi,\sigma}|h_\xi(\sigma)|+\|\overline{\Phi}\|_\kappa \tag{7.2.21}$$

be small, then we can treat the real and imaginary parts of Φ_ξ in the same way and, by the above proof, we obtain the same results of the proposition under a smallness condition on $\|\Phi\|_\kappa$. In the latter case the average in (7.2.20) becomes the average with weight 1 over the spin values.

The proof of the above proposition leads easily to the result that we would expect for $\overline{\Phi}$ small (see remark (4) to proposition (7.2.1)) in the cases in which the interaction $\overline{\Phi}$ only involves "finitely many bodies" or has "finite range": notions that we introduce more formally as follows.

D7.2.1 **(7.2.1) Definition:** (Finite range or finitely many bodies potentials) *A potential Φ is said an N–*body *or N–*spin interaction *if $\Phi_X\equiv 0$ if $|X|>N$. If $N=2$ it is called a* pair interaction*. A potential is said a R-*range potential *if $\Phi_Y\equiv 0$ for* $\mathrm{diam}(Y)>R$*. If, with the notation (7.2.21), $\|\overline{\Phi}\|_\kappa<\infty$ for some $\kappa>0$ we say that Φ is a* short range potential.

The following result is in many respects weaker compared to the second statement in proposition (7.2.1) (which we have not yet proved, see Appendix (7.2)), because it requires the potential to be a finitely many-spin one, it is nevertheless a fairly immediate consequence of the techniques that we are exploring and it is worth discussing as it is relevant for the theory of dynamical systems.

C7.2.1 **(7.2.1) Corollary:** (Any activity, finite range and small interactions) *Let Φ be a real R–range potential for a lattice spin system with spin $\sigma = 0,\ldots,n$ and without hard cores. Given $\kappa > 0$ if $\|\overline{\Phi}\|_\kappa + b < \varepsilon$, see (7.2.14), for a suitably small ε the Gibbs distribution μ_Φ is real analytic in Φ (*cf.* definition (7.1.1)).*

Proof: One simply remarks that if $|\overline{\Phi}_Y| \equiv 0$ for $\mathrm{diam}(Y) > R$ then a polymer that can be covered by q sets Y_i, $i = 1,\ldots,q$, with $\Phi_{Y_i} \not\equiv 0$ is such that $q \ge \delta(\gamma)/R$ and $\delta(\gamma) \ge |\gamma| - 1 \ge \frac{1}{2}|\gamma|$. We see that the assumption in the third point of proposition (7.2.3) holds with $c = 1/(2R)$. ■

In the same way one can prove the following similar statement.

C7.2.2 **(7.2.2) Corollary:** (Any activity, finitely many-body small interactions) *Let Φ be a real N–bodies potential for a lattice spin system with spin $\sigma = 0,\ldots,n$ and without hard cores. If $\|\overline{\Phi}\|_\kappa + b < \varepsilon$, see (7.2.14), for a suitably small ε, the Gibbs distribution μ_Φ is real analytic in Φ (*cf.* definition (7.1.1)).*

Proof: One simply remarks that if $|\overline{\Phi}_Y| \equiv 0$ for $|Y| > N$ then a polymer which can be covered by q sets Y_i, $i = 1,\ldots,q$, with $\Phi_{Y_i} \not\equiv 0$ is such that $q \ge |\gamma|/N$ and one can take $c = \frac{1}{N}$ in proposition (7.2.3). ■

Remark: We see that in the case of either finite range or finitely many bodies interactions $h, \overline{\Phi}$ we obtain regularity (in fact analyticity) in h and $\overline{\Phi}$ under a condition of smallness of $||\overline{\Phi}||_\kappa$, for some $\kappa > 0$, and of $b = \sup_{\xi,\sigma} |\mathrm{Im}\, h_\xi(\sigma)|$, no matter how large is the real part of $|h|$. In Physics applications potentials in Gibbs distributions appear in multiplied by $\beta = \frac{1}{k_B T}$ where T is the temperature and k_B is a constant (*Boltzmann's constant*). Therefore the smallness condition on $\beta\overline{\Phi}$ is always achieved at "high temperature", *i.e.* at small β, and the same holds for the smallness condition on the imaginary part of βh as long as we confine the values of h to a fixed *a priori* given strip along the real axis, of course provided the essential condition of absence of hard core interactions holds. High temperature regularity in the cases of systems without hard cores and interacting via particle potentials with only one species of particles besides the vacuum (*i.e.* $\sigma = 0, 1$) can also be derived from part (ii) of proposition (7.2.1): the interested reader will find the classical proof of proposition (7.2.1) in the Appendix (7.2) below and the high temperature analyticity, for coreless potentials that decay sufficiently rapidly at ∞, is derived as a consequence of it in problems [7.2.7] and [7.2.9], [Do68b], [GMR68]: hence the difference between the first part of proposition (7.2.1) and its second part or the above corollaries is substantial.

It will be important to note the following corollary of the proof of the last proposition which allows us to reach the same conclusions under somewhat different assumptions which turn out to be also interesting for the theory of dynamical systems.

C7.2.3 **(7.2.3) Corollary:** (Short range case)
Consider a one-dimensional spin system without hard core interactions. Let Φ be a real potential with $\Phi_X(\underline{\sigma}_X)$ different from 0 only if X is an interval. With the notations (7.2.14), there is $0<\varepsilon<1$, independent of the spin size n, such that

e7.2.22
$$b+e^{2\|\overline{\Phi}\|_\kappa}e^{-\frac{1}{4}\kappa}\,\kappa^{-1}\le\varepsilon \tag{7.2.22}$$

implies, if ε is small enough, the conclusions of points (iii) and (iv) of proposition (7.2.3) with (7.2.16) replaced by

e7.2.23
$$|\zeta(\gamma)|\le e^{-L_0|\gamma|}e^{-\frac{1}{2}\kappa\delta(\gamma)}. \tag{7.2.23}$$

with $e^{-L_0}\equiv\varepsilon\,c$ for a suitable constant c.

Remarks: (1) The peculiarity with respect to the proposition (7.2.3) is that the dimension is $d=1$ and that here we do not require existence of the constant $c>0$ such that $\nu(\gamma)\ge c\,|\gamma|$, hence we allow potentials with exponentially decreasing strength. It is also important to remark that (as in the above cases) there is no condition on the number n of values that the spin σ can take. The proof below works because, by definition of tree distance, it is always true that $\delta(\gamma)=|\gamma|-1$ and therefore nontrivial polymers are such that $\delta(\gamma)\ge\frac{1}{2}|\gamma|$ while the trivial ones (with $|\gamma|=1$) have small activity which is bounded proportionally to b.
(2) The corollary (7.2.3) is interesting because the potentials that we have shown to arise in the theory of Anosov systems will satisfy its assumptions (see Section §(7.3)).

Proof: If one looks at the proof of proposition (7.2.3) one realizes that in this case the polymers for which $\delta(\gamma)\ge1$, hence $\delta(\gamma)\ge\frac{1}{2}|\gamma|$, satisfy the estimate

e7.2.24
$$|\zeta(\gamma)|\le(e^{2(b+\|\overline{\Phi}\|_\kappa)+1}e^{-\frac{1}{2}\kappa})^{|\gamma|}\,e^{-\frac{1}{2}\kappa\delta(\gamma)} \tag{7.2.24}$$

(see remark (2) following proposition (7.2.3)). For $|\gamma|=1$ (hence $\delta(\gamma)=0$) one has $|\zeta(\gamma)|\le b$. Setting $b_0^2=1$, $\nu_0^2=e^{2(\|\overline{\Phi}\|_\kappa+b)+1-\frac{1}{2}\kappa}+b\le e^{-L_0}$, $\kappa_0=\frac{1}{4}\kappa$ the result again follows from proposition (7.1.1) under the condition

e7.2.25
$$\big(3e^{B_0\nu_0}+k(1,\frac{\kappa_0}{2})\big)\nu_0<1, \tag{7.2.25}$$

with $B_0=2$ and $k(1,\kappa)=2e\sum_{\xi\in\mathbb{Z}^1,|\xi|\ge1}e^{-\kappa|\xi|}\le4e^{-\kappa}(1+\kappa^{-1})<8\frac{e^{-\frac{\kappa}{2}}}{\kappa}$ (see (7.1.5)), *i.e.* if $\|\overline{\Phi}\|_\kappa$, κ and b verify

e7.2.26
$$e^{-L_0}\stackrel{def}{=}8^2\left[e^{2(\|\overline{\Phi}\|_\kappa+b)+1}\frac{e^{-\frac{1}{4}\kappa}}{\kappa}+b\right]\le c\,(b+e^{2\|\overline{\Phi}\|_\kappa}e^{-\frac{1}{4}\kappa}\,\kappa^{-1}) \tag{7.2.26}$$

with L_0 large enough, which is implied by (7.2.22). ■

We conclude the section with a few general remarks.

Remarks: (1) *We stress again that the results of propositions (7.2.1), (7.2.2) and (7.2.3) hold even when the spins are imagined located on a d–dimensional lattice* $\mathbb{Z}^d$ *with* $d \geq 1$. This is interesting not only for the study of spatio–temporal chaos (in Chapter X below), but also because the same problems arise in Statistical Mechanics where the lattice points have the interpretation of "space" points rather than of time variables.

(2) The assumptions in propositions (7.2.1), (7.2.2) and (7.2.3) are too strong for one dimensional systems and for applications to dynamical systems. In fact transitivity of the vacuum (cf. proposition (7.2.1)) and the smallness conditions in proposition (7.2.3) are typically not satisfied by the potentials Φ that arise (for instance) in the description of the SRB distributions for Anosov systems. Furthermore the results of proposition (7.2.3) are too strong to hold without the latter assumptions in dimension higher than 1 where, in general, the structure of the "multidimensional Gibbs states" is much deeper and interesting.

(3) There is a remarkable case, however, in which a simple extension of proposition (7.2.3) valid for arbitrary lattice dimension is possible even allowing for hard core interaction and without excessive smallness conditions on the potential. Namely if the potential $\Phi_X(\underline{\sigma}_X)$ is a potential of arbitrary strength when X lies on a single line parallel to a given "time" direction, *e.g.* the $(d+1)$-th coordinate axis, but it decreases exponentially in all directions and is sufficiently small for all X which are not on a single line parallel to the time direction, see problems and Section §(7.3). The one-dimensional cases as well as the latter case arise in the theory of Anosov systems as we have seen so far ($d = 1$) while discussing the theory of Markov pavements and symbolic dynamics and as we shall see in Chapter X in more involved and interesting cases.

We shall devote the next section to the weakening of the vacuum transitivity, *i.e.* to eliminate assumptions about hard cores in one dimension and to the study of a multidimensional hard core cases of the kind considered in remark (3) above). The simple idea underlying the possibility of the mentioned extensions can be mastered by studying the problems of the present section which treat the particular cases from which the general results will follow (as independently discussed in Section §(7.3)) with a few technical additions.

Appendix 7.2: *The classical expansion*

Here we sketch the classical cluster expansion also known as the theory of the Kirkwood-Salsburg equations (which are the relations (A7.2.3) below in a somewhat improved form [Do68b], [GM68]). We describe it quickly and without too many details although we have paid attention to not forgetting any of the key steps.

The partial inversion of the equations that is discussed below ((A7.2.5)) is the main tool that allows us to obtain the second statement in proposition (7.2.1): the idea arose independently in [Do68b] and [GMR68].

We consider a compatibility matrix T which is mixing and transitive on the vacuum ($T_{0\sigma} = T_{\sigma 0} = 1$ for all $\sigma = 1, \ldots, n$) and a potential $\Phi \in \widetilde{\mathcal{B}}$, where $\widetilde{\mathcal{B}}$ is the space of the particles potentials on $\{0, \ldots, n\}_T^{\mathbb{Z}}$ with $\|\,\overline{\Phi}\| < \infty$. We imagine the system enclosed in a finite region Λ which will be kept fixed and whose size will be irrelevant for the discussion so that the results will apply to finite as well as to infinite systems. We follow [GM68] (where the case of spin $\sigma = 0, 1$ is discussed) and, to simplify the analysis, we shall also consider only Φ real valued.

Let $\rho(X, \underline{\sigma}_X)$ be the probability in the Gibbs distribution that the spins found at the sites of $X \subset \Lambda$ are precisely $\underline{\sigma}_X$. We suppose that $\sigma = 0$ describes the vacuum (*i.e.* $\Phi_X(\underline{\sigma}_X) \equiv 0$ if for at least one $x \in X$ one has $\sigma_x = 0$). More generally it will be convenient to use the notation $\rho(X, \underline{\sigma}_X | Y, \underline{\sigma}_Y | \ldots | Z, \underline{\sigma}_Z)$ to denote the probability that the spins in X are $\underline{\sigma}_X$, those in Y are $\underline{\sigma}_Y$, …, those in Z are $\underline{\sigma}_Z$, if $, Y, \ldots, Z$ are disjoint sets: clearly $\rho(X, \underline{\sigma}_X | Y, \underline{\sigma}_Y | \ldots | Z, \underline{\sigma}_Z) \equiv \rho(X \cup Y \cup \ldots \cup Z, \underline{\sigma}_{X \cup Y \cup \ldots \cup Z})$.

Let $x_1 \in X$ be a point arbitrarily chosen among those of X (*e.g.* lexicographically); define, for $X \cap V = \emptyset$,

eA7.2.1

$$\begin{aligned}
&U_1(X, \underline{\sigma}_X) \stackrel{def}{=} \sum_{Y \subset X, Y \ni x_1} \overline{\Phi}_Y(\underline{\sigma}_Y),\\
&W_1(X, \underline{\sigma}_X | V, \tau_V) \stackrel{def}{=} \sum_{S \subset X, S \ni x_1} \overline{\Phi}_{S \cup V}(\underline{\sigma}_S, \underline{\tau}_V), \qquad (A7.2.1)\\
&U_1(X, \underline{\sigma}_X | V, \tau_V) \stackrel{def}{=} \sum_{S \subset X, x_1 \in S, Y \subset V} \overline{\Phi}_{S \cup Y}(\sigma_S, \tau_Y) = \sum_{Y \subset V} W_1(X, \underline{\sigma}_X | Y, \tau_Y),\\
&e^{-U_1(X, \underline{\sigma}_X | V, \tau_V)} = 1 + \sum_{q=1}^{\infty} \sum_{T \subset V} \sum_{Y_1 \cup \ldots \cup Y_q = T} \prod_{i=1}^{q} \big(e^{-W_1(X, \underline{\sigma}_X | Y_i, \underline{\tau}_{Y_i})} - 1\big) \stackrel{def}{=}\\
&\stackrel{def}{=} \sum_{T \subset V} K(X, \underline{\sigma}_X \,|\, T, \underline{\tau}_T),
\end{aligned}$$

where the term with $T = \emptyset$ in the last line corresponds to 1.

Note that

eA7.2.2

$$\begin{aligned}
&|U_1(X, \underline{\sigma}_X)|,\ |W_1(X, \underline{\sigma}_X | V, \underline{\tau}_V)| \le \|\,\overline{\Phi}\|,\\
&\sum_{T, T \cap X = \emptyset} |W_1(X, \underline{\sigma}_X | T, \underline{\tau}_T)| \le \|\,\overline{\Phi}\|, \qquad (A7.2.2)\\
&\sum_{T \cap X = \emptyset, T \ne \emptyset} |K(X, \underline{\sigma}_X \,|\, T, \underline{\tau}_T)| \le e^{\|\,\overline{\Phi}\|}(e^{\|\,\overline{\Phi}\|} - 1) \stackrel{def}{=} \delta(\,\overline{\Phi}).
\end{aligned}$$

Setting $X^{(1)} \stackrel{def}{=} X \setminus x_1$ one deduces from the definition of $\rho(X, \underline{\sigma}_X)$ the

Kirkwood-Salsburg equations, [Ga00],

eA7.2.3

$$
\begin{aligned}
\rho(X,\underline{\sigma}_X) &= e^{-h(\sigma_{x_1})-U_1(X,\underline{\sigma}_X)}\,\cdot\\
&\cdot\Big(\rho(X^{(1)},\underline{\sigma}_{X^{(1)}})-\sum_{\sigma>0}\rho(x_1,\sigma|X^{(1)},\underline{\sigma}_{X^{(1)}})\\
&+\sum_{T\cap X=\emptyset, T\neq\emptyset}\sum_{\underline{\tau}_T}K(X,\underline{\sigma}_X\,|\,T,\underline{\tau}_T)\cdot\\
&\cdot\big(\rho(X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\tau_T)-\sum_{\sigma>0}\rho(x_1,\sigma|X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\underline{\tau}_T)\big)\Big).
\end{aligned}
\tag{A7.2.3}
$$

Introducing

eA7.2.4

$$
\xi=\sum_{\sigma>0}\rho(x_1,\sigma|X^{(1)},\underline{\sigma}_{X^{(1)}}),\qquad \zeta=\sum_{\sigma>0}e^{-h(\sigma)+U_1(X,\underline{\sigma}_X)},
\tag{A7.2.4}
$$

and summing both sides of the equations over $\sigma_{x_1}>0$ we can derive an expression for ξ

eA7.2.5

$$
\xi=\frac{\zeta}{1+\zeta}\rho(X^{(1)},\sigma_{X^{(1)}})+\frac{1}{1+\zeta}\big(e^{-h-U_1},K(\rho'-\xi')\big),
\tag{A7.2.5}
$$

where $K(\rho'-\xi')$ is a short hand for the last sum in (A7.2.3) and the scalar product $\big(e^{-h-U_1},K(\rho'-\xi')\big)$ is

eA7.2.6

$$
\begin{aligned}
&\sum_{\sigma_{x_1}>0}e^{-h(\sigma_{x_1})-U_1(X,\underline{\sigma}_X)}\,\cdot\sum_{T\cap X=\emptyset, T\neq\emptyset}\sum_{\tau_T}K(X,\underline{\sigma}_X\,|\,T,\underline{\tau}_T)\cdot\\
&\cdot\big(\rho(X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\underline{\tau}_T)-\sum_{\sigma>0}\rho(x_1,\sigma|X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\underline{\tau}_T)\big).
\end{aligned}
\tag{A7.2.6}
$$

We imagine replacing the above expression for ξ in the equations (A7.2.3): the latter can be regarded as equations for the quantities $\rho(X,\sigma_X)$, with $X\neq\emptyset$ and $\sigma_x>0\ \forall x\in X$. If we define $\alpha(X,\underline{\sigma}_X)=\frac{e^{-h(\sigma)}}{1+\zeta}$ if $|X|=1$, with $X=x$ and $\sigma=\sigma_x$, and 0 otherwise, the equations take the form

eA7.2.7

$$
\rho=\alpha+K\rho,
\tag{A7.2.7}
$$

where the operator $K\rho(X,\underline{\sigma}_X)$ is defined as

eA7.2.8

$$
\begin{aligned}
&\frac{e^{-h(\sigma_{x_1})-U_1(X,\underline{\sigma}_X)}}{1+\zeta}\big(\rho(X^{(1)},\underline{\sigma}_{X^{(1)}})\delta_{|X^{(1)}|>0}\big)+\\
&-\frac{e^{-h(\sigma_{x_1})-U_1(X,\underline{\sigma}_X)}}{1+\zeta}\big(e^{-h-U_1},K(\rho'-\xi')\big)+\\
&+e^{-h(\sigma_{x_1})-U_1(X,\underline{\sigma}_X)}\sum_{T\cap X=\emptyset, T\neq\emptyset}\sum_{\underline{\tau}_T}K(X,\underline{\sigma}_X\,|\,T,\underline{\tau}_T)\cdot\\
&\cdot\big(\rho(X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\underline{\tau}_T)-\sum_{\sigma>0}\rho(x_1,\sigma|X^{(1)},\underline{\sigma}_{X^{(1)}}|T,\underline{\tau}_T)\big).
\end{aligned}
\tag{A7.2.8}
$$

Setting $\|\rho\| = \max|\rho(X,\sigma_X)|$ one checks that the operator K has the property that

eA7.2.9
$$\frac{\|K\rho\|}{\|\rho\|} \le \frac{ze^{\|\,\overline{\Phi}\|}}{1+ze^{\|\,\overline{\Phi}\|}} + \frac{ze^{\|\,\overline{\Phi}\|}}{1+ze^{\|\,\overline{\Phi}\|}} ze^{\|\,\overline{\Phi}\|}\delta(\,\overline{\Phi})n + ze^{\|\,\overline{\Phi}\|}\delta(\,\overline{\Phi})n \le$$
$$\le \frac{z\,e^{\|\,\overline{\Phi}\|}}{1+ze^{\|\,\overline{\Phi}\|}} + 2zne^{2\|\,\overline{\Phi}\|}(e^{\|\,\overline{\Phi}\|}-1). \qquad (A7.2.9)$$

where the reality assumption on Φ, hence on h and $\overline{\Phi}$ has been used in an essential way (*e.g.* in using the monotonicity of the function $x \to x/(1+x)$).

This shows that the equation $\rho = \alpha + K\rho$ can be solved by iteration and proves regularity in the domain where the *r.h.s.* is < 1. A number of consequences follow: they are most easily exhibited by pursuing the above analysis to show that $\frac{\|K\rho\|}{\|\rho\|} < 1$ implies that, in spite of the apparently different approach, one can define a suitable "polymers representation" of the solutions to the equations (A7.2.8). See [GK69], [DIS73], [GM68]. We do not discuss the matter further as we shall not need here the theory of the classical expansion.

Problems for §7.2

Q7.2.1 **[7.2.1]:** (*Hole–particle symmetry*)
If $n = 1$ and $\Phi \in \widetilde{\mathcal{B}}$ then $\Phi_X(\underline{\sigma}_X) \ne 0$ only if $\sigma_i = 1$, for all $i \in X$; having set $\Phi(X) = \Phi_X(\underline{1}_X)$ we shall identify Φ through the sequence $\{\Phi(X)\}_{X\subset\mathbb{Z}^d}$. One has $\|\Phi\| = \sum_{X\ni 0}|\Phi(X)| < +\infty$ and we shall denote $\mathcal{B}^h_s$ the space of the exponentially decaying potentials such that $\sum_{X\ni 0}\exp(\kappa\,\mathrm{diam}(X))|\Phi(X)| < +\infty$ for some $\kappa > s$. We shall denote B_f the space of the potentials for which $\Phi(X) = 0$ if $|X|$ is large enough (called *potentials with a finite number of bodies*).
Consider the map $\mathcal{L}$ defined on B_f by

$$(\mathcal{L}\Phi)(X) = (-1)^{|X|}\sum_{Y\supset X}\Phi(Y),$$

and show that $\mathcal{L}^2$ = identity and that if $C_\Phi = \sum_{X\ni 0}\frac{\Phi(X)}{|X|}$ the pressures of Φ and of $\mathcal{L}\Phi$ are related by

$$P(\mathcal{L}\Phi) = P(\Phi) + C_\Phi \qquad \forall\,\Phi\in B_f.$$

(*Hint:* Compare $U^\Phi_\Lambda(\underline{\sigma})$ and $U^{\mathcal{L}\Phi}_\Lambda(\underline{\sigma}^c)$ where $\underline{\sigma}^c$ is the configuration in $\{0,\ldots,n\}^\Lambda$ complementary of $\underline{\sigma}\in\{0,\ldots,n\}^\Lambda : \sigma^c_i = 1-\sigma_i$, $i\in\Lambda$; deduce, from this relation, a relation between $Z_{\mathcal{L}\Phi}(\Lambda)$ and $Z_\Phi(\Lambda)$ *etc.*, [GMR68].)

Q7.2.2 **[7.2.2]:** (*Gibbs states transformations under hole–particle symmetry*)
In the context of problem [7.2.1] find the relation between $G(\mathcal{L}\Phi)$ and $G(\Phi)$ by providing the expression of the functions $\rho_{\Phi,\mu}(X) = \mu\Big(\prod_{i\in X}\sigma_i\Big)$ in terms of the analogous functions relative to a Gibbs distribution for $\mathcal{L}\Phi$. (*Hint:* Use the relation between the pressures found in problem [7.2.1] and the fact that the pressure is the generating functional of the average values of the local observables, cf. proposition (6.1.2) and the related remark.)

Q7.2.3 **[7.2.3]:** Find a condition that implies that $\Phi = \mathcal{L}\Phi$ if Φ is a two-body potential.

Q7.2.4 **[7.2.4]:** (*Spin potentials*)
A potential Ψ for $\{0,1\}^{\mathbb{Z}}$ is called a *spin potential* if it can be written as $\Psi_X(\underline{\sigma}_X) = J(X)\prod_{i\in X}(2\sigma_i - 1)$.
If Ψ is a spin potential and μ_Ψ is a Gibbs state for it show, having set $(\mathcal{L}'\Psi)_X = (-1)^{|X|}\Psi_X$, that $\mathcal{L}'\Psi$ is a spin potential that admits among its Gibbs states the probability distribution μ such that $\mu(\prod_{i\in X}(2\sigma_i-1)) = (-1)^{|X|}\mu_\Psi(\prod_{i\in X}(2\sigma_i-1))$. (*Hint:* Proceed as in the analysis of problem [7.2.1].)

Q7.2.5 **[7.2.5]:** Prove that the map defined in problem [7.2.4] between measures in $G(\mathcal{L}'\Psi)$ and in $G(\Psi)$ is one-to-one and invertible.

Q7.2.6 **[7.2.6]:** (*Equivalence between spin systems and particle systems*)
Which relation exists between $\mathcal{L}$ and $\mathcal{L}'$? (*Hint:* Show that it is possible to establish a correspondence between spin potentials with a finite number of bodies and particle potentials with a finite number of bodies so that, via this correspondence, $\mathcal{L}$ is transformed into $\mathcal{L}'$).

Q7.2.7 **[7.2.7]:** Show that if $n = 1$, $T_{\sigma\sigma'} = 1$, condition (7.2.5) written for $\mathcal{L}\Phi$ instead of Φ leads to the determination of a new region of analyticity for μ_Φ via the relation between $G(\Phi)$ and $G(\mathcal{L}\Phi)$ derived in problem [7.2.2].

Q7.2.8 **[7.2.8]:** Estimate the mixing rate of μ_Φ in terms of the decay property at ∞ of Φ under the hypotheses that the potential is a particle potential and that z is small enough. More generally estimate the mixing rate for cylindrical functions. (*Hint:* Study first the case of finite range Φ: in this case it follows from the definitions that the $\varphi^T(\xi_1\ldots\xi_p\ldots)$ (cf. (7.1.23)) are zero if the points $\xi_1\ldots\xi_p$ are not close enough; then by using the expressions (7.2.14), (7.2.13) deduce an exponential bound for the mixing rate, *etc.*)

Q7.2.9 **[7.2.9]:** (*High temperature analyticity in lattice systems of particles*)
Making use of the method of problem [7.2.7] to further enlarge the region of analyticity show that if $(h,\overline{\Phi})$ is a potential with finitely many bodies and if $(h',\Phi') = \mathcal{L}(h,\overline{\Phi})$ then the region $z\|\overline{\Phi}\| < B$ or $z'\|\overline{\Phi}'\| < B$ contains a region having the form $\|\overline{\Phi}\| < \varepsilon$ with a suitable $\varepsilon > 0$ (independent of the value of z). Deduce that, given $\overline{\Phi}$, this proves analyticity at high temperature (*i.e.* small $\overline{\Phi}$) for all values of h. The weaker form of proposition (7.2.1) proved in proposition (7.2.2) suffices for such conclusion, [GMR68], [Do69], see footnote 7 of [Ru69], p. 112.

Q7.2.10 **[7.2.10]:** (Analiticity for strong decay in $\mathbb{Z}^d$)
Consider a spin system on $\mathbb{Z}^d$ without hard core interactions. Show that there is $\kappa_0 > 0$ such that if $\|\overline{\Phi}\|_\kappa + b \le \varepsilon$ for $\kappa \ge \kappa_0$ and ε small then the Gibbs distribution μ_Φ is analytic in Φ.(*Hint:* Use remark (2) after proposition (7.2.3).)

Q7.2.11 **[7.2.11]:** (*Decimation in* $d = 1$)
Let T be a compatibility matrix which mixes over a time a and let $h > a$. Let $\underline{\sigma}$ be a T–compatible sequence and define the sequence $\underline{\eta}(\underline{\sigma}) = (\ldots,\eta_{-1}(\underline{\sigma})\eta_0(\underline{\sigma}),\eta_1(\underline{\sigma}),\ldots)$ by

$$\eta_k(\underline{\sigma}) = (\sigma_{ka+kh},\ldots,\sigma_{(k+1)a-1+kh}),\ k = 0,\pm1,\pm2,\ldots$$

Check that all sequences $\underline{\eta}$ consisting of compatible strings of length a have the form $\eta_k(\underline{\sigma})$ for some $\underline{\sigma}$ which is T–compatible. In other words if one leaves a "gap" larger than a between compatible strings of length $\tau > a$ no further compatibility condition is required in order that the strings be part of an infinite compatible sequence. The strings η_k regarded as new spins can be freely put down spaced by h, or more, giving rise to a configuration of symbols that can be regarded as part of a T–compatible sequence $\underline{\sigma}$.

Q7.2.12 **[7.2.12]:** (*Reduction by decimation of hard core finite range systems to coreless systems in* $d = 1$)
In the context of problem [7.2.11] consider a short range real potential $\Phi \in \mathcal{B}$ with range r (*i.e.* $\Phi_X \equiv 0$ if the diameter of X exceeds r). Let $\tau = \max(a,r)+1$ and $h = m\tau$ with m large (to be determined later; see problem [7.2.14]). Write a spin configuration

in the interval $[1, N(\tau+h)+\tau]$ as a sequence $\underline{\eta}_1, \underline{\sigma}_1, \underline{\eta}_2, \ldots, \underline{\eta}_N, \underline{\sigma}_N, \underline{\eta}_{N+1}$ where $\underline{\eta}_j$ are compatible strings of length τ and $\underline{\sigma}$ is a compatible string of length h. Consider the matrix $M_{\underline{\eta},\underline{\eta}'}$ defined for η, η' being T–compatible sequences of length τ, $\underline{\eta} = (\eta_1, \ldots, \eta_\tau)$ and $\underline{\eta}' = (\eta'_1, \ldots, \eta'_\tau)$, and given by

$$M_{\underline{\eta},\underline{\eta}'} = 0, \qquad \text{unless} \quad \eta'_1 \equiv \eta_2, \ldots, \eta'_{\tau-1} = \eta_\tau.$$

Show that the number Z_0 of compatible strings of length $N(\tau+h)+\tau$ with $h = m\tau$ is

$$Z_0 = \sum_{\underline{\eta}_1, \ldots, \underline{\eta}_{N+1}} \prod_{j=1}^{N} \langle\, \underline{\eta}_j \,|\, M^h \,|\underline{\eta}_{j+1} \,\rangle\,,$$

where we use the notation $\langle\, \underline{\eta} \,|\, M \,|\underline{\eta}' \,\rangle \stackrel{def}{=} M_{\underline{\eta},\underline{\eta}'}$.
Likewise let $(\underline{\eta}, \underline{\eta}') = (\eta_1, \ldots, \eta_\tau, \eta'_1, \ldots, \eta_\tau)$ be a pair of compatible strings of length τ (not necessarily such that the string $(\underline{\eta}, \underline{\eta}')$ of length 2τ obtained by merging the two is compatible). Let $\underline{\sigma} = (\sigma_1, \ldots, \sigma_{\tau+1}) \stackrel{def}{=} (\eta_1, \ldots, \eta_\tau, \eta'_\tau)$ and define $W(\underline{\eta}, \underline{\eta}') \stackrel{def}{=} \sum_X^* \Phi_X(\underline{\sigma}_X)$ where the $*$ over the sum indicates that it runs over the subsets X of the labels $(1, \ldots, \tau+1)$ which contain *the extra site* $\tau+1$. Define also $U(\underline{\eta}) \stackrel{def}{=} \sum_X \Phi_X(\underline{\eta}_X)$ where the sum runs over the subsets of the labels $(1, \ldots, \tau)$ and $\overline{M}_{\underline{\eta},\underline{\eta}'} = e^{-W(\underline{\eta},\underline{\eta}')} M_{\underline{\eta},\underline{\eta}'}$. Check that the partition function Z for the system $\{0, \ldots, n\}^\Lambda$ on $\Lambda = [1, N(\tau+h)+\tau]$ is given by

$$\sum_{\underline{\eta}_1, \ldots, \underline{\eta}_{N+1}} e^{-U(\underline{\eta}_1)} \prod_{j=1}^{N} \langle\, \underline{\eta}_j \,|\, \overline{M}^h \,|\underline{\eta}_{j+1} \,\rangle$$

Q7.2.13 **[7.2.13]:** (*Decimation in $d=1$ by transfer matrix*)
In the context of problem [7.2.11] let $\overline{M}_{\underline{\eta},\underline{\eta}'} = \lambda\, \overline{w}_{\underline{\eta}} w_{\underline{\eta}'} + \Lambda P_{\underline{\eta},\underline{\eta}'}$ be the spectral decomposition of the mixing matrix $\overline{M}$ (see definition in problem [2.3.12]) along the eigenvector with largest eigenvalue λ which is simple and with an eigenvector denoted $w_{\underline{\eta}}$ which has also positive components (by Perron–Frobenius theorem, see problems [2.3.10] through [2.3.12] and [4.1.12]), while the eigenvalue of the transposed matrix of $\overline{M}$ is denoted $\overline{w}_{\underline{\eta}}$: they are chosen with (strictly) positive components and normalized so that $\sum_{\underline{\eta}} \overline{w}_{\underline{\eta}} w_{\underline{\eta}} = 1$) and P is a matrix with norm $\|P\| = 1$ and $\Lambda < \lambda$. Deduce from the above remark on the structure of Z, setting $\varepsilon = \Lambda/\lambda < 1$, that Z is given by

$$\lambda^{Nh} \sum_{\underline{\eta}_1, \ldots, \underline{\eta}_{N+1}} e^{-U(\underline{\eta}_1)} \prod_{i=1}^{N} \Big(\overline{w}_{\underline{\eta}_i} w_{\underline{\eta}_{i+1}} + \Big(\frac{\Lambda}{\lambda}\Big)^h P_{\underline{\eta}_i, \underline{\eta}_{i+1}} \Big)$$

and show that the sum can be written as a sum over $\underline{\eta}_1, \ldots, \underline{\eta}_{N+1}$ of

$$e^{-U(\underline{\eta}_1)}\, \overline{w}_{\underline{\eta}_1} w_{\underline{\eta}_{N+1}} \prod_{j=2}^{N} \overline{w}_{\underline{\eta}_j} w_{\underline{\eta}_j} \prod_{j=1}^{N} e^{\log\big(1+(\Lambda\lambda^{-1})^h P_{\underline{\eta}_j,\underline{\eta}_{j+1}}/(\overline{w}_{\underline{\eta}_j} w_{\underline{\eta}_{j+1}})\big)}$$

Show that this means that the *decimated* system of the spins $\underline{\eta}$ separated by $m\tau$ sites can be regarded, in the thermodynamic limit (*i.e.* in the limit as $N \to \infty$) as a spin system in which the spins are the compatible strings $\underline{\eta}$ of length τ and the interaction potential is simply a single-site and nearest neighbor interactions given by

$$h(\underline{\eta}) = -\log(\overline{w}_{\underline{\eta}} w_{\underline{\eta}}), \qquad \overline{\Phi}(\underline{\eta}, \underline{\eta}') = -\log\Big(1 + (\Lambda\lambda^{-1})^h \frac{P_{\underline{\eta},\underline{\eta}'}}{\overline{w}_{\underline{\eta}} w_{\underline{\eta}'}}\Big),$$

with suitably modified contributions for the sites 1 and $N+1$.

Q7.2.14 **[7.2.14]:** (*Cluster expansion with hard cores and finite range forces, $d = 1$*) Check that problem [7.2.13] implies the validity of proposition (7.2.3) when the transition matrix is mixing and the interaction potential has finite range. (*Hint:* The above problems reduce, by choosing h large enough, the case considered here to the case discussed in the proof of proposition (7.2.3) of interaction without hard core ans with arbitrary one-body potential $h(\underline{\eta})$ and finite range interaction without hard core.)

Bibliographical note for §7.2

Among the most remarkable extensions are perhaps those in [GK78], [Ku78], [Sy79], [Is76].

The classical theory of Appendix 7.2 is originally due to Groeneveld, Penrose and Ruelle, [Gr62],[Pe63],[Ru63]. The above papers were preceded by the work of Morrey, [Mo55], who also solved the same problem, by essentially the same arguments, in an ambitious attempt to derive the equations of fluidodynamics from microscopic classical dynamics: the importance of this paper for the theory of the cluster expansion was realized only about twenty years later.

The proliferation of alternative or independent and different proofs, or of nontrivial extensions, shows that in reality the problem is a natural one and that the methods to study it with the techniques of this section are also natural although they are still considered by many as not elegant (and not really natural), and they are avoided when possible or commented by saying that "it must be possible to obtain the same result in a simpler way" (often not followed by any actual work in this direction). An elegant and more general analysis is in [Ca82].

§7.3 Renormalization by decimation in one-dimensional systems

The following definition is particularly relevant for the theory of dynamical systems (see also problem [5.2.1]).

D7.3.1 **(7.3.1) Definition:** (Fisher potentials)
Consider a $\frac{n}{2}$-spin system on the one-dimensional lattice $\mathbb{Z}$ and consider the space $\mathcal{B}_F$ of the potentials Φ with the property

e7.3.1
$$\Phi_X \equiv 0 \text{ unless } |X| = 1 + \text{diam}(X) = 1 + \delta(X), \qquad (7.3.1)$$

i.e. Φ_X vanishes unless X is an interval. We shall also call such potentials extended nearest neighbor potentials. *We shall split the potentials into their* single spin part h *and the* many-spin part $\overline{\Phi}$ *and write* $\Phi = (h, \overline{\Phi})$, *where* $h_\xi(\sigma_x) = \Phi_\xi(\sigma_x)$ *and* $\overline{\Phi}_\xi \equiv 0$.

The name of "Fisher potentials" is used in the Physics literature: such potentials are quite interesting as they can be used to produce examples and counterexamples to various phenomena concerning Gibbs states, [Fi67], [Ga77]. Not least: the potentials that arise in the theory of Anosov systems are in this class, cf. (4.3.15), and have short range.

The main step toward the results on smoothness of Gibbs distributions relevant for dynamical systems concerns one-dimensional systems without hard core and with an extended nearest neighbor potential (see definition (7.3.1)).

P7.3.1 **(7.3.1) Proposition:** (Weakly Markovian chains)
Let $\mathcal{B}_F$ be the space of the extended nearest neighbor real translation invariant potentials Φ for a $\frac{n}{2}$-spin one-dimensional system and define for $\kappa > 0$

e7.3.2
$$\|\Phi\|_\kappa \stackrel{def}{=} \sum_{\xi\in C\subset\mathbb{Z},\,|C|\ge 1}^{C} \sup_{\underline{\sigma}_C} |\Phi_C(\underline{\sigma}_C)|\, e^{\kappa\,\mathrm{diam}(C)} \equiv ||h||_0 + ||\,\overline{\Phi}||_\kappa. \tag{7.3.2}$$

The compatibility matrix is supposed to be $T_{\sigma\sigma'} \equiv 1$ (i.e. no hard core interaction).
Then the correlation functions $\mu_\Phi(C^V_{\sigma_V})$ are real analytic functions of $\Phi \in \Sigma$, where $\Sigma \subset \mathcal{B}_F$ denotes the set of translation invariant Fisher potentials with $\|\Phi\|_\kappa < \infty$ for some κ. Furthermore the Gibbs distribution μ_Φ is exponentially mixing.

Remark: Note that $||h||_\kappa \equiv ||h||_0$ because if $|C| = 1$ then $\mathrm{diam}\,(C) = 0$. We stress that no condition on the size $||\Phi||_\kappa$ is assumed here. The latter quantity only affects the size of the complex analyticity domain.

Proof: The idea behind our discussion is simple and can be grasped by looking at the problems following problem[7.2.11] which deal with the simpler case in which $\Phi_C \equiv 0$ if the diameter $\mathrm{diam}(C) = \delta(C)$ is large enough: however here we deal directly, and independently of the mentioned problems, with the new case. We first study the case in which Φ is *real.*

Let Λ be a large interval and consider the strings $\underline{\sigma} \in \{1, \ldots, n\}^\Lambda$.

Divide the interval Λ into a sequence of intervals of lengths alternating between $\tau \ge 1$ (which means an interval consisting of at least a pair of nearest neighbors) and $h > \tau$, here h, τ are lengths to be chosen later. The intervals will be denoted

e7.3.3
$$B_0, H_0, B_1, H_1, \ldots, B_{\ell-1}, H_{\ell-1}, B_\ell, \tag{7.3.3}$$

and will be called of B–type and of H–type respectively. Of course this means that we suppose that the number of points in Λ is $|\Lambda| = \ell(\tau + h + 2) + \tau + 1$ (note that an interval of length h contains $h+1$ lattice points). Correspondingly we imagine that a string $\underline{\sigma} \in \{1, \ldots, n\}^\Lambda$ is a sequence

e7.3.4
$$\underline{\sigma} = (\beta_0, \eta_0, \beta_1, \eta_1, \ldots, \beta_{\ell-1}, \eta_{\ell-1}, \beta_\ell) \tag{7.3.4}$$

of shorter strings containing alternatively sets of $\tau+1, h+1, \tau+1, \ldots, \tau+1, h+1, \tau+1$ spins. The Gibbs state with potential Φ is unique, by the analysis of Section §5.2 and it can be generated by taking the limit as $\Lambda \to \infty$ of the probability distribution

e7.3.5
$$\mu_\Phi^\Lambda(\underline{\sigma}) = \frac{e^{-U_\Phi^\Lambda(\underline{\sigma})}}{Z_\Phi(\Lambda)} \tag{7.3.5}$$

over a sequence of intervals Λ of the above kind, *i.e.* of length $\ell(h+\tau+2)+\tau+1$.

The spins β_i have size $n^{\tau+1}$, *i.e.* they can take $n^{\tau+1}$ different values, and will be called *block spins*. They can be imagined to be located on the points of a lattice whose sites are labeled $B_0, B_1, \ldots$ and which will be called the *decimated lattice* or B–lattice: on the new lattice the distances will be computed by thinking of it as having its sites spaced by 1. The distances computed in this way, *i.e.* essentially in units of $\tau+h$, will be called "renormalized distances" when it will be necessary to avoid confusion with the distances of points and sets in the original lattice.

The distribution μ_Φ^Λ generates, by summation over the variables $\underline{\eta} = \{\eta_j\}_{j=0}^{\ell-1}$, a probability distribution over the variables $\underline{\beta} = \{\beta_j\}_{j=0}^{\ell}$ (which is a sequence of $(\tau+1)$–ples of spins)

e7.3.6
$$\overline{\mu}_\Phi^\Lambda(\underline{\beta}) = \sum_{\eta_0,\ldots,\eta_{\ell-1}} \mu_\Phi^\Lambda(\underline{\sigma}), \tag{7.3.6}$$

and our first task will be to check that the limit as $\Lambda \to \infty$ of $\overline{\mu}_\Phi^\Lambda(\underline{\beta})$ is a Gibbs distribution with an exponentially decreasing *renormalized* potential $\Phi_{\rm ren}$ on the sequences $\underline{\beta} = \{\beta_j\}_{j\in\mathbb{Z}}$, with $\beta_j \in \{1,\ldots,n\}^{\tau+1}$, *which has a range* κ_1^{-1} *much shorter than* κ^{-1} *if* τ *is large* ($\kappa_1 \xrightarrow[\tau\to\infty]{} \infty$) *and a size* $\|\,\overline{\Phi}_{\rm ren}\|_{\kappa_1} \stackrel{def}{=} \nu_1 \xrightarrow[\tau\to\infty]{} 0$.

In other words on the new space of sequences $\underline{\beta}$ of symbols $\beta_j \in \{1,\ldots,n\}^{\tau+1}$ the distribution $\overline{\mu}$ is still a Gibbs distribution with a new suitable potential that we shall call $\Phi_{\rm ren}$ verifying the property that $||\,\overline{\Phi}_{\rm ren}||_{\kappa_1}$ is as small as wished and κ_1 is as large as wished. *This means that the (possibly strong) initial interaction can be eliminated at the cost of dealing with a spin system of somewhat larger spin, i.e.* of size $n^{\tau+1}$ rather than n. In this way we can derive the stated result as a special case of proposition (7.2.3).

The energy $U_\Phi^\Lambda(\underline{\sigma})$ can be expressed in terms of the energy of the part of the potential Φ_τ obtained by cutting off from Φ the components $\Phi_X(\underline{\sigma}_X)$ with $\delta(X) > \tau$ (recall that $\tau \ge 1$ and, in the present one-dimensional case, $\delta(X)$ denotes the diameter of X) plus the part of the potential containing only the latter components. Thus we can write

$$U_\Phi^\Lambda(\underline{\sigma}) = \sum_{i=0}^{\ell} U_\tau^B(\beta_i) + \sum_{i=0}^{\ell-1} U^H(\eta_i) + \sum_{i=0}^{\ell-1} (W_\tau(\beta_i,\eta_i) + W_\tau(\eta_i,\beta_{i+1})) +$$

e7.3.7

$$+\sum_{\substack{X\,:\,\delta(X)>\tau\\ X\not\subset H_i}}\Phi_X(\underline{\sigma}_X). \tag{7.3.7}$$

Here U^B_τ and U^H denote the energies of the spins in the sets B_i and H_i, respectively, due to the cut off potential Φ_τ and to the full potential Φ; the terms $W_\tau(\beta_i,\eta_i)$ and $W_\tau(\eta_i,\beta_{i+1})$ yield the contribution to the energy from the potentials Φ_X with $\delta(X)\le\tau$ and with the set X containing points of H_i and either of B_i or of B_{i+1} respectively. The last terms are the contributions to the energy from potentials Φ_X involving only sets X with $\delta(X)>\tau$ which do not lie entirely inside a set H_i. Therefore the partition function can be written as

e7.3.8

$$\begin{aligned}&\sum_{\beta_0,\ldots,\beta_\ell}\sum_{\eta_0,\ldots,\eta_{\ell-1}}\Big(\prod_{i=0}^{\ell}e^{-U^B_\tau(\beta_i)}\Big)\cdot\\ &\quad\cdot\Big(\prod_{i=0}^{\ell-1}e^{-U^H(\eta_i)-W_\tau(\beta_i,\eta_i,\beta_{i+1})}\Big)\Big(\prod_{\substack{X\,:\,\delta(X)>\tau\\ X\not\subset H_i}}e^{-\Phi_X(\underline{\sigma}_X)}\Big),\end{aligned}\tag{7.3.8}$$

where $W_\tau(\beta_i,\eta_i\,\beta_{i+1})\overset{def}{=}W_\tau(\beta_i,\eta_i)+W_\tau(\eta_i,\beta_{i+1})$. Setting

e7.3.9

$$Z_h(\beta_i,\beta_{i+1})=\sum_\eta e^{-U^H(\eta)}e^{-W_\tau(\beta_i,\eta,\beta_{i+1})}, \tag{7.3.9}$$

the partition function expression in (7.3.8) becomes

e7.3.10

$$\begin{aligned}&\sum_{\beta_0,\ldots,\beta_\ell}\Big(\prod_{i=0}^{\ell}e^{-U^B_\tau(\beta_i)}\Big)\Big(\prod_{i=0}^{\ell-1}Z_h(\beta_i,\beta_{i+1})\Big)\cdot\\ &\cdot\frac{\sum_{\eta_0,\ldots,\eta_{\ell-1}}\prod_{i=0}^{\ell-1}e^{-U^H(\eta_i)-W_\tau(\beta_i,\eta_i,\beta_{i+1})}e^{-\sum^*_X\Phi_X(\underline{\sigma}_X)}}{\prod_{i=0}^{\ell-1}Z_h(\beta_i,\beta_{i+1})},\end{aligned}\tag{7.3.10}$$

where the $*$ means that the sum is over the X such that $\delta(X)>\tau$ and $X\not\subset H_i$ for any set H_i.

It is possible to interpret the last ratio as a partition function of "large" spins $\eta_0,\ldots,\eta_{\ell-1}$ (the word *large* refers to the fact that there are n^{h+1} different values that each η can take) interacting via a potential (suitably) generated by the functions $\Phi_X(\underline{\sigma}_X)$ with $\delta(X)\ge\tau+1$ (*i.e.* which does not involve the components of Φ involving up to τ neighbors).

In attempting an application of corollary (7.2.3) we must note that we can think of the sequence of "spins" $\underline{\eta}$ as a sequence of spins on a lattice whose points are $H_0,\ldots,H_{\ell-1}$ and will be thought as spaced by τ: then the potentials between the spins $\eta_0,\ldots,\eta_{\ell-1}$ at fixed $\beta_0,\ldots,\beta_\ell$ are many-body potentials

e7.3.11

$$\widetilde{\Phi}_{H_i,\ldots,H_{i+q}}(\eta_i,\ldots,\eta_{i+q})=\sum_{X\cap H_k=\emptyset,\,k=i,\ldots,i+q}\Phi_X(\underline{\sigma}_X), \tag{7.3.11}$$

which, by the assumption that $\Phi_C \not\equiv 0$ only if C is an interval, are the only nonvanishing potentials; in other words the potential $\widetilde\Phi$ is still a (*non-translation invariant and $\underline\beta$–dependent*) Fisher potential, cf. definition (7.3.1).

To take advantage of corollary (7.2.3) we need, *at fixed* $\beta_0,\ldots,\beta_\ell$, an estimate of a quantity like

e7.3.12
$$\sup_H \sum_{\mathcal{C}\ni H\,:\,|\mathcal{C}|>1} \sup_{\underline\eta_{\mathcal{C}}} |\widetilde\Phi_{C_1,\ldots,C_k}(\eta_{C_1},\ldots,\eta_{C_k})| e^{\frac12 (k-1)\tau\kappa}, \tag{7.3.12}$$

where $\mathcal{C}=(C_1,\ldots,C_k)$ is an "interval" on the lattice formed by H-type intervals, and $\underline\eta_{\mathcal{C}}=(\eta_{C_1},\ldots,\eta_{C_k})$. The case $|\mathcal{C}|=2$, *i.e.* $k=2$, corresponds to the interaction between nearest neighbors pairs of H-type intervals.

By (7.3.11) $|\widetilde\Phi_{C_1,\ldots,C_k}(\underline\eta_{C_1},\ldots,\underline\eta_{C_k})| \le \sum_{X\cap C_i\ne\emptyset,\,\forall i} |\Phi_X(\underline\sigma_X)|$; hence if $k\ge 2$ and H,H' are two different H-type intervals, we can bound (7.3.12) by

e7.3.13
$$\begin{aligned}&\sup_{H,H'} \sum_{\substack{X\cap H,\, X\cap H'\ne\emptyset\\ \delta(X)>\tau}} |\Phi_X(\underline\sigma_X)| e^{\kappa\delta(X)} e^{-\frac{\kappa}{2}\delta(X)} \le \\ &\le ||\,\overline\Phi||_\kappa \sum_{p=\tau+3}^{\infty} (p-(\tau+2))\, e^{-\frac{\kappa}{2}(p-1)} \le \frac{||\,\overline\Phi||_\kappa e^{-\frac{\kappa}{2}\tau}}{(1-e^{-\kappa/2})^2} \le m\,||\,\overline\Phi||_\kappa\, e^{-\frac{\kappa}{2}\tau},\end{aligned} \tag{7.3.13}$$

having taken into account that the sets X are, by assumption, intervals and having defined $m=(1-e^{-\kappa/2})^{-2}$.

N7.3.1 We choose τ so that[1] we can apply, if h is large enough, the corollary (7.2.3) to proposition (7.2.3) with $\|\,\overline\Phi\|_\kappa$ replaced by $m\,||\,\overline\Phi||_\kappa e^{-\kappa\tau/2}$ and $L_0=\frac18\kappa\tau$ (in (7.2.22) there is a factor $e^{-\kappa\tau/4}$ and one can suppose that h, hence τ, is so large that one can take into account the presence of the other factors simply by replacing 4 with 8 in the bound). This means that we can rewrite the ratio in (7.3.10) as

e7.3.14
$$\exp\sum_\Gamma \varphi^T(\Gamma)\zeta(\Gamma), \tag{7.3.14}$$

where Γ is a polymer configuration consisting of nontrivial (*i.e.* γ with more than one point, $|\gamma|>1$) polymers $\gamma_1,\gamma_2,\ldots$ which are intervals on the lattice of the B's; φ^T are the universal combinatorial coefficients whose existence and properties are proved in proposition (7.1.1). Therefore $\widetilde\Gamma$, *i.e.* the set of points covered by the polymer configuration Γ, is also necessarily an interval. The activities of the polymers verify an inequality like (7.2.23), *i.e.* (for $||\,\overline\Phi||_\kappa<1$)

e7.3.15
$$|\zeta(\gamma)| \le e^{-\frac18\kappa\tau|\gamma|} e^{-\frac14\kappa\tau\delta(\gamma)} \le e^{-\frac18\kappa\tau|\gamma|} e^{-\frac18\kappa\tau\delta(\gamma)}, \tag{7.3.15}$$

[1] Actually we also want that $\tau+1$ divides $h+1$. This can be assumed without loss of generality.

where the tree distance $\delta(\gamma)$ *has to be measured on the "decimated lattice" of the* B*'s*, *i.e.* the $\delta(\gamma)$ of a polymer consisting of k successive B–intervals $(k \geq 2)$ is $k-1$ (while the actual distance between the extreme points of the interval would be $k(\tau+1)+(k-1)(h+1)$: hence the unit of length has to be regarded as essentially equal to $\tau+h+2$). Therefore we can define

e7.3.16
$$\Psi_{B_j,\ldots,B_{j+p}}(\beta_j,\ldots,\beta_{j+p}) = \sum_{\Gamma\cap B_i\neq\emptyset, i=0,\ldots,p} \varphi^T(\Gamma)\zeta(\Gamma) \tag{7.3.16}$$

and we find by (7.1.19) that

e7.3.17
$$\|\Psi_{B_j,\ldots,B_{j+p}}\| \leq A_0 e^{-\frac{\kappa\tau}{16}} e^{-(p-1)\frac{\kappa h}{16}} \tag{7.3.17}$$

where A_0 depends only on κ ($A_0 = (1+\sum_{q>0} q e^{-\frac{\kappa\tau q}{16}})$, cf. (7.1.19)). The exponent of (7.3.14) has therefore the form of a short range potential Ψ between the spins $\underline{\beta}$ and

e7.3.18
$$\|\Psi\|_{\kappa\tau/32} \leq B_0 e^{-\frac{\kappa\tau}{32}} \stackrel{def}{=} \varepsilon_0, \tag{7.3.18}$$

for B_0 suitably chosen and Ψ is an exponentially decaying, (mildly) non-translation invariant[2], interaction between the $\underline{\beta}$'s (here the tree distances have to be measured in the above mentioned units of length for the decimated lattice) with the property that $\Psi_C \neq 0$ only if C is an interval: note that $\|\Psi\|_{\kappa\tau/32}$ includes bounds on the one-body and nearest neighbors blocks. The range of Ψ is much shorter than the original κ because the latter is now replaced by a quantity of the order of $\kappa\tau$.

N7.3.2a

The sum, *i.e.* the partition function, in (7.3.8) can be written as

e7.3.19
$$\begin{aligned}\sum_{\underline{\beta}} e^{-\sum_{X,\,|X|>1}\Psi_X(\underline{\beta}_X)}\Big(\prod_{i=0}^{\ell} e^{-U^B_\tau(\beta_i)}\Big)\\ \Big(\prod_{i=0}^{\ell-1} Z_h(\beta_i,\beta_{i+1})\Big)\Big(\prod_{i=0}^{\ell} e^{-\Psi_{\xi_i}(\beta_i)}\Big),\end{aligned} \tag{7.3.19}$$

where here and in the remaining part of the proof the sets X will denote subsets of the decimated lattice, *i.e.* of the lattice of the B intervals.

We fix τ: and we see that we have to consider a spin chain of arbitrary length with spins β interacting via a single spin potential $a_\xi(\beta)$ and a nearest neighbor pair potential $a_{\{\xi,\xi'\}}(\beta,\beta')$ both of arbitrary size and translation invariant plus a very weak, possibly non translation invariant, spin potential $a_X(\beta_X)$, $|X|\geq 2$ which satisfies

e7.3.20
$$|a_X(\underline{\beta}_X)| \leq \varepsilon e^{-\kappa h\delta(X)}, \tag{7.3.20}$$

[2] Translation invariance violation comes only from the B blocks at the boundaries.

with h prefixed *arbitrarily*, cf. (7.3.18). We call this potential A and refer to this model as the "supplementary model".

In our case $\beta\in\{1,\ldots,n\}^\tau$ and $a_\xi(\beta)=\Psi_\xi(\beta)+U^B_\tau(\beta)$ and $a_{\{\xi,\xi'\}}(\beta,\beta')=-\log Z_h(\beta,\beta')+\Psi_{\{\xi,\xi'\}}(\beta,\beta')$.

Therefore we see that the real problem is just to show that by taking h large the pair potential between nearest neighbors generated above between the blocks of β spins is the sum of a suitable one-spin potential plus a nearest neighbor potential which is as small as we please. This means that we have to show that the above supplementary model is such that its partition function in a box $\Lambda=\{1,\ldots,h\}$ verifies

e7.3.21
$$-\log Z_h(\beta,\beta')=z(\beta)+z(\beta')+e^{-\kappa_0 h}\Delta(\beta,\beta') \tag{7.3.21}$$

for a suitable $z(\beta)$ and with $\Delta(\beta,\beta')\le D$ for some $D,\kappa_0>0$.

The latter is a well known property of exponentially decreasing interactions in one-dimensional systems and we guide to its check in the hints to the problems at the end of the section.

If $V\subset H_i$ for some i and $\mu_{\Phi,\underline{\beta}}$ denotes the Gibbs distribution conditioned to the spins $\underline{\beta}$ the above analysis implies, in fact, that the probabilities $\mu_\Phi(C^V_{\underline{\sigma}^0_V})$ of cylinders $C^V_{\underline{\sigma}^0_V}$ can be written as

e7.3.22
$$\mu_\Phi(C^V_{\underline{\sigma}^0_V})=\int\overline{\mu}_\Phi(d\underline{\beta})\int\mu_{\Phi,\underline{\beta}}(d\underline{\eta})\chi_{C^V_{\underline{\sigma}^0_V}}(\underline{\eta}_V), \tag{7.3.22}$$

The latter expression admits a polymer expansion of the type (7.2.11) with the polymer activities ζ which are analytic and verify the appropriate bounds like (7.2.12) together with their derivatives. The inclusion $V\subset H_i$ is not restrictive because of the arbitrariness of h; the details are left to the reader.

Turning to the smoothness and analyticity statement we assume now that Φ is complex and call $\nu\stackrel{def}{=}\|\mathrm{Im}\,\overline{\Phi}\|_\kappa+\sup_\sigma|\mathrm{Im}\,\Phi_\xi(\sigma)|$. We then proceed through the algebra leading to (7.3.10). However we write (7.3.10) by replacing $U^H(\eta_i),W_\tau(\beta_i,\eta_i,\beta_{i+1})$ with their real parts $\mathrm{Re}\,U^H(\eta_i)$ and $\mathrm{Re}\,W_\tau(\beta_i,\eta_i,\beta_{i+1})$; the corresponding imaginary parts will be taken into account by interpreting the $\sum^*$ in (7.3.10) as containing besides the terms with $|X|>\tau$ also the imaginary parts of the Φ_X which occur in $U^H_\tau(\eta_i)$ and $W_\tau(\beta_i,\eta_i,\beta_{i+1})$. As a consequence in (7.3.12) there will be terms with $|C|=1$ and (7.3.12) will receive contributions also from the imaginary parts of the potentials Φ_X with $|X|\le\tau$. The factor $m||\Phi||_\kappa e^{-\kappa\tau/2}$ in the r.h.s. of inequality (7.3.13) will be modified into $m||\Phi||_\kappa e^{-\kappa\tau/2}+\nu h$. Hence if the imaginary parts of Φ_X are small enough ($e.g.$ $h\nu<m||\Phi||_\kappa\, e^{-\kappa\tau/2}$) (7.3.13) is changed by replacing $||\Phi||_\kappa$ by $2||\Phi||_\kappa$ and we get (7.3.16).

The analysis continues in the same way as in the real Φ case because the new $Z_h(\beta,\beta')$ only depends on the real part of the potential. The uniformity of the bounds in the (tiny) complex strip $\|\mathrm{Im}\,\Phi\|_\kappa+\max_\sigma|\mathrm{Im}\,\Phi_\xi(\sigma)|=$

$O(e^{-\kappa\tau/2})$, with h chosen large enough, implies the analyticity statement. ■

Consider a general *one-dimensional system with hard core interaction* given by a mixing compatibility matrix T for sequences of symbols $\sigma\in\{1,\ldots,n\}$ and by a real Fisher potential Φ such that for some $\kappa>0$ one has $\|\Phi\|_\kappa<\infty$, cf. (7.3.2), (without smallness requirement). The new setting, *i.e.* taking into account hard core interactions, extends considerably in the one-dimensional lattice cases that of Section §7.2 and its analysis will be directly applicable to the theory of SRB distributions, which as we have seen in the preceding sections, are precisely Gibbs states with such kind of interactions. Then the following proposition, [CO81], holds.

P7.3.2 **(7.3.2) Proposition:** (Decimation in one dimension)
*The correlations $\mu_\Phi(C^V_{\sigma_V})$ of a one-dimensional spin system are analytic functions of Φ if the interaction is given by a mixing compatibility matrix T for sequences of symbols $\sigma\in\{1,\ldots,n\}$ and by a translation invariant extended nearest neighbor potential Φ (*cf.* definition (7.3.1)) satisfying the condition that $||\,\overline{\Phi}||_\kappa$ is $<\infty$. Furthermore the correlations between events of the form $C^V_{\sigma_V},C^W_{\sigma_W}$ decay at exponential rate.*

Proof: This is almost an immediate consequence of the proof of proposition (7.3.1): we can proceed as in its proof. As soon as $h>\tau>a_0$ where a_0 is the mixing time of the compatibility matrix there will be no compatibility to be fulfilled in fixing the configurations of distinct β-blocks; nor there will be any compatibility to be fulfilled in fixing the configurations of distinct η-blocks: of course there will be compatibility conditions between β and η blocks.

Since no compatibility needs to be checked between the β–block configurations we fix them arbitrarily and perform the summation over the η blocks by using, as in proposition (7.3.1), that the interaction between such blocks is small although not translation invariant. The configurations allowed to the various blocks η are not the same in different blocks (because they must be compatible with the adjacent β–blocks). This is however irrelevant by the remark (3) to proposition (7.2.3). Therefore one gets exactly the same bounds on the potential Ψ between the β–blocks as in (7.3.17),(7.3.18).

The only novelty arises in treating the transfer matrix $M_{\sigma,\sigma'}$. However mixing implies that M^{a_0} has all matrix elements positive. Therefore if h is a multiple of τa_0, which we can assume without loss of generality, we can continue, replacing τ by $a_0\tau$. Since a_0 is a fixed quantity we are back in the situation studied in the previous proof if h is taken a large multiple of τa_0, rather than just of τ. ■

Remarks: The above method can be extended to several cases of $(d+1)$–dimensional spin systems with $d\ge 1$ verifying one among the following properties.
(1) There is a privileged coordinate direction that we call *time* (*e.g.* the $(d+1)$–th dimension) and the potential can have a hard core in the *time*

direction only, *i.e.* the spins σ, σ' which are on nearest neighbor sites that lie on the same line λ parallel to the privileged direction have to verify $T_{\sigma\sigma'} = 1$ $\forall \sigma, \sigma'$ for a given mixing compatibility matrix.
(2) The potential is short ranged and small enough, *i.e.* for ν, κ^{-1} small enough one has

e7.3.23
$$\sup_{\xi\in\mathbb{Z}^{d+1}} \sum_{\xi\in X\subset\mathbb{Z}^{d+1}, |X|>1} \sup_{\underline{\sigma}_X} \nu^{-|X|} |\Phi_X(\underline{\sigma}_X)| \, e^{\kappa\,\delta(X)} < 1. \tag{7.3.23}$$

(3) The extension proceeds by considering a box Λ with sides of size L and, assuming that the direction λ is the direction parallel to the axis $d+1$ of $\mathbb{Z}^{d+1}$, cutting each of the L^d lines parallel to λ into intervals $B_0, H_0, \ldots, H_{\ell-1}, B_\ell$. One then imagines to fix all spins in the intervals of B-type (which can be arbitrarily assigned among the T–compatible strings β of length τ) and to sum over the spins in the H–type intervals. This is a $(d+1)$–dimensional system of weakly coupled spins (as it follows by adapting the proof of the corresponding statement in the proof of proposition (7.3.2) making use of proposition (7.2.2)). The resulting distribution for the β spins is weakly coupled Gibbs state *without hard core* and proposition (7.3.1) (as well as proposition (7.2.2)) applies to it. We discuss below a result that can be obtained, along the above lines, in a special case of relevance for our later applications to dynamical systems; it is however useful to pose a formal definition

D7.3.2 **(7.3.2) Definition:** (Oriented hard core)
Consider the configurations $\underline{\sigma}$ of a $\frac{n}{2}$-spin system on a $(d+1)$–dimensional lattice $\mathbb{Z}^{d+1}$. Consider a mixing matrix $T_{\sigma\sigma'}$ and denote $\{1,\ldots,n\}_T^{\mathbb{Z}^{d+1}}$ the space of the configurations such that $T_{\sigma_x\sigma'_{x'}} = 1$ if the pair of nearest neighbor sites x, x' lie in the direction of the $(d+1)$–axis, i.e. if $x = (\xi, t)$ and $x' = (\xi', t')$ with $|t - t'| = 1$. We call $\{1,\ldots,n\}_T^{\mathbb{Z}^{d+1}}$ the configuration space of a spin system subject to an oriented *or* timelike *hard core. We say that the potential Φ has spatial and temporal ranges $\kappa^{-1}, \kappa_0^{-1}$ if*

e7.3.24
$$\|\,\overline{\Phi}\|_{\kappa,\kappa_0} \stackrel{def}{=} \sup_{x\in\mathbb{Z}^{d+1}} \sum_{x\in X\subset\mathbb{Z}^{d}, |X|>1} \|\Phi_X\| e^{\kappa\,\delta_\perp(X)+\kappa_0\,\delta_\parallel(X)} < \infty, \tag{7.3.24}$$

*where $\|\Phi_X\| = \sup_{\underline{\sigma}_X} |\Phi_X(\underline{\sigma}_X)|$ with the supremum taken over the $\underline{\sigma}_X$ which are T–compatible; $\delta_\perp(X)$ (*spatial tree length *of X) and $\delta_\parallel(X)$ (*temporal tree length *of X) are obtained by considering the sum of the lengths of the projections on the plane $\mathbb{Z}^d$ and, respectively, on the axis orthogonal to it in $\mathbb{Z}^{d+1}$ of the segments constituting the shortest tree linking the points of X.*

Likewise we can define oriented Fisher potentials:

D7.3.3 **(7.3.3) Definition:** (Oriented Fisher potentials)
A oriented hard core potential Φ is a oriented Fisher potential if $\Phi_X \neq 0$

only if the set X consists of a union of intervals located on pairwise distinct time lines.

Remarks: (1) We call "time" the last direction in $\mathbb{Z}^{d+1}$ and "timelike" any line in $\mathbb{Z}^{d+1}$ parallel to the last coordinate axis.
(2) A general oriented hard core potential Φ is often equivalent to a suitable oriented Fisher potential Ψ. Indeed let a generic set $X \subset \mathbb{Z}^{d+1}$ be the union of vertical intervals $\cup_{i,j} J_j^{\xi_i}$, with $i = 1, \ldots, m$ and $j = 1, \ldots, n_{\xi_i}$, lying on distinct time lines $\xi_1, \ldots, \xi_p \in \mathbb{Z}^d$ and ordered so that $J_j^{\xi_i} < J_{j+1}^{\xi_i}$. Let $I^{\xi_1}, \ldots, I^{\xi_p}$ be the *smallest* timelike intervals such that $\cup_j J_j^{\xi_i} \subset I^{\xi_i}$: we say that the $J_j^{\xi_i}$ generate I^{ξ_i} (in other words I^{ξ_i} is the interval delimited by the lowest point in $J_1^{\xi_i}$ and the highest point in $J_{n_{\xi_i}}^{\xi_i}$). Given a $Y = \cup_i I^{\xi_i}$, with I^{ξ_i} timelike intervals lying on distinct time lines, for each compatible confguration $\underline{\sigma}_Y$ we set

e7.3.25
$$\Psi_Y(\underline{\sigma}_Y) = \sum{}^{*} \Phi_X(\underline{\sigma}_X) \tag{7.3.25}$$

with the sum running over all set $X = \cup_{i,j} J_j^{\xi_i}$ such that $J_j^{\xi_i}$ generates I^{ξ_i}. Moreover we set $\Psi_Y = 0$ for all sets Y which are not the union of timelike intervals located on distinct timelike lines. We see that in this way we define a oriented Fisher potential that is equivalent to Φ, provided $\|\Psi\|_{\kappa',\kappa_0'}$ is finite for some κ', κ_0'.
(3) A sufficient condition for the finiteness of (7.3.25) is that the oriented potential verifies

e7.3.26
$$\|\,\overline{\Phi}\|_{\kappa,\kappa_0,\kappa_1} \stackrel{def}{=} \sup_{\xi\in\mathbb{Z}^d} \sum_{\xi\in X\subset\mathbb{Z}^d,\,|X|>1} \|\Phi_X\| e^{\kappa\,\delta_\perp(X)+\kappa_0\,\delta_\parallel(X))} e^{-\kappa_1\,n(X)} < \infty, \tag{7.3.26}$$

where $n(X)$ is the number of distinct vertical intervals whose union is X (*i.e.* $n(X) = \sum_i n_{\xi_i}$). In this case if κ_1 is large enough we get $\|\,\overline{\Psi}\|_{\frac12\kappa,\kappa_0} < \infty$.

The following proposition says that if the spatial range is short enough the system behaves as if it was one dimensional.

P7.3.3 **(7.3.3) Proposition:** (Cluster expansion for oriented hard core systems) *Consider a $\frac{n}{2}$-spin system subjected to a Fisher oriented hard core potential such that there are $\kappa, \kappa_0 > 0$ for which $\|\,\overline{\Phi}\|_{\kappa,\kappa_0} < \infty$. Then given κ_0 there are constants $\overline{\kappa}, b > 0$ such that if $\kappa > \overline{\kappa}$ the correlation functions $\mu_\Phi(C_{\sigma_V}^V)$ are analytic functions of Φ in the region $\|\,\overline{\Phi}\|_{\kappa,\kappa_0} < 1$ and $|\mathrm{Im}\,\Phi_\xi(\sigma)| < b$. Furthermore the Gibbs distribution μ_Φ is exponentially mixing.*

The proof follows closely the hint in remark (3) above and is made easier by the fact that the length κ^{-1} characterizing the diameter decay in (7.3.24) is very short for κ large. Hence the interaction is very small and short ranged for κ large apart from the oriented hard core. See problems [7.3.10] and [7.3.11] for details.

More general cases in which the interaction is not supposed small in the direction of the hard core orientation (*i.e.* in the case considered in remark (2) above) have been studied in [BFG03].

Problems for §7.3

Q7.3.1 **[7.3.1]:** Show that if $\mathcal{C} = (C_1, \ldots, C_k)$ with $\delta(\mathcal{C}) > 1$ the estimate (7.3.13) can be improved as

$$\sup_H \sum_{H\in\mathcal{C},\,\delta(\mathcal{C})>1} \sup_{\underline{\eta}_\mathcal{C}} |\widetilde{\Phi}_{C_1,\ldots,C_k}(\eta_{C_1},\ldots,\eta_{C_k})| e^{\frac12 (k-1)hk} < m\| \,\overline{\Phi}\|_\kappa e^{-\frac{\kappa}{2}h}.$$

(*Hint:* Note that the distance of the extreme points of the sets C_i, C_j is of the order of $(k-2)h + (k-1)\tau$ and $k > 2$ if $\delta(\mathcal{C}) > 1$.)

Q7.3.2 **[7.3.2]:** Let A be a Fisher potential for a spin $\frac{s}{2}$ system on a finite lattice $\Lambda = \{1, \ldots, h\}$ with h sites. Call $a(\beta) = a_\xi(\beta)$ and $a(\beta, \beta') = a_{\{\xi,\xi'\}}(\beta, \beta')$, respectively, the single-spin and the nearest neighbor potentials supposed translationally invariant. Suppose that there is also a potential $a_X(\beta_X)$, possibly non translation invariant, satisfying

$$|a_X(\underline{\beta}_X)| \le \varepsilon e^{-\kappa\delta(X)}$$

for some $\varepsilon, \kappa > 0$. Then the partition function $Z_h(\beta, \beta') \stackrel{def}{=} \sum_{\beta_1,\ldots,\beta_h} e^{-U(\underline{\beta})}$ with $\beta_1 = \beta, \beta_h = \beta'$ can be written as

$$\sum_{\substack{\beta_1,\ldots,\beta_h\\ \beta_1=\beta,\beta_h=\beta'}} e^{-\frac12 a(\beta)} \Big(\prod_{j=1}^{h-1} M_{\beta_j\beta_{j+1}}\Big) e^{-\frac12 a(\beta')} e^{-\sum_{X\subset\Lambda} a_X(\underline{\beta}_X)},$$

where $M_{\beta\beta'} \stackrel{def}{=} e^{-\frac12 a(\beta)} e^{-a(\beta,\beta')} e^{-\frac12 a(\beta')}$ and the last sum is over the intervals $X \subset \Lambda$ with $\delta(X) > 1$.

Q7.3.3 **[7.3.3]:** (*A transfer matrix approach to long range interactions*)
In the context of problem [7.3.2] let $Z_h^0(\beta, \beta')$ be defined as $Z_h(\beta, \beta')$ in problem [7.3.2] with $a_X(\beta_X) \equiv 0$. Check that

$$Z_h^0(\beta,\beta') = \lambda^{h-1} e^{-\frac12 a(\beta)} v(\beta) v(\beta') e^{-\frac12 a(\beta')} \exp\big(-e^{-\kappa_0 h}\Delta(\beta,\beta')\big),$$

where κ_0 is the logarithm of the ratio between the maximal eigenvalue λ of the matrix M and the largest modulus of the others, $|\Delta(\beta, \beta')| \le D_0$ for some $D_0 > 0$, end $v(\beta) > 0$ is the β component of the eigenvector with eigenvalue λ of M (cf. the Perron–Frobenius theorem). (*Hint:* This is just one more application of the transfer matrix, see for instance problem [7.2.13].)

