Review on the book Gradient flows in metric spaces and in the space of probability measures by **Ambrosio, Gigli and Savaré**

Wilfrid Gangbo *

February 2006

This book consists of two parts which can be read independently. It is written by authors who are on top of the topics they discuss. Throughout the reading of this book the reader will absorb a lot of fine and subtle information on the subjects discussed. This book contains known results, recent results and new results which never appeared before. The first part is about nonsmooth analysis and ordinary differential equations on metric spaces. The tools developed in this first part apply mainly to gradient flow differential equations. We recommend to the impatient reader to spend some time understanding the introduction and, if possible, chapter one. The proofs here are very accessible and the best way to penetrate the heart of the topic is to try to understand some of the proofs in that chapter. The theories developed in this book are very instructive for anybody who wants to have a deep understanding of parabolic equations.

The second part focuses on some particular metric spaces, the set of probability measures, endowed with a Wasserstein distance. The results presented in both parts improve our understanding of parabolic partial differential equations. Most of the recent developments of the Monge-Kantorovich theory, especially those related to partial differential equations, are collected in this book. It is an excellent source of information for people who have joint interest in probability, partial differential equations, fluid mechanics and probably geometry. Many remarkable old and new results obtained at the cross section of these fields, are incorporated in the book.

1 Analysis on metric spaces

1.1 A new concept as a substitute of gradient flow: curves of maximal slope

The first part of this book develops basic tools in metric spaces $(\mathcal{S}, dist)$. These tools are needed to give a sense to ordinary differential equations which are the analogue of gradient flows on metric spaces, which may not have any differential structure. It starts with the concept of metric derivative of an absolutely continuous curve $v : (a, b) \subset \mathbf{R} \to \mathcal{S}$ given by

$$|v'|(t) = \lim_{h \to 0} \frac{dist(v(t+h), v(t))}{|h|}. \tag{1}$$

For $p \in [1, +\infty]$, let $AC^p(a, b; \mathcal{S})$ be the set of $v : (a, b) \subset \mathbf{R} \to \mathcal{S}$ such that

$$dist(v(t), v(s)) \leq \int_s^t m(r)dr, \qquad \forall a < s \leq t < b \tag{2}$$

*Georgia Institute of Technology,Atlanta, GA 30332, USA, Atlanta, GA 30332, USA, gangbo@math.gatech.edu.
Key words: Gradient in metric spaces, parabolic equations, mass transfer, Wasserstein metric. AMS code: 35, 49J40, 82C40.

for some $m \in L^p(a,b)$. If $\mathcal{L}^1$ is the one dimensional Lebesgue measure, the authors prove that the limit in (1) exists for $\mathcal{L}^1$–almost every $t \in (a,b)$ provided that $v \in AC^p(a,b;\mathcal{S})$. The set $AC^1(a,b;\mathcal{S})$ is called the set of absolutely continuous curves. In that case, $|v'|$ is the smallest function m in $L^p(a,b)$ such that (2) holds.

If $\phi : \mathcal{S} \to \mathbf{R}$ and $u \in AC^p(a,b;\mathcal{S})$, u' and $\nabla\phi(u(t))$ may not make any sense since $\mathcal{S}$ may not have a differential structure. Thus, the equation

$$u'(t) = -\nabla\phi(u(t)) \tag{3}$$

may not be well-defined. Assume for a moment that $\mathcal{S}$ is a Hilbert space with an inner product $< \cdot ; \cdot >$ and denote its norm by $|| \cdot ||$. The equation $u'(t) = -\nabla\phi(u(t))$ is then equivalent to

$$0 \geq \frac{1}{2}||u'(t) + \nabla\phi(u(t))||^2 = \frac{1}{2}||u'(t)||^2 + \frac{1}{2}||\nabla\phi(u(t))||^2 + < u'(t); \nabla\phi(u(t)) >$$

which is equivalent to

$$0 \geq \frac{1}{2}||u'(t)||^2 + \frac{1}{2}||\nabla\phi(u(t))||^2 + \frac{d}{dt}\phi(u(t)) \tag{4}$$

provided that the chain rule $\frac{d}{dt}\phi(u(t)) =< u'(t); \nabla\phi(u(t)) >$ holds. Hence, (4) suggests that in case of lack of differential structure on $\mathcal{S}$, the equation $u'(t) = -\nabla\phi(u(t))$ could still make sense, if we find the right substitute for $||u'(t)||$, $||\nabla\phi(u(t))||$, and if $\frac{d}{dt}\phi(u(t))$ is well-defined. A candidate for an upper bound for $||\nabla\phi||$ is the local slope of ϕ defined by

$$|\partial\phi|(v) = \limsup_{w \to v} \frac{(\phi(v) - \phi(w))^+}{d(v,w)}. \tag{5}$$

We can convince ourselves that the expression in (5) is a good candidate for $||\nabla\phi(v)||$ when $\mathcal{S}$ is say, a Hilbert space and ϕ is differentiable at v. Indeed, in that case, the expression in (5) is clearly bounded above by $||\nabla\phi(v)||$. Since we identify $\mathcal{S}$ and its dual and

$$\lim_{r \to 0^+} \frac{\Big(\phi(v) - \phi(w_r)\Big)^+}{dist(v, w_r)} = ||\nabla\phi(v)||, \quad \text{where} \quad w_r = v - r\nabla\phi(v)/||\nabla\phi(v)||,$$

we conclude that $|\partial\phi|(v) = ||\nabla\phi(v)||$.

Let us suppose that ϕ never assumes the value $-\infty$ and is a proper function in the sense that $\phi(v_o) < +\infty$ for at least a $v_o \in \mathcal{S}$. Setting $g = |\partial\phi|$, then g is a weak upper gradient for ϕ in the following sense. For all $v \in AC^1(a,b;\mathcal{S})$, if $|v'| g \circ v \in L^1(a,b)$ and $\phi \circ v = \varphi$ $\mathcal{L}^1$ a.e. in (a,b) and φ is a function of finite pointwise variations in (a,b), then $|\varphi'| \leq |v'|\, g \circ v$. Note that there are several weak upper gradients. For instance if g is one of them then $g + \lambda$ is another one for $\lambda \geq 0$. Another candidate of upper bound for the modulus of the gradient is the strong upper gradient. A function $g : \mathcal{S} \to [0, +\infty]$ is called strong upper gradient for ϕ if for all $v \in AC^1(a,b;\mathcal{S})$, $g \circ v$ is a Borel map and

$$|\phi(v(t)) - \phi(v(s))| \leq \int_s^t |v'|(r) g \circ v(r) dr, \qquad \forall a < s \leq t < b.$$

If ϕ is lower semicontinuous, then a strong upper gradient for ϕ is l_ϕ, the global slope of ϕ, defined by

$$l_\phi(v) = \sup_{w \neq v} \frac{(\phi(v) - \phi(w))^+}{d(v,w)}.$$

If $(\mathcal{S}, ||\cdot||)$ is a Banach space, $(\mathcal{S}^*, ||\cdot||_*)$ is its dual space and $\phi : \mathcal{S}$ is convex and lower semicontinuous, then

$$l_\phi(v) = |\partial\phi|(v) = \min\{||\xi||_* : \ \xi \in \partial\phi(v)\}$$

where $\partial\phi(v)$ is the subdifferential of ϕ at v. We have the characterization

$$\xi \in \partial\phi(v) \subset \mathcal{S}^* \quad \text{if and only if} \quad \liminf_{w\to v} \frac{\phi(w) - \phi(v) - < \xi; w - v >}{||v - w||} \geq 0.$$

