Math 6112: Fall 2011. Advanced Linear Algebra

Luca Dieci [1]

December 7, 2011

[1]School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 U.S.A. (dieci@math.gatech.edu).Please bring mistakes to my attention, thanks!

Abstract

These notes are for the exclusive use of students in Math 6112 and cannot be distributed, copied, or posted. By and large, the material is not original and it is extracted from the books and references listed below. These notes have been collected so that students can avoid buying several books.

Contents

Chapter 1

Refresher and all the way to the Jordan Form

1.1 Preamble

Our goal in this Chapter is to arrive at obtaining the Jordan form of a linear transformation over an arbitrary field. Chances are that some of you have already seen the Jordan form of a matrix, most likely over the complex field. You will then appreciate the derivation of it over an arbitrary field, since it is obtained just about in the same way! At a high level, this is a consequence of the fact that elementary row operations keep us on the field we work in.

Still, to obtain the Jordan form, we will need to introduce the key concepts of eigenvalues and eigenvectors. No doubt we have all encountered these before, and we are quite used to define eigenvalues from the characteristic equation

$$\lambda : \quad \det(A - \lambda I) = 0 \to \textit{eigenvalues}$$

and associate eigenvector(s) to each eigenvalue, so that each eigenvalue has an associated algebraic and geometric multiplicity. Of course, this is absolutely fine, but for the time being we will pretend that we do not know this (yet).

If not the Jordan form itself, you have probably seen already the following prototypical result of *diagonalization* of matrices with real or complex entries by real valued or complex valued matrices, respectively:

$$A \in \begin{cases} \mathbb{R}^{n\times n}, & \text{with distinct real eigenvalues} \\ \mathbb{C}^{n\times n}, & \text{with distinct eigenvalues} \end{cases} .$$

Indeed, quite frequently, in introductory courses in Linear Algebra, these are the results which are presented, leaving the most misleading expectation that all matrices are diagonalizable! **But**, let me stress two key properties spelled out above, and whose relevance will soon be clarified: (a) Matrices were assumed to have distinct eigenvalues, and (b) Matrices with real entries were assumed to have real eigenvalues.

Observation 1.1.1 *As you know, or will soon see, $\mathbb{R}^{n\times n}$ can be thought of as a transformation of vectors from $\mathbb{R}^n$ to $\mathbb{R}^n$.*

- In the next few pages, we want to keep ourselves somewhat disconnected from the specificity of the *entries* in our matrices and obtain results independent of the numerical type of the entries in the matrices. To do this, we need to brush up on some abstraction.

  Our next goal is to introduce the concept of **vector space** V **over a field** $\mathbb{F}$ ($\mathbb{F}$ will always indicate a general field).

For the remaining part of this chapter, I have used material from [2] and [4].

1.2 Field

A field is a "commutative division ring".

1. *Ring* (associative, always). It is a set R with two operations "+" and "·" such that $\forall\, a, b, c \in R$
   (1) $a + b \in R$
   (2) $a + b = b + a$
   (3) $(a + b) + c = a + (b + c)$
   (4) $\exists\, 0 \in R : a + 0 = a$ (zero element)
   (5) $\exists\, -a \in R : a + (-a) = 0$ (opposite element)
   (6) $a \cdot b \in R$
   (7) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
   (8) $a \cdot (b + c) = a \cdot b + a \cdot c$, and $(a + b) \cdot c = a \cdot c + b \cdot c$
2. Moreover, if $a \cdot b = b \cdot a \Rightarrow$ it is called a *commutative ring*.
3. *Division ring*, if the nonzero elements form a group under multiplication.
   [NB: This implies that there is a *unit* element "1" in ring and every non-zero element a has an inverse, write it $1/a$.]

Examples 1.2.1 *Examples of fields.*

- *Familiar fields we work with:* $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ *(with the usual* $+$ *and* $\cdot$*). These are fields with infinitely many elements.*

- $\mathbb{Z}$ *(usual + and $\cdot$) is a commutative ring, it has unit element, but it is not a field ($\nexists$ "inverse" under multiplication).*
- $R = \{\mathbb{Z} \pmod 7\} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}\}$ *is a finite field.*
  (NB: if we had taken, say, $\{\mathbb{Z} \pmod 6\} \Rightarrow$ not a field.)

Recall. Remember that we say that a field $\mathbb{F}$ is of "*characteristic* 0" if it holds that $na \neq 0$, for any $a \in \mathbb{F}$, $a \neq 0$, and $n > 0$, any integer. On the other hand, $\mathbb{F}$ is called of *finite characteristic* n if there exist a smallest positive integer n such that $na = 0$, for all $a \in \mathbb{F}$.

Exercises 1.2.2

(1) *Give an example of a ring without unit element.*
(2) *Explain why $\{\mathbb{Z} \pmod 7\}$ is a field, but $\{\mathbb{Z} \pmod 6\}$ is not. [*Hint*: Can one have (or not) $a \cdot b = 0$ without $a = 0$ or $b = 0$?]*
(3) *Take the field $\{\mathbb{Z} \pmod 5\} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}\}$. Complete the addition and multiplication Tables. What is $\bar{1} - \bar{2}$?*
(4) *Give an example of a field of characteristic* 2.
(5) *Show that if $\mathbb{F}$ is of finite characteristic n, then n is a prime number.*

1.3 Vector Space

This will always be indicated as V, over $\mathbb{F}$ (V contains the vectors, $\mathbb{F}$ the scalars).

V is a set ($\neq \emptyset$) with "addition" + such that with respect to +, V is an Abelian group. That is, for all $w, v, z \in V$ we have

$$\begin{aligned} v + w &= w + v \in V \\ \exists\, 0 : \ v + 0 &= 0 + v = v \\ v + (-v) &= 0 \\ v + (w + z) &= (v + w) + z \end{aligned}$$

Moreover, $\forall\, \alpha \in \mathbb{F}$, $v \in V \Rightarrow \alpha v \in V$ (multiplication by a scalar) and $\forall\, \alpha, \beta \in \mathbb{F}$:

$$\alpha(v + w) = \alpha v + \alpha w, \quad (\alpha + \beta)v = \alpha v + \beta v, \quad \alpha(\beta v) = \alpha\beta v, \quad 1\ v = v\ .$$

Exercise 1.3.1 *Refresh yourself with the concepts of linear independence, bases, dimensions, subspaces, etc., for a vector space V.*

Agreement. For us, V will always be **finite dimensional**, the default dimension of V being the value n.

Example 1.3.2 *Examples of familiar vector spaces are:* $V = \mathbb{R}^n$ *(*$\mathbb{F} = \mathbb{R}$*),* $V = \mathbb{C}^n$ *(*$\mathbb{F} = \mathbb{C}$ *(technically, also* $\mathbb{R}$ *is legitimate)),* $V = \mathbb{Q}^n$ *(*$\mathbb{F} = \mathbb{Q}$*). Also the set of real valued* (m, n) *matrices* $V = \mathbb{R}^{m \times n}$*, etc.. To avoid unneeded confusion, we will henceforth restrict to the case in which the elements of the set* V *have numerical type from the field* $\mathbb{F}$*; e.g.,* $V = \mathbb{F}^{m \times n}$*.*

Definition 1.3.3 *Consider a mapping between vector spaces* V *and* W *(both over* $\mathbb{F}$*):*

$$T : V \to W\,.$$

If

(1) $T(v_1 + v_2) = Tv_1 + Tv_2$*, for all* $v_1, v_2 \in V$*,*
(2) $T(\alpha v) = \alpha T v$*, for all* $v \in V$ *and* $\alpha \in \mathbb{F}$*,*
$\Rightarrow T$ *is called homomorphism (or linear mapping, or linear transformation). The set of all linear mappings between* V *and* W *is written as Hom*(V, W)*.*

- Very important are the linear maps of V into itself, $T : V \to V$. These are elements of $\mathrm{Hom}(V, V)$.
- It is simple to see that under the obvious operation of "+", $\mathrm{Hom}(V, W)$ is itself a vector space over $\mathbb{F}$:

$$(T + S)(v) = Tv + Sv, \quad T(\alpha v) = \alpha T(v)\,.$$

Agreement. We will call two mappings S and T *equal*, and write $S = T$, if S and T agree on a basis of V.

Theorem 1.3.4 *Hom*(V, W) *is a vector space of dimension* mn*, where* $n = \dim(V)$ *and* $m = \dim(W)$*. In particular,* $\dim\big(Hom(V, V)\big) = n^2$*.*

Pf. We build a basis of $\mathrm{Hom}(V, W)$. Let $\{v, \dots v_n\}$ be a basis for V, and $\{w_1, \dots, w_m\}$ be a basis for W. For $i = 1, \dots, m$, $j = 1, \dots, n$, define

$$T_{ij} : V \to W \quad \text{s.t. if } v \in V\ ,\ \ v = \lambda_1 v_1 + \cdots + \lambda_n v_n \Rightarrow T_{ij} v = \lambda_j w_i\ ,$$

that is:

$$T_{ij}v_k = \begin{cases} 0, & k \neq j \\ w_i, & k = j \end{cases}$$

There are mn such T_{ij}. Now we show that they are a basis, that is:

(a) They span $\mathrm{Hom}(V, W)$ (over $\mathbb{F}$);

(b) they are linearly independent.

Take any $S \in \mathrm{Hom}(V, W)$. Now, for $i = 1 : n$, take $Sv_i \in W \Rightarrow Sv_i = \alpha_{1i}w_1 + \cdots + \alpha_{mi}w_m$ ($\alpha_{ji} \in \mathbb{F}$). Also, take $S_0 = \alpha_{11}T_{11} + \alpha_{21}T_{21} + \cdots + \alpha_{m1}T_{m1} + \alpha_{12}T_{12} + \cdots + \alpha_{n2}T_{n1} + \cdots + \alpha_{mn}T_{mn}$. Then: $S_0v_i = (\alpha_{11}T_{11} + \cdots + \alpha_{m1}T_{m1} + \cdots + \alpha_{mn}T_{mn})v_i = \alpha_{1i}T_{1i}v_i + \alpha_{2i}T_{2i}v_i + \cdots + \alpha_{mi}T_{mi}v_i = \alpha_{1i}w_1 + \cdots + \alpha_{mi}w_m$.

$\therefore$ $Sv_i = S_0v_i$, $i = 1 : n$ and S is arbitrary, $\therefore$ S and S_0 agree on a basis of V, and $\therefore$ $\{T_{ij}\}$ span.

Next, show they are linearly independent. Suppose not. Then, there $\exists\ \beta_{ij} \in \mathbb{F}$, $i = 1, \ldots, m; j = 1, \ldots, n$, not all 0 such that $\sum_{i,j} \beta_{ij}T_{ij} = 0$. For any $i = 1, \ldots, n$, we then have

$$0 = (\sum_{i,j} \beta_{ij}T_{ij})v_i = \beta_{1i}w_1 + \cdots \beta_{mi}w_m \ ,$$

but $\{w_1, \ldots, w_m\}$ are linearly independent $\therefore$ $\beta_{1i} = \cdots = \beta_{mi} = 0$, for all $i = 1 : n$. Contradiction. $\square$

Examples 1.3.5 *$V = \mathbb{R}^n$, $W = \mathbb{R}^m$; Hom$(V, W) = \{\mathbb{R}^n \to \mathbb{R}^m\}$ (which will soon be identified with $\mathbb{R}^{m\times n}$), has dimension mn. $V = W = \mathbb{C}^n$, Hom$(V, V) = \{\mathbb{C}^n \to \mathbb{C}^n\}$ (which will soon be identified with $\mathbb{C}^{n\times n}$), has dimension n^2 (over $\mathbb{C}$).*

Remark 1.3.6 *There exists $I \in$ Hom(V, V) such that $IT = TI = T$ (identity element). This is simply because there exists I such that $Iv = v$, $\forall\ v \in V$; just take $\alpha_{ij} = 0$, $j \neq i$, $\alpha_{ii} = 1$ in previous construction. [In other words, Hom(V, V) is an algebra with unit element.]*

Definition 1.3.7 *A linear transformation T is invertible –or nonsingular– if $\exists\ S \in$ Hom$(V, V) : TS = ST = I$. We write it as $S = T^{-1}$. If T is not invertible, it is also called singular.*