Q7.3.4 **[7.3.4]:** (*Toward polymerization of the problem of checking (7.3.21)*)
In the context of problems [7.3.2] and [7.3.3] check that $Z_h(\beta, \beta')$ can be written as

$$\sum_{\substack{\beta_1,\ldots,\beta_h\\ \beta_1=\beta,\beta_h=\beta'}} e^{-\frac12 a(\beta)} \Big(\prod_{j=1}^{h-1} M_{\beta_j\beta_{j+1}}\Big) e^{-\frac12 a(\beta')} \sum_{p\ge0} \sum_{I_1,\ldots,I_p\subset\Lambda} \prod_{j=1}^{p} \zeta(I_j, \beta_{I_j}),$$

where $I_1, \ldots, I_p$ are p disjoint consecutive intervals. Find conditions on ε, κ such that there are $b, C > 0$ for which

$$|\zeta(I, \beta_I)| \le \varepsilon C\, e^{-b\kappa|I|}.$$

(*Hint:* Note that $\zeta(I,\beta_I)=\sum_{q\ge1}\sum^*_{X_1,\ldots,X_q}\prod(e^{-a_{X_j}(\underline{\beta}_{X_j})}-1)$, where the $*$ means that the $X_1,\ldots,X_q$ are q pairwise distinct intervals with the interval I as union and which are "chain connected" (*i.e.* for every pair of intervals there is a chain of pairwise intersecting intervals among the $X_1,\ldots,X_q$ starting with the first interval and ending with the last). Then apply the method used to obtain (7.2.9) or (7.2.16). If $B=\sum_{0\in X\in\mathbb{Z}}e^{-\kappa\delta(X)}$ a sufficient condition will be for instance $\max_{z\in\mathbb{Z}}Bze^{\varepsilon Bz}e^{-\frac14\kappa z}\le C$ for $b=1/16$, say.)

Q7.3.5 **[7.3.5]:** In the context of the above problems let $I_1,\ldots,I_p$ be p disjoint consecutive intervals in Λ. The intervals can be specified by the sequence of integers $r_1,i_1,r_2,i_2,\ldots,r_p,i_p,r_{p+1}$ giving the "lengths" of the "empty" sites (r_j) and of the
N7.3.2 "occupied sites ($i_j\equiv|I_j|$).[3] Denote β^-_j,β^+_j the spins located at the initial and final sites of I_j and let $\beta^+_0=\beta,\beta^-_{p+1}=\beta'$. Check that the function $Z_h(\beta,\beta')$can be written as

$$\sum_{\substack{p,I_1,\ldots,I_p\\ \beta_{I_1},\ldots,\beta_{I_p}}}\Big(\prod_{j=1}^{p-1}Z^0_{r_j+1}(\beta^+_j,\beta^-_{j+1})\cdot Z^0_{i_j-1}(\beta^-_j,\beta^+_j)\Big)$$

$$Z^0_{r_1}(\beta,\beta^-_1)Z^0_{r_{p+1}}(\beta^+_p,\beta')\Big(\prod_{j=1}^{p}\frac{\zeta(I_j,\beta_{I_j})\prod_{\xi,\xi+1\in I_j}M_{\beta_\xi,\beta_{\xi+1}}}{Z^0_{i_j-1}(\beta^-_j,\beta^+_j)}\Big),$$

with the natural interpretation of the extreme cases (*e.g.* $p=0$ corresponds to no sum over β's and gives $r_1=h$ and a term equal to $Z^0_h(\beta,\beta')$, *etc*). (*Hint:* This is read off the first expression for $Z_h(\beta,\beta')$ in problem [7.3.4].)

Q7.3.6 **[7.3.6]:** In the context of the above problems and making use of the result of problem [7.3.3] check that setting $w(\beta)=e^{\frac12a(\beta)}v(\beta)$, cf. problem [7.3.3], $Z_h(\beta,\beta')$ is

$$Z_h(\beta,\beta')=\lambda^{h-1}w(\beta)w(\beta')\sum_{\substack{p;\{\beta^\pm_j\}_{j=1}^{p+1}\\ I_1,\ldots,I_p}}\Big(\prod_{j=1}^{p}w(\beta^+_j)^2w(\beta^-_j)^2\Big)\cdot$$

$$\cdot\Big(\prod_{j=1}^{p+1}\exp(-e^{-\kappa_0r_j}\Delta(\beta^+_{j-1},\beta^-_j)\Big)\Big(\prod_{j=1}^{p}\exp(-e^{-\kappa_0|I_j|}\Delta(\beta^-_j,\beta^+_j))\Big)\cdot$$

$$\cdot\prod_{j=1}^{p}\frac{\zeta(I_j,\beta_{I_j})\prod_{\xi,\xi+1\in I_j}M_{\beta_\xi,\beta_{\xi+1}}}{Z^0_{i_1-1}(\beta^-,\beta^+_j)},$$

where $r_1,i_1=|I_1|,\ldots$ are the successive lengths of empty and filled lattice points (cf. problem [7.3.5]). (*Hint:* This is simply a rewriting of the expression in problem [7.3.5]).

Q7.3.7 **[7.3.7]:** In the context of the above problems let L be such that $e^{-\kappa_0L}D_0<\varepsilon$. Given a configuration of successive intervals $I_1,\ldots,I_q$ in Λ we say that a subsequence of $k+1$ consecutive intervals $I_x,I_{x+1},\ldots,I_{x+k}$ is L*–connected* if the distance between consecutive intervals is $\le L$ and if it is maximal (given $I_1,\ldots,I_q$) with the latter property. Given such a subsequence we call J the interval between the first point of I_x and the last point of I_{x+k} *enlarged* by adding $L/2$ points to the right and to the left or (if there are not enough such points) as many as possible. In other words J is the part of the union of the intervals I_{x+j} enlarged to the right and to the left by $L/2$ which is inside Λ: we say that $I_x,I_{x+1},\ldots,I_{x+k}$ "generate" J. Define

---

[3] One must have $r_2,\ldots,r_p>0$ while $r_1,r_{p+1}\ge0$ and $i_j>0$ unless $p=0$, in which case $r_1=h$.

$$\zeta(J) = \sum_{\substack{q;I_1,\ldots,I_q \text{ generate } J\\ \beta_{I_1},\ldots,\beta_{I_q}}} \prod_{i=1}^{q-1} \exp(-e^{-\kappa_0 r_j}\Delta(\beta_j^+,\beta^- j+1))\cdot$$
$$\cdot \prod_{i=1}^{q} \frac{\zeta(I_j,\beta_{I_j})\prod_{\xi,\xi+1\in I_j} M_{\beta_\xi,\beta_{\xi+1}}}{Z^0_{i_1-1}(\beta^-,\beta_j^+)},$$

and check that for a suitable $b_1 > 0$ one has

$$|\zeta(J)| \le e^{\beta_1\kappa|J|}\frac{\varepsilon C}{1-\varepsilon C e^{\frac18 L}L} \stackrel{def}{=} e^{\beta_1\kappa|J|}C_0\varepsilon$$

if $\varepsilon C e^{\frac18 L}L < 1/2$. (*Hint:* Proceed as in the hint to problem [7.3.4].)

Q7.3.8 **[7.3.8]:** In the context of the above problems show that if ε is small enough and k is large enough proposition (7.1.1) can be applied to deduce that (with the notation of the polymers of Section §7.1

$$Z_h(\beta,\beta') = \lambda^{h-1}w(\beta)w(\beta')\exp\Big(\sum_X \varphi^T(X)\zeta(X)\Big),$$

where X is a polymer configuration built with polymers J with activity $\zeta(J)$ defined in problem [7.3.7]. Note that $\zeta(X)$ is, therefore, independent of β unless X contains the first point of Λ and independent of β unless one of the polymers of X contains the last point of Λ. (*Hint:* By problem [7.3.7] one has

$$Z_h(\beta,\beta') = \lambda^{h-1}w(\beta)w(\beta')\sum_{s\ge0}\sum_{\substack{J_1,\ldots,J_s\\ J_i\cap J_j=\emptyset}}\prod_{j=1}^{s}\zeta(J_j),$$

and $\zeta(J)$ verifies (7.1.16) with $b_0 = C_0$, $\nu_0 = e^{-\frac12 b_1\kappa}$, $\kappa_0 = \frac12 b_1\kappa$.)

Q7.3.9 **[7.3.9]:** Find a proof of the result in problem [7.3.8] simpler than performing the analysis in problems [7.3.2] through [7.3.8]. (*Hint:* It should be possible.)

Q7.3.10 **[7.3.10]:** (*Decimation in higher dimension. I*)
Given $\kappa_0 > 0$ and a potential Φ for a system of oriented hard cores (see definition (7.3.2)) such that $\|\Phi\|_{\kappa,\kappa_0} < \infty$ for some $\kappa > 0$. Let τ, h be two integers larger than the mixing time of the compatibility matrix T. Let $\Lambda = Q\times I$ where $Q\subset\mathbb{R}^d$ is a cube of side L and I is the interval $[0,(\tau+h)n+\tau]$: we regard the parallelepiped Λ as a "horizontal" strip consisting of $|Q|$ "vertical" intervals of size $(\tau+h)n+\tau$. Divide the space–time region $\Lambda = Q\times I$ into $(n+1)$ "horizontal" strips made of $|Q|$ "vertical" intervals of size τ alternating with n "horizontal" strips made of $|Q|$ vertical intervals of size h. Call $B_1, B_2,\ldots,B_{|Q|(n+1)}$ and $H_1,\ldots,H_{|Q|n}$ the vertical intervals of length τ or h into which each strip is divided; show that by holding fixed the block spin configurations $\underline\beta_{B_j}$ the *decimated energy* $U^{\rm dec}(\underline\beta) \stackrel{def}{=} \log\sum_{\underline\eta} e^{-U_\Lambda(\underline\beta,\underline\eta)}$ can be studied via the technique of the proof of proposition (7.3.1) (and proposition (7.2.3), corollary (7.2.3)). By choosing τ, h suitably large ($e^{-\kappa_0\tau} < e^{-\kappa}$), one finds that $U^{\rm dec}(\underline\beta)$ is expressed as an interaction without hard cores between the spins $\underline\beta$ which has the property that

$$\sup\sum_{\substack{B_1,\ldots,B_q\\ \delta^*(B_1,\ldots,B_q)>1}} |\Phi^{\rm dec}_{B_1,\ldots,B_q}(\underline\beta_{B_1},\ldots,\underline\beta_{B_q})|e^{\frac12\kappa\delta^*(B_1,\ldots,B_q)} < 1,$$

where $\delta^*(B_1,\ldots,B_q)$ denotes the tree distance of the centers of the sets $B_1,\ldots,B_q$ with the vertical components of the distances measured in units of $(\tau+h)$. (*Hint:* Note that $U_\Lambda(\underline\beta,\underline\eta)$ can be written as a sum of $\underline\beta$–only dependent terms $\sum_i\sum_{X_i\subset B_i}\Phi_{X_i}(\underline\beta_{X_i})$ plus $\sum_{j_1,\ldots,j_q}\Phi_{X_{j_1},\ldots,X_{j_q}}(\underline\beta_{X_{j_1}},\ldots,\underline\beta_{X_{j_q}})$ plus

$$\sum_i\sum_{X_i\subset H_i}\Phi_{X_i}(\underline\eta_{X_i})+\sum^*_{\substack{q>1,H_{i_1},\ldots,H_{i_q}\\ X_i\subset H_i}}\Phi_{X_{i_1},\ldots,X_{i_q}}(\underline\eta_{X_{i_1}},\ldots,\underline\eta_{X_{i_q}}),$$

where the * reminds that one has $X_{i_j}\cap H_{i_j}\ne\emptyset$ for *all* $j=1,\ldots,q$. The sets X_{i_j} possibly intersect also some of the B's so that the last expression does depend also on the $\underline\beta$'s. The spins $\underline\eta_H$ are "high spins" and their interaction is not translation invariant (for instance because the $\underline\beta$ variables have non translation invariant values). Calling their interaction Ψ we see that its size is such that

$$\sum_{q>1,H_1,\ldots,H_q}\|\Psi_{H_1,\ldots,H_q}\|e^{+\frac12\kappa\delta^*(H_1,\ldots,H_q)}<1,$$

so that we can apply proposition (7.2.3), corollary (7.2.3) and the result of problem [7.2.10] if κ is large enough.)

Q7.3.11 **[7.3.11]:** (*Decimation in higher dimension. II*)
In the context of problem [7.3.10] note that the potential $\Phi^{\rm dec}$ has no hard core. Furthermore the components with $\delta^*(B,B')=1$ (hence $q=2$) of the potential consist of a contribution that can be bounded $\sup\sum_{B_1,B_2,\delta^*(B_1,B_2)=1}|\Phi^{dec}_{B_1,B_2}(\underline\beta_{B_1},\underline\beta_{B_2})|e^{\frac12\kappa}<1$ plus a component that can be quite large equal to $\log Z_h(\underline\beta_{B_1},\underline\beta_{B_2})$ defined in (7.3.9). The latter is representable as a sum of a nearest neighbor potential which can be made as small as wished if h is large enough plus a one spin component, as studied following (7.3.14). Hence we can apply corollary (7.2.3): show that this yields a proof of proposition (7.3.2).)

Bibliographical note for §7.3

The analysis in this section is a technical extension of the simple problems proposed in Section §7.2.

The renormalization by decimation method used to study one dimensional Gibbs states with short range forces, and their analyticity, followed here has been introduced in [CO81], where a new proof of Dobrushin's optimal analyticity result for interactions with polynomial decay has been obtained.

§7.4 Absence of phase transitions: more criteria

Availability of simple criteria for uniqueness of the Gibbs state associated with a potential is often important. We have so far met several such criteria, *e.g.* the general result for one dimensional systems in proposition (5.2.1) and various criteria in propositions (7.2.1), (7.2.2), (7.2.3) and (7.3.1), and corollaries (7.2.1) and (7.2.2).

However there are applications of the theory, even to dynamical systems of Anosov type, in which the mentioned criteria do not apply because their assumptions are, or appear to be, too restrictive. We shall meet an important example in problem [10.4.1] and here we mention a simple criterion which turns out useful in its study.

P7.4.1 **(7.4.1) Proposition:** (Diameter decay of potential and no phase transitions conditions)
Let Φ be a potential for a spin system without hard core on the d–dimensional lattice $\mathbb{Z}^d$. Assume that $\Phi_X(\underline{\sigma}_X)$ can be different from 0 only if X is connected by nearest neighbors and define for $\kappa > 0$

e7.4.1
$$\|\Phi\|_\kappa^{\rm diam} = \sum_{0\in X} \|\Phi_X\| e^{\kappa D(X)}, \tag{7.4.1}$$

where $D(X) = \mathrm{diam}(X)$ is the diameter of X. Given $\kappa > 0$ if $\|\Phi\|_\kappa^{\rm diam}$ is small enough the Gibbs distribution is unique, depends smoothly on Φ and the correlations of local observables decay exponentially.

Proof: Suppose that the system is confined in a box Λ. Let F, G be two local observables, $F_A(\underline{\sigma}) = f(\underline{\sigma}_A)$ and $G_B(\underline{\sigma}) = g(\underline{\sigma}_B)$, where A, B are two finite sets in Λ. We can write the correlation $\mu(F_A G_B) - \mu(F_A)\mu(G_B)$ as

e7.4.2
$$\frac{1}{2}\frac{\sum_{\underline{\sigma},\underline{\tau}}(f(\underline{\sigma}_A) - f(\underline{\tau}_A))(g(\underline{\sigma}_B) - g(\underline{\tau}_B))e^{-\sum_X(\overline{\Phi}_X(\underline{\sigma}_X)+\overline{\Phi}_X(\underline{\tau}_X))}}{\sum_{\underline{\sigma},\underline{\tau}} e^{-\sum_X(\overline{\Phi}_X(\underline{\sigma}_X)+\overline{\Phi}_X(\underline{\tau}_X))}}, \tag{7.4.2}$$

where $\overline{\Phi}_X(\underline{\sigma}_X) \stackrel{def}{=} \Phi_X(\underline{\sigma}_X) - \max_{\underline{\sigma}_X} \Phi_X(\underline{\sigma}_X) \le 0$. The numerator can be expanded "as usual" by writing

e7.4.3
$$e^{-\sum_X(\overline{\Phi}_X(\underline{\sigma}_X)+\overline{\Phi}_X(\underline{\tau}_X))} = \prod_X \big(\rho(X, \underline{\sigma}_X, \underline{\tau}_X) + 1\big), \tag{7.4.3}$$

where $\rho(X, \underline{\sigma}_X, \underline{\tau}_X) \stackrel{def}{=} \big(e^{-(\overline{\Phi}_X(\underline{\sigma}_X)+\overline{\Phi}_X(\underline{\tau}_X))} - 1\big)$ and

e7.4.4
$$0 \le \rho(X, \underline{\sigma}_X, \underline{\tau}_X) \le e^{2\|\Phi\|_\kappa^{\rm diam}} 2\|\Phi\|_\kappa^{\rm diam} e^{-\kappa D(X)}. \tag{7.4.4}$$

By using (7.4.3) we get for the numerator of (7.4.2)

e7.4.5
$$\frac{1}{2}\sum_{\underline{\sigma},\underline{\tau}}(f(\underline{\sigma}_A) - f(\underline{\tau}_A))(g(\underline{\sigma}_B) - g(\underline{\tau}_B)) \sum_{n, Y_1,\ldots,Y_n} \prod_{i=1}^n \rho(Y_i, \underline{\sigma}_{Y_i}, \underline{\tau}_{Y_i}). \tag{7.4.5}$$

In the above sum we distinguish the n–ples $Y_1, \ldots, Y_n$ which *do not contain* a chain $C = (Y_1^*, \ldots, Y_p^*)$ of sets such that $Y_k^* \cap Y_{k+1}^* \ne \emptyset$ for $k = 1, \ldots, n-1$ and $Y_1^* \cap A \ne \emptyset$, $Y_n^* \cap B \ne \emptyset$ and the others.

The total contribution of the first terms is 0 by their odd symmetry $\underline{\sigma} \longleftrightarrow \underline{\tau}$: exchanging the configurations $\underline{\sigma}, \underline{\tau}$ *inside* the union of the sets Y which are parts of chains $Y_1^*, \ldots, Y_p^*$ intersecting A and not B changes the sign of (7.4.5) (note that such an operation of partial interchange of $\underline{\sigma}$ and $\underline{\tau}$ is only possible in general if there are no hard cores). The contribution of the other terms is bounded simply by summing over all chains $C = (Y_1^*, \ldots, Y_p^*)$ connecting A and B

e7.4.6
$$4\|F_A\|\,\|G_B\| \sum_C 2^p e^{2\|\Phi\|_\kappa^{\rm diam} p} (\|\Phi\|_\kappa^{\rm diam})^p \prod_{i=1}^p e^{-\kappa D(Y_i^*)}\, Z_\Lambda^2, \qquad (7.4.6)$$

where (7.4.4) has been used together with the remark that the sum over the other sets Y adds up at most to Z_Λ.

Hence, if $D(A,B)$ is the distance between A, B,

e7.4.7
$$\begin{aligned}&\sum_C 2^p e^{2\|\Phi\|_\kappa^{\rm diam} p} (\|\Phi\|_\kappa^{\rm diam})^p \prod_{i=1}^p e^{-\kappa D(Y_i^*)} \le \\ &\le e^{-\frac13\kappa D(A,B)} \sum_C (2 e^{2\|\Phi\|_\kappa^{\rm diam}} \|\Phi\|_\kappa^{\rm diam})^p \prod_{i=1}^p e^{-\frac23\kappa D(Y_i^*)}.\end{aligned} \qquad (7.4.7)$$

The sum over the sets Y which intersect a given set Y' of $e^{-\frac13\kappa D(Y)}$, *i.e.* $\sum_{Y\cap Y'\ne\emptyset} e^{-\frac13\kappa D(Y)}$, is bounded by $D(Y')^d C_0$ for some constant C_0 (depending only on κ). Therefore, if $C_1 \stackrel{def}{=} \max_D D^d e^{-\frac13\kappa D}$,

e7.4.8
$$\begin{aligned}S_p \stackrel{def}{=} &\sum_{Y_1^*,\ldots,Y_p^*} \prod_{i=1}^p e^{-\frac23\kappa D(Y_i^*)} \le C_0 \sum_{Y_1^*,\ldots,Y_{p-1}^*} D(Y_{p-1}^*)^d e^{-\frac13\kappa D(Y_{p-1}^*)}\cdot \\ &\cdot e^{-\frac13\kappa D(Y_{p-1}^*)} \prod_{i=1}^{p-2} e^{-\frac23\kappa D(Y_i^*)} \le C_0^p C_1^{p-1},\end{aligned} \qquad (7.4.8)$$

and from (7.4.7) we conclude that for some $C > 0$ one has

e7.4.9
$$|\mu(F_A G_B) - \mu(F_A)\mu(G_B)| \le C\|F_A\|\,\|G_B\|\,\|\Phi\|_\kappa^{\rm diam}\, e^{-\frac13\kappa D(A,B)}, \qquad (7.4.9)$$

provided $\|\Phi\|_\kappa^{\rm diam}$ is small enough (the factor Z_Λ in (7.4.6) is canceled out by the denominator).

This also shows that the Gibbs distribution is differentiable at Φ because if $\mu_{\Phi+\lambda\Psi,\Lambda}$ denotes the finite volume Gibbs distribution with potential $\Phi+\lambda\Psi$ then, if we chose $\Psi \in \mathcal{B}_0$,

e7.4.10
$$\begin{aligned}&\Big|\frac{d}{d\lambda}\mu_{\Phi+\lambda\Psi,\Lambda}(F_A)\Big|_{\lambda=0}\Big| = \Big|\sum_X (\mu_{\Phi,\Lambda}(F_A\Psi_X) - \mu_{\Phi,\Lambda}(F_A)\mu_{\Phi,\Lambda}(\Psi_X))\Big| \le \\ &\le C\|F_A\|\|\Psi\|_\kappa^{\rm diam}\|\Phi\|_\kappa^{\rm diam} \sum_X e^{-\kappa\big(D(X)+\frac13 D(A,X)\big)} \le E,\end{aligned} \qquad (7.4.10)$$

where E is a Λ–independent constant and the sum is finite as $\Psi \in \mathcal{B}_0$. This means that the tangent plane to $P_\Lambda(\Phi)$ varies continuously with Φ uniformly in Λ so that the Gibbs state is unique and it also follows that

e7.4.11
$$\frac{d}{d\lambda}\mu_{\Phi+\lambda\Psi}(F_A)\bigg|_{\lambda=0} = \sum_X (\mu(F_A\Psi_X) - \mu(F_A)\mu(\Psi_X)), \tag{7.4.11}$$

for $\|\Psi\|_\kappa^{\rm diam} < \infty$.

Likewise one can express the second derivative (and the higher ones): the above exponential decay can be used to check that also the second (and higher) derivatives are uniformly bounded so that μ depends in an infinitely smooth way on Φ as long as Φ varies keeping $\|\Phi\|_\kappa^{\rm diam}$ small enough, for a given $\kappa > 0$. The check requires dividing the space into regions and it is somewhat laborious. For instance in the case of the second derivative the regions to consider are $D(A,Y), D(X,Y) > R$, $D(X,Y) < R$ and $D(A,Y) < R$ for $R > 0$ large compared to the diameters of X, Y, A. Hence one first considers the sum over X, Y with diameter not exceeding R and checks that the result is finite. Then one proceeds to relax the restriction on the diameters by taking advantage of the exponential decay of Ψ_X, Ψ_Y with the diameters. We omit further details. ∎

Remarks: (1) The analysis requires absence of hard cores. Otherwise exchanging the configurations $\underline\sigma, \underline\tau$ inside the sets of a chain that intersect A but not B may lead to incompatible spin configuration thus ruining the cancellation mechanism.

(2) Uniqueness implies that the tangent plane varies with continuity along paths in the space of the potentials with the norm (7.4.1) small enough: hence the Gibbs distribution is differentiable in this region.

An important uniqueness criterion is the following, see [Si93].

P7.4.2 **(7.4.2) Proposition:** (Dobrushin's uniqueness criterion)
Let Φ be a potential for a spin system without hard cores in $\mathbb{Z}^d$. If

e7.4.12
$$\|\Phi\|_1 \stackrel{def}{=} \sum_{X \ni \underline 0} (|X| - 1)\, \|\Phi_X\| \stackrel{def}{=} \alpha < 1, \tag{7.4.12}$$

where $\|\Phi_X\| = \sup_{\underline\sigma_X} |\Phi_X(\underline\sigma_X)|$, then the Gibbs distribution with potential Φ is unique.

Proof: Let η be a spin configuration and we denote $\eta_{/i}$ the spin configuration over the sites of $\mathbb{Z}^d$ which are different from i. In this way $\eta_{/i}\sigma_i$ will denote the configuration obtained by replacing η_i with σ_i. Likewise we define $\eta_{/X}$ the configuration η over the sites $\mathbb{Z}^d/X$ for any set X, and $\eta_{/X}\sigma_X$ will denote the configuration obtained from η by replacing the spins of η with σ_X in X.

We assume that the sites of $\mathbb{Z}^d$ have been labeled $i = 0, 1, 2\ldots$ in such a way that the set $Q_n = \cup_{i=0}^n \{i\}$ eventually contains any prefixed cube K centered at the origin; in particular we denote with $i = 0$ the origin.

It is easy to compare the conditional probabilities $m(\sigma_i|\eta_{/ij}\eta'_j)$ and $m(\sigma_i|\eta_{/ij}\eta'_j)$. In fact let D be

e7.4.13
$$\begin{aligned} D &= \Big|\sum_{\sigma_i}(m(\sigma_i|\eta_{/ij}\sigma') - m(\sigma_i|\eta_{/ij}\sigma''))f(\sigma_i)\Big| = \\ &= \Big|\int_0^1 dt\,\sum_{\sigma}\frac{d}{dt}m_t(\sigma)f(\sigma)\Big|, \end{aligned} \tag{7.4.13}$$

where m_t is the probability distribution

e7.4.14
$$m_t(\sigma) = Z_t^{-1}e^{-\sum_{X\ni i,j}\big(\Phi_X(\sigma\eta_{/ij}\sigma') + t(\Phi_X(\sigma\eta_{/ij}\sigma'') - \Phi_X(\sigma\eta_{/ij}\sigma'))\big)}, \tag{7.4.14}$$

with Z_t being the normalization factor equal to the sum over σ of the exponential in (7.4.14).

We call $d(\sigma) \stackrel{def}{=} \sum_{X\ni ij}(\Phi_X(\sigma\eta_{/ij}\sigma'') - \Phi_X(\sigma\eta_{/ij}\sigma'))$ and denote with $\langle\cdot\rangle_t$ the average with the probability m_t. Then $\|d\| \equiv \max_\sigma |d(\sigma)| \le 2\sum_{X\ni i,j}\|\Phi_X\| \stackrel{def}{=} 2\rho_{ij}$; note that we can define $\rho_{ii} \equiv 0$, which is consistent with the definition of $d(\sigma)$. If $\|f\|$ denotes the maximum of $|f(\sigma)|$ then one has

e7.4.15
$$\begin{aligned} D &\le \int_0^1 dt(\langle f\,d\rangle_t - \langle f\rangle_t\langle d\rangle_t) \le \int_0^1 dt\langle f\,(d - \langle d\rangle_t)\rangle_t \le \\ &\le \|f\|\int_0^1 dt\langle (d-\langle d\rangle_t)^2\rangle_t^{\frac12} \le \|f\|\,\|d\| \le 2\|f\|\rho_{ij}. \end{aligned} \tag{7.4.15}$$

Therefore by the arbitrariness of f one has $\sum_\sigma |m(\sigma|\eta_{/ij}\sigma') - m(\sigma|\eta_{/ij}\sigma'')| \le 2\rho_{ij}$. Define

e7.4.16
$$(\tau_j f)(\eta) \stackrel{def}{=} \sum_\sigma m(\sigma|\eta_{/j})f(\sigma), \tag{7.4.16}$$

and note that the Gibbs distributions μ with potential Φ verify the DLR equations, *i.e.* $\mu(\tau_j f) \equiv \mu(f)$. Of course $(\tau_j F)(\eta)$ depends only on $\eta_{/j}$, as the spin on the site j is summed over.

Let F be a cylindrical function depending only on the spins in a finite cube K: then if n is such that $Q_n \supset K$ one has $(\tau_n F)(\eta) \equiv F(\eta)$. To estimate the dependence of F on the spins with large label we introduce

e7.4.17
$$\delta_i(F) = \max_{\sigma,\sigma',\eta}|F(\sigma\eta_{/i}) - F(\sigma'\eta_{/i})|, \qquad \Delta(F) = \sum_{i=0}^\infty \delta_i(F), \tag{7.4.17}$$

and call $\delta_i(F)$ and $\Delta(F)$ the variation of F at site i and, respectively, the variation of F. Then one has $\delta_i(\tau_i F) = 0$, while, for $i \ne j$,

$$\delta_i(\tau_j F) = \max_{\sigma'_i,\sigma''_i,\eta}$$

e7.4.18

$$\Big|\sum_{\sigma_j}\big(F(\sigma_j\eta_{/ij}\sigma'_i)m(\sigma_j|\eta_{/ij}\sigma'_i)-F(\sigma_j\eta_{/ij}\sigma''_i)m(\sigma_j|\eta_{/ij}\sigma''_i)\big)\Big|\le$$
$$\le\delta_i(F)+\max_{\sigma'_i,\sigma''_i,\eta}\Big|\sum_{\sigma_j}[F(\sigma_j\eta_{/ij}\sigma'_i)]\,[m(\sigma_j|\eta_{/ij}\sigma'_i)-m(\sigma_j|\eta_{/ij}\sigma''_i)]\Big|\stackrel{def}{=}$$
$$\stackrel{def}{=}\delta_i(F)+\max\Big|\sum_{\sigma_j}[G(\sigma_j)]\,[m(\sigma_j)]\Big|,\tag{7.4.18}$$

where m and G are the two functions in square brackets in the third line, and $\sum_\sigma m(\sigma)=0$, while by (7.4.15) one has $\sum_\sigma|m(\sigma)|\le 2\rho_{ij}$. On the other hand $\sum_\sigma G(\sigma)m(\sigma)\equiv\sum_\sigma(G(\sigma)-a)m(\sigma)$ for all a so that choosing a to be half way between the maximum and the minimum of $G(\sigma)$ we see that $|\sum_{\sigma_j}G(\sigma_j)m(\sigma_j)|\le\delta G\rho_{ij}$, where δG is the width of the range of the function G; since G varies because the spin σ_j varies one has $\delta G\le\delta_j(F)$ and this implies

e7.4.19

$$\delta_i(\tau_jF)\le\delta_i(F)+\rho_{ij}\delta_j(F).\tag{7.4.19}$$

Therefore noting that $(\tau_jF)(\eta)$ is independent of η_j and setting $\alpha\stackrel{def}{=}\sup_j\sum_{i\ne j}\rho_{ij}=\|\Phi\|_1$ we see that one has

e7.4.20

$$\begin{aligned}\Delta(\tau_0\ldots\tau_nF)\le&\sum_{i=1}^\infty\delta_i(\tau_1\ldots\tau_nF)+\,\alpha\,\delta_0(\tau_1\ldots\tau_nF)\le\\
\le&\sum_{i=2}^\infty\delta_i(\tau_2\ldots\tau_nF)+\sum_{i=2}^\infty\rho_{i1}\delta_1(\tau_2\ldots\tau_nF)+\\
&+\,\alpha\,\delta_0(\tau_2\ldots\tau_nF)+\alpha\,\rho_{01}\delta_1(\tau_2\ldots\tau_nF)\le\\
\le&\sum_{i=2}^\infty\delta_i(\tau_2\ldots\tau_nF)+\,\alpha\,\sum_{i=0}^1\delta_i(\tau_2\ldots\tau_nF)\le\\
\le&\sum_{i=n+1}^\infty\delta_i(F)+\,\alpha\,\sum_{i=0}^n\delta_i(F)\xrightarrow[n\to\infty]{}\alpha\,\Delta(F).\end{aligned}\tag{7.4.20}$$

The limit as $n\to\infty$ of $\tau_0\tau_1\ldots\tau_nF$ is defined for all functions F which are cylindrical (note that if n is large $\tau_nF\equiv F$), hence, by density, for all continuous functions on $\{1,\ldots,n\}^{\mathbb{Z}^d}$: we shall denote it by $TF\stackrel{def}{=}\lim_{n\to\infty}\tau_0\tau_1\ldots\tau_nF$. Then (7.4.20) proves that the variation of TF is smaller by a factor of $\alpha=\|\Phi\|_1<1$ than the variation of F: $\Delta(TF)\le\alpha\,\Delta(F)$ and $\Delta(T^kF)\le\alpha^k\Delta(F)\xrightarrow[k\to\infty]{}0$. Hence T^kF tends to a constant which is $\mu(F)$ if μ is a Gibbs distribution with potential Φ (*i.e.* $\mu(\tau_jF)\equiv\mu(F)$, so that $\mu(T^kF)\equiv\mu(F)$ for all k). Hence the value of $\mu(F)$ is uniquely determined by the conditional probabilities associated with the potential Φ, *i.e.* the Gibbs state is unique. ∎

Remark: Convexity of the function $P(\Phi)$ and uniqueness of its tangent plane imply that the tangent plane varies with continuity along paths in the region $\mathcal{B}_1$ of the potentials on which $\sum_{X\ni 0}(|X|-1)\,\|\Phi_X\| < 1$: hence the Gibbs distribution is differentiable in this region. In fact in [Gr81] it is shown that, very remarkably, under the same conditions the pressure $P(\Phi)$ is C^2 (twice continuously differentiable for $\Phi \in \mathcal{B}_1$) and one has an explicit bound on the size of its second derivative. And under the same condition $\|\Phi\|_1 < 1$ the further assumption that $\sum_{X\ni 0}|X|^{k-1}\,\|\Phi_X\| < \infty$ implies that $P(\Phi)$ is of class C^k in $\mathcal{B}_1$, [Pr83].

A review of *many other* criteria of uniqueness (and smoothness) that are known can be found, together with the related proofs and many comments, in [Si93].

Bibliographical note to §7.4

The above is only one more example of uniqueness criteria. A further important uniqueness criterion (*complete analyticity criterion* and *method of cleanings*) can be found in [DS87]; we do not discuss it here, in spite of its originality and importance in Statistical Mechanics, because we shall not have occasion to employ it in our dynamical systems studies. It is a criterion which is closely related to the cluster expansion methods discussed so far, cf. [Ol88]. On the probabilistic side the method of *cleanings* of [DS87] is related to the method of *shift of conditionings* introduced in discussing Statistical Mechanics aspects of field theory in [BCGNOPS78], [Ga80], [BGN80]. The above presentation follows closely [Si93].

§7.5 Phase transitions

The cases in which $G(\Phi)$ consists of more than one point are very interesting: such cases do not seem to arise in the theory of smooth dynamical systems because smooth dynamical systems generate one-dimensional Gibbs distributions with potential Φ which decays exponentially, as discussed in the previous sections. The situation is somewhat different when one considers systems which are *extended*, like lattices of coupled maps: in this case it is far less clear that the symbolic dynamics associated with such systems does not correspond to a spin system with a potential which admits several Gibbs distributions.

An example of an extended dynamical system that admits a symbolic dynamics with more that one Gibbs state was given in [GMK00]. The system consists of a 2-dimensional lattice of coupled expansive maps on $[0,1]$, like those studied in section §5.4. The key point is to use a coupling that keeps invariant the Markov partition and allows an easy computation of the corresponding potential Φ. In this way it is possible to build explicitly a dynamical system whose potential Φ is among the classical examples known to admit more than one Gibbs state. The problem with such examples with respect to the theory presented in this book is that the resulting dynamical systems are quite singular, indeed not even continuous. No example of such a phenomenon has been found for smooth dynamical system.

In this section we want to discuss some cases of potentials for one-dimensional lattice spin systems which admit several Gibbs distributions. It is in fact of interest to note that, in one-dimensional systems, if one considers potentials that decay slowly at infinite distance then one gets potentials Φ that admit several Gibbs distributions. Such potentials can, perhaps, arise in systems which, although not hyperbolic, still admit a symbolic dynamic description: one typically thinks to maps with a finite number of fixed or periodic points of marginal stability, *i.e.* with a Lyapunov coefficient which has modulus 1 (recall that the modulus of derivative of the map enters into the description of the SRB distributions and it logarithm provides an estimate of the decay of the potential in the symbolic dynamics description, cf. Section §5.4), and this makes them interesting in dynamics.

To find such a Φ it is sufficient, by the analysis of Section §6.1 (cf. proposition (6.1.2) and the remark following it) to find points $\Phi \in \mathcal{B}$, where $\mathcal{B}$ is the space of the potentials on $\{0,\ldots,n\}^{\mathbb{Z}}$, in which the tangent plane to the graph of P is not unique, cf. definition (5.1.1).

Since P is convex, cf. proposition (6.1.1), it is clear that a way to find such points is to consider lines, say straight lines, $h \to \Phi(h)$, $h \in \mathbb{R}$, in $\mathcal{B}$ and to find a value h_0 of h in which the function $h \to P(\Phi(h))$ is not differentiable: in such a point $\Phi(h_0)$ the plane tangent to the graph of P will not be unique and hence there will be at least two Gibbs states, cf. Section §6.1.

A second method is to find a potential $\Phi \in \mathcal{B}$ that admits a Gibbs state that is not ergodic: it is clear from the ergodic decomposition theorem for Gibbs states (cf. corollary (6.1.2)) that in such cases $G(\Phi)$ contains at least two points or, as one says in Physics, two *pure phases*.

One of the interesting problems is studying $G_e(\Phi)$, as its elements have the physical interpretation of pure phases, and characterizing in some way its elements. This kind of problem is called the *phase transitions problem* for Gibbs states.

The reason of the name "transition" lies in the fact that the Gibbs states are studied, usually, as Φ varies on a one parameter curve $h \to \Phi(h)$ or on a two parameters surface $(h,\beta) \to \Phi(h,\beta)$ (which are often just a line or, respectively, a plane in B). In the interesting situations for Physics the *phase rule* holds: for instance in the plane (h,β) the values of h and β to which there corresponds a unique Gibbs state are simply all points *except* those lying on certain regular curves where $G_e(\Phi)$ consists of two or more elements. Moving along an arbitrary curve transversal to such curves (*coexistence curves*) one finds a unique Gibbs state except when one crosses one of the coexistence curves. The properties of the Gibbs states are, or may be, very different to the right and to the left of such transition points (for example $\int \sigma_0 d\mu = \mu(\sigma_0)$ might have different values on the two sides of these lines).

At the transition point one has coexistence of states with different qualitative properties but which are simultaneously Gibbs states for the *same*

N7.5.1 *potential.*[1]

The problem of the phase transitions (or that of their absence) is rather complex and largely open. A key result (Dyson, Lee–Yang, Ruelle) is the following one.

P7.5.1 **(7.5.1) Proposition:** (Phase transitions in one dimension)
Let $\varepsilon \in (0,1)$, $\beta > 0$, $h \in \mathbb{R}$ and consider the potential for $\{0,1\}^{\mathbb{Z}}$ defined by $\Phi_X = 0$ if $|X| \ge 3$ and

e7.5.1
$$\Phi_{\{i\}}(\sigma) = (2\sigma - 1)\,h\ , \quad \Phi_{\{i,j\}}(\sigma,\sigma') = -\beta\tfrac{(2\sigma-1)(2\sigma'-1)}{|i-j|^{1+\varepsilon}}\ ,\ \ i \ne j, \tag{7.5.1}$$

(i) If $h \ne 0$, $G_e(\Phi)$ contains a unique point.
(ii) If $h = 0$ and β is large enough $G_e(\Phi)$ contains at least two points.

Remark: The condition $\varepsilon > 0$ is required to apply proposition (6.1.1) (*i.e.* so that the potential verifies (5.1.5). The condition $\varepsilon < 1$ is necessary because if $\varepsilon > 1$ the uniqueness result of proposition (5.2.1) applies excluding phase transitions.

Proof: We shall not discuss the proof of item (i) because it concerns the problem of the absence of phase transitions: nevertheless the proof is very interesting and it is based on a celebrated theorem of algebra due to Lee–Yang, [LY52], later improved and extended by Asano, [As70] and by Ruelle, [Ru71].

Part (ii) ([Dy69]) will be treated in detail. Let $h = 0$ and let μ a Gibbs distribution for Φ obtained as weak limit of some sequence of approximate probability distributions. The sequence will be built as follows.

Let $\Lambda_N = \{-N,\ldots,N\}$, let $\underline{\sigma} \in \{0,1\}^{\Lambda_N}$ and let $(\underline{0}\,\underline{\sigma}\,\underline{0}) \in \{0,1\}^{\mathbb{Z}}$ be the sequence obtained by continuing $\underline{\sigma}$ with 0 outside Λ_N (*i.e.* $(\underline{0}\,\underline{\sigma}\,\underline{0})_i = \sigma_i$ for $i \in \Lambda_N$ and $(\underline{0}\,\underline{\sigma}\,\underline{0})_i = 0$ for $i \notin \Lambda_N$). Then we can repeat the argument used in the proof of proposition (5.1.1) (cf. also problems [5.1.10] and [5.1.11]). Defining the finite volume Gibbs state with $\underline{0}$ boundary conditions as the
N7.5.2 distribution which attributes to the configuration $\underline{0}\,\underline{\sigma}\,\underline{0}$ the probability [2]

e7.5.2
$$\widetilde{\mu}_N(\underline{\sigma}) = \frac{e^{-U^{\Phi}_{\Lambda_N}(\underline{\sigma})}}{\sum_{\underline{\sigma}'\in\{0,1\}^{\Lambda_N}} e^{-U^{\Phi}_{\Lambda_N}(\underline{\sigma}')}} \qquad \underline{\sigma} \in \{0,1\}^{\Lambda_N}, \tag{7.5.2}$$

we see that the limit for $N \to \infty$ of (7.5.2) is a Gibbs state $\mu \in G(\Phi)$, provided it exists and is invariant under translation; here $U^{\Phi}_{\Lambda_N}$ is given, (see (7.5.1)), by

e7.5.3
$$U^{\Phi}_{\Lambda_N}(\underline{\sigma}) = -\beta \sum_{\substack{i,j\in\Lambda_N\\ i\ne j}} \frac{1}{|i-j|^{1+\varepsilon}}\,(2\sigma_i - 1)(2\sigma_j - 1). \tag{7.5.3}$$

[1] This phenomenon is connected with, but it should be considered different from, the phenomenon of the possibility of coexistence of a gas with its liquid or its crystal like vapor in presence of water. The phases that we consider here correspond, in the physical interpretation of Gibbs states, to pure vapor or pure liquid (or pure crystal). The states of physical coexistence usually correspond, instead, to Gibbs states which are not invariant under translation.

[2] We use the notations for U^{Φ} introduced in (5.1.11) in definition (5.1.3).

It is, however, necessary to show preliminarily the existence of the limit as $N\to\infty$ of the $\widetilde\mu_N$ and its translation invariance. To show the existence of the limit of (7.5.2) it is convenient to set

e7.5.4
$$\begin{aligned} S_i=(2\sigma_i-1)\qquad &\text{if}\qquad \sigma_i=0,1,\\ S_\emptyset=1\ ,\qquad S_X=\prod_{i\in X}S_i\quad &\text{if}\quad X\ne\emptyset,\ X\subset\mathbb{Z},\ |X|<\infty,\end{aligned}\tag{7.5.4}$$

and note that the variables S_i, S_X assume the values $+1$ and -1. The functions $\underline\sigma\to S_X$ on $\{0,1\}^{\mathbb{Z}}$ form a set that spans a linear manifold dense in the space $C(\{0,1\}^{\mathbb{Z}})$ of the continuous functions on $\mathbb{Z}$.

To see that the limit μ of the sequence of distributions in (7.5.2) exists (in
N7.5.3 the sense that $\mu_N(f)\xrightarrow[N\to\infty]{}\mu(f)$ for all $f\in C(\{0,1\}^{\mathbb{Z}})$,[3] it will suffice to check, for all $X\subset\mathbb{Z}$, existence of the limit

e7.5.5
$$\langle S_X\rangle=\lim_{N\to\infty}\langle S_X\rangle_{\widetilde\mu_N}=\lim_{N\to\infty}\int\widetilde\mu_N(d\underline\sigma)S_X=\lim_{N\to\infty}\widetilde\mu_N(S_X)\tag{7.5.5}$$

(with obvious implicit definitions of the symbols). For this purpose note that

e7.5.6
$$\widetilde\mu_N(S_X)=\frac{\sum_{\underline\sigma\in\{0,1\}^{\Lambda_N}}S_Xe^{\beta\sum_{i,j\in\Lambda_N,\,i\ne j}|i-j|^{-1-\varepsilon}S_iS_j}}{\sum_{\underline\sigma\in\{0,1\}^{\Lambda_N}}e^{\beta\sum_{i,j\in\Lambda_N,\,i\ne j}|i-j|^{-1-\varepsilon}S_iS_j}}.\tag{7.5.6}$$

Therefore (7.5.6) is a particular case of an expression of the type (if $\Lambda\subset\mathbb{Z}$, $|\Lambda|<+\infty$)

e7.5.7
$$\langle S_X\rangle_{\underline J}=\frac{\sum_{\underline S\in\{-1,1\}^\Lambda}S_X\,e^{\sum_{Y\subset\Lambda}J(Y)\,S_Y}}{\sum_{\underline S\in\{-1,1\}^\Lambda}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}},\tag{7.5.7}$$

with $J(X)\ge0$: equation (7.5.6) is obtained, indeed, from (7.5.7) by setting $\Lambda=\Lambda_{N'}=\{-N',\ldots,N'\}$, with N' arbitrarily chosen such that $N'\ge N$, $J(Y)=0$ if $|Y|\ge3$, $J(i)=0$, $J(i,j)=\beta|i-j|^{-1-\varepsilon}$ if $i\ne j$ and $i,\,j\in\Lambda_N$, while $J(i,j)=0$ if i or $j\in\Lambda_{N'}/\Lambda_N$ and if $i=j\in\Lambda_N$.

If in (7.5.6) we change the value of N, say from N to $N'\ge N$, this is equivalent to increasing in (7.5.7) some values of $J(i,j)$ in the following way: one interprets them as 0 for the pairs i,j which are not both in $[-N,N]$ and then one raises their values from 0 to $\beta|i-j|^{-1-\varepsilon}$. Therefore the existence of the limit (7.5.5) will be guaranteed if we shall show that $\langle S_X\rangle_{\underline J}$ is monotonic in $J(Y)$, for all Y, under the condition that all the values $J(Y)$ are ≥0. This means, by an easy check, that we must verify the following inequality: if for every Y one has $J(Y)\ge0$ then for every X,Y

e7.5.8
$$\frac{\partial\langle S_X\rangle_{\underline J}}{\partial J(Y)}=\langle S_XS_Y\rangle_{\underline J}-\langle S_X\rangle_{\underline J}\langle S_Y\rangle_{\underline J}\ge0.\tag{7.5.8}$$

[3] *i.e.* , more mathematically, in the weak $C(\{0,1\}^{\mathbb{Z}})$–topology on $M^0(\{0,1\}^{\mathbb{Z}})$ thought of as the unit ball of the space of probability distributions on $\{0,1\}^{\mathbb{Z}}$, dual space of $C(\{0,1\}^{\mathbb{Z}})$.

This is a very well known inequality (*second Griffiths inequality*), the first of a family of astute inequalities that keeps proliferating. Its proof is surprisingly simple, in the version due to Ginibre (the original proof was much more involved), [Gi70]. Note, for this purpose, that

e7.5.9

$$\begin{aligned}\langle S_{X_0}S_{Y_0}\rangle_{\underline{J}}-\langle S_{X_0}\rangle_{\underline{J}}\langle S_{Y_0}\rangle_{\underline{J}}=&\frac{\sum_{\underline{S}\in\{-1,1\}^\Lambda}S_{X_0}S_{Y_0}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}}{\sum_{\underline{S}\in\{-1,1\}^\Lambda}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}}-\\ -\frac{\sum_{\underline{S}\in\{-1,1\}^\Lambda}S_{X_0}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}}{\sum_{\underline{S}\in\{-1,1\}^\Lambda}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}}&\cdot\frac{\sum_{\underline{S}'\in\{-1,1\}^\Lambda}S'_{Y_0}e^{\sum_{Y\subset\Lambda}J(Y)S'_Y}}{\sum_{\underline{S}'\in\{-1,1\}^\Lambda}e^{\sum_{Y\subset\Lambda}J(Y)S'_Y}}=\\ =\frac{\sum_{\underline{S},\underline{S}'\in\{-1,1\}^\Lambda}(S_{X_0}S_{Y_0}-S_{X_0}S'_{Y_0})e^{\sum_{Y\subset\Lambda}J(Y)(S_Y+S'_Y)}}{\big(\sum_{\underline{S}\in\{-1,1\}^\Lambda}e^{\sum_{Y\subset\Lambda}J(Y)S_Y}\big)^2}&.\end{aligned}\tag{7.5.9}$$

The denominator of the r.h.s. of (7.5.9) is positive. The numerator can be conveniently transformed by setting $S'_i=S_iT_i$, with $T_i=\pm1$, and by transforming the sum on $\underline{S},\underline{S}'\in\{-1,1\}^\Lambda$ into the sum over $\underline{S}$, $\underline{T}\in\{-1,1\}^\Lambda$; after the latter change the numerator of the r.h.s. of (7.5.9) becomes

e7.5.10

$$\begin{aligned}&\sum_{\underline{S},\underline{T}\in\{-1,1\}^\Lambda}(1-T_{Y_0})S_{X_0}S_{Y_0}e^{\sum_{Y\subset\Lambda}J(Y)(1+T_Y)S_Y}=\\ &=\sum_{\underline{T}\in\{-1,1\}^\Lambda}(1-T_{Y_0})\Big(\sum_{\underline{S}\in\{-1,1\}^\Lambda}S_{X_0}S_{Y_0}e^{\sum_{Y\subset\Lambda}J(Y)(1+T_Y)S_Y}\Big),\end{aligned}\tag{7.5.10}$$

and since $1-T_{Y_0}\ge0$ it will suffice to check, in order to obtain (7.5.8), that the sum in parentheses in (7.5.10) is not negative. One has

e7.5.11

$$\begin{aligned}&\sum_{\underline{S}\in\{-1,1\}^\Lambda}S_{X_0}S_{Y_0}\exp\Big(\sum_{Y\subset\Lambda}J(Y)(1+T_Y)S_Y\Big)=\\ &=\sum_{k=0}^\infty\frac{1}{k!}\sum_{\underline{S}\in\{-1,1\}^\Lambda}\Big(\sum_{Y\subset\Lambda}J(Y)(1+T_Y)S_Y)\Big)^kS_{X_0}S_{Y_0}=\\ &=\sum_{k=0}^\infty\frac{1}{k!}\sum_{Y_1,\ldots,Y_k\subset\Lambda}\Big(\prod_{i=1}^k\big(J(Y_i)(1+T_{Y_i})\big)\Big)\Big(\sum_{\underline{S}\in\{-1,1\}^\Lambda}S_{Y_1}\ldots S_{Y_k}S_{X_0}S_{Y_0}\Big),\end{aligned}\tag{7.5.11}$$

and, since $J(Y_i)(1+T_{Y_i})\ge0$, it suffices to show that the term in the last parentheses is ≥0. A moment of reflection makes it manifest that such a term has value either $2^{|\Lambda|}$ or 0: hence (7.5.8) is proved.

Together with the (7.5.8) we obtain also, as already noted, the existence of the limit in (7.5.5) for all X, and hence such a limit defines a probability distribution $\mu\in G^0(\Phi)$. The same inequality (7.5.8) shows that μ is invariant under translation. Consider indeed the set X and its translated

τX; by denoting here with $\langle\cdot\rangle_\Lambda$ the average with respect to the probability distribution (7.5.2), with Λ replacing Λ_N, one has

e7.5.12
$$\begin{aligned}\langle S_X\rangle_{\Lambda_N}&\equiv\langle S_X\rangle_{\{-N,\ldots,N\}}\le\langle \underline{S}_X\rangle_{\{-N,\ldots,N+2\}}\equiv\\&\equiv\langle S_{\tau X}\rangle_{\{-N-1,\ldots,N+1\}}\equiv\langle S_{\tau X}\rangle_{\Lambda_{N+1}}\le\\&\le\langle S_{\tau X}\rangle_{\{-N-3,\ldots,N+1\}}\equiv\langle S_X\rangle_{\{-N-2,\ldots,N+2\}}\equiv\langle S_X\rangle_{\Lambda_{N+2}},\end{aligned}\tag{7.5.12}$$

which implies, in the limit $N\to\infty$, that $\mu(S_X)=\mu(S_{\tau X})$, for all X, and therefore the invariance of μ.

It is also obvious, for symmetry reasons, that (7.5.6) implies

e7.5.13
$$\mu(S_i)=\widetilde\mu_N(S_i)\equiv\langle S_i\rangle_{\Lambda_N}=0.\tag{7.5.13}$$

Were μ ergodic we should have

e7.5.14
$$\mu\big(\big(M^{-1}\sum_{j=0}^{M-1}S_j\big)^2\big)\xrightarrow[M\to\infty]{}0,\tag{7.5.14}$$

because, by Birkhoff theorem, $M^{-1}\sum_{j=0}^{M-1}S_j$ would converge μ–almost everywhere to $\mu(S_i)=0$ (staying uniformly bounded by 1).

Hence we shall proceed to show that the limit (7.5.14) is positive if β is large enough. This will complete the proof of (ii) by the comment in the sixth paragraph at the beginning of this section (it will imply, indeed, that μ is not ergodic).

By the remarked monotonicity of $\langle S_X\rangle_{\Lambda_N}$ with respect to N it will suffice
N7.5.3a to show that [4]

e7.5.15
$$\widetilde\mu_M\big(\big(M^{-1}\sum_{j=1}^{M}S_j\big)^2\big)\ge\overline{m}^2>0\tag{7.5.15}$$

for a divergent sequence of choices of M. Indeed

e7.5.16
$$\mu\big(\big(M^{-1}\sum_{j=1}^{M}S_j\big)^2\big)\ge\widetilde\mu_M\big(\big(M^{-1}\sum_{j=1}^{M}S_j\big)^2\big),\tag{7.5.16}$$

as we see by developing the squares into monomials and by applying (7.5.8) to each of them (*i.e.* $\mu(S_X)\ge\widetilde\mu_M(S_X)$ for all M and for all $X\subset\mathbb{Z}$).

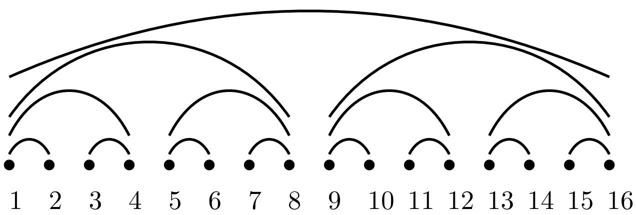
We shall choose the sequence $M=2^N$, $N=0,1,\ldots$ and set

e7.5.17
$$\overline{f}(N)=\widetilde\mu_{2^N}\big(\big(2^{-N}\sum_{j=1}^{2^N}S_j\big)^2\big).\tag{7.5.17}$$

[4] Just for convenience we shall consider the sum in (7.5.14) running from 1 to M instead that from 0 to $M-1$ (see also comments about Fig.(7.5.1) below): of course such a shift of the summation label it is quite irrelevant.

It is not easy to compute $\overline{f}(N)$: however we can estimate such a quantity by still using (7.5.8) and thinking of $\widetilde{\mu}_{2^N}(S_X)$ as a sum of the type (7.5.7) with the choice of the constant $\underline{J}$ indicated there. The idea is to set equal to zero or to decrease some of the nonzero constants $\underline{J}$ in order to obtain a controllable expression. Obviously it is necessary to make this without making too much worse the bound of $\overline{f}(N)$. The choice of Dyson is to define a lower bound of $\beta|i-j|^{-1-\varepsilon}$ via a "hierarchical" procedure.

For convenience in the following we shall imagine "translating" the set Λ_M in such a way that the first point is at 1. Divide the points of $[1,2^N]$ into 2^{N-p} blocks of 2^p consecutive points each, for $p=0,1,\ldots,N$; see Fig.(7.5.1).



F7.5.1 **Fig.(7.5.1)** Hierarchical grouping of the sites in $[1,\ldots,2^N]$.

Such blocks (p–blocks) will be denoted with the symbol (p,k); the "k–th block between the p–blocks", $p=0,1,\ldots,N$, $k=1,2,\ldots,2^{N-p}$ will be

e7.5.18
$$(p,k)=\{i\,|\,(k-1)2^p<i\le k2^p\}. \tag{7.5.18}$$

Note that

e7.5.19
$$\beta\,|\,i-j\,|^{-1-\varepsilon}>\beta\,2^{-p(i,j)(1+\varepsilon)}, \tag{7.5.19}$$

if $p(i,j)=\{$minimum value of p for which i and j are found in the same p–block$\}$ (and it is clear that $p(i,j)>\log_2|\,i-j\,|$).

Set, in (7.5.7), $J(Y)=0$ if $|\,Y\,|\ge 3$ or $|\,Y\,|=1$ and

e7.5.20
$$\begin{aligned}&J(i,j)=\beta 2^{-p(i,j)(1+\varepsilon)}\quad\text{if}\quad 0<i,j\le 2^N,\\&J(i,j)=0\qquad\text{otherwise,}\end{aligned} \tag{7.5.20}$$

We see that $\beta\,|\,i-j\,|^{-1-\varepsilon}>J(i,j)$ for all $i,j\in\{1,\ldots,2^N\}$, with $i\neq j$. Furthermore, for $N\ge 1$,

e7.5.21
$$\begin{aligned}&\sum_{i,j\in\{1,\ldots,2^N\}}J(i,j)S_iS_j=\beta\sum_{p=0}^{N-1}\sum_{r=1}^{2^{N-p-1}}2^{-(p+1)(1+\varepsilon)}S(p,2r-1)S(p,2r)\\&\text{if}\quad S(p,k)=\sum_{j=1+(k-1)2^p}^{k2^p}S_j\end{aligned} \tag{7.5.21}$$

and $S(p,k)$ is the sum of the variables of the (p,k)–block. Set then

e7.5.22
$$W(\underline{S}) = \exp\Big(\beta \sum_{p=0}^{N-1} \sum_{r=1}^{2^{N-p-1}} 2^{-(p+1)(1+\varepsilon)} S(p,2r-1)S(p,2r)\Big) \tag{7.5.22}$$

and

e7.5.23
$$f(N) = \frac{\sum_{\underline{S}\in\{-1,1\}^{2^N}} W(S)\big(\sum_{j=1}^{2^N} 2^{-N} S_j\big)^2}{\sum_{\underline{S}\in\{-1,1\}^{2^N}} W(S)}. \tag{7.5.23}$$

By (7.5.8) and the remark following (7.5.20), we get

e7.5.24
$$\overline{f}(N) \ge f(N) \qquad\qquad \text{for all}\, N \ge 1. \tag{7.5.24}$$

Then (7.5.23) can be cast into a recursive form apt for a bound.

Let us call R the random variable $R = 2^{-N}\sum_{j=1}^{2^N} S_j$, distributed according to the distribution induced by that of the variables $\underline{S}$,

e7.5.25
$$\begin{aligned} P(\underline{S}) &= \frac{e^{\beta \sum_{p=0}^{N-1} \sum_{r=1}^{2^{N-p-1}} 2^{-(p+1)(1+\varepsilon)} S(p,2r-1)S(p,2r)}}{\sum_{\underline{S}'\in\{-1,1\}^{2^N}} e^{\sum_{p=0}^{N-1} \sum_{r=1}^{2^{N-p-1}} 2^{-(p+1)(1+\varepsilon)} S'(p,2r-1)S(p,2r)}}\\ &= \frac{W(\underline{S})}{\sum_{\underline{S}'\in\{-1,1\}^{2^N}} W(\underline{S}')}, \end{aligned} \tag{7.5.25}$$

and let $\pi_N(R)dR$ be the distribution of the variable R (with a symbolic notation because $\pi_N(R)dR$ is a sum of Dirac measures).

It is immediate, from the structure of (7.5.25), that there is a simple relation between π_N and π_{N-1}. In fact, for $N \ge 1$,

e7.5.26
$$\begin{aligned} \pi_N(R) =& C_N \int \delta\Big(\frac{R_1+R_2}{2} - R\Big) dR_1 dR_2 \cdot\\ &\cdot e^{\frac{\beta}{4} 2^{N(1-\varepsilon)} R_1 R_2} \pi_{N-1}(R_1)\pi_{N-1}(R_2), \end{aligned} \tag{7.5.26}$$

where C_N is a normalization constant (*i.e.* it is the inverse of the integral over R of the remaining part of the r.h.s. of (7.5.26)). One has

e7.5.27
$$\begin{aligned} f(N) = \int R^2 \pi_N(R) dR, \qquad f(0) = 1,\\ \pi_0(R) = 2^{-1}(\delta(R-1) + \delta(R+1)), \end{aligned} \tag{7.5.27}$$

and therefore

e7.5.28
$$\begin{aligned} f(N) = C_N \int \Big(\frac{R_1+R_2}{2}\Big)^2 e^{\frac{\beta}{4} 2^{N(1-\varepsilon)} R_1 R_2} \pi_{N-1}(R_1)\pi_{N-1}(R_2) dR_1 dR_2 =\\ = \frac{1}{2} C_N \int R_1^2 e^{\frac{\beta}{4} 2^{N(1-\varepsilon)} R_1 R_2} \pi_{N-1}(R_1)\pi_{N-1}(R_2) dR_1 dR_2 +\\ + \frac{1}{2} C_N \int R_1 R_2 e^{\frac{\beta}{4} 2^{N(1-\varepsilon)} R_1 R_2} \pi_{N-1}(R_1)\pi_{N-1}(R_2) dR_1 dR_2. \end{aligned} \tag{7.5.28}$$

The first of the last two integrals is $\sum_{\underline{S}} P(\underline{S})\, S(N-1,1)^2$, which, by (7.5.8), again, is bounded below by $f(N-1)$ (*i.e.* by the value that it takes setting $\beta=0$ in the exponential that appears explicitly in (7.5.28)). Hence

e7.5.29

$$f(N) > \frac{1}{2}\, f(N-1) + \\ + \frac{1}{2}\frac{\int R_1R_2\big[e^{\frac{\beta}{4}\,2^{N(1-\varepsilon)R_1R_2}}\big]\pi_{N-1}(R_1)\pi_{N-1}(R_2)dR_1dR_2}{\int\big[e^{\frac{\beta}{4}\,2^{N(1-\varepsilon)R_1R_2}}\big]\pi_{N-1}(R_1)\pi_{N-1}(R_2)dR_1dR_2}. \tag{7.5.29}$$

For a lower bound on the second term we shall think of $x = R_1R_2$ as a random variable with the distribution induced by the distribution of R_1 and R_2 defined by $\pi_{N-1}(R_1)\pi_{N-1}(R_2)$. Then the ratio that appears in (7.5.8) can be written as

e7.5.30

$$\varphi(a) = \frac{\int P(x)xe^{ax}dx}{\int P(x)e^{ax}dx} \qquad \text{with } a = \frac{\beta}{4}\,2^{N(1-\varepsilon)}, \tag{7.5.30}$$

if $P(x) =$ distribution of $x = \int \pi_{N-1}(R_1)\pi_{N-1}(R_2)\delta(R_1R_2-x)dR_1dR_2$.

Remarking that $P(x) = P(-x)$, *i.e.* that the distribution P is symmetric, we can apply a simple inequality (Dyson) to give a lower bound to (7.5.30): if $\widehat{x} = \big(\int P(x)x^2dx\big)^{1/2}$ one has

e7.5.31

$$\varphi(a) \ge \widehat{x}\,\tanh(a\widehat{x}) \stackrel{def}{=} \psi(a), \tag{7.5.31}$$

valid if $P(x) = P(-x)$ (cf. problem [7.5.1] at the end of this section).

Then (7.5.31) can be applied to (7.5.29) to deduce

e7.5.32

$$f(N) > \frac{1}{2}f(N-1) + \frac{1}{2}\widehat{x}\,\tanh(a\widehat{x}), \tag{7.5.32}$$

and, by remarking that $\widehat{x}^2 = \int(R_1R_2)^2\pi_{N-1}(R_1)\pi_{N-1}(R_2)dR_1dR_2 = f(N-1)^2$, we get

e7.5.33

$$f(N) > \frac{1}{2}f(N-1) + \frac{1}{2}f(N-1)\,\tanh(af(N-1)) = \\ = f(N-1)\Big(1+\exp\Big[-\frac{\beta}{2}\,2^{N(1-\varepsilon)}f(N-1)\Big]\Big)^{-1}. \tag{7.5.33}$$

Suppose inductively that $f(k) \ge \frac{1}{4}\big(1+\frac{1}{2+k}\big)$ for $k=0,\ldots,N-1$. The hypothesis, true for $k=0$ ($f(0)=1$), gives, combined with (7.5.33),

e7.5.34

$$f(N) \ge \frac{1}{4}\Big(1+\frac{1}{1+N}\Big)\Big(1+e^{\left[-\frac{\beta}{2}\,2^{N(1-\varepsilon)}\frac{1}{4}\left(1+\frac{1}{1+N}\right)\right]}\Big)^{-1} = \\ = \frac{1}{4}\Big(1+\frac{1}{1+N}\Big)\Big(1+e^{\left[-\frac{\beta}{8}\,2^{N(1-\varepsilon)}\frac{N+2}{N+1}\right]}\Big)^{-1} \ge \\ \ge \frac{1}{4}\Big(1+\frac{1}{2+N}\Big) \tag{7.5.34}$$

for all $N \geq 1$, *provided β is large enough.* Hence if β is large enough one has

e7.5.35
$$f(N) \geq 1/4, \tag{7.5.35}$$

which concludes the proof. ∎

Remarks: (1) The requirement that $\varepsilon > 0$ comes from the requirement that the potential be summable which is necessary in order that the Gibbs state be defined and enjoys the properties that we expect on the basis of the general theory. The requirement that $\varepsilon < 1$ is fundamental: as we have seen the proof is essentially reduced to the theory of the recursive relation

e7.5.36
$$\begin{aligned}\pi'(R) = \frac{\int \exp[\beta b_N R'(2R-R')]\pi(R')\pi(2R-R')dR'}{\int \exp[\beta b_N R'(2R''-R')]\pi(R')(2R''-R')dR'dR''} \stackrel{def}{=} \\ \stackrel{def}{=} (K_N\pi)(R),\end{aligned} \tag{7.5.36}$$

or, better, to some aspects of it. In the case $\varepsilon < 1$, $b_N = \frac{1}{4} 2^{+N(1-\varepsilon)} \xrightarrow[N\to\infty]{} +\infty$ very rapidly, while $b_N \xrightarrow[N\to\infty]{} 0$ if $\varepsilon > 1$.
(2) The cases $b_N = 0$ and $b_N = +\infty$ are extreme cases. In the first case the relation that links π to π' links the distribution of the average of two equally distributed random variables to that of the individual variables

e7.5.37
$$\pi'(R) = \int \pi(R_1)\pi(R_2)\delta\Big(\frac{R_1+R_2}{2} - R\Big)dR_1dR_2, \tag{7.5.37}$$

while in the second case the relation becomes rather degenerate: for example every π of the form

e7.5.38
$$\overline{\pi}(R) = \frac{1}{2}\left(\delta(R+m) + \delta(R-m)\right) \tag{7.5.38}$$

is transformed into itself by the recurrence relation.
If β is large and if $b_N \to \infty$ we can think that we are very close to the case $b_N = +\infty$ and that the distribution $K_N K_{N-1} \ldots K_1 \pi_0$ conserves the initial structure that presents two maxima in -1 and 1 and converges to a distribution $\overline{\pi}$ with some $m \cong 1$.
If instead $b_N \to 0$, for all $\beta > 0$ and for N large enough the recurrence relation for K_N should not differ substantially from (7.5.37). This case corresponds to the study of the distribution of the average of 2^N *equally distributed independent variables* which converges, by the "0–1 law", to 0 if such variables are symmetrically distributed: hence we should have the *trivial* result that $K_N K_{N-1} \ldots K_1 \pi_0 \xrightarrow[N\to\infty]{} \delta$. The connection between the above remarks and the preceding proof should be clear and motivates intuitively the several steps of the proof.
(3) The theory of the recurrence relations of the type (7.5.36) is another interesting chapter of the theory of Gibbs states and it appears as a natural generalization of the classical problems on the distributions of sums or of linear combinations of independent random variables. As we already see

from the content of this section such theory is, perhaps surprisingly, rich of results and interesting phenomena. The modern approach to the study of critical phenomena is based on developments that followed the analysis by Wilson of the properties of "hierarchical" relations like (7.5.36) which came out remarkably at about the same time as Dyson's work illustrated here, [Wi70]. For a review see [Wi83], [Ga85], and [Ga99].

Problems for §7.5

Q7.5.1 **[7.5.1]:** Consider the functions $\varphi(a)$ and $\psi(a)$, $a \ge 0$, introduced in (7.5.30) and (7.5.31). If $\langle\cdot\rangle$ denotes the average value with respect to P note that

$$\varphi(a) = \frac{\langle x \sinh ax\rangle}{\langle \cosh ax\rangle}, \qquad \frac{\partial\varphi}{\partial a}(a) = \frac{\langle x^2 \cosh ax\rangle}{\langle \cosh ax\rangle} - \varphi^2(a),$$

and, furthermore,

$$\begin{aligned}
&\frac{\langle x^2 \cosh ax\rangle}{\langle \cosh ax\rangle} - \langle x^2\rangle = \frac{\langle x^2 \cosh ax\rangle - \langle x^2\rangle\langle \cosh ax\rangle}{\langle \cosh ax\rangle} = \\
&= \langle \cosh ax\rangle^{-1} \int P(x)P(y)(x^2 \cosh ax - x^2 \cosh ay)dxdy = \\
&= \langle \cosh ax\rangle^{-1} \int P(x)P(y)x^2(\cosh ax - \cosh ay)dxdy = \\
&= \frac{1}{2}\langle \cosh ax\rangle^{-1} \int P(x)P(y)(x^2 - y^2)(\cosh ax - \cosh ay)dxdy \ge 0.
\end{aligned}$$

Hence $\partial\varphi(a)/\partial a \ge \langle x^2\rangle - \varphi^2(a)$. Show that ψ is instead a solution of $\frac{\partial\psi}{\partial a} = \langle x^2\rangle - \psi^2$, and that from $\varphi(0) = \psi(0) = 0$ it follows $\varphi - \psi \ge 0$, for all $a \ge 0$.

Q7.5.2 **[7.5.2]:** (*Hierarchical model*)
Give a definition of the measure formally defined on $\mathbb{R}^{\mathbb{Z}^+}$ by

$$\text{const}\,\left(e^{\beta\sum_{p=1}^{\infty} 2^{-\alpha p}\sum_{k=0}^{\infty} S(p,k)^2}\right)\prod_{i\in\mathbb{Z}^+}\pi(S_j)dS_j = \mu(d\underline{S}),$$

where $\pi(S)dS$ is a measure with compact support and $1 < \alpha$ is an exponent measuring the strength of the interaction at large distances. (*Hint:* Proceed as in the construction of the Gibbs states by realizing μ as limit point of a sequence:

$$\widetilde{\mu}_N(d\underline{S}) = \text{ const }\left(e^{\beta\sum_{p=1}^{N} 2^{-\alpha p}\sum_{k=0}^{2^{N-p}} (S(p,k))^2}\right)\prod_{i=1}^{2^N}\pi(S_j)dS_j \prod_{j>2^N}$$

of finite volume distributions. Compute the conditional probability $\mu(S_1 \ldots S_{2^N} \mid S_{2^N+1}, \ldots)$ and define $G(\alpha,\beta,\mu)$ as the set of the measures on $\mathbb{R}^{\mathbb{Z}^+}$ that have these same conditional probabilities).

Q7.5.3 **[7.5.3]:** (*Renormalization in the hierarchical model*)
Let $\mu \in G(\alpha,\beta,\pi)$ (cf. preceding problem), $\rho \in (0,1)$ and set

$$S'_j = \frac{S_{1,j}}{2^{\rho/2}} = \frac{S_{2j} + S_{2j-1}}{2^{\rho/2}}.$$

We define the transformation K_ρ mapping a distribution μ on $\underline{S}$ into the distribution μ' of the new variables S' above: $\mu' = K_\rho \mu$. Show that the distribution of the variables S', which can be thought of as a probability distribution on $\mathbb{R}$, is the function $\pi' \in G(\alpha, \beta 2^{\rho-\alpha}, \pi')$, where

$$\pi'(S) = \text{const}\, e^{\beta 2^{\alpha-\rho} S^2} \int \pi(S_1)\pi(S_2)\delta\Big(\frac{S_1+S_2}{2^{\rho/2}} - S\Big) dS_1 dS_2.$$

The semigroup of transformations given by the iterates of K_ρ is an example of what is called a *renormalization group.*

Q7.5.4 **[7.5.4]:** (*Triviality of fixed points for the renormalization map in hierarchical models*) Making use of the central limit theorem show that if $\alpha > 2$ and $\rho = 1$ the limit: $\lim_{n\to\infty} K_1^n \mu = \overline{\mu}$ exists in a suitable (and natural) weak sense and it is Gaussian: $\overline{\mu}(d\underline{S}) = \prod_{j=1}^\infty \frac{e^{-S_j^2/2C}}{\sqrt{2\pi C}}\, dS_j$, where $C > 0$ is a suitable constant.

Q7.5.5 **[7.5.5]:** Develop the analogues of problems [7.5.2], [7.5.3] and [7.5.4] for

$$\text{const}\,\Big(e^{2^{1-\alpha}\beta \sum_{p=1}^\infty 2^{-\alpha p} \sum_{k=1}^\infty S(p,2k-1)S(p,k)}\Big) \prod_{j\in\mathbb{Z}_+} \pi(S_j) dS_j,$$

where π is a smooth rapidly decreasing even density distribution on $\mathbb{R}$.