The equivalence between (3) and (4) suggests that the analogue of gradient flows in case of lack of differential structure on $\mathcal{S}$ is

$$0 \geq \frac{1}{2}|u'|^2(t) + \frac{1}{2}g^2(u(t)) + \frac{d}{dt}\phi(u(t)). \tag{6}$$

Here, g is a weak upper gradient for ϕ, used as a substitution of what would have been the modulus of the gradient of $\nabla\phi$. The metric derivative $|u'|$ has been used as a substitution of what would have been the modulus of the tangent vector u'. Any solution $u \in AC^1(a, b; \mathcal{S})$ satisfying (6) will be called a 2–curve of maximal slope for ϕ, with respect to the upper gradient g of ϕ. If $p \in (1, +\infty)$ and $q = p/(p-1)$ is the conjugate of p, the authors exploit Young's inequality $|x|^p + \frac{1}{q}|y|^q$ $q = p/(p-1)$ to motivate the definition of p–curve of maximal slope as follows. Assume that $u \in AC^1(a, b; \mathcal{S})$. We say that u is a p–curve of maximal slope for ϕ, with respect to the its weak upper gradient g, if $\phi \circ u = \varphi$ $\mathcal{L}^1$ a.e. in (a, b) and φ is a non-increasing function such that

$$0 \geq \frac{1}{p}|u'|^p(t) + \frac{1}{q}g^q(u(t)) + \varphi'(t) \qquad \text{for} \quad \mathcal{L}^1 \quad \text{a.e.} \quad t \in (a, b). \tag{7}$$

Since $|\varphi'| \leq |v'|\, g \circ v$, one can combine Young's inequality to conclude that when (7) holds, then

$$|u'|^p(t) = g^q(u(t)) = -\varphi'(t) \qquad \text{for} \quad \mathcal{L}^1 \quad \text{a.e.} \quad t \in (a, b).$$

The authors further motivate the definition in (7) by showing that when $\mathcal{S}$ is a Banach space, their definitions coincide with the classical definitions in the literature. For the sake of brevity we lay down their arguments by further imposing that the Banach space is reflexive. We refer the reader to section 1.4 for complete statements on these issues. The first author has established in a previous work with B. Kirchheim that $v \in AC^p(a, b; \mathcal{S})$ is equivalent to v is differentiable at $\mathcal{L}^1$ a.e. $t \in (a, b)$, and its derivative $v' \in L^p(a, b; \mathcal{S})$. In that case the metric slope of v coincides with the norm of v' at $\mathcal{L}^1$ a.e. $t \in (a, b)$. Suppose that $\phi : \mathcal{S} \to (-\infty, +\infty]$ is a proper, lower semicontinuous function and $u \in AC^p(a, b; \mathcal{S})$ is such that $\phi \circ u$ is $\mathcal{L}^1$ a.e. equal to a non-increasing function. It is shown that u is p–curve of maximal slope for ϕ, with respect to the weak upper gradient of ϕ, if and only if

$$\emptyset \neq -\partial^{min}\phi(u(t)) \subset \mathcal{J}_p(u'(t)) \qquad \text{for} \quad \mathcal{L}^1 \quad \text{a.e.} \quad t \in (a, b). \tag{8}$$

Here, $\partial^{min}\phi(u(t))$ consists of the elements of minimal norm in $\partial\phi(u(t))$, the set of Fréchet differential of ϕ at $u(t)$ and $\mathcal{J}_p : \mathcal{S} \to 2^{\mathcal{S}'}$ is defined by $\xi \in \mathcal{J}_p(v)$ if and only if

$$< \xi; v >= ||v||^p = ||\xi||_*^q = ||v||\,||\xi||_*.$$

Note that if the norm $||\cdot||$ on $\mathcal{S}$ is differentiable, ϕ is differentiable and $p = 2$, then (8) is equivalent to

$$-\nabla\phi(u(t)) = u'(t) \qquad \text{for} \quad \mathcal{L}^1 \quad \text{a.e.} \quad t \in (a, b).$$

1.2 Existence of curves of maximal slope

For the sake of simplicity, assume for a moment that $\mathcal{S} = \mathbf{R}^d$ and $\phi : \mathbf{R}^d \to \mathbf{R}$ is twice differentiable. One can solve (3) by a fixed point method or a constructive algorithm which could be an explicit or implicit scheme. Implicit schemes have many advantages. Very often, they have a better convergence property. The authors show how to adapt implicit schemes to metric spaces $\mathcal{S}$ which are not flat. Assume we are given $u_o \in \mathbf{R}^d$ and we are to find $u : [0,1] \to \mathbf{R}^d$ satisfying (3) with $u(0) = u_o$. First, we fix a time step size $0 < \tau << 1$ such that $N = 1/\tau$ is an integer. Assume that we can define inductively u_{k+1} as a solution to

$$\frac{u_{k+1} - u_k}{\tau} = -\nabla\phi(u_{k+1}) \quad k = 0, \cdots, N-1. \tag{9}$$

We may interpolate the discrete values $\{u_k\}_{k=0}^N$ to define $u^\tau : [0,1] \to \mathbf{R}^d$ by

$$u^\tau(t) = u_k \qquad t \in (k\tau, (k+1)\tau], \quad k = 0, \cdots, N-1.$$

If we can interpret $(u_{k+1} - u_k)/\tau$ as $(u^\tau)'(t)$ for $t \in (k\tau, (k+1)\tau)$, (9) reads off

$$(u^\tau)' = -\nabla\phi(u^\tau) + 0(\tau). \tag{10}$$

The main issues are:

(i) given u_k, under what conditions can we find u_{k+1} that solves (9) even when $\mathcal{S}$ is a Hilbert space? What is the analogue of (9) when $\mathcal{S}$ is an arbitrary metric space?

(ii) Can one establish the analogue (10) on metric spaces?

(iii) Suppose the analogue of (10) holds on metric spaces. Can we show that $\{u^\tau\}_{\tau>0}$ is pre-compact for a topology (which might be) weaker than the topology of the metric *dist*? If $\{u^\tau\}_{\tau>0}$ converges to some u for the topology σ, as τ tends to 0 , does u satisfy (3)?