1.3.1 Minimal Polynomial

An important consequence of the fact that $\mathrm{Hom}(V, V)$ is a vector space is that "Given any element $T \in \mathrm{Hom}(V, V)$, there is a nontrivial polynomial $q(x) \in \mathbb{F}[x]$ of degree at most n^2 such that $q(T) = 0$". Let us verify this fact.
Pf. Take $T \in \mathrm{Hom}(V, V)$, $T \neq 0$. Form $I, T, T^2, \ldots, T^{n^2}$. These are $n^2 + 1$ elements $\Rightarrow$ they are linearly dependent.
$\Rightarrow \exists\, \alpha_0, \alpha_1, \ldots, \alpha_{n^2} \in \mathbb{F}$ (not all 0) such that $\alpha_0 I + \alpha_1 T + \cdots + \alpha_{n^2} T^{n^2} = 0$ that is, T satisfies the nontrivial polynomial $q(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_{n^2} x^{n^2}$. $\square$

Definition 1.3.8 *The monic*[1] *polynomial of lowest degree $p(x)$ such that $p(T) = 0$ is called minimal polynomial for T (over $\mathbb{F}$).*

Exercises 1.3.9 *(Most from [2].) Below, $T \in Hom(V, V)$ and V is n-dimensional, unless otherwise stated.*

(1) Show that the minimal polynomial of T is unique. [Note that there must exist one, since V is finite dimensional.]
*(2) Show that $T \in Hom(V, V)$ is invertible $\Leftrightarrow$ the constant term of the minimal polynomial is not 0. [*Hint*: Seek directly a form for T^{-1}.]*
(3) Show that T singular $\Leftrightarrow \exists\, v \in V,\ v \neq 0\ : Tv = 0$.
(4) This is about nonsingular linear maps.
  (a) Prove that the nonsingular elements in $Hom(V, V)$ form a group.
  *(b) Let $\mathbb{F} = \{\mathbb{Z} \pmod 2\}$, and let V be 2-dimensional over $\mathbb{F}$. Compute the group of nonsingular elements in $Hom(V, V)$. [*Hint*: Consider S_3, the symmetric group of order 3.]*
(5) The transformation $T \in Hom(V, V)$ is called nilpotent, *if $T^k = 0$, for some (integer) $k > 0$. [$T^k(v)$ is defined as the action $T(T \cdots (T(v)) \cdots)$ k-times. For completeness, we also define $T^0 = I$.]*
  (a) Show that if T is nilpotent and $Tv = \alpha v$, for some $v \neq 0$ and $\alpha \in \mathbb{F}$, then $\alpha = 0$.
  (b) Show that if T is nilpotent, that is $T^k = 0$ for some k, then $T^n = 0$.
  (c) Show that if T is nilpotent and $\alpha_0 \neq 0$, then $S = \alpha_0 I + \alpha_1 T + \cdots + \alpha_p T^p$ is invertible.
(6) Show that if T satisfies a polynomial $q(x) \in \mathbb{F}[x]$, and S is invertible, then also $S^{-1}TS$ satisfies $q(x)$. In particular, T and $S^{-1}TS$ have same minimal polynomial.

[1] The leading coefficient is $1 \in \mathbb{F}$

1.4 Eigenvalues and Eigenvectors

Here, let $T \in \mathrm{Hom}(V, V)$ and V is n-dimensional.

Definition 1.4.1 *$\lambda \in \mathbb{F}$ is eigenvalue of T if $T - \lambda I$ is singular. [NB: we are asking $\lambda \in \mathbb{F}$].*
A vector $v \in V$, $v \neq 0$, is an eigenvector of T associated to the eigenvalue $\lambda \in \mathbb{F}$, if $Tv = \lambda v$.

Exercises 1.4.2
(1) $\lambda \in \mathbb{F}$ is an eigenvalue of $T \Leftrightarrow$ for some $v \neq 0$, $v \in V$, $Tv = \lambda v$. In particular, 0 is an eigenvalue of T if and only if T is singular.
*(2) Show that if $\lambda \in \mathbb{F}$ is eigenvalue of $T \in Hom(V, V)$, then λ is a root of the minimal polynomial. [*Hint*: You know that $P(T) = 0$.]*
NB*: As a consequence of this exercise, T has finitely many eigenvalues in $\mathbb{F}$.*
(3) If $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}$ are distinct eigenvalues of $T \in Hom(V, V)$ and $v_1, v_2, \ldots v_k$ are the associated eigenvectors, then $v_1, v_2, \ldots, v_k$ are linearly independent.

A consequence of this last exercise is that:
"If $\dim(V) = n \Rightarrow T$ has at most n distinct eigenvalues in $\mathbb{F}$".
(Any k distinct eigenvalues correspond to k linearly independent eigenvectors. But V is n-dimensional.)

Further, this last fact can be also reformulated as saying:

Basis of Eigenvectors
"If $\dim(V) = n$ and T has n distinct eigenvalues in $\mathbb{F} \Rightarrow V$ has a basis of eigenvectors of T".

1.5 Matrices and canonical forms

As you may have seen before, to a linear transformation there is associated a matrix. Let us see this correspondence.

Consider $T \in \mathrm{Hom}(V, W)$, and let V be n-dimensional with basis $\{v_1, \ldots, v_n\}$, and W be m-dimensional with basis $\{w_1, \ldots, w_m\}$. Then, we have

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i \ , \ \ j = 1, \ldots, n \, ,$$

that is we associate to T the matrix $A = (a_{ij})_{i=1:m,j=1:n} \in \mathbb{F}^{m\times n}$. In the particular case of $T \in \mathrm{Hom}(V,V)$, we obtain a *square* matrix $A = (a_{ij})_{i=1:n,j=1:n} \in \mathbb{F}^{n\times n}$.

Remark 1.5.1 *By virtue of the above correspondence, we will freely identify linear transformations and matrices. Any result for matrices can be equivalently phrased for linear transformations. Definitions of eigenvalues and eigenvectors are likewise transported immediately to the matrix setting.*

Exercise 1.5.2 *Let $\{e_i^{(n)}\}_{i=1,\dots,n}$ and $\{e_i^{(m)}\}_{i=1,\dots,m}$ be the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. (As usual, these are the standard unit vectors: $\left(e_i^{(n,m)}\right)_j = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$, the only difference being how many entries there are in $e_i^{(n)}$ and $e_i^{(m)}$.)*

Find the basis $\{T_{ij}\}$ of which in Theorem 1.3.4 for Hom(V,W) (equivalently, for $\mathbb{R}^{m\times n}$).

Remark 1.5.3 *We stress that the matrix associated to T depends on the basis chosen for V. So, while the transformation viewed as mapping from V to V, say, is not ambiguous, the matrix representation of T is.*

The last remark is actually deeper than it sounds, because it suggests an **Idea**: "*Choose appropriate bases to get nice matrices*"!

What precisely is meant by "nice" is in the eye of the beholder, but let us tentatively agree that a nice matrix is one which has a form which reveals important information, such as the eigenvalues, or the rank, etc..

As a matter of fact, this innocent sounding consideration above is at the heart of much linear algebra: we usually (for convenience, simplicity, or necessity) use simple bases to express linear transformations, and end up with matrices which are not necessarily revealing structural information on the transformation itself. Afterwards, we try to transform these matrices into a nice form which reveals structural information on the transformation. The operations we do (see Theorem 1.5.5) are tantamount to changing bases. If we were able to guess the right basis before associating a matrix to the transformation, there would be little scope for further work!

Example 1.5.4 *For example, a nice matrix A, when T has n distinct eigenvalues* *in $\mathbb{F}$, will be obtained upon realizing that relatively to its eigenvectors we have*

$$\text{For } i=1,\dots,n: \; Tv_i = \lambda_i v_i \Rightarrow \text{ in this basis } A = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

We will consider this diagonal *form very nice indeed. As a matter of fact, we may as well agree that a nice matrix is one which is as close as possible to being diagonal.*

- An important result is that matrices representing the same linear transformation in different bases, are *similar* matrices. That is:

Theorem 1.5.5 *Let $T \in Hom(V,V)$, and suppose that A is the matrix associated to T in the basis $\mathcal{V} = \{v_1,\dots,v_n\}$, and B in the basis $\mathcal{W} = \{w_1,\dots,w_n\}$. Then, there exists $C \in \mathbb{F}^{n\times n}$, invertible, such that $B = C^{-1}AC$.[2] The matrix C is the matrix associated to the linear transformation S which changes bases: $Sw_i = v_i$, $i = 1:n$.*

Exercises 1.5.6

(1) *Prove Theorem 1.5.5.*
(2) *Show that similarity is an equivalence relation, and that it does not change eigenvalues of T.*

As a consequence, we have

Corollary 1.5.7 (First Canonical Form: Diagonalization over $\mathbb{F}$.) *If $A \in \mathbb{F}^{n\times n}$ has n distinct eigenvalues $\lambda_1,\dots,\lambda_n$ in $\mathbb{F}$, then $AV = V\Lambda$, where* $\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$ *and $V = [v_1,\dots,v_n]$ is the matrix of the eigenvectors.*

Remark 1.5.8 *Corollary 1.5.7 expresses a sufficient condition, but naturally a matrix may be diagonalizable even if it has repeated eigenvalues. All one needs to have are n linearly independent eigenvectors.*

[2] This is called a *similarity transformation* and A and B are called similar matrices.

Exercises 1.5.9

(1) *Show that the only matrices in $\mathbb{F}^{n\times n}$ commuting with all matrices in $\mathbb{F}^{n\times n}$ are αI.*

(2) *Show that there cannot be $A, B \in \mathbb{R}^{n\times n} \,:\, AB - BA = I$.*

(3) *Show that if $A \in \mathbb{F}^{n\times n}$, $A = \begin{pmatrix} \alpha & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ 0 & & & \alpha \end{pmatrix}$, $\alpha \neq 0$, then A is invertible and find its inverse.*

(4) *Let $A \in \mathbb{F}^{n\times n}$, and suppose $A^k = 0$ for some k. Show that $I + A$ is invertible and find $(I + A)^{-1}$.*

(5) *Let $M \in \mathbb{F}^{n\times n}$ be $M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ with $A \in \mathbb{F}^{n_1\times n_1}$ and $B \in \mathbb{F}^{n_2\times n_2}$, $n_1 + n_2 = n$. Then, show that C is diagonalizable (over $\mathbb{F}$) if and only if A and B are.*

The next theorem characterizes, fully, the situation when two matrices can be diagonalized by the same similarity transformation.

Theorem 1.5.10 (Simultaneous Diagonalizability.) *Let $A, B \in \mathbb{F}^{n\times n}$ be diagonalizable (over $\mathbb{F}$). Then, they are simultaneously diagonalizable if and only if $AB = BA$.*

Pf. Suppose there exist V such that $V^{-1}AV = D_A$ and $V^{-1}BV = D_B$, with D_A and D_B both diagonal. Then, obviously $D_AD_B = D_BD_A$ and thus

$$VD_AV^{-1}VD_BV^{-1} = VD_BV^{-1}VD_AV^{-1} \Rightarrow AB = BA\,.$$

Next, suppose that $AB = BA$ and let V be the similarity that diagonalizes A: $V^{-1}AV = D_A$, ordering the eigenvalues so that equal eigenvalues appear consecutively along the diagonal of D_A: $D_A = \operatorname{diag}(\lambda_k I_{n_k},\, k = 1, \ldots, p)$. Let $C = V^{-1}BV$, so that $AB = BA$ rewrites as $D_AC = CD_A$. Partitioning C in block form conformally to D_A's partitioning, we immediately get that $C = \operatorname{diag}(C_{kk},\, k = 1, \ldots, p)$. Since B is diagonalizable, so is C, and thus so are the blocks C_{kk}, $k = 1, \ldots, p$. Let T_k be the similarities diagonalizing these blocks C_{kk}, $k = 1, \ldots, p$, and let $T = \operatorname{diag}(T_k,\, k = 1, \ldots, p)$, so that $T^{-1}CT$ is diagonal. Now, given the block diagonal structure of T, we also have $T^{-1}D_AT = D_A$ and so:

$$T^{-1}\big(V^{-1}AV\big)T \quad \text{and} \quad T^{-1}\big(V^{-1}BV\big)T$$

are both diagonal. $\square$

Although diagonal form is very nice, we may not be able to achieve it for a certain transformation. In other words, for a certain given transformation T, there may be no basis leading to a diagonal matrix representation of T. The next "nice" form we look at is *triangular* (upper). This means that $R \in \mathbb{F}^{n\times n}$, $R = (r_{i,j})_{i,j=1,\dots,n}$ is such that $r_{ij} = 0$, $i > j$.