Q7.5.6 **[7.5.6]:** Call $G'(\alpha, \beta, \widetilde{\pi})$ the set of the measures constructed as in problem [7.5.1] starting from the formal expression of problem [7.5.5] show that given $\alpha, \beta > 0$ there exists $\widetilde{\pi}$ for which $G'(\alpha, \beta, \widetilde{\pi}) = G(\alpha, \beta, \pi)$ (*Hint:* Compare the conditional probabilities of the elements of G' and of G).

Bibliographical note to §7.5

Modern phase transitions theory in systems with dimension $d \geq 2$ originated with the works of Peierls, [Pe36], van der Waerden, [Wa41] and Onsager, [On44], on the Ising model (with nearest neighbour potential: $J(2\sigma-1)(2\sigma'-1)$, which is the simplest Gibbs process on $\{0,1\}^{\mathbb{Z}^2}$). In the latter work the pressure $P(J)$ is computed in terms of elementary functions and their quadratures. Another contemporary work (by van der Waerden, [VW41]), went unnoticed although it gave a proof of the existence of phase transitions in the same model considered by Onsager. Phase transitions theory in one-dimensional systems began with a negative theorem by Van Hove, [VH50]: the theorem was later substantially extended by Ruelle, [Ru67], to cover the "most general " possible case in which we could reasonably expect to show absence of phase transitions by mathematically showing the validity of a famous heuristic argument of Landau and Lifshitz in the last page of [LL67].

Violating Van Hove–Ruelle condition, (5.2.1), on absence of phase transitions makes it possible to obtain examples of potentials for one dimensional systems with phase transitions. These results are in a series of papers by Dyson; the third of them is particularly simple and readable, and it can be useful as a first reading, see [Dy69]; the analysis in Section §7.5 is based on it.

An essential instrument in the phase transitions theory are the Griffiths inequalities. The very simple proof described above is due to Ginibre, [Gi70]: the previous proofs were much more involved as one can realize by looking, for instance, to the proof in [Ru69]. Theory of inequalities between average values with respect to Gibbs distribution has had, after the works of Griffiths and Ginibre, a tumultuous development. See as an introduction [FKGC71], [Ho74], [Le74]. Another very interesting type of result related to the phase transitions theory (a result that we do not use here but that would have been necessary if we discussed the proof of item (i) of proposition (7.3.1)) is the Lee-Yang theorem and its generalizations due to Asano and Ruelle. It is a theorem of algebra which is quite surprising: in it one determines the locus of the zeroes of a certain class of polynomials of interest in the theory of Gibbs states. It is perhaps surprising that this theorem has received limited applications in other areas of mathematics. See the appendix of [LY52] and [Ru71], [Ru72], [Ru73], [SF72].

Phase transitions theory is quite developed in systems in more dimensions. An introduction can be the review article [Ga72], and [Ga00]. The phenomenon of coexisting phases for systems in a several dimensional space has been studied, sometimes in great detail; an analogous theory for one dimensional systems would be really interesting. Phase transitions theory in systems with a simple potential is easier in one dimension because most systems simply do not have phase transitions. However they have transitions in more dimensions. In one dimension only *remarkably complex* systems can present phase transitions, like the long range system considered in this section.

.

CHAPTER VIII

Special ergodic theory problems in nonchaotic dynamics

§8.1 Theory of quasi-periodic Hamiltonian motions

A very natural and important question that one can ask about stability in Hamiltonian systems is what becomes of the simple foliation of phase space into invariant tori when a perturbing force is switched on.

The simplest case is when the *unperturbed Hamiltonian* has the form

e8.1.1
$$\mathcal{H}_0(\underline{A},\underline{\alpha}) \stackrel{def}{=} K(\underline{A}) = \frac{1}{2J}\underline{A}^2, \qquad \begin{cases} \underline{A} = (A_1,\ldots,A_\ell) \in \mathbb{R}^\ell, \\ \underline{\alpha} = (\alpha_1,\ldots,\alpha_\ell) \in \mathbb{T}^\ell, \end{cases} \tag{8.1.1}$$

where $J > 0$ and $(\underline{A},\underline{\alpha}) \in \mathbb{R}^\ell \times \mathbb{T}^\ell$ are ℓ canonical momenta and their ℓ conjugate angles. This is the Hamiltonian representing the motion of ℓ point masses (*rotators*) rotating on ℓ unit circles with moment of inertia J, angular momenta A_j and positions α_j: $K(\underline{A})$ is their kinetic energy.

The motion trivially takes place on ℓ–dimensional invariant tori parameterized by the momenta $\underline{A}$: calling $\underline{\omega} = (\omega_1,\ldots,\omega_\ell) = J^{-1}\underline{A}$, the point $(\underline{A},\underline{\alpha})$ evolves in time t into $S_t(\underline{A},\underline{\alpha}) = (\underline{A},\underline{\alpha}+\underline{\omega}\,t)$.

If $V(\underline{\alpha})$ is an analytic function (*perturbing potential*) one might expect that the perturbed system described by the Hamiltonian function $K(\underline{A})+\varepsilon V(\underline{\alpha})$ still has, at least for small ε, quasi-periodic motions developing on tori which foliate the entire phase space and are close to the unperturbed ones in the sense that their parametric equations have the form

e8.1.2
$$\begin{cases} \underline{\alpha} = \underline{\psi} + \underline{h}_\varepsilon(\underline{B},\underline{\psi}), \\ \underline{A} = \underline{B} + \underline{H}_\varepsilon(\underline{B},\underline{\psi}), \end{cases} \qquad \underline{\psi} \in \mathbb{T}^\ell, \tag{8.1.2}$$

where $\underline{H}_\varepsilon, \underline{h}_\varepsilon$ are 2ℓ analytic functions of ε small as $\varepsilon \to 0$ and $\underline{B}$ are ℓ parameters that are introduced to parameterize the various tori. On the tori the motion should be $(\underline{A}, \underline{\alpha}) \to S_t^\varepsilon(\underline{A}, \underline{\alpha}) = (\underline{A}(t), \underline{\alpha}(t))$ with $(\underline{A}(t), \underline{\alpha}(t))$ obtained by replacing $\underline{\psi}$ with $\underline{\psi} + \underline{\omega}'\,t$ in equation (8.1.2), *i.e.* the motion should be given by

e8.1.3
$$\begin{cases} \underline{\alpha}(t) = \underline{\alpha} + \underline{\omega}' t + \underline{h}_\varepsilon(\underline{B}, \underline{\alpha} + \underline{\omega}' t), \\ \underline{A}(t) = \underline{B} + \underline{H}_\varepsilon(\underline{B}, \underline{\alpha} + \underline{\omega}' t), \end{cases}, \tag{8.1.3}$$

with $\underline{B}$ determined by the condition that the initial data $\underline{A}$ should be contained in the torus, $\underline{A} = \underline{B} + \underline{H}_\varepsilon(\underline{B}, \underline{\alpha})$. Moreover the velocities $\underline{\omega}'$ (*rotation vectors*) should be suitable functions of $\underline{B}$, *i.e.* of the torus containing the initial data $(\underline{A}, \underline{\alpha})$.

In fact the parameters $\underline{B}$ could be defined quite arbitrarily (by changing correspondingly the definition of the function $\underline{H}_\varepsilon(\underline{B}, \underline{\alpha})$). However in the unperturbed case there is a direct proportionality between velocities and angular momenta so that a natural choice of the parameters $\underline{B}$ could be simply $\underline{B} = J\,\underline{\omega}'$. This would mean choosing to parameterize the perturbed invariant tori by the rotation vector of the quasi-periodic motion on them.

Another natural choice could be that of determining the invariant tori by their *average action*:

e8.1.4
$$\underline{B} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \underline{A}(t)\,dt; \tag{8.1.4}$$

in this case $\underline{\omega}'$ would be a suitable function of $\underline{B}$, not necessarily equal to $J^{-1}\underline{B}$. However as $\underline{B}$ varies the velocities $\underline{\omega}'$ should assume all possible values, because in the unperturbed case they take all possible values as functions of the angular momenta. In other words the perturbed motion would exhibit all unperturbed motions with the only difference that they take place on slightly deformed trajectories.

The above *naive* picture representing the perturbed evolution as taking place on a phase space smoothly foliated into invariant tori on which a peaceful quasi-periodic motion takes place is clearly impossible as the simple case $\ell = 2$, $V(\underline{\alpha}) = \cos(\alpha_1 - \alpha_2)$ shows. In this case the perturbed motions with velocities $\omega_1 = \omega_2 = \omega$ do not form a family of closed trajectories filling up a 2–dimensional torus, as equation (8.1.2) would imply. It is easy to check that while the unperturbed motions with $\omega_1 = \omega_2 = \omega$ consist in a *continuum family of periodic motions* covering a torus the perturbed motions *only contain two* such motions which are close to the unperturbed ones; see problem [8.1.1].

In other words the perturbation "breaks" the invariant torus with $\omega_1 = \omega_2 = \omega$ and the only trace left of it are two isolated periodic orbits and no motion near them is periodic with period $2\pi/\omega$. All this can be easily checked because the perturbed system is still integrable in the sense of Hamiltonian mechanics in most of phase space and its theory reduces essentially to that of the classical pendulum.

It is very interesting that the above example in a way describes "all that can go wrong" and points out that essentially the only exceptions occur at *resonances*, *i.e.* the only motions that may be "absent" among the perturbed motions and which were present among the unperturbed ones are motions with "resonant frequencies" or with frequencies "too close" to resonant ones. A resonance is defined as a quasi-periodic motion with rotation vector $\underline{\omega} \in \mathbb{R}^\ell$ such that

e8.1.5
$$\underline{\omega} \cdot \underline{\nu} = 0 \qquad \text{for some } \underline{\nu} = (\nu_1, \ldots, \nu_\ell) \in \mathbb{Z}^\ell, \quad \underline{\nu} \neq \underline{0}. \tag{8.1.5}$$

A rotation vector $\underline{\omega}$ can be "far" from resonant in various senses. Here we shall consider, for simplicity, one of the strongest senses in which "far" can be understood and select the following vectors.

D8.1.1 **(8.1.1) Definition:** (Diophantine vectors)
A vector $\underline{\omega} \in \mathbb{R}^\ell$ such that there exist two constants $C, \tau > 0$ for which

e8.1.6
$$C|\underline{\omega} \cdot \underline{\nu}| > |\underline{\nu}|^{-\tau}, \qquad \textit{for all } \ \underline{\nu} \in \mathbb{Z}^\ell, \qquad \underline{\nu} \neq \underline{0}, \tag{8.1.6}$$

where $|\underline{\nu}| \stackrel{def}{=} \sum_{i=1}^{\ell} |\nu_i|$, *is called a* Diophantine vector *with constant C and exponent τ.*

Remark: It can be shown, see problem [8.1.7], that if $\tau > \ell - 1$ almost all $\underline{\omega} \in \mathbb{R}^\ell$ verify (8.1.6) for some suitable C (depending on $\underline{\omega}$), while in any ball of radius r in $\mathbb{R}^\ell$ the set of $\underline{\omega}$'s which do not satisfy (8.1.6) for given C and τ has volume $< B_{\tau,\ell} C^{-1} r^{\ell-1}$, for a suitable $B_{\tau,\ell} < \infty$.
In other words "most vectors $\underline{\omega} \in \mathbb{R}^\ell$ are far from resonances and satisfy a Diophantine condition" (hence they are rationally independent and enjoy the properties discussed in Section §2.2).

The key result is the following one.

P8.1.1 **(8.1.1) Proposition:** (KAM theorem)
Let $V(\underline{\alpha})$ be even and analytic*, let $K(\underline{A}) = \underline{A}^2/2J$, $J > 0$, and let $\underline{\omega}_0 = \underline{A}_0/J$ be a Diophantine vector with constant C and exponent τ. Then there exists $\varepsilon_0 > 0$ depending on C and τ and two functions, denoted $\underline{H}(\underline{\psi}, \varepsilon), \underline{h}(\underline{\psi}, \varepsilon)$, analytic for $(\varepsilon, \underline{\psi}) \in (-\varepsilon_0, \varepsilon_0) \times \mathbb{T}^\ell$ and divisible by ε, such that the points $(\underline{A}, \underline{\alpha})$ that lie on the tori*

e8.1.7
$$\begin{cases} \underline{\alpha} = \underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon), \\ \underline{A} = \underline{A}_0 + \underline{H}(\underline{\psi}, \varepsilon), \end{cases} \qquad \underline{\psi} \in \mathbb{T}^\ell, \tag{8.1.7}$$

form invariant surfaces for the evolution generated by the Hamiltonian

e8.1.8
$$H(\underline{A}, \underline{\alpha}) = K(\underline{A}) + \varepsilon V(\underline{\alpha}), \tag{8.1.8}$$

and their motion is simply $\underline{\psi} \to \underline{\psi} + \underline{\omega}_0 t$. Furthermore the time average of $\underline{A}(t) = \underline{A}_0 + \underline{H}(\underline{\psi} + \underline{\omega}_0 t, \varepsilon)$ is precisely $\underline{A}_0$.

Remarks: (1) This means that the Diophantine tori of the unperturbed
$N8.1.1$ system[1] are deformed into Diophantine tori of the perturbed system with the same rotation vector and with the same average action.
(2) The coincidence between the average action and the rotation vector times the moment of inertia J is a special property of Hamiltonians of the form (8.1.8). This property has been qualified with the adjective *twistless* in [Ga94a]. The interest of analyzing the KAM theorem for the special models in (8.1.8) has been remarked by Thirring, see p. 133 in [Th83], and for this reason they have been called *Thirring models.*
(3) The parity assumption on $V(\underline{\alpha})$, *i.e.*

$e8.1.9$
$$V(\underline{\alpha}) = \sum_{\underline{\nu}\in\mathbb{Z}^\ell} f_{\underline{\nu}}\, e^{i\underline{\nu}\cdot\underline{\alpha}}, \qquad \text{with } f_{\underline{\nu}} = f_{-\underline{\nu}}, \tag{8.1.9}$$

is a simplicity assumption that can be released as discussed in the problems for Section §(8.4).
(4) The proof presented here is elementary and it is based on the proof of Eliasson, [El96]. It is an interpretation of [El96] and provides a detailed analysis of the "cancellations" following [Ga94a].
(5) The proof will be discussed under slightly stronger assumptions in order to clarify its conceptual structure: the two extra assumptions that will be made below can be released as discussed in Section (8.2). The first assumption will be that $V(\underline{\alpha})$ is a *trigonometric polynomial*, *i.e.* the sum in (8.1.9) is, for some $N<\infty$, restricted to

$e8.1.10$
$$|\underline{\nu}| \le N < \infty. \tag{8.1.10}$$

The second assumption will be that the rotation vector $\underline{\omega}_0$ satisfies a property slightly stronger than the Diophantine inequality (8.1.7) that we shall introduce later (see definition (8.3.2)).

The proof will be split into several parts:
(i) We first derive the expression for the Taylor coefficients of the function $\underline{h}$ in powers of ε.
Calling $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$, $k\ge 1$, the k-th order coefficients of the Taylor expansion of $\underline{H}(\underline{\psi},\varepsilon), \underline{h}(\underline{\psi},\varepsilon)$ in powers of ε we derive recursive expressions for such coefficients supposing that $\underline{h}, \underline{H}$ exist, are analytic at $\varepsilon=0$ and have zero average over $\underline{\psi}$: we show that they are uniquely determined. In
$N8.1.2$ this way we obtain a formal power series[2] and we have to show its convergence to prove the proposition. Existence of the coefficients is a result due to Lindstedt and Newcomb in special cases and to Poincaré, [Po93] Ch. IX, in our case as well as in more general cases. The formal series is sometimes

[1] *i.e.* the tori on which run quasi-periodic motions with Diophantine rotation vectors.

[2] A formal power series of a field F is an infinite sequence $\{a_k\}_{k=1}^\infty$ over F; it is often written as $a_1\varepsilon + a_2\varepsilon^2 + a_3\varepsilon^3 + \ldots$, but with the understanding that no value is assigned to the symbol ε.

called the *Lindstedt series* although the same name is also given to a particular construction of the coefficients $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$ as sums of suitably defined "values" (cf. Section §(8.2)).
(ii) Subsequently, in Section §(8.2), we show that the coefficients of the Lindstedt series can be represented as sums of numerical values assigned to certain diagrams or graphs. Furthermore we check that the difficulty in proving convergence of the formal power series is only due to certain classes of diagrams whose members can give contributions to the coefficients of order k whose size could attain the order of a power of $k!$. This of course causes a problem, known as the *small divisors problem.*
(iii) In the only part which requires some work, Section §(8.3), we identify the diagrams which give too large contributions to the coefficients of order k and the way in which they can be collected together so that cancellations between their values become manifest. Finally we check that their sums is indeed far smaller than feared *a priori* so that the final bound on the coefficients of order k will be only of the order of an exponential in k, Section §(8.4), thus achieving the proof of proposition (8.1.1).

The first step is to show that the formal series for $\underline{h}, \underline{H}$ in powers of ε can be constructed and are well defined to all orders in ε.

L8.1.1 **(8.1.1) Lemma:** (Definition of the Lindstedt series)
There exists at most one solution $\underline{h}(\underline{\psi}), \underline{H}(\underline{\psi})$ of (8.1.7) which is analytic in ε near $\varepsilon = 0$, divisible by ε and with zero average over $\underline{\psi}$. The Taylor coefficients at $\varepsilon = 0$ of order $k \geq 1$ of such a solution are uniquely determined by (8.1.7). The resulting formal power series in ε will be called the Lindstedt series.

Proof: We suppose that $V(\underline{\alpha})$ is a *trigonometric polynomial* of some degree N: this means that the sum in (8.1.9) is restricted by (8.1.10).

Call $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$, $k \geq 1$, the k-th order coefficients of the Taylor expansion of $\underline{H}(\underline{\psi}, \varepsilon), \underline{h}(\underline{\psi}, \varepsilon)$ in powers of ε. By imposing that (8.1.7) with $\underline{\psi} = \underline{\omega}_0 t$ solve the equations of motion by substitution into the equations of motion $\dot{\underline{\alpha}} = J^{-1}\underline{A}$, $\dot{\underline{A}} = -\varepsilon \partial_{\underline{\alpha}} f(\underline{\alpha})$ we find the necessary (and sufficient) condition

e8.1.11
$$\begin{cases} (\underline{\omega}_0 \cdot \partial_{\underline{\psi}}) \underline{h}(\underline{\psi}, \varepsilon) = J^{-1} \underline{H}(\underline{\psi}, \varepsilon), \\ (\underline{\omega}_0 \cdot \partial_{\underline{\psi}}) \underline{H}(\underline{\psi}, \varepsilon) = -\varepsilon \partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)), \end{cases} \tag{8.1.11}$$

where $\partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)) = \partial_{\underline{\alpha}} V(\underline{\alpha})|_{\underline{\alpha} = \underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)}$. By equating the coefficients of the expansion of both sides in powers of ε we get immediately recursion relations for $\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$.

In fact the second set of equations in (8.1.11) shows that the only unknown is $\underline{h}(\underline{\psi}, \varepsilon)$ and that it is a solution of the "Hamilton–Jacobi equation":

e8.1.12
$$(\underline{\omega}_0 \cdot \partial_{\underline{\psi}})^2 \underline{h}(\underline{\psi}, \varepsilon) = -\varepsilon J^{-1} \partial_{\underline{\alpha}} V(\underline{\psi} + \underline{h}(\underline{\psi}, \varepsilon)), \tag{8.1.12}$$

because $\underline{H}(\underline{\psi}, \varepsilon)$ is immediately determined from the first equation in the

second set in (8.1.11). Clearly $\underline{h}(\underline{\psi},\varepsilon)$ is determined up to a constant which
N8.1.3 we choose to fix by requiring that $\underline{h}(\underline{\psi},\varepsilon)$ has zero average over $\underline{\psi}$.[3]

Writing $\underline{h}(\underline{\psi},\varepsilon)=\varepsilon\underline{h}^{(1)}(\underline{\psi})+\varepsilon^2\underline{h}^{(2)}(\underline{\psi})+\ldots$ and substituting into (8.1.12) we find for $k=1$

e8.1.13
$$\underline{h}^{(1)}(\underline{\psi})=-\sum_{\underline{\nu}\neq\underline{0}}\frac{iJ^{-1}\underline{\nu}}{(i\underline{\omega}_0\cdot\underline{\nu})^2}f_{\underline{\nu}}\,e^{i\underline{\nu}\cdot\underline{\psi}}. \tag{8.1.13}$$

And more generally one finds the k-th order from (8.1.11) simply by expanding in powers of $\underline{h}(\underline{\psi},\varepsilon)$ the function $\partial_{\underline{\alpha}}V(\underline{\psi}+\underline{h}(\underline{\psi},\varepsilon))$ and, subsequently, by expanding $\underline{h}(\underline{\psi},\varepsilon)$ in powers of ε. The result is, for $k>1$,

e8.1.14
$$\begin{aligned}&(\underline{\omega}_0\cdot\underline{\partial}_{\underline{\psi}})^2\,\underline{h}^{(k)}(\underline{\psi})=\\&=-J^{-1}\sum_{m=1}^{\infty}\frac{1}{m!}\partial_{\underline{\alpha}}^{m+1}V(\underline{\psi})\sum_{\substack{k_1,\ldots,k_m\ge1\\k_1+\ldots+k_m=k-1}}\prod_{r=1}^{m}\underline{h}^{(k_r)}(\underline{\psi}).\end{aligned} \tag{8.1.14}$$

Here we made an effort in trying to reduce the quantity of indices: therefore $\partial_{\underline{\alpha}}^{m+1}$ means a tensor with $m+1$ indices $j_0,j_1,\ldots,j_\ell$ in $\{1,2,\ldots,\ell\}$. If we imagine to use (8.1.14) to evaluate the component j_0 of the vector $\underline{h}^{(k)}(\underline{\psi})$ and set $\partial_j\equiv\partial_{\alpha_j}$, then $\partial_{\underline{\alpha}}^{1+m}V(\underline{\psi})$ denotes the tensor $\partial_{j_0}\partial_{j_1}\cdots\partial j_m V(\underline{\alpha})$ evaluated at $\underline{\alpha}=\underline{\psi}$; the indices $j_1,\ldots,j_m$ of the latter tensor must be contracted with the indices $j_1,\ldots,j_m$ determining the components of the vectors $\underline{h}^{(k_r)}(\underline{\psi})$, *i.e.* if $\partial_{j_1\cdots,j_m}\stackrel{def}{=}\partial_{j_1}\ldots\partial_{j_m}$ and $k_r\ge1$

e8.1.15
$$\begin{aligned}&\big(\partial_{\underline{\alpha}}^{m+1}V(\underline{\psi})\prod_{r=1}^{m}\underline{h}^{(k_r)}(\underline{\psi})\big)_{j_0}=\\&\qquad=\sum_{j_1,\ldots,j_m=1}^{\ell}\;\sum_{k_1+\ldots+k_m=k-1}\partial_{j_0j_1\ldots,j_m}V(\underline{\psi})\prod_{r=1}^{m}h_{j_r}^{(k_r)}(\underline{\psi}),\end{aligned} \tag{8.1.15}$$

while naturally the first index j_0 remains uncontracted as it represents the component of $\underline{h}^{(k)}(\underline{\psi})$ that we compute.

The equation (8.1.14) can be solved by Fourier transform: calling $\underline{\Phi}^{(k)}(\underline{\psi})$ the r.h.s. we see that the Fourier transforms of $\underline{h}^{(k)}$ and $\underline{\Phi}^{(k)}$ are related by

e8.1.16
$$-(\underline{\omega}_0\cdot\underline{\nu})^2\,\underline{h}_{\underline{\nu}}^{(k)}=\underline{\Phi}_{\underline{\nu}}^{(k)},\qquad \underline{\nu}\in\mathbb{Z}^\ell, \tag{8.1.16}$$

which can be solved if (and only if) $\underline{\Phi}_{\underline{0}}^{(k)}=\underline{0}$, *i.e.* if $\underline{\Phi}^{(k)}(\underline{\psi})$ has zero average over $\underline{\psi}$. The latter property holds for $k=1$ because in this case $\underline{\Phi}^{(1)}\equiv\underline{\partial}_{\underline{\alpha}}V$

[3] In fact the equation for $\underline{h}(\underline{\psi},\varepsilon)$ in (8.1.11) can be solved by Fourier transform only if $\underline{H}(\underline{\psi},\varepsilon)$ has zero average over $\underline{\psi}$ and, if it can be solved, it is then determined up to an additive constant equal to its average over $\underline{\psi}$.

is a gradient (cf. (8.1.13)). In general suppose that $\underline{h}^{(k)}(\underline{\psi})$ is trigonometric polynomial in $\underline{\psi}$ of degree $\le kN$, odd in $\underline{\psi}$, for $1 \le k < k_0 - 1$. Then we see immediately that the r.h.s. of (8.1.14), with $k = k_0$, is odd in $\underline{\psi}$ and of degree $\le k_0 N$. Therefore the r.h.s. of (8.1.14) has zero average in $\underline{\psi}$, and (8.1.14) can be also solved for $k = k_0$. This means that the equation for $\underline{h}^{(k_0)}(\underline{\psi})$ can be solved for all $k_0 \ge 1$, and its solution is a trigonometric polynomial of degree $\le k_0\, N$ which is odd in in $\underline{\psi}$, if $\underline{h}^{(k_0)}(\underline{\psi})$ is uniquely determined by imposing that its average over $\underline{\psi}$ vanishes. ■

Problems for §8.1

Q8.1.1 **[8.1.1]:** (*A resonance phenomenon*)
Show that for any $\omega \in \mathbb{R}$ the two-dimensional system described by the Hamiltonian $K(\underline{A}) + \varepsilon V(\underline{\alpha})$, with $K(\underline{A}) = \frac{1}{2}\underline{A}^2, V(\alpha) = \cos(\alpha_1 - \alpha_2)$, admits only two periodic motions with rotation vector (ω, ω) corresponding to the stationary points of a suitable function (this is an example of a *resonance*). (*Hint:* Use the (canonical) change of variable $\beta_1 = \alpha_1 + \alpha_2$, $\beta_1 = \alpha_1 - \alpha_2$ and similar for $\underline{A}$. Note that this is a simple example of the general set-up of the forthcoming Section §(9.1) about low-dimensional tori, of which the periodic orbits are an (easy) subcase.)

Q8.1.2 **[8.1.2]:** (*Rationally independent* 3*–vectors,* [Ch99])
Suppose that the two equations $n^3 = an \pm b$, with a, b integers, do not admit integer solutions, and let ω be real and a root of the equation $\omega^3 = a\omega + b$. Show that the vector $\underline{w} = (1, \omega, \omega^2)$ has rationally independent components. The case $a = b = 1$ defines a real root ω which is called the *spiral mean.* (*Hint:* If not $\underline{\nu} \cdot \underline{w} = \nu_1 + \nu_2\,\omega + \nu_3\,\omega^2 = 0$ with $\underline{0} \ne \underline{\nu} \in \mathbb{Z}^3$ and ω would be a quadratic number, *i.e.* a number of the form $\omega = (x + \sqrt{y})$ with x, y rational. Since ω is also a solution of the third order equation $\sqrt{y}$ must be rational: in fact if $\sqrt{y}$ is irrational then the equation $\omega^3 = a\omega + b$ implies

$$x^3 + 3xy - ax - b + \sqrt{y}\,(y + 3x^2 - a) = 0,$$

and necessarily $y + 3x^2 - a = 0$ and $x^3 + 3xy - ax - b = 0$, which means that $8x^3 - 2ax + b = 0$ admits a rational root $x = p/q$, with p, q relatively prime, *i.e.* $8p^3 - 2apq^2 + bq^3 = 0$. This says that $(2p)^3$ is divisible by q^2 so that $q = 1$ or $q = 2$ and $n = \frac{2p}{q}$ is integer: therefore $n^3 = an - b$ admits an integer solution, in contradiction with the hypothesis. Hence ω has to be rational. If $\omega = p/q$, p, q relatively prime integers, then $p^3 = a\,p\,q^2 + q^3$ and q^2 divides p^3, hence $q = 1$ and p is an integer solution of $n^3 = a\,n + b$, which again contradicts the hypothesis.)

Q8.1.3 **[8.1.3]:** (*Tartaglia's formula and rational independence*)
Let $3\,p, 2\,q$ be integers with $q^2 + p^3 > 0$ and suppose that the equations $z^3 = 3pz \pm 2q$ do not admit integer solutions. Show that $\underline{w} = (1, \omega, \omega^2)$ is a rationally independent vector if $\omega = (q + (p^3 + q^2)^{\frac{1}{2}})^{\frac{1}{3}} + (q - (p^3 + q^2)^{\frac{1}{2}})^{\frac{1}{3}}$. Show that, therefore, $\underline{w} = (1, 2^{\frac{1}{3}}, 4^{\frac{1}{3}})$ is a rationally independent vector. Check that the spiral mean is

$$\omega = \sqrt[3]{\tfrac{1}{2} + \tfrac{1}{2}\sqrt{\tfrac{23}{27}}} + \sqrt[3]{\tfrac{1}{2} - \tfrac{1}{2}\sqrt{\tfrac{23}{27}}}.$$

(*Hint:* By Tartaglia's formula, divulged by Cardano, ω is a real root of $\omega^3 = -3p\omega + 2q$: hence one applies problem [8.1.2]. Note that $2^{1/3}$ is a root of $\omega^3 = 2$.)

Q8.1.4 **[8.1.4]:** (*An example of a Diophantine* 3*–vector,* [Ch99])
To obtain an extension of problem [2.2.3] let

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \qquad N^k \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} p_k \\ q_k \\ r_k \end{pmatrix} \stackrel{def}{=} \underline{v}_k$$

Show that $\rho_j = p_j, q_j, r_j$ verify the recursion $\rho_j = \rho_{j-2} + \rho_{j-3}$, $j \ge 3$. Let $\lambda > 1, \lambda_+, \lambda_-$ be the three eigenvalues of N, $|\lambda_+| = |\lambda_-| = \lambda^{-1/2}$ and let $\underline{w}_0, \underline{w}_+, \underline{w}_- = (1, \lambda_j, \lambda_j^2)$, $j = 0, \pm$ be the respective eigenvectors (note that $\lambda_j^3 = \lambda_j + 1$). Let N^T be the transposed of N. Given $\underline{\nu} \in \mathbb{Z}^3$ define, if it exists, k so that for $0 \le h \le k$

$$|(N^T)^h \underline{\nu} \cdot \underline{w}_0| \le \max_{\pm} |(N^T)^h \underline{\nu} \cdot \underline{w}_\pm|, \qquad |(N^T)^{(k+1)} \underline{\nu} \cdot \underline{w}_0| > \max_{\pm} |(N^T)^{(k+1)} \underline{\nu} \cdot \underline{w}_\pm|,$$

otherwise set $k = -1$. Check that $|\underline{\nu}| > B |\lambda_\pm|^{-k} = B\,\lambda^{\frac{k}{2}}$ for some $B > 0$ (which only depends on the matrix N); deduce that therefore $\underline{w}_0$ is a Diophantine vector with exponent $\tau = 2$. (*Hint:* The recursion can be derived from the remark that $N^3 \underline{v}_0 = \underline{v}_0 + N \underline{v}_0$. To check the Diophantine property remark that $1 \le |(N^T)^k \underline{\nu}|$ because N has no zero eigenvalue; hence by the definition of k and for a suitable $b > 0$ depending on the basis $\underline{w}_0, \underline{w}_\pm$, one has $1 \le |(N^T)^k \underline{\nu}| \le b \max_\pm |\underline{w}_\pm \cdot (N^T)^k \underline{\nu}| \le b' \, |\underline{\nu}| \, \lambda^{-\frac{k}{2}} \max_\pm |\underline{w}_\pm|$ (in fact in R^3 all metrics are equivalent). Therefore, by the definition of k, one has $|\underline{w}_0 \cdot \underline{\nu}| = \lambda^{-(k+1)} |\underline{w}_0 \cdot (N^T)^{(k+1)} \underline{\nu}| \ge b'' \lambda^{-(k+1)} |(N^T)^{(k+1)} \underline{\nu}| \ge b'' \lambda^{-(k+1)} \ge b''' |\underline{\nu}|^{-2}$.)

Q8.1.5 **[8.1.5]:** (*Hyperbolicity, Anosov maps of* $\mathbb{T}^3$ *and quasi periodicity*)
Show that the matrix T in problem [4.3.6] yields a natural example of a algebraic hyperbolic map of the torus $\mathbb{T}^3$ and of a Diophantine 3 vector.

Q8.1.6 **[8.1.6]:** (*More examples of Diophantine* 3*–vectors,* [Ch99]: an extension of problem [2.2.3])
Show that the example of problem [8.1.4] can be extended by considering 3×3 integer entries matrices N with only one eigenvalue with modulus greater than 1 and two others with modulus less than 1 and with eigenvectors with rationally independent components. Find a few examples by studying matrices whose characteristic equation has roots that can be discussed by the techniques of problems [8.1.2] and [8.1.3]. The vectors constructed in this way are all in a class called *Pisot–Vijayaraghavan Diophantine* 3*–vectors.* (*Hint:* For instance consider matrices obtained from N in problem [8.1.4] by replacing the last row $(0, 1, 0)$ by $(0, p, q)$.)

Q8.1.7 **[8.1.7]:** (*Genericity of Diophantine vectors*)
Given C and $\tau > \ell - 1$ show that the volume of the points $\underline{\omega} \in \mathbb{R}^\ell$ with $|\underline{\omega}| < r$ that do not verify property (8.1.6) is bounded by $B_\tau C^{-1} r^{\ell-1}$, for a suitable constant B_τ depending on τ. Show that this implies that the set of $\underline{\omega}$ which do not satisfy a Diophantine property with a prefixed exponent $\tau > \ell - 1$ and some constant $C < \infty$ has complement of zero volume in $\mathbb{R}^\ell$. (*Hint:* Consider the slab $|\underline{\omega}_0 \cdot \underline{\nu}| < C^{-1} |\underline{\nu}|^{-\tau}$, $\underline{\nu} \ne \underline{0}$, in the sphere $|\underline{\omega}_0| < r$. Note that its volume is bounded proportionally to $C^{-1} |\underline{\nu}|^{-\tau-1} r^{\ell-1}$ hence the set of points *not* verifying (8.1.6) has volume bounded proportionally to $C^{-1} r^{\ell-1}$. The theorem of Borel–Cantelli, for instance, then implies the second statement.)

Q8.1.8 **[8.1.8]:** (*Examples of isochronous systems*, [CF02])
Consider in $\mathbb{C}^n$ the equation $\dot{\underline{z}} = \Lambda \underline{z} + \underline{Q}(\underline{z})$, where Λ is a diagonal matrix with diagonal $\lambda_1, \ldots, \lambda_n$ and $\underline{Q}$ is a polynomial with a second order 0 at the origin. Suppose $\lambda_j \equiv i$. Prove that all solutions with initial data $\underline{z}_0$ small enough are periodic with period 2π. (*Hint:* Define $\tau = (e^{it} - 1)/i$ and set $\underline{\zeta}(\tau) \stackrel{def}{=} e^{-it} \underline{z}(t)$: then $\underline{\zeta}(\tau)$ satisfies the equation $\frac{d\underline{\zeta}(\tau)}{d\tau} = (1 + i\tau)^{-2} \underline{Q}\big((1 + i\tau)\, \underline{\zeta}(\tau)\big)$. The latter equation has solutions holomorphic in τ for $|\tau| < R$ provided the initial data are small enough. Therefore choosing $R > 2$ we see that the solution is $z(t) = e^{it} \zeta((e^{it} - 1)/i)$ which is periodic in t.)

Q8.1.9 **[8.1.9]:** Find the condition under which the equation for $\underline{\zeta}$ in problem [8.1.8] is autonomous.

Q8.1.10 **[8.1.10]:** (*Perturbation expansion without divisor problems and isochrony*)
Derive the result of problem [8.1.8] by a perturbation expansion showing the existence of a function $\underline{h}(\underline{\zeta})$ holomorphic near $\underline{0}$ and vanishing to second order such that the change

of variables $\underline{z} = \underline{\zeta} + \underline{h}(\underline{\zeta})$ transforms the differential equation into $\dot{\underline{\zeta}} = \Lambda\underline{\zeta}$. (*Hint:* No small divisors problems arise in the perturbative determination of $\underline{h}$.)

Q8.1.11 **[8.1.11]:** Consider in $\mathbb{C}^n$ the equation $\dot{\underline{z}} = \Lambda\underline{z} + \underline{Q}(\underline{z})$, where Λ is a diagonal matrix with diagonal $\lambda_1, \ldots, \lambda_n$ and $\underline{Q}$ is a polynomial with a second order 0 at the origin. Suppose that $\min|\underline{\nu}\cdot\underline{\lambda} - \lambda_j| > c > 0$ with the minimum evaluated over all *non-negative integer components* vectors $\underline{\nu}$ with $\sum_j \nu_j \ge 2$ and over all $j = 1, \ldots, n$. Show the existence of a function $\underline{h}(\underline{\zeta})$ holomorphic near $\underline{0}$ and vanishing to second order such that the change of variables $\underline{z} = \underline{\zeta} + \underline{h}(\underline{\zeta})$ transforms the differential equation into $\dot{\underline{\zeta}} = \Lambda\underline{\zeta}$. Hence if $\mathrm{Re}\,\lambda_j \le 0$ all its solutions with small enough initial data are asymptotic to a quasi-periodic solution. Exhibit cases in which all solutions are quasi-periodic. (*Hint:* No small divisors problems arise in the perturbative determination of $\underline{h}$. Consider in particular the cases $\lambda_j = i + \varepsilon_j$, with $\varepsilon_j \le 0$, and the cases $\lambda_j = i + i\varepsilon_j$, with $|\varepsilon_j|$ small enough.)

§8.2 Graphs and diagrams for the Lindstedt series

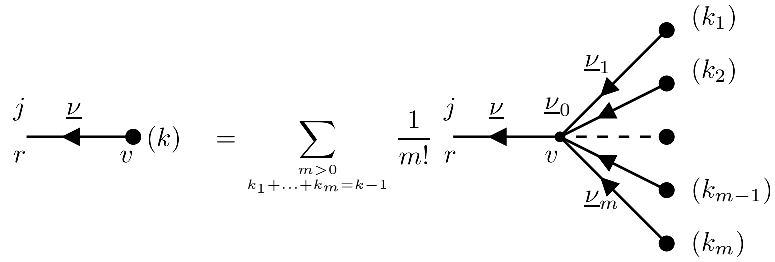
The analysis of the previous section provides us with a recursive algorithm to evaluate a formal power series solution to our problem. However it is very convenient to interpret the algorithm in terms of graphs and diagrams. An explicit construction of the values of the coefficients $\underline{h}^{(k)}(\underline{\psi})$ of the Lindstedt series can be obtained by simply "iterating" (8.1.15) until only $\underline{h}^{(1)}(\underline{\psi})$, cf. (8.1.13), appears. We get what we shall call a *tree representation* of $\underline{h}^{(k)}(\underline{\psi})$.

The construction is most easily performed through the diagrammatic expansion and therefore we dedicate some care to its detailed discussion. Note that the computation of $\underline{h}^{(k)}$ involves the inversion of the operator $(\underline{\omega}_0\cdot\partial_{\underline{\psi}})^2$ which is easily done if instead of $\underline{h}^{(k)}(\underline{\psi})$ we study its Fourier transform $\underline{h}^{(k)}_{\underline{\nu}}$ as this involves simply a division by $(\underline{\omega}_0\cdot\underline{\nu})^2$, cf. (8.1.16). Hence we rewrite (8.1.14) as

$$\underline{h}^{(k)}_{\underline{\nu}} = \frac{1}{J(\underline{\omega}_0\cdot\underline{\nu})^2}\sum_{m=1}^{\infty}\sum_{\underline{\nu}_0+\underline{\nu}_1+\ldots\underline{\nu}_m=\underline{\nu}}\frac{1}{m!}\,i\underline{\nu}_0\,f_{\underline{\nu}_0}\sum_{\substack{k_1,\ldots,k_m\ge1\\ k_1+\ldots+k_m=k-1}}\prod_{s=1}^{m} i\underline{\nu}_0\cdot\underline{h}^{(k_s)}_{\underline{\nu}_s}. \tag{8.2.1}$$

e8.2.1

And the latter relation can be represented by Fig.(8.2.1), where the l.h.s is a symbol for $\underline{h}^{(k)}_{\underline{\nu}}$ and in the r.h.s. we have a "simple tree" consisting of a "root" r, a "root branch" $\lambda_v \equiv rv$ coming from the "node" (or "vertex") v and $m = m_v$ branches "entering v".



F8.2.1 **Fig.(8.2.1)** Diagrammatic interpretation of (8.2.1). The tensor labels are not marked.

The length of the branches and the angles at which they are drawn (with respect to the sides of the present sheet of paper) are irrelevant.

The node v symbolizes the tensor with entries

e8.2.2
$$\Phi_{v;j_0,j_1,\ldots,j_m}=\frac{1}{m_v!}J^{-1}(i\underline{\nu}_0)_{j_0}f_{\underline{\nu}_0}\prod_{s=1}^{m}(i\underline{\nu}_0)_{j_s}, \tag{8.2.2}$$

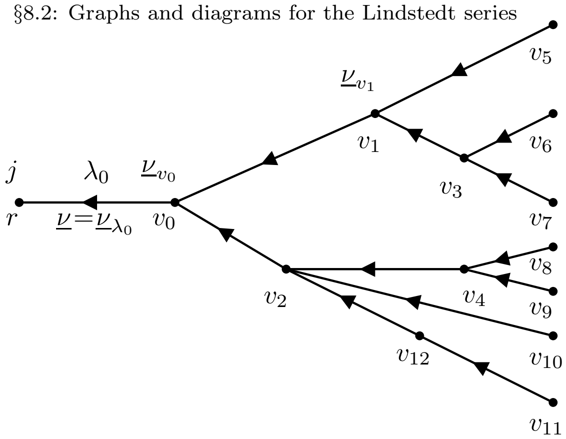
The line entering from the node v in the r.h.s. represents the *propagator*, *i.e.* the matrix

e8.2.3
$$\frac{\delta_{jj_0}}{(\underline{\omega}_0\cdot\underline{\nu})^2} \tag{8.2.3}$$

The line exiting from the bullet of the l.h.s represents $\underline{h}^{(k)}_{\underline{\nu},j}$; the branches exiting from the bullets of the r.h.s. with label (k_s), see figure (8.2.1), represent $\underline{h}^{(k_s)}_{\underline{\nu}_s,j'_s}$.

We should imagine that the labels $j_0,j_1,\ldots,j_m$ are affixed on the line exiting the node v and, respectively, on the lines entering the node v (and close to v). The labels $j'_1,\ldots,j'_m$ should be imagined affixed on the lines exiting the bullets on the r.h.s. of the drawings in figure (8.2.1). And the root label j can be imagined affixed over the root of the tree. Note that in such a way we associate a pair of labels (j'_v,j_v) to each branch $v'v$ of the tree.

The labels $j_0,j_1,\ldots,j_m,j'_1,\ldots,j'_m$ will be called *tensor labels* and the label j *root label*. However the drawing in the figure does not carry the tensor labels explicitly: the tensor labels on the same line are meant to be "contracted" (*i.e.* set to equal values) and summed over their possible values in $\{1,\ldots,\ell\}$.



$F8.2.2$ **Fig.(8.2.2)** A tree ϑ with $m_{v_0} = 2, m_{v_1} = 2, m_{v_2} = 3, m_{v_3} = 2, m_{v_4} = 2, m_{v_{12}} = 1$ and $k = 13$, and some decorations. Only two mode labels and two momentum labels are explicitly marked on the lines λ_0, λ_1; the number labels, distinguishing the branches, are not shown. The arrows represent the partial ordering on the tree.

At this point, if one notes that (8.1.14) is multilinear in the $\underline{h}^{(k_s)}_{\underline{\nu},j'_s}$ (*i.e.* of degree 1 in each of them), it is clear that we can just replace each of the branches exiting from a bullet with the same graphical expression in the r.h.s. of the above figure. And so on, until the labels (k) on all the branches exiting from a bullet ("top branches") become equal to 1. In this case the branches will represent $\underline{h}^{(1)}_{\underline{\nu}_{j'}} = (i\underline{\nu}_{j'})\, J^{-1}\, (\underline{\omega}_0 \cdot \underline{\nu}_{j'})^{-2} f_{\underline{\nu}_{j'}}$ for some j'.

Thus we have represented our $\underline{h}^{(k)}_{\underline{\nu}}$ as a "sum over trees", with k branches and k nodes (we shall not regard the root as a node), of suitable "tree values".

The combinatorial factors (coming from the factorial in (8.2.2)) can be simplified if we decide to distinguish the lines of the trees with k branches by affixing on them a *number label* taking values from 1 to k. The set of labels attached to the trees, other than the tensor labels, will be called *decorations.*

We shall regard as identical two decorated trees which can be overlapped (with all labels matching) by adjusting the lengths of the lines and by pivoting the branches merging into a node around it.

The rule to construct $\underline{h}^{(k)}_{\underline{\nu}}$ can be summarized as follows.

$D8.2.1$ **(8.2.1) Definition:** (Trees and tree values)
To evaluate the Fourier transform $h^{(k)}_{\underline{\nu},j}$ *of* $h^{(k)}_j(\underline{\psi})$ *the following definitions are given.*
(i) A tree ϑ *consists of a family of* oriented branches *(i.e. oriented segments or arrows)* $\Lambda(\vartheta)$ *numbered from* 1 *to* k *arranged to form a (rooted) tree as in Fig.(8.2.2) (i.e. all oriented towards the root* r*); the initial points of the branches are the* nodes $V(\vartheta)$. *A partial ordering relation, denoted* $\preceq$, *is induced by the lines orientations between the lines and between the nodes.*
(ii) At each node v *we attach a* mode label $\underline{\nu}_v \in \mathbb{Z}^\ell$, $|\underline{\nu}_v| \le N$ *(cfr. (8.1.10)). At the root* r *we attach a unit vector* $\underline{\nu}_r$ *in a selected direction* $j \in (1,\ldots,\ell)$. *The* order *of the tree will be the number* k *of nodes or*

N8.2.1 *the (equal) number of branches.*[1]
(iii) At each branch $\lambda_v \equiv v'v$, connecting a node v with the following node v', we attach a momentum

e8.2.4
$$\underline{\nu}_{\lambda_v} \equiv \sum_{w \preceq v} \underline{\nu}_w, \tag{8.2.4}$$

i.e. we can think that into each node "enters" a momentum $\underline{\nu}_v$ which then "flows" toward the root so that the the momentum flowing through the branch
N8.2.2 *emerging from v is $\underline{\nu}_{\lambda_v}$.*[2]
(iv) Near each node v we write $m_v + 1$ tensor labels *$j_{v,0}, j_{v,1}, \ldots, j_{v,m_v}$ in $1, \ldots \ell$: the first on the line exiting the node v and the remaining m_v on the*
N8.2.3 *lines entering the node v. We imagine such labels written very close to v.*[3]
(v) With each line λ joining two nodes $v'v$ or joining the root r and the first node before it v_0 we associate a propagator *matrix $\delta_{ij}(\underline{\omega}\cdot\underline{\nu}_\lambda)^{-2}$,* cf. *(8.2.3), whose indices i, j are contracted, respectively, with the labels $j_{v'}$ and j_v on the line and close to the nodes $v'v$ (defined in (ii) above). In the case of the root line the label j_r is taken equal to j, the index of the component of $h^{(k)}_j$ that we want to study (and no sum on the root label is performed).*
(vi) The mode labels and the number labels "decorating" the tree will be called also decorations *(i.e. the labels attached to the tree and the decorations are synonymous below). In particular the component indices j of points (iv) and (v) are not decoration since they are contracted, i.e. summed over.*
(vii) With each k–th order tree we associate a combinatorial factor $k!^{-1}$. We shall call $\Theta_{k,\underline{\nu},j}$ the set of all decorated trees of order k with momentum $\underline{\nu}_{\lambda_0} = \underline{\nu}$ flowing through the root branch and such that for each branch λ one has $\underline{\nu}_\lambda \neq \underline{0}$ and carrying a root label j.

The above combinatorial analysis has led us to a final result which is, perhaps surprisingly, very simple.

L8.2.1 **(8.2.1) Lemma:** (Tree representation of the Lindstedt series)
Given the above definition $\underline{h}^{(k)}_{\underline{\nu},j}$ will have the following representation for the j–th component of $\underline{h}^{(k)}(\underline{\psi}) = \sum_{\underline{\nu}\in\mathbb{Z}^\ell} \underline{h}^{(k)}_{\underline{\nu}} e^{i\underline{\nu}\cdot\underline{\psi}}$:

e8.2.5
$$h^{(k)}_{\underline{\nu},j} = -i\frac{1}{k!}\sum_{\vartheta\in\Theta_{k,\underline{\nu},j}} \Big(\prod_{v\in V(\vartheta)} J^{-1} f_{\underline{\nu}_v}\Big) \prod_{\lambda=(v'v)\in L(\vartheta)} \frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}_0\cdot\underline{\nu}_{\lambda_v})^2}, \tag{8.2.5}$$

where $\Theta_{k,\underline{\nu},j}$ is defined in item (vii) of definition (8.2.1).

Remarks: (1) The factor $\underline{\nu}_r \cdot \underline{\nu}_0$ in the contribution from the factor corresponding to the root line $\lambda_0 \equiv rv_0$ has to be interpreted by taking $\underline{\nu}_r$ to be

---

[1] In Fig.(8.2.2) the mode labels are marked only above the nodes v_0 and v_1; the number labels and the other labels are not marked explicitly, for simplicity.

[2] In Fig.(8.2.2) the momentum labels are marked only on the root branch λ_0 and on the branch λ_1.

[3] In Fig.(8.2.2) the tensor labels are not marked.

the unit vector in the direction j (see item (v) of definition (8.2.1)).
(2) The introduction of the number labels that distinguish the tree lines of course greatly increases the number of terms contributing to a given order k because there are many terms that give the same contribution simply because the value of a tree, defined as

e8.2.6
$$\mathrm{Val}\,(\vartheta)=-i\frac{1}{k!}\Big(\prod_{v\in V(\vartheta)}J^{-1}f_{\underline{\nu}_v}\Big)\prod_{\lambda=(v'v)\in L(\vartheta)}\frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}_0\cdot\underline{\nu}_{\lambda_v})^2},\tag{8.2.6}$$

does not depend on the number labels. However the advantage is that the combinatorial factor is $k!^{-1}$ for all trees, while using trees without number labels would require a combinatorial factor $\prod_v m_v!^{-1}$ which depends on the structure of each tree: the advantage being that one would need to consider far less trees.

The above formulae remain valid also in the case of a perturbation $V(\underline{\alpha})$ which is not even: however in such case one needs to check (a result by Poincaré, [Po93]) that this is correct because it does not follow immediately from a parity property that the graphs in which a zero current flows have to be discarded; see problem [8.4.8] in Section §(8.4). The formulae also extend to the case in which $V(\underline{\alpha})$ is analytic in $\underline{\alpha}$.
It is sometimes useful to rewrite (8.2.6) more compactly (however somewhat more symbolically) as

e8.2.7
$$\mathrm{Val}(\vartheta)=\frac{1}{k!}\Big(\prod_{v\in V(\vartheta)}F_v\Big)\Big(\prod_{\lambda\in\Lambda(\vartheta)}G_\lambda\Big),\qquad h^{(k)}_{\underline{\nu},j}=\sum_{\vartheta\in\Theta_{k,\underline{\nu},j}}\mathrm{Val}\,(\vartheta)\tag{8.2.7}$$

where the symbols are introduced in the following definition.

D8.2.2 **(8.2.2) Definition:** (Propagator and node factors)
Given a node v and a line λ the node factor F_v *(a tensor) and the* propagator G_λ *(a $\ell\times\ell$–matrix proportional to the identity) will be defined, respectively as*

e8.2.8
$$F_v=J^{-1}(i\underline{\nu}_v)^{m_v+1}f_{\underline{\nu}_v},\qquad G_\lambda=\frac{1}{(\underline{\omega}_0\cdot\underline{\nu}_\lambda)^2}.\tag{8.2.8}$$

Remarks: (1) The above analysis is well known in graph theory. Since the order of the tree is k, equal to the number of nodes, the above trees are analogous to Feynman graphs of a perturbation theory and the evaluation of the values "corresponds" to the Feynman rules for the KAM problem (in fact a literal meaning could be given to the mentioned analogy, [Ga01]).
(2) In the general case the representation in (8.2.5) is known as the *Lindstedt series* and it was introduced, at low orders only, in Celestial Mechanics problems by Lindstedt and Newcomb (independently); it has been extended and studied to all orders by Poincaré: about the name see comment following (8.1.10). The difficulty in the extension, solved by Poincaré, is that of proving that the algorithm makes sense to all orders, *i.e.* the equations for

$\underline{H}^{(k)}(\underline{\psi}), \underline{h}^{(k)}(\underline{\psi})$ can be solved by Fourier transform "without ever dividing by 0". The latter difficulty is absent here because of the parity assumption on $V(\underline{\alpha})$ which excludes that $(\underline{\omega}_0 \cdot \partial_{\underline{\psi}})\underline{H}^{(k)}(\underline{\psi})$, as given by (8.1.11), has non-zero average over $\underline{\psi}$ (being odd), see problems [8.4.8] and [8.4.9] for the non-even cases.

Formula (8.2.5) would provide immediately a proof of proposition (8.1.1), the KAM theorem, if certain trees were not present. The idea is that the values of the "unwanted" trees cancel each other to the extent that their sums behave well enough not to spoil the bounds. *See problem [8.2.4] for an example of an "unwanted" tree* with value of extremely large size.

To see this, *for the purpose of illustration*, we shall first restrict the sum in (8.2.7) to a sum over trees which, besides the property that for all branches one has $\underline{\nu}_\lambda \ne \underline{0}$ (as expressed by the definition of the set $\Theta_{k,\underline{\nu},j}$ in equation (8.2.7)), satisfy the property

[P] $\underline{\nu}_{\lambda_v} \ne \underline{\nu}_{\lambda_{v'}}$ *for all pairs of comparable nodes* v, v' *(not necessarily next to each other in the tree order), with* $v' \preceq v \preceq v_0$.

This property is remarkable because of the following lemma.

*L*8.2.2 **(8.2.2) Lemma:** (Summation of values of trees with property [P]) *Consider the formal power series in* ε *obtained by considering the sum of the trees values of the trees in* $\Theta_{k,\underline{\nu},j}$ *which verify property [P] above. Then the series has a positive convergence radius.*

Remark: The purpose of this lemma is to make clear where the problem of proving proposition (8.1.1), *i.e.* the KAM theorem, is really located. To prove the theorem we shall have to understand how to bound the sum of the values of tree graphs which do not satisfy property [P].

Proof: There are at most $2^{2k}k!$ trees (see problem [8.2.2]) and the contraction of the tensors labels F_v, cf. (8.2.8), gives at most ℓ^k terms, while the node momenta $\underline{\nu}_v$ can be chosen in a number of ways bounded by $(2N+1)^{\ell k} < (3N)^{\ell k}$. Therefore, if $f_0 = \max_{\underline{\nu}} |f_{\underline{\nu}}|$,

*e*8.2.9
$$\begin{aligned}\sum_{\underline{\nu}} \Big|\underline{h}^{(k)}_{\underline{\nu}}\Big| &\le (3N)^{\ell k} 2^{2k} \ell^k \frac{f_0^k C^{2k}}{J^k} N^{2k-1} \max_{\vartheta\in\Theta_{k,\underline{\nu},j}} \prod_{\lambda\in V(\vartheta)} (C\underline{\omega}_0\cdot\underline{\nu}_\lambda)^{-2} \\ &\le (f_0 C^2 J^{-1})^k N^{(\ell+2)k-1} (4\ell 3^\ell)^k M,\end{aligned} \tag{8.2.9}$$

where M is an estimate of the indicated maximum which is over the k-th order trees ϑ verifying property [P] above. Hence the whole problem is reduced to find a value for M, a "small divisors problem".

Let q be large: then by the Diophantine condition in (8.1.6) one has $C|\underline{\omega}_0 \cdot \underline{\nu}| \ge q^{-1}$ if $0 < |\underline{\nu}| \le q^{1/\tau}$. We say that *the harmonic with Fourier label* $\underline{\nu} \in \mathbb{Z}^\ell$ *is "q–singular"* if $C|\underline{\omega}_0 \cdot \underline{\nu}| < q^{-1}$ and the following (extension) of a lemma by Bryuno holds for trees of degree k verifying property [P] above. Fixed $q \ge 1$ let $N(k,q)$ be the number of "q–singular branches" (*i.e.* of branches corresponding to q–singular harmonics) in a tree ϑ with k nodes. Then we can have recourse to the following lemma

$L8.2.3$ **(8.2.3) Lemma:** (Simple Bryuno–Siegel bound)
The number $N(k,q)$ of "q–singular branches" of a tree of order k in the above sense is equal to

$e8.2.10$
$$N(k,q) \le \text{const}\, \frac{k}{q^{1/\tau}}, \tag{8.2.10}$$

and the constant could be taken $2N2^{3/\tau}$.

Remark: (1) We shall not prove here the lemma as the present argument is being carried over only for illustration purposes (in any event (8.2.10) and the value of the constant are an immediate corollary of the proof of the lemma discussed and proved below).
(2) The intuition behind (8.2.10) is very simple. In order to achieve a branch momentum $\underline{\nu} = \underline{\nu}_{\lambda_v}$ with $C\underline{\omega}_0 \cdot \underline{\nu}$ of size q^{-1} one needs at least $|\underline{\nu}| \ge q^{1/\tau}$, *i.e.* the node v must be preceded by *at least* $N^{-1}q^{1/\tau}$ nodes. Once a q–singular branch has been generated, the branches following it will have non-q–singular momentum and we must collect *about as many new nodes* to generate a second q–singular branch and so on. Since the total number of nodes is k it follows that the number of q–singular branches is bounded proportionally to $k/q^{1/\tau}$. The actual estimate of the constant in (8.2.10) is irrelevant for our immediate purposes.

Assuming lemma (8.2.3) we can continue the proof of lemma (8.2.2). We fix an exponentially decreasing sequence γ^n, $n = 1, 0, -1, -2, \ldots$; we choose $\gamma = 2$. The number of 2^{-n}–singular harmonics among the line momenta which are not $2^{-(n-1)}$–singular is bounded by $2N2^{3/\tau}\, k\, 2^{-n}$, (being trivially bounded by the number of 2^{-n}–singular harmonics!). Hence we can take as a bound on M in (8.2.9) the quantity

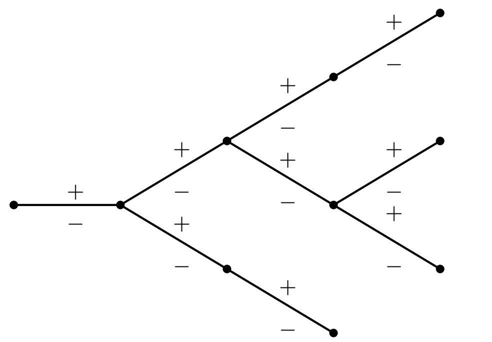
$e8.2.11$
$$\prod_{\lambda\in\Lambda(\vartheta)} \frac{1}{(C\underline{\omega}_0 \cdot \underline{\nu}_\lambda)^2} \le \prod_{n=-\infty}^{-1} 2^{-(n-1)4N2^{3/\tau}2^{n/\tau}k} = e^{cN\tau k} \equiv M, \tag{8.2.11}$$

where $c > 0$ is a suitable constant (see problem [8.2.5]); therefore the series for the approximation to $h^{(k)}_{\underline{\nu}}$, that we are considering because of the extra restriction in the sum equation (8.2.4), has radius of convergence in ε bounded below by ε'_0 given by

$e8.2.12$
$$(\varepsilon'_0)^{-1} = (f_0 C^2 J^{-1})^k N^{(\ell+2)} (4\ell 3^\ell) e^{cN\tau}. \tag{8.2.12}$$

Hence the proof of the lemma (8.2.2) is concluded. ■

Lemma (8.1.1) shows that we must understand the contributions to the value of $\underline{h}^{(k)}$ coming from the tree graphs that violate property [P] above.



$$W = \{+ + + + - - + + - + - - - + + - - -\}$$

F8.2.3 **Fig.(8.2.3)** A rooted tree and the corresponding random walk W.

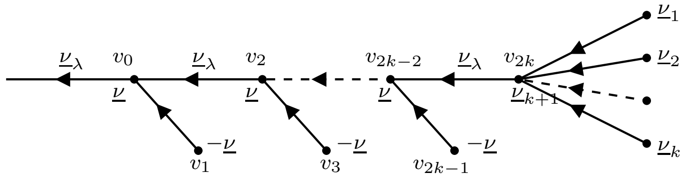
Problems for §8.2

Q8.2.1 **[8.2.1]:** Prove that the number of trees of order k with assigned labels $\{m_v\}$ is bounded by $(k-1)!/\prod_v m_v!$ (*Hint:* It is just problem [7.1.4]).

Q8.2.2 **[8.2.2]:** (*Counting tree graphs*)
Show that the number of elements in $\Theta_{k,\nu}$ is bounded by $k!2^{2k}$. (*Hint:* The number of non-numbered trees is bounded by the number of random walks on the one-dimensional lattice with $2k-1$ steps, that is by 2^{2k}. In fact we can imagine walking on the tree starting at the root and marking $+1$, *i.e.* a step forward, every time we pass a node; once we reach an endpoint we imagine reversing the direction and continue walking on the "other side" of the branch (we imagine that each branch has an upper and a lower side) marking a -1, *i.e.* a step backward until we reach a non trivial node where we reverse the walk direction etc.; see Fig.(8.2.3). At the end we reach the root and the sum of the number of $+1$'s equals that of the -1 and the sequence of ±1's determines uniquely the non-numbered tree. Furthermore the number of ways of assigning the numbers $1,\ldots,k$ to the branches is $k!$.)

Q8.2.3 **[8.2.3]:** Prove inductively, by using the result of [8.2.2], that if in (8.2.7) we replace G_λ with 1 then the bound $|h^{(k)}_{\underline{\nu}}| < C^k$ does follows for a suitable constant C. (*Hint:* Prove that one has $|h^{(k)}_{\underline{\nu}}| < D^k N(k)$, where D is a suitable constant and $N(k)$ is the number of not numbered trees of order k, then use the bound on $N(k)$ proved in problem [8.2.2].)

Q8.2.4 **[8.2.4]:** (*Factorial growth of (some) trees*)
Consider the tree θ:



Show that one can have $\mathrm{Val}(\vartheta) \simeq C^k k^{\alpha k}$ for suitable positive constants C and α, where the order of θ is $3k+1$. (*Hint:* : The momentum $\underline{\nu}_\lambda$ flowing through the line $v_{2k}v_{2k-2}$ is bounded by Nk, and it can happen that $C\underline{\omega}_0\cdot\underline{\nu}_\lambda$ is of the order of $(Nk)^{-\tau}$. Then choose $\underline{\nu}_{2p}+\underline{\nu}_{2p+1}=0$ for $p=0,\ldots,k-1$ (hence obtaining $k-1$ self-energy graphs with the terminology introduced in the forthcoming Section §(8.3)), and use that there are k line carrying a momentum $\underline{\nu}_\lambda$.)

Q8.2.5 **[8.2.5]:** Estimate the constant c in (8.2.11).

§8.3 Cancellations

The key remark in order to take into account the values of the trees that we have excluded by imposing the *unphysical property* [P] in lemma (8.2.2) is that they cancel *almost exactly.* This is the core of the theorem: indeed the previous sections basically contain only definitions necessary to set up the formalism and the following contains bounds based on the maximum principle and the analyticity properties of tree values.

The reason behind the cancellation is very simple. If $\underline{\nu}_{\lambda_v} = \underline{\nu}_{\lambda_{v'}}$ for two comparable nodes $v' \succ v$, and if we denote with ϑ the tree with $\lambda_{v'}$ as root line, we imagine to detach from the tree ϑ the subtree ϑ_2 with last node v. Then attach it to all the *remaining* nodes $w \preceq v'$, $w \in \vartheta/\vartheta_2$, (this means the tree ϑ with the subtree ϑ_2 removed). The simplest case is illustrated in Fig.(8.3.1). We can, furthermore, imagine changing all the signs, simultaneously, of the node momenta of the nodes $w \in \vartheta/\vartheta_2$. Then we add together all the values of the family $\mathcal{D}$ of trees constructed in this way.

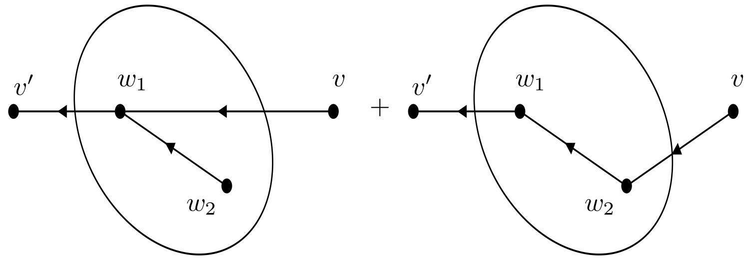
Note that the line momenta of the lines λ_w emerging from $w \in \vartheta/\vartheta_2$ have either the form $\underline{\nu}_{\lambda_w} = \underline{\nu}^0_{\lambda_w} + \underline{\nu}$ with $\underline{\nu} = \underline{\nu}_{\lambda_v}$ and $\underline{\nu}^0_{\lambda_w} = \sum_{w' \preceq w, w' \in \vartheta_2} \underline{\nu}_{w'}$ or the form $\underline{\nu}_{\lambda_w} = \underline{\nu}^0_{\lambda_w}$: the first case arises when the path from v to v' passes through w.

Therefore if $\delta = \underline{\omega}_0 \cdot \underline{\nu}$ the denominators of the propagators of the lines $\lambda_w \in \vartheta_2$ (*i.e.* preceding v' but not preceding v) will have the form $(\underline{\omega}_0 \cdot \underline{\nu}^0_{\lambda_w} + \delta)^{-2}$ or, respectively, $(\underline{\omega}_0 \cdot \underline{\nu}^0_{\lambda_w})^{-2}$. Hence the above trees can be considered only if no one among the divisors vanishes: which cannot happen *as we suppose temporarily* if $\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w}$ is large compared to δ. By construction if we sum the values of all trees in the family $\mathcal{D}$ we get an even function of δ.

The family of trees $\mathcal{D}$ consists of trees whose contributions to $\underline{h}^{(k)}_{\underline{\nu}}$ differ because

(1) some of the branches below v' have changed total momentum by the amount $\underline{\nu} \equiv \underline{\nu}_{\lambda_v}$: this means that some of the denominators $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w})^{-2}$ have become $(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w} + \delta)^{-2}$ or $-(\underline{\omega}_0 \cdot \underline{\nu}_{\lambda_w} + \delta)^{-2}$ (see the branch $\lambda_{w_2} \equiv w_1 w_2$ in Fig.(8.3.2));

(2) there is one of the node factors which changes by taking successively the



F8.3.1 **Fig.(8.3.1)** The simplest cancellation: the circle encloses a subgraph which violates of property [P] (which we shall call later a *self-energy graph*), provided $\underline{\nu}_{w_1}+\underline{\nu}_{w_2}=\underline{0}$. The parts of the tree ϑ above v' and below v are not drawn. Imagine that the branch momentum $\underline{\nu}$ of the branch coming out of v is very large and that $\delta\equiv\underline{\omega}_0\cdot\underline{\nu}$ is very small and note that in the two trees one has $\underline{\omega}_0\cdot\underline{\nu}_{\lambda_{w_2}}=\underline{\omega}_0\cdot\underline{\nu}_{w_2}$ and $\underline{\omega}_0\cdot\underline{\nu}_{\lambda_{w_2}}=\underline{\omega}_0\cdot\underline{\nu}_{w_2}+\delta$, respectively.

value $\underline{\nu}_w$, where $w\in\vartheta/\vartheta_2$ is the node to which such a branch is reattached.

Hence if $\delta=\underline{\omega}_0\cdot\underline{\nu}_{\lambda_v}$ is replaced by 0 and if we sum the values of all the trees considered we would build, in this *resummation*, a quantity proportional to $\sum\underline{\nu}_w=\underline{\nu}_{\lambda_{v'}}-\underline{\nu}_{\lambda_v}$ *which is zero.* Since $\delta\ne 0$ we can expect to see a sum of order δ^2, as the sum that we are considering is even in δ.

However this can be advantageous only if $|\delta|\ll|\underline{\omega}_0\cdot\underline{\nu}^0_\lambda|$ for any branch λ in $\vartheta\setminus\vartheta_2$. If the latter property does not hold then $\underline{\omega}_0\cdot\underline{\nu}_{\lambda_w}$ must be small of order δ at least for some $w\in\vartheta/\vartheta_2$ and $(\underline{\omega}\cdot\underline{\nu}^0_{\lambda_w}+\delta)^{-2}$ may be an extremely small divisor spoiling the gain due to the fact that the sum vanishes for $\delta=0$ to second order in δ.

By applying carefully the argument above one sees that the extreme case in which all lines λ_w are such that $\underline{\omega}_0\cdot\underline{\nu}_{\lambda_w}$ is close to δ would be essentially also treatable. Therefore the problem is to show that the two regimes just envisaged (and their "combinations") do exhaust all possibilities.

Such problems are very common in renormalization theory where they are called *overlapping divergences* problems. Their systematic analysis is made through the "renormalization group methods". We argue here that Eliasson's method can be interpreted in the same way. The following definition of scale of a line with momentum $\underline{\nu}$ will play a key role.

D8.3.1 **(8.3.1) Definition:** (Scale of a propagator)
Given a scaling factor γ and setting $\underline{\omega}=C\underline{\omega}_0$ *we say that a propagator* $G_\lambda=(\underline{\omega}_0\cdot\underline{\nu}_\lambda)^{-2}$ *of a line* λ *has* scale n *if*

e8.3.1
$$\gamma^{n-1}\le|\underline{\omega}\cdot\underline{\nu}_\lambda|<\gamma^n,\qquad\text{for } n\le 0, \tag{8.3.1}$$

and we set $n=1$ *if* $1\le|\underline{\omega}\cdot\underline{\nu}_\lambda|$.

Remark: (1) The notion of scale of a line of a graph was introduced in the theory of (overlapping) divergences in the renormalization of the perturbation series arising in quantum field theory by Hepp, [He61].

(2) In the following we fix $\gamma = 2$: this is an arbitrary choice which recommends itself. However any other value $\gamma > 1$ would be suitable in order to carry the analysis that we need.

We make at this point a second simplifying assumption (which can be removed quite easily as discussed in Section §(8.4)). Namely we want to suppose more than the Diophantine condition (8.1.6), *i.e.* that the vector $\underline{\omega}_0$ is *strongly Diophantine* in the sense of the following definition.

D8.3.2 **(8.3.2) Definition:** (Strongly Diophantine vectors)
Let $\underline{\omega}_0 \in \mathbb{Z}^\ell$ *and suppose that there exists* $\gamma > 1$ *and* $C, \tau > 0$ *such that for all integers* $n \le 0$

e8.3.2
$$\begin{aligned} &(1) \qquad C\,|\underline{\omega}_0 \cdot \underline{\nu}| \ge |\underline{\nu}|^{-\tau}, \qquad && \underline{0} \ne \underline{\nu} \in \mathbb{Z}^l, \\ &(2) \qquad \min_{0 \ge p \ge n} \big| C\,|\underline{\omega}_0 \cdot \underline{\nu}| - \gamma^p \big| > \gamma^{n+1} \qquad && 0 < |\underline{\nu}|^\tau \le (\gamma^{-(n+3)}). \end{aligned} \tag{8.3.2}$$

We shall say that $\underline{\omega}_0$ *satisfies a* strong Diophantine property *with scale factor* γ*, constant* C *and exponent* τ*.*

Remark: (1) The notion means that the values of the numbers $C|\underline{\omega}_0 \cdot \underline{\nu}|$ are "far" from the prescribed sequence of scale factors γ^p for all $n \le p \le 0$ provided $|\underline{\nu}|$ is not too large, *i.e.* it does not exceed a constant times $\gamma^{-n\tau^{-1}}$. The constant is chosen $\gamma^{-3/\tau}$ for later convenience. In other words, taking $\gamma = 2$ a line of momentum $\underline{\nu}$ and scale $n+3$ is such that the value $C|\underline{\omega}_0 \cdot \underline{\nu}|$ is far on the scale 2^n from the "marks" 2^p, for $n \ge p \ge 0$.
(2) For a given γ one can prove that the set of strong Diophantine vectors contained in any ball of radius r in $\mathbb{R}^\ell$ has relative measure which tends to 1 for $C \to \infty$; see problem [8.3.1].

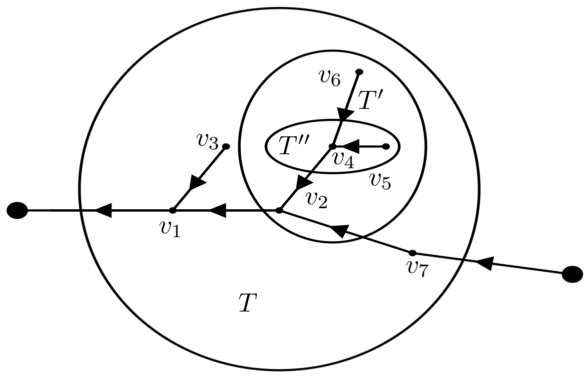
In view of the last remark (2) and for simplicity we shall take $\gamma = 2$ and suppose that $\underline{\omega}_0$ is strongly Diophantine with scale factor 2, constant C and exponent τ, (see problems for Section §(8.4) for a discussion on releasing the assumption).

Proceeding as in quantum field theory, given a tree ϑ we can attach a *scale label* to each branch $\lambda \in \Lambda(\vartheta)$: it is equal to n if n is the scale of the branch propagator. Note that the labels thus attached to a tree are *uniquely determined by the tree*: they will have only the role of helping to visualize the orders of magnitude of the divisors associated with the various tree branches. The scale labels allow us to organize naturally the lines of a tree into "clusters" defined formally as follows.

D8.3.3 **(8.3.3) Definition:** (Clusters of lines and nodes)
Given a tree in which each line carries its scale label one can identify the largest connected clusters T *of nodes that are linked by continuous paths of branches carrying the same scale label* n_T *or a higher one. We shall say that the* cluster T has scale n_T*. We shall denote by* $V(T)$ *the set of nodes in* T*, and by* $\Lambda(T)$ *the set of branches connecting them; by extension we shall say that such branches are contained, or internal, in* T*. We also denote*

by $\Lambda_1(T)$ the set of branches in $\Lambda(T)$ plus the exiting branch of T (if any). Finally call $\mathcal{T}(\vartheta)$ the set of all clusters in ϑ.