Even when $\mathcal{S} = \mathbf{R}^d$ and ϕ is of class C^∞, unless an upper bound is imposed on $|\nabla\phi|$ (or $|\nabla^2\phi|$), a solution to (3) may exist only for $t \in [0, \epsilon)$ for some $\epsilon > 0$ which may be small. Existence of a solution $u(t)$ for all $t \geq 0$ can be established when $\mathcal{S} = \mathbf{R}^d$ and $\nabla^2\phi \geq \lambda I$ for some $\lambda \in \mathbf{R}$. Here, I is the $d \times d$ identity matrix. The inequality $\nabla^2\phi \geq \lambda I$ which expresses that all the eigenvalues of $\nabla^2\phi$ are greater than or equal to λ is equivalent to saying that $u \to \phi(u) - \lambda||u||^2/2$ is convex. In other words, for all $u, v \in \mathbf{R}^d$,

$$\phi((1-t)u + tv) \leq (1-t)\phi(u) + t\phi(v) - \frac{\lambda}{2}t(1-t)||u-v||^2 \qquad \forall t \in (0,1). \tag{11}$$

When $(\mathcal{S}, dist)$ is a complete metric space and $\lambda \in \mathbf{R}$, $\phi : \mathcal{S} \to [-\infty, +\infty]$ is said to be λ–convex if the analogue of (11) holds:

$$\phi(\gamma_t) \leq (1-t)\phi(\gamma_o) + t\phi(\gamma_1) - \frac{\lambda}{2}t(1-t)\, dist^2(\gamma_o, \gamma_1), \tag{12}$$

for all minimal geodesics $t \to \gamma_t$ of constant speed.

When $\mathcal{S} = \mathbf{R}^d$, any solution u_{k+1} of (9) is a critical point of the functional $u \to ||u - u_k||^2/2\tau + \phi(u)$. When $\mathcal{S}$ is an arbitrary metric space, the authors introduce the functional

$$\Phi(\tau, u, v) = \frac{dist^2(u,v)}{2\tau} + \phi(v). \tag{13}$$

They also introduce a Hausdorff topology σ, weaker than the *dist*–topology and such that if $\{u_n\}_{n=1}^\infty$, (resp. $\{v_n\}_{n=1}^\infty$) converges to u (resp. v) then

$$dist(u,v) \leq \liminf_{n \to +\infty} dist(u_n, v_n).$$

Given u_k, a sufficient condition for a point u_{k+1} to be a critical point of $\Phi(\tau, u_k, \cdot)$ is that u_{k+1} be a minimizer. Standard conditions in the calculus of variations which ensure existence of a minimizer are:

(H 1) σ–lower semicontinuity of ϕ on bounded subsets of $\mathcal{S}$.

(H 2) If $\{u_n\}_{n=1}^{\infty}$ is a bounded sequence in $\mathcal{S}$ and $\{\phi(u_n)\}_{n=1}^{\infty}$ is a bounded sequence in $\mathbf{R}$, then $\{u_n\}_{n=1}^{\infty}$ is σ–precompact.

To be hopeful that $\Phi(\tau, u_k, \cdot)$ will admit a minimizer for all $0 < \tau << 1$ and all $u_k \in \mathcal{S}$, one should at least impose that there exists $\tau^* > 0$ and $u^* \in \mathcal{S}$ such that

$(H3)\quad \phi_{\tau_*}(u_*) = \inf_{v \in \mathcal{S}} \Phi(\tau_*, u_*, v) > -\infty$

The author introduce the number

$$\tau_*(\phi) = \sup\{\tau > 0 : \ \phi_\tau(u) > -\infty \quad \text{for some} \quad u \in \mathcal{S}\}$$

One can now check that if $0 < \tau < \tau_* \leq \tau_*(\phi)$ and $u \in \mathcal{S}$, then

$$\phi_\tau(u) \geq \phi_{\tau_*}(u_*) - \frac{1}{\tau_* - \tau} dist^2(u_*, u) \tag{14}$$

and so, $\phi_\tau(u) > -\infty$ provided that $\phi_{\tau_*}(u_*) > -\infty$. Also, for $u, v \in \mathcal{S}$,

$$dist^2(u, v) \leq \frac{4\tau\tau_*}{\tau_* - \tau}\Big(\Phi(\tau, u, v) - \phi_{\tau_*}(u_*) + \frac{1}{\tau_* - \tau} dist^2(u_*, u)\Big) \tag{15}$$

Since (H 1, 2, 3) imply (14) and (15), existence of a minimizer $u_{k+1} \in \mathcal{S}$ is ensured for the functional $\Phi(\tau, u_k, \cdot)$ if $\tau < \tau_*/2$. Hence, given u_o, one can inductively define a sequence $\{u_k\}_{k=0}^{\infty} \subset \mathcal{S}$ such that $\Phi(\tau, u_k, u_{k+1}) = \phi_\tau(u_k)$. In fact, the authors study a more subtle problem, where they allow the starting point of the algorithm to be a point $u_{o,\tau}$, close to u_o in the sense that

$$\lim_{\tau \to 0^+} \phi(u_{o,\tau}) = \phi(u_o), \quad u_{o,\tau} \text{ tends to } \ u_o \text{ in the } \sigma \text{ topology, as } \tau \text{ tends to } \ 0$$

One can interpolate the u_k's in time and obtain a function $u_\tau : [0, +\infty) \to \mathcal{S}$ such that the analogue of (10) holds. To do that, it is important to study the finest properties of Φ and ϕ_τ. There are several possible ways of interpolating the u_k's to obtain an approximate solution u^τ. The authors choose a deep interpolation which goes back to De Giorgi [17]. We hereby describe a simplified version of the algorithm and the interpolation used in the manuscript under review. First, we introduce

$$J_\tau[u] = argmin\Phi(\tau, u, \cdot), \quad d_\tau^+(u) = \sup_{v \in J_\tau[u]} dist(v, u), \quad d_\tau^-(u) = \inf_{v \in J_\tau[u]} dist(v, u).$$

Set $t_k = k\tau$ (in the book under review, the authors do not impose that $t_{k+1} - t_k$ is independent of k). Given $\delta \in (0, t_{k+1} - t_k) = (0, \tau)$, one chooses the interpolation

$$u_\tau(t + \delta) \in argmin_{u \in \mathcal{S}} \Phi(\delta, u_k, u). \tag{16}$$

That interpolation satisfies the energy equality

$$\frac{1}{2} \int_{t_i}^{t_j} \Big(|u_\tau'|^2(t) + G_\tau^2(t)\Big) dt + \phi(u_\tau(t_j)) = \phi(u_\tau(t_i)). \tag{17}$$

where,

$$|u_\tau'|(t) = \frac{dist(u_\tau(t_k), u_\tau(t_{k-1}))}{t_k - t_{k-1}}, \qquad G_\tau(t) = \frac{d_{t - t_{k-1}}^+(u_\tau(t_{k-1}))}{t - t_{k-1}} \qquad t \in (t_{k-1}, t_k].$$

Set

$$\bar{u}_\tau(t) = u_\tau(t_k) \qquad t \in (t_{k-1}, t_k]$$

The equality in (17) is a powerful property which is used to show existence of a $u \in AC^2_{loc}([0,+\infty);\mathcal{S})$ and existence of a subsequence $\{\tau_n\}_{n=1}^\infty$ (independent of t) such that for all $t \geq 0$

$$u_{\tau_n}(t), \bar{u}_{\tau_n}(t) \text{ tend to } u(t) \text{ in the } \sigma \text{ topology, as } n \text{ tends to } +\infty. \tag{18}$$

Exploiting again (17), the authors obtain an energy inequality for u :

$$\frac{1}{2}\int_s^t \Big(|u'|^2(r) + |\partial^{relaxed}\phi(u(r))|^2\Big)dr \leq \varphi(s) - \varphi(t) \tag{19}$$

where,

$$\varphi(t) = \lim_{n\to+\infty} \phi(\bar{u}_{\tau_n}(t)), \qquad |\partial^{relaxed}\phi(w)| = \inf_{\{w_n\}_{n=1}^\infty} \{\liminf_{n\to+\infty} |\partial\phi(w_n)|\}.$$