Fact 1.5.11 (Backward substitution) *If $R \in \mathbb{F}^{n\times n}$ is triangular and nonsingular, then we can always uniquely solve the linear system*

$$Rx = b, \quad \text{for } b \in \mathbb{F}^n \quad \text{getting } x \in \mathbb{F}^n\,.$$

Remark 1.5.12 *This is a very useful result. The algorithm to obtain x is called "backward substitution." Key ingredients are nonsingularity and the fact we are doing the arithmetic in a field.*

Exercises 1.5.13

(1) Develop the backward substitution algorithm of which above. Convince yourself that all arithmetic operations keep you on the field $\mathbb{F}$.

(2) Show that if R is triangular $\in \mathbb{F}^{n\times n}$, then its eigenvalues are precisely the elements on diagonal of R, and only these.

(3) Suppose that $M \in \mathbb{F}^{n\times n}$ is in the block form $M = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$, where $A \in \mathbb{F}^{n_1\times n_1}$ and $B \in \mathbb{F}^{n_2\times n_2}$, $n_1 + n_2 = n$, and that M has eigenvalues in $\mathbb{F}$. Then, show that $\lambda \in \mathbb{F}$ is an eigenvalue of M if and only if it is an eigenvalue of either A or B, or both.

- The next result tells us that "If $T \in \mathrm{Hom}(V, V)$ has all eigenvalues in $\mathbb{F}$, then there is a basis of V in which T is triangular" (i.e., its matrix representation in this basis is). This result is effectively akin to the famous Schur's form of a matrix.

Theorem 1.5.14 (Second Canonical Form: Triangularization over $\mathbb{F}$.) *Suppose $A \in \mathbb{F}^{n\times n}$ has all its eigenvalues in $\mathbb{F}$. Then, there is an invertible matrix $B \in \mathbb{F}^{n\times n}$ such that $R = B^{-1}AB$ is (upper) triangular. Moreover, we can arrange the entries on the diagonal of R (the eigenvalues) in any order we like.*

Pf. Suppose we have eigenvalues $\lambda_1, \lambda_2, \ldots,$ and that this is the ordering we are after.

The proof is by induction on n. Let $c_1 \neq 0$, $Ac_1 = \lambda_1 c_1$. Complete c_1 to a basis:

$$\Rightarrow A \underbrace{[c_1, \tilde{c}_2, \ldots, \tilde{c}_n]}_{E} = \underbrace{[c_1, \tilde{c}_2, \ldots, \tilde{c}_n]}_{E} \begin{bmatrix} \lambda_1 & * \cdots * \\ 0 & C \end{bmatrix}$$

Now, eigenvalues of C must be also eigenvalues of A (because $E^{-1}AE = \begin{pmatrix} \lambda_1 & * \\ 0 & C \end{pmatrix}$). By induction, $\exists\ D$ such that $D^{-1}CD$ is triangular with eigenvalues on the diagonal ordered as we please. But then

$$\underbrace{A\, E \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix}}_{B} = \underbrace{E \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & D \end{bmatrix}}_{R}$$

and R is upper triangular of the type we wanted (note that B is invertible since it is the product of invertible matrices.) $\square$

Corollary 1.5.15 *If all eigenvalues of A (equivalently, of T) are in $\mathbb{F}$, then $\exists$ a polynomial $q(x) \in \mathbb{F}[x]$ of degree n such that $q(A) = 0$ (equivalently, $q(T) = 0$).*

Pf. Take basis given by columns of B in Theorem 1.5.14, call these $v_1, \ldots, v_n$:

$$A[v_1, \ldots, v_n] = [v_1, \ldots, v_n] \begin{bmatrix} \lambda_1 & r_{12} & \cdots & r_{1n} \\ & \lambda_2 & \cdots & \vdots \\ 0 & & \ddots & \lambda_n \end{bmatrix} :$$

$$\begin{cases} Av_1 = \lambda_1 v_1 \\ Av_2 = r_{12} v_1 + \lambda_2 v_2 \\ \quad \vdots \\ Av_i = r_{1i} v_1 + r_{2i} v_2 + \cdots + r_{i-1,i} v_{i-1} + \lambda_i v_i, \quad i = 1, 2, \ldots, n \end{cases}$$

$$\Rightarrow \begin{cases} (A - \lambda_1 I) v_1 = 0 \\ (A - \lambda_2 I) v_2 = r_{12} v_1 \\ \quad \vdots \\ (A - \lambda_i I) v_i = r_{1i} v_1 + r_{2i} v_2 + \cdots + r_{i-1,i} v_{i-1}, \quad i = 1, 2, \ldots, n. \end{cases}$$

Now, observe that

$$(A-\lambda_1 I)(A-\lambda_2 I) = (A-\lambda_2 I)(A-\lambda_1 I) \quad \text{etc.},$$

$\therefore$ $(A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_i I)v_k = 0, \quad k = 1,\dots,i$
$\therefore$ $\underbrace{(A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_n I)}_{S}v_k = 0, \quad k = 1:n$
$\therefore$ S annihilates a basis of V $\therefore$ $S \equiv 0$
$\therefore$ A satisfies the polynomial equation $(x-\lambda_1)\cdots(x-\lambda_n) = 0$. $\square$

Remarks 1.5.16

(1) *To get triangular form, we did not assume distinct eigenvalues, but still assumed all eigenvalues in $\mathbb{F}$.*
(2) *The polynomial in Corollary 1.5.15 is effectively the* characteristic polynomial, *to be introduced below. Corollary 1.5.15, then, is the celebrated* Cayley–Hamilton *theorem.*

- Our next step will be to further simplify, if possible, the triangular structure we achieved. (Remember that we can get diagonal over $\mathbb{F}$ if all eigenvalues are distinct and in $\mathbb{F}$). The key step will be developed in Exercise 1.5.18 below.

Exercise 1.5.17 *This is a referesher of frequent matrix manipulations.*

- *Block multiplication. We will make repeated use of multiplying matrices partitioned in compatible blocks. For example, let $A_{ii}, B_{ii} \in \mathbb{F}^{n_i,n_i}$, $i = 1,2$, and let $A_{12}, B_{12} \in \mathbb{F}^{n_1,n_2}$ and $A_{21}, B_{21} \in \mathbb{F}^{n_2,n_1}$, then:*

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{pmatrix}.$$

- *Transpose. If $A \in \mathbb{F}^{m\times n}$ with entries $(A)_{ij} = a_{ij}$, then $A^T \in \mathbb{F}^{n\times m}$ is the matrix with entries $(A^T)_{ij} = a_{ji}$. A^T is called* transpose *of A.*

Exercise 1.5.18 *We are going to put together several of the previous ingredients. Given $R \in \mathbb{F}^{(n+m)\times(n+m)}$, $R = \begin{pmatrix} A & -C \\ 0 & B \end{pmatrix}$ where $C \in \mathbb{F}^{n\times m}$, and*

$$A \in \mathbb{F}^{n\times n} = \begin{pmatrix} \lambda & a_{12} & \cdots & a_{1n} \\ & \lambda & \cdots & \vdots \\ 0 & & \ddots & \lambda \end{pmatrix}, \quad B \in \mathbb{F}^{m\times m} = \begin{pmatrix} \mu & b_{12} & \cdots & b_{1m} \\ & \mu & \cdots & \vdots \\ 0 & & \ddots & \mu \end{pmatrix}, \quad \lambda \neq \mu.$$

Then, $\exists$ *unique* $X \in \mathbb{F}^{n\times m}$ *such that* $\begin{pmatrix} I_n & -X \\ 0 & I_m \end{pmatrix}\begin{pmatrix} A & -C \\ 0 & B \end{pmatrix}\begin{pmatrix} I_n & X \\ 0 & I_m \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$.

Solution: Note that X must solve

$$AX - XB = C\ .$$

Rewrite this matrix equation as

$$\begin{pmatrix} \lambda & a_{12} & \cdots & a_{1n} \\ & \ddots & \ddots & \vdots \\ & & \lambda & a_{n-1,n} \\ & & & \lambda \end{pmatrix}\begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix} - \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}\begin{pmatrix} \mu & b_{12} & \cdots & b_{1m} \\ & \ddots & \ddots & \vdots \\ & & \mu & b_{m-1,m} \\ & & & \mu \end{pmatrix}$$
$$= \begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nm} \end{pmatrix}$$

Let us solve for the entries of X from the bottom to the top, and left to right.

From n-th row of X

$$\begin{aligned} \text{1st column} &: \lambda x_{n1} - \mu x_{n1} = c_{n1} \to x_{n1} = c_{n1}/(\lambda-\mu) \\ \text{2nd column} &: \lambda x_{n2} - x_{n1}b_{n2} - x_{n2}\mu = c_{n2} \to x_{n2} = (c_{n2} + x_{n1}b_2)/(\lambda-\mu) \\ &\vdots \\ m\text{-th column} &: \Rightarrow x_{nm} \end{aligned}$$

$\Rightarrow$ n-th row of X found.

Next, use the $(n-1)$-st row X. 1st column: $\lambda x_{n-1,1} + a_{n-1,n}x_{n1} - x_{n-1,1}\mu = c_{n-1,1} \to x_{n-1,1}$. Keep going. $\square$

Exercise 1.5.19 *Given* $R = \begin{array}{c|c} & \begin{matrix} n_1 & n_2 & \cdots & n_p \end{matrix} \\ \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_p \end{matrix} & \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ & R_{22} & \ddots & \vdots \\ & & \ddots & R_{p-1,p} \\ & O & & R_{pp} \end{pmatrix} \end{array}$ *where* $R_{ii} = \begin{bmatrix} \lambda_i & * & \cdots & * \\ & \ddots & & \vdots \\ & & & * \\ & O & & \lambda_i \end{bmatrix}$,

$i = 1, \ldots, p$, *and* $\lambda_i \neq \lambda_j$, $i \neq j$. *Then, there exist -unique-* $T = \begin{bmatrix} I_{n_1} & X_{12} & \cdots & X_{1p} \\ & I_{n_2} & \ddots & \vdots \\ & & \ddots & \\ & O & & I_{n_p} \end{bmatrix}$,

$X_{ij} \in \mathbb{F}^{n_i \times n_j}$, *such that* $T^{-1}RT = \begin{bmatrix} R_{11} & & O & \\ & R_{12} & & \\ & & \ddots & \\ & O & & R_{pp} \end{bmatrix}$.

As a consequence of Exercise 1.5.19, we have

Corollary 1.5.20 (Block Diagonal Form.) *If* $A \in \mathbb{F}^{n\times n}$ *has all eigenvalues in* $\mathbb{F}$, *then there* $\exists\, T \in \mathbb{F}^{n\times n}$, *invertible, such that* $T^{-1}AT$ *is block diagonal,* $\begin{pmatrix} R_{11} & & 0 \\ & \ddots & \\ 0 & & R_{pp} \end{pmatrix}$, *where each* R_{ii} *is upper triangular with constant diagonal:* $R_{ii} = \begin{pmatrix} \lambda_i & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ 0 & & & \lambda_i \end{pmatrix}$, *and* $\lambda_i \neq \lambda_j$, $i \neq j$.

- Is this as "simple" a form as we can get? The answer has a lot to do with multiple eigenvalues, and nilpotent matrices. To appreciate this phrase, let us first give the desired target result, then we'll see how to get it.

1.5.1 Jordan Form

Definition 1.5.21 *A matrix* $J_k(\lambda) \in \mathbb{F}^{k\times k}$ *is called a Jordan block of size* k, *relatively to the eigenvalue* λ, *if it is of the form* $J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix}$. *A matrix* $J \in \mathbb{F}^{n\times n}$ *is said to be in Jordan form if it is the direct sum of Jordan blocks:* $J = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & O \\ & & \ddots & \\ O & & & J_{n_k}(\lambda_k) \end{pmatrix}$ *where* $n_1 + n_2 + \cdots + n_k = n$. *The* λ_i*'s need not be distinct.*

Remarks 1.5.22

(1) If $k = n$ (hence $n_i = 1$, for all i) $\Rightarrow J$ is diagonal.