Remark: (1) This can be visualized by drawing a box enclosing all the nodes and lines internal to a cluster T. In this way the boxes are hierarchically ordered by inclusion, cf. Figure (8.3.2).
(2) The definition implies that a cluster T of scale n_T is *maximal*, *i.e.* it cannot be enlarged by adding new lines of scale $n \geq n_T$.
(3) The branches of the tree carry an arrow pointing to the root: this gives a meaning to the expressions "entering" or "exiting a cluster".
(4) Each cluster can have at most one exiting line. One can also imagine to define the entering lines (if any). Note that the number of entering lines can be arbitrary: hence figure (8.3.2) is a rather special case, so chosen for its convenience in the later illustration of other features of clusters of a graph.



*F*8.3.2 **Fig.(8.3.2)** An example of three clusters symbolically delimited by circles, as visual aids, inside a tree (whose remaining branches and clusters are not drawn and are indicated by the bullets); not all labels are explicitly shown. The scales (not marked) of the branches increase as one crosses inward the circles boundaries: recall, however, that the scale labels are integers ≤ 1 (hence typically ≤ 0). If the mode labels of (v_4, v_5) add up to $\underline{0}$ the cluster T'' is a self-energy graph. If the mode labels of (v_2, v_4, v_5, v_6) add up to $\underline{0}$ the cluster T' is a self-energy graph and such is T if the mode labels of $(v_1, v_2, v_3, v_4, v_5, v_6, v_7)$ add up to $\underline{0}$. The cluster T' is maximal in T.

Among the clusters we consider the ones with the property that there is only one tree branch entering them and only one exiting and both carry the same momentum. We set the following definition.

*D*8.3.4 **(8.3.4) Definition:** (Self-energy clusters and divergence seeds)
Let T be a cluster with just one entering and one exiting lines. Denote λ_T the entering branch: its scale $n = n_{\lambda_T}$ is smaller than the smallest scale n_T of the branches inside T. We call w_1 the node into which the branch λ_T ends inside T. We say that such a T is a self-energy subgraph if the following conditions are verified.
(i) $\sum_{w\in T} \underline{\nu}_w = \underline{0}$: hence the entering and exiting branches carry the same momentum.
(ii) If $n = n_{\lambda_T}$, and E, η are defined by $E \equiv 2^{-3\eta} N^{-1}$, $\eta = \tau^{-1}$ then the

number $M(T)$ of branches contained in T is not too large:

e8.3.3
$$M(T)\stackrel{def}{=}\textit{number of branches contained in } T\le E\,2^{-n\eta}, \tag{8.3.3}$$

We call n_{λ_T} the self-energy-scale *of T, and λ_T a* self-energy branch*.*
A graph containing a self-energy graph will be said to contain a divergence seed*.*

Remarks: (1) The self-energy graphs are called *resonances* in Eliasson's terminology, [El96].
(2) Note that the self-energy-scale n of a self-energy graph T is different from the scale n_T of T as a cluster: one has $n<n_T$.
(3) The reason for the name of divergence seed that we give to self-energy subgraphs is due to the fact that the values of graphs of order k which contain many divergence seeds can only be, in general, bounded proportionally to a power of $k!$ as the key example in problem [8.2.4] shows. The appropriateness of the name is even more clear when one considers the theory of invariant tori of dimension lower than the maximal, *i.e.* lower than the number of degrees of freedom, [GG02], of the approach to KAM theory based on "resummations" of the Lindstedt series; see Sections §(9.1) and §(9.1).

Let us consider a tree ϑ and its clusters. We wish to estimate the number $N_n(\vartheta)$ of branches in $\Lambda(\vartheta)$ with scale $n\le 0$.
Denoting by T a cluster of scale n let q_T be the number of self-energy graphs of self-energy-scale n contained in T (hence with entering branches of scale n). Recalling that $\Theta_{k,\underline{\nu},j}$ is defined in item (vii) of definition (8.2.1), we have the following inequality.

L8.3.1 **(8.3.1) Lemma:** *For all trees $\vartheta\in\Theta_{k,\underline{\nu},j}$ one has*

e8.3.4
$$N_n(\vartheta)\le\frac{4k}{E\,2^{-\eta n}}+\sum_{\substack{T\in\mathcal{T}(\vartheta)\\ n_T=n}}(-1+q_T), \tag{8.3.4}$$

with $E=N^{-1}2^{-3\eta},\eta=\tau^{-1}$.

Remarks: (1) This is a version of Bryuno's lemma which goes back to Bryuno's work on a theorem by Siegel (in fact Siegel found in a seminal paper the first solution of a small divisor problem, cf. problem [8.3.2]); a proof of (8.3.1) is deferred to Appendix (8.3) below.
(2) Note that a cluster T of scale n can contain self-energy graphs of self-energy scale n which in turn contain other self-energy graphs (necessarily of higher self–energy scale): the number q_T in lemma (8.3.1) counts *only* the self–energy graphs of self-energy scale n.

Besides an estimate of the number of self-energy graphs that can be found in a tree in $\Theta_{k,\underline{\nu},j}$ we shall need a precise description of the families of trees containing divergence seeds whose values we want to sum together to exhibit the cancellations that show that their sums are not as large as they could *a priori* be. The families are defined below: a definition whose usefulness relies upon he validity of the following lemma (8.3.2). The formulation is quite involved and the reader is advised to follow it on a drawing of a graph, *e.g.* on the graph of figure (8.3.2).

*D*8.3.5 **(8.3.5) Definition:** (Resummation families)
Given a labeled tree $\vartheta\in\Theta_{k,\underline{\nu},j}$ and a self-energy graph T of ϑ detach the part of ϑ which has λ_T as root line and attach it successively to the points $w\in\widetilde T$, where $\widetilde T$ is the set of nodes of $V(T)$ outside the self-energy graphs contained in T if any (note that the endpoint $w_1\in V(T)$ of λ_T is among them). This defines a family $\mathcal{F}_1(\vartheta)$ of trees that have the same self energy graphs. Denote by $\Lambda(\widetilde T)$ the set of branches λ contained in T and with at least one point in $\widetilde T$, and by $\Lambda_1(\widetilde T)$ the set of branches in $\Lambda(\widetilde T)$ plus the self-energy branches λ_T; note that all branches $\lambda\in\Lambda(\widetilde T)$ have a scale $n_\lambda\ge n_T$. Select another self energy graph T^1 of ϑ and reapeat the above operations of detaching and attaching the line λ_{T^1} on all the trees $\vartheta^1\in\mathcal{F}_1(\vartheta)$ and thus define a larger family of trees $\mathcal{F}_2(\vartheta)$. Imagine the procedure repeated for all self-energy graphs in ϑ. For each self-energy graph T of ϑ we shall call V_T the number of nodes in $\widetilde T$, i.e. $V_T=|V(\widetilde T)|$. To the just defined set of trees we add the trees obtained by reversing simultaneously the signs of the node modes $\underline{\nu}_w$, for $w\in\widetilde T$: the change of sign is performed independently for the various self-energy graphs. This defines a family of $\prod 2V_T$ trees that we call $\mathcal{F}(\vartheta)$ (the product is over all self-energy graphs in ϑ).

Remarks: (1) The number $\prod 2V_T$ is bounded by $\exp\sum 2V_T\le e^{2k}$.
(2) It is *important* to note that the definition of self-energy graph is such that the above operation (of shift of the node to which the branch entering the self-energy graph is attached) has the property that it cannot change too much the sizes of the propagators of the branches inside the self-energy graphs. The reason is simply that inside a self-energy graph of self-energy-scale n the number of branches is not very large, being $\le\ \overline{N}_n\equiv E\,2^{-n\eta}$ and the number E has been chosen to make the property true.
Indeed let λ be a branch contained inside the self-energy graphs $T=T_1\subset T_2\subset\ldots$ of self-energy-scales $n=n_1>n_2>\ldots$; then the shifting of the branches λ_{T_i} can cause a change in the size of the propagator of λ by at most

*e*8.3.5
$$2^{n_1}+2^{n_2}+\ldots<2^{n+1}. \tag{8.3.5}$$

For any branch λ in $\Lambda(T)$ the quantity $\underline{\omega}_0\cdot\underline{\nu}_\lambda$ has the form $\underline{\omega}_0\cdot\underline{\nu}^0_\lambda+\sigma_\lambda\underline{\omega}_0\cdot\underline{\nu}_{\lambda_T}$ if $\underline{\nu}^0_\lambda$ is the momentum of the branch λ "inside the self-energy graph T", *i.e.* it is the sum of all the node modes of the nodes preceding λ, in the sense of the branch arrows, but contained in T; and $\sigma_\lambda=0,1$.
Therefore not only $|\underline{\omega}\cdot\underline{\nu}^0_\lambda|\ge 2^{n+3}$ (because $\underline{\nu}^0_\lambda$ is a sum of $\le\ \overline{N}_n$ node modes, so that $|\underline{\nu}^0_\lambda|\le N\,\overline{N}_n$) but $\underline{\omega}_0\cdot\underline{\nu}^0_\lambda$ is "in the middle" of the dyadic interval containing it and by (8.3.2) does not get out of it if we add a quantity bounded by 2^{n+1} (like $\sigma_\lambda\underline{\omega}_0\cdot\underline{\nu}_{\lambda_T}$), cf. remark (1) to definition

(8.3.2). Hence no branch changes scale as ϑ varies in $\mathcal{F}(\vartheta^1)$, if $\underline{\omega}_0$ verifies (8.3.2).

By the strong Diophantine property (8.3.2) imposed on the rotation vector ω_0 we have therefore achieved a proof of the following lemma.

$L8.3.2$ **(8.3.2) Lemma:** (Invariance of the scales of the lines of trees in the resummation families)
The self-energy graphs of the trees ϑ in $\mathcal{F}(\vartheta^1)$ all contain the same sets of branches, and the same branches enter or exit each self-energy graph (although they are attached to generally distinct nodes inside the self-energy graphs: the identity of the branches is here defined by the number label that each of them carries in ϑ^1). Furthermore the scales of the self-energy graphs, and of all the branches, do not change as ϑ varies in $\mathcal{F}(\vartheta^1)$.

Remarks: (1) The above proof of this key lemma relies in an essential way on the strong Diophantine property assumed for the rotation vector $\underline{\omega}_0$. It is therefore desirable to show that one can release the latter assumption. In the problems for Section §(8.4) we shall see that if $\underline{\omega}_0$ does not satisfy the strong Diophantine property for a sequence of scale factors γ^n with $\gamma = 2$ it essentially satisfies it for another sequence of scaling factors also exponentially shrinking to 0 (as $n \to -\infty$) and this is all one really needs to achieve a proof of the proposition (8.1.1).
(2) The lemma allows us to consider the collection of the trees in $\Theta_{k,\underline{\nu},j}$ organized as follows. Let ϑ^2 be a tree not in $\mathcal{F}(\vartheta^1)$ and construct $\mathcal{F}(\vartheta^2)$, *etc.* We define a collection $\{\mathcal{F}(\vartheta^i)\}_{i=1,2,\ldots}$ of pairwise disjoint families of trees. We shall sum all the contributions to $h^{(k)}_{\underline{\nu},j}$ coming from the individual members of each family and we shall use that no line "changes scale" as the tree varies in the collection in the sense of lemma (8.3.2).
(3) Collecting tree values in this way is a realization of *Eliasson's resummation*: it is more detailed than the original one in [El96], where no subdivision of the trees in classes was considered and the cancellation that we exhibit below was derived from an argument involving all graphs at the same time. Thus the Eliasson cancellation can be regarded as a cancellation due to a special symmetry of the problem (analogous to the Ward identities of field theory) and the above analysis shows that more symmetry is present in the problem and the cancellation takes place in a more detailed fashion.

Appendix 8.3: Siegel-Bryuno bound on the number of self-energy graphs.

We give here a proof of the lemma (8.3.1). The argument followed below is a minor adaptation of Bryuno's proof of Siegel's theorem, as remarkably exposed by Pöschel, [Po86].

Proof of lemma (8.3.1). Call $N^*_n(\vartheta)$ the number of non-self-energy branches carrying a scale label $\leq n$ in a tree ϑ with k nodes. We shall prove first that $N^*_n(\vartheta) \leq 2k(E2^{-\eta n})^{-1} - 1$ if $N_n(\vartheta) > 0$ (recall that $E = N^{-1}2^{-3\eta}$ and

$\eta = 1/\tau$). We fix n and denote $N^*_n(\vartheta)$ as $N^*(\vartheta)$.

If ϑ has the root branch λ_0 with scale $> n$ then calling $\vartheta_1, \vartheta_2, \ldots, \vartheta_m$ the subtrees of ϑ emerging from the last node of ϑ and with $k_j > E\,2^{-\eta n}$ branches, one has $N^*(\vartheta) = N^*(\vartheta_1) + \ldots + N^*(\vartheta_m)$ and the statement is inductively implied from its validity for $k' < k$ provided it is true that $N^*(\vartheta) = 0$ if $k < E2^{-\eta n}$, which is is certainly the case if E is chosen as in
N8.3.1 equation (8.3.4).[1]

In the other case, call $\lambda_1, \ldots, \lambda_m$ the $m \ge 0$ branches on scale $\le n$ which
N8.3.2 are the nearest to λ_0:[2] such branches are the entering branches of a cluster T on scale $n_T > n$. If ϑ_i is the tree with λ_i as root branch one has $N^*(\vartheta) \le 1 + \sum_{i=1}^m N^*(\vartheta_i)$, and if $m = 0$ the statement is trivial, while if $m \ge 2$ the statement is again inductively implied by its validity for $k' < k$.

If $m = 1$ we once more have a trivial case unless the order k_1 of ϑ_1 is $k_1 > k - E\,2^{-n\eta}/2$. Finally, and this is the real problem as the analysis of a few examples shows, we claim that in the latter case either the root branch of ϑ_1 is a self-energy branch or it cannot have scale $\le n$.

To see this, note that $|\underline{\omega} \cdot \underline{\nu}_{\lambda_0}| \le 2^n$ and $|\underline{\omega} \cdot \underline{\nu}_{\lambda_1}| \le 2^n$, hence $\delta \equiv |(\underline{\omega} \cdot (\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1})| \le 2^{n+1}$, and the Diophantine condition implies that either $|\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1}| > 2^{-(n+1)\eta}$ or $\underline{\nu}_{\lambda_0} = \underline{\nu}_{\lambda_1}$. The latter case being discarded as $k - k_1 < E\,2^{-n\eta}/2$ (and we are not considering the self-energy graphs), it follows that $k - k_1 < E\,2^{-n\eta}/2$ is inconsistent: it would in fact imply that $\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1}$ is a sum of $k - k_1$ node modes and therefore $|\underline{\nu}_{\lambda_0} - \underline{\nu}_{\lambda_1}| < NE\,2^{-n\eta}/2$, hence $\delta > 2^3\,2^n$ which contradicts the above opposite inequality.

A similar, far easier, induction can be used to prove that if $N^*_n(\vartheta) > 0$ then the number $p_n(\vartheta)$ of clusters of scale n verifies the bound $p_n(\vartheta) \le 2k\,(E2^{-\eta n})^{-1} - 1$; see problem [8.3.3]. Thus equation (8.3.4) is proved. ■

Problems for §8.3

Q8.3.1 **[8.3.1]:** (*Full measure of strongly Diophantine vectors*)

Show that the volume of the points $\underline{\omega} \in \mathbb{R}^\ell$ that do not verify property (8.3.2) for some $C > 0$ has zero volume if $\tau > \ell - 1$. (*Hint:* Follow a path similar to the hint for problem [8.1.7].)

Q8.3.2 **[8.3.2]:** (*Siegel's theorem*)

Given a function $\underline{F}(\underline{z})$ of ℓ complex variables holomorphic in $|z_i| < R$ and with $\underline{F}(\underline{0}) = \underline{0}$, $\partial_{z_i} F_j(\underline{0}) = \delta_{ij} e^{i\omega_j}$, consider the map

$$z'_j = F_j(\underline{z}) = e^{i\omega_j} z_j + O(|\underline{z}|^2)$$

and the problem of finding new variables $\underline{\zeta} = \underline{\Phi}(\underline{z}) = \underline{z} + O(|\underline{z}|^2)$ such that near the origin the map takes the form

$$\zeta'_j = e^{i\omega_j} \zeta_j.$$

---

[1] Note that if $k \le E\,2^{-n\eta}$ one has, for all momenta $\underline{\nu}$ of the branches, $|\underline{\nu}| \le NE\,2^{-n\eta}$, *i.e.* $|\underline{\omega} \cdot \underline{\nu}| \ge (NE\,2^{-n\eta})^{-\tau} = 2^3\,2^n$ so that there are *no* clusters T with $n_T = n$ and $N^*(\vartheta) = 0$. The choice $E = N^{-1}2^{-3\eta}$ is convenient: but this, as well as the whole lemma, remains true if 3 is replaced by any number larger than 1. The choice of 3 is made only to simplify some of the arguments based on the self-energy graph concept.

[2] *i.e.* such that no other branch along the paths connecting the branches $\ell_1, \ldots, \ell_m$ to the root is on scale $\le n$.

Show that if there exist $C, \tau > 0$ such that $|\underline{\omega}\cdot\underline{\nu} + 2\pi n| > C^{-1}(|\underline{\nu}| + |n|)^{-\tau}$ for all $\underline{\nu} \in \mathbb{Z}^\ell, n \in \mathbb{Z}$ with $|\underline{\nu}|+|n| > 0$ then $\underline{\Phi}$ exists and is analytic near the origin $\underline{z} = \underline{0}$. (*Hint:* Setting $\underline{m} = (m_1, \ldots, m_\ell) \in \mathbb{Z}^\ell_+$ and $|\underline{m}| = m_1+\ldots+m_\ell$, write $\underline{\Phi}(\underline{z}) = \underline{z}+\sum_{|\underline{m}|>1} c_{\underline{m}}\, \underline{z}^{\underline{m}}$ in Taylor series and determine $c_{\underline{m}}$ recursively via a tree expansion: note that $\underline{m} \in \mathbb{Z}^\ell_+$ plays here the role of $\underline{\nu}$ in the KAM theory: however no self-energy graphs can arise because $m_j \ge 0$. Therefore the result follows from a simpler version of lemma (8.3.1), *i.e.* from lemma (8.2.2) which is also implicit in lemma (8.3.1).)

Q8.3.3 **[8.3.3]:** (*Estimate of the number of clusters of scale n*)
Show that the number $p_n(\vartheta)$ of clusters on scale n contained in a tree ϑ verifies the bound $p_n(\vartheta) \le 2k(E2^{-\eta n})^{-1} - 1$, with $\eta = \tau^{-1}$ and $E = N^{-1}2^{-3\eta}$. (*Hint:* The bound is true for $k \le E2^{-\eta n}$. If the last tree node v_0 is not in a cluster of scale n one has $p_n(\vartheta) = p_n(\vartheta_1) + \ldots + p_n(\vartheta_m)$, with the notation in Appendix 8.3 , and the statement follows by induction. If v_0 is in a cluster of scale n we call $\vartheta_1, \ldots, \vartheta_m$ the subtrees entering the cluster containing v_0 and with orders $k_j > E2^{-\eta n}$: then one has $p_n(\vartheta) = 1 + p_n(\vartheta_1) + \ldots + p_n(\vartheta_m)$. As in Appendix 8.3 we can assume $m = 1$, the other cases being trivial. But in such a case there will be only one branch entering the cluster V of scale n containing v_0 and it will have a momentum of scale $\le n-1$. Therefore the cluster V must contain at least $E2^{-\eta n}$ nodes. This means that $k_1 \le k - E2^{-\varepsilon n}$.)

Q8.3.4 **[8.3.4]:** In the context of problem [8.1.11] replace the condition that the minimum (evaluated over all non-negative integer components vectors $\underline{\nu}$ with $\sum_j \nu_j \ge 2$ and over all $j = 1, \ldots, n$) $\min|\underline{\nu}\cdot\underline{\lambda} - \lambda_j|$ is $> c > 0$ with the condition $\min|\underline{\nu}\cdot\underline{\lambda} - \lambda_j| > C|\underline{\nu}|^{-\tau}$ for suitable constants $C, \tau > 0$ and prove that the same conclusions hold. (*Hint:* Follow the proof of Siegel's theorem in problem [8.3.2].)

§8.4 Convergence and KAM theorem

Having exhibited, cf. remark (2) to lemma (8.3.2) and definition (8.3.5), the class of graphs whose values we want to add up before attempting an estimate we turn to proving that the value of their sums admits, at fixed order k, a much better bound than the one they satisfy individually.

Let $\eta_T = \underline{\omega}\cdot\underline{\nu}_{\lambda_T}$ if $\underline{\nu}_{\lambda_T}$ is the momentum of the line λ_T of scale n_{λ_T}, $n_{\lambda_T} < n_T$ and $\underline{\omega} = C\,\underline{\omega}_0$. If λ is a branch in $\Lambda(\widetilde{T})$, *i.e.* by definition (8.3.3) a branch of a line outside the inner self–energy clusters in T if any, we can imagine to write the quantity $\underline{\omega}\cdot\underline{\nu}_\lambda$ as $\underline{\omega}\cdot\underline{\nu}^0_\lambda + \sigma_\lambda\eta_T$, with $\sigma_\lambda = 0, 1$ (cf. figure (8.3.2)). The product of the propagators of the branches in $\widetilde{T}$ is $C^{2|\Lambda(\widetilde{T})|}$ times

e8.4.1
$$\prod_{\lambda\in\Lambda(\widetilde{T})} \frac{1}{(\underline{\omega}\cdot\underline{\nu}^0_\lambda + \sigma_\lambda\eta_T)^2}. \tag{8.4.1}$$

D8.4.1 **(8.4.1) Definition:** (Height of tress and of self-energy graphs)
If the tree does not contain any self-energy graphs, we say that it has height 0*; if the only self-energy graphs do not contain other self-energy graphs, we say that the tree has height* 1*; more generally if the maximum number of self-energy graphs that contain a given self-energy graph is p, we say that the tree has height p.*
Likewise we say that a self-energy graph has height 0 *if it does not contain other self-energy graphs. Recursively we say that a self-energy graph has*

height p if it contains only self-energy graph of height $\le p-1$ and at least one self-energy graph of height $p-1$. Given a tree ϑ, call $\mathcal{V}(\vartheta)$ the set of all self-energy graphs in ϑ, and set

e8.4.2
$$\Lambda(\mathcal{V}(\vartheta)) = \bigcup_{T\in\mathcal{V}(\vartheta)} \Lambda(T), \qquad \Lambda_1(\mathcal{V}(\vartheta)) = \bigcup_{T\in\mathcal{V}(\vartheta)} \Lambda_1(T). \tag{8.4.2}$$

Remark: Of course in (8.4.2) the union could just be restricted to the maximal self-energy graphs (the sets $\Lambda(T)$ and $\Lambda_1(T)$ have been introduced in definition (8.3.5)).

Consider first the case of a tree ϑ of height 1 and let us denote by T any of its self-energy graphs: if we regard the quantities η_T as independent variables we see that (8.4.1) is holomorphic in η_T for $|\eta_T| < 2^{n_T-3}$. While η_T varies in such complex disk the quantity $|\underline{\omega}\cdot\underline{\nu}^0_\lambda + \sigma_\lambda\eta_T|$ does not become
N8.4.1 smaller than 2^{n_T-3}.[1] The main point here is that the quantity 2^{n_T-3} will usually be $\gg 2^{n_{\lambda_T}}$ which is the value that η_T actually can reach in every tree in $\mathcal{F}(\vartheta)$; this can be exploited in applying the maximum principle, as done below.

Note that the quantities η_T do not depend on the element of the family $\mathcal{F}(\vartheta)$, so that we can factor out of the sum of the values of the graphs in $\mathcal{F}(\vartheta)$ the product $\prod_{T\in\mathcal{V}(\vartheta)} \eta_T^{-4}$ (because each self-energy graph has one branch entering and one exiting with the same propagator) and write the product of the propagators of any tree as $C^{2|\Lambda(widetilde\vartheta)|}$

e8.4.3
$$\Big(\prod_{\lambda\in\Lambda(\vartheta)\setminus\Lambda_1(\mathcal{V}(\vartheta))} \frac{1}{(\underline{\omega}\cdot\underline{\nu}_\lambda)^2}\Big)\cdot\Big(\prod_{\substack{T\in\mathcal{V}(\vartheta)\\ \lambda\in\Lambda(T)}} \frac{1}{(\underline{\omega}\cdot\underline{\nu}^0_\lambda+\sigma_\lambda\eta_T)^2}\Big)\cdot\Big(\prod_{T\in\mathcal{V}(\vartheta)} \frac{1}{\eta_T^2}\Big), \tag{8.4.3}$$

where the first product is over the branches λ which neither enter nor are inside a self-energy graph of ϑ (so that their momentum is the same in all trees of the family $\mathcal{F}(\vartheta)$), the second product is over the branches λ contained in V, and the third product is over the self-energy graphs $T\in\mathcal{V}(\vartheta)$ and takes into account the branches entering T, *i.e.* the self-energy branches. As said above the last product factors out of the sum of the values of the trees in $\mathcal{F}(\vartheta)$ at fixed ϑ.

Consider the sum of the $\prod 2V_T \le e^{2k}$ products of propagators of the members of the family $\mathcal{F}(\vartheta)$ *divided by the last factor in* (8.4.3). Each such product relative to the tree ϑ is holomorphic in the region $|\eta_T| < 2^{n_T-3}$ and it is bounded there by $\prod 2^{-2(n_\lambda-3)} \le 2^{6k}\prod_\lambda 2^{-2n_\lambda}$, if n_λ the scale of the branch λ in ϑ and if the product is over the branches neither entering nor exiting a self-energy graph. This even holds if the η_T are regarded as independent complex parameters.

[1] In fact $|\underline{\omega}\cdot\underline{\nu}^0_\lambda| \ge 2^{n+3}$ because T is a self-energy graph; therefore $|\underline{\omega}\cdot\underline{\nu}_\lambda| \ge 2^{n+3}-2^{n+1} > 2^{n+2}$ so that $n_T \ge n+3$. On the other hand we note that $|\underline{\omega}\cdot\underline{\nu}^0_\lambda| > 2^{n_T-1}-2^{n+1}$, so that it follows that $|\underline{\omega}\cdot\underline{\nu}^0_\lambda+\sigma_\lambda\eta_T| \ge 2^{n_T-1}-2^{n+1}-2^{n_T-3} \ge 2^{n_T-3}$, for $|\eta_T| < 2^{n_T-3}$.

By construction it is clear that the just considered sum of the $\prod 2V_T \le e^{2k}$ terms from the trees in $\mathcal{F}(\vartheta)$, vanishes to second order in the η_T parameters (by the approximate cancellation discussed above). *By the maximum principle* this means that if we bound the sum by the number of terms times the maximum among them (easy to estimate because the propagators have all well defined seizes) we can multiply the result by a further factor of the order of $2^{-2n_{\lambda_T}}/2^{-2(n_T-3)}$ and still obtain a valid bound.

Hence by the maximum principle, and recalling that each $\underline{\nu}_v$ in (8.2.8) can be bounded by N, and that (as it is straightforward to check) one has

e8.4.4
$$\sum_{v\in V(\vartheta)} m_v = k-1, \tag{8.4.4}$$

we can bound the contribution to $\underline{h}^{(k)}_{\underline{\nu}}$ from the family $\mathcal{F}(\vartheta)$ by

e8.4.5
$$\Big[\frac{1}{k!}\Big(\frac{f_0 C^2 N^2}{J}\Big)^k 2^{6k} e^{2k} \prod_{n\le 0} 2^{-2nN_n}\Big]\Big[\prod_{n\le 0}\prod_{\substack{T\in\mathcal{T}(\vartheta)\\ n_T=n}}\prod_{i=1}^{q_T} 2^{2(n-n_i+3)}\Big], \tag{8.4.5}$$

where
(1) $N_n = N_n(\vartheta)$ is the number of propagators of scale n in ϑ ($n=1$ does not appear as $|\underline{\omega}\cdot\underline{\nu}| \ge 1$ in such cases);
(2) the first square bracket is the bound on the product of individual elements in the family $\mathcal{F}(\vartheta)$ times the bound e^{2k} on their number: this takes into account *also* the last product in (8.4.3);
(3) the second square bracket is the part coming from the maximum principle, applied to bound the resummations, and is explained as follows.
(i) The dependence on the variables $\eta_{T_i} \equiv \eta_i$ relative to self-energy graphs $T_i \subset T$ with self-energy-scale $n_{\lambda_{T_i}} = n$ is holomorphic for $|\eta_i| < 2^{n_i-3}$, if $n_i \equiv n_{T_i}$, provided $n_i > n+3$ (see above).
(ii) The resummation says that the dependence on the η_i's has a second order zero in each. Hence the maximum principle tells us that we can improve the bound given by the third factor in (8.4.3) by the product of factors $(|\eta_i|\,2^{-n_i+3})^2$ if $n_i > n+3$. If $n_i \le n+3$ we cannot gain anything: but since the contribution to the bound from such terms in (8.4.3) is >1 we can leave them in it to simplify the notation, (of course this means that the gain factor can be important only when $\ll 1$).

We shall shortly see that the above would be sufficient if there were no trees of height higher than 1. To treat the general case we can proceed inductively and suppose that the bound in (8.4.5) holds for trees of height $1,2,\ldots,p-1$ and for values of the $\eta_T = \underline{\omega}\cdot\underline{\nu}_{\lambda_T}$ of the branches that enter the maximal self-energy graphs T which are in the complex disk $|\eta_T| < 2^{n_T-3}$.

Let ϑ be a tree with height p: then each of its maximal self-energy graphs V contains a tree of height $< p$. We imagine that all the resummations relative to the branches that enter the self-energy graphs that are not maximal have been performed so that we only have to consider the trees that are obtained by attaching the branches that enter the maximal self-energy graphs T to the nodes in $\widetilde{T}$.

Suppose for simplicity that there is only one maximal self-energy graph T of height p. Then the sum of the values of the trees of the family $\mathcal{F}(\vartheta)$ obtained by shifting the entrance node into the self-energy graphs of lower height will have the form

e8.4.6

$$\begin{aligned}&\Big(\prod_{\lambda\in\Lambda(\vartheta)\backslash\Lambda_1(T)}\frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}\cdot\underline{\nu}_\lambda)^2}\Big)\cdot\\&\qquad\cdot\Big(\prod_{\lambda\in\Lambda(\widetilde{T})}\frac{\underline{\nu}_{v'}\cdot\underline{\nu}_v}{(\underline{\omega}\cdot\underline{\nu}^0_\lambda+\sigma_\lambda\eta_T)^2}\Big)\cdot\frac{1}{\eta_T^2}\cdot\Big(\prod_{T_i}F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v'_i})\Big),\end{aligned}\tag{8.4.6}$$

where the last product is over all the maximal self-energy graphs T_i contained in T, v'_i and v_i are the nodes in $\widetilde{T}_i$ from which exits (or enters, respectively) the branch that enters (or exits) the self-energy graph T_i, and $F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v'_i})$ is the sum of the values of all the trees that we have to sum in shifting the entrance node of the branches that enter the self-energy graphs of lower order inside T_i. Note that the branches λ in the first product are the branches external to T (but not entering T), while the ones in the second product are the branches internal to T (*i.e.* branches connecting to nodes in $\widetilde{T}$).

We can then remark that when η_T varies in the complex disk $|\eta_T|\le 2^{n_T}$ the divisors of the branches that enter the inner self-energy graphs T' (of any height) do not exceed in modulus $2^{n_{\lambda_{T'}}}$. Therefore we can bound the quantities $F(T_i,\underline{\nu}_{v_i},\underline{\nu}_{v'_i})$ via the inductive bound and obtain that (8.4.5) is valid also for trees of height p which contain only one maximal self-energy graph.

Since the case in which there are many self-energy graphs of height p is clearly reducible to the case in which there is only one such cluster the conclusion is the validity of the inequality (8.4.5) in general. It would also possible to give a proof of the inequality that is not base on an inductive argument but we leave it as a problem for the reader.

Hence substituting (8.3.2) into (8.4.5) we see that the q_T is taken away by the first factor in $2^{2n}2^{-2n_i}$, while the remaining 2^{-2n_i} are compensated by the -1 before the $+q_T$ in (8.3.2), taken from the factors with $T=T_i$, (note that there are always enough -1's).

Hence the product (8.4.5) is bounded by

e8.4.7

$$\frac{1}{k!}\,(C^2J^{-1}f_0N^2)^k e^{2k}2^{12k}\prod_{n\le0}2^{-8nkE^{-1}\,2^{\eta n}}\le\frac{1}{k!}\,B_0^k,\tag{8.4.7}$$

with B_0 suitably chosen.

To sum over the trees we note that fixed ϑ the collection of clusters is fixed. Therefore we only have to multiply (8.4.7) by the number of tree shapes for ϑ, ($\le 2^{2k}k!$), by the number of ways of attaching mode labels, ($\le(3N)^{\ell k}$), by the number of ways of contacting the tensor labels, ($\le \ell^k$), so that we can bound $|\underline{h}_{\underline{\nu}}^{(k)}|$ by

e8.4.8
$$\varepsilon_0^{-k} \equiv (b_\ell J^{-1} C^2 f_0 N^{2+\ell} e^{cN} e^{\xi N})^k, \tag{8.4.8}$$

with b_ℓ suitably chosen. ■

Problems for §8.4 *(Complements to KAM theory)*

Q8.4.1 **[8.4.1]:** (*Diophantine spacing*)
Let $\underline{\omega}$ be a Diophantine vector, cf. (8.1.6), with constants C and exponent τ. Consider the set B_n of the values of $|\underline{\omega}\cdot\underline{\nu}|$ as $\underline{\nu}\in\mathbb{Z}^\ell$ varies in the set $0\le|\underline{\nu}|<(2^{n+3})^{-1/\tau}$, with $n=0,-1,-2,\ldots$. The sets B_n verify the inclusion relation $B_n\subset B_m$ if $m<n$. Check that the spacing between the points of each of the sets B_n is at least $2^\tau(2(2^{n+3})^{-1/\tau})^{-\tau}\ge 2^{n+3}$. Check also that $x\in B_n$ is such that $x>2^{n+3}$, if $x\ne 0$.

Q8.4.2 **[8.4.2]:** In the context of problem [8.4.1] let, more abstractly, B_n, $n=0,-1,\ldots$, be a sequence of sets such that (i) $0\in B_n$, (ii) $B_n\subset B_m$ if $m<n$ and (iii) the spacing between the points in B_n is at least 2^{n+3} (the latter will be the *spacing property*). Show that there exists a sequence $\gamma_0,\gamma_{-1},\ldots$ with $\gamma_p\in[2^{p-1},2^p]$ such that

e8.4.9
$$|x-\gamma_p|\ge 2^{n+1} \qquad \text{if} \quad n\le p\le 0 \qquad \text{and} \quad x\in B_n, \tag{8.4.9}$$

for all $n\le 0$. (*Hint:* The reader who does not want to build the proof can read it in the appendix (8.4).)

Q8.4.3 **[8.4.3]:** (*Arithmetic properties related to the Diophantine property*)
In the context of problem [8.4.2] let $C_0=2^\tau C$ and $\underline{\omega}=C_0\underline{\omega}$ ehre $\underline{\omega}_0$ is a Diophantine vector with constant C and exponent τ; show that it is possible to find a sequence γ_p with $\gamma_p\in[2^{p-1},2^p]$, $p\le 0$, such that

e8.4.10
$$||\underline{\omega}\cdot\underline{\nu}|-\gamma_p|\ge 2^{n+1} \qquad \text{if} \quad 0<|\underline{\nu}|\le(2^{n+3})^{-\tau^{-1}} \tag{8.4.10}$$

for all $n\le 0$ and for all $p\ge n$. Furthermore $|\underline{\omega}\cdot\underline{\nu}|\ne\gamma_n$, for all $n\le 0$. (*Hint:* This is implied by the arithmetic result of problem [8.4.2].)

Q8.4.4 **[8.4.4]:** (*Elimination of the assumption of strongly Diophantine property*, [GG95])
Starting from the result of problem [8.4.3] consider the notion of self–energy subgraph introduced in Section 8.3 assuming rotation vectors $\underline{\omega}$ satisfying the conditions (8.3.2). Adapt the notion to the case in which only the Diophantine property (8.1.6) is assumed. To define the scales let (as above) $\underline{\omega}=2^\tau C\underline{\omega}$; a propagator will be said to be on scale n if one has $\gamma_{n-1}\le|\underline{\omega}\cdot\underline{\nu}|<\gamma_n$, for $n\le 0$, and set $n=1$ if $\gamma_0<|\underline{\omega}\cdot\underline{\nu}|$. Show that the line scales do not change as a line entering a self-energy graph is detached and reattached to another node of the subgraph. (*Hint:* The shift of the lines exemplified in Fig.(8.3.1) causes a change of the size of the propagator by at most $\gamma_{n_1}+\gamma_{n_2}+\ldots$ (instead of $2^{n_1}+2^{n_2}+\ldots$, see (8.3.5)), but such a quantity is still bounded by 2^{n+1}. By proceeding as in Section 8.4, we see that the product of the propagators is holomorphic in $\eta_T\equiv\underline{\omega}\cdot\underline{\nu}_{\lambda_T}$ for $|\eta_T|<\gamma_{n_T-3}\ldots$.)

Q8.4.5 **[8.4.5]:** (*Extension of Siegel–Bryuno bound*)
Imagine that the perturbation $f(\underline{\alpha})$ is not a trigonometric polynomial but it is just analytic in $\underline{\alpha}$. Extend the definition of self–energy graph by changing the condition that the number of branches is $\le E2^{-n\eta}$, cf. (8.3.3), into

e8.4.11
$$M(T)\stackrel{def}{=}\sum_{v\in V(T)}|\underline{\nu}_v|\le E2^{-n\eta}, \tag{8.4.11}$$

where $E = 2^{-3\eta}$. Show that a lemma analogous to lemma (8.3.1) holds in the form: *for all trees ϑ one has*

e8.4.12
$$N_n(\vartheta) \le \frac{4M(\vartheta)}{E\,2^{-\eta n}} + \sum_{\substack{T\in\mathcal{T}(\vartheta)\\ n_T=n}} (-1+q_T)\,, \qquad M(\vartheta) = \sum_{v\in V(\vartheta)} |\underline{\nu}_v|, \tag{8.4.12}$$

with $E = 2^{-3\eta}$ and $\eta = 1/\tau$. Check that the above statement implies as a particular case lemma (8.3.1) if the perturbation is a trigonometric polynomial.(*Hint:* The proof is again by induction, and it is essentially identical to that of lemma (8.3.1), apart from the obvious changes in the notations.)

Q8.4.6 **[8.4.6]:** (*Non-divergent graphs in the case of analytic perturbations*)
Check that the bound in problem [8.4.5] can be directly used to obtain the result of proposition (8.1.1) under the only assumption that the perturbation is analytic. (*Hint:* In (8.4.3) the bound f_0^k has to be replaced by the bound $F^k\prod_{v\in V(\vartheta)} e^{-\kappa|\underline{\nu}_v|}$ for some $F,\kappa > 0$ (which comes from the analyticity assumption on V). The product $\prod_{n\le 0} 2^{-2nN_n}$ will be replaced by $2^{-2n_0k}\prod_{n\le n_0} 2^{-2nN_n}$ where n_0 is an arbitrary negative integer. Of course it is no longer possible to bound the product $\prod_{v\in V(\vartheta)} |\underline{\nu}_v|^{m_v+1}$ with N^{2k} as in (8.4.3), but we can use that $\frac{1}{m!}|\underline{\nu}|^m \le \left(\frac{8}{\kappa}\right)^m$ and that the number of trees with given $\{m_v\}$ and without labels is bounded by $k!/\prod_v m_v!$; see problem [8.2.1]. From the above bound by $F^k\prod_{v\in V(\vartheta)} e^{-\kappa|\underline{\nu}_v|}$ we can extract a factor $\le \exp[-\kappa|\underline{\nu}|/2]$, with $\underline{\nu} = \underline{\nu}_{\lambda_0}$, if λ_0 is the root branch, while the remaining factor $\exp[-\kappa M(\vartheta)/2]$ (with $M(\vartheta) \equiv \sum_v |\underline{\nu}_v|$, see (8.4.12)), will be kept as such. Collecting together the bounds the sum over all non divergent graphs (*i.e.* without self-energy subgraphs) is

$$F^k e^{-\frac{1}{2}\kappa|\underline{\nu}|} \sum_{\{\underline{\nu}_v\}_{v\in V(\vartheta)}} \Big(\prod_v e^{-\frac{1}{2}\kappa|\underline{\nu}_v|}|\underline{\nu}_v|\Big)\, 2^{-2n_0k} \left(\frac{8}{\kappa}\right)^{2k} \prod_{n\le n_0} 2^{\sum_v |\underline{\nu}_v| 4E^{-1}2^{n/\tau}}$$

can be bounded by a constant to the power k if n_0 is so chosen that $\sum_{n\le n_0} 4(\log 2)\,E^{-1}2^{n/\tau} < \kappa/4$, which assures the summability over the Fourier labels.)

Q8.4.7 **[8.4.7]:** (*Cancellation in the case of analytic perturbations*)
Prove that, by considering the quantities $\eta_T \stackrel{def}{=} \underline{\omega}\cdot\underline{\nu}_{\lambda_T}$, cf. lines preceding (8.4.1), as independent variables, the product of propagators is holomorphic in η_T for $|\eta_T| < \gamma_{n_T-3}$. Check that this can be combined with the results of problems [8.4.5] and [8.4.6], to proceed as in Section §8.4 and to achieve a proof of proposition (8.1.1) under the only assumption the perturbation $V(\underline{\alpha})$ is analytic in $\underline{\alpha}$. (*Hint:* If λ is a branch on scale n_T, one has $\gamma_{n_T-1} \le |\underline{\omega}\cdot\underline{\nu}_\lambda| < \gamma_{n_T}$; then $|\underline{\omega}\cdot\underline{\nu}_\lambda^0| > 2^{n+3}$ implies $|\underline{\omega}\cdot\underline{\nu}_\lambda| > 2^{n+3} - 2^n > 2^{n+2}$, so that $n_T \ge n+3$. On the other hand, if $n_T > n+3$, *i.e.* $n_T = n+m$, for some $m > 3$, one has $|\underline{\omega}\cdot\underline{\nu}_\lambda^0| > \gamma_{n_T-1} - \gamma_n$, because the the scales of all the branches do not change, so that it follows that, for $|\eta_T| < \gamma_{n_T-3}$, one has $|\underline{\omega}\cdot\underline{\nu}_\lambda^0 + \sigma_\lambda\eta_T| \ge \gamma_{n_T-1} - \gamma_n - \gamma_{n_T-3}$ $\ge (2^{n_T-2} - 2^{n_T-m}) - \gamma_{n_T-3} \ge (2^{n_T-3} + 2^{n_T-4} + \ldots + 2^{n_T-m+1}) - 2^{n_T-3} \ge 2^{n_T-4}$; otherwise, if $n_T = n+3$, one has $|\underline{\omega}\cdot\underline{\nu}_\lambda^0 + \sigma_\lambda\eta_T| > 2^{n+3} - \gamma_{n_T-3} \ge 2^{n_T-1}$, for $|\eta_T| < \gamma_{n_T-3}$. Therefore we can conclude that, for $|\eta_T| < \gamma_{n_T-3}$, the quantity $|\underline{\omega}\cdot\underline{\nu}_\lambda^0 + \sigma_\lambda\eta_T|$ does not become smaller than 2^{n_T-4}.)

Q8.4.8 **[8.4.8]:** (*Lindstedt series for non-even perturbations*)
Check that the assumption $f_{\underline{\nu}} = f_{-\underline{\nu}}$ can be avoided in order to obtain the formal solubility of the equations of motions (8.1.12). (*Hint:* We can prove by induction on the tree order that no branch can have a vanishing momentum. Consider all contributions arising from the trees $\vartheta \in \Theta_{k,\underline{0}}$ deprived of the root branch[2] (so that $\mathrm{Val}(\vartheta)$ is a well-
N8.4.2 defined quantity by the inductive hypothesis): we group together all trees obtained from

[2] This means that the propagator of the root line is replaced with 1.

each other by shifting the root branch, *i.e.* by changing the node which the root branch exits and orienting the arrows in such a way that they still point toward the root. We call $\mathcal{F}(\vartheta)$ such a class of trees (here ϑ is any element inside the class). The values $\mathrm{Val}(\vartheta')$ of such trees $\vartheta' \in \mathcal{F}(\vartheta)$ differ because (1) there is a factor $i\underline{\nu}_v$ depending on the node v to which the root branch is attached, and (2) some arrows change their directions. More precisely, when the root branch is detached from the node v_0 and reattached to the node v, if $\mathcal{P}(v_0,v)=\{w\in V(\vartheta)\,:\, v_0\succeq w\succeq v\}$ denotes the path joining the node v_0 to the node v, all the momenta flowing through the branches λ along the path $\mathcal{P}(v_0,v)$ change their signs: nevertheless the propagators do not change. Then by summing the values of all possible trees inside the class $\mathcal{F}(\vartheta)$ we obtain a common value times i times $\sum_{v\in V(T)}\underline{\nu}_v$, and the sum gives zero.)

Q8.4.9 **[8.4.9]:** (*KAM theorem for non-even perturbations*)
Check that the assumption $f_{\underline{\nu}}=f_{-\underline{\nu}}$ can be avoided in the proof of the KAM theorem. (*Hint:* Given a tree ϑ with a self-energy graph T, set $\delta=\underline{\omega}\cdot\underline{\nu}$, with $\underline{\nu}=\underline{\nu}_{\lambda_T^2}$. Consider all trees which can be obtained by shifting the entrance node of the entering branch λ_T^2: the sum of the values of all the so obtained trees is zero when the propagators of the branches in $\Lambda(T)$ are computed at $\delta=0$, by the same argument as the one used in Section 8.1. Then consider the corresponding contributions to first order in δ: one has to derive the product of propagators of the branches $\lambda\in\Lambda(T)$ and compute it at $\delta=0$. The derivative gives $|\Lambda(T)|$ terms, each of which has the derivative acting on a propagator G_λ, while all the other propagators $G_{\lambda'}$, $\lambda'\neq\lambda$, are not derived. The branch λ divides $V(T)$ into two disjoint set of nodes V_1 and V_2, such that λ_T^1 exits from a node inside V_1 and λ_T^2 enters a node inside V_2: if $\lambda=\lambda_v$ one has $V_2=\{w\in V(T)\,:\,w\preceq v\}$ and $V_1=V(T)\setminus V_2$. By setting $\underline{\nu}_1=\sum_{v\in V_1}\underline{\nu}_v$ and $\underline{\nu}_2=\sum_{v\in V_2}\underline{\nu}_v$, one has $\underline{\nu}_1+\underline{\nu}_2=\underline{0}$. Then consider the families $\mathcal{F}_1(\vartheta)$ and $\mathcal{F}_2(\vartheta)$ of trees obtained as follows: $\mathcal{F}_1(\vartheta)$ is obtained from ϑ by detaching λ_T^1 then reattaching to all the nodes $w\in V_1$ and by detaching λ_T^2 then reattaching to all the nodes $w\in V_2$, while $\mathcal{F}_2(\vartheta)$ is obtained from ϑ by reattaching the branch λ_T^1 to all the nodes $w\in V_2$ and by reattaching the branch λ_T^2 to all the nodes $w\in V_2$. As a consequence of such an operation the arrows of some branches $\lambda\in\Lambda(T)$ change their directions: this means that for some branch λ the momentum $\underline{\nu}_\lambda$ is replaced with $-\underline{\nu}_\lambda$. The derived propagator G_λ can change sign, while the non-derived propagators $G_{\lambda'}$, with $\lambda'\neq\lambda$, do not change by parity: one has a different sign for the trees in $\mathcal{F}_1(\vartheta)$ with respect to the trees in $\mathcal{F}_2(\vartheta)$. Then by summing over all the possible trees in $\mathcal{F}_1(\vartheta)$ we obtain a value $i^2\underline{\nu}_1\underline{\nu}_2$ times a common factor, while by summing over all the possible trees in $\mathcal{F}_2(\vartheta)$ we obtain $-i^2\underline{\nu}_1\underline{\nu}_2$ times the same common factor, so that the sum of two sums gives zero.)

Appendix 8.4: Weakening the strong Diophantine condition.

Here we prove the statement in problem [8.4.2]. Note that (8.4.10) is very similar to the second condition introduced in (8.3.2) to define the *strong Diophantine condition.* Our point is that all that is really needed (see the following discussion) to prove the KAM theorem are (8.1.6) (Diophantine property) and (8.4.10): the latter is a simple arithmetic property which is in fact a consequence of the first.

Proof: Note that if $\gamma_p\in[2^{p-1},2^p]\equiv I_p$ the equations (8.4.9) are obviously verified for $n>p-3$, hence we can suppose $p\geq n+3$.

Fix $p\leq 0$ and let $G=[a,b]$ be an interval verifying what we shall call below the *property* $\mathcal{P}_n$:

eA8.4.1
$$\begin{aligned}\mathcal{P}_n\ : \qquad & |x-\gamma|\ge 2^{m+1} \qquad \text{for all} \quad n\le m\le -3,\\ & \text{for all}\, x\in B_n \quad\text{and}\quad \text{for all}\,\gamma\in G,\ |G|\ge 2^{n+1}.\end{aligned} \tag{A8.4.1}$$

Let $G_{p-3}\equiv[2^{p-1},b_{p-3}]$, with $b_{p-3}\in[2^{p-1},2^p]$, be a *maximal* interval verifying property $\mathcal{P}_{p-3}$ (note that G_{p-3} exists because $x\in B_{p-3}$, $x\ne 0$ implies $x\ge 2^p$ by the spacing property, and it is $b_{p-3}\ge 2^{p-1}+2^{p-2}$).

Assume inductively that the intervals $[a_n,b_n]=G_n$ can be so chosen that $G_{n'}\subseteq G_{n''}$ if $n'<n''$ and G_n is maximal among the intervals contained in G_{n+1} and verifying the property $\mathcal{P}_n$.

If we can check that the hypothesis implies the existence of an interval $G\subseteq G_n$ verifying $\mathcal{P}_{n-1}$ we shall be able to define G_{n-1} to be a maximal interval among the ones contained in G_n and with the property $\mathcal{P}_{n-1}$: in case of ambiguity we shall take G_{n-1} to coincide with the rightmost possible choice. And as a consequence we shall be able to define $\gamma_p=\lim_{n\to-\infty}b_n$, which will verify (8.4.10).

To check the existence of $G\subseteq G_n$ verifying $\mathcal{P}_{n-1}$ we consider first the case in which B_{n-1} has one and only one point x in G_n. If $|G_n|\ge 2^{n+2}$ and x is in the first half of G_n we can take, by the spacing property, $G=[x+2^n,\min\{b_n,\,x+2^{n+2}-2^n\}]$; if it is in the second half we take $G=[\max\{a_n,x-2^{n+2}+2^n\},x-2^n]$.

If, on the other hand, $|G_n|<2^{n+2}$ it is $G_n\subset G_{n+1}$ strictly and, furthermore, the interval $(x-2^{n+2},x+2^{n+2})$ does not contain points of B_{n-1} other than x itself (by the spacing property). The strict inclusion implies that there is a point $y\in B_n$ at distance exactly 2^{n+1} from G_n (recall the maximality of G_n).

Suppose that x is in the first half of G_n and $y<a_n$, *i.e.* $y=a_n-2^{n+1}$; then $x-y<2^{n+1}+2^{n+1}$ contradicting the spacing property. Hence $y>b_n$, *i.e.* $y=b_n+2^{n+1}$: in such case it cannot be, again by the spacing property, that $x+2^{n+2}>y=b_n+2^{n+1}$, so that $b_n-x\ge 2^{n+1}$ and we can take $G=[b_n-2^n,b_n]$. If x is in the second half the roles of left and right are exchanged.

This completes the analysis of the case in which only one point of B_{n-1} falls in G_n. The cases in which either no point or at least two points of B_n fall in G_n are analogous but easier. If two consecutive points $x<y$ of B_{n-1} fall inside G_n we must have $y-x\ge 2^{n+2}$ by the spacing property: hence $G=[x+2^n,y-2^n]\subset G_n$ enjoys the property $\mathcal{P}_{n-1}$. If no point of B_{n-1} falls in G_n let $y\in B_{n-1}$ be the closest point to G_n; if its distance to G_n exceeds 2^n we take $G=G_n$. Otherwise suppose that $y>b_n$: the spacing property implies that the interval $(y-2^{n+2},y)$ is free of points of B_{n-1}. Hence if $a=\max\{a_n,y-2^{n+2}+2^n\}$, $b=y-2^n$ then $G=[a,b]$ has the property $\mathcal{P}_{n-1}$. If, instead, $y<a_n$ we set $a=y+2^n>a_n$ and $b=\min\{b_n,y+2^{n+2}-2^n\}$ and $G=[a,b]$ enjoys the property $\mathcal{P}_{n-1}$. ■

Bibliographical note for §8.1, §8.2, §8.3 and §8.4

The stability problem of Hamiltonian motions is ancient; introductions to this problem are the first two chapters of the book by Moser, [Mo63], the article [Mo78] and the article by Arnold, [Ar63], from p. 85 to p. 99. For a "classical" proof of the KAM theorem see Appendix 34 of [AA68] or Section 12 in [Ga82]. More recently a new proof of the theorem has been derived by Eliasson, [El96]: we have followed the interpretation given to it in [Ga94]. The proof presented in this section deals with a special case but it can be done in full generality along the same lines, see [GM96] and Section §8.4. The technique used is inspired to the perturbation theory in quantum field theory and renormalization group, [Ga01].

The analytic case considered in this section has been studied, in great generality, by Kolmogorov and then applied to celestial mechanics by Arnold. The case in which $V(\underline{\alpha})$ is only assumed differentiable has been studied by Moser: the methods of this section can also be applied to some differentiable cases, see [BGGM98].

As remarked in problem [8.4.4] almost all $\underline{\omega}$ verify (8.3.2) for some C and some $\tau > \ell - 1$, with a sequence γ_p that can be prescribed *a priori* as $\gamma_p = 2^p$. This, however, leaves out important cases like a quadratic irrational as rotation number in $\ell = 2$. And it has the very unfortunate drawback of being non-constructive, as the set of full measure of the $\underline{\omega}$ verifying the strong Diophantine condition is obtained by abstract arguments (*e.g.* the Borel-Cantelli lemma), see problem [8.1.7]. Nevertheless considering strongly Diophantine vectors is natural as it leads to the simplified proof of the previous section, eliminating the "secondary" difficulty we have discussed here.

The Siegel–Bryuno bound is not the only way to approach the problem of bounds on products of small divisors. Instead one can use that, for trees ϑ without self-energy graphs one has

$$\Big|\prod_{\lambda\in\Lambda(\vartheta)} G_\lambda\Big| \le D^k\Big(\sum_{v\in V(\vartheta)} |\underline{\nu}_v|\Big)^{-1}\prod_{v\in V(\vartheta)} |\underline{\nu}_v|^{3\tau},$$

for a suitable constant D. This is an extension of a bound due to Siegel and Eliasson. and it leads to the same results while it can be fruitfully applied also to related problem for which the bound in lemma (8.3.1) is not well suited. Such a bound would allow us to obtain the same conclusions: for further details we refer to the original papers (see [El96] and [BGGM98]).

The "tree expansion technique" can be applied to most perturbation theory problems providing a simple and unifying way of seeing them (however simplicity is not objective and this point of view does not seem to be shared by everybody).

A further natural question could be what happens if the perturbation is not an even function or, even more, if it depends also on the action variables.

If the function $V(\underline{\alpha})$ is not even, than the formal solubility of the equations of motions and the second order cancellation of the self-energy graphs does not follow anymore from parity considerations. In such a case the cancellation mechanism is a little more complicated, and it requires also the shifting of the exiting branch of the self-energy graphs (as originally remarked in [CF94]; see [GM96] for an implementation within a formalism closer to the one described here).

The generalization to perturbations depending also on the action variables requires some minor extensions of the cancellations described in Section §8.1, which can be found in [CF96] and in [GM96]; also the case of unperturbed Hamiltonians $K(\underline{A})$ with a more general dependence on the action variables can be dealt with provided that one has $\det \partial^2_{\underline{A}} K(\underline{A}) \neq 0$ (*anisochrony condition*). But no real new difficulty arises; we shall not discuss further such a case here, for which we refer to the original papers. For a discussion of the optimal conditions under which the KAM theorem can be formulated we refer to a recent exhaustive paper by Rüssmann [Ru01].

CHAPTER IX

Some special topics in KAM theory

§9.1 Resummation and renormalized series for invariant tori

In Chapter VIII the proof of the KAM theorem was based on diagrams and notions becoming visually clear if the diagrammatic interpretation of the various terms building the Lindstedt series was kept in mind. The analogy with the diagrams used in perturbation theory in Quantum Field Theory and in Statistical Mechanics is, we feel, quite striking. Therefore one can wonder whether one could go ahead and apply other techniques widely employed in those fields.

One key technique is that of transforming power series solutions to perturbation problems parameterized by a small parameter ε into series which are no longer power series but which are series of functions that depend in a nontrivial way on ε. This is interesting particularly in the cases (virtually all cases in the quoted fields) in which the power series solutions are either very unlikely to converge even for small ε or are even known to diverge: although they often are asymptotic series. The new series in terms of functions of the parameters are usually obtained by collecting together parts of *different* orders in ε identified by suitable properties of graphs that are used to represent them: in the best cases the new series may even be absolutely convergent and from them one can obtain informations about the nature of the singularity at $\varepsilon = 0$, as well as much simpler expressions for the solutions to perturbation problems.

In the case of the KAM theory however we have seen that the results are analytic in ε and no resummation is, strictly speaking, necessary. Nevertheless the analogies allow us to apply ideas that had so much impact in the understanding of quantum and statistical theories to the "simpler" case of KAM.

This is interesting enough in itself, and its interest becomes even more manifest if one attempts studying the invariant tori of dimension lower than the maximal one (*i.e.* lower than the number of degrees of freedom). Such tori exist and are conjectured to have parametric equations that are non analytic functions of the perturbation parameter. This will be discussed briefly later: we show here how *resummations* of the Lindstedt series can be actually performed in the case of the KAM theory leading to simpler *renormalized* series, for instance, for the parametric equations of the invariant tori $\underline{h}, \underline{H}$ considered in proposition (8.1.1).

We consider the case of perturbations which are even trigonometric polynomials depending only on the angle variables, and impose the strong Diophantine condition, definition (8.3.2), with scale factor $\gamma \equiv 2$ on the rotation vector $\underline{\omega}_0$ of the invariant torus that we study. In this case we know from the analysis in Chapter VIII that the power expansion in ε for the parametric equation $\underline{h}(\underline{\psi})$ converges for small ε. However, conceptually, *the analysis here is independent of the one developed in Chapter VIII* and it develops an independent proof of the KAM theorem, which at the same time leads to a natural approach to the study of some invariant tori of dimension lower than the maximal one, namely the hyperbolic ones.

In terms of the power expansion envisaged in the previous Chapter, we can define a solution "approximated to order k" as

e9.1.1
$$\underline{h}^{(\le k)}(\underline{\psi},\varepsilon)=\sum_{\underline{\nu}\in\mathbb{Z}^\ell} e^{i\underline{\nu}\cdot\underline{\psi}}\underline{h}^{(\le k)}_{\underline{\nu}}(\varepsilon),\qquad \underline{h}^{(\le k)}_{\underline{\nu}}(\varepsilon)=\sum_{k'=1}^{k}\varepsilon^{k'}\underline{h}^{(k')}_{\underline{\nu}},\tag{9.1.1}$$

where $\underline{h}^{(k)}_{\underline{\nu}}$ is defined in (8.2.5) and $\underline{h}^{(k)}_{\underline{0}}=\underline{0}$.

The propagators G_λ introduced in (8.2.3) are matrices (proportional to the identity). Given a branch $\lambda = v'v$ it carries two tensor labels $j'_{v'}, j_v$ associated with the nodes v' and v, respectively. The propagator labels are contracted with the tensor labels of the line λ_v which are then summed over
N9.1.1 all the ℓ possible values.[1]

Note in particular that the propagators $G_\lambda \equiv G(\underline{\omega}_0 \cdot \underline{\nu}_\lambda)$ satisfy (trivially) the relations

e9.1.2
$$G^T(-x)=G^\dagger(x)=G(x),\tag{9.1.2}$$

which will play a crucial rôle in the following; here and henceforth superscripts ^{T} and † on a matrix denote, respectively, the transposed and the adjoint (*i.e.* the conjugated transposed) of a matrix.

We define scales and clusters exactly as in Chapter VIII using $\gamma \equiv 2$ as a scale factor.

[1] Since G_λ is diagonal this means the we have to take $j_v = j'_{v'}$ and sum over this common value.

If T is a self-energy graph in a tree ϑ, we denote by $V(T)$ the set of nodes in T, by $\Lambda(T)$ the set of branches in T (see definition (8.3.3) for details), by k_T the number of nodes in T (*i.e.* $k_T = |V(T)|$), and by λ_T^1 and λ_T^2 the branches, respectively, exiting and entering T.

Let ϑ be a tree $\vartheta \in \Theta_{k,\underline{\nu},j}$, cf. item (vii) in definition (8.2.1), with a self-energy subgraph T. Define $\vartheta_0 = \vartheta \setminus T$ as the set of nodes and branches of ϑ outside T (of course ϑ_0 is not a tree), we define $V(\vartheta_0) = V(\vartheta) \setminus V(T)$ and $\Lambda(\vartheta_0) = \Lambda(\vartheta) \setminus \Lambda(T)$. Consider simultaneously all trees such that the structure ϑ_0 outside of the self-energy graph is the same, while the self-energy graph itself can be arbitrary, *i.e.* T can be replaced by any other self-energy graph T' with $k_{T'} \geq 1$. This allows us to define as a formal power series the matrix

e9.1.3
$$
\begin{aligned}
&M(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon) = \sum_{\vartheta=\vartheta_0 \cup T'} \mathcal{V}_{T'}(\underline{\omega}_0 \cdot \underline{\nu}), \qquad \text{where} \\
&\mathcal{V}_T(\underline{\omega}_0 \cdot \underline{\nu}) \stackrel{def}{=} \varepsilon^{k_T} \Big(\prod_{v \in V(T)} F_v \Big) \Big(\prod_{\lambda \in \Lambda(T)} G_\lambda \Big),
\end{aligned}
\tag{9.1.3}
$$

where the sum is over all trees ϑ such that $\vartheta \setminus T$ is fixed to be ϑ_0 and the mode labels of the nodes $v \in V(T)$ have to satisfy the conditions defining the self-energy graphs (see (8.3.3)).

Recall that a self-energy graph T has *height* $p = 0$ if it does not contain any other self-energy graphs, and that it has height $p \in \mathbb{N}$, recursively, if it contains maximal self-energy graphs with height $p - 1$ (see definition following (8.4.1)).

The following property holds as an algebraic identity between formal power series.

L9.1.1 **(9.1.1) Lemma:** (Symmetries of the propagators)
The following two properties hold:
(1) $(M(x;\varepsilon))^T = M(-x;\varepsilon)$, *and*
(2) $(M(x;\varepsilon))^\dagger = M(x;\varepsilon)$;
The latter means that the matrix $M(x;\varepsilon)$ *is self-adjoint.*

Proof: Consider a graph computed with propagators verifying the properties (9.1.2), which is trivially valid in our case since the propagator is a real and diagonal matrix, cf. (8.2.3). Given a self-energy graph T with momentum $\underline{\nu}$ flowing through the entering branch λ_T^2, call $\mathcal{P}$ the path connecting the
N9.1.2 exiting branch λ_T^1 to the entering branch λ_T^2.[2] Then consider also the self-energy graph T' obtained by taking λ_T^1 as entering branch and λ_T^2 as exiting branch and by taking $-\underline{\nu}$ as momentum flowing through the (new) entering branch λ_T^1: in this way the arrows of all branches along the path $\mathcal{P}$ change orientation, while all the subtrees (internal to T) having the root in $\mathcal{P}$ are

[2] *i.e.* the minimal connected set of branches joining the branches λ_T^1 and λ_T^2 along the lines of the graph T.

left unchanged. This implies that the momenta of the branches belonging to $\mathcal{P}$ change sign, while all the other momenta do not change. Since all propagators G_λ corresponding to the branches $\lambda \in \mathcal{P}$ are transformed into G_λ^T, the property (9.1.2) implies that the entry ij of the matrix $M(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$ corresponding to the self-energy graph T is equal to the entry ji of the matrix $M(-\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$; then the property (1) is proved.

Given a self-energy graph T, consider also the self-energy graph T' obtained by reversing the sign of the mode labels of the nodes $v \in V(T)$, and by swapping the entering branch with the exiting one. In this way the arrows of all the branches along the path $\mathcal{P}$ joining the two external branches are reversed, while all the subtrees (internal to T) having the root in $\mathcal{P}$ are left unchanged (as before). One then realizes that the complex conjugate of $\mathcal{V}_{T'}(\underline{\omega}_0 \cdot \underline{\nu})$ equals $\mathcal{V}_T(\underline{\omega}_0 \cdot \underline{\nu})$, by using the form of the node factors in (8.2.8), and the fact that one has $f^*_{\underline{\nu}} = f_{-\underline{\nu}}$ (as $V(\underline{\alpha})$ is real) and $G^\dagger(\underline{\omega}_0 \cdot \underline{\nu}) = G(\underline{\omega}_0 \cdot \underline{\nu})$; this proves the property (2). ∎

Remark: The lemma has been proved without making use of the exact form of the propagator, but only exploiting the fact that it satisfies the property (9.1.2). Therefore it has a more general validity, a fact that will be exploited below.

The function $M(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$ depends on ε but, by construction, it is independent of ϑ_0: hence we can rewrite (9.1.3) as

e9.1.4
$$M(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon) = \sum_{T'} \mathcal{V}_{T'}(\underline{\omega}_0 \cdot \underline{\nu}), \tag{9.1.4}$$

where the sum is over all self-energy graphs of order $k \geq 1$ with external branches with momentum $\underline{\nu}$.

Note that, by setting $\underline{\omega} = 2^\tau C \underline{\omega}_0$, in (9.1.4), if $2^{n-1} \leq |\underline{\omega} \cdot \underline{\nu}| < 2^n$, the sum is restricted to the self-energy graphs T' on scale $n_{T'} \geq n+3$. If we define n^0_λ so that, if $\underline{\nu}^0_\lambda$ is the momentum which *would* flow on the line λ *if* the entering line had momentum $\underline{\nu} = \underline{0}$ (cf. comment following ♣8.3.5),

e9.1.5
$$2^{n^0_\lambda - 1} \leq \left|\underline{\omega} \cdot \underline{\nu}^0_\lambda\right| < 2^{n^0_\lambda}, \tag{9.1.5}$$

then, for all branches $\lambda \in \Lambda(\vartheta)$ one has $n_\lambda = n^0_\lambda$ because, for the same reasons discussed, in Sections §8.1 and §8.2, when shifting the branches external to the self-energy graphs of a tree ϑ, the scale labels n_λ of all branches $\lambda \in \Lambda(\vartheta)$ do not change.

So far we considered formal power expansions in ε. Instead of the function in (9.1.1), we can define a different sequence $\{\overline{\underline{h}}^{[d]}(\underline{\psi}, \varepsilon)\}_{d \in \mathbb{N}}$ of *approximating functions* converging to the solution (as we shall see below), by defining it iteratively as follows.

D9.1.1 **(9.1.1) Definition:** (Renormalized graphs and clusters)
Denote by $\Theta^{\mathcal{R}}_{k,\underline{\nu},j}$ *the set of all trees of order* k *without self-energy graphs*

and with labels $\underline{\nu}_{\lambda_0} = \underline{\nu}$ *and* $j_{\lambda_0} = j$ *associated with the root branch; we shall call* $\Theta^{\mathcal{R}}_{k,\underline{\nu}}$ *the set of* renormalized trees *of order* k *and with label* $\underline{\nu}$ *associated with the root branch. Given a tree* $\vartheta \in \Theta^{\mathcal{R}}_{k,\underline{\nu}}$ *and a cluster* T *in* ϑ*, by extension we shall say that* T *is a* renormalized cluster*.*
We can also consider a self-energy graph which does not contain any other self-energy graph: we shall say that such a self-energy graph is a renormalized self-energy graph*; of course no one of such clusters can appear in any tree in* $\Theta^{\mathcal{R}}_{k,\underline{\nu},j}$*.*

For a renormalized tree ϑ of arbitrary order k, define

e9.1.6
$$\overline{\underline{\rm Val}}^{[d]}(\vartheta) \stackrel{def}{=} \Big(\prod_{v\in V(\vartheta)} F_v\Big)\Big(\prod_{\lambda\in\Lambda(\vartheta)} \overline{G}^{[d-1]}_\lambda\Big), \tag{9.1.6}$$

with the *dressed propagators* given by the matrices

e9.1.7
$$\begin{cases} \overline{G}^{[0]}_\lambda = (\underline{\omega}_0\cdot\underline{\nu}_\lambda)^{-2}\,, \\ \overline{G}^{[d]}_\lambda = \Big[(\underline{\omega}_0\cdot\underline{\nu}_\lambda)^2 - M^{[d]}(\underline{\omega}_0\cdot\underline{\nu}_\lambda;\varepsilon)\Big]^{-1}, \quad \text{for } d\ge 1\,, \end{cases} \tag{9.1.7}$$

and the sequence $\{M^{[d]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon)\}_{d\in\mathbb{N}}$ is iteratively defined as sum of the values of all renormalized self-energy graphs that can be inserted on a line of momentum $\underline{\nu}$ and that are computed by using the propagators $\overline{G}^{[d-1]}_\lambda$, *i.e.* as

e9.1.8
$$\begin{aligned} M^{[d]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon) &= \sum_{\text{renormalized}\,T} \mathcal{V}^{[d]}_T(\underline{\omega}_0\cdot\underline{\nu}), \\ \mathcal{V}^{[d]}_T(\underline{\omega}_0\cdot\underline{\nu}) &= \varepsilon^{k_T}\Big(\prod_{v\in V(T)} F_v\Big)\Big(\prod_{\lambda\in\Lambda(T)} \overline{G}^{[d-1]}_\lambda\Big); \end{aligned} \tag{9.1.8}$$

We set $M^{[0]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon)\equiv 0$. If a line λ has scale $-d_\lambda$ the self energy graphs that can be inserted on λ necessarily consist of lines with scales $< -d_\lambda$ *therefore* $M^{[d]}(\underline{\omega}\cdot\underline{\nu}_\lambda)\equiv M^{[d_\lambda]}(\underline{\omega}\cdot\underline{\nu})$ for $d>d_\lambda$.

To avoid confusing the value of a renormalized tree with the tree value introduced in (8.2.7), we shall call (9.1.6) the *renormalized value* of the (renormalized) tree. Then we shall write

e9.1.9
$$\begin{aligned} &\overline{\underline{h}}^{[d]}(\underline{\psi},\varepsilon) = \sum_{\underline{\nu}\in\mathbb{Z}^\ell} e^{i\underline{\nu}\cdot\underline{\psi}}\,\overline{\underline{h}}^{[d]}_{\underline{\nu}}(\varepsilon), \\ &\overline{\underline{h}}^{[d]}_{\underline{\nu}}(\varepsilon) = \sum_{k'=1}^{\infty} \varepsilon^{k'}\,\overline{\underline{h}}^{[d,k]}_{\underline{\nu}}(\varepsilon), \qquad \overline{\underline{h}}^{[d,k]}_{\underline{\nu}}(\varepsilon) = \sum_{\vartheta\in\Theta^{\mathcal{R}}_{k,\underline{\nu}}} \overline{\underline{\rm Val}}^{[d]}(\vartheta), \end{aligned} \tag{9.1.9}$$

where the last formula holds for $\underline{\nu}\ne\underline{0}$, because for $\underline{\nu}=\underline{0}$ one has $\overline{\underline{h}}^{[d,k]}_{\underline{0}}\equiv 0$, cf. (9.1.1).

Remark: Note that if we expand the quantity $M^{[d]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon)$ in powers of ε, by expanding the propagators $\overline{G}^{[d-1]}_\lambda$, we reconstruct the sum of the values of all self-energy graphs containing only self-energy graphs with height $p < d$ (recall the definition of height given after (9.1.3)). Therefore if we expand $M^{[d+1]}(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$ in powers of ε we obtain the same terms that we would obtain by expanding $M^{[d]}(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$, plus the sum of the values of all the self-energy graphs containing also self-energies graphs with height d, which are absent in the self-energy graphs contributing to $M^{[d]}(\underline{\omega}_0 \cdot \underline{\nu}; \varepsilon)$. Such a result can be used in order to prove that the power series defining the functions $\overline{\underline{h}}^{[d]}_{\underline{\nu}}(\varepsilon)$, truncated at order $k < d$, coincide with the functions $\underline{h}^{(\le k)}_{\underline{\nu}}(\varepsilon)$ given by (9.1.1), as it is not difficult to check (see problem [8.3.1]).

In the coming section we shall derive the following renormalization result.

P9.1.1 **(9.1.1) Proposition:** (Resummation and renormalization of Lindstedt series)
(i) The matrices $M^{[d]}(x, \varepsilon)$ defined in (9.1.8) admit a limit $M^{[\infty]}((x, \varepsilon)$ as $d \to \infty$ which is analytic in ε and is defined for all real $x \ne 0$ and verifies, for all ε small enough

e9.1.10
$$\|M^{[d]}(x, \varepsilon)\| \le C\, \varepsilon^2\, x^2 \tag{9.1.10}$$

with an ε–independent constant C.
(ii) The series obtained by considering all renormalized tree graphs (i.e. all graphs without any self–energy subgraph) and computing their values by the rules of definition (8.2.1) but replacing the propagator $\delta_{j'j}(\underline{\omega}_o \cdot \underline{\nu})^{-2}$ by the matrix

e9.1.11
$$((\underline{\omega}_0 \cdot \underline{\nu})^2 + M^{[\infty]}(\underline{\omega}_0 \cdot \underline{\nu}, \varepsilon))^{-1} \tag{9.1.11}$$

is convergent for $|\varepsilon|$ small.
(iii) The functions $t \to \underline{\psi} + \underline{\omega}_0 t + \underline{h}^{[\infty]}(\underline{\psi} + \underline{\omega}_0 t)$, where $\underline{h}^{[\infty]}$ is the sum of the series in item (ii), is analytic in ε for ε small and in $\underline{\psi} \in \mathbb{T}^\ell$ and solve the equations of motion for the Hamiltonian (8.1.1).
(iv) The function $\underline{h}^{[\infty]}$ defines the parametric equations for an invariant torus with rotation vector $\underline{\omega}_0$. Therefore it coincides with the function $\underline{h}$ considered in proposition (8.1.1).