Here, the infimum is performed over the set of bounded sequences $\{w_n\}_{n=1}^\infty \subset \mathcal{S}$ such that $\{\phi(w_n)\} \subset \mathbf{R}$ is bounded and $\{w_n\}_{n=1}^\infty$ tends to w in the σ topology, as n tends to $+\infty$. They prove that

$$\varphi(t) = \phi(u(t)) \quad \text{for } \mathcal{L}^1 \text{ a.e. } t \in (0,+\infty), \tag{20}$$

provided that $|\partial^{relaxed}\phi|$ is a weak upper gradient. Now (19) and (20) readily yield that

$$-\varphi'(t) \geq \frac{1}{2}\Big(|u'|^2(t) + |\partial^{relaxed}\phi(u(t))|^2\Big) \quad \text{for } \mathcal{L}^1 \text{ a.e. } t \in (0,+\infty) \tag{21}$$

and so, u is a curve of maximal slope for ϕ with respect to its weak upper gradient $|\partial^{relaxed}\phi|$.

It is worth mentioning that when ϕ is λ–convex, its slope $|\partial\phi|$ is lower semicontinuous (so it is equal to the relaxed slope $|\partial^{relaxed}\phi|$) and it is a strong upper gradient. Hence, not only are all the inequalities in (19) and (21) equalities, but also, $t \mapsto \phi(u(t))$ is locally absolutely continuous.

1.3 Uniqueness of curves of maximal slope

Assume $(\mathcal{S}, dist)$ is a complete metric space and $\phi : \mathcal{S} \to (-\infty,+\infty]$ is convex, proper, lower semicontinuous. Let $D(\phi)$ be the domain of ϕ, which is the set of $u \in \mathcal{S}$ such that $\phi(u) < +\infty$. Assume temporarily that $\mathcal{S}$ is a Hilbert space and ϕ satisfies the coercivity condition

$$\exists u_* \in D(\phi), r_* > 0 : \quad \inf_{v\in\mathcal{S}}\{\phi(v) : \ d(v,u_*) \leq r_*\} > -\infty. \tag{22}$$

Under these restrictive assumptions, uniqueness of solutions of (3) is well understood. The uniqueness property has been established on *nonpositively curved metric spaces:* these are length spaces such $1/2\,dist^2(\cdot,w)$ is 1–convex for every $w \in \mathcal{S}$. This notion was introduced by Aleksandrov and can be found in the book by Jost [25]. As done in section 4 of the manuscript under review, let us come back to the assumption that $(\mathcal{S}, dist)$ is solely a complete metric space. To obtain uniqueness of solutions, the authors impose that the functional Φ introduced in (13) satisfies the following property: for every $w, u_o, u_1 \in D(\phi)$ there exists a curve $\gamma : [0,1] \to \mathcal{S}$ such that $\gamma(0) = u_o$, $\gamma(1) = u_1$ and

$$\Phi(\tau, w, \cdot) \ \text{ is } (\frac{1}{\tau} + \lambda) - \text{convex for each } \tau \in (0, \frac{1}{\lambda^-}). \tag{23}$$

Clearly, (23) holds when ϕ is λ–convex and $\mathcal{S}$ is a *nonpositively curved metric spaces* as considered by Mayer [32]. The results obtained here are stronger than a simple uniqueness result. The authors show that no matter what subsequence is chosen in (18), we can obtain u as

$$u(t) = S[u(0)](t) = \lim_{n\to+\infty} (J_{t/n})^n[u(0)].$$

Uniqueness is now a consequence of the fact that

$$dist\Big(S[u(0)](t), S[v(0)](t)\Big) \leq e^{-\lambda t} dist(u(0), v(0)) \qquad \forall u(0), v(0) \in \overline{D(\phi)}.$$

Also, when $\lambda = 0$ and ϕ is nonnegative, they obtain the optimal a priori estimate (see (4.0.15))

$$dist^2\Big(S[u(0)](t), (J_{t/n})^n[u(0)]\Big) \leq \frac{t}{n}\phi(u(0)) \qquad \forall u(0) \in D(\phi).$$

2 Mass transport and weak riemannian structure on the set of probability measures

Section 5, reviews some basic measure-theoric tools. If X is a separable metric space, $\mathcal{B}(X)$ denotes the collection of Borel subsets of X and $\mathcal{P}(X)$ denotes the set of Borel probability measures on X. If $\mu \in \mathcal{P}(X)$, we say that μ is tight if for every every $\epsilon > 0$ there exists a compact set $K \subset X$ such that $\mu(X \setminus K) < \epsilon$. If X is a separable metric space then every $\mu \in \mathcal{P}(X)$ is tight, by Ulam theorem. More generally, we say that X is a Radon space if every $\mu \in \mathcal{P}(X)$ is tight.

In 1781, the great geometer G. Monge formulated a mass transport problem which consists in finding the optimal way to transport a pile of dirt to an excavation. Here, optimality is measured against the cost function $|x-y|$ which represents the cost of transporting a unit mass from the point x to the point y. It has long been suspected that Monge has formulated a central problem which would drastically impact mathematics. This suspicion went far beyond expectation as justified by the endless ramifications of the Monge problem in the past two hundred years. The Academy of Paris established the 'Prix Bordin' which was offered to Appell a hundred years later, for a formal solution to Monge problem. The title of [2] alone says much more about the importance of this problem than our short sell. We also refer the reader to [16] for more comments on the 'Prix Bordin' . The first impact of Monge's problem was first on developable surfaces. A century elapsed before a first rigorous solution to a relaxation of Monge's problem was proposed by Kantorovich as a problem in economics with applications in statistics and probability. In the past twenty years, the Monge-Kantorovich problem and its variants have been used as a powerful tool in partial differential equations, fluid mechanics. Many scientists from fields ranging from geometry to functional analysis, geophysics to kinetic theory have recognized that subject as a useful area which has advanced their research.

One of the goals of the second part of this book is to collect known and basic results of the Monge-Kantorovich theory and establish many other new ones, which have never appeared elsewhere. Especially, this book develops a rigorous setting which endows the set of probability measures with a weak Riemannian structure, formally introduced by Otto in his seminal paper [35]. This Riemannian structure is used to prove existence and uniqueness for parabolic partial differential equations. These equations can be expressed as gradient flows of functionals, on the set $\mathcal{P}(\mathbf{R}^d)$, of Borel probability measures on $\mathbf{R}^d$.