(2) Relatively to the same eigenvalue, customarily one chooses (as we will do below) the sizes of the Jordan blocks in decreasing order.

(3) If $k < n \Rightarrow J$ cannot *be transformed to* diagonal *form.*
(If $k < n \Rightarrow$ there is one block, say $J_p \in \mathbb{F}^{p\times p}$, $p > 1$. We show that this block cannot be diagonalized. For, suppose the contrary: $\exists\ S : S^{-1}J_pS = \Lambda_p$, diagonal $\Rightarrow \Lambda_p = \begin{pmatrix} \lambda_p & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} = \lambda_p I \Rightarrow J_p - \lambda_p I = S\Lambda_p S^{-1} - \lambda_p I = S(\lambda_p I - \lambda_p I)S^{-1} = 0$, which is not true if $p > 1$.)

We are ready for the **important** result that matrices with eigenvalues in $\mathbb{F}$ can be brought to Jordan form. In light of the last Remark above, this form cannot be further simplified; that is, the form is as close to being diagonal as possible.

Theorem 1.5.23 (Jordan canonical form.) *Given $A \in \mathbb{F}^{n\times n}$ with all eigenvalues in $\mathbb{F}$. Then, there exist $S \in \mathbb{F}^{n\times n}$, invertible, such that $S^{-1}AS$ is in Jordan form. This form (that is, the values $n_1, \dots, n_k$) is unique, aside from trivial reordering of the diagonal blocks.*

To prove this result we make some preliminary observations, which will simplify our life.

Remarks 1.5.24

(1) We can assume – without loss of generality – that the matrix A is in block diagonal form and the diagonal blocks are upper triangular corresponding to the different eigenvalues. Clearly, if we can bring to their respective Jordan forms all of these blocks on the diagonal of A, then we can bring A to Jordan form!

(2) So, we may as well assume that $A \in \mathbb{F}^{n\times n}$ is given by a single triangular block: $A = \lambda I + N$, where N is the strictly upper triangular part of A. That is, $A = \begin{pmatrix} \lambda & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & & * \\ 0 & & & \lambda \end{pmatrix} = \lambda I + N$. Since $S^{-1}AS = \lambda I + S^{-1}NS$, we will only need to prove that N can be taken to Jordan form as claimed. So, we will focus on

transforming N to Jordan form,

$$N = \begin{pmatrix} 0 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & & * \\ 0 & & & 0 \end{pmatrix}. \tag{1.5.1}$$

Observe that obviosuly N is nilpotent and for sure $N^n = 0$.

Definition 1.5.25 *A given nilpotent matrix $N \in \mathbb{F}^{n\times n}$ is said to have index of nilpotency n_1 if $N^{n_1} = 0$ but $N^{n_1-1} \neq 0$.*

Example 1.5.26 *Take*

$$N = \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}.$$

Easily, we always have $N^3 = 0$, but $N^2 = \begin{pmatrix} 0 & 0 & ab \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, i.e., $N^2 = 0$ if and only if a or b is 0.

$\therefore$ *N has index of nilpotency 3 if $a \neq 0$, $b \neq 0$, nilpotency index 2 if $a = 0$ or $b = 0$, and the nilpotency index is 1 if $a = b = c = 0$.*

In light of Remarks 1.5.24, Theorem 1.5.23 will follow from the following:

Theorem 1.5.27 *Let $N \in \mathbb{R}^{n\times n}$ be strictly upper triangular as in* (1.5.1)*. Then, there is an invertible $V \in \mathbb{F}^{n\times n}$ and indices $n_1, n_2, \ldots, n_p : n_1 \geq n_2 \geq \cdots \geq n_p \geq 1$, $n_1 + n_2 + \cdots + n_p = n$ such that*

$$V^{-1}NV = \begin{bmatrix} J_{n_1}(0) & & & 0 \\ & J_{n_2}(0) & & \\ & & \ddots & \\ & 0 & & J_{n_p}(0) \end{bmatrix}.$$

Here, n_1 is the nilpotency index of N. The indices $n_1, \ldots, n_p$ are unique.

Example 1.5.28 *Take $N = \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$ and let us examine different cases.*

1) $a = b = c = 0 \Rightarrow N = 0$ *and* $n_1 = n_2 = n_3 = 1$.

2) $a = b = 0,\ c \neq 0 \Rightarrow N = \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ N^2 = 0.$

Observe that N *is similar to* $\tilde{N} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1/c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} N \begin{pmatrix} c & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *which is similar to* $\begin{pmatrix} 0 & 1 & | & 0 \\ 0 & 0 & | & 0 \\ - & - & | & - \\ 0 & 0 & | & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tilde{N} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \right)$, *then* $n_1 = 2, n_2 = 1$.

3) $a = 0,\ b \neq 0,\ N = \begin{pmatrix} 0 & 0 & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow N^2 = 0.$

N *is similar to* $\begin{pmatrix} 0 & 0 & c \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *which is similar to* $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ $\left(\text{use } \begin{pmatrix} 1 & -c & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *which is similar to* $\begin{pmatrix} 0 & 1 & | & 0 \\ 0 & 0 & | & 0 \\ - & - & | & - \\ 0 & 0 & | & 0 \end{pmatrix}$, *so* $n_1 = 2,\ n_2 = 1$.

4) $a \neq 0,\ b = 0 \Rightarrow$ *as in 3).*

5) $a \neq 0,\ b \neq 0,\ N^2 \neq 0,\ N^3 = 0;\ N = \begin{pmatrix} 0 & a & c \\ 0 & 0 & b \\ 0 & 0 & 0 \end{pmatrix}$, *which is similar to*

$\begin{pmatrix} 0 & ab & c \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/b & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *which is similar to*

$\begin{pmatrix} 0 & 1 & c/ab \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1/ab & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *which is finally similar to*

$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \left(\text{use } \begin{pmatrix} 1 & -c/ab & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right)$, *that is* $n_1 = 3$. $\square$

Proof of Theorem 1.5.27. [The proof is adapted from [4].]

By induction on n. $n = 1 \Rightarrow N = [0]$, and the statement is trivially true.

So, suppose that the result is true for all strictly upper triangular matrices of size $< n$.

Write $N = \begin{array}{c|cc} & 1 & n-1 \\ 1 & 0 & a^T \\ n-1 & 0 & N_1 \end{array}$ i.e. $N = \begin{pmatrix} 0 & a^T \\ 0 & N_1 \end{pmatrix}$ (we write a^T to clarify that it is a row vector).

By induction, $\exists\, V_1 : V_1^{-1} N_1 V_1 = \begin{bmatrix} J_{k_1} & & 0 \\ & \ddots & \\ 0 & & J_{k_s} \end{bmatrix} = \begin{bmatrix} J_{k_1} & 0 \\ 0 & J \end{bmatrix}$ with $k_1 \geq \cdots \geq k_s \geq 1$ and $k_1 + \cdots + k_s = n-1$, and note that $J^{k_1} = 0$. So

$$\begin{pmatrix} 1 & 0 \\ 0 & V_1^{-1} \end{pmatrix} N \begin{pmatrix} 1 & 0 \\ 0 & V_1 \end{pmatrix} = \begin{array}{c} \\ 1 \\ k_1 \\ k_2+\cdots+k_s \end{array} \begin{array}{c} \begin{array}{ccc} 1 & k_1 & k_2+\cdots+k_s \end{array} \\ \begin{pmatrix} 0 & a_1^T V_1 & a_2^T V_2 \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix} \end{array} =: \begin{pmatrix} 0 & b_1^T & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}.$$

This is further transformed by similarity as follows

$$\begin{pmatrix} 1 & -b_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} 0 & b_1^T & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix} \begin{pmatrix} 1 & b_1^T J_{k_1}^T & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix}$$
$$= \begin{pmatrix} 0 & b_1^T (I - J_{k_1}^T J_{k_1}) & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix} = \begin{pmatrix} 0 & (b_1^T e_1) e_1^T & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix},$$

since (verify!)

$$I - J_{k_1}^T J_{k_1} = e_1 e_1^T \quad \left(e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{F}^{k_1}, \quad \text{with} \quad e_2, e_3, \ldots \quad \text{similarly defined} \right).$$

There are two cases to consider.

1) $b_1^T e_1 = 0 \Rightarrow$ have $\begin{pmatrix} 0 & 0 & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}$ which is similar to (via permutation) $\left(\begin{array}{c|cc} J_{k_1} & 0 & 0 \\ \hline 0 & 0 & b_2^T \\ 0 & 0 & J \end{array} \right)$,

but by induction there is $V_2 \in \mathbb{F}^{(n-k_1)\times(n-k_1)}$ such that $V_2^{-1} \begin{pmatrix} 0 & b_2^T \\ 0 & J \end{pmatrix} V_2 = \tilde{J}$ is in Jordan form.

$\therefore$ N is similar to $\begin{pmatrix} J_{k_1} & 0 \\ 0 & \tilde{J} \end{pmatrix}$ which is in Jordan form (it may be that blocks are not arranged in non-increasing order, in which case we can permute them to obtain desired ordering).

2) $b_1^T e_1 \neq 0 \Rightarrow$ have $\begin{pmatrix} 0 & [*, 0, \ldots, 0] & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}$. Now, first we make the "$*$" entry become 1:

$$\begin{pmatrix} \frac{1}{b_1^T e_1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} 0 & (b_1^T e_1) e_1^T & b_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix} \begin{pmatrix} (b_1^T e_1) & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} =: \begin{pmatrix} 0 & e_1^T & c_2^T \\ 0 & J_{k_1} & 0 \\ 0 & 0 & J \end{pmatrix}$$

Note that $\hat{J} = \begin{pmatrix} 0 & e_1^T \\ 0 & J_{k_1} \end{pmatrix}$ is a Jordan block of size $k_1 + 1 : \hat{J} = J_{k_1+1}(0)$. So, we have

$$\begin{array}{r c} & \begin{array}{cc} k_1 + 1 & k_2 + \cdots + k_s \end{array} \\ \begin{array}{r} k_1 + 1 \\ k_2 + \cdots + k_s \end{array} & \begin{pmatrix} \hat{J} & e_1 c_2^T \\ 0 & J \end{pmatrix} \end{array}.$$

Next, we "chase away" $e_1 c_2^T$ by a sequence of similarities:

$$\begin{pmatrix} I & e_2 c_2^T \\ 0 & I \end{pmatrix} \begin{pmatrix} \hat{J} & e_1 c_2^T \\ 0 & J \end{pmatrix} \begin{pmatrix} I & -e_2 c_2^T \\ 0 & I \end{pmatrix} = \begin{pmatrix} \hat{J} & -\hat{J} e_2 c_2^T + e_1 c_2^T + e_2 c_2^T J \\ 0 & J \end{pmatrix} = \begin{pmatrix} \hat{J} & e_2 c_2^T J \\ 0 & J \end{pmatrix},$$

since $\hat{J} e_2 = e_1$. And, recursively, for $i = 1, 2, 3, \ldots, k_1$:

$$\begin{aligned} &\begin{pmatrix} I & e_{i+1} c_2^T J^{i-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \hat{J} & e_i c_2^T J^{i-1} \\ 0 & J \end{pmatrix} \begin{pmatrix} I & -e_{i+1} c_2^T J^{i-1} \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} \hat{J} & -\hat{J} e_{i+1} c_2^T J^{i-1} + e_i c_2^T J^{i-1} + e_{i+1} c_2^T J^i \\ 0 & J \end{pmatrix} = \begin{pmatrix} \hat{J} & e_{i+1} c_2^T J^i \\ 0 & J \end{pmatrix}, \end{aligned}$$

since $\hat{J} e_{i+1} = e_i$. But $J^{k_1} = 0$, and so we eventually get the form

$$\begin{pmatrix} \hat{J} & 0 \\ 0 & J \end{pmatrix} = \begin{pmatrix} J_{k_1+1}(0) & & 0 \\ & J_{k_2}(0) & \\ & \ddots & \\ 0 & & J_{k_s}(0) \end{pmatrix}$$

as desired. (Here, $n_1 = k_1 + 1$, $n_j = k_j$, $j = 2, \ldots, s$ and $s = p$.)