Remark: This means that the resummation procedure leads to a renormalized series with propagators (9.1.11) which is a representation of the function $\underline{h}$ whose existence is proved by the KAM theorem. At the same time this yields an independent proof of the KAM theorem. We devote Section §(9.2) to a proof of the above proposition.

§9.2 Bounds on renormalized series. Convergence

Let $\|\cdot\|$ be an algebraic matrix norm (*i.e.* a norm which verifies $\|AB\| \le \|A\|\,\|B\|$ for all matrices A and B); for instance $\|\cdot\|$ can be $\|A\| =$

$\max_w \|Aw\|$, with the maximum taken over all vectors w of norm 1, *i.e.* $||w||^2 = \sum_j |w_j|^2 = 1$.

L9.2.1 **(9.2.1) Lemma:** (Uniform bounds on resummed series)
Assume that the propagators $\overline{G}^{[d]}_\lambda \equiv \overline{G}^{[d]}(\underline{\omega}\cdot\underline{\nu}_\lambda;\varepsilon)$ *satisfy*

e9.2.1
$$\Big(\overline{G}^{[d]}(x;\varepsilon)\Big)^T = \overline{G}^{[d]}(-x;\varepsilon), \qquad \big\|\overline{G}^{[d]}(x;\varepsilon)\big\| < \frac{2}{x^2} \tag{9.2.1}$$

for all $|\varepsilon| < \varepsilon_0$*, if* ε_0 *is small enough. Then given* $\kappa > 0$ *there is a constant* B_f *such that one has, setting* $M(\vartheta) = \sum_{v\in V(\vartheta)} |\underline{\nu}_v|$,

e9.2.2
$$\begin{aligned} &\overline{h}^{[d,V]}_{\underline{\nu}} \le \frac{1}{V!}\sum_{\vartheta\in\Theta^{\mathcal{R}}_{V,\underline{\nu}}} \Big|\overline{\mathrm{Val}}^{[d]}(\vartheta)\Big| \le \left(|\varepsilon|B_f\right)^V,\\ &\frac{1}{V!}\sum_{\substack{\vartheta\in\Theta^{\mathcal{R}}_{V,\underline{\nu}}\\ M(\vartheta)=s}} \Big|\overline{\mathrm{Val}}^{[d]}(\vartheta)\Big| \le \left(|\varepsilon|B_f\right)^V e^{-\kappa s/4}, \end{aligned} \tag{9.2.2}$$

for all $s > 0$.

Proof: The hypothesis (9.2.1) implies that for all propagators $\overline{G}^{[d]}_\lambda$ one has

e9.2.3
$$\Big\|\overline{G}^{[d]}_\lambda\Big\| \le C_1 2^{-2n_\lambda}, \qquad C_1 = 2\left(2^{\tau+2}C\right)^2. \tag{9.2.3}$$

where C is the constant in definition (8.3.2). Therefore the contribution from a single tree is bounded for all $n_0 \le 0$ by

e9.2.4
$$|\varepsilon|^V \left(C_1\, 2^{-2n_0}\right)^V \prod_{v\in V(\vartheta)} \Big[N^{m_v+1} f_0\Big(\prod_{n=-\infty}^{n_0} C_1 2^{-ck2^{n/\tau}}\Big)\Big], \tag{9.2.4}$$

where f_0 is a bound of the size of the Fourier coefficients of the perturbation, cf. (8.4.4), $V = |V(\vartheta)|$, having used that, for all trees $\vartheta \in \Theta^{\mathcal{R}}_{k,\underline{\nu},j}$, the number $N_n(\vartheta)$ of branches with scale n in ϑ satisfy the bound

e9.2.5
$$N_n(\vartheta) \le c\,k\,2^{n/\tau} \tag{9.2.5}$$

for some constant c (see lemma (8.3.1) where c is shown to be bounded by $4\,E^{-1}$.

We then can simply proceed as we did in the analysis of the convergence of the Lindstedt series when the tree values of trees not verifying property [P], introduced before lemma (8.2.2), were ignored. The only difference being that the constant C_1 which bounds the propagators is now different from the constant C and is given by (9.2.3).

Finally the last factor in the second of Eq. (9.2.2) can be inserted by further increasing the constant that multiplies ε because $\sum_v |\underline{\nu}_v| \le kN$ as

f is a trigonometric polynomial of degree N and $1 = e^{\kappa\sum_v |\underline{\nu}_v|} e^{-\kappa\sum_v |\underline{\nu}_v|} \le e^{N\kappa k} e^{-\kappa\sum_v |\underline{\nu}_v|}$. ■

Remark: Note that, although the propagators are no longer diagonal, they still satisfy the same property as (9.1.2), which was the crucial one which is used in order to prove the formal cancellations between tree values.

L9.2.2 **(9.2.2) Lemma:** (Symmetry and cancellations properties)
The matrices $M^{[d]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon)$ satisfy the relation

e9.2.6
$$\Big(M^{[d]}(x;\varepsilon)\Big)^T = M^{[d]}(-x;\varepsilon). \tag{9.2.6}$$

Moreover the matrix $M^{[-q]}(x;\varepsilon)$ is the restriction to the x's of the form $x=\underline{\omega}\cdot\underline{\nu}$ with $\underline{\nu}$ of scale $\le q$ of an analytic function of x in the disk $|x|\le 2^q$ if ε_0 is small enough; it satisfies the bound

e9.2.7
$$\Big\|M^{[d]}(x;\varepsilon)\Big\| \le Dx^2|\varepsilon|^2, \tag{9.2.7}$$

for all $d\in\mathbb{N}$ and for a suitable d–independent constant D. As a consequence $\overline{G}^{[d]}_\lambda$ verify (9.2.1) for all $d\ge 1$, and therefore (9.2.2) holds for all $d\ge 1$.

Proof. We consider the matrices $M^{[d]}$ defined in (9.1.8) and suppose *inductively* that $M^{[d]}$ verifies (9.2.7) and the analyticity property preceding it for $0\le d\le p-1$; note that the assumption holds trivially for $d=0$. Note also that (9.2.6) imply that the propagators $\overline{G}^{[d]}_\lambda$ verify (9.2.1) for ε_0 small enough.

To define $M^{[p]}$ we must consider the renormalized self-energy graphs T and evaluate their values by using the propagators $\overline{G}^{[p-1]}_\lambda$, according to (9.1.8).

Given $\underline{\omega}\cdot\underline{\nu}$ such that $2^{q-1}\le|\underline{\omega}\cdot\underline{\nu}|<2^q$ for some $q\le 0$, the propagators $\overline{G}^{[p-1]}_\lambda$ have an analytic extension to the disk $|\underline{\omega}\cdot\underline{\nu}|<2^{q+2}$ and, under the hypotheses (9.2.6) and (9.2.7), verify the symmetry property and the bound in (9.2.1).

We have

e9.2.8
$$M^{[p]}(x;\varepsilon) = \sum_{h=q+3}^{0} \sum_{\substack{\text{renormalized } T\\ n_T=h}} \mathcal{V}^{[p]}_{T,h}(x), \tag{9.2.8}$$

where by appending the label h to $\mathcal{V}^{[p]}_T(x)$ we distinguish the contributions to $M^{[p]}(x;\varepsilon)$ coming from self-energy graphs T on scale h (which is constrained to be $\ge q+3$).

The value $\mathcal{V}^{[p]}_T(\underline{\omega}_0\cdot\underline{\nu})$ is analytic in $\underline{\omega}\cdot\underline{\nu}$ for $|\underline{\omega}\cdot\underline{\nu}|\le 2^{h+2}$, and the sum over all the self-energy graphs T with V nodes is bounded by

e9.2.9
$$\sum_{\substack{T\\ k_T=V}} \Big|\mathcal{V}^{[p]}_{T,h}(\underline{\omega}_0\cdot\underline{\nu})\Big| \le \frac{(|\varepsilon|B_f)^V}{1-e^{-\kappa/8}} e^{-\kappa 2^{-h/\tau}/8}, \tag{9.2.9}$$

because the mode labels $\underline{\nu}_v$ of the nodes $v \in V(T)$ must satisfy $\sum_{v\in V(T)} |\underline{\nu}_v| > 2^{-h/\tau}$ (recall that we are dealing with renormalized trees, so that for all clusters T one has $\widetilde{T} = T$, *i.e.* T contains no self-energy graph, cf. definition (8.3.5)).

Since the symmetry property expressed by (9.2.1) for $d = p-1$ is implied by(9.2.6) and this is the only property of the propagators that one needs in order to check the algebraic cancellations, we can conclude that the same cancellation mechanisms extend to the renormalized self-energy values $\mathcal{V}_T^{[p]}(\underline{\omega}_0 \cdot \underline{\nu})$. Therefore we see that $\mathcal{V}_{T,h}^{[p]}(x)$ will vanish at $x = 0$ to order 2.

By the analyticity in $\underline{\omega} \cdot \underline{\nu}$ for $|\underline{\omega} \cdot \underline{\nu}| \le 2^{h+2}$ and by the maximum principle we deduce from (9.2.9) that one has

e9.2.10 $$\sum_{\substack{T\\ k_T = V}} \left|\mathcal{V}_{T,h}^{[p]}(x)\right| \le \frac{(|\varepsilon| B_f)^V}{1 - e^{-\kappa/8}} e^{-\kappa 2^{-h/\tau}/8} \left(\frac{2^\tau C x}{2^{h+2}}\right)^2, \qquad (9.2.10)$$

Therefore we can use that $\sum_{h=q+3}^{0} e^{-\kappa 2^{-h/\tau}} 2^{-2h} < B_1 < \infty$ and that $V \ge 2$, and the proof is complete. ■

It also follows that there exists the limit (reached at a finite value of d if x is fixed because $M^{[d]}(x;\varepsilon)$ becomes identically equal to $M^{[d_0]}(x;\varepsilon)$ if if $x = \underline{\omega} \cdot \underline{\nu}_\lambda$ with λ on scale $-d_0$))

e9.2.11 $$\lim_{d\to\infty} M^{[d]}(x;\varepsilon) = M^{[\infty]}(x;\varepsilon), \qquad (9.2.11)$$

with $M^{[\infty]}(x;\varepsilon)$ analytic in ε for $|\varepsilon| < \varepsilon_0$: in fact the following result holds.

L9.2.3 **(9.2.3) Lemma:** (Further uniform bounds on resummed Lindstedt series) *For all $d \ge 1$ one has*

e9.2.12 $$\left\| M^{[d+1]}(x;\varepsilon) - M^{[d]}(x;\varepsilon) \right\| \le \hat{B}_1 \hat{B}_2^d \varepsilon^{2d} x^2, \qquad (9.2.12)$$

for some constants $\hat{B}_1$ and $\hat{B}_2$ and for $|\varepsilon| < \varepsilon_0$ with ε_0 small enough.

Proof: For $-d <$ scale of x the difference is 0. For $-d \ge$ scale of x this is implied by (9.2.10). ■

We can now define the "fully renormalized" expansion of the parametric equations of the invariant torus as the sum of the values of the renormalized trees evaluated according to (9.1.8) with $\overline{G}_\lambda^{[d-1]}$ replaced by

e9.2.18 $$\overline{G}^{[\infty]}(x;\varepsilon) = \left(x^2 - M^{[\infty]}(x;\varepsilon)\right)^{-1}, \qquad x = \underline{\omega}_0 \cdot \underline{\nu}_\lambda. \qquad (9.2.13)$$

The above discussion shows that the series

e9.2.19 $$\begin{aligned} \underline{h}^{[\infty]}(\underline{\psi},\varepsilon) &= \sum_{k=1}^{\infty} \sum_{\underline{\nu}\in\mathbb{Z}^\ell} \varepsilon^k e^{i\underline{\nu}\cdot\underline{\psi}} \sum_{\vartheta\in\Theta^{\mathcal{R}}_{k,\underline{\nu}}} \underline{\mathrm{Val}}^{[\infty]}(\vartheta), \\ \underline{\mathrm{Val}}^{[\infty]}(\vartheta) &= \Big(\prod_{v\in V(\vartheta)} F_v\Big)\Big(\prod_{\lambda\in\Lambda(\vartheta)} \overline{G}_\lambda^{[\infty]}\Big), \end{aligned} \qquad (9.2.14)$$

converges for $|\varepsilon| < \varepsilon_0$ and that it coincides with the limit for $d \to \infty$ of $\overline{\underline{h}}^{[d]}(\underline{\psi};\varepsilon)$, which therefore exists. Moreover the function $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$, which we have just shown to be well defined and analytic for ε small enough, solves the equations of motion, as the following lemma shows.

L9.2.4 **(9.2.4) Lemma:** Convergence of renormalized series)
One has, formally (i.e. order by order in the expansion in ε around $\varepsilon = 0$)

e9.2.20
$$\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon) \equiv \lim_{d\to\infty} \overline{\underline{h}}^{[d]}(\underline{\psi};\varepsilon) = \underline{h}(\underline{\psi};\varepsilon), \tag{9.2.15}$$

where $\underline{h}(\underline{\psi};\varepsilon)$ is the formal power series which solves the equations of motion. The function $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ solves, therefore, the equations of motion.

Remark: (1) In the formulation of the above lemma we combine the information about the existence and analyticity of the function $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ which has been proved in the previous lemmata and the existence of the formal power series for $\underline{h}(\underline{\psi},\varepsilon)$, about whose convergence *we make no statement here* as we do not want to make use of the KAM result of Section §8.1, but we want to derive it again by the alternative approach that we are following.
(2) If we show that the analytic function in the l.h.s. of (9.2.15) has a power series at $\varepsilon = 0$ whose coefficients coincide with those of the Lindstedt series (which is a formal series) for $\underline{h}(\underline{\psi})$ then it follows that the Lindstedt series necessarily converges and its sum coincides with $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$. Furthermore the function $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ solves the equations of motion in the sense that $t \to \underline{\alpha}(t) = \underline{\psi} + \underline{\omega}t + \underline{h}^{[\infty]}(\underline{\psi} + \underline{\omega}t,\varepsilon)$ is a solution of the equations of motion $\ddot{\underline{\alpha}} = -\underline{\partial}V(\underline{\alpha})$: here we use that if two analytic functions (in our case the two sides of (8.1.12) evaluated with $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ in place of $\underline{h}(\underline{\psi},\varepsilon)$) have equal derivatives of all orders at $\varepsilon = 0$ then they coincide.
(3) The above remarks show the validity of the lemma: it is however interesting to check its validity directly by substituin of the series for $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ into the equation (8.1.12) that it solves. Therefore we discuss how to perform the check. This provides a useful method in similar cases in which it is not known or it is not true that generalizations of Lindstedt series converge and their solution is produced by summation rules of divergent series: see Appendix 9.2 for an example in which this situation arises.

Proof: We denote by $\overline{G}^{[\infty]}$ the operator with kernel $\overline{G}^{[\infty]}(\underline{\omega}_0\cdot\underline{\nu};\varepsilon)$ in Fourier space, and we represent (9.2.14), in a more compact notations, as

e9.2.21
$$\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon) = \sum_{\vartheta\in\Theta^{\mathcal{R}}} \overline{\underline{\rm Val}}^{\mathcal{R}}(\vartheta;\underline{\psi},\varepsilon), \tag{9.2.16}$$

where $\Theta^{\mathcal{R}}$ is the set of all renormalized trees, and, for $\vartheta \in \Theta^{\mathcal{R}}_{k,\underline{\nu}} \subset \Theta^{\mathcal{R}}$, we have defined

e9.2.22
$$\overline{\underline{\rm Val}}^{\mathcal{R}}(\vartheta;\underline{\psi},\varepsilon) = \varepsilon^k e^{i\underline{\nu}\cdot\underline{\psi}}\,\overline{\underline{\rm Val}}^{[\infty]}(\vartheta). \tag{9.2.17}$$

The function $\underline{h}(\underline{\psi},\varepsilon)$ solving the equations of motion is formally defined as the solution of the functional equation

e9.2.23
$$\underline{h}(\underline{\psi},\varepsilon) = G\partial_{\underline{\psi}} f\left(\underline{\psi}+\underline{h}(\underline{\psi},\varepsilon)\right), \tag{9.2.18}$$

where $G = (i\underline{\omega}_0\cdot\partial)^{-2} = \overline{G}^{[0]}$ is the operator with kernel $G(x) = x^{-2}$.

We have (see problem [8.3.1])

e9.2.24
$$G(x)\left(M^{[\infty]}(x;\varepsilon) + (G^{[\infty]}(x;\varepsilon))^{-1}\right) = 1. \tag{9.2.19}$$

Then it is easy to realize that the function $\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)$ solves the equation of motions; one can reason as follows. One has

e9.2.25
$$\begin{aligned} G\partial_{\underline{\psi}} f\left(\underline{\psi}+\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)\right) &= G\sum_{m=0}^{\infty}\frac{1}{m!}\partial_{\underline{\psi}}^{m+1} f(\underline{\psi})\left(\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)\right)^m \\ &= G\sum_{m=0}^{\infty}\frac{1}{m!}\partial_{\underline{\psi}}^{m+1} f(\underline{\psi})\sum_{\vartheta_1\in\Theta^{\mathcal{R}}}\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta_1;\underline{\psi},\varepsilon)\ldots\sum_{\vartheta_m\in\Theta^{\mathcal{R}}}\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta_m;\underline{\psi},\varepsilon) \\ &= G\left(\overline{G}^{[\infty]}\right)^{-1}\sum_{\vartheta\in\Theta^{\mathcal{R}}_*}\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta;\underline{\psi},\varepsilon), \end{aligned} \tag{9.2.20}$$

where $\Theta^{\mathcal{R}}_*$ differs from $\Theta^{\mathcal{R}}$ as it contains also trees which can have only one self-energy graph with exiting branch λ_0, if, as usual, λ_0 denotes the root branch of ϑ; the operator $G(\overline{G}^{[\infty]})^{-1}$ takes into account the fact that, by construction, to the root branch λ_0 an operator G is associated, while in $\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta;\underline{\psi},\varepsilon)$, by definition, a propagator $\overline{G}^{[\infty]}$ is associated.

Then we can write (9.2.20), by explicitly separating the trees containing such a self-energy graph from the others,

e9.2.26
$$\begin{aligned} &G\partial_{\underline{\psi}} f\left(\underline{\psi}+\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)\right) \\ &\quad= G\left(\overline{G}^{[\infty]}\right)^{-1}\left(\overline{G}^{[\infty]}M^{[\infty]}\sum_{\vartheta\in\Theta^{\mathcal{R}}}\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta,\underline{\psi},\varepsilon) + \sum_{\vartheta\in\Theta^{\mathcal{R}}}\overline{\underline{\mathrm{Val}}}^{\mathcal{R}}(\vartheta,\underline{\psi},\varepsilon)\right) \\ &\quad= G\Big(M^{[\infty]}\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon) + (\overline{G}^{[\infty]})^{-1}\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon)\Big) \\ &\quad= G\Big(M^{[\infty]} + (\overline{G}^{[\infty]})^{-1}\Big)\overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon) = \overline{\underline{h}}^{[\infty]}(\underline{\psi},\varepsilon), \end{aligned} \tag{9.2.21}$$

where the property (9.2.19) has been used in the last line.

Note that at each step only absolutely converging series have been dealt with; the lemma is thus proved. ∎

This completes the alternative proof of proposition (8.1.1). We can take full advantage of such a different approach when dealing with hyperbolic lower-dimensional tori (see Appendix (9.2)).

Appendix 9.2: Resonances and low dimensional invariant tori

Consider the Hamiltonian

eA9.2.1
$$\mathcal{H}(\underline{A},\underline{\alpha}) = \underline{\omega}_0 \cdot \underline{A} + \frac{1}{2}\underline{A}\cdot\underline{A} + \frac{1}{2}\underline{B}\cdot\underline{B} + \varepsilon f(\underline{\alpha},\underline{\beta}), \tag{A9.2.1}$$

where $(\underline{\alpha},\underline{A}) \in \mathbb{T}^r \times \mathbb{R}^r$ and $(\underline{\beta},\underline{B}) \in \mathbb{T}^s \times \mathbb{R}^s$ are conjugated variables, and $\underline{\omega}_0$ is a vector in $\mathbb{R}^r$ satisfying the Diophantine property $C\,|\underline{\omega}_0 \cdot \underline{\nu}| > |\underline{\nu}|^{-\tau}$, $\forall \underline{\nu} \in \mathbb{Z}^r \setminus \{\underline{0}\}$, with $C > 0$ and $\tau \geq r-1$, and $f(\underline{\alpha},\underline{\beta}) = \sum_{\underline{\nu}\in\mathbb{Z}^r} e^{i\underline{\nu}\cdot\underline{\alpha}} f_{\underline{\nu}}(\underline{\beta})$.

The motions $\underline{\alpha}(t) = \underline{\alpha}_0 + \underline{\omega}_0 t$, $\underline{\beta}(t) = \underline{\beta}_0$, $\underline{A} = \underline{0}$, $\underline{B} = \underline{0}$ define, if $\varepsilon = 0$, an invariant torus of dimension $r+s$ which is called *resonant* and which is foliated into invariant tori of dimension r on which the motion is quasi-periodic and ergodic (parameterized by the s angles $\underline{\beta}_0$).

The question is whether the resonant tori continue to exist, slightly deformed, when ε is > 0 and small. This means that we ask whether there are solutions of the form

eA9.2.2
$$\begin{cases} \underline{\alpha}(t) = \underline{\psi}(t) + \underline{a}(\underline{\psi}(t),\underline{\beta}_0;\varepsilon), \\ \underline{\beta}(t) = \underline{\beta}_0 + \underline{b}(\underline{\psi}(t),\underline{\beta}_0;\varepsilon), \end{cases} \tag{A9.2.2}$$

for some functions $\underline{a}$ and $\underline{b}$, real analytic and 2π-periodic in $\underline{\psi} \in \mathbb{T}^r$, such that the motion in the variable $\underline{\psi}$ is $\dot{\underline{\psi}} = \underline{\omega}_0$.

The problem can be naturally studied via the resummation method used for the maximal tori: however not all invariant tori will still exist for $\varepsilon > 0$, at least not in general. Only invariant tori which are close to the unperturbed ones which have $\underline{\beta}_0$ coinciding with a stationary point for the function $f_{\underline{0}}(\underline{\beta}) = (2\pi)^{-r}\int d\underline{\alpha}\, f(\underline{\alpha},\underline{\beta})$ and which correspond to strict maxima of $f_{\underline{0}}(\underline{\beta})$ can be shown to exist for $\varepsilon > 0$ small and to be analytically reducible to the unperturbed ones.

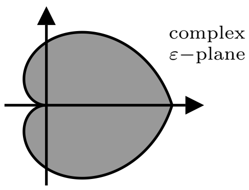
P9.2.1 **(9.2.1) Proposition:** (Hyperbolic tori, [GG02])
Consider the equations of motion $\ddot{\underline{\alpha}} = -\underline{\partial}_{\underline{\alpha}} f(\underline{\alpha},\underline{\beta})$, $\ddot{\underline{\beta}} = -\underline{\partial}_{\underline{\beta}} f(\underline{\alpha},\underline{\beta})$ *(the Hamiltonian equations for (A9.2.1)), and suppose* $\underline{\omega}_0$ *to satisfy (8.1.6) and* $\underline{\beta}_0$ *to be such that*

eA9.2.3
$$\partial_{\underline{\beta}} f_{\underline{0}}(\underline{\beta}_0) = \underline{0}, \qquad \partial^2_{\underline{\beta}} f_{\underline{0}}(\underline{\beta}_0) \text{ is negative definite.} \tag{A9.2.3}$$

There exist a constant $\varepsilon_0 > 0$ *and, for all* $\varepsilon \in (0,\varepsilon_0)$*, two functions* $\underline{a}(\underline{\psi},\underline{\beta}_0;\varepsilon)$ *and* $\underline{b}(\underline{\psi},\underline{\beta}_0;\varepsilon)$*, real analytic and* 2π*-periodic in* $\underline{\psi} \in \mathbb{T}^r$*, such that (A9.2.2) is a solution of (A9.2.1) with* $\dot{\underline{\psi}} = \underline{\omega}_0$*. Moreover* $\underline{a}(\underline{\psi},\underline{\beta}_0;\varepsilon)$ *and* $\underline{b}(\underline{\psi},\underline{\beta}_0;\varepsilon)$ *are analytic in* ε *for* $\varepsilon \in (0,\varepsilon_0)$.

Remarks: (1) The solutions whose existence is stated by the theorem do not seem to be analytic in ε at $\varepsilon = 0$: however they are certainly analytic

for $\varepsilon > 0$ and small: it appears to be an open problem to prove their actual divergence (or convergence) for $|\varepsilon|$ small. Furthermore, if the second condition in (A9.2.3) is replaced by $\partial^2_{\underline{\beta}} f_{\underline{0}}(\underline{\beta}_0)$ *is positive definite* then the same conclusions hold for $\varepsilon \in (-\varepsilon_0, 0)$.
(2) The proof based on the resummation methods is not the only possible one, nor it has been the first in historical order, however it yields rather detailed information on the regularity of the considered tori. In particular the analyticity domain is much larger, see the heart-like domain D_0 in Fig.(9.2.1).
(3) From the point of view of stability theory the negative definiteness assumption in (A9.2.3) and $\varepsilon > 0$ imply that the invariant torus with $\underline{\beta}$ close to $\underline{\beta}_0$ is unstable and has positive Lyapunov exponents which, to first order, coincide with the eigenvalues of $-\varepsilon\partial^2_{\underline{\beta}} f(\underline{\beta}_0)$: for this reason we call them *hyperbolic tori*.



F9.2.1 **Fig.(9.2.1)** Analyticity domain D_0 for the hyperbolic invariant torus. The cusp at the origin is a second order cusp. The figure corresponds to the case in (A9.2.3) of the proposition (9.2.1).

A tempting conjecture is that the analytic functions $\underline{a}(\underline{\psi}), \underline{b}(\underline{\psi})$ admit a more extended domain of analyticity which touches the negative ε axis on a set Δ which has 0 as a density point and such that the limiting value, as ε approaches points on the negative real axis in Δ, of the functions $\underline{a}(\underline{\psi}), \underline{b}(\underline{\psi})$ is real and describes the parametrization of an "*elliptic invariant torus*". The conjectured form of the extended analyticity domain is represented in figure (9.2.2).

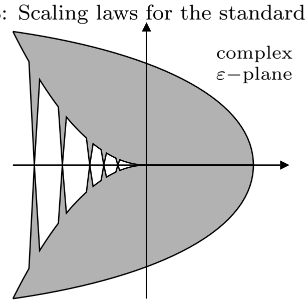
The above proposition is proved by the method of resummations in [GG02].

Problems for §9.2

Q9.2.1 **[9.2.1]:** Check that for any self-energy graph T with self-energy-scale n, one has $n_T \geq n+3$. (*Hint:* Writing, for any branch $\lambda \in \Lambda(T')$, $\underline{\nu}_\lambda$ as $\underline{\nu}_\lambda = \underline{\nu}^0_\lambda + \sigma_\lambda \underline{\nu}$, one has $|\underline{\omega}\cdot\underline{\nu}^0_\lambda| > 2^\tau C^{-1} C |\underline{\nu}^0_\lambda|^{-\tau} \geq 2^\tau (\sum_{v\in V(\widetilde{T})} |\underline{\nu}_v|)^{-\tau} \geq 2^\tau 2^{n+3}$, while $|\underline{\omega}\cdot\underline{\nu}| < 2^n$, so that one obtains $|\underline{\omega}\cdot\underline{\nu}_\lambda| > 2^\tau 2^{n+3} - 2^n > 2^{n+2}$, which implies $n_\lambda \geq n+3$.)

Q9.2.2 **[9.2.2]:** Check that the power series defining the functions $\overline{\underline{h}}^{[d]}_{\underline{\nu}}(\varepsilon)$, truncated at any order $k < d$, coincide with the functions $\underline{h}^{(\leq k)}_{\underline{\nu}}(\varepsilon)$ given by (9.1.1). (*Hint:* Compare the diagrammatic representations of the two functions.)

Q9.2.3 **[9.2.3]:** Prove (9.2.19). (*Hint:* One has $\overline{G}^{[\infty]}(x;\varepsilon) = (G^{-1}(x) - M^{[\infty]}(x;\varepsilon))^{-1}$, so that $G^{-1}(x) = (\overline{G}^{[\infty]}(x;\varepsilon))^{-1} + M^{[\infty]}(x;\varepsilon)$.)



F9.2.2 **Fig.(9.2.2)** The domain D_0 of Figure 1 can be further extended? The conjecture above asks whether the extended analyticity domain could possibly be represented (close to the origin) as here: the domain reaches the real axis at cusp points which are in a set Δ which has 0 as a density point: the points of Δ correspond to elliptic tori which are analytic continuations of the hyperbolic tori in the complex ε-plane. The analytic continuation would be continuous through the real axis at the points of Δ. The cusps would be at least quadratic.

Q9.2.4 **[9.2.4]:** (*Lindstedt series for resonant invariant tori*)
Look for an expansion

$$\underline{a}(\underline{\psi};\varepsilon)=\sum_{k=1}^{\infty}\varepsilon^k\underline{a}^{(k)}(\underline{\psi})=\sum_{\underline{\nu}\in\mathbb{Z}^r}e^{i\underline{\nu}\cdot\underline{\psi}}\underline{a}_{\underline{\nu}}(\varepsilon)=\sum_{k=1}^{\infty}\varepsilon^k\sum_{\underline{\nu}\in\mathbb{Z}^r}e^{i\underline{\nu}\cdot\underline{\psi}}\,\underline{a}_{\underline{\nu}}^{(k)},$$

$$\underline{b}(\underline{\psi};\varepsilon)=\sum_{k=1}^{\infty}\varepsilon^k\underline{b}^{(k)}(\underline{\psi})=\sum_{\underline{\nu}\in\mathbb{Z}^r}e^{i\underline{\nu}\cdot\underline{\psi}}\underline{b}_{\underline{\nu}}(\varepsilon)=\sum_{k=1}^{\infty}\varepsilon^k\sum_{\underline{\nu}\in\mathbb{Z}^r}e^{i\underline{\nu}\cdot\underline{\psi}}\,\underline{b}_{\underline{\nu}}^{(k)},$$

of the parametric equations for the resonant invariant tori in the context of proposition (9.2.1): the dependence on $\underline{\beta}_0$ has not been explicitly written. Check that to order k the equations of motion for the Hamiltonian (A9.2.1) have solutions that, written in the form (A9.2.2), become

$$(\underline{\omega}\cdot\underline{\nu})^2\,\underline{a}_{\underline{\nu}}^{(k)}=\big[\partial_{\underline{\alpha}}f\big]_{\underline{\nu}}^{(k-1)},$$
$$(\underline{\omega}\cdot\underline{\nu})^2\,\underline{b}_{\underline{\nu}}^{(k)}=\big[\partial_{\underline{\beta}}f\big]_{\underline{\nu}}^{(k-1)},$$

and show that they can be uniquely solved to all orders, recursively, under the only assumption that $\underline{\beta}_0$ is a *non degenerate* stationarity point for $f_{\underline{0}}(\underline{\beta})$ (*i.e.* $\det\partial^2_{\underline{\beta}}f_{\underline{0}}(\underline{\beta}_0)\neq 0$) and that $\underline{\alpha}^{(k)}$ has $\underline{0}$ average over $\underline{\psi}$. (*Hint:* it is essential that one does not assume that $\underline{\beta}^{(k)}$ has zero average over $\underline{\psi}$ and therefore also its average is determined (uniquely) recursively; the definiteness positive or negative of $\underline{\partial}^2_{\underline{\beta}}f(\underline{\beta}_0)$ is *not necessary* for this result.)

§9.3 Scaling laws for the standard map

In this section we shall consider another case in which the diagrammatic formalism developed in Chapter VIII can be naturally applied. The problem consists in considering what happens in some simple dynamical systems for complex values of the rotation vector; for simplicity one can take systems with two degrees of freedom, and with one component of the rotation vector fixed to some value (for instance 1), so that only one free component ω is left, and one can refers to it as the rotation number.

The main motivation comes from studying the optimal arithmetic dependence on the rotation number in order to have an invariant curve. It is known that, for analytic perturbations of two-dimensional integrable area-preserving maps, the best condition one can impose on the rotation number ω, whose continued fraction convergents are p_k/q_k (see problem [2.2.3]), is that the function $B_1(\omega) = \sum_{k=0}^{\infty} q_k^{-1} \log q_{k+1}$ is finite: if $B_1(\omega) < \infty$ one says that ω satisfies the Bryuno condition (which is weaker than the Diophantine condition). The function $B_1(\omega)$ is related to the *Bryuno function* $B(\omega)$ introduced by Yoccoz in studying Siegel's problem [Yo95]: the difference $B_1(\omega) - B(\omega)$ is an essentially bounded function.

Then one asks if it is possible to interpolate the radius of convergence $\rho(\omega)$ of the Lindstedt series for the standard map in terms of the Bryuno function, in the sense that there exists $\beta \in \mathbb{R}^+$ such that the quantity $Q_\beta(\omega) \equiv \log \rho(\omega) + \beta B(\omega)$ remains uniformly bounded, independently of ω (provided that ω satisfies the Bryuno condition, so that both $\rho(\omega)$ and $B(\omega)$ are well defined).

The possibility that $Q_2(\omega)$ be uniformly bounded is related, see Remark (1) to proposition (9.3.1), to the divergence rate of the conjugating function for $\omega = p/q + i\eta$, with $\gcd(p, q) = 1$ and $\eta \neq 0$. This is the problem that we shall consider in this and next sections.

The standard map is a discrete-time, one-dimensional dynamical system generated by the iteration of the area-preserving (symplectic) map of the cylinder into itself, $T_\varepsilon : \mathbb{T} \times \mathbb{R} \to \mathbb{T} \times \mathbb{R}$, given by

e9.3.1
$$T_\varepsilon : \begin{cases} x' = x + y + \varepsilon \sin x, \\ y' = y + \varepsilon \sin x. \end{cases} \tag{9.3.1}$$

It was first introduced by Greene [Gr79] and Chirikov [Ch79], and it can be considered as one of the simplest Hamiltonian dynamical systems with a nontrivial behaviour (see problem [9.3.1] for the interpretation of the standard map as the Poincaré map of a continuous-time Hamiltonian system).

The homotopically non-trivial invariant curves $\mathcal{C}_{\varepsilon,\omega}$ with rotation number ω of the map T_ε may be determined by changing coordinates on $\mathbb{T} \times \mathbb{R}$,

e9.3.2
$$\begin{cases} x = \alpha + u(\alpha, \varepsilon, \omega), \\ y = 2\pi\omega + v(\alpha, \varepsilon, \omega), \end{cases} \tag{9.3.2}$$

and imposing that the dynamics induced in the variables (α, ω) is given by the unperturbed map

e9.3.3
$$\begin{cases} \alpha' = \alpha + 2\pi\omega, \\ \omega' = \omega. \end{cases} \tag{9.3.3}$$

By (9.3.1) one has $y' = x' - x$ so that $v(\alpha, \varepsilon, \omega) = u(\alpha, \varepsilon, \omega) - u(\alpha - 2\pi\omega, \varepsilon, \omega)$; we can therefore consider only the function u.

The coordinate transformation (9.3.2) *conjugates* the dynamics on the invariant curve to a rotation, and the function u is called the *conjugating function.* It satisfies the functional equation

e9.3.4
$$D^2_\omega u(\alpha,\varepsilon,\omega) = \varepsilon \sin(\alpha + u(\alpha)), \tag{9.3.4}$$

where the operator D^2_ω acts on functions of α as follows:

e9.3.5
$$D^2_\omega \phi(\alpha) = \phi(\alpha + 2\pi\omega) - 2\phi(\alpha) + \phi(\alpha - 2\pi\omega). \tag{9.3.5}$$

Assuming that equations (9.3.4) admit ε-analytic solutions near $\varepsilon = 0$ and imposing that the average of u over α be 0 one checks that the Taylor coefficients in ε of $u(\alpha,\varepsilon,\omega)$ are uniquely determined and are odd as functions of α. The proofs of such statements can be carried out as in Section §8.1. However, as in the cases considered in Chapter VIII one needs to study, and prove, the convergence of the series.

To each smooth solution to (9.3.4) corresponds an invariant curve $\mathcal{C}_{\varepsilon,\omega}$ whose parametric equations are

e9.3.6
$$\mathcal{C}_{\varepsilon,\omega} : \begin{cases} x = \alpha + u(\alpha,\varepsilon,\omega), \\ y = 2\pi\omega + u(\alpha,\varepsilon,\omega) - u(\alpha - 2\pi\omega,\varepsilon,\omega), \end{cases} \tag{9.3.6}$$

and it has the same smoothness properties as those of $u(\cdot,\varepsilon,\omega)$.

To simplify the notations, we shall drop the dependence on ω and write just $u(\alpha,\varepsilon)$. The expansion in powers of ε of the conjugating function u (that we naturally call *Lindstedt series* because it is derived by following arguments analogous to those repeatedly discussed in Chapter VIII) has the form

e9.3.7
$$u(\alpha,\varepsilon) = \sum_{\nu\in\mathbb{Z}} u_\nu(\varepsilon)\, e^{i\nu\alpha} = \sum_{k\ge1} u^{(k)}(\alpha)\varepsilon^k = \sum_{k\ge1}\sum_{\nu\in\mathbb{Z}} u^{(k)}_\nu e^{i\nu\alpha}\,\varepsilon^k. \tag{9.3.7}$$

By inserting in (9.3.4) the series (9.3.7) and equating the Fourier and Taylor coefficients of both sides (see Section 8.1), one finds that the coefficients $u^{(k)}_\nu$ are defined by the recursion relations

e9.3.8
$$u^{(k)}_\nu = \frac{1}{\gamma(\nu)} \sum_{m\ge0} \frac{1}{m!} \sum_{\substack{\nu_0+\ldots+\nu_m=\nu \\ k_1+\ldots+k_m=k-1}} \frac{1}{2}(-i\nu_0)(i\nu_0)^m \prod_{j=1}^m u^{(k_j)}_{\nu_j}, \tag{9.3.8}$$

with $\nu_0 = \pm1$ and

e9.3.9
$$\gamma(\nu) = 2\,(\cos 2\pi\omega\nu - 1) = -4\sin^2 \pi\omega\nu, \tag{9.3.9}$$

for $\nu \ne 0$, while $u^{(k)}_0 = 0$ for all $k \ge 1$. The case $m = 0$ in (9.3.8) has to be interpreted as $u^{(k)}_\nu = (-i\nu_0)/\gamma(\nu)$, which imposes $k = 1$ and $\nu = \nu_0$.

The radius of convergence of the Lindstedt series is defined as

e9.3.10
$$\rho = \inf_{\alpha\in\mathcal{T}} \left(\limsup_{k\to\infty} \left| u^{(k)}(\alpha) \right|^{1/k} \right)^{-1} . \tag{9.3.10}$$

Of course the Lindstedt series is plagued by the *small divisors* problem, due to the fact that $\gamma(\nu)$ can be arbitrarily close to 0 for $\omega \in \mathbb{R}\setminus\mathbb{Q}$ and can be 0 for $\omega \in \mathbb{Q}$ (here $\mathbb{Q}$ denotes the real rational numbers). If $\omega \in \mathbb{R}$ satisfies a Diophantine property and ε is sufficiently small, ρ is strictly greater than 0 and therefore we have an analytic invariant curve and an analytic conjugation to a circle rotation. Here we are interested in the behavior of ρ as the rotation number ω tends to a *resonant* value, *i.e.* $\omega \to p/q$, with $p, q \in \mathbb{Z}$ and $\gcd(p, q) = 1$. The behavior of the radius of convergence near a resonant value of the rotation number is related to the problem of *Bryuno's interpolation* and to the problem of determining the optimal arithmetic condition on ω to have an analytic invariant curve, see below. We consider

e9.3.11
$$\omega = \frac{p}{q} + i\eta, \tag{9.3.11}$$

with $p, q \in \mathbb{Z}$, $\gcd(p, q) = 1$ and $\eta \in \mathbb{R}$, in the limit $\eta \to 0$. Later on we shall show how to extend our results to the case in which ω tends to p/q *along any path on the complex ω plane non-tangential to the real axis.*

We are interested in the exact (asymptotic) dependence of ρ on η, in the limit $\eta \to 0$. In particular, we shall prove the following result.

P9.3.1 **(9.3.1) Proposition:** (Asymptotics near resonances in the standard map) *Consider the standard map (9.3.1) with $\omega = p/q+i\eta$, $p, q \in \mathbb{Z}$, $\gcd(p, q) = 1$ and $\eta \in \mathbb{R}$. Then the following results hold.*
(i) For fixed $\eta \neq 0$ the function $u(\alpha, \varepsilon)$, defined by (9.3.2), is divisible by ε and jointly analytic in (α, ε) in the product of a strip of width $\delta_0 > 0$ around the real axis in the complex α-plane and a neighborhood $|\varepsilon| < \varepsilon_0$ of the origin in the complex ε-plane, with $\varepsilon_0 = O(\eta^{2/q}e^{-\delta_0})$.
(ii) The function $U(\alpha, \varepsilon) \equiv u(\alpha, (2\pi\eta)^{2/q}\varepsilon)$ is well defined for $\eta \to 0$ and converges to a function $\bar{u}(\alpha, \varepsilon)$, divisible by ε^q and analytic in ε^q in a neighborhood of the origin. Furthermore $\overline{u}$ is $2\pi/q$ periodic and solves the differential equation

e9.3.12
$$\frac{\mathrm{d}^2\bar{u}(\alpha)}{\mathrm{d}\alpha^2} = C_{p/q}\varepsilon^q \sin\left(q(\alpha + \bar{u}(\alpha))\right), \tag{9.3.12}$$

with boundary conditions $\bar{u}(0) = \bar{u}(2\pi) = 0$, for some nonvanishing explicitly computable constant $C_{p/q}$.

Remarks: (1) The bound $\varepsilon_0 = O(\eta^{2/q})$ of proposition (9.2.1) *is consistent* with the law $|Q_2(\omega)| < C$, because the logarithm of the complex extension of the Bryuno function diverges as $\eta \to 0$ in the same way as the $\log\varepsilon_0$, [Yo95]; but of course it does not implies such a law. The proof of the latter bound can be found in [BG01].
(2) The analyticity strip in α width can be prefixed arbitrarily.

The proof will be completed in Section §(9.4). Here we derive explicit expressions for $u_\nu^{(k)}$ as sums of suitable quantities and exhibit cancellations that occur if the quantities used to express the value of $u_\nu^{(k)}$ are suitably collected showing the validity of item (i) of proposition (9.3.1).

We can envisage a tree expansion for the conjugating function as in Section §8.1, with some simplifications, due to the fact that mode labels take only the values $\nu_v = \pm 1$ (because in (9.3.1) the non linear terms, *i.e.* $\sin x$ have only harmonics ± 1) has to be associated with the nodes v of the trees. Then the momentum flowing through a branch λ_v will be defined as

e9.3.13
$$\nu_{\lambda_v} = \sum_{w \le u} \nu_w, \qquad \nu_w = \pm 1; \tag{9.3.13}$$

and the condition $u_0^{(k)} = 0$ implies that no branch λ can have momentum $\nu_\lambda = 0$.

So we can write

e9.3.14
$$u_\nu^{(k)} = \frac{1}{2^k} \sum_{\vartheta \in \Theta_{k,\nu}} \mathrm{Val}(\vartheta), \qquad \mathrm{Val}(\vartheta) = -i \prod_{v \in V(\vartheta)} \frac{1}{m_v!} \frac{\nu_v^{m_v+1}}{\gamma(\nu_{\lambda_v})}, \tag{9.3.14}$$

where $\Theta_{k,\nu}$ is (unlike in the previous Sections) the set of *non-numbered trees* with k nodes and momentum $\nu_{\lambda_{v_0}} = \nu$, if v_0 is the last node of the tree, and such that $\nu_\lambda \neq 0$ for all branches $\lambda \in \Lambda(\vartheta)$ (see Section §8.1).

Remark: Unlike what was done in the previous Sections, we consider here non-numbered trees. This means that we do not introduce the number labels for the branches: in order to count the number of non-numbered trees of a given order k, one can imagine to assign to the m_v subtrees with root on a node v a number from 1 to m_v from up to down, and to consider as distinct only trees which cannot be overlapped by permuting such subtrees (for instance if two subtrees with numbers $1 \le m < m' \le m_v$ are equal to each other, then by exchanging the two subtrees we obtain no further tree). This corresponds to the the factorials $m_v!$ associated, in evaluating tree values, with the nodes of the trees. The number of elements in $\Theta_{k,\nu}$ is now bounded by 2^{2k} (see problem [8.2.2]). The reason to proceed in this way shall become clear in the following where the new combinatorial factors will turn out to be quite convenient to prove that the limit function $\bar u(\alpha)$
N9.3.1 solves the differential equation (9.3.12).[1]

For $\eta \neq 0$ the power series (9.3.7) is well defined and no small divisors appear: the denominators in $\mathrm{Val}(\vartheta)$ are all bounded from below by $|\eta|^2$. Nevertheless the convergence radius ρ is not uniform in η, and it shrinks to 0 when $\eta \to 0$. The small divisor $\gamma(\nu)$ satisfies the bound

e9.3.15
$$|\gamma(\nu)| \ge \begin{cases} c|\nu\eta|^2, & \text{for } \nu \in q\mathbb{Z} \setminus \{0\}, \\ cq^{-2}, & \text{otherwise}, \end{cases} \tag{9.3.15}$$

[1] The enumeration of trees considered in Chapter VIII and in Sections 9.1, 9.2 will still be useful later in checking cancellations. See also [Ga94b].

for some positive constant c; one can take $c = 2\pi^2$.

Given a tree ϑ, we can associate with each branch λ of ϑ a scale label n_λ, setting $n_\lambda = 0$ if its momentum ν_λ is a multiple of q, and $n_\lambda = 1$ otherwise. Hence in the present analysis *only two scales* arise, unlike the infinitely many arising in the previous sections and in Chapter VIII. Given a tree ϑ, a *cluster* T of ϑ is a maximal connected set of branches on scale $n = 1$; we shall say that such branches are *internal* to T, and write $\lambda \in \Lambda(T)$, so denoting by $\Lambda(T)$ the set of branches in T. A node v will be considered internal to T if the exiting branch or at least one of its entering branches is in T, and we shall write $v \in V(T)$ so denoting by $V(T)$ the set of nodes in T. The branches outside the clusters are all on scale $n = 0$, and each cluster has an arbitrary number $m_T \geq 0$ of entering branches but only one exiting branch.

A cluster T will be called a *null graph* if

e9.3.16
$$\sum_{v\in V(T)} \nu_v = 0, \tag{9.3.16}$$

At least one entering branch *must* be present otherwise the exiting branch or the root branch would have momentum $\nu = 0$. For the same reason at least one entering branch must exist. In such a case, the exiting branch of the cluster T will be called a *null branch*; we also denote by k_T the number of nodes internal to T.

Remark: Note that the definition of null graph given here is *not* equivalent to that of self-energy graph of the previous Sections, because a null graph can have an arbitrary number of entering branches, and no constraint is imposed on the number of internal nodes. Here we shall call *null branch* a branch entering a null graph.

A null graph T in a tree graph ϑ_0 determines a subtree ϑ attached to the rest of the tree graph by its entering lines and its exiting line: we can define its *null graph factor* $\mathcal{V}_T(\vartheta)$ as

e9.3.17
$$\mathcal{V}_T(\vartheta) = \Big(\prod_{v\in V(T)} \frac{1}{m_v!}\Big)\Big(\prod_{v\in V(T)} \nu_v^{m_v+1}\Big)\Big(\prod_{\lambda\in\Lambda(T)} \frac{1}{\gamma(\nu_{\lambda_v})}\Big) \tag{9.3.17}$$

the null graph factor will depend on ϑ_0 only through the momenta of the incoming branches of T. By (9.3.15), we have the bound

e9.3.18
$$|\mathcal{V}_T(\vartheta)| \leq c^{-k_T} q^{2k_T}, \tag{9.3.18}$$

because all branches inside T are on scale $n = 1$.

Let $N_n(\vartheta_0)$, $n = 0, 1$, denote the number of branches in a tree graph ϑ_0 which have scale n: using again (9.3.15) it follows that for a fixed tree ϑ_0 one has

e9.3.19
$$|\mathrm{Val}(\vartheta_0)| \leq c^{-k} q^{2N_1(\vartheta_0)} |\eta|^{-2N_0(\vartheta_0)}. \tag{9.3.19}$$

Let $N_0^*(\vartheta_0)$ be the number of branches on scale $n=0$ which are not null branches in a given tree ϑ. We then have the bound

e9.3.20
$$N_0^*(\vartheta_0) \le \Big[\frac{k}{q}\Big], \tag{9.3.20}$$

where k is the order of the tree and $[x]$ is the largest integer smaller or equal to x. This is a statement analogous to the one that in Chapter VIII was called "Siegel–Bryuno" bound and its check is very similar too, see problem [9.3.4].

We now show how to construct a suitable partial resummation of the Lindstedt series. This means that we shall define families of trees to be grouped and bounded together to obtain "extra" η factors in the bounds of the sums of their values.

Given m trees $\vartheta_1,\ldots,\vartheta_m$ with root lines $\lambda_1\ldots\lambda_m$ on scale 0 and a tree $\overline{\vartheta}$ which has among its lowest lines (in its partial order) a line $\overline{\lambda}$ on scale 0 as well, we can imagine to construct a tree ϑ_0 which contains a null graph T with k_T nodes, $m_T\equiv m$ entering lines $\lambda_1\ldots\lambda_{m_T}$ and with $\overline{\lambda}$ as exiting line. The k_T nodes $v_1,\ldots,v_{k_T}$ carry modes $\nu_{v_1},\ldots,\nu_{v_{k_T}}$ such that $\sum\nu_{v_i}=0$.

We shall call $\mathcal{F}$ the family of trees formed by all trees that can be built in this way (it depends on $\overline{\vartheta},\vartheta_1,\ldots,\vartheta_m$). Each element of $\mathcal{F}$ is entirely defined by the tree ϑ determined by the null graph T (see comments before (9.3.17)) and its value is the product of a ϑ–independent factor times the null graph factor $\mathcal{V}_T(\vartheta)$ defined in (9.3.17).

It is now convenient, for enumeration purposes, to strip $\mathcal{V}_T(\vartheta)$ of the first factor $\prod_v\frac{1}{m_v!}$ and to attach a label $1,2,\ldots,k_T+m$ to the k_T+m lines forming ϑ. We obtain a family of "numbered null graphs" which contain many more graphs: we want to regard them as distinct unless they can be overlapped (number labels included) by pivoting, around their ending nodes, the lines ending on the nodes of ϑ.

If we define the value of the new numbered null graphs as in (9.3.17) with $\prod_v\frac{1}{m_v!}$ replaced by $\frac{1}{(k_T+m)!}$ we realize that the sum of the values of the distinct numbered null graphs equals the sum of the values of the null graphs previously considered in the family $\mathcal{F}$.

The momentum flowing through a branch λ_v inside a null graph will depend on the modes of the nodes $w\in T$ such that $w\preceq v$ and on the momenta of the entering branches of the null graph only if the latter end into nodes which precede v; we call L_v the set of such branches.

For any λ_v internal to a null graph T we have

e9.3.21
$$\nu_{\lambda_v}=\sum_{\substack{w\in V(T)\\ w\le v}}\nu_w+\sum_{\lambda'\in L_v}\nu_{\lambda'}, \tag{9.3.21}$$

and, in the corresponding propagator, we can write the argument of the cosine as

e9.3.22
$$\frac{2\pi p}{q}\sum_{\substack{w\in V(T)\\ w\le v}}\nu_w+2\pi i\Big[\sum_{\substack{w\in V(T)\\ w\le v}}\eta\nu_w+\sum_{\lambda'\in L_v}\eta\nu_{\lambda'}\Big]. \tag{9.3.22}$$

because the lines entering the graph have scale 0 so that $\sum_{\lambda'\in L_v}\nu_{\lambda'}$ is a multiple of q.

We can consider the null graph factor $\mathcal{V}_T(\vartheta)$ as a function of the quantities $\mu_1\equiv\eta\nu_{\lambda_1}$, ..., $\mu_{m_T}\equiv\eta\nu_{\lambda_{m_T}}$ of the incoming branches λ_1, ..., λ_{m_T}, *i.e.* $\mathcal{V}_T(\vartheta)\equiv\mathcal{V}_T(\vartheta;\eta\nu_{\lambda_1},\ldots,\eta\nu_{\lambda_{m_T}})$. For any $v\in V(T)$, we have $L_v\subseteq\{\lambda_1,\ldots,\lambda_{m_T}\}$. We can write

e9.3.23

$$\begin{aligned}&\mathcal{V}_T(\vartheta;\eta\nu_{\lambda_1},\ldots,\eta\nu_{\lambda_{m_T}})=\mathcal{V}_T(\vartheta;0,\ldots,0)\\&+\sum_{m=1}^{m_T}\eta\nu_{\lambda_m}\frac{\partial}{\partial\mu_m}\mathcal{V}_T(\vartheta;0,\ldots,0)\\&+\sum_{m,m'=1}^{m_T}\eta^2\nu_{\lambda_m}\nu_{\lambda_{m'}}\int_0^1\mathrm{d}t\,(1-t)\frac{\partial^2}{\partial\mu_m\partial\mu_{m'}}\mathcal{V}_T(\vartheta;t\eta\nu_{\lambda_1},\ldots,t\eta\nu_{\lambda_{m_T}}),\end{aligned}\tag{9.3.23}$$

where $[\partial\mathcal{V}_T/\partial\mu_m](\vartheta;0,\ldots,0)$ denotes the first derivative of $\mathcal{V}_T(\vartheta;\mu_1,\ldots,\mu_{m_T})$ with respect to the argument μ_m, computed in $\mu_1=\ldots=\mu_{m_T}=0$, while the term in the second line is the integral interpolation formula for the second order remainder.

We call $\mathcal{F}_T(\vartheta)$ the collection of all trees which are in the family $\mathcal{F}$ defined above and which contain in the null graph T a tree ϑ' of topological shape which can differ from that of ϑ because of the shift of the entering lines of
N9.3.2 T, and with given mode labels $\{\nu_v\}_{v\in\vartheta}$ assigned to the nodes $v\in\vartheta$;[2] we further enlarge the family $\mathcal{F}_T(\vartheta)$ by adding to it the trees in which the signs of the mode labels are all *simultaneously* reversed. Therefore the elements of the family $\mathcal{F}_T(\vartheta)$ differ from each other because the entering lines can be attached in various ways to the nodes of ϑ or because the signs of the node modes are reversed simultaneously, *i.e.* the trees of $\mathcal{F}$ are obtained by "shifting the external lines entrance nodes" in the null graph T or by the "reversing all signs of the node modes".

For the same reasons discussed in Section 8.3 we see that the sum of all the numbered null graphs in $\mathcal{F}_T(\vartheta)$ is an *even* function of η (because if the modes ν_v can be attached to the nodes of ϑ also the modes $-\nu_v$ can) which vanishes as $\eta\to0$ to order η^{m_T}: therefore it vanishes to order η^{m_T} if m_T
N9.3.3 is even and η^{m_T+1} if m_T is odd.[3] Hence, since $m_T\ge1$, it vanishes at least to second order in η so that the zero-th and first orders in η in the r.h.s. of (9.3.23) vanish. This implies the following lemma which is the crucial one, where cancellations between trees in the same family are exploited. The lemma will be immediately used to prove the main estimate on the radius of convergence of the Lindstedt series.

L9.3.1 **(9.3.1) Lemma:** (Cancellations : single null graph case)
Given a tree ϑ_0, with a null graph T with k_T nodes containing a tree of the

[2] Two numbered trees have the same topological shape if, disregarding the number labels, they can be overlapped by pivoting the branches around the nodes into which they enter.

[3] The higher order of the zero in η is due to the fact that one can independently shift the m_T entering lines, while in Section 8.3 there was only one entering line.

topological shape of some ϑ*, let* $\nu_1, \ldots, \nu_{m_T}$ *be the momenta flowing through the entering branches* $\lambda_1, \ldots, \lambda_{m_T}$ *of* T*. We consider the family* $\mathcal{F}_T(\vartheta)$*: we have the bound*

e9.3.24
$$\frac{1}{\mathcal{F}_T(\vartheta)}\Big|\sum_{\vartheta'\in\mathcal{F}_T(\vartheta)} \mathcal{V}_T(\vartheta')\Big| \le c^{-k_T} q^{2k_T} D_0 k_T^2 \sum_{m,m'=1}^{m_T} \big|\nu_m \nu_{m'} \eta^2\big|\,, \tag{9.3.24}$$

for some constant D_0*.*

Remark: The improved bound (9.3.24) is due to the above discussed remarkable cancellations which extend those considered in Section 8.3. Without exploiting such cancellations, the value of a null graph could only be bounded through (9.3.18). The cancellations produce the extra factors appearing in the sum: the factor k_T^2 is due to the fact that in order to exhibit the cancellation we use an interpolation formula, see (9.3.23), which involves a second order derivative with respect to the k_T momenta of the entering lines of the null graph. The latter is proportional to the product of k_T propagators.

Proof: The first two terms in (9.3.23) give a vanishing contribution to the sum over $\mathcal{F}_T(\vartheta)$, as discussed above; the integral appearing in the third term in (9.3.23) gives, for each $\vartheta' \in \mathcal{F}_T(\vartheta)$, a contribution bounded by $d_0 q^2$ times the original bound on $|\mathcal{V}_T(\vartheta)|$, for a suitable d_0. In fact the propagators of all branches inside V are bounded by $c^{-1}q^2$, and their first and second derivatives, respectively, by $d_1 q^3$ and $d_2 q^4$ for some constants d_1 and d_2 (such branches remain on scale $n = 1$: the incoming branches contribute a quantity that modifies only the imaginary part of the momenta of branches inside V). Then the lemma follows, with $D_0 = \max\{c d_1^2, d_2\} q^2$. ■

However a typical tree may contain more than one null graph: hence we shall need the following easy corollary of the above lemma.

C9.3.1 **(9.3.1) Corollary:** (Cancellations: several null graphs case)
Given a tree ϑ_0 *with null graphs* $T_1, \ldots, T_s$*, containing numbered trees with shapes* $\vartheta_1, \ldots, \vartheta_s$*; consider the families* $\mathcal{F}_1 = \mathcal{F}_{T_1}(\vartheta_0)$*,* $\mathcal{F}_2 = \cup_{\vartheta'\in\mathcal{F}_1}\mathcal{F}_{T_2}(\vartheta')$ *and so on recursively till* $\mathcal{F}_s$*; the number of trees in the family* $\mathcal{F}$ *is given by* $k_{\mathcal{F}_s} = \prod_{i=1}^s |\mathcal{F}_{T_i}(\vartheta_i)|$*, and*

e9.3.25
$$\frac{1}{k_{\mathcal{F}_s}}\Big|\sum_{\vartheta'\in\mathcal{F}_s} \mathrm{Val}(\vartheta')\Big| \le c^{-k} q^{2N_1(\vartheta_0)} |\eta|^{-2N_0(\vartheta_0)} D_0^s k^2 |\eta|^{2s}, \tag{9.3.25}$$

with a suitable D_0*.*

Proof: One remarks that the cancellation mechanisms operating for each null graph do not interfere with each other (*i.e.* there is no cancellations overlap because here we have only two scales), as noted (in a more general context) in Section §8.1. ■

The above corollary implies the main result on the convergence radius of the Lindstedt series (9.3.7), which can be formulated as follows.

Let $N_0^R(\vartheta_0)$ be the number of null branches (necessarily on scale $n=0$), so that $N_0(\vartheta_0) = N_0^*(\vartheta_0) + N_0^R(\vartheta_0)$. For the null graphs an overall gain $D_0^s k^2 |\eta|^{2s}$ is obtained, by (9.3.25). Then we have

e9.3.26
$$\frac{1}{k_{\mathcal{F}_s}} \Big| \sum_{\vartheta' \in \mathcal{F}_s} \mathrm{Val}(\vartheta') \Big| \le c^{-k} q^{2N_1(\vartheta_0)} |\eta|^{-2N_0^*(\vartheta_0)} |\eta|^{-2N_0^R(\vartheta_0)} D_0^{N_0^R(\vartheta_0)} k^4 |\eta|^{2N_0^R(\vartheta_0)} \\ \le \left(c^{-1} e^2 D_0 q^2\right)^k |\eta|^{-2k/q}, \tag{9.3.26}$$

as $k^2 \le e^{2k}$ and the number of null graphs is equal to the number of null branches, *i.e.* $s = N_0^R(\vartheta_0)$. Therefore, writing the sum over all trees as

e9.3.27
$$\sum_{\vartheta \in \Theta_{\nu,k}} \mathrm{Val}(\vartheta) = \sum_{\vartheta \in \Theta_{\nu,k}} \frac{1}{k_{\mathcal{F}_s}} \sum_{\vartheta' \in \mathcal{F}_s} \mathrm{Val}(\vartheta'), \tag{9.3.27}$$

we can conclude that one has, for α real,

e9.3.28
$$\left| u_\nu^{(k)} \right| \le D_1^k |\eta|^{-2k/q}, \qquad \Big| \sum_{\nu \in \mathbb{Z}} e^{i\nu\alpha} u_\nu^{(k)} \Big| \le D_1^k |\eta|^{-2k/q}, \tag{9.3.28}$$

for some constants D_1. The above analysis gives $D_1 = 4cD_0q^2$; this implies that $\rho(\eta) \ge r|\eta|^{2/q}$, for some explicitly computable constant r.

Since the momentum ν of a k–th order tree is $|\nu| \le k$ we see that analyticity in the strip $|\mathrm{Im}\,\alpha| < \delta_0$ has also been obtained in the domain $|\varepsilon| < \rho(\eta) e^{-\delta_0}$ because $|e^{i\alpha\nu}| \le e^{\delta_0 k}$. This proves item (i) of proposition (9.3.1). Item (ii) will be studied in Section §(9.4).

Problems for §9.3

Q9.3.1 **[9.3.1]:** (*Kicked rotator*)
Show that (9.3.1) can be seen as the map at time 1 of the continuous-time Hamiltonian system (known in Physics literature as the *kicked rotator*) described by the singular Hamiltonian

$$2\pi A + \frac{I^2}{2} + 2\pi\varepsilon \sum_{\nu \in \mathbb{Z}} \delta(\alpha - 2\pi n)\,(\cos\varphi - 1)\,,$$

where δ is the delta function. (*Hint:* Simply write down the Hamilton equations and integrate them between $t=0$ and $t=1$, then define $x=\varphi$ and $y = I - \varepsilon \sin\varphi/2$.)

Q9.3.2 **[9.3.2]:** (*Kicked rotator and KAM*)
Prove the existence of the analytical invariant curves (9.3.6) for the standard map, for ε small enough. (*Hint:* Simply adapt the proof of KAM theorem in Section §8.1 to the standard map by using the tree expansion (9.3.14)).

Q9.3.3 **[9.3.3]:** Prove the bound (9.3.15). (*Hint:* Use that $|\cos z - 1| \ge |z|^2/4$, and $|\cos z - 1| \ge |\mathrm{Im}\, z|^2/2$ for $|\mathrm{Re}\, z| \le \pi/4$, while $|\cos z - 1| \ge 1/2$ for $|\mathrm{Re}\, z \bmod 2\pi| \ge \pi/4$.)

Q9.3.4 **[9.3.4]:** Prove the bound (9.3.20). (*Hint:* The cases $q=1$ and $q \ge 2$, $k \le q$ are trivial. So consider the case $k > q$. If the tree ϑ has the root branch either on scale $n=1$ or on scale $n=0$ and resonant, then the bound $N_0^*(\vartheta) \le k/q$ follows inductively. If the root branch is on scale $n=0$ and non-resonant, then the branches entering v_0 cannot be all on scale $n=0$, otherwise $\nu_{v_0} = 0$ (as $q > 1$), which is not allowed. Then at least one branch will have scale $n=1$: let T be the cluster containing it. The cluster T will have

m_T entering branches, with $m_T \geq 0$, and it is not a null graph; then $\sum_{v \in V(T)} \nu_v \neq 0$, so there must be at least q nodes, hence $q-1$ branches, inside T. The subtrees entering into T will have, respectively, $k_1, \ldots, k_{m_T}$ nodes, with $q + \sum_{j=1}^{m_T} k_j \leq k$ so that, again inductively, one obtains $N_0^*(\vartheta) = 1 + \sum_{j=1}^{m_T} N_0^*(\vartheta_j) \leq 1 + (k-q)/q \leq k/q$. Then N_0^* will be bounded by k/q. As N_0^* has to be an integer number, the assertion follows.)

Bibliographical note to §9.3

The possible relations between radius of convergence of the conjugating function and some function depending only on the arithmetic properties of the rotation number was first considered by Yoccoz for Siegel's problem, [Yo95], who also introduced the Bryuno function $B(\omega)$ as the solution of a suitable functional equation. In particular he considered the problem of the optimal dependence of the radius of convergence on the Bryuno function. The analogous problem for the *semistandard map*, which is a simplified model sharing with the standard maps many interesting features, was considered by Davie, [Da94], who found that, by denoting with $\rho_0(\omega)$ the radius of convergence of the function which takes the role of the conjugating function for the standard map, a bound $|\log \rho_0(\omega) + \beta B(\omega)| < C$ holds with $\beta = 2$. A natural question is then whether a bound of the same kind can be obtained also for the standard map, *i.e.* if, by defining $Q_\beta(\omega) \equiv \log \rho(\omega) + \beta B(\omega)$, there exists a universal constant C such that $|Q_\beta(\omega)| < C$ for some value of β and for all irrational numbers ω such that $B(\omega) < \infty$ (the so called *Bryuno numbers*). Davie's result implies the upper bound $Q_2(\omega) < C$

The results stated in proposition (9.2.1) were first conjectured in [BM94], supported by numerical results, and a proof was given for the resonances $p/q = 0/1$ and $p/q = 1/2$. Validity of Bryuno's interpolation formula $|Q_2(\omega)| < C$ was recently proved in [BG01], where a bound $Q_2(\omega) > C$ was found via a refinement of the techniques discussed in the present section.

§9.4 Scaling laws for the standard map

In Section §9.3 we have proved that at least a rescaling $\varepsilon \to (2\pi\eta)^{2/q}\varepsilon$ is needed in order to obtain a well defined limit of the conjugating function as $\eta \to 0$. Without exploiting any cancellations a rescaling $\varepsilon \to (2\pi\eta)^2\varepsilon$ might seem to be necessary, but of course, by taking into account the cancellations *a posteriori*, one sees that such a rescaling would have produced a vanishing function. The analysis performed so far, however, does not exclude that further cancellations are still possible, so that it could happen that the limit of the rescaled conjugating function is still zero: in other words it could happen that we are still rescaling "too much". In this section we want to show that this is not the case, and, furthermore, we want to provide an explicit expression for the rescaled function.

Let us consider the coefficient $u_\nu(\varepsilon)$ in (9.3.7). By definition of momentum, only trees of order $k \ge |\nu|$ can contribute to $u_\nu(\varepsilon)$, so that $u_\nu^{(k)} = 0$ for $k < |\nu|$. Therefore we can write

e9.4.1
$$u_\nu(\varepsilon) = \sum_{k\ge|\nu|} \varepsilon^k u_\nu^{(k)} = \varepsilon^{|\nu|} u_\nu^{(|\nu|)} + \sum_{k=|\nu|+1}^{\infty} \varepsilon^k u_\nu^{(k)}, \tag{9.4.1}$$

and use the first bound in (9.3.28) for $u_\nu^{(k)}$, $k > |\nu|$, in order to bound the last sum in (9.4.1) by

e9.4.2
$$\Big|\sum_{k=|\nu|+1}^{\infty} \varepsilon^k u_\nu^{(k)}\Big| \le 2\left(D_1|\eta|^{-2/q}\varepsilon\right)^{|\nu|+1}, \tag{9.4.2}$$

provided that $|\varepsilon| < D_1^{-1}|\eta|^{2/q}/2$. The coefficient $u_\nu^{(|\nu|)}$ in (9.4.1) can be expressed in terms of trees having all the modes $\nu_v = \sigma$, where $\sigma = \operatorname{sign}\nu$. From the definition of tree value it is easy to see that such trees have the same product $\prod_{v\in V(\vartheta)} \nu_v^{m_v+1}$ appearing in $\mathrm{Val}(\vartheta)$, which is given by $\sigma^{2|\nu|+1} = \sigma$. Furthermore among such trees there will also be trees having $[|\nu|/q]$ propagators with momentum respectively q, $2q$, …, $[|\nu|/q]q$: for instance the linear tree (*i.e.* the tree whose nodes have all only one entering branch). Therefore there will be trees whose value will be bounded *from below* by $O\left(|\eta|^{-2[|\nu|/q]}\right)$. When the limit $\eta \to 0$ is taken, remarking that the quantities $\gamma(\nu)$ for $\eta = 0$ are not positive, cf. (9.3.9), each of such trees ϑ has a value which is:

e9.4.3
$$\begin{aligned} &\mathrm{Val}(\vartheta) = -i\sigma B(\vartheta)\eta^{-2[|\nu|/q]},\\ &B(\vartheta) = (-1)^k\Big(\prod_{v\in V(\vartheta)} \frac{1}{m_v!}\Big)\Big(\prod_{\substack{\lambda\in\Lambda(\vartheta)\\ n_\lambda=0}} \frac{1}{(2\pi\nu_\lambda)^2}\Big)\Big(\prod_{\substack{\lambda\in\Lambda(\vartheta)\\ n_\lambda=1}} \frac{1}{|\gamma(\nu_\lambda)|}\Big), \end{aligned} \tag{9.4.3}$$

because the propagators $\gamma(\nu)^{-1}$ diverge as $(2\pi i\nu\eta)^{-2}$ for $\eta \to 0$.

This means that no cancellation is possible among them, so that

e9.4.4
$$u_\nu^{(|\nu|)} = A_\nu|\eta|^{-2[|\nu|/q]}, \qquad |A_\nu| > 0. \tag{9.4.4}$$

If we want that, in the limit $\eta \to 0$, the coefficient $u_\nu(\varepsilon)$ non only does not diverge but also does not vanish, we have to impose that

e9.4.5
$$0 < \lim_{\eta\to 0} \varepsilon^{|\nu|} u_\nu^{(|\nu|)} < \infty, \tag{9.4.5}$$

so that (9.4.4) implies that ε has to be taken of order $O(|\eta|^{2/q})$. In such a way all the limits (9.4.5) exist, for any ν, and they are vanishing but for $|\nu|$ multiple of q.

Moreover, when we compute $u_\nu(\varepsilon)$, with ν multiple of q, only the coefficients $u_\nu^{(k)}$ with k multiple of $|\nu|$ will contribute to the limit $\eta \to 0$, because all other contributions arise from trees containing null graphs, and the analysis above shows that the propagators corresponding to the resonant branches do not introduce new denominators small in η, while each new node contributes a factor $|\eta|^{2/q}$, by (9.4.4) and (9.4.5).

The function $U(\alpha,\varepsilon) = u(\alpha,(2\pi\eta)^{2/q}\varepsilon)$ admits a Taylor series in ε convergent for $|\varepsilon| < \varepsilon_0 = O(1)$ uniformly in η. The k-th order Fourier coefficients of U are defined by

e9.4.6
$$U_\nu^{(k)} = u_\nu^{(k)}(2\pi\eta)^{2k/q}, \tag{9.4.6}$$

so that, if we introduce the function

e9.4.7
$$\bar{u}(\alpha,\varepsilon) = \lim_{\eta\to 0} U(\alpha,\varepsilon), \tag{9.4.7}$$

we have that

e9.4.8
$$\begin{aligned} \bar{u}(\alpha,\varepsilon) &= \sum_{k=1}^{\infty}\sum_{\nu\in\mathbb{Z}} \varepsilon^k e^{i\nu\alpha}\bar{u}_\nu^{(k)}, \\ \bar{u}_\nu^{(k)} &= \lim_{\eta\to 0} u_\nu^{(k)}(2\pi\eta)^{2k/q}, \end{aligned} \tag{9.4.8}$$

and the first series in (9.4.8) converges absolutely by (9.3.28): this means that the coefficients $\bar{u}_\nu^{(k)}$ are well defined, and, from the analysis above, we know that only the Fourier coefficients with modes multiples of q survive when the limit $\eta \to 0$ is taken; furthermore, the function $\bar{u}(\alpha,\varepsilon)$ is analytic in ε for ε small enough and η independent, and periodic in α with period $2\pi/q$. Summarizing, we have

e9.4.9
$$\bar{u}_\nu^{(k)} = \begin{cases} -\dfrac{(2\pi)^{k/q}}{2^k} \displaystyle\sum_{\vartheta\in\Theta_{\nu,k}}{}' \mathrm{Val}'(\vartheta), & \text{if } \nu \in q\mathbb{Z}\setminus\{0\} \text{ and } k \in q\mathbb{Z}\setminus\{0\}\ , \\ 0, & \text{otherwise}\ , \end{cases} \tag{9.4.9}$$

where $\sum'$ means that only trees without null graphs have to be summed over and $\mathrm{Val}'(\vartheta)$ differs from $\mathrm{Val}(\vartheta)$ inasmuch as, for ν multiple of q, the denominator $\gamma(\nu)$ has to be replaced with $(2\pi i\nu)^2$.