2.1 The mass transportation problem

Let X and Y be two Radon spaces and let $c : X \times Y \to [0, +\infty]$ be a Borel function. We call c a cost function and interpret $c(x, y)$ as the cost of transporting a unit mass from the point x to the point y. If $\mathbf{t} : X \to Y$ and $\mu \in \mathcal{P}(X)$, we define $\mathbf{t}_{\#}\mu$, the push forward measure of μ through $\mathbf{t}$ by

$$\mathbf{t}_{\#}\mu(B) = \mu\Big(\mathbf{t}^{-1}(B)\Big) \qquad B \in \mathcal{P}(Y).$$

Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the Monge problem is the following variational problem:

$$\inf_{\mathbf{t}} \Big\{ \int_X c(x, \mathbf{t}(x)) d\mu(x) : \ \mathbf{t}_{\#}\mu = \nu \Big\}. \tag{24}$$

At the time Monge first formulated his problem, measure theory was not a well-developed field and so, it is clear that he must be assuming to have 'nice' measures. For instance these measures could have been chosen to have densities. Also, the above formulation of the Monge problem is so general that it may be ill posed unless we impose additional conditions on the measures μ and ν. Indeed, there may not exist a Borel map $\mathbf{t} : X \to Y$ such that $\mathbf{t}_{\#}\mu = \nu$. For instance if $x \in X$, y_1, y_2 are two distinct elements of Y, there is no $\mathbf{t}$ such that $\mathbf{t}_{\#}\delta_x = 1/2(\delta_{y_1} + \delta_{y_2})$. It is a standard fact in measure theory that if that μ has no atoms ($\mu\{x\} = 0$ for all $x \in X$) then there exists a Borel map $\mathbf{t}$ that pushes μ forward to ν.

The first breakthrough on Monge's problem was achieved by Kantorovich in 1942 [26], [27], who formulated a linear programing problem and was awarded the nobel price for related works [34]. Kantorovich linear programing problem (25) turned out to be the relaxation of the Monge problem when μ does not have any atoms [21]. Let $\Gamma(\mu, \nu)$ be the set of Borel probability measures on $X \times Y$ that have μ and ν as their marginals: $\gamma(A \times Y) = \mu(A)$ and $\gamma(X \times B) = \nu(B)$ for all $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$. The so-called Monge-Kantorovich problem is:

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \Big\{ \int_{X \times Y} c(x, y) d\gamma(x, y) \Big\}. \tag{25}$$

It is easy to show that when c is lower semicontinuous then (25) admits a minimizer. If $\mathbf{t}_{\#}\mu = \nu$, and $\mathbf{id}_X$ denote the identity map on X, one can identify $\mathbf{t}$ with $(\mathbf{id}_X \times \mathbf{t})_{\#}\mu \in \Gamma(\mu, \nu)$. Since

$$\int_X c(x, \mathbf{t}(x)) d\mu(x) = \int_{X \times Y} c(x, y) d(\mathbf{id}_X \times \mathbf{t})_{\#}\mu(x, y)$$

we conclude that

$$\inf_{\mathbf{t}} \Big\{ \int_X c(x, \mathbf{t}(x)) d\mu(x) : \ \mathbf{t}_{\#}\mu = \nu \Big\} \geq \inf_{\gamma \in \Gamma(\mu,\nu)} \Big\{ \int_{X \times Y} c(x, y) d\gamma(x, y) \Big\}. \tag{26}$$

The support of $\gamma \in \mathcal{P}(X \times Y)$ is the smallest closed subset $K \subset X \times Y$ such that $\gamma(K) = 1$. A set $\Gamma \subset X \times Y$ is c–cyclically monotone if for all n integers, all $\{(x_i, y_i)\}_{i=1}^n \subset X \times Y$ and all σ permutation of n letters,

$$\sum_{i=1}^n c(x_i, y_i) \leq \sum_{i=1}^n c(x_i, y_{\sigma(i)}). \tag{27}$$

When for instance $X = Y = \mathbf{R}^d$ and $c(x, y) = |x - y|^2$, by a well known result of Rockafeller, c–cyclically monotone sets are characterized by the fact that they are contained in the subdifferential of convex functions. In theorem 6.1.4, the authors assume that

$$\mu\Big(\Big\{x \in X : \ \int_Y c(x, y) d\nu(y) < +\infty\Big\}\Big) > 0. \tag{28}$$

and

$$\nu\Big(\Big\{y \in Y : \ \int_Y c(x, y) d\mu(x) < +\infty\Big\}\Big) > 0. \tag{29}$$

The authors show that if $\gamma \in \Gamma(\mu, \nu)$ is optimal in (25) and $\int_{X \times Y} c(x, y) d\gamma(x, y) < +\infty$, then γ is concentrated on a c–cyclically monotone Borel set. If in addition c is continuous, then the support of γ is c–cyclically monotone. Conversely, if c is real-valued, $\gamma \in \Gamma(\mu,\nu)$ is concentrated on a c–cyclically monotone Borel set and (28 –29) hold, then γ is optimal and $\int_{X\times Y} c(x,y)d\gamma(x,y) < +\infty$. The current result is slightly finer than what was known in the literature.

A second breakthrough on Monge's problem was achieved by Brenier [3] in 1987 when he proved existence and uniqueness of an optimal map in (24), when $X = Y = \mathbf{R}^d$, $c(x,y) = ||x-y||^2$ and $\mu << \mathcal{L}^d$. That paper is responsible for the revival of the mass transport theory in the partial differential equations community. The topic has then stayed vibrant for the past 15 years. It has been applied to fluid mechanics [4], [6] and miraculously found a strong connection with the semigeostrophy systems introduced by Eliasssen [18] in 1948 and rediscovered by Hoskins [24] in 1975. The first studies of the semigeostrophy systems were done by meteorologists [8][14] [15] who were not at first aware of the mass transport theory. However, since the connection between these two fields has been observed, many of the recent works in the mathematics community [5] [7] [10] [11] [12] [13] [31] rely on the mass transport theory.

Let $C_b(X)$ be the set of continuous, bounded functions on X. Theorem 6.1.1 establishes that when c is proper and lower semicontinuous then a dual to (25) is

$$\sup_{(\varphi,\psi)} \Big\{ \int_X \varphi(x)d\mu(x) + \int_Y \psi(y)d\nu(y) \Big\}. \tag{30}$$

Here the infimum is performed over the set of pairs $(\varphi,\psi) \in C_b(X)\times C_b(Y)$ such that $\varphi(x)+\psi(y) \leq c(x,y)$ for all $(x,y) \in X\times Y$. We write $(\varphi,\psi) \in \mathcal{C}$. This duality result was probably due to Kellerer [28]. The proof in this book is different from the original proof and is succinct. Many other authors contributed to a better understanding of (24) in the case when $X = Y = \mathbf{R}^d$, $c(x,y) = ||x-y||^2$ and $\mu << \mathcal{L}^d$. A refined version of the work of Brenier, due to McCann [30], allows μ to vanish on $(d-1)$–rectifiable sets. A method for establishing existence of optimal maps, which solely focuses on understanding (30) without using as an ingredient that it is a dual to (24), was proposed by the reviewer of the current manuscript in [19]. That method was later applied to more general cost functions in [20] and [22]. A new geometric argument was introduced in [23] and is the basis of many of the methods used by various authors including the one of the current manuscript.