The last thing that we need to justify is the uniqueness part. We have obtained the Jordan form for N:

$$J = \begin{pmatrix} J_{n_1}(0) & & \\ & \ddots & \\ & & J_{n_p(0)} \end{pmatrix}; \quad n_1 + \cdots + n_p = n, n_1 \geq \cdots \geq n_p \geq 1 \, .$$

We'll show that the set of Jordan blocks (i.e., the values $n_1, \ldots, n_p$) is completely specified by the values $\text{rank}(J^m)$, $m = 1, 2, \ldots, n$. [This will give the result, because if A and B are similar $\Rightarrow A^m$ and B^m are too and, thus, their ranks are the same.]

Now: $J_{n_k}^m = 0$ if $m \geq n_k$ and $\text{rank}(J_{n_k}^{m-1}) - \text{rank}(J_{n_k}^m) = 1$ if $m < n_k$. Set $r_m = \text{rank}\ (J^m)$, $m = 1, 2, \ldots$, $r_0 = n$ and $d_m = r_{m-1} - r_m$.

$\therefore$ d_m is # of Jordan blocks in J of size $k \geq m$ and surely $d_m = 0$, if $m > n$.

$\therefore$ the # of Jordan blocks in J of exact size $k = m$ is given by $d_m - d_{m+1} = r_{m-1} - 2r_m + r_{m+1}$, $m = 1, 2, \ldots, n$ and so the # of Jordan blocks of size $k = m$ is given by $\text{rank}(J^{m-1}) - 2\ \text{rank}(J^m) + \text{rank}(J^{m+1}) = \text{rank}(N^{m-1}) - 2\ \text{rank}(N^m) + \text{rank}(N^{m+1})$. $\square$

Remarks 1.5.29

(1) *The Jordan form is a very useful theoretical result, which further allows one to create equivalency classes for matrices with the same Jordan form, in the sense that matrices are similar if and only if they have the same Jordan form.*

(2) *The number of Jordan blocks corresponding to the same eigenvalue gives the* geometric multiplicity *of that eigenvalue. The* algebraic multiplicity *of that eigenvalue is the sum of the sizes of all Jordan blocks relative to that eigenvalue. [As an aside observation, this fact tells us that the algebraic multiplicity of an eigenvalue cannot be less than its geometric multiplicity.]*

(3) *Unfortunately, the Jordan form of a matrix is not continuous in the entries of the matrix, and small changes in the entries of a matrix can produce Jordan forms with different blocks sizes. For example, consider the Jordan form of* $\begin{bmatrix} \varepsilon & 0 \\ 1 & 0 \end{bmatrix}$.

Exercise 1.5.30 *Verify the above remarks (1) and (3).*

1.5.2 Extending the field, characteristic polynomial

We must stress that to obtain the Jordan form (over an arbitrary field $\mathbb{F}$), we have assumed all the eigenvalues to be in $\mathbb{F}$. This restriction cannot be removed.

So, to proceed further, we need to understand where can be the eigenvalues of a matrix in $\mathbb{F}^{n\times n}$, if not in $\mathbb{F}$. Or, alternatively, can we have some situations (that is, some special classes of transformations/matrices, and/or of $\mathbb{F}$) which are guaranteed to give eigenvalues in $\mathbb{F}$?

To make progress, we take two steps back and revisit our concept of eigenvalue.

This is the idea. We have $A \in \mathbb{F}^{n\times n}$, but we can also think of $A \in \mathbb{K}^{n\times n}$, where $\mathbb{K}$ is an *extension* of the field $\mathbb{F}$. And, we will want to think of the extension $\mathbb{K}$ so that **all** the **eigenvalues** of A will be in $\mathbb{K}$. That is, so that all the values λ such that

$$A - \lambda I \quad \text{is singular} \quad (\text{in } \mathbb{K}^{n\times n}),$$

will be in $\mathbb{K}$. To decide what $\mathbb{K}$ should be, we finally resort to the concepts of **determinant** and **characteristic polynomial**. For completeness, let us refresh these concepts and the relevant properties.

1st. Recall that for $A \in \mathbb{F}^{n\times n}$, $\det A = \sum_{\sigma\in S}(-1)^\sigma a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}$ where S is the set of permutations on $\{1,2,\dots,n\}$ (symmetric group). Here, σ is a permutation, and

$$(-1)^\sigma = \begin{cases} 1, & \text{for even permutations} \\ -1, & \text{for odd permutations} \end{cases}$$

(recall that "even" means that the minimal number of pairwise exchanges (transpositions) needed to produce σ from $\{1,2,\dots,n\}$ is even). Alternatively, you may want to repeatedly use the (*Laplace Expansion*) formula for the determinant:

$$\det A = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij} = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}\ , \tag{1.5.2}$$

where $A_{ij} \in \mathbb{F}^{n-1,n-1}$ is the matrix obtained by deleting the i-th row and j-th column of A.

2nd. Recall that A is nonsingular $\Leftrightarrow$ $\det A \neq 0$. Therefore, $A - \lambda I$ singular $\leftrightarrow$ $\det(A-\lambda I) = 0$. Finally, note that $\det(\lambda I - A)$ is a polynomial in λ of degree n with coefficients in $\mathbb{F}$ (it is called the *characteristic polynomial*).

Exercises 1.5.31

(1) *Show that* $\det A = \prod_{j=1}^n \lambda_j$ *is the constant term in the characteristic polynomial. This also shows that* $\prod_{j=1}^n \lambda_j \in \mathbb{F}$ *and that A is singular if and only if* $\det A = 0$.

(2) *Show that if A is invertible, then the eigenvalues of A^{-1} are the reciprocal of the eigenvalues of A. In particular,* $\det A^{-1} = 1/\det A$.

(3) *Let* $\mathrm{tr}(A) = (a_{11} + a_{12} + \cdots + a_{nn})$ *[*trace *of A]. Show that* $\mathrm{tr}(A)$ *is the coefficient (possibly, except for the sign) relative to degree $(n-1)$ in the characteristic polynomial. Moreover, show that* $\mathrm{tr}(A) = \sum_{j=1}^{n} \lambda_j$ *(which also shows that* $\sum_{j=1}^{n} \lambda_j \in \mathbb{F}$*).*

Remark 1.5.32 *A consequence of the above Exercises is that both* $\det A$ *and* $\mathrm{tr}\, A$ *are invariant under similarity transformations.*

To sum up, the extension field $\mathbb{K}$ should be chosen so to include all the roots of the characteristic polynomial, that is of all monic algebraic equations of degree n with coefficients in $\mathbb{F}$. In general (that is, unless we are willing to restrict to special classes of matrices), this effectively forces us to consider the complex field $\mathbb{C}$ (which is a *closed* field).

In particular, from our previous construction, we have proven the fundamental result that

Theorem 1.5.33 *Any matrix* $A \in \mathbb{C}^{n\times n}$ *is similar to a Jordan form as in Definition 1.5.21. That is, for any given* $A \in \mathbb{C}^{n\times n}$*, there exists invertible* $V \in \mathbb{C}^{n\times n}$ *such that* $V^{-1}AV$ *is in Jordan form.*

As a consequence of Theorem 1.5.33, one may as well assume that a matrix $A \in \mathbb{C}^{n\times n}$ is in Jordan form, in that it can be brought to Jordan form if it is not so to begin with. This line of thought allows us to prove useful facts in a simple way, as exemplified by the next exercise.

Exercise 1.5.34 *Let* $A \in \mathbb{C}^{n\times n}$*. Suppose there exist an integer* $k > 1$ *such that* $A^k = A$*. Show that A is diagonalizable.*

Solution. Without loss of generality, we can assume that A is in Jordan form

$$J = \mathrm{diag}(J_{n_k}(\lambda_k),\ k = 1, \ldots, p)\,.$$

By contradiction, we suppose that at least one Jordan block is non-diagonal (i.e., at least one $n_k > 1$, for some $k = 1, \ldots, p$). Since $A^k = A$ implies $J^k = J$, and J is block diagonal, we can assume that we are dealing with just one Jordan block:

$J = \begin{pmatrix} \lambda & 1 & & & 0 \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix}$. It is immediate to realize (use induction) that J^k has the form $J^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \cdots & \cdots & \\ & \lambda^k & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & k\lambda^{k-1} \\ 0 & & & & \lambda^k \end{pmatrix}$. Therefore, to have $J^k = J$, we need to have

$$\lambda^k = \lambda \quad \text{and} \quad k\lambda^{k-1} = 1\,.$$

So, either $\lambda = 0$ or $\lambda^{k-1} = 1$. If $\lambda = 0$, then we surely cannot satisfy $k\lambda^{k-1} = 1$, but if $\lambda^{k-1} = 1$, then $k = 1$. Either way, we would reach a contradiction. Therefore, all Jordan blocks must be diagonal. □

Remark 1.5.35 *We will not look at other canonical forms, the so-called rational canonical forms (nothing to do with $\mathbb{Q}$), which maintain us in the original field $\mathbb{F}$ (and not an extension), since they are seldom used. If interested, please refer to [2, 4].*

Exercises 1.5.36 *Here, we explore the connection between Jordan form and minimal and characteristic polynomials. You may assume that $A \in \mathbb{C}^{n\times n}$ or that $A \in \mathbb{F}^{n\times n}$ with all eigenvalues in $\mathbb{F}$.*

(1) Suppose that the distinct eigenvalues of A are $\lambda_1, \ldots, \lambda_k$. Show that the minimal polynomial of A is $p(x) = \prod_{j=1}^{k}(x-\lambda_j)^{n_j}$, where n_j is the size of the largest Jordan block relative to λ_j.[3] Conclude that A is diagonalizable if and only if its minimal polynomial is simply $(x-\lambda_1)\cdots(x-\lambda_k)$. What is the characteristic polynomial of A?

[3] How is n_j related to the nilpotency index?

(2) *Consider the following matrix C:*

$$C = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & 1 & 0 & -a_{n-2} \\ 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix}.$$

Show that the characteristic and minimal polynomials of C are the same. [C is called the companion matrix *of its characteristic polynomial.] Note that we are not saying that a certain matrix A with given characteristic polynomial $p(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$ is similar to C; however, show that A is similar to C if and only if the minimal and characteristic polynomials of A are the same.*

We conclude this first set of lectures with a question.

Question. Suppose we have a matrix in $A \in \mathbb{R}^{n\times n}$. We find its eigenvalues as roots of the characteristic polynomial, and have that some of these are complex conjugate. Surely we can triangularize it with respect to $\mathbb{C}$. But, what can we achieve if we insist that the similarities be real? Of course, we could ask the same question relatively to the Jordan form.

Chapter 2

Schur decomposition, SVD, and their consequences

This chapter contains the core material of linear algebra. At a high level, the theoretical results in this chapter are consequences of having an inner product. At a finer level, we will obtain very specialized results for important classes of matrices and important inequalities for eigenvalues.

Most of the material in this chapter can be found in [4] and [7].

Hereafter, we will always work with matrices in $\mathbb{C}^{n\times n}$ or $\mathbb{R}^{n\times n}$.

2.1 Inner Product

Let us recall the concept (Euclidean or standard) of inner product $\langle x, y\rangle$ on vectors x, y in $\mathbb{C}^n$ (or $\mathbb{R}^n$):

$$x, y \in \mathbb{C}^n \Rightarrow \langle x, y\rangle = y^* x \ ,$$

$$x, y \in \mathbb{R}^n \Rightarrow \langle x, y\rangle = y^T x \ ,$$

where we are using the notation: $y^* = \bar{y}^T$ (conjugate transpose, also called Hermitian of y).

Note that in $\mathbb{R}^n$ we have $\langle x, y\rangle = \langle y, x\rangle$, but the same is not true in $\mathbb{C}^n$, in general. In fact, in $\mathbb{R}^n$, inner product is bilinear, that is for any $\alpha, \beta \in \mathbb{R}$:
a) $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$;
b) $\langle x, \alpha y + \beta z\rangle = \alpha\langle x, y\rangle + \beta\langle x, z\rangle$;

But in $\mathbb{C}^n$, the inner product is linear in the first argument, but "conjugate-linear" in the second:

c) $\langle x, \alpha y + \beta z\rangle = \bar{\alpha}\langle x, y\rangle + \bar{\beta}\langle x, z\rangle$, for any $\alpha, \beta \in \mathbb{C}$. In particular, $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

Regardless of being in $\mathbb{C}$ or $\mathbb{R}$, the norm of a vector x (Euclidean norm, 2-norm, or simply length) induced by the above inner product is

$$\|x\|_2 = (\langle x, x\rangle)^{1/2}\,.$$

Accordingly, we define the induced matrix norm (2-norm) as

$$\|A\|_2 = \max_{x:\|x\|_2=1} \|Ax\|_2\,.$$

A most important result is the following inequality.