Let us now consider all trees of order q contributing to $\nu = \sigma q$, with $\sigma = \operatorname{sign}\nu$: the values of such trees can be read from (9.3.14), and one sees that the numerator is identically $-i\sigma$, so that

e9.4.10
$$\sum_{\vartheta\in\Theta_{\sigma q,q}} \mathrm{Val}'(\vartheta) = (2\pi)^2 \sum_{\vartheta\in\Theta_{\sigma q,q}} (-i\sigma) \prod_{v\in V(\vartheta)} \frac{1}{m_v!}\frac{1}{\gamma(\nu_{\lambda_v})} = \frac{1}{(i\nu)^2}(-i\sigma)S_{p/q}, \tag{9.4.10}$$

thus defining the expression $S_{p/q}$ (which does not depends on σ): the factor $(i\nu)^{-2}$ arises from $(2\pi)^2$ times the propagator $\gamma(\nu)$, which appears in all tree values of trees in $\Theta_{\sigma q,q}$ so it can be factored out. For instance, for $p/q = 0/1$, $p/q = 1/2$ and $p/q = 1/3$, by explicit computation, we find

e9.4.11
$$\begin{aligned} S_{0/1} &= 1, \\ S_{1/2} &= \frac{1}{2(\cos\pi - 1)} = -\frac{1}{4}, \\ S_{1/3} &= \frac{1}{2(\cos(4\pi/3)-1)}\frac{1}{2(\cos(2\pi/3)-1)} + \frac{1}{2}\frac{1}{[2(\cos(4\pi/3)-1)]^2} = \frac{1}{6}, \end{aligned} \tag{9.4.11}$$

as one checks from the definition (9.4.10).

We have the following result.

L9.4.1 **(9.4.1) Lemma:** (Limit function near the resonances)
The coefficients $\bar u_\nu^{(k)}$ satisfy the recursion relation

e9.4.12
$$\bar u_\nu^{(k)} = \frac{1}{2^{q-1}}\frac{1}{(i\nu)^2}\sum_{m=1}^{\infty}\frac{1}{m!}\frac{S_{p/q}}{2}(-i\sigma)(i\sigma q)^m \sum_{\substack{\sigma q+\nu_1+\ldots+\nu_m=\nu \\ k_1+\ldots+k_m=k-q}} \prod_{i=1}^{m}\bar u_{\nu_i}^{(k_i)}, \tag{9.4.12}$$

where the constant $S_{p/q}$ is defined in (9.4.10).

Proof: A generic tree of order $k = \kappa q$, $\kappa \geq 1$, and momentum $\nu = nq$, $n \geq 1$, can be obtained starting from a tree ϑ_0 of order q by attaching to its nodes $m \geq 0$ trees $\vartheta_1, \ldots, \vartheta_m$ of orders $k_1 = \kappa_1 q, \ldots, k_m = \kappa_m q$, with $\kappa_1 + \ldots + \kappa_m = \kappa - 1$ and total momenta $\nu_1 = n_1 q, \ldots, \nu_m = n_m q$ with $n_1 + \ldots + n_m = n - 1$. Each tree can be attached to any node of ϑ_0 so that the combinatorial factor associated to any node $v \in V(\vartheta_0)$will be $m_v!^{-1}$, with $m_v = s_v + r_v$, if s_v is the number of branches connecting v to other nodes of ϑ_0 and r_v is the number of trees attached to v. Then by construction $\sum_{v\in V(\vartheta_0)} r_v = m$. Note that two trees in which one of the first s_v subtrees (with root branch belonging to ϑ_0) is permuted with one of the remaining r_v cannot be identical, as they have a different number of nodes: only the latter will have a number of nodes which is multiple of q.

If we sum together all trees which can be obtained from each other by choosing in a different way the s_v subtrees with root branch belonging to ϑ_0 and the remaining r_v subtrees, we have $\prod_{v\in V(\vartheta_0)} \frac{m_v!}{s_v!\,r_v!}$ terms, so that, by taking into account that (see (9.4.10))

e9.4.13
$$\mathrm{Val}'(\vartheta_0) = (2\pi)^2(-i\sigma)\prod_{v\in V(\vartheta_0)}\frac{1}{s_v!}\frac{1}{\gamma(\nu_{\lambda_v})}. \tag{9.4.13}$$

Furthermore, by shifting the subtrees attached to the nodes of ϑ_0, the momentum flowing through any branch of ϑ_0 can vary by an amount proportional to a multiple of q, so that the corresponding propagator does not change, we can write

e9.4.14
$$\mathrm{Val}'(\vartheta) = \sum_{\vartheta_0\in\Theta_{\sigma q,q}} \mathrm{Val}(\vartheta_0) \sum_{m=0}^{\infty} \sum_{\substack{\{r_v\ge 0\}\\ \sum_{v\in V(\vartheta_0)} r_v = m}} \prod_{v\in V(\vartheta_0)} \frac{1}{r_v!}(i\sigma)^m \qquad {\sum_{\vartheta_1,\ldots,\vartheta_m}}' \prod_{i=1}^{m} \mathrm{Val}'(\vartheta_i), \tag{9.4.14}$$

where $'$ recalls the constraint on the trees $\vartheta_1, \ldots, \vartheta_p$ described above. Since

e9.4.15
$$\prod_{\substack{\{r_v\ge 0\}\\ \sum_{v\in V(\vartheta_0)} r_v = m}} \frac{1}{r_v!} = \frac{q^m}{m!}, \tag{9.4.15}$$

we deduce from (9.4.14) that

e9.4.16
$$\mathrm{Val}'(\vartheta) = \sum_{\vartheta_0\in\Theta_{\sigma q,q}} \mathrm{Val}'(\vartheta_0) \sum_{m=0}^{\infty} \frac{1}{m!}(i\sigma q)^m {\sum_{\vartheta_1,\ldots,\vartheta_m}}' \prod_{i=1}^{m} (i\sigma q)\mathrm{Val}'(\vartheta_i). \tag{9.4.16}$$

Then from (9.4.10) and (9.4.16) we read that

e9.4.17
$$\bar{u}_{\nu}^{(k)} = \frac{1}{2^q} \sum_{\vartheta_0\in\Theta_{q,q}} \mathrm{Val}'(\vartheta_0) \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{\substack{k_1+\ldots+k_m=k-q\\ \sigma+\nu_1+\ldots+\nu_m=\nu}} \prod_{i=1}^{m} (i\sigma q)\bar{u}_{\nu_i}^{(k_i)}. \tag{9.4.17}$$

and using (9.4.10) we obtain (9.4.12). ∎

The above results yield that the function $\bar{u}(\alpha) \equiv \bar{u}(\alpha, \varepsilon)$ in (9.4.7) satisfies the differential equation

e9.4.18
$$\frac{\mathrm{d}^2\bar{u}(\alpha)}{\mathrm{d}\alpha^2} = C_{p/q}\varepsilon^q \sin(q(\alpha + \bar{u}(\alpha))), \tag{9.4.18}$$

with boundary conditions $\bar{u}(0) = \bar{u}(2\pi) = 0$, and $C_{p/q}$ given by

e9.4.19
$$C_{p/q} = 2^{-(q-1)} S_{p/q} = 2^{-(q-1)} \sum_{\vartheta\in\Theta_{q,q}} (2\pi i\nu)^2 \prod_{v\in V(\vartheta)} \frac{1}{m_v!} \frac{1}{\gamma(\nu_{\lambda_v})}, \tag{9.4.19}$$

where the factor $(2\pi i\nu)^2$ simply cancels the propagator of the root branch of ϑ. This is seen by a straightforward check: write (9.4.18) in Fourier space, and write the recursion relations defining the coefficients, by taking into account that only orders and momenta multiples of q can occur (this can be seen inductively), obtaining

e9.4.20
$$\bar{u}_{\nu}^{(k)} = \frac{1}{(i\nu)^2} \sum_{m=0}^{\infty} \frac{1}{m!} \frac{C_{p/q}}{2} (-i\sigma)(i\sigma q)^m \sum_{\substack{\sigma q+\nu_1+\ldots+\nu_m=\nu\\ k_1+\ldots+k_m=k-q}} \prod_{i=1}^{m} \bar{u}_{\nu_i}^{(k_i)}, \tag{9.4.20}$$

with $\sigma = \pm 1$. This is the same expression as (9.4.12) provided the constant $\mathcal{C}_{p/q}$ is chosen as (9.4.19).

Now we come back to the proof of proposition (9.3.1), and, to conclude the proof, we collect the results obtained so far.

Since for $\eta \neq 0$ there are no small divisors and the convergence of the Lindstedt series can be proved by elementary means, the only non obvious part of item (i) is the behavior of the radius of convergence as $\eta \to 0$. This has been studied in detail at the end of Section 9.3, and the conclusion is that the radius of convergence is strictly positive; to exclude that the radius of convergence could be infinite (see also problem [9.4.2]) one can consider the limit function, whose existence is stated in item (ii), and use the properties of the elliptic functions to conclude that its radius of convergence is finite
*N*9.4.1 (see [BM94] for an explicit discussion).[1]

Remark: The restriction on the path over which we take the limit $\omega \to p/q$ in the complex plane (*i.e.* approaching the real axis perpendicularly) is taken only for the sake of simplicity: in fact, it is easy to modify the proofs in such a way that any path in the complex plane, *provided it is not tangent to the real axis*, can be taken. More precisely, let

*e*9.4.21
$$\omega = \frac{p}{q} + \zeta + i\eta, \tag{9.4.21}$$

with $p, q \in \mathbb{Z}$, $\gcd(p, q) = 1$, and $\zeta, \eta \in \mathbb{R}$, with

*e*9.4.22
$$|\eta| \geq a|\zeta|, \qquad a > 0, \tag{9.4.22}$$

in the limit $\eta \to 0$. Condition (9.4.22) defines a cone in the complex ω plane, with its vertex in p/q and its slope equal to a: any path inside this cone tends to p/q non-tangentially.

First we show that inequalities like (9.3.15) can be derived under the condition (9.4.22), with the only difference that now $c = 2\pi^2 a^{-2}$.

Since the analysis above was based on the inequalities (9.3.15) and on the definition of null graph, it can be repeated essentially unchanged in the case (9.4.21), and the same results hold. In fact, the proof of lemma (9.3.1) can be carried out in a similar manner, by expressing the null graph value as a function of the quantities $\xi\nu_{\lambda_1}, \ldots, \xi\nu_{\lambda_{m_T}}$, with $\xi = \zeta + i\eta$; then by taking into account that, for any ν such that $\nu = 0 \pmod q$, one has $|\gamma(\nu)| \geq 4\pi^2|\nu\eta|^2$, and $|\nu\zeta| \leq |\nu|\sqrt{\zeta^2+\eta^2} \leq |\nu\eta|(1+a^{-1})$, one sees that the cancellation mechanisms operate exactly in the same way as before, and the second order terms can be dealt with as before, with the only difference that now $D_0 = \max\{cd_1^2, d_2\}q^2a^{-2}(1+a^{-1})$.

Once the perturbation parameter ε has been scaled to $(2\pi\xi)^{2/q}\varepsilon$, the surviving terms are exactly the same as before, so that all of the above discussions

[1] Alternatively one can use an argument due to Davie [Da94]; see also [BG01] for the implementation of such an idea in a more general context.

apply *verbatim*; in fact

e9.4.23
$$\lim_{\eta\to 0} \gamma(\nu)[2\pi(\zeta+i\eta)]^{-2} = 1, \tag{9.4.23}$$

for ν multiple of q and ζ, η satisfying (9.4.22).

Our main results therefore still apply provided the path taken by ω while tending to p/q is not tangential to the real axis, so that (9.4.22) applies for some a.

Problems for §9.4

Q9.4.1 **[9.4.1]:** Show that for rotation numbers ω satisfying (9.4.21) and (9.4.22), the bound (9.3.15) still holds, for a different constant c. (*Hint:* The first inequality holds as, for ν multiple of q, we can write $\cos(2\pi(p/q+\zeta+i\eta)\nu) = \cos(2\pi(\zeta+i\eta)\nu)$, so that $2|\cos(2\pi(p/q+\zeta+i\eta)\nu)-1| \ge 4\pi^2|\nu\eta|^2$. If $\nu \ne 0 \pmod q$, we can deduce that, by denoting $x=(p/q+\zeta)\nu$, one has $|\gamma(\nu)| \ge 1/2$ for $2\pi|x| \ge \pi/4$. Then, if $2\pi|x| \le \pi/4$ we have $|\gamma(\nu)| \ge 2\pi^2(|x|^2+|\eta\nu|^2)$, so that, if $|x| \le (2q)^{-1}$, one has $|\zeta\nu| \ge (2q)^{-1}$ as $|p\nu/q| \ge 1$, hence $|\nu| \ge (2q|\zeta|)^{-1}$, *i.e.* $|\eta\nu| \ge (2q)^{-1}|\eta/\zeta|$, which implies $|\gamma(\nu)| \ge \pi^2 q^{-2} a^2/2$, while, if $|x| \ge (2q)^{-1}$, then $|\gamma(\nu)| \ge \pi^2 q^{-2}/2$.)

Q9.4.2 **[9.4.2]:** Provide an example of a power series in ε depending on a parameter η such that, for $\eta \to 0$, there is only one possible rescaling of ε as a function of η making the radius of convergence different from zero such that the rescaled function is different from zero and its radius of convergence id infinite. (*Hint:* Consider $\sum_{k=1}(\varepsilon/\eta)^k/k!$.)

CHAPTER X

Special ergodic theory problems in chaotic dynamics

§10.1 Perturbing Arnold's cat map

A general theorem on Anosov maps allows us to say that in a certain sense Anosov maps that are close enough in C^2 can be considered as derived one from the other by a "change of coordinates", which, however, is not really smooth. This is the theorem of structural stability of Anosov that can be formulated as follows.

P10.1.1 **(10.1.1) Proposition:** (Structural stability) *Let S, S' be two Anosov diffeomorphisms of a manifold Ω. If they are close enough together with their*
N10.1.1 *first two derivatives*[1] *then there exists a homeomorphism $H : \Omega \longleftrightarrow \Omega$ such that*

e10.1.1
$$S \circ H = H \circ S'. \tag{10.1.1}$$

The homeomorphism is Hölder continuous but, in general, not differentiable.

Remarks: (1) The lack of real smoothness, and even of differentiability, is however a considerable obstacle to arguments based on the naive interpretation of the theorem suggesting that S and S' "just" differ by the change of coordinates H.
(2) It is important to realize that if μ is an invariant measure for S then $\mu' = H\mu$ is an invariant measure for S'. Notwithstanding this nice property, in general, the image under H of the SRB distribution μ_{SRB} of S is not the SRB distribution μ'_{SRB} of S'. Indeed $H\mu_{SRB}$ is the weak limit under the

[1] One also says "close enough in the C^2 topology"

action of S' of $H\lambda$ where λ is the volume measure on Ω. Due to the above remarks $H\lambda$ is, in general, singular with respect to λ.

We suggest a proof for the simple two-dimensional case in the problems at the end of this section. Here we discuss how to construct concretely the map H in a class of special cases. An explicit construction of H as well as of the stable and unstable manifolds of an Anosov system and a detailed study of the SRB distribution is of course a very difficult task. Nevertheless it can be performed with remarkable depth in some special cases.
Here we consider maps that are perturbations of Arnold's cat map and the following proposition (or better its proof) is the key result [BFG03].

P10.1.2 **(10.1.2) Proposition:** (Arnold's cat map perturbations) *Let $\underline{f}(\underline{\varphi})$ be a real trigonometric polynomial, $\underline{f}(\underline{\varphi}) = \sum_{\underline{\nu}\in\mathbb{Z}^2,|\underline{\nu}|\le N} e^{i\underline{\nu}\cdot\underline{\varphi}}\underline{f}_{\underline{\nu}}$, defined on the two-dimensional torus $\mathbb{T}^2$ and let*

e10.1.2
$$S_\varepsilon\underline{\varphi} = S_0\underline{\varphi} - \varepsilon\underline{f}(\underline{\varphi}), \qquad \textit{with} \qquad S_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{10.1.2}$$

For $\beta\in(0,1)$ there exist $C(\beta)<\infty$ and $\varepsilon_0(\beta)>0$ such that for $|\varepsilon|<\varepsilon_0(\beta)$ the equation

e10.1.3
$$H\circ S_0 = S_\varepsilon\circ H \tag{10.1.3}$$

defines a unique homeomorphism $\underline{\varphi}\to H(\underline{\varphi})$ which is analytic in ε in the complex disk $|\varepsilon|<\varepsilon_0(\beta)$ and Hölder continuous with exponent at least as large as β and with Hölder continuity modulus bounded by $C(\beta)$.

Remarks: (1) Proposition (10.1.2) tells us that we can conjugate the map S_ε to the map S_0 via a function H that is analytic in ε. However, in general *it is not true that we can conjugate S_0 to S_ε via an analytic homeomorphism.* Indeed the equation

e10.1.4
$$\widetilde{H}\circ S_\varepsilon = S_0\circ\widetilde{H} \tag{10.1.4}$$

cannot be studied with the method developed in the proof below because it would require an expansion of $\widetilde{H}$ in power of ψ.
(2) Clearly, in the hypotheses of proposition (10.1.2), $\widetilde{H}$ exists and is the inverse of H. Conversely, (10.1.4) can be solved explicitly by using the special properties of S_0. Moreover, if S_ε is Anosov, the solution is a homeomorphism and is Hölder continuous in ε, as can be directly checked from $\widetilde{H}\circ H = \mathrm{Id}$. See problems [10.1.6], [10.1.9] for a more detailed discussion.

Proof: We shall write $\underline{\varphi} = H(\underline{\psi}) = \underline{\psi} + \underline{h}(\underline{\psi})$, $\underline{\psi}\in\mathbb{T}^2$; then the relation (10.1.3) becomes an equation for $\underline{h}$, namely

e10.1.5
$$S_0\underline{h}(\underline{\psi}) - \underline{h}(S_0\underline{\psi}) = \varepsilon\underline{f}(\underline{\psi}+\underline{h}(\underline{\psi})), \tag{10.1.5}$$

and the analogies with (8.1.12) or (9.3.4) suggest employing the same method to solve it. Hence we look for a solution which is analytic in ε:

$\underline{h}(\underline{\psi}) = \varepsilon \underline{h}^{(1)}(\underline{\psi}) + \varepsilon^2 \underline{h}^{(2)}(\underline{\psi}) + \cdots$ with $\underline{h}^{(k)}$ an ε–independent function. For instance the equation for the first order is

e10.1.6
$$S_0 \underline{h}^{(1)}(\underline{\psi}) - \underline{h}^{(1)}(S_0 \underline{\psi}) = \underline{f}(\underline{\psi}). \tag{10.1.6}$$

We call $\underline{v}_+, \underline{v}_-$ the two normalized eigenvectors of S_0 relative to the eigenvalues $(1 \pm \sqrt{5})/2$ and we call λ the inverse of the largest one ($\lambda = (\sqrt{5}-1)/2$), so that $\lambda_+ = \lambda^{-1}, \lambda_- = -\lambda$: although the fact that λ is the golden mean might be intellectually nice in the following the only property that we shall use is $\lambda < 1$.

The functions $\underline{f}, \underline{h}$ can be split into two components along the vectors $\underline{v}_\pm$:

e10.1.7
$$\begin{aligned} \underline{f}(\underline{\psi}) &= f_+(\underline{\psi})\underline{v}_+ + f_-(\underline{\psi})\underline{v}_-, \\ \underline{h}(\underline{\psi}) &= h_+(\underline{\psi})\underline{v}_+ + h_-(\underline{\psi})\underline{v}_-, \end{aligned} \tag{10.1.7}$$

and the equations (10.1.6) for $h^{(1)}_\pm$ are

e10.1.8
$$\begin{aligned} \lambda_+ h^{(1)}_+(\underline{\psi}) - h^{(1)}_+(S_0\underline{\psi}) &= f_+(\underline{\psi}), \\ \lambda_- h^{(1)}_-(\underline{\psi}) - h^{(1)}_-(S_0\underline{\psi}) &= f_-(\underline{\psi}). \end{aligned} \tag{10.1.8}$$

The equations (10.1.8) can be solved by simply setting

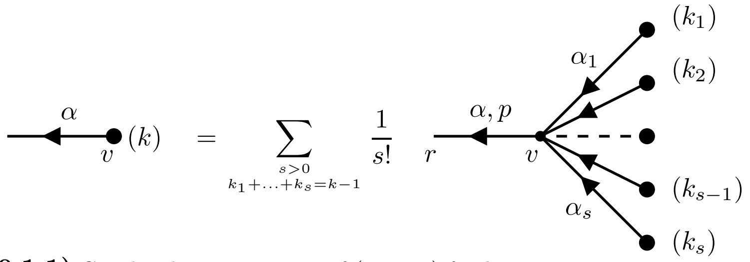
e10.1.9
$$h^{(1)}_\alpha(\underline{\psi}) = -\sum_{p\in\mathbb{Z}_\alpha} \alpha \lambda_\alpha^{-|p+1|\alpha} f_\alpha(S_0^p \underline{\psi}), \qquad \alpha = \pm, \tag{10.1.9}$$

where $\mathbb{Z}_+ = [0,\infty)\cap\mathbb{Z}$ and $\mathbb{Z}_- = (-\infty, 0)\cap\mathbb{Z}$ are subsets of the integers $\mathbb{Z}$ and advantage is taken of the inequality $\lambda = \lambda_+^{-1} = |\lambda_-| < 1$ to ensure convergence.

Therefore the equations for $h^{(k)}_\pm$ become

e10.1.10
$$\begin{aligned} h^{(k)}_\alpha(\underline{\psi}) = \sum_{s=0}^\infty \frac{1}{s!} \sum_{\substack{k_1+\cdots+k_s=k-1,\, k_i\ge 0 \\ \alpha_1,\ldots,\alpha_s=\pm}} \sum_{p\in\mathbb{Z}_\alpha} \alpha \lambda_\alpha^{-|p+1|\alpha} \cdot \\ \cdot \Big(\prod_{j=1}^s (\underline{v}_{\alpha_j} \cdot \partial_{\underline{\varphi}}) \Big) f_\alpha(S_0^p \underline{\psi}) \Big(\prod_{j=1}^s h^{(k_j)}_{\alpha_j}(S_0^p \underline{\psi}) \Big), \end{aligned} \tag{10.1.10}$$

and proceeding as in Section §8.1 we obtain a similar graphical representation



F10.1.1 **Fig.(10.1.1)** Graphical interpretation of (10.1.10) for $k \ge 1$.

where the l.h.s. represents $h^{(k)}_\alpha(\underline\psi)$. Representing again, in the same way, the graph elements that appear on the r.h.s. one obtains an expression for $h^{(k)}_\alpha(\underline\psi)$ in terms of *trees*, oriented toward the root, just like we already saw in Section §8.1, cf. Fig.(8.2.1), (8.2.2) and (8.3.1), in the KAM theory.

*N*10.1.2 A tree ϑ with k nodes will carry on the branches[2] ℓ a pair of labels α_ℓ, p_ℓ, with $p_\ell\in\mathbb Z$ and $\alpha_\ell\in\{-,+\}$, and on the nodes v a pair of labels α_v, p_v, with $\alpha_v=\alpha_{\ell_v}$ and $p_v\in\mathbb Z_{\alpha_v}$ such that

*e*10.1.11
$$p(v)\equiv p_{\ell_v}=\sum_{w\succeq v}p_v, \tag{10.1.11}$$

where the sum is over the nodes following v (*i.e.* over the nodes along the path connecting v to the root), ℓ_v denotes the branch $v'v$ exiting from the node v, and to each tree we shall assign a *value* given by

*e*10.1.12
$$\mathrm{Val}(\vartheta)=\prod_{v\in V(\vartheta)}\frac{\alpha_v}{s_v!}\lambda_{\alpha_v}^{-|p_v+1|\alpha_v}\Big(\prod_{j=1}^{s_v}\partial_{\alpha_{v_j}}\Big)f_{\alpha_v}(S_0^{p(v)}\underline\psi), \tag{10.1.12}$$

where $\partial_\alpha\stackrel{def}{=}\underline v_\alpha\cdot\underline\partial_{\varphi}$, $V(\vartheta)$ is the set of nodes in ϑ, the nodes $v_1,\ldots,v_{s_v}$ are the s_v nodes preceding v (if v is a top node then the derivatives are simply missing). If $\Theta_{k,\alpha}$ denotes the set of all trees with k nodes and with label α associated to the root line, then one has

*e*10.1.13
$$h_\alpha(\underline\psi)=\sum_{k=1}^\infty\varepsilon^k\sum_{\vartheta\in\Theta_{k,\alpha}}\mathrm{Val}(\vartheta), \tag{10.1.13}$$

and the "only" problem left is to estimate the radius of convergence of the above formal power series. For this purpose it is convenient to study the Fourier transform of the function $h_\alpha(\underline\psi)$. This is easily done graphically because it is enough to attach a label $\underline\nu_v\in\mathbb Z^2$ to each node and define the momentum that flows on the tree branch $v'v$ as in Section §8.1, *i.e.* $\underline\nu_{\ell_v}\stackrel{def}{=}\sum_{w\preceq v}\underline\nu_w$, see (8.2.4).

Then (10.1.13) becomes

*e*10.1.14
$$h_\alpha(\underline\psi)=\sum_{k=1}^\infty\varepsilon^k\sum_{\underline\nu\in\mathbb Z^2}e^{i\underline\nu\cdot\underline\psi}\,h^{(k)}_{\alpha,\underline\nu}, \tag{10.1.14}$$

with

*e*10.1.15
$$\begin{aligned}h^{(k)}_{\alpha,\underline\nu}=&\sum_{\vartheta\in\Theta_{k,\underline\nu,\alpha}}\sum_{p_v\in\mathbb Z_{\alpha_v}}\Big(\prod_{v\in V(\vartheta)}\frac{\alpha_v}{s_v!}\lambda_{\alpha_v}^{-|p_v+1|\alpha_v}f_{\alpha_v,S_0^{-p(v)}\underline\nu_v}\Big)\cdot\\&\cdot\prod_{\substack{v\in V(\vartheta)\\ v'\ne v_0}}\big(-S_0^{-p(v')}\underline\nu_{v'}\cdot\underline v_{\alpha_v}\big),\end{aligned} \tag{10.1.15}$$

[2] We do not use the letter λ as in Section §8.1 to denote the lines in order to avoid confusion with the parameter λ introduced after (10.1.6).

where $\Theta_{k,\underline{\nu},\alpha}$ denotes the set of all trees with k nodes and with labels $\underline{\nu}$ and α associated with the root line.

Calling $F=\max_{\underline{\nu}}|f_{\underline{\nu}}|$ we can estimate $\sum_{\underline{\nu}}|\underline{\nu}|^\beta|h^{(k)}_{\alpha,\underline{\nu}}|$. The only problem is given by the presence of the factor $|\underline{\nu}|^\beta$. In fact consider first the case $\beta=0$: since we are assuming that $\underline{f}$ is a trigonometric polynomial there are only $(2N+1)^2$ possible choices for each $\underline{\nu}_v$, given p_v, such that $|S_0^{-p_v}\underline{\nu}_v|\le N$. Hence fixed ϑ, $\{\alpha_v\}_{v\in V(\vartheta)}$ and $\{p_v\}_{v\in V(\vartheta)}$ the remaining sum of products in (10.1.15) is bounded by (if $\lambda\equiv\lambda_+^{-1}\equiv-\lambda_-$)

e10.1.16
$$(3N)^{2k}N^kF^k\prod_{v\in V(\vartheta)}\frac{\lambda^{|p_v|}}{s_v!}. \tag{10.1.16}$$

The sum over the p_v is a geometric series bounded by $(2/(1-\lambda))^k$.

The combinatorial problem is identical to the one discussed in Section §8.1 and in the other sections of Chapter VIII, so that the factor $\prod_v(1/s_v!)$ becomes, after summing over all the trees, simply bounded by 2^{3k}, (2^{2k} due to the number of trees for fixed labels, see Section §8.1, and 2^k due to the sum of labels α_v), noting that in the present problem all the intricacies due to the small divisors are just absent. In fact we have not used the $(s_v!)^{-1}$ and we have bounded them by 1: they would be necessary if we had supposed f to be only analytic rather than a trigonometric polynomial; see problems [10.1.3], [10.1.4] and [10.1.5], and bear in mind the problems of Section §8.4.

Therefore for $\beta=0$ we have proved that the conjugating function H exists and that inside the complex domain $|\varepsilon|<\varepsilon_0(0)\stackrel{def}{=}(3N)^{-3}F^{-1}2^{-4}(1-\lambda)$ it is uniformly continuous and uniformly bounded with a uniformly summable Fourier transform.

Taking $\beta>0$ requires estimating $|\underline{\nu}|^\beta$: we bound it by $\sum_v|\underline{\nu}_v|^\beta$. Then we can make use of the fact that $|S_0^{-p(v)}\underline{\nu}_v|\le N$ to infer that $|\underline{\nu}_v|\le\lambda^{-|p(v)|}BN$, where $B\ge1$ is a suitable constant; see problem [10.1.2]. The sum $\sum_v|\underline{\nu}_v|^\beta$ is over k terms which can be estimated separately so that we can write $\sum_v|\underline{\nu}_v|^\beta\le k|\underline{\nu}_{\overline{v}}|^\beta$ where $|\underline{\nu}_{\overline{v}}|=\max_v|\underline{\nu}_v|$. This can be taken into account by multiplying (10.1.16) by an extra factor $(BN)^\beta\lambda^{-\beta|p(\overline{v})|}\le BN\lambda^{-\beta\sum_v|p_v|}$. Therefore if $\beta<1$ the bound that we found for $\beta=0$ is modified into

e10.1.17
$$\varepsilon_0(\beta)=(3N)^{-3}F^{-1}(1-\lambda^{1-\beta})2^{-5}. \tag{10.1.17}$$

This shows that $H(\underline{\psi})$ analytic in ε in the disk with radius $\varepsilon_0(\beta)$. Furthermore, since in (10.1.17) we inserted (for simplicity) an extra factor 2^{-1} in excess of the result obtained by the procedure described, the Hölder modulus is also uniformly bounded by a suitable function $C(\beta)$ of β. Note that $\varepsilon_0(\beta)\to0$ for $\beta\to1$.

The map H is a homeomorphism. In fact it is one-to-one because if $\underline{\varphi}=\underline{\psi}_i+\underline{h}(\underline{\psi}_i)$ for $i=1,2$ and $\underline{\psi}_1\ne\underline{\psi}_2$ one would have $S_0^k\underline{\psi}_1+\underline{h}(S_0^k\underline{\psi}_1)=S_0^k\underline{\psi}_2+\underline{h}(S_0^k\underline{\psi}_2)$ for all k, which is impossible being incompatible with the hyperbolicity of S_0 (as a matrix). Furthermore given any $\underline{\varphi}$ there is a $\underline{\psi}$ such that $H(\underline{\psi})=\underline{\varphi}$: this is a general property of injective maps of the torus $\mathbb{T}^d$ of the form $\underline{\varphi}=\underline{\psi}+\underline{h}(\underline{\psi})$, with $\underline{h}$ Hölder continuous. Indeed if $d=1$ we see that as ψ runs around the circle the point $\varphi=\psi+h(\psi)$ follows and must pass through all points of the circle as well: hence if ψ is given there is a φ and a function $k(\varphi)$ such that $\psi=\varphi+k(\varphi)$ and the function k is continuous. If $d=2$ one repeats twice the argument: first we compute the partial inverse of $\varphi_1=\psi_1+h_1(\psi_1,\psi_2)$ by fixing ψ_2 and determining $k'(\varphi_1,\psi_2)$ such that $\psi_1=\varphi_1+k'(\varphi_1,\psi_2)$, with k' continuous; then one considers the map $\varphi_2=\psi_2+h_2(\varphi_1+k'(\varphi_1,\psi_2),\psi_2)$ and repeats the argument obtaining $\psi_2=\varphi_2+k_2(\varphi_1,\varphi_2)$. One finally sets $\psi_1=\varphi_1+k_1(\varphi_1,\varphi_2)$ with $k_1(\varphi_1,\varphi_2)=k'(\varphi_1,\varphi_2+k_2(\varphi_1,\varphi_2))$. ■

Problems for §10.1

Q10.1.1 **[10.1.1]:** (*A property of the golden mean*)
Prove that since λ is Diophantine the correlations of Arnold's cat map decay superexponentially. (*Hint:* Given two functions f and g, analytic on $\mathbb{T}^2$, write the correlations in Fourier space, $S=\sum_{\underline{\nu}} f_{\underline{\nu}} g_{-S^{-p}\underline{\nu}} \le FG\sum_{\underline{\nu}} e^{-\kappa|\underline{\nu}|}e^{-\kappa'|S^{-p}\underline{\nu}|}$, where F,G,κ,κ' are constants depending on f,g. Then use the Diophantine condition to deduce the inequality $|S^{-p}\underline{\nu}|>C\lambda^p/|\underline{\nu}|^{-\tau}$, for suitable constants C and τ. If λ is the golden mean one can take $\tau=1$.)

Q10.1.2 **[10.1.2]:** Show that if $|S_0^p\underline{\nu}|\le N$, then there is a constant $B=O(N)$ such that $|\underline{\nu}|\le\lambda^{-|p|}B$ and $B=N$ is a possible choice. (*Hint:* Call the eigenvectors $\underline{v}_\alpha,\alpha=\pm$, and let $2B=\max_{\underline{\mu}\in\mathbb{Z}^2,|\underline{\mu}|\le N,\alpha=\pm}|\underline{v}_\alpha\cdot\underline{\mu}|$; then note that, by the spectral decomposition of the matrix S_0, one has $|\underline{\nu}|=|S_0^{-p}\underline{\mu}|\le 2|\lambda|^{-|p|}|\underline{\mu}|\le B|\lambda|^{-|p|}$.)

Q10.1.3 **[10.1.3]:** (*Alternative to the convergence proof for H*)
Show without using the Fourier transform that the series for H defined by (10.1.12) converges. (*Hint:* Bound the s-th derivatives by the maximum of F of the Fourier coefficients of $\underline{f}$ times N^s.)

Q10.1.4 **[10.1.4]:** (*Homeomorphism in the case of analytic $\underline{f}$*)
Show that the series for H defined by (10.1.12) converges under the only assumption that $\underline{f}$ is analytic. (*Hint:* Bound the m-th derivatives by the maximum F_∞ of $|\underline{f}(\underline{\varphi})|$ in a strip $|\mathrm{Im}\,\varphi_j|<\xi$ (for some $\xi>0$) times $m!\xi^{-m_v}F_\infty$. The factorial is compensated by the $s_v!$'s in (10.1.12) and the estimate proceeds as in the trigonometric polynomial case.)

Q10.1.5 **[10.1.5]:** (*Hölder continuity in the case of analytic $\underline{f}$*)
Show that the series for H defined by (10.1.12) is Hölder continuous under the only assumption that f is analytic. (*Hint:* Make use of the Fourier transform and do more carefully the same bounds.)

Q10.1.6 **[10.1.6]:** (*The inverse conjugation $\widetilde{H}$*)
Show that the solution of (10.1.4) can be written as

$$\tilde{h}_\alpha(\underline{\psi})=\varepsilon\sum_{p\in\mathbb{Z}_\alpha}\alpha\lambda_\alpha^{-|p+1|\alpha}f_\alpha(S_\varepsilon^p\underline{\psi})\qquad\alpha=\pm$$

where $\widetilde{H}(\underline{\psi})=\underline{\psi}+\tilde{h}(\underline{\psi})$. Show that if we assume that S_ε is still Anosov then $\widetilde{H}$ is a homeomorphism. What can you say of the regularity of $\widetilde{H}$ as a function of $\underline{\varphi}$ and ε? (*Hint:* Writing $\widetilde{H}(\underline{\varphi})=\underline{\varphi}+\underline{\tilde{h}}(\underline{\varphi})$ we get that $\underline{\tilde{h}}(\underline{\varphi})$ satisfies $S_0\underline{\tilde{h}}(\underline{\varphi})-\underline{\tilde{h}}(S_\varepsilon(\underline{\varphi}))=\varepsilon\underline{f}(\underline{\varphi})$.

This equation look very much like (10.1.6) and the linearity of S_0 allows us to write the solution as in (10.1.8) with S_ε in the place of S_0. Invertibility of $\widetilde{H}(\underline{\varphi})$ follows from an argument similar to the one used for H.)

Q10.1.7 **[10.1.7]:** (*Linear part of a homeomorphism on the torus*)
Consider a homeomorphism S of $\mathbb{T}^2$. Show that it can be written in a unique way has $S = L + f$ where L is a linear homeomorphism, *i.e.* is given by a invertible 2×2 matrix with integer entries, and f is a periodic function on $\mathbb{R}^2$. We call L the linear part of S.(*Hint:* consider S as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$ and use periodicity.)

Q10.1.8 **[10.1.8]:** (*Anosov system with the same linear part are conjugated*)
Show that if S_1 and S_2 are two Anosov homeomorphisms of $\mathbb{T}^2$ with the same linear part L and L is also Anosov, then there exist $\widehat{H}$ such that $\widehat{H} \circ S_1 = S_2 \circ \widehat{H}$. (*Hint:* Use the result of problem [10.1.6] to construct $\widetilde{H}_i$ such that $\widetilde{H} \circ S_i = L \circ \widetilde{H}$. Then $\widehat{H} = \widetilde{H}_2 \circ \widetilde{H}_1^{-1}$.)

Q10.1.9 **[10.1.9]:** (*Domain of existence of $\widetilde{H}$ and H: a simple case*)
Assume that $\underline{f}(\underline{\varphi}) = \underline{v}_+ g(\underline{\varphi})$ where g is a trigonometric polynomial. Find a condition on ε that assure that $S_\varepsilon = S_0 + \varepsilon \underline{v}_+ g$ is still Anosov. This will give a condition for the existence and invertibility of $\widetilde{H}$ in problem [10.1.6]. Compare it with (10.1.17). (*Hint:* The differential of S_ε has the same eigendirections of S_0. The relative eigenvalues are $\lambda_{\varepsilon,-}(\underline{\varphi}) \equiv \lambda_-$ and $\lambda_{\varepsilon,+}(\underline{\varphi})$. If $|\lambda_{\varepsilon,+}(\underline{\varphi})| > 1$ problem [10.1.6] allows us to construct $\widetilde{H}$ and shows that it is invertible. The Anosov property for S_ε follows from proposition (4.2.1) where points (i) and (ii) are evident and point (iii) follows from the existence of $\widetilde{H}$.)

Q10.1.10 **[10.1.10]:** (*Domain of existence of $\widetilde{H}$ and H: general case*)
Generalize problem [10.1.9] to a generic trigonometric polynomial $\underline{f}(\underline{\varphi})$.(*Hint:* The lines defined by the two vectors $\underline{c}_\pm = \underline{v}_+ \pm \underline{v}_-$ split $\mathbb{R}^2$ in two cones $\Gamma_\pm$ with the property that $DS_0\Gamma_\pm$ is well inside $\Gamma_\pm$, see problem [4.2.2] for a precise formulation. Write a condition on $\varepsilon \underline{f}(\underline{\varphi})$ so that $DS_\varepsilon \Gamma_\pm$ is still well inside $\Gamma_\pm$. Compare it with (10.1.17).)

Q10.1.11 **[10.1.11]:** (*Continuity of Markov pavements in $d = 2$*)
Consider a two-dimensional Anosov map (Ω, S) and assume that it admits a fixed point x_0. Let S_ε be a small perturbation of S (in class C^∞) depending on a parameter ε. Show that the map S_ε admits a Markov pavement $\mathcal{P}_\varepsilon = \{P_{0\varepsilon}, \ldots, P_{n\varepsilon}\}$ with the same compatibility matrix T as that of $\mathcal{P}_0$. (*Hint:* Construct a Markov pavement $\mathcal{P} = \{P_0, \ldots, P_n\}$ for (Ω, S) by the method of problem [4.3.9]. Note that S_ε will have a fixed point x_ε which merges differentially with x_0 as $\varepsilon \to 0$ together with any finite portion of its stable and unstable manifolds. Note that the construction of $\mathcal{P}$ in problem [4.3.9] is based on two *finite* connected portions of the stable and of the unstable manifolds of x_0.)

Q10.1.12 **[10.1.12]:** (*Anosov structural stability in $d = 2$*)
In the context of problem [10.1.11] show that there is a Hölder continuous map H which conjugates S and S_ε as in (10.1.1) if ε is small enough. (*Hint:* Define a map of Ω into itself by setting $Hx = x'$ if $x' = X_\varepsilon(\underline{\sigma}(x))$ where $\underline{\sigma}(x)$ is the symbolic history of x on the pavement $\mathcal{P}$ under the map S and X_ε is the map that associates with a compatible sequence $\underline{\sigma}$ a point $X_\varepsilon(\underline{\sigma}) \in \Omega$, see definition (4.1.3). Hölder continuity follows from the Hölder continuity of the correspondence between points and symbolic histories on Markov pavements, see proposition (4.1.1), (4.1.6). In other words x, x' are mapped into each other by H if they have the same symbolic representation in the pavements $\mathcal{P}, \mathcal{P}_\varepsilon$ under the maps S, S' respectively.)

§10.2 Extended systems. Lattices of Arnold's cat maps

Let $\Lambda = [-L/2, L/2]^d$ be the cube of side L in $\mathbb{Z}^d$. If $\Omega_\Lambda \stackrel{def}{=} (\mathbb{T}^2)^\Lambda$ we call

$\underline{\varphi} \in \Omega_\Lambda$, $\underline{\varphi} = (\underline{\varphi}_\xi)_{\xi\in\Lambda}$ a *microstate* of the *lattice system* Ω_Λ. The microstates $\underline{\varphi}$ are considered with *periodic boundary conditions*, *i.e.* $\underline{\varphi} = (\underline{\varphi}_\xi)_{\xi\in\mathbb{Z}^d}$ with $\underline{\varphi}_{\xi+\underline{d}_i L} = \underline{\varphi}_\xi$ for every ξ and i, where $\underline{d}_i$ is the unit vector in the direction i in $\mathbb{Z}^d$. We define a map $\mathcal{S}_0 : \Omega_\Lambda \longleftrightarrow \Omega_\Lambda$

e10.2.1 $$(\mathcal{S}_0(\underline{\varphi}))_\xi = S_0\underline{\varphi}_\xi, \qquad S_0 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \tag{10.2.1}$$

and we call $\mathcal{S}_0$ the *unperturbed evolution map* of the microstates.

Let $\underline{f}(\underline{\varphi}_{nn})$ be a $\mathbb{R}^2$-valued function of $2d+1$ arguments in $\mathbb{T}^2$: we label the $2d+1$ arguments $\underline{\varphi}_{nn} = (\underline{\varphi}_0, \underline{\varphi}_1, \cdots, \underline{\varphi}_{2d})$. We imagine that $\underline{\varphi}_0$ is associated with the origin in $\mathbb{Z}^d$ and that the remaining $2d$ arguments are associated with the lattice sites which are nearest neighbors of the origin ordered in a arbitrary way, *e.g.* lexicographically. We shall call $nn(\xi)$ the set formed by ξ and by its nearest neighbors so that $\underline{f}(\underline{\varphi}_{nn(\xi)})$ make sense and depend only on the values of the microstate $\underline{\varphi}$ at ξ and at its neighboring sites.

We shall suppose that we are given a $\underline{f}(\underline{\varphi}_{nn})$ which is a trigonometric polynomial of degree $\le N$ with values in $\mathbb{T}^2$, *i.e.*

e10.2.2 $$\underline{f}(\underline{\varphi}_{nn}) = \sum_{\underline{\nu}_0,\underline{\nu}_1,\ldots,\underline{\nu}_{2d},\,|\underline{\nu}_j|\le N} \underline{f}_{\underline{\nu}_0,\underline{\nu}_1,\ldots,\underline{\nu}_{2d}} e^{i\sum_{j=0}^{2d} \underline{\nu}_j\cdot\underline{\varphi}_j}. \tag{10.2.2}$$

We shall call $\underline{f}$ a *nearest neighbor interaction* and define

e10.2.3 $$(\mathcal{S}_\varepsilon\underline{\varphi})_\xi = S_0\underline{\varphi}_\xi - \varepsilon\underline{f}(\underline{\varphi}_{nn(\xi)}), \tag{10.2.3}$$

which maps $\Omega_\Lambda \longleftrightarrow \Omega_\Lambda$ if ε is small enough. We call $\mathcal{S}_\varepsilon$ the *perturbed evolution map*.

For small ε the system will be an Anosov system by the general structural stability theorem in proposition (10.1.1) and it will be conjugated with the unperturbed system $(\Omega_\Lambda, \mathcal{S}_0)$. However in order to insure that this happens it might be necessary to take ε smaller and smaller as the *infrared cutoff* Λ tends to ∞. Therefore it is important to note that it is not so: *the system remains chaotic enough no matter how spatially extended it is.*

P10.2.1 **(10.2.1) Proposition:** (Uniform structural stability of lattices of maps) *Given a nearest neighbors interaction $\underline{f}$ as above and $\beta \in (0,1)$, there exist $\varepsilon_0(\beta) > 0$ and $C(\beta) < \infty$ such that for all $\Lambda \subset \mathbb{Z}^d$ and for all $|\varepsilon| < \varepsilon_0(\beta)$ there is a homeomorphism $H : \Omega_\Lambda \longleftrightarrow \Omega_\Lambda$ and a constant $\kappa = \kappa(\beta,\varepsilon)$ such that*

e10.2.4 $$H \circ \mathcal{S}_0 = \mathcal{S}_\varepsilon \circ H, \tag{10.2.4}$$

where H is analytic in the disk $|\varepsilon| < \varepsilon_0(\beta)$ and Hölder continuous with range $\kappa(\beta,\varepsilon)^{-1}$ and modulus $C(\beta)$, in the sense that if $\underline{\varphi}, \underline{\varphi}' \in \Omega_\Lambda$ and $\underline{\varphi}_\xi \equiv \underline{\varphi}'_\xi$ for all ξ but for ξ' then

e10.2.5 $$|(H(\underline{\varphi}))_\xi - (H(\underline{\varphi}'))_\xi| \le C(\beta) e^{-\kappa|\xi-\xi'|} |\underline{\varphi}_{\xi'} - \underline{\varphi}'_{\xi'}|^\beta \tag{10.2.5}$$

for all ε *in the complex disk* $|\varepsilon| < \varepsilon_0(\beta)$. *The range constant* $\kappa = \kappa(\beta,\varepsilon)$ *can be taken* $-(2d)^{-1}\log(|\varepsilon|/2\varepsilon_0(\beta))$.

Remarks: (1) As we did in Section §10.1 we can study the equation for $\widetilde{H} = H^{-1}$ and obtain directly a solution that turns out to be invertible if we assume that $\mathcal{S}_\varepsilon$ is still an Anosov system, see problem [10.2.1].
(2) Using an argument similar to the one used in problem [10.1.10] one can prove that $\mathcal{S}_\varepsilon$ is Anosov uniformly in Λ for ε small enough so that $\widetilde{H}$ exists and is invertible, see problem [10.2.2]. As before this construction is not suitable to study the regularity of H as a function of ε. Moreover our proof give us a detailed description of H that will be fundamental in the construction of the SRB measure, see corollary (10.2.1).

Proof: The proof of this proposition is essentially a repetition of the corresponding "single site" $(L = 0, \Lambda = \{0\}, \Omega = \mathbb{T}^2)$ case discussed in proposition (10.1.1). We write H as $(H(\underline{\psi}))_\xi = \underline{\psi}_\xi + (\underline{h}(\underline{\psi}))_\xi$ and decompose $\underline{f}_\xi, \underline{h}_\xi$ along the eigenvectors $\underline{v}_\pm$ of S_0, see (10.1.7). The equations become

e10.2.6

$$\begin{aligned}\lambda_+ h_+(\underline{\psi})_\xi - h_+(S_0\underline{\psi})_\xi &= f_+((\underline{\psi} + \underline{h}(\underline{\psi}))_{nn(\xi)}),\\ \lambda_- h_-(\underline{\psi})_\xi - h_-(S_0\underline{\psi})_\xi &= f_-((\underline{\psi} + \underline{h}(\underline{\psi}))_{nn(\xi)}).\end{aligned} \tag{10.2.6}$$

We can solve the equation (10.2.6) by power series in ε. The first order equation has a solution similar to the corresponding (10.1.9) and the general order recursion can also be written in a form similar to (10.1.10), namely

e10.2.7

$$\begin{aligned}h^{(k)}_{\xi,\alpha}(\underline{\psi}) = \sum_{s=0}^{\infty} \frac{1}{s!} \sum_{\substack{k_1+\cdots+k_s = k-1,\, k_i \ge 0\\ \alpha_1,\ldots,\alpha_s=\pm;\ \xi_1,\ldots,\xi_s \in nn(\xi)}} \sum_{p\in\mathbb{Z}_\alpha} \alpha \lambda_\alpha^{-|p+1|\alpha}\cdot\\ \cdot\Big(\prod_{j=1}^{s} (\partial_{(\alpha_j,\xi_j)})\Big) f_\alpha((S_0^p\underline{\psi})_{nn(\xi)}) \cdot \Big(\prod_{j=1}^{s} h^{(k_j)}_{\xi_j,\alpha_j}(S_0^p\underline{\psi})\Big),\end{aligned} \tag{10.2.7}$$

where $\partial_{(\alpha,\xi)} \stackrel{def}{=} \underline{v}_\alpha \cdot \partial_{\underline{\psi}_\xi}$.

Therefore we can represent the complete $\underline{h}^{(k)}_{\xi,\alpha}(\underline{\psi})$ again in terms of the same tree graphs introduced in Section §10.1 with a few more labels attached to the branches. Namely we must add to each node v a new label ξ_v with the restriction that if the line $\lambda = v'v$ emerging from v enters a node $v' \succ v$ then $\xi_v \in nn(\xi_{v'})$ (which reflects the nearest neighbor property of the interaction). We add also a label ξ_ℓ to the line $\ell = v'v$ by setting $\xi_\ell = \xi_v$.
Hence the *value* of a tree ϑ will be given, see (10.1.12) for comparison, by

e10.2.8

$$\mathrm{Val}(\vartheta) = \prod_{v\in V(\vartheta)} \frac{\alpha_v}{s_v!} \lambda_{\alpha_v}^{-|p_v+1|\alpha_v} \Big(\prod_{j=1}^{s_v} \partial_{(\alpha_{v_j},\xi_{v_j})}\Big) f_{\alpha_v}((S_0^{p(v)}\underline{\psi})_{nn(\xi_v)}), \tag{10.2.8}$$

and it will be a contribution to $h^{(k)}_{\xi,\alpha}$ if α, ξ are the labels attached to the root branch of ϑ and k is the number of nodes of ϑ.

Remark: By construction the sets $nn(\xi_v)$ must all intersect $nn(\xi_{v'})$ if v' is the node immediately following v. Hence the set $\cup_v nn(\xi_v)$ is connected (by nearest neighbors) and, in fact, its connectivity reflects that of the tree to which it is associated.

The estimates are done as in Section §10.1 provided that one takes care of the fact that for each site ξ there are $2d+1$ points in $nn(\xi)$ rather than just 1 as in the case of Section §10.1. This gives an extra factor $(2N+1)^{2(2d+1)k}(2d+1)^k$, that we shall bound (for simplicity) by $(3Ne)^{2(2d+1)k}$, so that, in Fourier space,

e10.2.9
$$\sum_{\vartheta\in\Theta_{k,\underline{\nu},\alpha}}\sum_{v\in V(\vartheta)}|\underline{\nu}_v|^\beta|\mathrm{Val}(\vartheta)|\le \\ \le(3Ne)^{2(2d+1)k}N^kF^k(2/(1-\lambda^{1-\beta}))^k2^{3k} \tag{10.2.9}$$

will be a bound on the sum of the values of the trees of order k (*i.e.* with k nodes): in fact the combinatorics is again the same as in Section §10.1 and it is accounted by the last factor 2^{3k}.

We have obtained a rather explicit expression for $\underline{h}$ which can be expressed as

e10.2.10
$$h_{\xi,\alpha}(\underline{\psi})=\sum_{X\ni\xi}\Phi_{X,\alpha}(\underline{\psi}_X), \tag{10.2.10}$$

where X is a subset, connected by nearest neighbors, of Λ, and Φ_X is a function of $\underline{\psi}_X$ which is translation invariant in the sense that $\Phi_{X,\alpha}(\underline{\psi}_X)=\Phi_{X+\eta,\alpha}(\underline{\psi}_X)$ for all $\eta\in\mathbb{Z}^d$ (the translations must be considered modulo L of course, because of the periodic boundary conditions on Λ). This is because (10.2.8) has the structure of (10.2.10) if $X=\cup_{v\in V(\vartheta)}nn(\xi_v)$.

By the estimate in (10.2.9) we conclude that the functions $\Phi_{X,\alpha}(\underline{\psi}_X)$ are analytic in ε in the complex disk with radius twice the quantity

e10.2.11
$$\varepsilon_0(\beta)=(3Ne)^{-2(2d+1)}N^{-1}F^{-1}(1-\lambda^{1-\beta})2^{-5}, \tag{10.2.11}$$

and that they verify the bounds

e10.2.12
$$\max_{|\varepsilon|\le\varepsilon_0(\beta),\,X,\,\underline{\psi}_X}|\Phi_{X,\alpha}(\underline{\psi}_X)|<B(\beta), \\ \max_{|\varepsilon|\le\varepsilon_0(\beta),\,X,\,\underline{\psi}_{X\setminus\xi'}}|\Phi_{X,\alpha}(\underline{\psi}_X)-\Phi_{X,\alpha}(\underline{\psi}'_X)|<B(\beta)\,|\underline{\psi}_{\xi'}-\underline{\psi}'_{\xi'}|^\beta, \tag{10.2.12}$$

if $\underline{\psi}_X,\underline{\psi}'_X$ coincide at all points but at $\xi'\in\Lambda$, with $B(\beta)$ finite because of
N10.2.1 the extra factor 1/2 we inserted in (10.2.11).[1]

[1] Note that (10.2.9) says that the analyticity disk can be taken to be the r.h.s. of (10.2.11) with 2^{-4} instead of 2^{-5}: it is convenient to give up a factor 2 in the size of $\varepsilon_0(\beta)$ in order to have uniform bounds in the disk of radius $\varepsilon_0(\beta)$.

A further key remark, that follows immediately from the formula (10.2.8) and from the remark on connectivity following it, is that the Taylor coefficient $\Phi^{(k)}_{X,\alpha}(\underline{\psi}_X)$ of order k in ε of $\Phi_{X,\alpha}(\underline{\psi}_X)$ verifies

*e*10.2.13
$$\Phi^{(k)}_{X,\alpha}(\underline{\psi}_X)\equiv 0 \qquad \text{if } k\le \delta(X)/2d, \tag{10.2.13}$$

where $\delta(X)$ is the tree length of the set X, see definition (7.1.3); therefore if $\kappa(\beta,\varepsilon)=-(2d)^{-1}\log(|\varepsilon|/2\varepsilon_0)$ the bound (10.2.12) can be improved into

*e*10.2.14
$$\begin{aligned}&\max_{|\varepsilon|\le\varepsilon_0(\beta),\,\underline{\psi}_X}|\Phi_{X,\alpha}(\underline{\psi}_X)|<B(\beta)\,e^{-\kappa(\beta,\varepsilon)\delta(X)} \qquad (10.2.14)\\ &\max_{|\varepsilon|\le\varepsilon_0(\beta),\,\underline{\psi}_{X\setminus\xi'}}|\Phi_{X,\alpha}(\underline{\psi}_X)-\Phi_{X,\alpha}(\underline{\psi}'_X)|<B(\beta)\,e^{-\kappa(\beta,\varepsilon)\delta(X)}\,|\underline{\psi}_{\xi'}-\underline{\psi}'_{\xi'}|^\beta,\end{aligned}$$

as a consequence of the maximum principle for holomorphic functions. Since $\kappa(\beta,\varepsilon)\to 0$ for $|\varepsilon|\to\varepsilon_0(\beta)$ in order to obtain the result stated in the proposition we have to reduce by an extra factor 2 the value of $\varepsilon_0(\beta)$ obtained in (10.2.11) to obtain the quantity named $\varepsilon_0(\beta)$ in the statement of proposition (10.2.1). ■

The above analysis has led to a result that it is convenient to state separately.

*C*10.2.1 **(10.2.1) Corollary:** (Conjugation potentials) *In the context of proposition (10.1.1) the homeomorphism H can be written as*

*e*10.2.15
$$H(\underline{\psi})_\xi=\underline{\psi}_\xi+\sum_{X\ni\xi}\underline{\Phi}_X(\underline{\psi}_X), \tag{10.2.15}$$

where the sum is over connected sets X, and $\Phi_X(\underline{\psi}_X)$ are translation invariant *functions (i.e. $\Phi_X(\underline{\psi}_X)=\Phi_{X+\eta}(\underline{\psi}_X)$ for $\eta\in\Lambda$) analytic in ε in the complex disk $|\varepsilon|<\varepsilon_0(\beta)$ and verifying, for a suitably chosen constant $B(\beta)$, the bounds*

*e*10.2.16
$$\begin{aligned}&\max_{|\varepsilon|\le\varepsilon_0(\beta),\,\underline{\psi}_X}|\underline{\Phi}_X(\underline{\psi}_X)|<B(\beta)\Big(\frac{|\varepsilon|}{2\varepsilon_0(\beta)}\Big)^{\frac{\delta(X)}{2d}}, \qquad (10.2.16)\\ &\max_{|\varepsilon|\le\varepsilon_0(\beta),\,\underline{\psi}_{X\setminus\xi}}|\underline{\Phi}_X(\underline{\psi}_X)-\underline{\Phi}_X(\underline{\psi}'_X)|<B(\beta)\Big(\frac{|\varepsilon|}{2\varepsilon_0(\beta)}\Big)^{\frac{\delta(X)}{2d}}|\underline{\psi}_\xi-\underline{\psi}'_\xi|^\beta,\end{aligned}$$

*if $\underline{\psi}'_X$ differs from $\underline{\psi}_X$ only at the site ξ and $\delta(X)$ is the tree length of X, for all X (*cf. *definition (7.1.3)).*

Remarks: (1) We can summarize the above analysis by saying that the lattice of coupled maps remains an Anosov system if the perturbation is small enough and the allowed maximum size of ε does not depend on the size of the lattice Λ "containing" the system (proposition (10.2.1)).

(2) The result of corollary (10.2.1) falls short of exhibiting a key property of the conjugation. Suppose that instead of the sequence $\{\underline\psi_\xi\}_{\xi\in\Lambda}$ we represent each coordinate $\underline\psi_\xi$ by an infinite symbolic sequence $\sigma_{\xi,t}$, $\xi\in\Lambda, t\in\mathbb{Z}$ where $\{\sigma_{\xi,t}\}_{t\in\mathbb{Z},\xi\in\mathbb{Z}^d}$ is the symbolic sequence that represents $H^{-1}(\underline\psi_\xi)$ on a Markovian pavement of S_0. Then we can use the corollary to obtain a representation of H in terms of the symbolic dynamics $\{\sigma_{\xi,t}\}_{t\in\mathbb{Z}}$ representing $\underline\psi$ as a sequence on a *space–time lattice*, see also section §(10.4) above proposition (10.4.2) for more details on this contraction. By following the "telescopic" procedure to express a Hölder continuous function in terms of potentials, cf. proposition (4.3.1), we find easily that H can be written

e10.2.17
$$H(\underline\psi)_\xi-\underline\psi_\xi=\sum_{\mathbb{Z}^{d+1}\supset X\ni(\xi,0)}\Phi_X(\underline\sigma_X),\tag{10.2.17}$$

where the summation is restricted to sets X which are rectangles on the $(d+1)$–dimensional lattice and there are constants F,κ such that $|\Phi_X(\underline\sigma_X)|\le Fe^{-\kappa\,\mathrm{diam}(X)}$. We see that H can be expressed in this way in terms of potentials which however do not decay exponentially as the tree length but "only" as the diameter. This would eventually create unsurmountable difficulties when we shall try to derive analyticity of the SRB distribution.
(3) However one can get a much better representation of the form (10.2.17) in which sets more general than rectangles appear in the sum, but the decay rate will be exponential in the tree length $\delta(X)$ rather than in the diameter. This is in fact implicit in the expression (10.2.8) and one just has to read it out: see the proof of proposition (10.4.2), where this is discussed and needed.

The above results will allow us to define a Markov pavement for the extended system quite easily, as we shall show in Section §(10.4), and a detailed construction of the SRB distribution. We proceed to perform the construction by first considering the case of a perturbation of a single map.

Problems for §10.2

Q10.2.1 **[10.2.1]:** (*The inverse conjugation in* $d\ge1$)
Show that the equation $\mathcal{S}_\varepsilon\circ\widetilde H=\widetilde H\circ\mathcal{S}_\varepsilon$ can be solved uniformly in L if one assumes that $\mathcal{S}_\varepsilon$ is Anosov uniformly in Λ. (*Hint:* See problem [10.1.6].)

Q10.2.2 **[10.2.2]:** (*Anosov property for slow decaying interaction*)
Give a condition on ε and $\underline f$ such that $\mathcal{S}_\varepsilon$ is Anosov uniformly in Λ. (*Hint:* Let w be a tangent vector to Ω_Λ. Define $|w|_+=\sup_\xi|(w_\xi,v_+)|$ and $|w|_-=\sup_\xi|(w_\xi,v_-)|$ and $|w|_\infty=\max\{|w|_+,|w|_-\}$. Proceed like in problem [10.1.10] using the cones $\Gamma_+=\{w\,|\,|w|_-\le\alpha|w|_+\}$ and $\Gamma_-=\{w\,|\,|w|_+\le\alpha|w|_-\}$ with $\alpha<1$. Show that if $|D\underline f w|_\infty\le C|w|_\infty$, where $D\underline f$ is the differential of $\underline f$, then it is possible to find ε such that $D\mathcal{S}_\varepsilon\Gamma_+$ is well inside Γ_+ and similarly for Γ_-.)

Q10.2.3 **[10.2.3]:** (*Homeomorphism in the case of analytic* $\underline f$)
Show that the series for H defined by (10.2.7) converges under the only assumption that f is analytic. In particular show that corollary (10.2.1) is still valid. (*Hint:* See [10.1.4]).

Q10.2.4 **[10.2.4]:** (*Exponential decay for* $\underline f$)

Let $\underline{f}$ be a bounded function from $(\mathbb{T}^2)^{\mathbb{Z}^d}$ in $\mathbb{T}^2$ that can be extended to a bounded analytic function on the complex neighbor of $(\mathbb{T}^2)^{\mathbb{Z}^d}$ defined by $|\mathrm{Im}\,\underline{\varphi}_\xi| < \gamma e^{\omega|\xi|}$. Show that for such an $\underline{f}$ the series for H defined by (10.2.7) converges. Is corollary (10.2.1) still valid? (*Hint:* : Use Cauchy estimate to bound the derivatives of $\underline{f}$ with respect to $\underline{\varphi}_\xi$).

Q10.2.5 **[10.2.5]:** (*Tree decay for* $\underline{f}$)
Consider a function $\underline{f}$ from $(\mathbb{T}^2)^{\mathbb{Z}^d}$ in $\mathbb{T}^2$ that can be written as

$$\underline{f}(\underline{\varphi}) = \sum_{X\ni 0} \underline{f}_X(\underline{\varphi}_X),$$

with the functions $\underline{f}_X(\underline{\varphi}_X)$ analytic in the complex neighbor of $(\mathbb{T}^2)^X$ defined by $|\mathrm{Im}\,\underline{\varphi}_\xi| < \gamma e^{\omega\delta(X)}$ and there bounded by a common constant C. Show that for such an $\underline{f}$ the series for H defined by (10.2.7) converges and that H can still be written as in (10.2.10) and Φ_X satisfy (10.2.14) with a suitable decay rate κ. (*Hint:* Use Cauchy estimate to bound the derivatives of $\underline{f}_X$ with respect to $\underline{\varphi}_\xi$).

§10.3 Chaos in time: an SRB distribution

We shall use the notations of Section §10.1, where we have constructed the conjugation H that transforms a perturbed cat map S_ε of the torus $\mathbb{T}^2$ into a "free" cat map S_0. The conjugation is not differentiable (in general) and we could only prove that it can be taken to be Hölder continuous with a prefixed exponent $\beta < 1$ for a perturbation strength that is suitably small.

In spite of the lack of regularity we can still use the homeomorphism $\underline{\varphi} = H(\underline{\psi})$ to construct the dynamics as well as the stable and unstable manifolds of each point. If $\underline{\varphi} = H(\underline{\psi})$ the latter manifolds are given by parametric equations of the form

e10.3.1
$$\underline{\varphi}(t) = H(\underline{\psi} + t\underline{v}_\alpha) \qquad t\in\mathbb{R}, \alpha = \pm, \tag{10.3.1}$$

where $t \to \underline{\psi} + t\,\underline{v}_\alpha$ is the unstable or stable manifold for the unperturbed map S_0 through $\underline{\psi}$ if $\alpha = +$ or $\alpha = -$.

However the above parameterization is not very useful because the function $H(\underline{\psi} + t\underline{v}_\alpha)$ is not regular as a function of t. This can be seen already from the fact that the first order term in its expansion in powers of ε cannot be (in general) differentiated with respect to t at $t = 0$. Indeed we see from (10.1.9) that the component $h_-^{(1)}(\underline{\psi})$ can be differentiated in the direction $\underline{v}_+$ because term by term differentiation enhances convergence since $p < 0$; on the other hand the component $h_+^{(1)}(\underline{\psi})$ cannot be differentiated in the direction $\underline{v}_+$ (unless special cancellations occur) because the convergence factor $\lambda_+^{-(p+1)}$ in (10.1.9) is compensated by the λ_+^p that the differentiation along $\underline{v}_+$ brings out since $p \geq 0$. To second order in ε not even the t–derivative of the component $h_-^{(2)}(\underline{\psi} + \underline{v}_+ t)$ can be shown to exist.

To construct the stable and unstable manifolds as well as the other necessary ingredients to define the SRB distribution we apply once more the technique discussed in the previous two sections. Calling $\widehat\Omega$ the (*non compact*) space $\mathbb{T}^2\times\mathbb{R}^2$ we define the dynamical system[1]

N10.3.1

e10.3.2
$$\widehat S_0(\underline\varphi,\underline v)=(S_0\underline\varphi,S_0\underline v). \tag{10.3.2}$$

This is a system that fails to be an Anosov system because the phase space is not compact. We can nevertheless consider its perturbation

e10.3.3
$$\widehat S_\varepsilon(\underline\varphi,\underline v)=(S_0\underline\varphi+\varepsilon\underline f(\underline\varphi),S_0\underline v+\varepsilon(\underline v\cdot\partial_{\underline\varphi})\underline f(\underline\varphi)), \tag{10.3.3}$$

and we can attempt at finding an isomorphism between $\widehat S_\varepsilon$ and $\widehat S_0$ or, since this turns out to be in general impossible (as it will be implicit in what follows), between $\widehat S_\varepsilon$ and $\widehat S_{0,\varepsilon}$ defined by

e10.3.4
$$\widehat S_{0,\varepsilon}(\underline\varphi,\underline v)=(S_0\underline\varphi,(S_0+\Gamma_\varepsilon(\underline\varphi))\underline v), \qquad (10.3.4).$$

with $\Gamma_\varepsilon(\underline\varphi)$ a matrix *diagonal* on the basis $\underline v_\pm$ (on which S_0 is diagonal too). Therefore we look for a map $\widehat H$ of a simple form and such that $\widehat S_\varepsilon\circ\widehat H=\widehat H\circ\widehat S_{0,\varepsilon}$, *i.e.*

e10.3.5
$$\widehat H\,:\,(\underline\psi,\underline w)\longleftrightarrow(\underline\varphi,\underline v)=(\underline\psi+\underline h(\underline\psi),\underline w+K(\underline\psi)\underline w), \tag{10.3.5}$$

as we already know how to conjugate S_ε with S_0, from the analysis of Section §10.1.

Remarks: (1) Let $\mathcal{K}_\varepsilon(\varphi)=1+K(\varphi)$ and $\mathcal{L}_\varepsilon(\varphi)=S_0+\Gamma_\varepsilon(\varphi)$, the above conjugation is equivalent to the following equation

e10.3.6
$$DS_\varepsilon(H(\underline\varphi))\mathcal{K}_\varepsilon(\underline\varphi)\underline v=\mathcal{K}_\varepsilon(S_\varepsilon\underline\varphi)\mathcal{L}_\varepsilon(\underline\varphi)\underline v. \tag{10.3.6}$$

This implies that the vector $\underline w_\pm(\underline\varphi)=\mathcal{K}_\varepsilon(\underline\varphi)\underline v_\pm$ satisfies:

e10.3.7
$$DS_\varepsilon(H(\underline\varphi))\underline w_\pm(\underline\varphi)=\lambda_\pm(\underline\varphi)\underline w_\pm(S_\varepsilon\underline\varphi) \tag{10.3.7}$$

where $\lambda_\pm(\underline\varphi)$ are the diagonal element of $\mathcal{L}_\varepsilon(\underline\varphi)$.
(2) A naive attempt to construct the stable and unstable directions for S_ε would lead to the equation:

e10.3.8
$$DS_\varepsilon(\underline\varphi)\underline{\tilde w}_\pm(\underline\varphi)=\tilde\lambda_\pm(\underline\varphi)\underline{\tilde w}_\pm(S_\varepsilon\underline\varphi) \tag{10.3.8}$$

Indeed (10.3.8) can be considered as a definition of the stable and unstable directions.
(3) From the general theory we know that $\underline{\widetilde w}_\pm(\underline\varphi)$ are, in general, only

[1] Note that, by allowing the space to be non compact, we are using a definition of dynamical system slightly more general of that given in Section §1.2 and used so far.

Hölder continuous as function of $\underline{\varphi}$, so that the solution of (10.3.8) cannot be analytic in ε, unless special cancellations occur. On the other hand, if $\widetilde{w}_\pm(\underline{\varphi})$ is a solution of (10.3.8) then $\underline{w}_\pm(\underline{\varphi}) = \underline{\widetilde{w}}_\pm(H(\underline{\varphi}))$ is a solution of (10.3.7) and is, as we shall see shortly, analytic in ε. In conclusion we can say that the conjugation defined in (10.3.5) is the right one to look at the stable and unstable directions and its solution give the stable and unstable directions as functions of the "unperturbed" point $H^{-1}(\underline{\varphi})$, cf. the corresponding remarks to proposition (10.1.1).

(4) Equation (10.3.7) does not determine $\Gamma_\varepsilon(\underline{\varphi})$ and $\mathcal{K}(\underline{\psi})$ uniquely. Indeed, if $l(\underline{\varphi})$ is a non zero function from $\mathbb{T}^2$ to $\mathbb{R}$ and $\lambda_\pm(\underline{\varphi})$, $\underline{w}_\pm(S_\varepsilon\underline{\varphi})$ solve (10.3.7), then also $\bar\lambda_\pm(\underline{\varphi}) = \frac{l(S_\varepsilon\underline{\varphi})}{l(\underline{\varphi})}\lambda_\pm(\underline{\varphi})$ and $\underline{\bar v}_\pm(\underline{\varphi}) = l(\underline{\varphi})\underline{w}_\pm(\underline{\varphi})$ solve it.[2]

N10.3.2 To fix this ambiguity we will require that the diagonal elements of $\mathcal{K}(\underline{\varphi})$, on the basis $\underline{v}_\pm$, are equal to 1, *i.e.* the matrix $K(\underline{\varphi})$ is completely off diagonal.

The equation that the matrix $K(\underline{\psi})$ has to verify is

e10.3.9
$$\begin{aligned}&(S_0K(\underline{\psi}) - K(S_0\underline{\psi})S_0)_{ij} = -\varepsilon\partial_{\varphi_j} f_i(\underline{\psi}+\underline{h}(\underline{\psi}))-\\ &-\varepsilon\partial_{\varphi_s} f_i(\underline{\psi}+\underline{h}(\underline{\psi}))K(\underline{\psi})_{sj} + \Gamma_\varepsilon(\underline{\psi})_{ij} + (K(S_0\underline{\psi})\Gamma_\varepsilon(\underline{\psi}))_{ij},\end{aligned} \tag{10.3.9}$$

where $\partial_{\underline{\varphi}}$ denotes a derivative of f with respect to its original argument and repeated indices mean implicit summation (to abridge notations).

We write the above matrix equation on the basis in which S_0 and Γ_ε are diagonal, *i.e.* on the basis formed by the two eigenvectors $\underline{v}_\pm$ of S_0 in which the matrices K, Γ have been assumed to take the form

e10.3.10
$$\Gamma(\underline{\psi}) = \begin{pmatrix}\gamma_+(\underline{\psi}) & 0\\ 0 & \gamma_-(\underline{\psi})\end{pmatrix}, \qquad K(\underline{\psi}) = \begin{pmatrix}0 & k_+(\underline{\psi})\\ k_-(\underline{\psi}) & 0\end{pmatrix}. \tag{10.3.10}$$

If $\alpha = \pm$ and $\beta = -\alpha$ (10.3.9) becomes, by setting $\partial_\alpha = \underline{v}_\alpha\cdot\partial_{\underline{\varphi}}$,

e10.3.11
$$\begin{aligned}&0 = -\varepsilon\partial_\alpha f_\alpha(\underline{\varphi}) - \varepsilon K_{\beta\alpha}(\underline{\psi})\partial_\beta f_\alpha(\underline{\varphi}) + \gamma_\alpha(\underline{\psi}),\\ &(\lambda_\alpha K_{\alpha\beta}(\underline{\psi}) - \lambda_\beta K_{\alpha\beta}(S_0\underline{\psi})) =\\ &\qquad = -\varepsilon\partial_\beta f_\alpha(\underline{\varphi}) - \varepsilon K_{\alpha\beta}(\underline{\psi})\partial_\alpha f_\alpha(\underline{\varphi}) + K_{\alpha\beta}(S_0\underline{\psi})\gamma_\beta(\underline{\psi}),\end{aligned} \tag{10.3.11}$$

where $\underline{\varphi}$ means $\underline{\psi}+\underline{h}(\underline{\psi})$, $\lambda_+ = \lambda^{-1}, \lambda_- = -\lambda$ are the eigenvalues of S_0 (with $\lambda = (\sqrt5-1)/2$), and $\gamma_\alpha(\underline{\psi}) = \underline{v}_\alpha\cdot\Gamma(\underline{\psi})\underline{v}_\alpha$.