If the supremum in (30) is finite, (28 –29) hold and we further assume that c is real-valued, then there exists (φ,ψ) maximizer in (30). In that case, it is not a loss of generality to assume that the supremum is attained at an extreme point of $\mathcal{C}$ so that $\varphi = \psi^c$, and $\psi = \varphi^c$ where

$$\psi^c(x) = \inf_{y\in Y} c(x,y) - \psi(y). \tag{31}$$

By duality, if $\gamma \in \Gamma(\mu,\nu)$ is a minimizer in (25) we must necessary have that

$$\varphi(x) + \varphi^c(y) = c(x,y) \qquad \gamma \text{ a.e. in } X\times Y. \tag{32}$$

Indeed, by duality and the fact that $\gamma \in \Gamma(\mu,\nu)$ and $(\varphi,\varphi^c) \in \mathcal{C}$, we have

$$0 = \int_{X\times Y} c(x,y)d\gamma(x,y) - \int_{X\times Y} (\varphi(x)+\varphi^c(y))d\gamma(x,y) = \int_{X\times Y} |c(x,y)-\varphi(x)-\varphi^c(y)|d\gamma(x,y).$$

This proves (32). That equality is fundamental for proving existence of a minimizer in (24) when μ is absolutely continuous with respect to Lebesgue measure. To illustrate that fact, let us make that additional assumption on μ, in the remainder of this paragraph. The authors state existence of an optimal map for the Monge problem for cost functions of the form $c(x,y) = h(x-y)$ when $X = Y = \mathbf{R}^d$. They only imposed that $h : \mathbf{R}^d \to [0,+\infty)$ is strictly convex. They don't impose any condition on the growth of $h(z)$ as $|z|$ tends to infinity. That growth condition was required in [23], which provided the first existence of optimal maps for general cost functions. The expression in (31) is used to prove that φ is locally Lipschitz and so is differentiable μ–almost everywhere. Let $x \in \mathbf{R}^d$ be a point of differentiability of φ such that (32) holds for some y. Then, $z \to c(z,y) - \varphi(z) - \varphi^c(y) = l(z)$ attains its minimum at x. Thus, $0 \in \partial l(x) = \partial h(x-y) - \nabla\varphi(x)$ and so, $\nabla\varphi(x) \in \partial h(x-y)$. The strict convexity of h yields that $y = x - (\nabla h)^{-1}(\nabla\varphi(x))$. One can readily conclude that except for a set of zero measure, the support of γ is contained in the graph of the map $x \to \mathbf{t}(x) = x - (\nabla h)^{-1}(\nabla\varphi(x))$ and so, $\mathbf{t}_{\#}\mu = \nu$. One verifies that $\gamma = (\mathbf{id}_X \times \mathbf{t})_{\#}\mu$ and so,

$$\int_X c(x, \mathbf{t}(x))d\mu(x) = \int_{X\times Y} c(x,y)d\gamma(x,y).$$

This, together with (26) gives that $\mathbf{t}$ is a minimizer for (24). It is also clear that $\mathbf{t}$ is uniquely determined μ–almost everywhere since its explicit expression was dictated to us from (32). An approximate differential property of $\mathbf{t}$ is established in theorem 6.2.7. It is also shown that the eigenvalues of $\nabla\mathbf{t}$ are nonnegative when h satisfies appropriate conditions (e.g $h(z) = ||z||^p$, $p \in (1,+\infty)$.) A first proof of that weak regularity result on $\mathbf{t}$ and the nonnegativeness of $\nabla\mathbf{t}$ is due to Otto when h is smooth. We also refer the reader to Cordero [9].

The authors consider a more general case than the one we describe earlier by assuming that $X = Y$ are Hilbert spaces, $c(x,y) = ||x-y||^p = h(x-y)$ and $p \in (1,+\infty)$. Note that $(\nabla h)^{-1}$ exists and is continuous. A measure $\nu \in \mathcal{P}(X)$ is a nondegenerate Gaussian measure if $L_{\#}\nu \in \mathcal{P}(\mathbf{R})$ has a Gaussian distribution for all linear bounded forms on X. A set $B \in \mathcal{B}(X)$ is a Gaussian null set if $\mu(B) = 0$ for every nondegenerate Gaussian measure $\nu \in \mathcal{P}(X)$. A analogue of Radermacher's theorem ensures that if X is a Hilbert separable space, every locally Lipschitz map is Gateaux differentiable everywhere, except maybe on a Gaussian null set. A measure $\mu \in \mathcal{P}(X)$ is regular if $\mu(B) = 0$ for all Gaussian null sets. Let us denote the set of regular measures by $\mathcal{P}(X)^r$. The argument of the previous paragraph can be readily adapted to this Hilbert case, to obtain existence of an optimal map for the Monge problem.

2.2 Weak Riemannian structure on a subset of $\mathcal{P}(X)$

In this subsection, $X = Y$ is a separable Hilbert space endowed with the inner product $< \cdot; \cdot >$. If $\{e_i\}_{i=1}^{\infty}$ is an orthonormal basis of X, one can define on X a norm $||\cdot||^2_{\bar\omega}$ whose topology is weaker than the original topology. For instance

$$||x||^2_{\bar\omega} = \sum_{i=1}^{\infty} \frac{1}{i^2} < x; e_i >^2 .$$

Every bounded sequence for the original norm is precompact for $||\cdot||^2_{\bar\omega}$. For $p \in [1,+\infty)$ the authors consider the Monge-Kantorovich problem with the cost function $c(x,y) = ||x-y||^p$. For $\mu, \nu \in \mathcal{P}(X)$, we define

$$W_p^p(\mu,\nu) = \inf_{\gamma\in\Gamma(\mu,\nu)} \Big\{ \int_{X\times Y} ||x-y||^p d\gamma(x,y) \Big\}. \tag{33}$$

Let $L^p(\mu, X)$ denote the set of μ–measurable Borel maps $\mathbf{v} : X \to X$ such that $||v||_{L^p(\mu,X)} = \int_X ||v||^p d\mu < +\infty$. For $W_p^p(\mu,\nu)$ to be finite, it suffices to assume that $\mathbf{id}_X \in L^p(\mu,X) \cap L^p(\nu,X)$. Indeed, setting $\gamma = \mu\otimes\nu$ and using that $||x-y||^p \leq 2^{p-1}(||x||^p + ||y||^p)$, one concludes that

$$W_p^p(\mu,\nu) \leq 2^{p-1}(||\mathbf{id}_X||^p_{L^p(\mu,X)} + ||\mathbf{id}_X||^p_{L^p(\nu,X)}).$$

This fact suggests that the sets $\mathcal{P}_p(X)$ of $\mu \in \mathcal{P}(X)$ such that $\mathbf{id}_X \in L^p(\mu, X)$ will play an important role. For $p \in (1, +\infty)$ and $q = p/(p-1)$, the continuous map $j_p : X \to X$ defined by $j_p(\mathbf{v}) = \mathbf{v}||\mathbf{v}||^{p-2}$ yields a map from $L^p(\mu, X)$ onto $L^q(\mu, X)$. If $\mu, \nu \in \mathcal{P}_p(X)$, the set of γ minimizers in (33), is nonempty and we denote it by $\Gamma^p_{opt}(\mu, \nu)$. When $p = 2$, we simply denote that set by $\Gamma_{opt}(\mu, \nu)$. The authors generalize an interpolation introduced by McCann [29] in his influential PhD dissertation, to Hilbert spaces. If $\mu_o, \mu_1 \in \mathcal{P}_2(X)$ and $\gamma \in \Gamma^p_{opt}(\mu, \nu)$, one can define the interpolant path $t \to \mu_t \in \mathcal{P}_2(X)$ starting at μ_o and ending at μ_1 by

$$\mu_t = \pi_t \# \gamma, \qquad \pi_t(x, y) = (1-t)x + ty, \quad (x, y) \in X \times X.$$

One can readily check that $t \to \mu_t$ is a geodesic (of constant speed) such that

$$W_p(\mu_s, \mu_t) = (t-s) W_p(\mu_o, \mu_1) \qquad \forall 0 \leq s < t \leq 1. \tag{34}$$

In lemma 7.2.1, the authors prove the remarkable result that if $t \in (0, 1)$ then $\Gamma^p_{opt}(\mu_o, \mu_t)$ (resp. $\Gamma^p_{opt}(\mu_o, \mu_t)$) contains a unique element $\mu^{0\,t}$ (resp. $\mu^{t\,1}$). The plans $\mu^{t\,1}$ and $(\mu^{0\,t})^{-1}$ are induced by transport maps. In other words, their supports are contained in the graph of a Borel map.