Lemma 2.1.1 (Cauchy-Schwartz inequality) *For any $x, y \in \mathbb{C}^n$ (or $\mathbb{R}^n$), we have*

$$|\langle x, y\rangle| \le \|x\|_2\|y\|_2\,, \tag{2.1.1}$$

with equality only if $y = \alpha x$.

Exercise 2.1.2 *Prove Lemma 2.1.1.*

Remark 2.1.3 *A consequence of the Cauchy-Schwartz inequality is that we can define an angle θ between x and y from the relation:*

$$\cos\theta = \frac{\langle x, y\rangle}{\|x\|_2\|y\|_2}\ , \tag{2.1.2}$$

customarily taking $-\pi < \theta \le \pi$.

In particular, we call two vectors *orthogonal* if their inner product is 0. Further, we call them *orthonormal* if each of them has unit length.

An extremely useful and important process is the "Gram-Schmidt orthogonalization", which allows to transform a given set of p linearly independent vectors into another set of p orthonormal vectors, spanning the same subspace as the original set.

Algorithm 2.1.4 (Gram-Schmidt) *Given $\{x_1, x_k, \dots, x_p\}$ linearly independent vectors in $\mathbb{C}^n$ (or $\mathbb{R}^n$). The procedure below will produce $\{u_1, u_2, \dots, u_p\}$ such that $\langle u_i, u_j\rangle = 0$, $i \ne j$, $\langle u_i, u_i\rangle = 1$, and span$\{u_1, u_2, \dots, u_p\}$ = span$\{x_1, x_2, \dots, x_p\}$.*

$i = 1:$

$$y_1 = x_1\ ,\ \ u_1 = y_1/\|y_1\|\ .$$

$i > 1$:

$$\begin{cases} y_2 = x_2 - \langle x_2, u_1 \rangle u_1 & \textit{(subtract from } x_2 \textit{ its component in direction of } u_1\textit{)} \\ u_2 = y_2/\|y_2\| \end{cases}$$

$$\vdots$$

$$\begin{cases} y_k = x_k - \langle x_k, u_{k-1} \rangle u_{k-1} - \cdots - \langle x_k, u_1 \rangle u_1 \\ u_k = y_k/\|y_k\| \end{cases}, \quad k = 2, \ldots, p.$$

If we write $U = [u_1, u_2, \ldots, u_p] \in \mathbb{C}^{n\times p}$ and $X = [x_1, x_2, \ldots, x_p] \in \mathbb{C}^{n\times p}$, then we have found that $X = UR$, where $R \in \mathbb{C}^{p\times p}$ is upper triangular. Moreover, the diagonal of R is positive (it is $\|y_k\|$, $k = 1, \ldots, p$). Finally, the matrix U has *orthonormal* columns: $u_j^* u_k = 0$, $j \neq k$, and $u_j^* u_j = 1$, $j = 1, \ldots, p$. For this reason, U is called *orthonormal*: $U^*U = I_p$. If $p = n \Rightarrow U$ is called *unitary*: $U^*U = I$ $(= UU^*)$.

Also, note that the Gram-Schmidt process holds unchanged in the real case. In this case, when $p = n$, U is called *orthogonal*: $U^TU = I$ $(= UU^T)$. [We will reserve the term orthogonal for real matrices (other texts use orthogonal also for $A \in \mathbb{C}^{n\times n}$ whenever $A^TA = I$).]

Exercise 2.1.5 *Give the formula for the non-zero entries of R obtained from the Gram-Schmidt process.*

We summarize the construction given by the Gram-Schmidt process in the following theorem.

Theorem 2.1.6 (QR factorization) *Let $A \in \mathbb{C}^{n\times p}$ (or $\mathbb{R}^{n\times p}$), $n \geq p$, be full rank. Then there is an orthonormal (n, p) matrix Q and a (p, p) upper triangular R, with real diagonal, such that $A = QR$. This factorization is unique for any assigned sign pattern on the diagonal of R (e.g., all positive). If $p = n$, then Q is unitary (orthogonal in the real case).* □

The following are important properties of unitary matrices U (similar properties hold for orthogonal matrices):

(1) $U^{-1} = U^*$ and U^* is unitary.
(2) UV is unitary if U and V are.
(3) The set $\{U \in \mathbb{C}^{n\times n},\ U \text{ unitary}\}$ is a group; it is a subgroup of $GL(n, \mathbb{C})$ (the orthogonal matrices are a subgroup of $GL(n, \mathbb{R})$).

(4) The set $\{U \in \mathbb{C}^{n\times n},\ U \text{ unitary}\}$ is closed and bounded (as subset of $\mathbb{C}^{n^2}$) and therefore compact.
(5) $\|Ux\|_2 = \|x\|_2$, $\forall\, x \in \mathbb{C}^n$.

Exercises 2.1.7

*(1) Show properties (1), (2), (3), (4), (5) above. (*Hint: *To show property (4), reason as follows. Let $\{U_k\}$ be a sequence of unitary matrices and suppose that $U_k \to U$. Show that U is unitary. Here, convergence is entrywise.)*

(2) Show that all the eigenvalues of a unitary matrix U are on the unit circle and that $|\det U| = 1$. The result is true also in the real case, but $\det U$ can only take values ± 1 in the real case. [As an aside remark, this means that there are two disjoint classes of orthogonal matrices: Those with determinant 1 and those with determinant -1. In the plane, these are rotations *and* reflections*, respectively.]*

(3) Suppose that $A \in \mathbb{C}^{n\times n}$ is similar to U, and that U is unitary. Show that A^{-1} is similar to A^*.

Let us now see some consequence of *unitary similarity* (orthogonal similarity in $\mathbb{R}^{n\times n}$). By this we mean that given A and B: $A = U^*BU$, where U is unitary.

First of all, observe that unitary similarity is an equivalence relation, and that the 2-norm of a matrix is trivially unitarily invariant, since unitary transformations maintain length. The following result tells us that also the Frobenius norm of a matrix is unitarily invariant, where

$$\|A\|_F^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$$

is the *Frobenius norm* of the matrix A (it is also called the Hilbert-Schmidt norm).

Theorem 2.1.8 (Frobenius invariance) *If $A = U^*BU$, U unitary, then*

$$\sum_{i,j} |a_{ij}|^2 = \sum_{i,j} |b_{ij}|^2 \ .$$

Pf. First, observe that $\sum_{i,j} |a_{ij}|^2 = \mathrm{tr}(\bar{A}^T A) = \mathrm{tr}(A^*A)$. But $\mathrm{tr}(A^*A) = \mathrm{tr}(U^*B^*UU^*BU) = \mathrm{tr}(U^*B^*BU)$ and the trace is invariant under similarity transformations (see Remark 1.5.32) $\therefore$ $tr(A^*A) = tr(B^*B)$. $\square$

Remark 2.1.9 *Careful: Similar matrices may fail to satisfy Frobenius invariance. For example, take* $A = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$, $\quad B = \begin{bmatrix} -9 & -110 \\ 1 & 12 \end{bmatrix}$, $\quad \left(B = T^{-1}AT,\ T = \begin{bmatrix} 1 & 10 \\ 0 & 1 \end{bmatrix} \right)$.

Exercise 2.1.10 *Let $A \in \mathbb{C}^{n\times n}$ (or also in $\mathbb{R}^{n\times n}$). Show that $\|A\|_2 \leq \|A\|_F$.*

2.2 Schur Decomposition

We know that $A \in \mathbb{C}^{n\times n}$ can be triangularized by similarity (see Theorem 1.5.14). A very important result is that unitary similarity is sufficient to achieve the triangularization of a matrix $A \in \mathbb{C}^{n\times n}$.

Theorem 2.2.1 (Schur) *Given $A \in \mathbb{C}^{n\times n}$. There is a unitary U such that $U^*AU = R$, where R is upper triangular. The eigenvalues on the diagonal of R can be arranged in any desired way, say $\lambda_1, \lambda_2, \dots, \lambda_n$.*

Pf. "Constructive." Let v_1 be such that $Av_1 = \lambda_1 v_1$ and $v_1^* v_1 = 1$. Extend v_1 to a basis of $\mathbb{C}^n$: $v_1, y_2, \dots, y_n$. Use the Gram-Schmidt process to produce an orthonormal basis: $v_1, z_2, \dots, z_n$. Form unitary $U_1 = [v_1, z_2, \dots, z_n]$, $U_1^* U_1 = I$. We have $U_1^* A U_1 = \begin{bmatrix} \lambda_1 & | & * \cdots * \\ -- & | & -- \\ 0 & | & A_1 \end{bmatrix}$, where $A_1 \in \mathbb{C}^{n-1,n-1}$ has eigenvalues $\lambda_2, \dots, \lambda_n$. Let $x_2 \in \mathbb{C}^{n-1}$: $A_1 x_2 = \lambda_2 x_2$, $x_2^* x_2 = 1$. Extend x_2 to a basis of $\mathbb{C}^{n-1}$ and use the Gram-Schmidt process on this to form $V_2 \in \mathbb{C}^{n-1,n-1}$ unitary $\Rightarrow V_2^* A_1 V_2 = \begin{bmatrix} \lambda_2 & | & * \cdots * \\ -- & | & -- \\ 0 & | & A_2 \end{bmatrix}$, $A_2 = \mathbb{C}^{n-2,n-2}$ with eigenvalues $\lambda_3, \dots, \lambda_n$. Form $U_2 = \begin{bmatrix} 1 & | & 0 \cdots 0 \\ -- & | & -- \\ 0 & | & V_2 \end{bmatrix} \Rightarrow U_2^* U_1^* A U_1 U_2 = \begin{bmatrix} \begin{matrix} \lambda_1 & * \\ 0 & \lambda_2 \end{matrix} & | & * \\ -- & | & -- \\ 0 & | & A_2 \end{bmatrix}$. Keep going. □

Remark 2.2.2 *In general, U in theorem 2.2.1 is not unique, nor is R. For example, take $R_1 = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$, and $R_2 = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix}$, which are unitarily equivalent via $U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$, $U^* R_1 U_2 = R_2$, and the off-diagonal entries of R_1 and R_2 are quite different! As another example, with repeated eigenvalues, consider $R_1 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, and $R_2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, which are trivially unitarily equivalent via $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.*

Note that the example in the last Remark also shows that two matrices which are simultaneously triangularized by unitary similarity, in general do not commute. The converse, however, is true.

Exercise 2.2.3 *Let $A, B \in \mathbb{C}^{n\times n}$. Show that if $AB = BA \Rightarrow A$ and B can be simultaneously triangularized by a unitary similarity transformation.*

To answer a question posed at the end of Chapter 1, let us see what happens when we have a matrix $A \in \mathbb{R}^{n\times n}$ and try to use orthogonal transformations to bring it to triangular form. The issue is how to deal with complex conjugate eigenvalues, since obviously they preclude us from being able to obtain a full triangularization maintaining the triangular factor in the real field!

Theorem 2.2.4 (Real Schur Theorem: Quasi-Upper Triangular Form) *Let $A \in \mathbb{R}^{n\times n}$. Then, there exists $Q \in \mathbb{R}^{n\times n}$, orthogonal, such that*

$$Q^T AQ = R = \begin{bmatrix} R_1 & * & \cdots & * \\ & R_2 & \ddots & * \\ & 0 & & R_k \end{bmatrix} , \quad 1 \le k \le n \ ,$$

where each R_i is either $(1,1)$ or $(2,2)$ containing a pair of complex conjugate eigenvalues.