Calling $k_\alpha(\underline{\psi}) = K_{\alpha\beta}(\underline{\psi})$, with $\alpha = \pm$ and $\beta = -\alpha$, we can rewrite the equations (10.3.11) as

e10.3.12
$$\begin{aligned}&\gamma_\alpha(\underline{\psi}) = \varepsilon\,\partial_\alpha f_\alpha(\underline{\varphi}) + \varepsilon\,k_\beta(\underline{\psi})\partial_\beta f_\alpha(\underline{\varphi}),\\ &k_\alpha(\underline{\psi}) + \lambda^2 k_\alpha(S_0^\alpha\underline{\psi}) =\\ &\qquad \alpha\lambda\big(-\varepsilon\partial_\beta f_\alpha(\underline{\varphi}^\alpha) - \varepsilon k_\alpha(\underline{\psi}^\alpha)\partial_\alpha f_\alpha(\underline{\varphi}^\alpha) + k_\alpha(S_0\underline{\psi}^\alpha)\gamma_\beta(\underline{\psi}^\alpha)\big),\end{aligned} \tag{10.3.12}$$

---

[2] $l(\underline{\varphi})$ is sometime called a *cocycle*. Observe that the expansion coefficient on the unstable manifold with respect to the parameterization defined by $\underline{\bar w}_+$ is linked to that of $\underline{w}_+$ by $l(\underline{\varphi})(\underline{w}_+(\underline{\varphi})\cdot\underline{\partial})S^n_\varepsilon(\underline{\varphi}) = (\underline{\bar w}_+(\underline{\varphi})\cdot\underline{\partial})S^n_\varepsilon(\underline{\varphi})l(S^n_\varepsilon(\underline{\varphi}))$.

where $\alpha = \pm$ and $\underline{\varphi}$ must be thought of as denoting $\underline{\psi} + \underline{h}(\underline{\psi})$ (hence $S_\varepsilon^{-1}\underline{\varphi}$ means $S_0^{-1}\underline{\psi} + \underline{h}(S_0^{-1}\underline{\psi})$), and we have set $\underline{\psi}^\alpha = S_0^{-(1-\alpha)/2}\underline{\psi}$ and likewise $\underline{\varphi}^\alpha = S_\varepsilon^{-(1-\alpha)/2}\underline{\varphi}$. These equations are in a form suitable for a recursive solution in powers of ε. For instance the first order is

e10.3.13

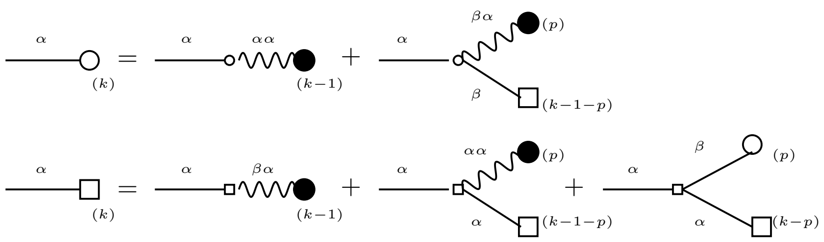
$$\begin{aligned}
&\gamma_\alpha^{(1)}(\underline{\psi}) = \partial_\alpha f_\alpha(\underline{\psi}), \qquad \alpha = \pm,\\
&k_+^{(1)}(\underline{\psi}) + \lambda^2 k_+^{(1)}(S_0\underline{\psi}) = -\lambda\,\partial_- f_+(\underline{\psi}),\\
&k_-^{(1)}(\underline{\psi}) + \lambda^2 k_-^{(1)}(S_0^{-1}\underline{\psi}) = \lambda\,\partial_+ f_-(S_0^{-1}\underline{\psi}),
\end{aligned} \tag{10.3.13}$$

which has the solution

e10.3.14

$$\begin{aligned}
&\gamma_\alpha^{(1)}(\underline{\psi}) = \partial_\alpha f_\alpha(\underline{\psi}) \qquad \alpha = \pm\\
&k_+^{(1)}(\underline{\psi}) = -\lambda \sum_{n=0}^{\infty} (-1)^n \lambda^{2n} \partial_- f_+(S_0^n \underline{\psi}),\\
&k_-^{(1)}(\underline{\psi}) = \lambda \sum_{n=0}^{\infty} (-1)^n \lambda^{2n} \partial_+ f_-(S_0^{-(n+1)} \underline{\psi}).
\end{aligned} \tag{10.3.14}$$

The equations (10.3.12) can be represented in graph form by suitably modifying the similar representation derived for $\underline{h}^{(k)}$ in Section §10.1:



F10.3.1 **Fig.(10.3.1)** Here $\alpha = \pm$ and $\beta = -\alpha$. All the lines have to imagined to carry arrows (not drawn) pointing toward the root. The line carrying a label α and emerging from a circle or a square with label k denotes $\gamma_\alpha^{(k)}$ or $k_\alpha^{(k)}$, respectively. The wavy line emerging from a bullet with label p, with $1 \le p < k-1$, carrying a pair of labels γ, δ, represents $[\partial_\gamma f_\delta]^{(p)}$, the p-th order in the power expansion in ε of $\partial_\gamma f_\delta$ (evaluated at a point dependent on ε, see below). The small circle or square in the the last node (*i.e.* in the node closest to the root) expresses that $\underline{\varphi}$ (circle) or $S_\varepsilon^{\widetilde{q\alpha}}\underline{\varphi}^\alpha$ (square) is the argument in which the functions $\partial_\gamma f_\delta$ are computed, with $\widetilde{q} = q$ if $\delta = +$ and $\widetilde{q} = q+1$ if $\delta = -$, or it expresses that $\underline{\psi}$ (circle) or $S_0^{\widetilde{q\alpha}}\underline{\psi}^\alpha$ (square) is the argument in which the functions γ_α, k_α are computed, with $\widetilde{q} = q$ if $\alpha = +$ and $\widetilde{q} = q+1$ if $\alpha = -$ (for the square in the last graph contributing to $k_\alpha^{(k-1-p)}(\underline{\psi})$ there is an extra S_0 in the argument). Furthermore a summation over $q = 0, 1, \ldots$ and a multiplication by $-\alpha(-1)^q\lambda^{1+2q}$ is understood to be performed over the nodes represented as small squares.

The representation is drawn in figure (10.3.1) and the symbols are explained in the corresponding caption: the reader will recognize in them a pictorial rewriting of (10.3.12).

In this case too we can continue the expansion until in the r.h.s. of (10.3.9) all endpoints of the graph are either squares or circles carrying a label (1) (*i.e.* they represent a first order contribution to Γ or to K (circle or square) bullets representing either $\partial_\alpha f_\beta(\underline{\varphi})$ or $\partial_\alpha f_\beta(\widetilde{S_0^{q\alpha}\underline{\varphi}^\alpha})$. Of course the latter quantities can themselves be represented by the tree expansion discussed in Section §10.1: if we do so then we obtain a full expansion in powers of ε in which the wavy lines with label p are replaced by a tree with p nodes.

The rule to construct the value of each tree graph is easily read from (10.3.9) and from the rules discussed above and in Section §10.1 to build the value of trees representing $\underline{h}$.

The estimate of the k–th order contribution is given by (10.1.16) with an extra factor N^k to take into account the extra derivatives due to the lines with two labels. Also the counting of the trees has to be modified but at the end result will be that K is expressed by a convergent series in ε for $|\varepsilon| < \varepsilon_0(\beta)$, where $\varepsilon_0(\beta)$ can be taken of the form (10.1.17) with a different numerical factor and with N replaced by N^2. The above analysis yields the following result.

P10.3.1 **(10.3.1) Proposition:** (Symbolic representation of the expansion rate) *Given a Markovian pavement $\mathcal{P}_0 = \{P_1, \ldots, P_n\}$ of $\mathbb{T}^2$ for S_0, let $\underline{\sigma}$ be the symbolic representation with respect to the Markov partition $\mathcal{P}_\varepsilon = H(\mathcal{P}_0)$ of a point $\underline{\varphi} = X_\varepsilon(\underline{\sigma})$. There exists $\varepsilon_0(\beta)$ such that the expansion rate $\lambda_u(\underline{\sigma})$ of S_ε along the unstable manifold of $\underline{\varphi}$, see definition (4.3.2), is defined and holomorphic in ε in the disk $|\varepsilon| < \varepsilon_0(\beta)$. As a function of $\underline{\sigma}$ it is Hölder continuous of exponent β and modulus $C(\beta)$.*

Proof: One just notes that if $\underline{\varphi} = H(\underline{\psi})$ then the unstable direction at $\underline{\varphi}$ will be $\underline{w}_+(\underline{\psi}) = \underline{v}_+ + K(\underline{\psi})\underline{v}_+$. Furthermore a Markov pavement for S_ε will be the image of a Markov pavement for S_0 under the map $\underline{\psi} \to H(\underline{\psi})$; therefore $\underline{\psi}$ evolved with S_0 and $\underline{\varphi}$ evolved with S_ε will have the *same symbolic history* $\underline{\sigma}$ on such pavements. Hence $X_\varepsilon(\underline{\sigma}) = \underline{\varphi}$ and $X_0(\underline{\sigma}) = \underline{\psi}$ if $\underline{\varphi} = \underline{\psi} + \underline{h}(\underline{\psi})$. The expansion rate along the unstable manifold will be, following the notations of definition (4.3.2), $\lambda_u(\underline{\sigma}) = e^{A_u(\underline{\sigma})}$ with

e10.3.15
$$A_u(\tau\underline{\sigma}) = \log\left(\big(\lambda_+ + \gamma_+(\psi)\big)\frac{|\underline{w}_+(S_0\underline{\psi})|}{|\underline{w}_+(\underline{\psi})|}\right) \tag{10.3.15}$$

where $|\underline{w}_+(\underline{\psi})| = \sqrt{1 + k_+(\psi)^2}$. The functions γ_+, k_+ are analytic in ε and Hölder continuous in $\underline{\psi}$ with a uniformly bounded modulus $C(\beta)$ if $\beta < 1$ and ε is small enough and since $\underline{\sigma}$ fixed means $\underline{\psi}$ fixed (because $\underline{\sigma}$ is the history of a point $\underline{\psi}$ on a Markov pavement for the ε–independent map S_0) we see that $A_u(\underline{\sigma})$ is analytic in ε at fixed $\underline{\sigma}$ and Hölder continuous in $\underline{\sigma}$ as well as in $\underline{\psi}$ and therefore in $\underline{\varphi}$. Hence the function $A_u(\underline{\sigma})$ generates a potential of Fisher type (cf. (4.3.16) and propositions (4.3.1) and (4.3.2) apply). ■

Remark: The above proposition allows us to conclude that the SRB distribution μ_ε will be a Gibbs state for the energy function $A_u(\underline{\sigma})$ and, therefore, it will mix at an exponential rate any pair of Hölder continuous function. Observe that if G is a Hölder continuous function then $G \circ H^{-1}$ has a representation on the Markov pavement $\mathcal{P}_\varepsilon$ that is independent of ε. Moreover Gibbs states with exponentially decaying Fisher potential have the property that the expectation values of observables depend analytically on the parameters on which the potential itself depends analytically, cf. proposition (7.3.1) ([CO81]). This can be summarized in the following corollary.

$C10.3.1$ **(10.3.1) Corollary:** (Mixing for SRB distributions)
Let F and G be two Hölder continuous "observables" (i.e. functions on $\mathbb{T}^2$). Then if μ_ε denotes the SRB distribution for S_ε the following properties hold.
(i) The expectation values $\mu_\varepsilon(F)$ and $\mu_\varepsilon(G)$ are defined and Hölder continuous in ε. Moreover the expectation values $\mu_\varepsilon(F \circ H^{-1})$ and $\mu_\varepsilon(G \circ H^{-1})$ are analytic function in ε for $|\varepsilon| \le \varepsilon_0$, with ε_0 independent of F or G.
(ii) If F is an analytic observable then the expectation value $\mu_\varepsilon(F)$ is analytic in ε for $\varepsilon \le \varepsilon_F$, with ε_F dependent on F.
(iii) The functions F, G mix at an exponential rate in the sense that the difference between the l.h.s. and the r.h.s. of

$e10.3.16$
$$\mu_\varepsilon((S_\varepsilon^n F)G) \equiv \int \mu_\varepsilon(d\underline{\varphi})F(S^n\underline{\varphi})G(\underline{\varphi}) \xrightarrow[n\to\pm\infty]{} \mu_\varepsilon(F)\mu_\varepsilon(G) \tag{10.3.16}$$

tends to 0 bounded by a constant (depending of F, G) times $e^{-\kappa_{F,G} n}$, for a suitable constant $\kappa_{F,G} > 0$.
(iv) The volume distribution $\mu_0(d\underline{\varphi}) = d\underline{\varphi}/(2\pi)^2$ "mixes with the SRB distribution" exponentially fast in the sense that given the functions F, G the difference between the l.h.s. and the r.h.s. of

$e10.3.17$
$$\mu_0((S_\varepsilon^n F)G) \equiv \int \mu_0(d\underline{\varphi})F(S^n\underline{\varphi})G(\underline{\varphi}) \xrightarrow[n\to+\infty]{} \mu_\varepsilon(F)\mu_0(G) \tag{10.3.17}$$

tends to 0 bounded by a constant (depending of F, G) times $e^{-\kappa_{F,G} n}$.
The latter limit has, in general, a different value if $n \to -\infty$ and one has to replace the SRB distribution μ_ε with the one for S_ε^{-1} (which essentially has the same properties as μ_ε).

We can therefore appreciate the power of the perturbation expansion method which allows us to obtain rather detailed and precise informations about the Markov partition (which of course can be defined as the H–image of the trivial partition for Arnold's cat map S_0); and, what is more important, it provides us with an analytic description of the SRB distribution (in the considered small perturbation problem).

In particular it is interesting to compute the derivatives of $\mu_\varepsilon(F)$ for $\varepsilon = 0$. Since $\mu_\varepsilon = \lim_{N\to\infty} \mu_\varepsilon^{(N)}$, with $\mu_\varepsilon^{(N)}$ being the finite volume Gibbs distribution in volume $\Lambda_N = [-N, N]$ with potential energy function $A(\underline{\sigma}) =$

$\lambda_u(\underline{\sigma})$, cf. (6.1.15), with Δ_{Λ_N} chosen to correspond to periodic boundary conditions; then the derivative at $\varepsilon = 0$ of $\mu_\varepsilon^{(N)}(F)$ is

e10.3.18
$$\begin{aligned}&\mu_0^{(N)}(\underline{\partial}F\cdot\underline{h}^{(1)})-\lambda_+\sum_{k=-N}^{N}(\mu_0^{(N)}(-(\partial_+f_+)\circ S_0^kF)-\\&-\mu_0^{(N)}(-(\partial_+f_+)\circ S_0^k)\mu_0^{(N)}(F)),\end{aligned}\tag{10.3.18}$$

because the first order of the derivative of $\lambda_u(\underline{\sigma})$ is (in the $\underline{\psi}$ coordinates $\partial_+f_+(\underline{\psi})$ (cf. (10.3.12))). But $\mu_0^{(N)}(-(\partial_+f_+)\circ S_0^k)\xrightarrow[N\to\infty]{}0$ because $\mu_0(\partial_+f_+)\equiv 0$. Hence, up to an interchange of limits which it is not difficult to justify, the derivative of $\mu_\varepsilon(F)$ is given by $\mu_0(\underline{h}^{(1)}\cdot\underline{\partial}F)+\lambda_+\sum_{k=-\infty}^{\infty}\mu_0(\partial_+f_+F\circ S_0^k)$. Note that

e10.3.19
$$\begin{aligned}\mu_0(\underline{h}^{(1)}\cdot\underline{\partial}F)&=\sum_{-\infty}^{k=0}\mu_0(\lambda_+^{k-1}f_+\partial_+F\circ S_0^k)-\sum_{k=1}^{\infty}\mu_0(\lambda_-^{k-1}f_-\partial_-F\circ S_0^k)\\&=\sum_{-\infty}^{k=-1}\mu_0(f_+\circ S_0^{-1}\partial_+(F\circ S_0^k))-\\&\quad\sum_{k=0}^{\infty}\mu_0(f_-\circ S_0^{-1}\partial_-(F\circ S_0^k))\end{aligned}\tag{10.3.19}$$

We can now write $\sum_{k=-\infty}^{\infty}\mu_0(\partial_+f_+F\circ S_0^k)=\sum_{k=-\infty}^{\infty}\mu_0(\partial_+(f_+F\circ S_0^k))-\sum_{k=-\infty}^{\infty}\mu_0(f_+\partial_+(F\circ S_0^k))$. After further elaboration and taking into account that $\mu_0(\partial_+G)=0$ for every differentiable G, we get $\lambda_+\sum_{k=-\infty}^{\infty}\mu_0(\partial_+f_+F\circ S_0^k)=\sum_{k=-\infty}^{\infty}\mu_0(f_+\circ S_0^{-1}\partial_+(F\circ S_0^k))$ so that

e10.3.20
$$\partial_\varepsilon\mu_\varepsilon(F)\big|_{\varepsilon=0}=-\sum_{k=0}^{\infty}\mu_0(\underline{\partial}(F\circ S_0^k)\cdot\underline{f}\circ S_0^{-1}).\tag{10.3.20}$$

This formula was proved, in a much more general setting, by Ruelle, see [Ru97], and is very interesting for applications because it closely resembles the standard Green-Kubo formula.[GR97]

Remark: We note that once $\underline{w}_+(\underline{\varphi})$ is known we can construct the unstable manifold of a point $\underline{\varphi}_0$ solving the differential equation

e10.3.21
$$\begin{cases}\dot{\mathcal{X}}(t,\underline{\varphi}_0)=\underline{\widetilde{w}}_+(\mathcal{X}(t,\underline{\varphi}_0))=\underline{w}_+(H^{-1}(\mathcal{X}(t,\underline{\varphi}_0))),\\ \mathcal{X}(0,\underline{\varphi}_0)=\underline{\varphi}_0,\end{cases}\tag{10.3.21}$$

where $\mathcal{X}(t,\underline{\varphi}_0)$ is a parameterization of the unstable manifold $W^u(\underline{\varphi}_0)$. Clearly this parameterization is differentiable in t. From Section §4.2 we know that $W^u(\underline{\varphi}_0)$ is at least a C^∞ manifold so that it should be possible to find a parameterization much smoother than the one given by (10.3.21).

Moreover (10.3.21) does not allow us to discuss the smoothness of $\mathcal{X}(t,\underline{\varphi})$ as a function of ε.

A general theory for $\mathcal{X}(t,\underline{\varphi})$ on the lines developed in this section can be achieved by generalizing equation (10.3.4) or (10.3.7), see problem [10.3.6].

Problems for §10.3

Q10.3.1 **[10.3.1]:** (Regularity of the SRB measure for analytic f.)
Extend the result of this section to the case of an analytic $\underline{f}$. (*Hint:* See problem [10.1.4]).

Q10.3.2 **[10.3.2]:** (*Generalization of (10.3.5) to the nonlinear part of $W^+(\underline{\psi})$: an attempt.*) Let $\widetilde{S}_\varepsilon$ be defined on $\mathbb{T}^2\times\mathbb{R}^2$ by

$$\widetilde{S}_\varepsilon(\underline{\varphi},\underline{v}) = (S_\varepsilon(\underline{\varphi}), S_\varepsilon(\underline{\varphi}+\underline{v}) - S_\varepsilon(\underline{\varphi}))$$

and $\widehat{S}_{0,\varepsilon}$ by (10.3.4). We can look for a conjugation $\widetilde{H}$ of the form

$$\widetilde{H}\,:\,(\underline{\psi},\underline{w})\longleftrightarrow(\underline{\varphi},\underline{v}) = (\underline{\psi}+\underline{h}(\underline{\psi}),\mathcal{X}(\underline{\psi},\underline{w})). \qquad (*)$$

Show that, if $\widetilde{H}$ satisfies $\widetilde{H}\circ\widehat{S}_{0,\varepsilon} = \widetilde{S}_\varepsilon\circ\widetilde{H}$ then, $\underline{\partial}\mathcal{X}(\underline{\psi},0)\underline{w} = \underline{w}+K(\underline{\psi})\underline{w}$ with $K(\underline{\psi})$ defined by (10.3.5).

Q10.3.3 **[10.3.3]:** (*Failure of the attempt in problem [10.3.2].*)
Write an equation for the $n+1$-order tensor $\underline{\partial}^n\mathcal{X}(\psi,0)$ for $n=2,3$. Expand this equation in series of ε and consider the equation for $\underline{\partial}^n\mathcal{X}^{(0)}(\psi,0)$. Are these equations solvable? (*Hint:* Check the equation for $\partial_+^2\partial_-\mathcal{X}_+(\psi,0)$.)

Q10.3.4 **[10.3.4]:** (*Tree expansion for the nonlinear part of $W^+(\psi)$.*)
Let $\overline{S}_{0,\varepsilon}$ be the restriction of $\widetilde{S}_\varepsilon$ to $\mathbb{T}^2\times\mathbb{R}v_+$, *i.e.*

$$\overline{S}_{0,\varepsilon}(\underline{\varphi},t) = (S_0(\underline{\varphi}),(\lambda_+ + \gamma_+(\underline{\varphi}))t)$$

and analogously

$$\overline{H}(\underline{\psi},t) = (\underline{\psi}+\underline{h}(\underline{\psi}),\mathcal{X}(\underline{\psi},t)),$$

where we want that

$$\overline{H}\circ\overline{S}_{0,\varepsilon} = \widetilde{S}_\varepsilon\circ\overline{H}.$$

Note that both sides of the equation represent a function from $\mathbb{T}^2\times\mathbb{R}v_+$ to $\mathbb{T}^2\times\mathbb{R}^2$. Assume that

$$\mathcal{X}(\underline{\psi},t) = \sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\mathcal{X}^{(n,k)}(\underline{\psi})t^n\varepsilon^k.$$

Write a tree expansion for $\mathcal{X}^{(n,k)}(\underline{\psi})$ and prove that $\mathcal{X}(\underline{\psi},t)$ is Hölder continuous of exponent β in $\underline{\psi}$ and analytic in t and ε for $|t|\le t_0$ and $|\varepsilon|\le\varepsilon(\beta)$.(*Hint:* We have that $\mathcal{X}(\underline{\psi},t)$ satisfies the equation:

$$S_\varepsilon(H(\underline{\psi})+\mathcal{X}(\underline{\psi},t)) = H(S_0(\underline{\psi}))+\mathcal{X}(S_0(\underline{\psi}),\lambda_{+,\varepsilon}(\underline{\psi})t).$$

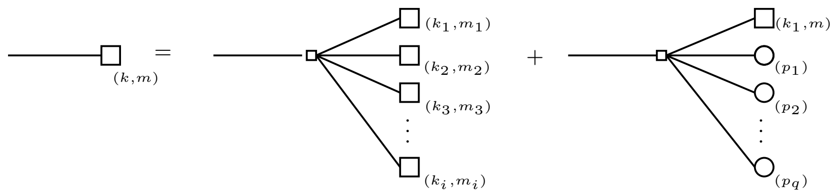
Expanding in serie of ε and t we get:

$$\lambda^m\mathcal{X}^{(m,k)}(\underline{\psi}) - \mathcal{X}^{(m,k)}(S_0(\underline{\psi})) =$$

$$\lambda^m\sum_i\frac{D^i f(\underline{\psi})}{i!}\sum_{\substack{m_1+m_2+\cdots+m_i=m\\ k_1+k_2+\cdots+k_i=k-1}}\mathcal{X}^{(m_1,k_1)}(\underline{\psi})\mathcal{X}^{(m_2,k_2)}(\underline{\psi})\cdots\mathcal{X}^{(m_i,k_i)}(\underline{\psi})+$$

$$+\sum_{q\le m}\lambda^q\sum_p\mathcal{X}^{(m,k-p)}(\underline{\psi})\sum_{p_1+p_2+\cdots+p_q=p}\lambda_+^{(p_1)}(\underline{\psi})\lambda_+^{(p_1)}(\underline{\psi})\cdots\lambda_+^{(p_q)}(\underline{\psi})$$

where $k_i \geq 1$, $m_i \geq 0$, $p_i \geq 0$. Finally $\lambda_+^{(0)}(\underline{\psi}) = \lambda_+$ and $\mathcal{X}^{(0,k)}(\underline{\psi}) = h^{(k)}(\underline{\psi})$. Moreover a proper contraction of the index of the tensor $D^i f(\underline{\psi})$ with the component of $\mathcal{X}^{(m_i,k_i)}(\underline{\psi})$ in understood. This can be represented graphically as in the following figure.



F10.3.2 **Fig.(10.3.2)** Graphical representation of the expansion for $\mathcal{X}^{(m,k)}(\underline{\psi})$. We call the second graphical element on the right hand side a *counterterm vertex*.

The representation for $\lambda_+^{(p)}$ is similar to the one in figure (10.3.1) where we must replace the bullet with label p with a square with label $(p, 0)$, the wavy line with a line and the square as to be considered with label $(p, 1)$.
We can immagine that the small square in the vertex carry a label m equal to the sum of the m labels of the square entering in it. With this convention the line exiting from a small square with label $m \neq 1$ represent the operator $\lambda^m(\mathbf{T}_m^{-1})$ where

$$\mathbf{T}_m \mathcal{X}^{(m)}(\underline{\psi}) = \lambda^m \mathcal{X}^{(m)}(\underline{\psi}) - \mathcal{X}^{(m)}(S_0(\underline{\psi}))$$

The inverse of $\mathbf{T}_m$ for $m = 0$ was computed while studying the conjugation H while the cases $m \geq 2$ can be easily solved and do not require to divide $\mathcal{X}^{(m)}$ in its $+$ and $-$ components.
The case $m = 1$ need a particular treatment. In this case the line exiting from a small square with label 1 represent the operator $\lambda \mathbf{T}_1^{-1}\mathbf{P}_+$ where $\mathbf{P}_+$ is the projection on the $+$ direction. In the same way the line exiting from a small circe (that has necessarily a label 1) represent $\lambda \mathbf{P}_-$, see figure (10.3.1). We can now iterate the graphical equation of figure (10.3.2) untill all square have label $(m, 1)$ or all circle have label 1. The value of the square with label $(m, 1)$ for $m = 0$,1 and for the circle with $p = 1$ has already been computed. For $m \geq 2$ it easy to find that

$$\mathcal{X}^{(m,1)}(\underline{\psi}) = \sum_{n=0}^{\infty} \lambda^{mn} S_0^n \partial_+^m f(S_0^{-n-1}(\underline{\psi})).$$

The analysis of the convergence goes has in the case of the conjugation. Note that the a counterterm vertex has no ε factor associated but does not contribute a $C^q q!$ to the estimates because it bear no derivative of f. Moreover the number of operator $\mathbf{T}_m^{-1}$ associated with a tree is equal to the number of vertex with a small square plus the number of final points, *i.e.* it is equal to k if the tree contributes to $\mathcal{X}^{(m,k)}$. Finally it is easy to check that $\|\mathbf{T}_m^{-1}\| \leq CK^m$ for suitable constant C and K. The only thing left is to estimate the number of tree that contributes to $\mathcal{X}^{(m,k)}$. To do this one can immagine that to every final point with label $(m, 1)$ are attached m incoming lines with a final points with label $(1, 0)$. This reduces the estimates on the number of tree to the case already studied.)

Q10.3.5 **[10.3.5]:** Show that equation $(*)$ in problem [10.3.2] is equivalent to the equation

$$S_\varepsilon(H(\underline{\varphi}) + \mathcal{X}(\underline{\varphi}, t)) = H(S_0\underline{\varphi}) + \mathcal{X}(S_0\underline{\varphi}, \lambda_+(\underline{\varphi})t), \qquad (**)$$

where $\lambda_+(\underline{\varphi}) = \lambda_+ + \gamma_+(\underline{\varphi})$. A natural definition of the unstable manifold would be given by

$$S_\varepsilon(\underline{\varphi} + \widetilde{\mathcal{X}}(\underline{\varphi}, t)) = S_0\underline{\varphi} + \mathcal{X}(S_0\underline{\varphi}, \widetilde{\lambda}_+(\underline{\varphi})t).$$

What is the relation between $\widetilde{\mathcal{X}}(\underline{\varphi},t)$ and $\mathcal{X}(\underline{\varphi},t)$?

Q10.3.6 **[10.3.6]:** (*Extension of the parametrization of* $W^+(\psi)$.)
Use equation $(**)$ of problem [10.3.5] to show that $\mathcal{X}(\underline{\varphi},t)$ can be extended to an analytic function of t in a complex strip around the real axis. How large is this strip? (*Hint:* Use $(**)$ to extend the domain of analyticity of $\mathcal{X}(\underline{\varphi},t)$ by a factor $\lambda_+(\underline{\varphi})$. Note that $\mathcal{X}(\underline{\varphi},t)$ should be in the domain of analyticity of $\underline{f}(\underline{\varphi})$.)

§10.4 Chaos in space–time and SRB distributions

The theory of Section §10.2 can be extended along the lines of Section §10.3 to lattices of "coupled maps" thereby answering various natural questions.

Adopting the notations of Section §10.2 we consider a lattice of Arnold's cat maps. Following the method of Section §10.3 we can construct the unstable manifold by studying the dynamical system (on a noncompact phase space $\widetilde{\Omega}_\Lambda = (\mathbb{T}^2\times\mathbb{R}^2)^\Lambda$, see footnote 1)

e10.4.1
$$(\mathcal{S}_\varepsilon(\underline{\varphi},\underline{v}))_\xi = \big(S_0\underline{\varphi}_\xi - \varepsilon\underline{f}(\underline{\varphi}_{nn(\xi)}), S_0\underline{v}_\xi - \varepsilon\underset{\sim}{v}_\xi\cdot\partial_{\underset{\sim}{\varphi}_\xi}\underline{f}(\underline{\varphi}_{nn(\xi)})\big). \tag{10.4.1}$$

Therefore a point $\underline{\varphi}$ in $(\mathbb{T}^2)^\Lambda$ can be identified by assigning for each $\xi\in\Lambda$ a pair of coordinates $\underline{\varphi}_\xi$ labeled by ξ which identify a point in $\mathbb{T}^2$; and a vector $\underline{w}$ in $(\mathbb{R}^2)^\Lambda$ can be identifies by assigning $|\Lambda|$ two-component vectors $\underline{w}_\xi$ labeled by a point in Λ.

We construct functions $\underline{h}(\underline{\psi})$ and $K(\underline{\psi})$ so that

e10.4.2
$$\begin{aligned}\underline{\varphi}_\xi =&\underline{\psi}_\xi + \underline{h}_\xi(\underline{\psi}) \stackrel{def}{=} H_\xi(\underline{\psi}),\\ \underline{v}_\xi =&\underline{w}_\xi + (K(\underline{\psi})\underline{w})_\xi\end{aligned} \tag{10.4.2}$$

defines a map $\widetilde{H}$ of $\widetilde{\Omega}_\Lambda$ into itself, where now $\underline{h}(\underline{\psi})$ is a function defined on $(\mathbb{T}^2)^\Lambda$ with values in $(\mathbb{T}^2)^\Lambda$ and K is defined on $(\mathbb{T}^2)^\Lambda$ with values in the $2|\Lambda|\times 2|\Lambda|$ matrices mapping $(\mathbb{R}^2)^\Lambda$ into itself. The construction is achieved by imposing that the map $\widetilde{H}$ transforms $\widetilde{\mathcal{S}}_\varepsilon$ into $\widetilde{\mathcal{S}}_{0,\varepsilon}$, *i.e.*

e10.4.3
$$\begin{aligned}&\widetilde{\mathcal{S}}_\varepsilon\circ\widetilde{H} = \widetilde{H}\circ\widetilde{\mathcal{S}}_{0,\varepsilon}, \qquad \text{with}\\ &\widetilde{\mathcal{S}}_{0,\varepsilon}(\underline{\psi},\underline{w})_\xi = \big(S_0\underline{\psi}_\xi, S_0\underline{w}_\xi + (\Gamma(\underline{\psi})\underline{w})_\xi\big),\end{aligned} \tag{10.4.3}$$

with $\Gamma(\underline{\psi})$ depending on ε. Let $\underline{B}$ be the basis formed by the vectors $(\underline{0},\ldots,\underline{v}_\pm,\ldots,\underline{0})$ in the $2|\Lambda|$–dimensional space $(\mathbb{R}^2)^\Lambda$. We suppose that, in the basis $\underline{B}$, the matrix $K(\underline{\psi})_{\xi i,\eta j}$ has vanishing matrix elements except for $K_{\xi+,\eta-}(\underline{\psi}) \stackrel{def}{=} k_{\xi\eta,+}(\underline{\psi})$ and $K_{\xi-,\eta+}(\underline{\psi}) \stackrel{def}{=} k_{\xi\eta,-}(\underline{\psi})$. The matrix Γ will be supposed to have zero matrix elements except for $\Gamma_{\xi\alpha,\eta\alpha}(\underline{\psi}) = \gamma_{\xi\eta,\alpha}$, $\alpha=\pm$. This implies that

e10.4.4

$$\begin{aligned}(K(\underline{\psi})\underline{w})_{\xi,\pm} &= \sum_{\eta} k_{\xi\eta,\pm}(\underline{\psi})\underline{w}_{\eta,\mp}\\ (\Gamma(\underline{\psi})\underline{w})_{\xi,\pm} &= \sum_{\eta} \gamma_{\xi\eta,\pm}(\underline{\psi})\underline{w}_{\eta,\pm},\end{aligned} \tag{10.4.4}$$

i.e. for each ξ,η, Γ is an ε–dependent diagonal matrix on the basis $\underline{v}_\pm$ of the eigenvectors of $\mathcal{S}_0$ and K is an ε–dependent off-diagonal matrix on the same basis. Inspired by the results of Section §10.3 we shall also try to show that $\widetilde{H}(\underline{\psi})$ be expressible as

e10.4.5

$$\begin{aligned}\underline{h}_\xi(\underline{\psi}) &= \sum_{X\ni\xi} \Phi_X(\underline{\psi}_X),\\ (K(\underline{\psi})\underline{w})_{\xi,\pm} &= \sum_{\xi'}\sum_{X\ni\xi,\,\xi'} \gamma_{X,\pm}(\underline{\psi}_X)(\underline{w})_{\xi',\pm}\\ (K(\underline{\psi})\underline{w})_{\xi,\pm} &= \sum_{\xi'}\sum_{X\ni\xi,\,\xi'} k_{X,\pm}(\underline{\psi}_X)(\underline{w})_{\xi',\mp},\end{aligned} \tag{10.4.5}$$

where, of course, we have already determined $\underline{h}$ in Section §10.2. The equations that should be obeyed by $\widetilde{H}$ become, cf. (10.3.9),

e10.4.6

$$\begin{aligned}&(\mathcal{S}_0 K(\underline{\psi}) - K(\mathcal{S}_0\underline{\psi})\mathcal{S}_0)_{\xi i,\eta j} = -\varepsilon\partial_{\varphi_{\eta,j}} f_i(\underline{\varphi}_{nn(\xi)}) -\\ &- \varepsilon\partial_{\underline{\varphi}_{\rho,s}} f_i(\underline{\varphi}_{nn(\xi)})K(\underline{\psi})_{\rho s,\eta j} + \Gamma(\underline{\psi})_{\xi i,\eta j} + (K(\mathcal{S}_0\underline{\psi})\Gamma(\underline{\psi}))_{\xi i,\eta j},\end{aligned} \tag{10.4.6}$$

where $\underline{\varphi} = \underline{\psi} + \underline{h}(\underline{\psi})$ and summation over the repeated indices ρ, s is understood. The labels i,j have values $\pm$ as we imagine that the equation is written in the basis $\underline{B}$.

The equations can be solved by writing them by components; defining

e10.4.7

$$\begin{aligned}&\underline{\psi}^{(\alpha)} \stackrel{def}{=} \mathcal{S}_0^{-(1-\alpha)/2}\underline{\psi}, \quad \underline{\psi}'^{(\alpha)} = \mathcal{S}_0^{(1+\alpha)/2}\underline{\psi}, \qquad \alpha = \pm,\\ &\underline{\varphi}^{(\alpha)} = \mathcal{S}_\varepsilon^{-(1-\alpha)/2}\underline{\varphi}, \qquad\qquad\qquad\qquad\qquad \alpha = \pm,\end{aligned} \tag{10.4.7}$$

one finds, for $\alpha = \pm, \beta = -\alpha$,

e10.4.8

$$\begin{aligned}\gamma_{\xi\eta,\alpha}(\underline{\psi}) =& \varepsilon\partial_{\underline{\varphi}_{\eta,\alpha}} f_\alpha(\underline{\varphi}_{nn(\xi)}) + \varepsilon\partial_{\underline{\varphi}_{\rho,\beta}} f_\alpha(\underline{\varphi}_{nn(\rho)})k_{\rho\eta,\beta}(\underline{\psi}),\\ k_{\xi\eta,\alpha}(\underline{\psi}) =& -\lambda^2 k_{\xi\eta,\alpha}(\mathcal{S}_0^\alpha\underline{\psi}) - \alpha\lambda\big[\varepsilon\partial_{\varphi_{\eta,\beta}} f_\alpha(\underline{\varphi}^{(\alpha)}_{nn(\xi)}) +\\ &+ k_{\xi\rho,\alpha}(\underline{\psi}'^{(\alpha)})\gamma_{\rho\eta,\beta}(\underline{\psi}^{(\alpha)}) +\\ &+ \varepsilon k_{\rho\eta,\beta}(\underline{\psi}^{(\alpha)})\partial_{\varphi_{\rho,\beta}} f_\alpha(\underline{\varphi}^{(\alpha)}_{nn(\xi)})\big],\end{aligned} \tag{10.4.8}$$

where $\lambda = \lambda_+^{-1} = -\lambda_- = (\sqrt{5}-1)/2 < 1$. Here the derivatives of $\underline{f}$ are meant to be performed with respect to the argument $\underline{\varphi}$ of $\underline{f}$ (*i.e.* to be precise one should everywhere write $(\partial_{\underline{\varphi}_{\xi,\alpha}} f_{\alpha'})$ rather than writing $\partial_{\underline{\varphi}_{\xi,\alpha}} f_{\alpha'}$ without the parentheses). This has the consequence that values $\gamma_{\xi\eta}, k_{\xi\eta}$ will be of order $O(\varepsilon^{|\xi-\eta|})$, because the recursion for γ, k involves only nearest neighbors.

Proceeding as in Section §10.3 we can interpret (10.4.8) in terms of suitable tree graphs by simply adding a few more labels to the graphs in Fig. (10.3.1) and we obtain the following result.

P10.4.1 **(10.4.1) Proposition:** (Stable and unstable planes for lattices of cat maps)
Given a nearest neighbors interaction $\underline{f}$ as above and $0<\beta<1$ there exist $\varepsilon_0(\beta)>0$, $C(\beta)<\infty$ and $\kappa(\beta)>0$ as in proposition (10.2.1), and functions Φ_X with values in $(\mathbb{T}^2)^\Lambda$ and $\gamma_{X,\alpha}$, $k_{X,\alpha}$ with values in the $2|\Lambda|\times 2|\Lambda|$ matrices such that defining $\underline{h}$, K and Γ by (10.4.5) and therefore $\widetilde{H}$ by (10.4.2) the following results hold.
(i) Equation (10.4.3) holds, i.e. $\widetilde{H}$ conjugates $\widetilde{\mathcal{S}}_\varepsilon$ with $\widetilde{\mathcal{S}}_0$.
(ii) The functions Φ_X verify the bounds in (10.2.16) and the functions $\gamma_{X,\alpha}$ and $k_{X,\alpha}$ verify, if $\kappa_{X,\alpha}$ is either $\gamma_{X,\alpha}$ or $k_{X,\alpha}$:

e10.4.9
$$\max_{\substack{|\varepsilon|\le\varepsilon_0(\beta)\\ \underline{\psi}_X}} |\kappa_{X,\alpha}(\underline{\psi}_X)| < B(\beta)\Big(\frac{|\varepsilon|}{2\varepsilon_0(\beta)}\Big)^{\frac{\delta(X)}{2d}}, \tag{10.4.9}$$
$$\max_{|\varepsilon|\le\varepsilon_0(\beta),\underline{\psi}_{X\setminus\xi}} |\kappa_X(\underline{\psi}_X)-\kappa_X(\underline{\psi}'_X)| < B(\beta)\Big(\frac{|\varepsilon|}{2\varepsilon_0(\beta)}\Big)^{\frac{\delta(X)}{2d}}\,|\underline{\psi}_\xi-\underline{\psi}'_\xi|^\beta,$$

where $\underline{\psi}_X,\underline{\psi}'_X$ differ only in the site $\xi\in X$ and $\delta(X)$ denotes the tree length of the set X.

It is now immediate to define a basis of tangent vectors to the unstable manifold at a point $\underline{\varphi}=\underline{\psi}+\underline{h}(\underline{\psi})$. If $\underline{v}_{\xi,+}=(\underline{v}_+\delta_{\xi'\xi})_{\xi'\in\Lambda}$ is, as ξ varies in Λ, a basis of vectors for the unstable manifold of the unperturbed system then a basis for the perturbed system will be given by the $|\Lambda|$ vectors

e10.4.10
$$(\underline{w}_{\xi,+}(\underline{\psi}))_{\xi'}=\underline{v}_+\delta_{\xi\xi'}+(K(\underline{\psi})\underline{v}_{\xi,+})_{\xi'},\qquad \xi\in\Lambda. \tag{10.4.10}$$

The situation is similar to the one discussed in Section §10.3: a Markovian pavement is constructed as the H–image of a fixed Markovian pavement for $\mathcal{S}_0$. The latter is simply obtained by fixing a generating Markov pavement $\mathcal{P}$ for the simple Arnold's cat map (see Section §4.2) and then defining the pavement $\mathcal{P}^\Lambda$ obtained as "product" of copies of the pavement $\mathcal{P}$ on each factor of the product $(\mathbb{T}^2)^\Lambda$. If $\underline{\varphi}=\underline{\psi}+\underline{h}(\underline{\psi})$ then $\underline{\varphi}$ and $\underline{\psi}$ have symbolic histories $X_\varepsilon(\underline{\varphi})$ and $X_0(\underline{\psi})$ which coincide, by construction:

e10.4.11
$$\underline{\varphi}=X_\varepsilon(\underline{\sigma}),\qquad \underline{\psi}=X_0(\underline{\sigma}),\qquad \underline{\varphi}=\underline{\psi}+\underline{h}(\underline{\psi}). \tag{10.4.11}$$

Therefore it is now interesting to evaluate the volume of the parallelepiped spanned by the $|\Lambda|$ vectors $\underline{w}_{\xi,+}(\underline{\psi})$ as ξ varies in Λ. The latter is

e10.4.12
$$D(\underline{\psi})=|\mathrm{Vol}(\underline{w}_{\xi_1,+}(\underline{\psi}),\ldots,\underline{w}_{\xi_{|\Lambda|},+}(\underline{\psi}))|=\det\Big((1+K_{++})(1+K^T_{++})\Big)^{\frac12}, \tag{10.4.12}$$

where K_{++} denotes the linear operator $K(\underline{\psi})$ regarded as a map between the linear space spanned by the vectors $\underline{v}_{\xi,+}$, *i.e.* the tangent plane P_+ to the unstable manifold of $\mathcal{S}^0$ and its image $1+K(\underline{\psi}))P_+$, and K^T_{++} denotes its adjoint.

The volume of the image under $\widetilde{\mathcal{S}}_\varepsilon$ of the latter plane P_+ is by (10.4.3) the volume $D'(\underline{\psi})$ of the parallelepiped spanned by the vectors $\underline{w}'_{\xi,+} \stackrel{def}{=} (1+$
N10.4.1 $K(\mathcal{S}_0\underline{\psi}))(\lambda_+ + \Gamma_+(\underline{\psi}))\underline{v}_{\xi,+}$, hence[1]

e10.4.13
$$\begin{aligned} D'(\underline{\psi}) = &|\mathrm{Vol}(\underline{w}'_{\xi_1,+}(\underline{\psi}),\ldots,\underline{w}'_{\xi_{|\Lambda|},+}(\underline{\psi}))| = \\ &= D(\mathcal{S}_0\underline{\psi})\det\Big((\lambda_+ + \Gamma^T_+(\underline{\psi}))(\lambda_+ + \Gamma_+(\underline{\psi}))\Big)^{\frac12}, \end{aligned} \tag{10.4.13}$$

where $\Gamma^T_+(\underline{\psi})$ is the transpose of $\Gamma_+(\underline{\psi})$ defined as the restriction to the plane P_+ of the matrix $\Gamma(\underline{\psi})$ (quite simple as the latter matrix is $\Gamma_{\xi\alpha,\eta\beta} = \Gamma_{\xi,\eta,\alpha}\delta_{\alpha\beta}$ (*i.e.* diagonal in the local label $\alpha=\pm$) so that $\mathcal{S}_0+\Gamma^T$ is a $|\Lambda|\times|\Lambda|$ matrix. The matrix $\widetilde{M} = ((1+\lambda_+^{-1}\Gamma^T)((1+\lambda_+^{-1}\Gamma))^{\frac12}$ has the form, as it is implied by (10.4.9),

e10.4.14
$$\widetilde{M}_{\xi,\xi'}(\underline{\psi}) = \sum_{X\subset\Lambda\,:\,\xi,\xi'\in X} \widetilde{m}_X(\underline{\psi}), \tag{10.4.14}$$

with the functions $m_X(\underline{\psi})$ admitting the bounds (10.4.9) with k_X replaced with m_X.

By construction the differential of $\mathcal{S}_\varepsilon(\underline{\psi}+\underline{h}(\underline{\psi})))$ on the bases generated by $\underline{w}_{\xi,+}(\underline{\psi})$ and $\underline{w}_{\xi,+}(\mathcal{S}_0\underline{\psi})$ is given by $\lambda_+ + \Gamma_+(\underline{\psi})$ so that the expansion coefficient of the surface area of the unstable manifold is given by

e10.4.15
$$\lambda_+^{|\Lambda|}\frac{D(\mathcal{S}_0\underline{\psi})}{D(\underline{\psi})}D_0(\underline{\psi}). \tag{10.4.15}$$

where $D_0(\underline{\psi}) = \det\widetilde{M}(\underline{\psi}) \equiv \det(1+\widetilde{\Gamma}_+(\underline{\psi}))$, and $\widetilde{\Gamma}_+(\underline{\psi})$ can be written in a form analogous to (10.4.14).

For the purpose of constructing the SRB distribution we can ignore the "cocycle" ratio $\frac{D(\mathcal{S}_0\underline{\psi})}{D(\underline{\psi})}$ and the potential energy for the SRB distribution is simply $A(\underline{\psi}) = -\log\det(1+\widetilde{\Gamma}_+(\underline{\psi}))$. Equations (10.4.9) imply that the determinant $D_0(\underline{\psi})$ can be expressed as

e10.4.16
$$\exp\Big(\sum_{X\subset\Lambda} p_X(\underline{\psi}_X)\Big). \tag{10.4.16}$$

Moreover $p_X(\underline{\psi}_X)$ still verify the bounds (10.4.9) with κ_X replaced by $p_X(\underline{\psi}_X)$. This follows from the form of the matrices K, Γ in (10.4.14) and from the following lemma.

[1] Here we note that the n–dimensional volume of the parallelepiped generated by n vectors $\underline{w}_1,\ldots,\underline{w}_n$ in $\mathbb{R}^m$, $m\ge n$, is the square root of the determinant of the matrix $(\underline{w}_i\cdot\underline{w}_j)$, $i,j=1,\ldots,n$.

L10.4.1 **(10.4.1) Lemma:** (Cluster expansion of a determinant)
Let $m_X(\underline{\psi}_X)$ be a $2|\Lambda|\times 2|\Lambda|$ matrix verifying the bounds (10.4.9) with m_X replacing κ_X. Then if $M=\sum_{X\subset\Lambda} m_X(\underline{\psi}_X)$ there exist scalars $n_X(\underline{\psi}_X)$ such that

e10.4.17
$$\det(1+M)=\exp\Big(\sum_{X\subset\Lambda} n_X(\underline{\psi}_X)\Big), \tag{10.4.17}$$

and the functions $n_X(\underline{\psi}_X)$ verify the same bounds with a different constant B' instead of B.

Remark: To check this property note that the determinant of $1+M$ can be written

e10.4.18
$$\det(1+M)=e^{\mathrm{Tr}\,\log(1+M)}=e^{\sum_{k=1}^\infty \frac{(-1)^{k+1}}{k}\mathrm{Tr}\,M^k}, \tag{10.4.18}$$

and $\mathrm{Tr}\,M^k=\sum_{X_1,\ldots,X_k}\mathrm{Tr}(m_{X_1}\cdots m_{X_k})$ has the desired form provided one collects together all terms whose X_j's form a connected set X and consider the result as a contribution to the value of n_X. The sum over k gives no problem because the terms of order k that arise in this way and which have the same X have size bounded at least by $(B''|\varepsilon|)^k$ for a suitable B'' because of the bounds (10.4.9).

Note that the sum in the exponent (10.4.17) is to be expected to be of size of the order of $|\Lambda|$ in spite of the bounds on n_X. And the matrix M is also to be expected to have large size. Since the (10.4.18) is a power series expansion one may be worried that the expression (10.4.17) is only formal and that it is affected by convergence problems. Imagine that instead of $\det(1+M)$ we consider $\det(1+\vartheta M)$ with ϑ a parameter. Then for ϑ small enough (at Λ fixed) the above formal calculation applied to $\det(1+\vartheta M)$ is certainly correct and n_X will become ϑ–dependent remaining bounded by $2^{\delta(X)}$ times the same quantities in (10.4.9) (with a B' replacing B) not only for ϑ small but also for $|\vartheta|\le 2$ and complex. Hence the relation (10.4.17) equates a polynomial in ϑ to the exponential of a function which is analytic for $|\vartheta|<2$: therefore setting $\vartheta=1$ we see that (10.4.17) is rigorously established.

We now have all the ingredients to discuss the SRB distribution for the coupled system. If $\mathcal{P}=\{P_1,\ldots,P_n\}$ is a Markov pavement for the map S_0 then we can construct a Markov partition for $\mathcal{S}_0$ simply considering, for every $\underline{\tau}\in\{1,\ldots,n\}^{\mathbb{Z}^d}$, the square $Q_{\underline{\tau}}=\times_{\xi\in\mathbb{Z}^d}P_{\underline{\tau}_\xi}$. Associated with this Markov partition $\mathcal{Q}$ there is a symbolic representation $X(\underline{\sigma})$ where now $\underline{\sigma}$ can be naturally thought as a point in $\{0,\ldots,n\}^{\mathbb{Z}^{d+1}}$. We indicate the coordinate on $\mathbb{Z}^{d+1}$ with (ξ,t) and call the ξ coordinates *spatial* or *horizontal* and the t coordinate *temporal* or *vertical*. As in the single map case the pavement $H(\mathcal{Q})$ is a Markov pavement for $\mathcal{S}_\varepsilon$ with a symbolic representation given by $X_\varepsilon(\underline{\sigma})=H(X(\underline{\sigma}))$. As an energy for the SRB distribution we can use $\lambda_u(\underline{\psi})$ given by (10.4.15). Note that because of proposition (6.4.1) we can use also $\lambda_u(\underline{\psi})=D_0(\underline{\psi})$.

We obtain the following key result (for more elementary statements, see[BK95], or alternative ones, see [JP99], can be found in the literature; see also problems [10.4.1], [10.4.2]).

$P10.4.2$ **(10.4.2) Proposition:** (SRB distribution potential for lattices of cat maps)
Fixed Λ for all $\beta \in (0,1)$ there exist a Λ–independent constant $\varepsilon_0(\beta) > 0$ and a Markovian pavement $\mathcal{Q} = \{Q_{\underline{\tau}}\}_{\tau \in \mathbb{Z}^d}$ of Ω such that, for $|\varepsilon| < \varepsilon_0(\beta)$, the expansion rate $\lambda_u(\underline{\sigma})$ of $\mathcal{S}_\varepsilon$ along the unstable manifold of a point $x = X_\varepsilon(\underline{\sigma})$ that has a symbolic representation $\underline{\sigma}$ is Hölder continuous in $\underline{\sigma}$ and has the form

$e10.4.19$
$$\lambda_u(\underline{\sigma}) = \sum_{X \in \Lambda} \Phi_X(\underline{\sigma}_X), \tag{10.4.19}$$

where X are sets in $\mathbb{Z}^{d+1}$ and $\Phi_X(\underline{\sigma}_X)$ is holomorphic in ε in the disk $|\varepsilon| < \varepsilon_0(\beta)$. The potential Φ verifies the bound

$e10.4.20$
$$|\Phi_X(\underline{\sigma}_X)| < C\Big(\frac{|\varepsilon|}{\varepsilon_0(\beta)}\Big)^{\frac12\delta_\perp(X)+\frac12 n(X)} e^{-\kappa\delta_\parallel(X)}, \tag{10.4.20}$$

*where $\delta_\parallel(X), \delta_\perp(X)$ denote the tree length of the projection of the set X on the horizontal plane $t=0$ and, respectively, the sum of the tree lengths of the intersections of X with the vertical lines; $n(X)$ denotes the number of the timelike intervals whose union is X (*cf. *remarks to definition (7.3.3) and equation (7.3.26)). The constant C is Λ–independent. Furthermore Φ_X is translation invariant. Note that the $\frac12$ in the exponent of (10.4.20) arises because by construction the number $n(X)$ of vertical intervals building X is necessarily $n(X) \ge \delta_\perp(X)$ so that the natural bound would be the stronger one with $\frac{|\varepsilon|}{\varepsilon_0(\beta)}\Big)^{\frac12\delta_\perp(X)+\frac12 n(X)}$ replaced by $\frac{|\varepsilon|}{\varepsilon_0(\beta)}\Big)^{\delta_\perp(X)}$.*
The SRB distribution μ_ε is a Gibbs state for a system on a $(d+1)$-dimensional lattice with a nearest neighbor hard core interaction in the direction of "time" and a longer range exponentially decreasing "many-body" potential Φ with $||\Phi||_\kappa$, cf. *(7.2.6), small with ε.*

Proof: For simplicity we shall give the proof by supposing that the lattice (*i.e.* spatial) dimension is $d=1$.
In the present case the representation in (10.2.8) yields a representation of $\underline{h}$ with the desired properties.
One starts from (10.2.8) in order to obtain a representation of $\underline{h}$ in terms of potentials. Proceeding as in the proof of propositions (10.4.1) and lemma (10.4.1) one represents, with the same technique, the conjugating map $\underline{h}(\underline{\psi})$ (see below), the stable and unstable planes which split the tangent plane, cf. proposition (4.2.1), then the determinant of the map along the unstable manifold and then its logarithm. At each step the above quantities are expressed as functions of the $d+1$–dimensional symbolic representations $\underline{\sigma}$ of the points $\underline{\psi}$ in the form analogous to (10.4.19) together with bounds like (10.4.20). We discuss only the representation of the map $\underline{h}$ leaving the

remaining analogous representations of the stable and unstable planes and of the determinant of the map along the unstable manifold.

Consider a single tree: its value is given by (10.2.8). We represent the quantity $\prod_j \partial_{(\alpha_{v_j},\xi_{v_j})} f_{\alpha_v}(\underline{\psi}_{nn(\xi_v)})$ via the symbols $\underline{\sigma}_{\xi,s}$ as a sum of "potentials" $\Phi_v^{\bf x}(\{\sigma_{\xi,t}\}) = \sum_{t=0}^\infty \varphi_t^{\bf x}(\{\sigma_{\xi,s}\}_{\xi\in nn(\xi_v),|s|<t})$ by means of the telescopic method used several times (*e.g.* to derive (4.3.8), see (4.3.10)). Here $bfx = \{(\alpha_i,\xi_i)\}$ indicates the derivative of the function f we are considering. The potentials φ will be bounded by $C_v e^{-\kappa t}$ where κ can be taken $\frac12 \log\lambda^{-1}$ (because the symbols sequences determine the corresponding points ψ "at rate" λ in the cat map we consider) and C_v is a constant.

Therefore we can represent $\lambda_{\alpha_v}^{-|p_v+1|\alpha_v} \prod_{j=1}^{s_v} \partial_{(\alpha_{v_j},\xi_{v_j})} f(S_0^{p(v)} \underline{\psi}_{nn(\xi_v)})$ as a sum of

e10.4.21
$$\lambda_{\alpha_v}^{-|p_v+1|\alpha_v} \varphi_{t_v}^{\bf x}(\{\sigma_{\xi,s}\}_{\xi\in nn(\xi_v),|s-p(v)|<t_v}) \tag{10.4.21}$$

over t_v. Repeating the same considerations for the other nodes v of the tree we represent the value of the tree ϑ of order k in (10.2.8) as a product of factors like (10.4.21) and, after properly multiplying it by ε^k, we can interpret it as a contribution to the total potential for the representation (10.2.10) of the conjugating function relative to the space–time set

e10.4.22
$$X = (\{\xi_{v_0}\}\times\{0\}) \cup \cup_{v\in V(\vartheta)\setminus v_0}\big(nn(\xi_v)\times[p(v)-t_v, p(v)+t_v]\big), \tag{10.4.22}$$

if v_0 is the first node of the tree ϑ. The contribution is bounded by $\varepsilon^k C^k \lambda^{-\sum_v (t_v+p(v))}$ for a suitable $C>0$ (note that the $\sup_v C_v$ can be bounded uniformly in all the labels attached to the node v because f contains only finitely many Fourier modes). Although this might seem awkward, it happens that $(\{\xi_{v_0}\}\times\{0\}) \in X$, with the above definition, but $\Phi_X(\underline{\sigma}_X)$ does not necessarily depend on $\sigma_{(\xi_0,0)}$.

Since $k \ge c\delta_\parallel(X)$ because the interaction involves only nearest neighbors this is a bound of the type (10.4.20). The convergence of the summation over the tree values implies that the bound holds, with different constants, also for the total potential Φ_X. Note that the set X has a rather arbitrary shape, unlike the simpler shapes that are considered in [BS88], [JM96], [BK94], [BK96], [BK97] and [JP99].

The expansion illustrated in Fig.(10.3.1) shows that the same representation can be given to the planes tangent to the unstable (and to the stable) manifolds. And continuing with the same arguments one gets, from (10.4.12) and from (10.4.18), the representation of $\lambda_u(\underline{\sigma})$ in the form (10.4.20).

This leads (after decimation to eliminate the hard core in the time direction by proposition (7.3.3) and in the remarks preceding and following it) to the analyticity result. We omit further details (for which we refer to [BFG03]). ■

Remarks: (1) The potentials Φ_X in (10.4.19) are "space–time" potentials because X is now a subset $\mathbb{Z}^{d+1}$.

(2) The representation of $\underline{h}$ in terms of potentials is not unique: proceeding more naively one obtains a potential in which the sets X are space–time "rectangles" which are therefore simpler: however the potentials decay less fast, essentially with the diameter rather than with the tree length (see problem [10.4.2]). This is not really useful if one looks for analyticity results, but it is already sufficient to get smoothness results ([JM96], [BK95], [BK96] and [BK97]).
(3) In our case aside from the hard core the unperturbed potential is exactly 0 so that the potentials can be made as small as wished and with range as short as wished, as shown by the bounds (10.4.9),(10.4.20), by taking ε small. Nevertheless we cannot apply the results of Section §7.2 immediately. In fact our potentials still contain a hard core so that proposition (7.2.3) cannot be applied. In absence of hard cores, however, it would solve the problem.
(4) Therefore one has first to eliminate the hard core by a decimation procedure. In fact by the above remark (3) the assumptions of proposition (7.3.3) can be satisfied with κ growing arbitrarily large for $\varepsilon \to 0$ (one can in fact take $\kappa = O(\log\log \varepsilon^{-1})$). The decimation argument, *i.e.* the proposition (7.3.3), implies several consequences: we explicitly list some among them because they are the conclusion of the theory of lattices of Arnold's cat maps envisaged in this book.

C10.4.1 **(10.4.1) Corollary:** (Regularity of SRB distributions) *Let F and G be two Hölder continuous "observables" (i.e. functions on $\Omega = (\mathbb{T}^2)^\Lambda$) and suppose that they depend on $\underline{\varphi}$ only through the $\underline{\varphi}_\xi$ with ξ in a finite region V. If μ_ε^Λ denotes the SRB distribution for $\mathcal{S}_\varepsilon$ the following properties hold.*

(i) The expectation values $\mu_\varepsilon^\Lambda(F)$ is well defined and Hölder continuous in ε for $|\varepsilon|$ small enough, uniformly in Λ, V. Moreover the expectation values $\mu_\varepsilon^\Lambda(F\circ H^{-1})$ is analytic in ε for $|\varepsilon|$ small enough, uniformly in Λ, V.

(ii) If F is an analytic observable the expectation values $\mu_\varepsilon^\Lambda(F)$ is analytic in ε for $|\varepsilon| \le \varepsilon_F$

(iii) Any two functions F, G mix at an exponential rate, i.e. there exists two positive constants κ and ℓ_0, depending on F and G, such that, if $\mu_\varepsilon^\Lambda((\mathcal{S}_\varepsilon^n F)G) \stackrel{def}{=} \int \mu_\varepsilon^\Lambda(d\underline{\varphi})F(\mathcal{S}_\varepsilon^n\underline{\varphi})G(\underline{\varphi})$ and $\|F\|, \|G\|$ denote the maximum modulus of F, G, one has

e10.4.23
$$|\mu_\varepsilon^\Lambda((\mathcal{S}_\varepsilon^n F)G) - \mu_\varepsilon^\Lambda(F)\mu_\varepsilon^\Lambda(G)| \le \|F\|\,\|G\| e^{-\kappa_{F,G}(n-\ell_0)} \tag{10.4.23}$$

for all Λ, V and, therefore, for the limit μ_ε as $\Lambda\to\infty$ of μ_ε^Λ.

(iv) The volume distribution $\mu_0(d\underline{\varphi}) = d\underline{\varphi}/(2\pi)^2$ "mixes with the SRB distribution" exponentially fast the functions F, G in the sense that $\mu_0((\mathcal{S}_\varepsilon^n F)G) \stackrel{def}{=} \int \mu_0(d\underline{\varphi})F(\mathcal{S}_\varepsilon^n\underline{\varphi})G(\underline{\varphi})$ verifies, for all Λ,

e10.4.24
$$|\mu_0((\mathcal{S}_\varepsilon^n F)G) - \mu_\varepsilon^\Lambda(F)\mu_0(G)| \le \|F\|\,\|G\| e^{-\kappa_{F,G}(n-\ell_0)}, \qquad n>0. \tag{10.4.24}$$

For $n < 0$, in general, (10.4.23) and (10.4.24) hold with a different distribution μ'_ε, which is the SRB distribution for S_ε^{-1} (which essentially has the same properties as μ_ε).

Remarks: (1) One could make the statements above without introducing the constant ℓ_0 and writing (10.4.23) and (10.4.24) with an extra positive constant B in front of the right hand sides and with no ℓ_0 in the exponent, provided that n is chosen large enough. Of course the two formulations are quite equivalent.
(2) Statement (iii) requires a new analysis based on the representation for the volume distribution μ_0 for Anosov systems discussed in Section §4.3. We saw, in fact, that in a single Anosov system the SRB distribution μ and the volume μ_0 are related because the restriction of μ_0 to the functions depending only on the symbols with time label ≥ 0 is absolutely continuous with respect to the restriction of μ to the same functions. For lattices of Anosov systems this is still true but the ratio $d\mu_0/d\mu$ is not uniformly finite, away from 0 and ∞, in the spatial size $|\Lambda|$ of the lattice $\Lambda \subset \mathbb{Z}^d$. However a careful examination of the above proofs implies that if one further restricts the distributions μ, μ_0 to functions which are spatially local in a region V and depend only on the symbols with positive time label then one gets two probability distributions $\mu^{+,V}, \mu_0^{+,V}$ which are absolutely continuous with respect to each other. And $d\mu_0^{+,V}/d\mu^{+,V} = \rho(\underline{\sigma})$ is a Hölder continuous function so that the statement (iii) follows from (ii).
(3) The study the SRB distribution for $\mathcal{S}_\varepsilon^{-1}$ is not immediate due to the fact that $\mathcal{S}_\varepsilon^{-1}$ in general is not given by a nearest neighbor perturbation of $\mathcal{S}_0^{-1}$. A very simple way to solve this problem is to observe that the expansion rate on the unstable manifold for $\mathcal{S}_\varepsilon^{-1}$ is the inverse of the contraction rate on the stable manifold for $\mathcal{S}_\varepsilon$, that has a representation similar to the one discussed in the proof of proposition (10.4.2). Moreover the Markov partition and the symbolic code for $\mathcal{S}_\varepsilon^{-1}$ are strictly related to those for for $\mathcal{S}_\varepsilon$. One can also prove this directly constructing $\mathcal{S}_\varepsilon^{-1}$ as a perturbation of $\mathcal{S}_0^{-1}$ with a perturbation that still decay fast enough in space and time, see problem [10.4.3].

In the above analysis the potentials $m_X(\underline{\psi}_X)$ do depend on Λ. However from the trees expansion formulae we see that the dependence of Φ_X, G_X and therefore all the other potentials we have introduced, in particular n_X, consists of a term which is Λ independent as soon as $\Lambda \supset X$ plus a small correction that while verifying uniformly in Λ the general bounds like (10.4.9) contains corrections of size of order $e^{-\mathrm{const}\,L}$. In our situation in which Gibbs states depend continuously on the potential this implies not only the mentioned *existence of the thermodynamic limit* for the SRB distribution but also a wealth of results that can be derived from the well established theory of Gibbs states with weak coupling. Here we mention only the following.

P10.4.3 **(10.4.3) Proposition:** (Space-time chaos in cat maps lattices) *Let μ_ε^Λ be* *the SRB distribution for the lattice of Arnold's cat maps considered above. For all $F(\underline{\varphi})$ defined on $(\mathbb{T}^2)^\Lambda$ which are analytic and dependent only on the microstates $\underline{\varphi}_\xi$ with $\xi\in\Lambda_0$ where Λ_0 is a finite region, the limit*

e10.4.25
$$\mu_\varepsilon(F)=\lim_{\Lambda\to\infty}\mu_\varepsilon^\Lambda(F) \tag{10.4.25}$$

exists and is analytic in ε for $|\varepsilon|<\varepsilon_F$. Functions F which depend only on the microstates $\underline{\varphi}_\xi$ with ξ contained in a given finite region are called spatially local.
(ii) If F,G are two smooth enough (e.g. Hölder continuous with some exponent $\beta>0$) functions which are local in space then

e10.4.26
$$\mu_\varepsilon((\tau^\xi\mathcal{S}_\varepsilon^tF)G)\xrightarrow[|\xi|+|t|\to\infty]{}\mu_\varepsilon(F)\mu_\varepsilon(G), \tag{10.4.26}$$

where τ^ξ denotes the lattice translation by $\xi\in\mathbb{Z}^d$ and the limit is on t or on ξ or both.

(iii) If F,G are two smooth enough (e.g. Hölder continuous with some exponent $\beta>0$) functions which are local in space then

e10.4.27
$$\mu_0((\mathcal{S}_\varepsilon^tF)G)\xrightarrow[t\to\infty]{}\mu_\varepsilon(F)\mu_0(G), \tag{10.4.27}$$

where τ^ξ denotes the lattice translation by $\xi\in\mathbb{Z}^d$.
(iv) If F is a smooth enough observable (e.g. Hölder continuous with some exponent $\beta>0$) which is local in space then

e10.4.28
$$\lim_{t\to+\infty}\frac{1}{t}\sum_{j=0}^{t}F(\mathcal{S}_\varepsilon^j\underline{\varphi})=\int F(\underline{\varphi})\mu_\varepsilon(d\underline{\varphi}) \tag{10.4.28}$$

for $\mu_0(d\underline{\varphi})$–almost all $\underline{\varphi}\in(\mathbb{T}^2)^{\mathbb{Z}^d}$.

Remarks: (1) The above result is sometimes read as saying "lattices of Anosov maps weakly coupled via short range interactions" show spatio–temporal chaos.
(2) The method of proof followed here can be made very elementary in the case of weakly interacting lattices $\Lambda\subset\mathbb{Z}^d$ of expansive maps S_0 of the interval like the ones considered in proposition (5.4.1) (strictly expansive and surjective) or of expansive maps of the circle: the theory is very similar but the lattice spin system that they generate is a seminfinite lattice system with spins located on $\Lambda\times\mathbb{Z}_+^d$. The great simplification is that there is no hard core in the spin interactions because of the Markovian assumption of strict surjectivity. In this case proposition (7.4.1) immediately applies and gives smoothness as well as mixing, cf. [BK95] and problem [10.4.1]. It does *not* however show the analyticity that follows from an (equally immediate) application of proposition (7.3.2).
(3) Smoothness in the case of lattices of cat maps can also be proved by a

method close to the one presented here (although it does not yield analyticity), cf. [JP99].

Problems for §10.4

Q10.4.1 **[10.4.1]:** (*Nonanalytic proof of regularity of SRB distributions for lattices of circle maps*, from [BK95].)
Consider a chain of circle maps or of interval maps instead of Arnold cats small perturbation of independent maps S_0 acting on each site variable. Assume the expansiveness for S_0 in the sense of Section §5.4 in the interval case and assume that $S_0\varphi_\xi = 2\varphi_\xi$ for $\xi \in \Lambda$ in the case of circle maps (for simplicity). Check that in this case one can transform the problem of determining a SRB distribution into a problem of a lattice spin system on a semi-infinite lattice $\Lambda \times \mathbb{Z}_+$ without hard core conditions. Check that the general uniqueness and smoothness results of Section §7.4 applies immediately, given the result of proposition (10.4.2), (10.4.20), to obtain the results of corollary (10.4.1) with analyticity replacing smoothness. The advantage of this approach is that it is elementary and one does not need the cluster expansion theory of Chapter VII. (*Hint:* The proposition (7.4.1) applies directly because of lack of hard cores. The seminfinite lattice is due to the fact that the maps in question are described symbolically by semiinfinite sequences of symbols, being not invertible).

Q10.4.2 **[10.4.2]:** (*Alternative potentials*)
Potentials $\Phi_X(\underline{\sigma}_X)$ can be immediately derived from the potentials $n_X(\underline{\psi}_X)$ of (10.4.17) by the method of Section §10.3: for fixed X we consider the sequence $\underline{\sigma}_X$ corresponding to $\underline{\psi}_X$ on the Markov pavement and call $\underline{\sigma}_{X,h}$ the sequence obtained from $\underline{\sigma}_X$ by truncating $\underline{\sigma}_X$ beyond the heights $[-h, h]$ and replacing the deleted $\sigma_{\xi,t}$ with $|t| > h$ by standard compatible sequences as done in the proof of proposition (4.3.1). Check that the potentials thus obtained are not zero only for sets of the form $H_0 \times I$ with $H_0 \subset \mathbb{Z}^d$ and $I \subset \mathbb{Z}$. Check that the bounds on $n_X(\underline{\psi}_X)$, $X \in \mathbb{Z}^d$ and the Hölder continuity of $n_X(\underline{\psi}_X)$ imply bounds on the potential that decays exponentially as $e^{-\kappa\delta(H_0)+|I|}$, *i.e.* if $d = 2$ as the diameters of the sets $H = H_0 \times I$.

Q10.4.3 **[10.4.3]:** (*Tree expansion for* $\mathcal{S}_\varepsilon^{-1}$.)
Show that $\mathcal{S}_\varepsilon^{-1}(\underline{\psi})$ can be written as $\mathcal{S}_0^{-1}(\underline{\psi}) + \varepsilon \underline{g}(\varepsilon, \underline{\psi})$ where

$$g_\xi(\varepsilon, \underline{\psi}) = \sum_{X \ni \xi} \Psi_X(\underline{\psi}_X)$$

where X is a connected set in $\mathbb{Z}^d$ and Ψ_X is of order $\varepsilon^{\frac{\delta(X)}{2d}}$. (*Hint:* write an series expansion for $\underline{g}(\varepsilon, \underline{\psi})$ as a function of ε and use $\mathcal{S}_\varepsilon^{-1} \circ \mathcal{S}_\varepsilon(\underline{\psi}) = \underline{\psi}$ to compute recursively the coefficients. The decay follows from an argument very similar to the one used for H.)

Q10.4.4 **[10.4.4]:** Extend the result of problem [10.2.5] to the construction of the unstable direction and of the expansion rate on the unstable direction.

Q10.4.5 **[10.4.5]:** Refine the result of problem [10.4.3] to show that the symbolic representation $\underline{\tilde{g}}(\sigma)$ of $\underline{g}(\varepsilon, \underline{\psi})$ can be written as

$$\tilde{g}_\xi(\underline{\sigma}) = \sum_{X \ni \xi} \widetilde{\Psi}_X(\underline{\sigma}_X)$$

where now X is a set in $\mathbb{Z}^{d+1}$ such that its projection on the temporal coordinate is connected and the intersection of X with any vertical line is a segment centered at 0. Moreover we have:

$$|\widetilde{\Psi}_X(\underline{\sigma}_X)| \le C e^{-\kappa_0 \delta_\perp(X)} e^{-\kappa_1 \delta_\parallel(X)}.$$

(*Hint:* simply write $\underline{\varphi}$ as a sum of potential in the expansion studied in problem [10.4.3].

Q10.4.6 **[10.4.6]:** Show that this is enough to obtain again estimate (10.4.20) for the potential of the SRB distribution of $\mathcal{S}_\varepsilon^{-1}$.

Bibliographical note §10.4

The theory of Sections §10.2 and §10.4 was initiated by [BS88]. The main technical tool in our approach is the cluster expansion discussed in Chapter VII in the form introduced for the decimation problem in [CO82]. The advantage over the original approaches to spatio–temporal chaos theory started in [BS88], and essentially completed in [JM96] is that here one obtains analyticity instead of infinite differentiability of the SRB distribution. Subsequent papers eliminated technical restrictions present in [JM96], at first at the price of restricting the theory to lattices of coupled expanding maps of the circle (which exhibit the important simplification of having a symbolic dynamics representation without hard cores, see [BK95], [BK96] and [BK97]). Later the restriction to expanding maps has been eliminated, [JP99], [BK96] and [JP99], by further developing the ideas in [JM96]. The approach discussed in Section §10.4, due to [BFG03], goes a little beyond the results in the just quoted papers as it also gives analyticity in the dependence of the Gibbs distribution on the perturbation size. The latter result is discussed in [BFG03] where also the theory of Sections §10.1 and §10.3 were developed. For an example of application of the above results see [Ga99].

§10.5 Isomorphisms

We can ask in which cases two Gibbs states μ and μ', one on $\{0,\ldots,n\}_T^{\mathbb{Z}}$ and the other on $\{0,\ldots,n'\}_T^{\mathbb{Z}}$, with respective potentials Φ and Φ' are isomorphic in the sense that the dynamical systems $(\{0,\ldots,n\}_T^{\mathbb{Z}},\tau,\mu)$ and $(0,\ldots,n')_T^{\mathbb{Z}},\tau,\mu')$ are isomorphic $\bmod\ 0$.