In the current manuscript, the authors introduce the set $C^\infty_{fin,\,c}(X)$ which consists of the functions

$$\psi \circ \pi, \quad \psi \in C^\infty_c(\mathbf{R}^d), \quad \pi(x) = (< x; e_1 >, \cdots, < x; e_n >),$$

where $\{e_1, \cdots, e_n\} \subset X$ is an orthonormal family and n is a positive integer. One can show that $\psi \circ \pi \in C^\infty_{fin,\,c}(X)$ is continuous for $\bar{\omega}$ and its Fréchet gradient $\nabla \psi$ exists everywhere. In theorem 8.3.1, the authors prove the following amazing and central result: If $I \subset \mathbf{R}$ is an open interval and $t \to \mu_t$ is an absolutely continuous curve and $|\mu'|$ is its metric derivative, then there exists a Borel vector field $(x, t) \to \mathbf{v}_t(x)$ such that

$$\mathbf{v}_t \in L^p(\mu_t, X), \quad ||\mathbf{v}_t||_{L^p(\mu_t, X)} \leq |\mu'| \qquad \text{for } \mathcal{L}^1 \text{ a.e. } t \in I. \tag{35}$$

and the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mu_t \mathbf{v}_t) = 0 \qquad \in X \times I \tag{36}$$

holds in the sense of distribution. Moreover, for $\mathcal{L}^1$–a.e. $t \in I$, $j_p(\mathbf{v}_t)$ belongs to the closure in $L^q(\mu_t, X)$, of the subspace generated by the gradients $\nabla \psi$ with $\psi \in C^\infty_{fin,\,c}(X)$. We denote that space by $T_{\mu_t} \mathcal{P}_q(X)$. They prove that $\mathbf{v}_t$ is uniquely determined

Conversely, if a narrowly continuous curve $t \in I \to \mu_t \in \mathcal{P}_p(X)$ satisfies (36) for some Borel velocity field $\mathbf{v}_t$ with $||\mathbf{v}_t||_{L^p(\mu_t, X)} \in L^1(I)$, then $t \to \mu_t$ is absolutely continuous and

$$|\mu'| \leq ||\mathbf{v}_t||_{L^p(\mu_t, X)} \qquad \text{for } \mathcal{L}^1 \text{ a.e. } t \in I. \tag{37}$$

The authors prove another amazing generalized version of p–Hodge decomposition which we describe only when $p = 2$: if $\mathbf{u} \in L^2(\mu, X)$ then there exists a unique pair $(\mathbf{v}, \mathbf{w}) \in L^2(\mu, X) \times L^2(\mu, X)$ such that

$$\mathbf{u} = \mathbf{v} + \mathbf{w}, \ \ \mu \quad \text{a.e.} \qquad \mathbf{v} \in T_\mu \mathcal{P}_2(X), \qquad \text{div}_\mu \mathbf{w} = 0,$$

where div_μ is the linear functional on $L^2(\mu, X)$ defined by

$$< \mathbf{w}; \nabla \psi >_\mu =: \int_X < \mathbf{w}; \nabla \psi > d\mu = -\int_X (\text{div}_\mu \mathbf{w}) \psi d\mu \qquad \forall \psi \in C^\infty_{fin;\,c}(X).$$

In other words, $\mathbf{v} = \Pi(\mathbf{u})$ where Π is the orthogonal projection onto $T_\mu \mathcal{P}_2(X)$. They apply this Hodge decomposition to show in proposition 8.4.5 that the vector field $\mathbf{v}_t$ identified by (35) and (36), is uniquely determined and belongs to $T_{\mu_t}\mathcal{P}_2(X)$. It becomes then legitimate to write that

$$\mu_t' = \mathbf{v}_t$$

Exploiting now (37) one concludes that

$$\int_I |\mu'|^2 dt = \int_I ||\mu_t'||^2_{L^2(\mu_t, X)} dt. \tag{38}$$

While the left handside of (38) is expressed in terms of the metric derivative of μ_t, its right handside is expressed in terms of the $L^2(\mu_t, X)$–norm of $\mathbf{v}_t$. Taking (38) into account and using (8.0.2), one realizes that the authors have established that the space $\mathcal{P}_2(X)$ is endowed with a weak Riemannian structure, consistent with the metric W_2 in the following sense. If $\mu_o, \mu_1 \in \mathcal{P}_2(X)$, then

$$W_2^2(\mu_o, \mu_1) = \min_{\{\mu_t\}} \Big\{ \int_0^1 < \mu_t'; \mu_t' >_{\mu_t} dt : \quad t \to \mu_t \in AC^2(0,1;X),\ \mu_{t=0} = \mu_o, \mu_{t=1} = \mu_1 \Big\}. \tag{39}$$

2.3 Sudifferential and convexity of functions on $\mathcal{P}_2(X)$

In chapter 9, for functions $\phi : \mathcal{P}_p(X) \to (-\infty, +\infty]$, the authors defined various notions of convexity, including λ–convexity as defined in (12), along the geodesics. For instance, they show that if $\nu \in \mathcal{P}_2(X)$ then $-\frac{1}{2}W_2^2(\cdot, \nu)$ is (-1)–convex whereas $\frac{1}{2}W_2^2(\cdot, \nu)$ fails to be λ–convex for any $\lambda \in \mathbf{R}$. In lemma 9.1.4, they prove that if a sequence of λ–convex functionals Γ–converges, then the limit is also λ–convex. They introduce notions of *generalized geodesics* and *generalized geodesics with a base point* . We suggest that the reader should spend sometime on remark 9.2.8 which comments on the convexity properties of $\frac{1}{2}W_2^2(\cdot, \nu)$ with respect to these various geodesics.