Sketch of proof. Whereas the real eigenvalues can be treated as in the previous Schur's theorem 2.2.1, let's see how to deal with a complex conjugate pair. So, suppose $\lambda, \bar\lambda = \alpha \pm i\beta$, and let $x \neq 0$ be an eigenvector associated to λ: $Ax = \lambda x$. Write $x \in \mathbb{C}^n$ as $x = u + iv$, $u, v \in \mathbb{R}^n$, and notice that also $A\bar x = \bar\lambda \bar x$. Therefore, we have $A[u, v] = [u, v]\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, all real. Since $\lambda \neq \bar\lambda \Rightarrow \{x, \bar x\}$ are linearly independent over $\mathbb{C}$ and thus $\{u, v\}$ are linearly independent over $\mathbb{R}$. Now extend $\{u, v\}$ to a basis for $\mathbb{R}^n$, and use the Gram-Schmidt process on this to obtain an orthonormal basis of $\mathbb{R}^n$. Then, there exists Q orthogonal such that $Q^T AQ = \begin{bmatrix} R_1 & | & * \\ -- & | & -- \\ 0 & | & A_1 \end{bmatrix}$, where R_1 has eigenvalue $\lambda, \bar\lambda$, and $A_1 \in \mathbb{R}^{n-2,n-2}$. Now, continue on A_1. □

Exercises 2.2.5 *Here we explore the degree to which unitary (respectively, orthogonal) similarities to upper triangular form (respectively, quasi-uppper triangular form) are unique.*

(1) *Suppose that all eigenvalues $\lambda_1, \dots, \lambda_n$, of $A \in \mathbb{C}^{n\times n}$ are distinct, and that we have transformed A, by unitary similarity, to upper triangular form with the eigenvalues appearing on the diagonal as $\lambda_1, \dots, \lambda_n$. Discuss uniqueness of the unitary transformation. What if the eigenvalues are not distinct?*

(2) *Suppose that all eigenvalues $\lambda_1, \dots, \lambda_n$, of $A \in \mathbb{R}^{n\times n}$ are distinct (though not necessarily all of them real), and that we transform A using orthogonal similarity to quasi-upper triangular form with a given ordering of the eigenvalues (complex conjugate eigenvalues appear on a $(2,2)$ diagonal block but are not otherwise distinguished). Discuss uniqueness of the orthogonal factor. What if the eigenvalues are not distinct?*

An important consequence of Schur theorem (complex case) is that "the set of diagonalizable matrices is dense in the set of matrices." This means that given any non-diagonalizable matrix, there is a matrix arbitrarily close to it which is diagonalizable. We will actually show a stronger result, that there is a matrix with distinct eigenvalues.

Theorem 2.2.6 *Given $A \in \mathbb{C}^{n\times n}$. Then, $\forall\, \varepsilon > 0$, there exist $A(\varepsilon) \in \mathbb{C}^{n\times n}$ with distinct eigenvalues and such that*

$$\|A - A(\varepsilon)\|_F^2 < \varepsilon.$$

Pf. Let U unitary be such that $U^*AU = R$ is upper triangular with eigenvalues $\lambda_1, \dots, \lambda_n$ along the diagonal. Let $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{C}$ such that $|\alpha_i| < \left(\frac{\varepsilon}{n}\right)^{1/2}$ and such that $\lambda_1+\alpha_1$, $\lambda_2+\alpha_2, \dots, \lambda_n+\alpha_n$ are all distinct. Take $E(\varepsilon) = \mathrm{diag}(\alpha_1, \dots, \alpha_n)$ and form $R(\varepsilon) = R + E(\varepsilon)$ and, from this, set $A(\varepsilon) = UR(\varepsilon)U^*$. Now, observe that

$$\|A(\varepsilon) - A\|_F^2 = \|R(\varepsilon) - R\|_F^2 = \sum_{i,j} |R_{ij}(\varepsilon) - R_{ij}|^2 = \sum_{i=1}^{n} |\alpha_i|^2 < \varepsilon. \qquad \square$$

Exercise 2.2.7 *State and prove an analog of Theorem 2.2.6 for $A \in \mathbb{R}^{n\times n}$, making $A(\varepsilon) \in \mathbb{R}^{n\times n}$. [*Hint*: Produce a real matrix with distinct eigenvalues, whether real or complex conjugate.]*

In light of Theorem 2.2.6, and Exercise 2.2.7, it is natural to ask why we should bother with non-diagonalizable matrices.

Example 2.2.8 *Think about this problem. Suppose we have the real valued function* $A(t) = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix}$*, which is clearly non-diagonalizable at* $t = 0$*. We claim that if we take any real perturbation* εB *of this, it will be non-diagonalizable in a neighborhood of* $t = 0$*. [***Exercise***: Verify this claim.] In other words, whereas each single matrix may be effectively perturbed into a diagonalizable one, the entire family cannot!*

2.3 Self-adjoint (symmetric) and normal matrices

In general, in Schur's theorem 2.2.1, U and R are not uniquely determined. However, the quantity

$$\nu = \sum_{i<j} |R_{ij}|^2 \tag{2.3.1}$$

is always uniquely determined. (This is simply a consequence of Frobenius invariance, see Theorem 2.1.8.) For reasons that will clarify below, ν is called "*departure from normality*" of A. Schur's theorem, and the fact that ν is invariant under unitary transformations, give us an important consequence: In general, we cannot expect being able to transform a matrix to diagonal form with unitary similarity. Or, to put it in other words, we are able to do so only if –during the process of triangularizing it– we have effectively diagonalized it!

Our next task is to specialize Schur theorem to extremely important classes of matrices, those which are diagonalizable by a unitary transformation.

Definition 2.3.1 *A matrix* $A \in \mathbb{C}^{n\times n}$ *such that* $A^* = A$ *is called* Hermitian *(or* self-adjoint*). If* $A^* = -A$*, then* A *is called* anti-Hermitian *(or* skew-Hermitian*). In the real case,* $A \in \mathbb{R}^{n\times n}$*, if* $A^T = A$ *then* A *is called* symmetric *(or, again,* self-adjoint*), whereas if* $A^T = -A$ *then* A *is called* anti-symmetric *(or* skew-symmetric*).*

Unless otherwise explicitly stated, the name symmetric by itself will always be used for real matrices; on occasion, we may use the term "complex symmetric" to identify a complex matrix A which happens to formally satisfy $A = A^T$.

Remark 2.3.2 *If* $A^* = -A$*, then* iA *is Hermitian (similarly, if* $A^* = A$ *then* iA *is anti-Hermitian). Therefore, a result proven for Hermitian matrices has an immediate counterpart for anti-Hermitian ones.*

Hermitian (symmetric) matrices appear pervasively in applications, and this justifies paying special attention to these matrices. In particular, symmetric (Hermitian) matrices are of fundamental importance in studying behavior of functions near critical points.

Example 2.3.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be sufficiently smooth (at least twice continuously differentiable). Consider the Taylor expansion of f near a point x_0: $f(x) = f(x_0) + (\nabla f(x_0))^T(x - x_0) + \frac{1}{2}Q(x - x_0) + \|x - x_0\|^2 r(\|x - x_0\|)$ (here, $r(\|x - x_0\|) \to 0$ as $\|x - x_0\| \to 0$). The function $Q(x - x_0)$ is a quadratic function of the form*

$$Q(y) = \sum_{i,j} H_{ij} y_i y_j \,, \quad \text{and} \quad H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}\Big|_{x=x_0} \,.$$

Therefore, $Q(x - x_0) = (x - x_0)^T H(x - x_0)$ where the Hessian H is symmetric: $H_{ij} = H_{ji}$.

Now, if x_0 is a critical point $\Rightarrow \nabla f(x_0) = 0 \therefore$ the quadratic function determines the nature of the critical point. Undoubtedly, it would be simpler if H happened to be diagonal! $\square$

The next exercise is useful to understand the nature of the eigenvalues of Hermitian (anti-Hermitian) matrices.

Example 2.3.4 *Let $A \in \mathbb{C}^{n\times n}$. Define $A_s = (A + A^*)/2$ (*Hermitian, or self-adjoint, part *of A) and $A_a = (A - A^*)/2$ (*anti-Hermitian part *of A). Clearly, $A = A_s + A_a$ and this is the unique decomposition of A into the sum of a Hermitian and an anti-Hermitian matrix. Then, we claim that:*

$$\begin{aligned} \mathrm{Re}\langle x, Ax\rangle &= \langle x, A_s x\rangle \\ \mathrm{Im}\langle x, Ax\rangle &= \langle x, A_a x\rangle \,. \end{aligned}$$

The verification of this claim is simple. Take

$$\begin{aligned} \langle x, Ax\rangle = x^* A^* x &= \frac{x^*(A^* - A + A + A^*)x}{2} \\ &= \frac{x^*(A^* + A)x}{2} + \frac{x^*(A^* - A)x}{2} = \langle x, A_s x\rangle + \langle x, A_a x\rangle. \end{aligned}$$

Now:

$$\langle x, A_s x\rangle = (\langle x, A_s x\rangle)^* \therefore \langle x, A_s x\rangle \in \mathbb{R}$$

$$\textit{and} \quad \langle x, A_a x\rangle = -(\langle x, A_a x\rangle)^* \therefore \langle x, A_a x\rangle \in i\mathbb{R} \ ,$$

as claimed.

Furthermore, if (λ, x) *is an eigenpair of* A *such that* $x^*x = 1$ *then* $Ax = \lambda x$ *and so* $\begin{cases} x^*Ax = \lambda \\ x^*A^*x = \bar{\lambda} \end{cases}$ *, from which* $\mathrm{Re}\lambda = x^*A_s x$ *, and* $\mathrm{Im}\lambda = x^*A_a x$ *.*

An immediate consequence of the above Example is the following result.

Corollary 2.3.5 *If* $A \in \mathbb{C}^{n\times n}$ *(or* $A \in \mathbb{R}^{n\times n}$*) is Hermitian, then its eigenvalues are real. If* $A \in \mathbb{C}^{n\times n}$ *(or* $A \in \mathbb{R}^{n\times n}$*) is anti-Hermitian, then its eigenvalues are purely imaginary.*

We are now ready for one of the key results in linear algebra.

Theorem 2.3.6 (Spectral Theorem for Hermitian matrices) *Let* $A \in \mathbb{C}^{n\times n}$ *be Hermitian:* $A = A^*$*. Then,* A *is unitarily diagonalizable. That is, there exist unitary* $U \in \mathbb{C}^{n\times n}$ *such that* U^*AU *is diagonal. The eigenvalues can be ordered in any desired way.*

Pf. From Schur's theorem 2.2.1, we have that there exists U unitary such that $U^*AU = R$, R upper triangular with eigenvalues ordered as we like. Therefore, $U^*A^*U = R^* \Rightarrow R = R^*$, $\therefore R_{ij} = 0$, $i \neq j$, $\therefore R$ diagonal (and the diagonal is real). $\square$

Much the same result holds if $A \in \mathbb{R}^{n\times n}$. Before stating the theorem in this case, let us make an observation.

Observation 2.3.7 *If* $A \in \mathbb{R}^{n\times n}$ *with eigenvalue* $\lambda \in \mathbb{R}$ *(e.g., if* $A = A^T$*), then we can (and will) take the corresponding eigenvector to be real. This is simply because if* $Ax = \lambda x$*,* $x \neq 0$ *and we chose* $x = u + iv$*, with* $u, v, \in \mathbb{R}^n$*, then we must have* $Au + iAv = \lambda u + 0 \cdot iv \therefore \begin{cases} Au = \lambda u \\ Av = 0 \end{cases}$ *and so we can take* $v = 0$ *and* u *to be a real eigenvector.*

Theorem 2.3.8 (Spectral theorem for symmetric matrices) *Let* $A \in \mathbb{R}^{n\times n}$ *be symmetric:* $A = A^T$*. Then, there exist orthogonal* $Q \in \mathbb{R}^{n\times n}$ *such that* Q^TAQ *is diagonal. The eigenvalues can be ordered in any desired way.*

Pf. Since all eigenvalues are real, the real Schur theorem 2.2.4 gives that there exists Q, orthogonal, such that $Q^TAQ = R$, with R upper triangular with eigenvalues ordered as we like. But then $Q^TA^TQ = R^T \Rightarrow R$ is diagonal. $\square$

Remark 2.3.9 *We have achieved diagonalization for a Hermitian (respectively, symmetric) matrix without the need to assume distinct eigenvalues. Hermitian (symmetric) matrices lead to a basis of orthonormal eigenvectors: Hermitian matrices are diagonalized by a unitary similarity transformation. So, conceptually at least, we can always think of a Hermitian (symmetric) matrix as being diagonal. For example, in Example 2.3.3, we can change coordinates and diagonalize the Hessian; if we do so, we would be using the so-called* principal axes *as coordinate system.*

- Finally, let us see how far we can push diagonalization by unitary similarity. To motivate the next definition, consider the following.