P10.5.1 **(10.5.1) Proposition:** (Isomorphims of Gibbs distributions with fast decreasin potential)
If T and T' are two mixing compatibility matrices, $(n+1)\times(n+1)$ and $(n'+1)\times(n'+1)$ respectively, and if Φ and Φ' are two potentials for $\{0,\ldots,n\}_T^{\mathbb{Z}}$ and for $\{0,\ldots,n'\}_T^{\mathbb{Z}}$ such that

e10.5.1
$$\sum_{X\ni 0}\frac{1+\mathrm{diam}(X)}{|X|}\|\Psi_X\|<+\infty,\qquad \text{if } \Psi=\Phi \text{ or } \Phi', \tag{10.5.1}$$

with $\|\Phi_X\|=\sup_{\underline{\sigma}_X}\|\Phi_X(\underline{\sigma}_X)\|$, then the dynamical systems $(\{0,\ldots,n\}_T^{\mathbb{Z}},\tau,\mu)$ and $(\{0,\ldots,n\}_T^{\mathbb{Z}},\tau,\mu)$ are isomorphic $\bmod\ 0$ *if and only if their entropies are equal: $s(\mu)=s(\mu')$.*

Remarks: (i) The average entropy is therefore a complete invariant for the isomorphisms $\bmod\ 0$ in a rather wide class of dynamical systems.

(ii) The proof of this theorem reaches already its maximal difficulty in the case in which $\Phi = \Phi' = 0, T_{\sigma\sigma'} = T'_{\sigma\sigma} = 1$: this is the case of the Bernoulli schemes.

(iii) The problem of the isomorphisms between Bernoulli schemes is a famous problem solved at the end of the 1960's by Ornstein who also established new techniques to attack and solve the problem of the isomorphism between large classes of dynamical systems. Proposition (10.5.1) gives us an example of a problem that can be solved with the ideas and the techniques of Ornstein.

Ornstein's theorem was preceded by another important result of Sinai that established the weak equivalence mod 0 (see below for a precise definition) between two Bernoulli schemes of equal entropy. It was followed by several deep explicit constructions of the codes that realize the isomorphisms. The first one, due to Monroy and Russo, [MR75], concerned a very special case, but it introduced, in that case, some new ideas: such a construction has been improved by the constructions of Keane and Smorodinski, who, with the use of new and deep ideas, explicitly realized the code of the isomorphism between two isentropic Bernoulli schemes, see [KS79]. Such codes are "constructive" in the sense that it is possible to construct an arbitrarily *prefixed* number of values of the elements of the sequence $\underline{\sigma}'$ image, in the isomorphism in question, of a sequence $\underline{\sigma}$ by making use of an algorithm that can be implemented on a computer, so that it requires a finite time for almost all the sequences $\underline{\sigma}$ (randomly chosen with respect to the measure of one of the two Bernoulli schemes).

From a practical point of view the description of the algorithm of Keane and Smorodinski is certainly the fastest way to achieve the proof of the isomorphism between isoentropic Bernoulli schemes. Nevertheless Ornstein's proof remains a monument and an instrument for whoever wants to study in more detail the abstract and conceptual aspects of the isomorphism theory and of the meaning of entropy. Furthermore it provides us with the means to deal with isomorphism problems between dynamical systems that until now, were not otherwise soluble.

(iv) We shall not discuss here the proof of Ornstein's theorem, and the reader should consult for this purpose the book by Ornstein, [Or74], and then meditate on the codes of Monroy-Russo and Keane-Smorodinski. It seems difficult to present Ornstein's theory in a more compact or simpler way than the one he himself chose and likewise it is difficult to present the constructions of Monroy-Russo and of Kean-Smorodinski without repeating word for word the relatively short and brilliant original works.

For completeness it is convenient to give a precise definition of weak isomorphism mod 0 and to state the theorem of Sinai.

D10.5.1 **(10.5.1) Definition:** (Weak isomorphism)
If $(\{0,\ldots,n\}^{\mathbb{Z}},\tau,\mu)$ *and* $(\{0,\ldots,n'\}^{\mathbb{Z}},\tau,\mu')$ *are two ergodic metric dynamical systems, they will be said to be weakly isomorphic* mod 0 *if there is a measurable partition* $\mathcal{P} = \{P_0,\ldots,P_{n'}\}$ *of* $\{0,\ldots,n\}^{\mathbb{Z}}$ *and a measurable*

partition $\mathcal{P}' = \{P'_0, \ldots, P'_n\}$ *of* $\{0, \ldots, n'\}^{\mathbb{Z}}$ *such that*

e10.5.2
$$\begin{aligned} \mu(C^{0\ldots M}_{\sigma_0\ldots\sigma_M}) &= \mu'(P'^{0\ldots M}_{\sigma_0\ldots\sigma_M}) \qquad \text{for all } M \ge 0, \underline{\sigma} \in \{0, \ldots, n\}^{M+1}, \\ \mu'(C^{0\ldots M}_{\sigma'_0\ldots\sigma'_M}) &= \mu(P^{0\ldots M}_{\sigma'_0\ldots\sigma'_M}) \qquad \text{for all } M \ge 0, \underline{\sigma}' \in \{0, \ldots, n'\}^{M+1}, \end{aligned} \tag{10.5.2}$$

where $P'^{0\ldots M}_{\sigma_0\ldots\sigma_M} = \cap_{j=0}^{M} \tau^{-j} P'_{\sigma_j}$ *and* $P^{0\ldots M}_{\sigma'_0\ldots\sigma'_M} = \cap_{j=0}^{M} \tau^{-j} P_{\sigma'_j}$. *This means that the "two dynamical systems are weakly isomorphic if each contains a copy of the other".*

An important result (due to Sinai) is the following one.

P10.5.2 **(10.5.2) Proposition:** (Sinai's theorem)
Let $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ *and* $(\{0, \ldots n'\}^P\mathbb{Z}, \tau, \mu'\}$ *be two Bernoulli schemes of equal entropy. Then they are weakly isomorphic* mod 0.

We conclude this very brief introduction to the isomorphisms theory with a comment and an interesting definition.
One of the key remarks in Ornstein's theory is the importance of the notion of *finite determination* of a dynamical system $(\{0, \ldots, \}^{\mathbb{Z}}, \tau, \mu)$. From a mathematical point of view the interest of this notion lies in the fact that, as shown by Ornstein, a system that enjoys this property is isomorphic mod 0 to a Bernoulli scheme and *vice versa*. From a"physical" point of view it is a property that has a very interesting interpretation that we illustrate after giving the precise notion of *finite determination* (due to Ornstein, see [Or74]).

D10.5.2 **(10.5.2) Definition:** (Finitely determined systems)
A dynamical system $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ *is* finitely determined *if, given* $\varepsilon > 0$, *there exist* $\delta_\varepsilon, N_\varepsilon$ *such that every other ergodic dynamical system* $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu')$, *for which*

e10.5.3
$$\begin{aligned} &\sum_{\sigma_0\ldots\sigma_{N_\varepsilon}} |\mu(C^{0\ldots N_\varepsilon}_{\sigma_0\ldots\sigma_{N_\varepsilon}}) - \mu'(C^{0\ldots N_\varepsilon}_{\sigma_0\ldots\sigma_{N_\varepsilon}})| < \delta_\varepsilon, \\ &|s(\mu) - s(\mu')| < \varepsilon, \end{aligned} \tag{10.5.3}$$

is such that we can construct a code $I_\varepsilon : \{0, \ldots, n\}^{\mathbb{Z}} \longleftrightarrow \{0, \ldots, n\}^{\mathbb{Z}}$ *such that* $\mu(I_\varepsilon(E)) = \mu'(E)$ *for all* μ'*–measurable set* $E \subset \{0, \ldots, n\}^{\mathbb{Z}}$ *with the property*

e10.5.4
$$\limsup_{N\to\infty} (2N)^{-1} \sum_{i=-N}^{N-1} |I_\varepsilon(\underline{\sigma})_i - \sigma_i|, \le \varepsilon \qquad \mu' - \text{almost everywhere.} \tag{10.5.4}$$

We shall say, briefly, that μ *is finitely determined if* $(\{0, \ldots, n\}^{\mathbb{Z}}, \tau, \mu)$ *is such.*

Remarks: (i) This means that if μ is a finitely determined every measure $\mu' \in \mathcal{M}(\{0, \ldots, n\}^{\mathbb{Z}})$ "close to it in entropy and in distribution" (cf. problem

[3.3.9])) produces "the same set of typical configurations up to errors that occur with a small density in time".

(ii) It is easy to construct ergodic distributions that have preassigned probability distributions for the cylinders of preassigned base length and having, furthermore, preassigned entropy (within a preassigned approximation), (cf. problem [3.3.9]). We then see that if μ is finitely determined such "approximate models" of μ can be used to generate sequences (via random extractions) that differ from those that one would obtain by using μ itself only in a fraction ε of sites. In other words by examining the statistics of short sequences it is possible to infer properties of infinitely long sequences, if such sequences are generated at random with a finitely determined distribution.

(iii) If μ_Φ is a Gibbs state with a potential Φ and if Φ varies in an open region $\Sigma \subset B$ where there are no phase transitions (*i.e.* for all $\Phi \in \Sigma, G(\Phi)$ contains a single element) then, from the variational principle (corollary (6.1.1)), from the convexity and continuity of $P(\Phi)$ and from the interpretation of μ_Φ as tangent plane to the graph of P, it follows that both $s(\mu_\Phi)$ and μ_Φ vary with continuity (with respect to the weak topology for the distributions). This means that as Φ varies in Σ the Gibbs state changes by a small amount in distribution and entropy if Φ changes by little.

Therefore if every Gibbs process μ_Φ with $\Phi \in B$ was finitely determined it would follow, from this observation, that also the typical configurations vary "with continuity": note that for the validity of such a property in general the continuity of μ_Φ in Φ would not be enough if the distributions μ_Φ were not finitely determined.

(iv) it appears difficult to think that there can exist Gibbs states with potential $\Phi \in B$ that are not finitely determined, because of the physical meaning of this property, that emerges from the remarks (iii) and (ii).

Hence we see the interest of proposition (10.5.1) that goes in the direction of a confirming the latter idea. And we also understand the interest of studying what happens in the case in which $\Phi \in B$ but it does not verify (10.5.1).

In reality no examples of Gibbs states that are not finitely determined (with $\Phi \in B$) are known; and in some cases in which one could perhaps have of doubts and that refer to the analogous problems for systems on lattices of dimension ≥ 2 finite determinacy has been established (an unpublished work of Ornstein-Weiss shows the finite determination of the Gibbs state for the 2–dimensional Ising model at the critical point).

(v) The property of finite determination allows us to speak of certain global properties of typical configurations in terms of local properties. It can therefore be used to formulate in a mathematically precise way various notions of physical nature that concern global properties of typical (*i.e.* randomly chosen) configurations. This fact han not until now been really exploited in the Physics literature. An example of an attempt to exploit this idea can be found in [EG73] in connection with a probabilistic interpretation of the "approximate symmetries".

Bibliographical note

Ornstein's theory is exposed in [Sh73] and in the book by D. Ornstein, [Or74]. Proposition (10.5.1) is based on the derivation of certain estimates needed to check a sufficient criterion in order that a dynamical system is isomorphic to a Bernoulli scheme and it can be found in [Ga73], [Le73]. The code of Monroy-Russo is found in [MR75]; the code of Kean-Smorodinski is found in [KM79]. The quoted work D. of Ornstein and B. Weiss is unfortunately still unpublished, [OW74].

.

Appendix: Nonequilibrium thermodynamics ? twentyseven comments

Here one of us (Gi.G.) adds a few (personal) comments on the physical meaning of entropy which appeared in november 2002, stimulated by continuous heated discussions held at Rutgers University in the course of the last few years: involving, among many others, in particular S.Goldstein, J.Lebowitz, D.Ruelle. The twenty-seven comments (marked by a (•)) on the Second Law and nonequilibrium systems are related particularly to the contents of Chapter X and give a very personal (and apparently controversial) outlook on the role of entropy in nonequilibrium thermodynamics. Appropriate extensions of the SRB distributions, cf. definition (6.2.2), seem to be the correct generalization of the equilibrium distributions (Gibbs distributions) to describe the statistics of the motions of *stationary* nonequilibrium states: here the view is held that attempting an extension of the definition of *entropy* to stationary nonequilibrium systems may not be the right direction to proceed. The discussion that follows suggests that the connection between the entropy notion that is useful in ergodic theory and information theory (as employed in this book) may be related to entropy in macroscopic Thermodynamics in a subtle way: and possibly not really identifiable with it. The problem becomes particularly evident when one tries to extend entropy to nonequilibrium Thermodynamics understood as the study of general properties of transformations in which a system evolves through a sequence of *stationary* states (while classical Thermodynamis deals with evolutions through sequences of equilibrium states). A stationary state can be an equilibrium state, when the system is controlled by conservative forces, or a nonequilibrium state, when the system is subject to nonconservative forces and can reach a stationary state only if the work done by the external forces is appropriately dissipated by "thermostats": as in the case of a constant electromotive field on closed conducting wire. In the following discussion one should carefully distinguish between the transient phenomena that occur while a system approaches a stationary state and the phenomena associated with the stationary state itself: the distinction is exemplified by an analysis of the Joule's expanding gas experiment.

Definitions.

The purpose, in the present discussion, is investigating the possibility of an extension of Thermodynamics to systems which are in a stationary state but are subject to the action of conservative and nonconservative positional forces $\underline{f}_{pos}$ and (therefore) also to the action of the forces $\underline{\vartheta}$ necessary to take away the heat thus generated.

We first consider systems for which a finite microscopic mechanical model exists and are, therefore, described by equations of the form

$$m\ddot{\underline{x}} = \underline{f}_{pos}(\underline{x}) + \underline{\vartheta}(\underline{x},\dot{\underline{x}}) \stackrel{def}{=} \underline{F}(\underline{x},\dot{\underline{x}})$$

and $\underline{x}$ is a point in an appropriate finite dimensional phase space (typically of very large dimension). If the force $\underline{f}_{pos}$ is conservative then no thermostat is needed and we suppose $\underline{\vartheta} = \underline{0}$ (for simplicity). In general we call the force law $\underline{\vartheta}$ a *mechanical thermostat.*

(•) A key notion will be the *phase space contraction rate* $\sigma(\dot{\underline{x}},\underline{x})$ which is defined as minus the divergence of the equations of motion:

$$\sigma(\dot{\underline{x}},\underline{x}) = -\sum_{\alpha=1}^{3N} \partial_{\dot{x}_\alpha} F_\alpha(\dot{\underline{x}},\underline{x})$$

(•) An equilibrium state will be a stationary probability distribution given by a density (one says a stationary "*absolutely continuous*" distribution) on the phase space of a system which is subject only to conservative forces. We also identify the distribution with any point which is typical with respect to it: by typical we mean that the time averages of observables evaluated on the trajectory of the point are the same as the averages with respect to the distribution.

(•) We suppose (loosely called an "ergodicity assumption") that the time averages of observables (just of the *few* physically relevant for macroscopic Physics) are computable from any of the (equivalent) statistical ensembles: like the microcanonical ensemble. Hence an equilibrium state is identified with a probability distribution on phase space. A "typical" microscopic configuration, *i.e.* an initial datum in phase space which is not in a set of "unlucky cases" which, however, form a zero volume set and are therefore
NA1 (believed to be)[1] unobservable, will evolve in time so that the time averages of the observables (at least the *few* relevant for macroscopic Physics) are computable by means of the equilibrium state which has the correct values of the macroscopic parameters: *e.g.* in the microcanonical case the energy U and the container volume V. When we speak of properties of a single

[1] This means that "typical" is any initial state chosen with a probability distribution absolutely continuous with respect to the volume on phase space. This is is not to be taken for granted, even though it is very often considered so: here I do not enter into discussing this mysterious assumption (as I have nothing to say). For instance an initial state chosen with a probability distribution which is uniform on the energy surface of energy U of a system that occupies half of a volume V is typical for the gas enclosed in the full box V because the configurations which occupy half the box have positive probability among the ones allowed to occupy the full box: see below. Of course the statistics of the configurations just considered is completely different if there is a physical wall separating the two halves of the box or if it is absent.

(typical) point in phase space, like of its "entropy", we always mean the same property of the equilibrium state for which the datum is a typical one.

Remarks: This already might be controversial: in fact the above (admittedly unconventional) definition of entropy has the following implications. A rarefied gas which initially happens to have all molecules located in the left half of a container, because a separation wall has just been removed, setting the gas in macroscopic motion and out of the previous equilibrium state, will be an initial datum in phase space whose entropy is that of the same gas occupying the entire container and at the same temperature[2] (the difference residing only in the different dynamics that follows the removal the wall in the middle of the container). Since a physicist would apply the Boltzmann equation to describe the evolution, the question arises about which is the place here of Boltzmann's H–function, which is different if evaluated for the initial datum or for a datum into which the initial one evolves after a moderately large time (and in both cases it equals the classical thermodynamic entropy of the initial and final equilibria).

$NA2$

(•) In trying to study nonequilibrium cases a conceptual difficulty must be met: if a system is subject to external non conservative forces then the thermostatting forces will have a non zero divergence and volume in phase space will not be preserved. The key idea (due to Ruelle) [Ru95] is that in this case (extending, and including, the previous case) the nonequilibrium states will be the stationary states which are generated by time averaging of initial states that have a density or more satisfactorily, perhaps, by time averaging on the evolution of initial data that are typical for probability distributions given by some (arbitrary) density. The latter states are called *SRB distributions* (Sinai,Ruelle,Bowen, [Ru95],[Ru99],[Ga00],[Ga02]). Their appearance is natural since there will be *no state* (*i.e.* no probability distribution) which is stationary and at the same time is also given by a density. As in the equilibrium state we shall attribute to each individual point in phase space the macroscopic properties of the stationary state which allows us to compute the averages of macroscopic observables on the motion of the given point.

Remark: Systems that are chaotic in a mathematical sense (*hyperbolic systems*) can be shown, on rather general grounds, to have the property that there is only one SRB distribution (with the correct values of the macroscopic parameters), [Ru95],[Ru99],[Ga02]. Therefore such systems, the only ones for which the above definition has a strict mathematical sharpness, verify an extension of the ergodic hypothesis: adopting the latter definition means believing that the system is *chaotic enough* so that typical initial data generate a unique stationary distribution with several features of the

[2] Assuming the gas to be ideal or the temperature would not be the same: we think here of Joule's experiment. In the following the temperature will have the dimension of energy, *i.e.* we call temperature T what is usually called k_BT with k_B being Boltzmann's constant

SRB distributions for hyperbolic systems: the assumption has been called *chaotic hypothesis*,[GC95], and it represents (in our analysis) the nonequilibrium analogue of the classical ergodic hypothesis.

(•) Given an initial state (a typical point or a distribution on phase space) it might be possible to define a function of it that, at least if the state is evolved with an equation which is a good approximation for very long times, will monotonically increase to a limit value which is the same for almost all data sampled with a distribution with a density on phase space (*i.e.* absolutely continuous). Since Boltzmann's H–function is an example of such a function we shall call H–functions all such functions: among them one should also mention the "*Resibois' H function*" for systems described by Enskog's equation (hard sphere systems) and the "*(Boltzmann) entropy*" for systems in local thermal equilibrium [Re78],[GL03]. We recall that Boltzmann's H–function is defined by a coarse graining of phase space into "macrostates" determined by the occupation numbers $f(p,q)d^3pd^3q$ of phase space cells d^3pd^3q around p,q and by defining $H = -\int f(p,q)\log f(p,q)\, d^3pd^3q$. Clearly here the cells size affects by an additive constant the actual value of H, which therefore should have no significance (at least in the classical mechanics context in which we are working) and only the variations of H can be meaningful. In a rarefied gas the Boltzmann equation applies approximately (and even exactly within the Grad's approximation, [Ga00]): so that in the latter cases the function just defined is a nontrivial example of an H–function. See also [Ru03].

(•) Here we propose that even in the case of rarefied gases it is neither necessary nor useful that the H–function is *identified with the entropy*: we want to consider it as a Lyapunov function whose role is to indicate which will be the final equilibrium state of an initial datum in phase space. This does not change, nor it affects, the importance of Boltzmann's discovery that, in rarefied gases, the H–function can be identified with the physical entropy whenever the latter is defined (*i.e.* in equilibria). Suppose that the initial state is chosen randomly with respect to a Gibbs distribution which is *not* the one that pertains to the given parameters that describe the system (*e.g.* volume V and energy U) but to other values: for instance it is chosen randomly with a Gibbs distribution $\mu_{U,V/2}$ that has the same energy but occupies half of the volume, as in Joule's experiment. Then the initial and final values of the H–function happen to coincide with the physical entropies of the Gibbs states $\mu_{U,V}$ and $\mu_{U,V/2}$ (at least in a rarefied gas) which are
NA3 given by the Gibbs' entropy.[3]

[3] One should note that in the whole Boltzmann's work he has been really concerned with the approach to equilibrium: in our terminology he has been concerned with the problem of determining the stationary equilibrium distribution to which a given initial datum gives rise at large time: restricting consideration only to such cases one could well call the Boltzmann's H–function the "entropy" of the state as it evolves towards equilibrium. And, whatever name we give it, it remains true that H is a measure of the disorder in the system. Our analysis here is intended to say that such an interpretation

(•) The latter property could possibly be used to attempt a definition of entropy for states which are neither equilibrium nor stationary states, [Le93] however such a definition would be useful only in the special situations in which an H-theorem could be proved. That seems effectively to reduce the cases in which the notion would be useful to the ones in which an initial equilibrium state identified by some parameters (like U, V) evolves towards a final one identified by other values of the parameters. And even in such cases it is severely restricted to the rarefied gases evolutions in which the $H-$theorem can be proved: a proposed model independent, universal, extension of the above H–function would have to be proved to have a monotonicity property, at least at a heuristic level and within approximations in which exceedingly long times are involved, to avoid that its assumed monotonicity becomes an *a priori* law of nature.

(•) In general I would think that there will always be a Lyapunov function which describes the evolution of an initial state and is maximal on the stationary state that its evolution will eventually reach: however such a Lyapunov function may not have a universal form (unlike the H–function in the rarefied gases cases) and it may depend on the particular way the system is driven by the external forces. After all the SRB distribution verifies, cf. Section §6.2, a variational principle (Ruelle), [Ru95],[Ru99],[Ga02], which *remarkably* has the same form both in equilibrium and nonequilirium systems and one may imagine that in general it will be possible to define (on a case by case basis, I am afraid) a quantity that, within a good approximation, will tend in a short time to a maximum reached on the eventual stationary state. This picture seems to me simpler than trying to guess a (possibly nonexistent) general definition of a quantity that would play the role of Boltzmann's H.

Entropy creation.

The second law of Thermodynamics, in classical Thermodynamic treatises, states:

It is impossible to construct a device that, operating in a cycle, will produce no effect other than the transfer of heat from a cooler to a hotter body

Of course this will be assumed to be a law of nature (Clausius), [Ze68].

(•) The law implies that one can define an *entropy* function S on all equilibrium states of a given system (characterized in simple bodies by energy U and available volume V) and if an equilibrium state 1 can be transformed into another equilibrium state 2 then

is not tenable when the system evolves towards a stationary nonequilibrium state. It must also to be said that even the H-theorem is *not general* because it applies only to rarefied gases, and even there it is an approximation (in which exceedingly long times are not considered): to extend it to general situations, even when one only deals with approach to equilibrium, is a profound statement which should be substantiated by appropriate arguments.

$$S_2 - S_1 \ge \int_1^2 \frac{đQ}{T}$$

using notations familiar from Thermodynamics: here the integral is over the transformation followed by the system in going from 1 to 2 and the $đQ$ is the heat that the system absorbs at temperature T from the outside reservoirs with which it happens to be in contact. The equality sign holds if the path followed is a reversible one.

(•) One should note that the principle really says that the $\int_1^2 \frac{đQ}{T}$ does not depend on the path followed, if the path is a reversible sequence of equilibrium states; and its maximum value is reached along such paths. Existence of a path connecting 1 with 2 with $S_2 - S_1 < \int_1^2 \frac{đQ}{T}$ would lead to a violation of the second law.

(•) Is there an extension of the $S_2 - S_1 \ge \int_1^2 \frac{đQ}{T}$ relation to nonequilibrium
NA4 Thermodynamics?[4] Since there seems to be no agreement on the definition of entropy of a system which is in a stationary nonequilibrium and since there seems to be no necessity in Physics of such a notion, at least I see none, I shall *not* define entropy of stationary nonequilibrium systems (in fact the analysis that follows indicates that if one really insisted in defining it then its natural value could, perhaps, be $-\infty$!).

(•) In Thermodynamics one interprets $-\int_1^2 \frac{đQ}{T}$ as the *entropy creation* in the process leading from 1 to 2, *be it reversible or not or through intermediate stationary states or not.* The name is chosen because it is thought as the entropy increase of the heat reservoirs with which the system is in contact and which are supposed to be systems in thermal equilibrium: so that their entropy variations are, in principle, well defined because they fall in the domain of equilibrium thermodynamics. In a general transformation from a state 1 to a state 2, both of which are stationary (non)equilibrium states, following a path of (non)equilibrium stationary states $\mu^{(t)}$ and in contact with purely mechanical thermostats one could consider the contribution to the *entropy creation* due to irreversibility in the process leading from 1 to 2 during a time interval $[0, \Theta]$ to be

$$\Delta = c \int_0^\Theta \langle \sigma(\underline{x}, \dot{\underline{x}}) \rangle_{\mu^{(t)}} dt = c \int_0^\Theta \sigma_t \, dt$$

where $\sigma_t \stackrel{def}{=} \langle \sigma(\underline{x}, \dot{\underline{x}}) \rangle_{\mu^{(t)}}$ is the average phase space contraction computed in the state $\mu^{(t)}$. This follows a recently proposed identification of $\sigma(\underline{x}, \dot{\underline{x}})$ as proportional to the *entropy creation rate* (here c is a proportionality constant), [EM90]. The quantity Δ is for mechanical thermostats the analogue of $-\int_1^2 \frac{đQ}{T}$ for the generic phenomenological thermostats characterized by a temperature T.

---

[4] Which in a sense is tantamount of asking "is there a nonequilibrium Thermodynamics?"

(•) In considering macroscopic systems one may imagine situations in which a system is partially thermostatted by mechanical forces for which a model
$NA5$ considered physically reasonable is available[5] and partially by phenomenologically defined "heat reservoirs" characterized by a temperature T and able to cede to the system quantities $\not{d}Q$ of heat ($\not{d}Q$ can have either sign or vanish).

In such more general settings a system in contact with several thermostats of which a few are modeled by mechanical equations and a few others are unspecified and are just assumed to exchange quantities of heat $\not{d}Q$ at temperature T the second principle will be extended as

$$-\Delta + \int_1^2 \frac{\not{d}Q}{T} \le 0$$

assuming that the cyclic path leading from 1 *to* 2 *consists entirely of nonequilibrium stationary states* and that it lasts a time interval $[0,\Theta]$ (*i.e.* 1 and 2 differ only because their times are different). Regarding the external thermostats as thermodynamic equilibrium systems $-\frac{\not{d}Q}{T}$ is the entropy increase of the reservoirs at temperature T and $\Delta = c\int\langle\sigma\rangle_t$ is interpreted as the entropy increase of the mechanical reservoir: if this interpretation is accepted the above relation becomes the ordinary second law for the external reservoirs and could be read as "the entropy of the rest of the universe does not decrease" (because $\Delta - \int_1^2 \frac{\not{d}Q}{T} \ge 0$), where "universe" is not the astronomical Universe but rather the collection of physical systems whose interaction with the system under study cannot be neglected.

If a system is a in a stationary state in which $\sigma_t = \langle\sigma\rangle > 0$ (t–independent) this essentially forces us to say that its entropy, if one insisted in defining it at the time t_0 of observation, could only be $-\int_{-\infty}^{t_0}\langle\sigma\rangle\,dt = -\infty$ as hinted
$NA6$ above.[6]

(•) Since the quantity σ_t is ≥ 0 (Ruelle), [Ru99], under very general conditions (in fact always if the chaotic hypothesis is assumed to hold true) $\Delta \ge 0$ and the proposed extension is compatible with the main consequences of the second law. The constant c will be taken 1 because the factor 1 can be computed by studying the expression of $\sigma(\underline{x},\dot{\underline{x}})$ in special models: at the moment, however, I see no immediate physical implications of the "universality" of this choice of c and for the purposes of what follows c could be any constant, even non universal.

(•) The above implies again that we shall *not* be able to define an entropy function *unless* $\sigma_t \equiv 0$. The latter is the condition under which equilibrium

[5] Typically these are models of friction, as in the Navier–Stokes equation case in which the viscosity plays the role of a thermostat. Or in granular matter where the restitution coefficient in the collisions produces energy dissipation. Another well known example is in Drude's theory of electrical conduction.

[6] If we imagine possible to replace a mechanical thermostat with a phenomenological thermostat at temperature T then the left hand side of the relation above remains unchanged but a part of $-\Delta$ becomes a contribution to the second addend.

Thermodynamics is set up: so that if one studies only transformations from equilibrium states to other equilibrium states it is possible to define not only the creation of entropy but the entropy itself (up to an additive constant). In other words *transformations between equilibrium states play the role for entropy that isochoric transformations play for heat*: if we only consider adiabatic transformations between equilibrium states then heat is a function of state, likewise if we restrict to equilibrium states then entropy is a function of state; for more general (stationary) states and their trasformations neith is a function of state but still it makes sense to talk about their creation.

(•) A number of compatibility questions arise: suppose that the system evolves between 1 and 2 under the action of a thermostat which is modeled by forces that act on the system. For instance we can imagine a container with periodic boundary conditions, we call it a *closed wire*, containing a lattice of obstacles, which we call a *crystal*, and N particles, which we call *electrons*, interacting between each other and with the lattice via hard core interactions (say) and subject also to a constant force, which we call *electromotive force*, of intensity E; furthermore the particles will be subject
NA7 to a thermostat force of Gaussian type[7] (as it is essentially the case in Drude's theory as described in classical electromagnetism treatises, [Be64]) which forces the particles to have an energy $U = u(E)$ which is an assigned function of E. Suppose that the value of E changes in time (very slowly compared to the microscopic time scales) following a profile $E(t)$ as drawn in Fig.1

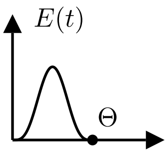


Fig.A1

In this case the wire performs a cycle which is irreversible and the integral $\int_1^2 \frac{đQ}{T}$ is 0 because the system is adiabatically isolated (the thermostat being only of mechanical nature). The entropy variation of the system is defined because the initial and final state are equilibria and, since they are the same, it is 0: but there has been entropy creation $\Delta > 0$.

(•) *It is* (perhaps) *natural to define the "temperature" of a mechanical thermostat* by remarking that in the models studied in the literature it turns out that σ_t is proportional to the work per unit time that the mechanical forces perform, the proportionality constant being in general a function of the point in phase space. Therefore we can call T_0^{-1} the time average of the proportionality constant between σ and the work W per unit time that the mechanical thermostatting forces perform: in this way $\sigma_t = \frac{W}{T_0}$ and $\int_0^\Theta \sigma_t \, dt = \int \frac{đQ_0}{T_0}$ where $đQ_0$ is the total work performed by the mechanical forces *which we can* (naturally) *call the heat absorbed by the mechanical*

[7] This is not the appropriate place to remind the Gauss' least constraint principle: it can be easily found in the literature, [Ga00].

reservoir.

(•) If one imagines that the above conducting wire model at the same time exchanges heat with two sources, absorbing Q_2 at temperature T_2 and ceding Q_1 at temperature T_1 via some unspecified mechanism, and assuming that the profile of $E(t)$ is as in Fig.2

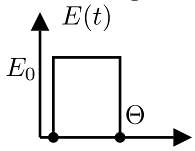


Fig.A2

where the value of E_0 corresponds to a temperature T_0 in the above sense. The inequality $-\Delta + \int_1^2 \frac{đQ}{T} \le 0$ becomes

$$-\int_0^\Theta \sigma_t\, dt + \frac{Q_2}{T_2} - \frac{Q_1}{T_1} \equiv -\frac{Q_0}{T_0} + \frac{Q_2}{T_2} - \frac{Q_1}{T_1} \le 0$$

For instance, we see that if $T_1 = T_2 = T_0$ we have realized a cycle which is irreversible. In it a quantity of heat $Q_2 - Q_1 - L$, with $L = Q_0 = T_0\Delta = T_0 \int_0^\Theta \sigma_t\, dt$ is absorbed at a single temperature $T = T_0$ and is transformed into the amount $Q_2 - Q_1 - L$ of work: however the inequality forbids this to be positive, as expected.

Mechanical and stochastic models.

A definition in Physics is interesting (only) if it is useful to describe properties of the systems in which we are interested. Therefore having set the above definitions one should expect to be asked why all the work was made.
In this case the whole matter was originated by efforts to interpret results that started to appear in the late 1970's concerning numerical experiments in molecular dynamics, [EM90].

(•) It is obvious that in numerical experiments one needs to deal with a finite system (and even not too large): hence various models of thermostats were devised for the purpose of obtaining equations that could be transformed into numerical codes and studied on electronic machines. This was a theoretical innovation with respect to previous models which either relied on stochastic boundary conditions or, in the more sophisticated cases, with (poorly understood) systems with infinitely many particles. And it opened the way to import the knowledge in the theory of dynamical systems that had been being developed in the two preceding decades or so.
The novelty with respect to stochastic thermostats was more conceptual than numerical. Given the number of particles a stochastic code is often only mildly more complex (and it could even be simpler) at least in the cases in which the noise is uncorrelated in time and acts on one particle at a time. This means that the resulting code does not require a longer running time than a deterministic one: a fact that can also be seen by noting that a stochastic system can be regarded as a deterministic system with more degrees of freedom (*i.e.* the ones needed to describe the random numbers

generators that one has to use and which, as it is well known, are simply suitably chaotic dynamical systems themselves).

From the point of view of code writing this amounts at adding a few more particles to the system.[8]

*NA*8

(•) It is by no means clear that by using mechanical thermostats one can obtain physically realistic models nor, assuming that the stochastic dynamic models are more realistic, models behaving in as complex a way as the stochastic ones (typically consisting in boundary collision laws in which the particles emerge with a Maxwellian velocity distribution with suitable variance, *i.e.* suitable temperature). Understandably the matter is controversial but quite a few researchers think that this has been positively demonstrated by large amounts of work done in the last thirty years.

Since stochastic models are just models with more degrees of freedom it is tautological that there is equivalence between all possible stochastic models and all deterministic ones. The real question is whether the rather simple deterministic thermostats models that have been used are able to simulate accurately the stochastic models believed to be more realistic.

(•) In my view it is likely that vast classes of thermostats, deterministic and stochastic as well, are equivalent in the sense that they produce motions which although very different when compared at equal initial conditions and at each time have, nevertheless, the same statistical properties, [Ga00],[Ga02]. And in my opinion there is already evidence that it is indeed possible to simulate the same system with simple deterministic thermostats or with some corresponding stochastic ones.

Here the equivalence is intended in a sense that is familiar in the theory of equilibrium ensembles: if one fixes suitably certain parameters then ensembles (*i.e.* time invariant probability distributions in phase space) that are apparently very different (*e.g.* microcanonical and canonical) give, nevertheless, the same statistical properties to vast (*not all*) classes of observables. If one fixes the energy U and the volume V in a microcanonical ensemble or the inverse temperature β and the volume V in the canonical ensemble then one obtains that local observables have the same statistical distribution in the two cases provided the value of β is chosen such that the average canonical energy is precisely U.

One among the most striking examples of such equivalence (She,Jackson), [SJ93], is the equivalence between the dissipative Navier–Stokes fluid and the Euler fluid in which the energy content of each shell of wave numbers is fixed (via Gauss' least constraint principle) to be equal to the value that Kolmogorov's theory predicts to be the energy content of each shell at a given (large) Reynolds number. Here one compares two very different mechanical thermostats. A more general view on the equivalence between different

[8] For instance if the stochastic thermostat is defined by requiring that upon collision with the boundary a particle rebounds with a Maxwellian velocity distribution with dispersion (temperature) depending only on the boundary point hit then one needs three Gaussian random number generators, *i.e.* essentially three more degrees of freedom.

thermostats has been developed since. In fact many instances in which Physicists say that "an approximation is reasonable" really correspond to equivalence statements about certain properties of different theories (and in the best cases the statements can be translated into proper mathematical conjectures).

(•) Coming therefore to consider more closely mechanical thermostat models the phase space contraction has turned out in many cases to be an interesting quantity often interpretable as the ratio between the work done by the thermostats on the systems and some kinetic energy average: this led since the beginning to identify the phase space contraction rate with the entropy creation rate. The above "philosophical" considerations have been developed to give some background interpretation to the vast phenomenology generated by the new electronic machines used as tools to investigate complex systems evolutions.

A collision between the previously held views, masterfully summarized by the book of De Groot–Mazur, [DGM84], based on *continuum mechanics* and the new approaches based on *transistors, chips and dynamical systems theory* ensued: often showing that the two communities give the impression of not really meditating on each other arguments.

(•) Setting aside controversies it is interesting that the mechanical thermostats approach has nevertheless led to a new perspective and to a few new results. Here I mention the *fluctuation relation*: the phase space contraction $\sigma(x(t),\dot{x}(t))$ which in various models has interesting physical meaning (like being related to conductivity or viscosity) is a fluctuating quantity as time goes on. Its average value in a time interval of size τ divided by its infinite time average in the future $\langle\sigma\rangle$ is a quantity p that still fluctuates. Of course it fluctuates less and less the larger is τ and its probability distribution (easily analyzable by observing it for a long time and by dividing the time into intervals of size τ and forming a histogram of the values thus observed) is expected on rather general grounds to be proportional to $e^{\tau\zeta(p)}$ for τ large with $\zeta(p)$ being a function with a maximum at $p = 1$ (*i.e.* at the average, hence most probable, value of p) *provided* $\langle\sigma\rangle > 0$, *i.e.* provided there is dissipation in the system and the system is, therefore, out of equilibrium.

If the dynamical equations are *reversible*, *i.e.* if there is an isometry I of phase space which anticommutes with the time evolution $(x,\dot{x}) \to S_t(x,\dot{x}) \equiv (x(t),\dot{x}(t))$ in the sense that $IS_t = S_{-t}I$ and furthermore $I^2 = 1$ then, provided the system motion is "very chaotic", it follows that

$$\zeta(-p) = \zeta(p) - p\langle\sigma\rangle$$

This is a *parameter free* symmetry relation that was discovered in a numerical experiment (Evans, Cohen, Morriss), [ECM93], and which has been checked in many cases. By very chaotic one means that the motion of the system can be assimilated to that of a suitable Anosov flow whose trajectories fill densely phase space (transitive Anosov flow).

(•) Indeed for transitive reversible Anosov systems the above relation holds as a theorem, [GC95]. Since models of physical systems are *not* Anosov systems from a strict mathematical point of view the above relation cannot be applied, not even to cases in which the model is reversible and the trajectories are dense on the allowed phase space: the chaotic hypothesis says that the fact that the system is not mathematically an Anosov system is not relevant for physical observations, in most cases. This is similar to the statement that in equilibrium systems the lack of ergodicity of motions is irrelevant in most cases and averages can be computed by assuming ergodicity (*i.e.* by using the microcanonical distribution).

If this is correct the above relation should hold: a non trivial fact to check due to the difficulty of observing such large fluctuations *and* to the lack of free parameters to fit the data, once they have been laboriously obtained.

(•) When the forcing of the system is let to 0 the above relation degenerates: not only $\langle\sigma\rangle \to 0$ but also p itself becomes ill defined as its definition involves division by $\langle\sigma\rangle$. Nevertheless by extracting the leading behavior of both sides the fluctuation relation leads to relations between average values of derivatives of dynamical quantities with respect to the intensity of the forcing, *evaluated at zero forcing*, and such relations can be interpreted as Onsager reciprocity relations and Green–Kubo expressions for suitably defined transport coefficients, [Ga02].

(•) Clearly a reversibility assumption on thermostats is a strong assumption and so is the chaotic hypothesis. Nevertheless the results are interesting and they seem to be among the few that can be obtained in a field which is well known for its imperviousness. The philosophical framework developed in Sections 1,2 helps keeping a unified view on a subject that is being developed although, strictly speaking, one could dispense with the philosophical view and concentrate on obtaining results that can be drily stated without appealing to entropy, entropy creation, thermostats *etc*.

(•) And one can go beyond various assumptions via the use of equivalence conjectures between different thermostats: for instance Drude's thermostat model which strictly speaking is not reversible is conjectured to be equivalent to a Gaussian thermostat which is reversible. The Navier–Stokes equation for incompressible fluids, clearly irreversible, is conjectured to be equivalent to a similar reversible equation, [Ga02], as the quoted experiment, [SJ93], shows and as other successive experiments seem to confirm, [GRS02]. The research along the just mentioned lines seems to go quite far and to lead not only to new perspectives but also to new results or confirmations (*i.e.* non contradictions) of the general views in Sections 1,2. Doubts about the whole approach can be legitimately raised, and have been raised, on the grounds that the results are too few and too meager to be really interesting: for instance one can hold against their consideration that they are not even sufficient to give some hint at a derivation of "elementary" relations like Fourier's law or Ohm's law. One can only say that time is not yet ripe to see whether the new methods and ideas lead really anywhere or at least to a better understanding of some of the problems that also the old ones have not been able to tackle, so far, (like the heat conduction laws or the electric conduction laws) in spite of intense research efforts.

.

Bibliography

[AA68] **Arnold, V.I., Avez, A.**: *Ergodic problems of classical mechanics*, Benjamin, New York, 1968.

[AW68] **Adler, R., Weiss, B.**: *Similarity of automorphism of the tours*, Memoirs of the American Mathematical Society, no. **98**, 1970.

[AS67] **Anosov, D.V., Sinai, Ya. G.**: *Certain smooth ergodic systems*, Russian Mathematical Surveys **22** (1967), no. 5, 103–167.

[Ar63] **Arnold, V.I.**: *Small denominators and problems of stability of motion in classical and celestial mechanics*, Russian Mathematical Surveys, **18**, 85–191, 1963.

[As70] **Asano, T.**: *Theorems on the partition function of the Heisenberg ferromagnet*, Journal of the Physical Society of Japan **29** (1970), 350–359.

[Av76] **Avez, A.**: *Sistemi dinamici*, Edited by C. Buzzanca, Quaderni del CNR-GNFM, (Istituto Matematico di Palermo), 1976.

[BCGOPS78] **Benfatto, G., Cassandro, M., Gallavotti G., Nicolò, F., Olivieri, E., Presutti, E., Scacciatelli, E.**: *Some probabilistic techniques in field theory*, Communications in Mathematical Physics **59** (1978), 143–166, **71** (1980), 95–130.

[Be64] **Becker, R.:** *Electromagnetic fields and interactions*, Blaisdell, New-York", 1964.

[BGN80] **Benfatto, G., Gallavotti, G., Nicolò, F.**: *Elliptic equations and gaussian processes*, Journal of Functional Analysis **36** (1980), 343–400.

[BG99] **Berretti, A., Gentile, G.**: *Scaling properties for the radius of convergence of Lindstedt series: the standard map*, Journal de Mathématiques Pures et Appliquées **78** (1999), 159–176.

[BG01] **Berretti, A., Gentile, G.**: *Bryuno function and the standard map*, Communications in Mathematical Physics **220** (2001), 623–656.

[BM94] **Berretti, A., Marmi, S.**: *Scaling near resonances and complex rotation numbers for the standard map*, Nonlinearity **7** (1994), 603–621.

[Bi31] **Birkhoff, G.D.**: *Proof of a recurrence theorem for strongly transitive systems*, Proceedings of the National Academy of Sciences **17** (1931), 650–655. And: *Proof of the ergodic theorem*, Proceedings of the National Academy of Sciences **17** (1931), 656–660. Also in *Collected mathematical papers*, Vol. II, p. 398, Dover, New York, 1968.

[Bi35] **Birkhoff, G.D.**: *Nouvelles rechèrchers sur les systèmes dynamiques*, Memoriae Pontificiae Academiae Lyncaei **1** (1935), 85–216.

[BFG03] **Bonetto, F., Falco, P.L., Giuliani, A.:** , *Analyticity and large deviations for the SRB measure of a lattice of coupled hyperbolic systems*, Preprint 2003.

[BGGM98] **Bonetto, F., Gallavotti, G., Gentile, G., Mastropietro, V.** *Lindstedt series, ultraviolet divergences and Moser's theorem*, Annali della Scuola Normale Superiore di Pisa, **26** (1998), 545–593. And: *Quasi linear flows on tori: regularity of their linearization*, Communications in Mathematical Physics, **192** (1998), 707–736. See also the review **Gallavotti, G.**: *Methods in the theory of quasi periodic motions*, *Recent advances in partial differential equations and applications*, Venezia, 1996, Eds. R. Spigler & S. Venakides, Proceedings of Symposia in Applied Mathematics, Vol. 54, 163–174, American Mathematical Society, 1997.

[BGW98] **Bourgain, J., Golse, F., Wennberg, S.**: *The ergodisation time for linear flows on tori: application to kinetic theory*, Communications in Mathematical Physics **190** (1998), 491–508.

[BK95] **Bricmont, J., Kupiainen, A.**: *Coupled analytic maps*, Nonlinearity **8** (1995), 379–396.

[BK96] **Bricmont, J., Kupiainen, A.**: *High temperature expansions and dynamical systems*, Communications in Mathematical Physics **178** (1996), 703–732.

[BK97] **Bricmont, J., Kupiainen, A.**: *Infinite dimensional SRB measures*, Physica D **103** (1997), 18–33.

[Bo09] **Boltzmann, L.**: *Ueber die Eigenschaften monocyklisher und andere damit verwandter systeme*, in *Wissenshftliche Abhandlungen*, Vol. III, Ed. P. Hasenhörl, Leipzig, 1909.

[Bo96] **Boltzmann, L.**: *Vorlesungen über Gastheorie*, Barth, Leipzig, 1896–1898.

[Bo70] **Bowen, R.**: *Markov partitions for Axiom A diffeormorphisms*, American Journal of Mathematics **92** (1970), 725-747.

[Bo75] **Bowen, R.**: *Equilibrium states and the ergodic theory of Anosov diffeormorphisms*, Lecture Notes in Mathematics, Vol. 470, Springer-Verlag, Berlin-Heidelberg, 1975.

[Br57] **Breiman, L.**: *The individual ergodic theorem of information theory*, Annals of Mathematical Statistics **28** (1957), 809–811.

[Br99] **Brush, S.**: *Gadflies and geniuses in the history of gas theory*, Synthese, 11–43, 1999.

[BS88] **Bunimovich, L., Sinai, Ya.G.**: *Space–time chaos in coupled map lattices*, Nonlinearity **1** (1998), 491–516.

[Ca82] **Cammarota, C.**: *Decay of correlations for infinite range interactions in unbounded spin systems*, Communications in Mathematical Physics **85** (1982), 517–528.

[Ca76] **Capocaccia, D**: *A definition of Gibbs states for a compact Z^v action*, Communications in Mathematical Physics **48** (1976), 85–88.

[CF02] Calogero, F., Françoise, J.P.: *Isochronous motions galore: nonlinearly coupled oscillators with lots of isochronous solutions*, in print in the *Proceedings of the workshop on superintegrability in classical and quantum systems*, Centre de Rechérches Mathématiques (CRM), Université de Montreal, September 2002.

[CO81] Cassandro, M., Olivieri, E.: *Renormalization group and analyticity in one dimension: a proof of Dobrushin's theorem* **80** (1981), 255–269.

[Ce99] Cercignani, C.: *Ludwig Boltzmann: the man who trusted atoms*, Oxford University Press, Oxford, 1999.

[CE80a] Collet, F., Eckmann, J.-P.: *Iterated maps of the interval*, Birkhauser, Boston, 1980.

[CE80b] Collet, P., Eckmann, J.-P.: *On the abundance of aperiodic behaviour for maps on the interval*, Communications in Mathematical Physics **73** (1980), 115-160.

[CEL80] Collet, P., Eckmann, J.-P., Lanford, O.: *Universal properties of maps on an Interval*, Communications in Mathematical Physics **76** (1980), 211–254.

[CF96] Chierchia, L., Falcolini, C.: *Compensations in small divisor problems*, Communications in Mathematical Physics, **175** (1996), 135–160.

[Ch99] Chandre, C.: PhD thesis, Physics Dept., Université de Bourgogne, 1999.

[CM01] Chandre, C., MacKay, R.S.D.: *Approximate renormalization with codimension-one fixed point for the break-up of some three-frequency tori*, Physics Letters A **275** (2000), no. 5-6, 394–400.

[Da94] Davie, A.M.: *The critical function for the semistandard map*, Nonlinearity **7** (1994), 219–229.

[DGM84] de Groot, S., Mazur, P.: *Non equilibrium thermodynamics*, Dover, Mineola, NY, 1984.

[DIS73] Duneau, M., Iagolnitzer, D., Soulliard, B.: *Decrease properties of truncated correlation functions and analyticity properties for classical lattices and continuous systems*, Communications in Mathematical Physics **31** (1973), 191–208.

[Do68a] Dobrushin, R.L.: *The description of a random field by means of conditional probabilities and conditions of its regularity*, Theory of Probability abd Applications **13** (1968), 197–224.

[Do68b] Dobrushin, R.L.: *Gibbsian Random fields for lattice systems with pairwise interactions*, Functional Analysis and Applications **2** (1968), 292–301.

[Do68c] Dobrushin, R.L.: *The problem of uniquenes of a gibbsian ran-*

dom field and the problem of phase transitions, FunctionalAnalysis and Applications **2** (1968), 302–312.

[Do69] Dobrushin, R.L.: *Gibbs fields: the general case*, Functional Analysis and Applications **3** (1969), 27–35.

[DS58] Dunford, N., Schwartz, J.: *Linear Operators*, Part I, Interscience, 1958.

[DS87] Dobrushin, R.L., Shlosman, S.B.: *Completely analytic interactions*, Journal of Statistical Physics **46** (1987), 971–1014.

[Dy69] Dyson, F.: *Existence of a phase-transition in a one-dimensional Ising ferromagnet*, Communications in Mathematical Physics **12** (1969), 91–107, 1969. And: *An Ising ferromagnet with discontinous long range order*, Communications in Mathematical Physics **21** (1971), 269–283.

[Ec75] Eckmann, J.-P.: *Lectures on constructive field theory*, Roma, 1975, Quaderni del GNFM, 1975.

[ECM93] Evans D.J., Cohen, E.G.D., Morriss, G.': *Probability of second law violations in shearing steady flows*, Physical Review Letters, **70**, 2401–2404, 1993.

[EE59] Ehrenfest, P., Ehrenfest, T.: *The conceptual foundations of the statistical approach in Mechanics*, Cornell University Press, Ithaca (reprinted), 1959.

[EG75] Esposito, R., Gallavotti, G.: *Approzimate symmetries and their spontaneous breakdown*, Annales de l'Institut H. Poincaré **22** (1975), 159–172.

[El96] Eliasson, L. H.: *Absolutely convergent series expansions for quasi–periodic motions*, Mathematical Physics Electronic Journal (MPEJ) **2** (1996), paper 4, 33 pp.

[EM90] Evans D.J., Morriss G.P.:,*Statistical Mechanics of Nonequilibrium Fluids*, Academic Press", New-York, 1990.

[Fe78] Feigenbaum, M.: *Quantitative universality for a class of nonlinear transformations*, Journal of Statistical Physics **19** (1978), no. 1, 25–52.

[Fe79] Feigenbaum, M.: *The universal metric properties of nonlinear transformations*, Journal of Statistical Physics **21** (1979), no. 6, 669–706.

[Fi67] Fisher, M.E.: *Theory of condensation and the critical point*, Physics–Physica–Fyzika **3** (1967), 255–283.

[FKG71] Fortuin, C., Kasteleyn, P., Ginibre, J.: *Correlaction inequalities for some partially ordered sets*, Communications in Mathematical Physics **22** (1971), 89-103.

[Ga72] Gallavotti, G.: *Instabilities and Phase transitions in the Ising model*, Rivista del Nuovo Cimento **2** (1972), 133–169.

[**Ga72a**] **Gallavotti, G.** *Phase separation line in the two–dimensional Ising model*, Communications in Mathematical Physics, **27** (1972), 103–136.

[**Ga77**] **Gallavotti, G.**: *Funzioni zeta ed insiemi basilari*, Rendiconti dell'Accademia Nazionale dei Lincei **51** (1977), 509–517.

[**Ga77**] **Gallavotti, G.**: *Elliptic operators and gaussian processes*, in *Statistical and physical aspects of Gaussian processes* (Saint-Flour, 1980), 349–360, Colloques internationaux du Centre National de la Rechérche scientifique, no. 307, Paris, 1981.

[**Ga82**] **Gallavotti, G.**: *Elementary Mechanics*, Springer-Verlag, New York, 1982.

[**Ga85**] **Gallavotti, G.**: *Renormalization theory and ultraviolet stability for scalar fields via renormalization group methods*, Reviews of Modern Physics **57** (1985), 471–562.

[**Ga94a**] **Gallavotti, G.**: *Twistless KAM tori*, Communications in Mathematical Physics **164** (1994), 145–156.

[**Ga94b**] **Gallavotti, G.**: *Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable hamiltonian systems. A review*, Reviews on Mathematical Physics **6** (1994), 343–411.

[**Ga99**] **Gallavotti, G.**: *A local fluctuation theorem*, Physica A **263** (1999), 39–50.

[**Ga00**] **Gallavotti, G.**: *Statistical Mechanics*, Springer-Verlag, Berlin, 2000.

[**Ga01**] **Gallavotti, G:**: *Renormalization group in statistical mechanics and mechanics: gauge symmetries and vanishing beta functions*, Physics Reports **352** (2001), 251–272.

[**Ga02**] **Gallavotti, G.**: *Fluid mechanics. Foundations*, Springer-Verlag, Berlin, 2002.

[**GC95**] **Gallavotti, G., Cohen, E.G.D.:** *Dynamical ensembles in nonequilibrium statistical mechanics*, Physical Review Letters, **74**, 2694–2697, 1995.

[**GG95**] **Gallavotti, G., Gentile, G.:** *Majorant series convergence for twistless KAM tori*, Ergodic theory and dynamical systems **15** (1995), 857–869.

[**GG02**] **Gallavotti, G., Gentile, G.:** *Hyperbolic low-dimensional invariant tori and summations of divergent series*, Communications in Mathematical Physics, **227**, 421–460, 2002.

[**GGM78**] **Gallavotti, G., Guerra, F., Miracle-Solé, S.**: *A comment on the talk by E. Seiler*, in *Mathematical problems in theoretical physics*, Eds. G. Dell'Antonio, S. Doplicher & G. Jona–Lasinio, Lecture notes in Physics, Vol. 80, pp. 436–438, Springer, Berlin, 1978.

[GJS73] Glimm, J., Jaffe, A., Spencer, T.: *The particle structure of the weakly coupled* $P(\varphi)_2$ *model and other applications of high temperature expansions. I. Physics of quantum field models*, pp. 132–198; *II. The cluster expansion*, pp. 199–242, in *Constructive field theory*, Eds. G. Velo & A. Wightman, Lecture Notes in Physics, Vol. 25, Springer, Berlin-New York, 1973.

[GL70] Gallavotti G., Lin, T.F.: *One dimensional lattice gases with rapidly decreasing interactions*, Archive for Rational Mechanics and Analysis **37** (1970), 181–191.

[GL03] Goldstein, S., Lebowitz J.: *On the (Boltzmann) entropy of Nonequilibrium systems", Preprint Rutgers University*, 2003.

[GLM02] Gallavotti G., Lebowitz, J., Mastropietro, V.: *Large deviations in rarefied quantum gases*, Journal of Statistical Physics **108** (2002), 831-861.

[GM68] Gallavotti, G., Miracle-Solé, S.: *Correlation functions of a lattice system*, Communications in Mathematical Physics **7** (1968), 271-288.

[GM70] Gallavotti, G., Miracle-Solé, S.: *Absence of phase transitions in hard core one–dimensional systems with long range interactions*, Journal of Mathematical Physics **11** (1970), 147–154, 1970. See also [GMR68].

[GMR68] Gallavotti, G., Miracle-Solé, S., Ruelle, D.: *Absence of phase transitions in one dimensional systems with hard core*, Physics letters A **26** (1968), 350–351.

[GM95] Gentile, G., Mastropietro, V.: *Tree expansion and multiscale analysis for KAM tori*, Nonlinearity **8** (1995), 1159–1178.

[GM96] Gentile, G., Mastropietro,V.: *KAM theorem revisited*, Physica D **90** (1996), 225–234. And: *Methods of analysis of the Lindstedt series for KAM tori and renormalizability in classical mechanics. A review with some applications*, Reviews in Mathematical Physics **8** (1996), 393–444. See also [GM95].

[GMR68] Gallavotti, G., Miracle-Solé, S., Robinson, D.: *Analyticity properties of a lattice gas*, Physics Letters A **25** (1967), 493–494.

[GMM73] Gallavotti, G., Martin Löf, A., Miracle-Solé, S.: *Some problems connected with the description of coexisting phases at low temperature in the Ising model*, Lecture Notes in Physics, Vol. 20, Ed. A. Lenard, Springer-Verlag, Heidelberg, 1973.

[Gi60] Gibbs, W.J.: *Elementary principles of Statistical Mechanics*, Dover, New York, 1960.

[GMK00] Gielis, G., MacKay, R.S.: *Coupled map lattices with phase* *transition* Nonlinearity, **13** (2000), 867-888.

[Gi70] Ginibre, J. *General formulation of Griffiths' inequalities*, Communications in Mathematical Physics **16** (1970), 310-328.

[GK78] Gruber, C., Kunz, H.: *General properties of polymer systems*, Communications in Mathematical Physics **22** (1971), 133-161.

[GP79] Griffiths, R., Pearce, E.P.: *Mathematical properties of position space renormalization group transformations*, Journal of Statistical Physics **20** (1979), 499-545.

[Gr62] Groeneveld, H.: *Two theorems on classical many particle systems*, Physics Letters, **3**, 50–51, 1962.

[GR71] Griffiths, R., Ruelle, D.: *Strict convexity (continuity) of the pressure in Lattice systems*, Communications in Mathematical Physics **23** (1971), 169–175.

[GR97] Gallavotti, G., Ruelle, D.: *SRB states and nonequilibrium statistical mechanics close to equilibrium*, Communications in Mathematical Physics, **190** (1997), 279–285.

[GRS02] Gallavotti G., Rondoni L., Segre E.: *Lyapunov spectra and nonequilibrium ensembles equivalence in 2d fluids*, Submitted to Physica D, 1–20, 2003.

[Gr81] Gross, L.: *Absence of second order phase transitions in the Dobrushin uniqueness theorem*, Journal of Statistical Physics **25** (1981), 57–72.

[He61] Hepp, K.: *Renormalization theory*, Lecture Notes in Physics, Springer, 1961.

[Ho74] Holley, R.: *Remarks on FKG inequalities*, Communications in Mathematical Physics **36** (1974), 227–235.

[Is76] Israel, R.: *High temperature analyticity in classical lattice systems*, Communications in Mathematical Physics **50** (1976), 245–257.

[Ja59] Jacobs, K.: *Die Ubertragung distrete Informationen durch periodische und fastperiodische Kanale*, Mathematische Annalen **137** (1959), 125–135.

[JLZ99] Jorba, A., Llave, R., Zou, M.: *Lindstedt series for lower dimensional tori*, in *Hamiltonian systems with more than two degrees of freedom* Ed. C. Simó, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., , Kluwer, **533**, 151–167, 1999.

[JM96] Jiang, M., Mazel, E.: *Uniqueness and exponential decay of correlations for some two–dimensional spin lattice systems*, Journal of Statistical Physics **82** (1996), 797–821.

[JP98] Jiang M., Pesin, Ya.B.: *Equilibrium Measures for Coupled Map Lattices: Existence, Uniqueness and Finite-Dimensional Approximations*, Communications in Mathematical Physics, **193**, 675–711, 1998.

[KH95] Katok, A., Hasselblatt, B.: *Introduction to the modern theory of dynamical systems*, Cambridge University Press, Cambridge, 1995.

[KS77] Keane, M., Smorodinsky, M.: *A class of finatary codes*, Israel J. Math. **26** (1977), no. 3–4, 352-371.

[KS79] Keane, M., Smorodinsky, M.: *Bernoulli schemes with the same entropy are finitarily isomorphic*, Annals of Mathematics **109** (1979), no. 2, 397-406.

[Ki57] Khinchin, A.Ya.: *Mathematical foundations of informations theory*, Dover, New York, 1957.

[Ki64] Khinchin, A.Ya.: *Continued fractions*, Dover, New York, 1964.

[Kr79] Krylov, N.S.: *Works on the foundations of statistical mechanics*, Princeton University Press, Princeton, 1979.

[Ku78] Kunz, H.: *Analyticity and clustering properties of unbounded spin systems*, Communications in Mathematical Physics **59** (1978), no. 1,53–69.

[La78] Lanford, O.: *Qualitative and statistical theory of dissipative systems*, Statistical Mechanics, C.I.M.E., I^o ciclo (1976), pp. 25-98, Liguori, Napoli, 1978

[Le11] Levi, E.E.: *Sur les équations différentielles périodiques*, Comptes Rendus de l'Academie des Sciences **153** (1911), – .

[Le93] Lebowitz, J.: *Boltzmann's entropy and time's arrow*, Physics Today, September, 32–38, 1993.

[Li01] Lindley, D.: *Boltzmann's atom: the great debate that launched a revolution in physics*, Free Press, New York, 2001.

[LR69] Lanford, O., Ruelle, D.: *Observables at infinity and states with short range correlations*, Communications in Mathematical Physics **13** (1969), 194–215.

[LR68] Lanford, O., Robinson, D.: *Statistical Mechanics of quantum spin systems*, Communications in Mathematical Physics **9** (1968), 327–338.

[LL68] Landau, L.D., Lifshitz, E.M.: *Statistical Physics*, Pergamon Press, Oxford-Edinburgh-New York, 1968.

[LY73] Lasota, A., Yorke, J.: *On the existence of invariant measures for a piecewise monotonic transformation*, Transactions of the American Mathematical Society **186** (1973), 481–488.

[Le74] Lebowitz, J.L.: *GHS and other inequalities*, Communications in Mathematical Physics **35** (1974), 87–92.

[Le73] Ledrappier, F.: *Mesure d'equilibre sur un reseau*, Communications in Mathematical Physics **33** (1973), 119–128.

[LY52] Lee, T.D., Yang, C.N.: *Statistical Theory of equations of state and phase transitions. II. Lattice gas and Ising model*, Physical Reviews **87** (1952), 410–419.

[Lo63] Lorenz, E.N.: *Deterministic non periodic flow*, Journal of the Athmospheric Sciences **20** (1963), 130–141, reprinted in [Cv84].

[Ma79] Manin, Yu.I.: *How convincing is a proof?*, Mathematica Intelligencer **2** (1979), no. 1, 17–18.

[Ma71] Manning, A.: *Axiom* A *diffeomorphisms have rational zeta functions*, Bullettin of the London Mathematical Society **3** (1971), 215–220.

[Ma65] Maxwell, J.: *Selected readings in Phisics*, Vol. 1 & 2, Ed. S. Brush, Pergamon Press, Oxford, 1965–66.

[Me65] Mel'nikov, V.K.: *On some cases of conservation of conditionally periodic motions under a small change of the Hamiltonian function*, Soviet Mathematics Doklady **6** (1965), 1592–1596.

[Me68] Mel'nikov, V.K.: *A family of conditionally periodic solutions of a Hamiltonian systems*, Soviet Mathematics Doklady **9** (1968), 882–886.

[Mi95] Miracle–Solé, S.: *On the microscopic theory of phase coexistence*, in *25 Years of Non-Equilibrium Statistical Mechanics*, Ed. J.J. Brey, Lecture Notes in Physics, Vol. 445, pp. 312–322, Springer, Berlin, 1995. And: *Surface tension, step free energy and facets in the equilibrium crystal shape*, Journal Statistical Physics **79** (1995), 183–214.

[Mi79] Misiurewicz, M.: *Absolutely continuous measures for certain maps of an interval*, Publications Mathématiques de l'Institut des Hautes Études Scientifiques **53** (1981), 17–51.

[Mo56] Morrey, C.B.: *On the derivation of the equations of hydrodynamics from Statistical Mechancs*, Communications in Pure and Applied Mathematics **8** (1955), 279–326.

[Mo67] Moser, J.: *Convergent series expansions for quasi periodic motions*, Mathematische Annalen **169** (1967), 136–176.

[Mo73] Moser, J.: *Stable and random motions in dynamical systems*, Annals of Matematical Studies, Princeton University Press, Princeton, 1973.

[Mo78] Moser, J.: *Is the solar system stable?*, Mathematical Intelligencer **1** (1978), 65-71.

[MR75] Monroy, G., Russo, L.: *A family of codes between Markov and Bernoulli schemes*, Communications in Mathematical Physics **43** (1975), 155–159.

[MS67] Minlos, R., Sinai, Ya: *The phenomenon of separation of phases at low temperatures in some lattice models of a gas, I*, Math. USSR Sbornik **2** (1967), 335–395. And: *The phenomenon of phase separation at low tem-*

peratures in some lattice models of a gas, II, Transactions of the Moscow Mathematical Society **19** (1968), 121–196. Both reprinted in [Si91].

[MS76] Magnen, J., Seneor, R.: *The infinite volume limit of φ_3^4 model*, Annales de l'Institut Henri Poincaré **24** (1976), 95–159.

[Ol88] Olivieri, E.: *Cluster expansion for lattice spin systems: a finite size condition for the convergence*, Journal of Statistical Physics **50** (1988), 1179–1200.

[On44] Onsager, L.: *Crystal statistics I. A two dimensional model with an order disorder transition*, Physical Review **65** (1944), 117–149.

[Or74] Ornstein, D.: *Ergodic Theory, randomness and dynamical systems*, Yale Mathematical Monographs, no.. 5, Yale University Press, New Haven, 1974.

[OW74] Ornstein, D., Weiss, B.: unpublished (1974).

[OS73] Osterwalder, K., Schrader, R.: *On the uniqueness of the energy density in the infinite volume limit for quantum field models*, Helvetica Physica Acta **45** (1973), 746–754.

[Pa67] Parthasarathy, K.R.: *Probability measures on metric spaces*, Probability and Mathematical Statistics, no. 3, Academic Press, New York-London, 1967.

[Pe63] Penrose, O.: *Convergence of fugacity expansions for fluids and lattice gases*, Journal of Mathematical Physics **4** (1963) 1312–1320.

[Pi79] Pianigiani, G.: *Absolutely continous invariant measures for the process $x_{n+1} = Ax_n(1 - x_n)$*, Bollettino dell'Unione Matematica Italiana A **16** (1979), no. 2, 374-378.

[Pi80] Pianigiani, G.: *First return map and invariant measures*, Israel Journal of Mathematics **35** (1980), no. 1-2, 32–48.

[Pi81] Pianigiani, G.: *Existence of invariant measures for piecewise continous transformations*, Annales Polonici Mathematici **40** (1981), no. 1, 39–45.

[PY79] Pianigiani, G., Yorke, J.: *Exanding maps on sets which are almost invariant: decay and chaos* (dedicated to J. Massera), Transactions of the American Mathematical Society **252** (1979), 351–366.

[Po85] Poincaré, H.: *Sur les courbes définies par les équations differentielles*, Journal de Mathématiques pures et appliquées **I**, 1885, reprinted in *Oeuvres Complètes* **1**, paper n. 57, p. 156, Gauthiers–Villars, Paris 1934–1954.

[Po93] Poincaré, H.: *Les Méthodes Nouvelles de la Mécanique Céleste*, Gauthiers–Villars, Paris 1893.

[Po86] Pöschel, J. *Invariant manifolds of complex analytic mappings*, Les Houches, XLIII (1984), vol. II, p. 949– 964, Eds. K. Osterwalder & R.

Stora, North Holland, 1986.

[Pr83] Prakash, Ch.: *High temperature differentiability of lattice Gibbs states by Dobrushin uniqueness techniques*, Journal of Statistical Physics **31** (1983), no. 1, 169–228.

[Ra99] Rawlins, D.: *Continued fractions decipherment: the Aristarchan ancestry of Hipparchos' yearlength & precession*, DIO **9** (1999), 30–42.

[Re57] Renij, A.: *Representations for real numbers and their ergodic properties*, Acta Mathematica, Academia Scientiae Hungarica **8** (1957), 477-493.

[Re78] Resibois, P.: *H–theorem for the (modified) nonlinear Enskog Equation*, Journal of Statistical Physics", **19**, 593, 1978.

[Ru63] Ruelle, D.: *Correlation functions of classical gases*, Annals of Physics **25** (1963), 109–120.

[Ru66] Ruelle, D.: *States of classical statistical mechanics*, Communications in Mathematical Physics **3** (1966), 133–150.

[Ru67] Ruelle, D.: *Statistical Mechanics of a one dimensional lattice gas*, Communications in Mathematical Physics **9** (1967), 267–278.

[Ru69] Ruelle, D: *Statistical Mechanics*, Benjamin, New York, 1969.

[Ru71] Ruelle, D.: *Extension of the Lee-Yang circle theorem*, Physical Review Letters **26** (1971), 303-304.

[Ru72] Ruelle, D.: *On the use of small external fields in the problem of symmetry breakdown in statistical Mechanics*, Annals of Physics **69** (1972), 364–374.

[Ru73] Ruelle, D.: *Some remarks on the location of zeroes of the partition function for lattice systems*, Communications in Mathematical Physics **31** (1973), 265-277.

[Ru76] Ruelle, D.: *A measure associated with axiom A attractors*, American Journal of Mathematics **98** (1976), 619–654.

[Ru77] Ruelle, D.: *Application conservant une mesure absolument continue par rapport a dx sur* $[0,1]$, Communications in Mathematical Physics **55** (1977), 47–51.

[Ru78] Ruelle, D.: *Thermodynamic formalism*, Encyclopedia of Mathematics and its applications, Vol. 5, Addison-Wesley, Boston, 1978.

[Ru95] Ruelle, D.: *Turbulence, strange attractors and chaos*, World Scientific, River Edge, 1995.

[Ru97] Ruelle, D.: *Differentiation of SRB states*, Communications in Mathematical Physics, , **187** (1997), 227–241.

[Ru99] Ruelle, D.: *Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics*, Journal of Statistical Physics **95** (1999), 393–468.

[Ru01] Rüssmann, H.: *Invariant tori in non–degenerate nearly integrable Hamiltonian systems*, Regular and Chaotic Dynamics **6** (2001), 119–204.

[Ru03] Ruelle D.: *Extending the definition of entropy to nonequilibrium*, Proceedings of the National Acadey of Sciences, USA, 2003, MathPhys archive 09-98.

[Sh49] Shannon, C.: *The mathematical theory of Communication*, Bell Systems Technology Journal, **27** (1948), 379–423, 623–656.

[Sh73] Shields, P.: *The theory of Bernoulli shifts*, University of Chicago Press, Chicago, 1973.

[SJ93] She, Z.S., Jackson E.:, *Constrained euler system for Navier-Stokes turbulence*, Physical Review Letters, **70**, 1255–1258, 1993.

[Si93] Simon, B.: *The statistical mechanics of lattice systems*, Princeton University Press, 1993.

[Si68a] Sinai, Ya.G.: *Markov partitions and C-Diffeormorphisms*, Functional Analysis and Applications **2** (1968), 64–69.

[Si68b] Sinai, Ya.G.: *Costruction of Markov partitions*, Functional analysis and Applications **2** (1968), no. 2, 70–80.

[Si72] Sinai, Ya.G.: *Gibbs measures in ergodic theory*, Russian Mathematical Surveys **166** (1972), 21–69.

[Si76] Sinai, Ya.G.: *Introduction to ergodic theory*, Mathematical Notes, Vol. 18, Princeton University Press, Princeton, 1976.

[Si91] Sinai, Ya.G.: *Mathematical problems of statistical mechanics. Collection of papers*, Advanced Series in Nonlinear Dynamics, Vol. 2, World Scientific, Teaneck, 1991.

[Si91b] Sinai, Ya.G.: *Dynamical systems.Collection of papers*, Advanced Series in Nonlinear Dynamics, Vol. 1, World Scientific, Teaneck, 1991.

[Si94] Sinai, Ya.G.: *Topics in ergodic theory*, Princeton Mathematical Series, Vol. 44, Princeton University Press, Princeton, 1994.

[Sy79] Sylvester, G.: *Weakly coupled Gibbs measures*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **50** (1979), 97–118.

[SF71] Suzuki, M., Fischer, M.: *Zeros of the partition function for the Heisenberg ferroelectric and general Ising models*, Journal of Mathematical Physics **12** (1971), 235–246.

[Sh72] Shen, C.: *A functional calculus approach to the Ursell-Mayer functions*, Journal Mathematical Physics **13** (1972), 754-759.

[Th83] Thirring, W.: *Course in Mathematical Physics*, Vol. 1, Springer, Wien, 1983.

[Tr89] Treshev, D.: *The mechanism of destruction of resonant tori of Hamiltonian systems*, Mathematics USSR-Sbornik, (translation), **68** (1991), 181–203. See also the MathSciNet review by M.B. Sevryuk, <http://www.ams.org>.

[UV47] Ulam, S., Von Neumann, J.: *On combination of stochastic and deterministic processes*, Bulletin of the American Mathematical Society **53** (1947), 1120–.

[VH50] Van Hove, L:: *Sur l'integrale de configuration pour les systèmes de particules á une dimension*, Physica **16** (1950), 137–143.

[VW41] van der Waerden, B.L.: *Die lange reichweite der regelmässigen atomanordnung in Mischcristallen*, Zeitschrift für Physik **118** (1941), 473–488.

[We78] Wehrl, A.: *General properties of entropy*, Reviews of Modern Physics **60** (1978), 221–260.

[Wi70] Wilson, K.G.: *Model of coupling constant renormalization*, Physical Review D **2** (1970), 1438–1472.

[Wi83] Wilson, K.G.: *The renormalization group and critical phenomena*, Reviews of Modern Physics **55** (1983), 583–600.

[Yo95] Yoccoz, J.-Ch.: *Théorème de Siegel, Nombres de Bruno et Polinômes Quadratiques*, Astérisque **231** (1995), 3–88.
[Ze68] Zemansky, M.W.: *Heat and thermodynamics*, McGraw-Hill, New-York, 1957.

Author index

Subject index

Citations index