The notion subdifferential of a functional $\phi : \mathcal{P}_2(X) \to (-\infty, +\infty]$ is crucial for studying ordinary differential equations on $\mathcal{P}_2(X)$. Since $\mathcal{P}_2(X)$ is neither a vector space, nor a Riemannian manifold in any classical sense, the definition of subdifferentials of functionals is not straightforward to guess. The authors propose various definitions of subdifferentials for the metric space $\mathcal{P}_p(X)$ in definition 10.1.1 and 10.3.1. Definition 10.1.1 is too restrictive and definition 10.3.1 requires additional notation. Because of that, I hereby give myself the liberty to suggest a definition (for $p = 2$) which appeared in a joint paper by the first author of the book under review and myself. If $\xi \in L^2(\mu, X)$ and $\phi(\mu) < +\infty$, they say that ξ belongs to $\partial\phi(\mu)$ the Fréchet subdifferential of ϕ at μ if

$$\phi(\nu) \geq \phi(\mu) + \sup_{\gamma \in \Gamma_{opt}(\mu,\nu)} \int_{X \times X} < \xi(x), y - x > d\gamma(x,y) + o(W_2(\mu, \nu),$$

as $\nu \to \mu$. The set $\partial\phi(\mu)$ is a convex closed subset of $\mathcal{P}_2(X)$ and so, admits an element of minimal norm which is necessary in $T_\mu \mathcal{P}_2(X)$. That element is simply denoted by $\nabla\phi(\mu)$; one could rather write $\nabla_{W_2}\phi(\mu)$ to distinguish it from gradient of functions defined on X.

With the notions of gradients and subdifferential of functions $\phi : \mathcal{P}_2(X) \to (-\infty, +\infty]$, the authors introduced a rigorous concept of gradient flow on $\mathcal{P}_p(X)$. This coincides with the notion of curves of maximal slopes studied in section 1.3. The fact that $\frac{1}{2}W_2^2(\cdot, \nu)$ is not λ–convex for any $\lambda \in \mathbf{R}$, led the authors to exploit its convexity along *generalized geodesics with a base point.* That property is used in chapter 11 to study existence and uniqueness of gradient flows solutions when the functional ϕ is λ–convex.

Conclusions. The importance of topics covered by this book and the care with which it is written, make it an excellent book of analysis/partial differential equations. Sections 5–11 complements another recent book on the Monge-Kantorovich theory, by C. Villani, which appeared in 2003 as a graduate text book in the AMS series. I anticipate that this book will establish itself for many years to come, as one of the main references on the geometry of the set of probability measures.

References

[1] L. Ambrosio and W. Gangbo. Hamiltonian ODE's in the Wasserstein space of probability measures. Submitted.

[2] P. Appell. Mémoire sur les déblais et les remblais des systèmes continues ou discontinues. Mémoire présentés par divers Savants à l'académie des sciences de l'Institut de France, Paris, I.N., 29, 1–208 (1887).

[3] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C.R. Acad. Sci. Paris Sér. I Math.*, 305:805–808 (1987).

[4] Y. Brenier. Derivation of the Euler equations from a caricature of Coulomb interaction. Comm. Math. Phys. 212, no. 1, 93–104 (2000).

[5] J.D. Benamou and Y. Brenier. Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère equations/transport problem. *SIAM J. Appl. Ana. Math.* **58**, no 5, 1450–1461 (1998).

[6] J.D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.* **84**, no. 3, 375–393 (2000).

[7] Y. Brenier and G. Loeper. A geometric approximation to the Euler equations: The Vlasov-Monge-Ampere equation. *Geom. Funct. Anal.*, (2004).

[8] P.Cloke and M.J.P. Cullen. A semi-geostrophic ocean model with outcropping . *Dyn. Atmos. Oceans*, 21:23–48 (1994).

[9] D. Cordero-Erausquin. Non-smooth differential properties of optimal transport. *AMS Contemp. Math.* Recent Advances in the Theory and Applications of Mass Transport. Summer School on Mass Transportation Methods in Kinetic Theory and Hydrodynamics, Vol 353, 61–71, Sep 04–09 2000.

[10] M.J.P. Cullen and M. Feldman. Lagrangian solutions of the semigeostrophic equations in physical space. *SIAM Jour. Analysis* vol. **37** issue 5, 1371–1395, 2006.

[11] M.J.P. Cullen and W. Gangbo. A variational approach for the 2-dimensional semi-geostrophic shallow water equations. *Arch. Ration. Mech. Anal.* **156** , 241–273 (2001).

[12] M.J.P. Cullen, W. Gangbo and G. Pisante. Semigeostrophic equations discretized in reference and dual variables. Submitted

[13] M.J.P. Cullen and H. Maroofi. The fully compressible semigeostrophic equations from meteorology. *Archive for Rational Mech. and Analysis* **167** no 4, 309–336 (2003).

[14] M.J.P. Cullen and R.J. Purser. An extended lagrangian theory of semigeostrophic frontogenesis. *J. Atmosph. Sciences* **41**, 1477–1497 (1984).

[15] M.J.P. Cullen and R.J. Purser. Properties of the lagrangian semigeostrophic equations. *J. Atmosph. Sciences* vol 40, **17**, 2684–2697 (1989).

[16] G. Darboux. Prix Bordin (géometrie) . *C.R. Acad. Sci. Paris* , 101 :1312–1316–808 (1885).

[17] E. De Giorgi. New problems on minimizing movements, in Boundary Value Problems for PDE and Applications. C. Baiocchi and J.L. Lions, *Eds Masson* vol 40, **17**, 81–98 (1993).

[18] A Eliassen. The quasi-static equations of motion. Geofys. Publ., **17**, No 3 (1948).

[19] W. Gangbo. An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational Mech. Anal.*, 128:381–399, (1994).

[20] W. Gangbo. Habilitations (Jan 1995): Xerox copy of the section on the Monge-Kantorovich theory. *http://www.math.gatech.edu/ gangbo/publications/*.

[21] W. Gangbo. The Monge Mass Transfer Problems and Its Applications. *NSF-CBMS Conference. Contemporary Mathematics*, Vol 226, 79–104 (1999).

[22] W. Gangbo and R.J. McCann. Optimal maps in Monge's mass transport problem. *C.R. Acad. Sci. Paris Sér. I Math.* **321** (1995) 1653–1658.

[23] W. Gangbo and R.J. McCann. The geometry of optimal transportation. *Acta Math.* **177**, 113–161 (1996).

[24] B.J. Hoskins. The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmosph. Sciences* **32**, 233–242 (1975).

[25] J. Jost. Nonpositive curvature: geometric and analytic aspects. Lectures in Mathematics, ETH Zurich, Birkhäuser, 1997.

[26] L. Kantorovich. On the translocation of masses. *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201 (1942).

[27] L. Kantorovich. On a problem of Monge (In Russian). *Uspekhi Math. Nauk.*, 3:225–226 (1948).

[28] H.G.. Kellerer. Duality theorems for marginal problems *Z. Wahrsch. Verw. Gebiete.*, 67 : 399–432 (1984).

[29] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.* **128** 153–179 (1997).

[30] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, **80** 309–323 (1995).

[31] R.J. McCann and A. Oberman. Exact semi-geostrophic flows in an elliptical ocean bassin . *Nonlinearity* **17** 1891–1922 (2004).

[32] U.F. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps *Comm. Ana. Geom.* **6** 199–253 (1998).

[33] G. Monge. Mémoire sur la théorie des déblais et de remblais. *Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année*, pages 666–704 (1781).

[34] B.S. Katz, editor. *Nobel Laureates in Economic Sciences: a Biographical Dictionary.* Garland Publishing Inc., New York (1989).

[35] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. P.D.E.*, **26**, 156–186 (2001).