Take $A \in \mathbb{C}^{n\times n}$ and write $A = A_s + A_a$. We know that $A_s = A_s^* \Rightarrow\ \exists\, U$ unitary such that $U^*A_sU = D$ (and real). Likewise (since iA_a is Hermitian) also A_a can be unitarily diagonalized. We also know (see Theorem 1.5.10) that two diagonalizable matrices are simultaneously diagonalizable $\Leftrightarrow$ they commute. Now, observe the following.

$$A_sA_a = \frac{A+A^*}{2}\frac{A-A^*}{2} = \frac{A^2 + A^*A - AA^* - (A^*)^2}{4}$$

$$A_aA_s = \frac{A-A^*}{2}\frac{A+A^*}{2} = \frac{A^2 - A^*A + AA^* - (A^*)^2}{4}$$

$\therefore\ A_sA_a = A_aA_s \Leftrightarrow AA^* - A^*A = A^*A - AA^* \Leftrightarrow AA^* = A^*A.$

Definition 2.3.10 *If $A \in \mathbb{C}^{n\times n}$ (or $\mathbb{R}^{n\times n}$) is such that $AA^* = A^*A$ (respectively, $AA^T = A^TA$) then A is called a* normal *matrix.*

So, we know that normal matrices are diagonalizable (since both A_s and A_a are). What we will show next is that the diagonalizing transformation can be taken to be unitary. The following observation is trivial to verify, and it will come in handy.

Observation 2.3.11 *Unitary similarity preserves normality.*

Theorem 2.3.12 (Spectral Theorem for normal matrices) *$A \in \mathbb{C}^{n\times n}$ is normal $\iff$ it is is unitarily diagonalizable.*

Pf. ($\Leftarrow$) Let U unitary be such that $U^*AU = D$, where D is diagonal. Thus, $U^*A^*U = D^* \Rightarrow AA^* = U(U^*AUU^*A^*U)U^* = UDD^*U^* = UD^*DU^* = UD^*U^*UDU^* = A^*A$.

($\Rightarrow$) We proceed by induction on n. Obviously the result is true for $n = 1$.

Next, let U be such that $U^*AU = R$ is upper triangular (Schur theorem). Then, since $A^*A = AA^*$, then also $RR^* = R^*R$.

Write $R = \begin{bmatrix} r_{11} & a^* \\ 0 & R_1 \end{bmatrix} \Rightarrow R^*R = \begin{bmatrix} r_{11}^* & 0 \\ a & R_1^* \end{bmatrix} \begin{bmatrix} r_{11} & a^* \\ 0 & R_1 \end{bmatrix} = \begin{bmatrix} r_{11}^*r_{11} & r_{11}^*a^* \\ ar_{11} & R_1^*R_1 \end{bmatrix}$ and $RR^* = \begin{bmatrix} r_{11} & a^* \\ 0 & R_1 \end{bmatrix} \begin{bmatrix} r_{11}^* & 0 \\ a & R_1^* \end{bmatrix} = \begin{bmatrix} r_{11}r_{11}^* + a^*a & a^*R_1^* \\ R_1a & R_1R_1^* \end{bmatrix}$. Thus, from $r_{11}^*r_{11} = r_{11}r_{11}^* + a^*a$ we immediately get $a = 0$ $\therefore$ $R_1R_1^* = R_1^*R_1$ and so R_1 is normal. Now use induction hypothesis. $\square$

Remark 2.3.13 *Careful! Theorem 2.3.12 says that there exists U, unitary, diagonalizing a normal matrix A. But it does not say that a unitary similarity diagonalizing A_s also diagonalizes A_a. For example, take $M \in \mathbb{C}^{n,n}, M = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$, $A^* = A \in \mathbb{C}^{n_1,n_1}$ and $B^* = -B \in \mathbb{C}^{n_2,n_2}$, $n_1 + n_2 = n$, $\Rightarrow M_s = \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix}$, $M_a = \begin{bmatrix} 0 & 0 \\ 0 & B \end{bmatrix}$. Let $U_1 \in \mathbb{C}^{n_1,n_1}, U_1^*U_1 = I$, be such that $U_1^*AU_1 = D_A (\in \mathrm{Re})$, and likewise let $U_2 \in \mathbb{C}^{n_2,n_2}, U_2^*U_2 = I$, be such that $U_2^*BU_2 = D_B (\in \mathrm{Im})$; then, $U = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}$ diagonalizes all of M_s, M_a, M. But, we could have also taken*

$$V_1 = \begin{pmatrix} U_1 & 0 \\ 0 & \hat{U} \end{pmatrix},\ V_1 \ \textit{unitary} : V_1^*M_sV_1 = \begin{pmatrix} D_A & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow V_1^*MV_1 = \begin{pmatrix} D_A & 0 \\ 0 & \hat{U}^*B\hat{U} \end{pmatrix}$$

and generally $\hat{U}^*B\hat{U}$ is not diagonal.

However, if the orthonormal eigenvectors of A_s and A_a are the same, then A is normal and it is diagonalized by these very same eigenvectors. $\square$

Exercise 2.3.14 *If the orthonormal eigenvectors of A_s are also eigenvectors of $A_a \Rightarrow A = A_s + A_a$ is normal.*

In the real case, similar results hold. To begin with, just as in the complex case, being normal is equivalent to $A_sA_a = A_aA_s$, where $A_s = \frac{A+A^T}{2}$, $A_a = \frac{A-A^T}{2}$. Moreover, normality is preserved by orthogonal similarity.

Theorem 2.3.15 (Spectral Theorem for real normal matrices) *Let $A \in \mathbb{R}^{n\times n}$. Then, A is normal $\iff$ there exists $Q \in \mathbb{R}^{n\times n}$, orthogonal, such that*

$$Q^T A Q = E = \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_p \end{bmatrix},$$

where each D_j is either a $(1,1)$ block (real eigenvalues) or a $(2,2)$ block of the type $\begin{pmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{pmatrix}$ (complex conjugate eigenvalues).

Pf. $(\Leftarrow)$ $Q^T A Q = E \Rightarrow Q^T A^T Q = E^T \Rightarrow A = QEQ^T$, $A^T = QE^TQ^T \Rightarrow AA^T = QEQ^TQE^TQ^T = QEE^TQ^T = QE^TEQ^T = QE^TQ^TQEQ^T = A^TA$.

$$\Big(\text{Note:} \quad \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix}\Big)$$

$(\Rightarrow)$ Using the real Schur Theorem 2.2.4, we can assume that A is in quasi-triangular form as

$$A = \begin{bmatrix} R_{11} & A_{12} & \cdots & \cdots & A_{1p} \\ & R_{22} & A_{22} & \cdots & A_{2p} \\ & & \ddots & \ddots & \\ & & & & R_{pp} \end{bmatrix},$$

where R_{11} is upper triangular and has all the real eigenvalues of A, and $R_{22}, \dots, R_{pp}$ are $(2,2)$ real matrices corresponding to complex conjugate eigenvalues.

Now, since $A^TA = AA^T$, then $R_{11}^T R_{11} = R_{11}R_{11}^T + A_{12}A_{12}^T + \cdots + A_{1p}A_{1p}^T$ and therefore also $\mathrm{tr}(R_{11}^TR_{11}) = \mathrm{tr}(R_{11}R_{11}^T + \cdots + A_{1p}A_{1p}^T) = \mathrm{tr}(R_{11}R_{11}^T) + \mathrm{tr}(A_{12}A_{12}^T) + \cdots + \mathrm{tr}(A_{1p}A_{1p}^T)$. But, $\mathrm{tr}(R_{11}R_{11}^T) = \mathrm{tr}(R_{11}^TR_{11})$, and so we must have $\mathrm{tr}(A_{12}A_{12}^T) + \cdots + \mathrm{tr}(A_{1p}A_{1p}^T) = \sum (A_{12})_{ij}^2 + \cdots + \sum (A_{1p})_{ij}^2 = 0$ from which it follows that $A_{12} = 0, \ \cdots \ , \ A_{1p} = 0$. So, A really has the form

$$A = \begin{bmatrix} R_{11} & 0 & \cdots & \cdots & 0 \\ & R_{22} & A_{22} & \cdots & A_{2p} \\ & & \ddots & \ddots & \\ & & & & R_{pp} \end{bmatrix}.$$

Now, reasoning as in the proof of Theorem 2.3.12, since R_{11} is upper triangular and normal, then it must be diagonal, call it D_1.

Now, since $A^TA = AA^T$, then we must also have $R_{22}^TR_{22} = R_{22}R_{22}^T + A_{23}A_{23}^T + \cdots + A_{2p}A_{2p}^T$ and again $\mathrm{tr}(R_{22}^TR_{22}) = \mathrm{tr}(R_{22}R_{22}^T) + \mathrm{tr}(A_{23}A_{23}^T) + \cdots + \mathrm{tr}(A_{2p}A_{2p}^T)$, from which at once we get $A_{23} = 0, \cdots, A_{2p} = 0$, and R_{22} is normal.

Continuing this way, we end up with

$$A = \begin{bmatrix} D_1 & & & \\ & R_{22} & & \\ & & \ddots & \\ & & & R_{pp} \end{bmatrix},$$

where each R_{jj} is $(2,2)$, normal, with complex conjugate eigenvalues. So, we now verify that a real matrix $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with complex conjugate eigenvalues and normal, $B^TB = BB^T$, must have the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. We have

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & ac+bd \\ ac+bd & c^2+d^2 \end{pmatrix}$$

$$\therefore \begin{cases} a^2+c^2 &= a^2+b^2 \\ d^2+b^2 &= d^2+c^2 \to c^2 = b^2 \Rightarrow b = c \text{ or } b = -c \\ ab+cd &= ac+bd \to a(b-c) = d(b-c) \end{cases}.$$

Now, if $b = c \Rightarrow B$ symmetric $\therefore$ it has real eigenvalues, which is excluded. So, it must be $b = -c \Rightarrow 2ba = 2bd \Rightarrow a = d$ ($b \neq 0$ otherwise again B would be symmetric) $\therefore B = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. □

Remark 2.3.16 *In Theorem 2.3.15, note the specific form of the* $(2,2)$ *blocks of complex conjugate eigenvalues.*

- Who is "normal"? For example, a diagonal matrix, or Hermitian, or anti-Hermitian, or unitary, or the direct sum of any of these is normal. In the appropriate basis, these types of matrices are all diagonal. In the real case, see Theorem 2.3.15, one can only achieve quasi-diagonal structure. For example, the next result holds as immediate consequence of Theorem 2.3.15.

Corollary 2.3.17 (Spectral Theorem for antisymmetric matrices) *Let $A \in \mathbb{R}^{n\times n}$, $A^T = -A$. Then, there exists $Q \in \mathbb{R}^{n\times n}$, orthogonal, such that*

$$Q^T A Q = E = \begin{bmatrix} 0_{n_1} & & & \\ & D_2 & & \\ & & \ddots & \\ & & & D_p \end{bmatrix},$$

where each D_j, $j = 2, \dots, p$, is a $(2,2)$ block of the type $\begin{pmatrix} 0 & \beta_j \\ -\beta_j & 0 \end{pmatrix}$.

Exercises 2.3.18

(1) $A \in \mathbb{C}^{n\times n}$ is normal $\Leftrightarrow \|Ax\|_2 = \|A^*x\|_2$, $\forall\, x \in \mathbb{C}^n$.

(2) $A \in \mathbb{C}^{n\times n}$ is normal $\Leftrightarrow \exists\, U$, unitary: $A^* = AU$.

(3) $A \in \mathbb{C}^{n\times n}$, $z \in \mathbb{C}$ be given. A is normal $\Leftrightarrow A + zI$ is normal.

(4) If A is normal and $q(t)$ is a polynomial $\Rightarrow q(A)$ is normal.

(5) Show that if $A \in \mathbb{C}^{n\times n}$ is normal, ker(A) = ker(A^*). [This is trivial after the next exercise, so do it without using it.]

(6) Show that $A \in \mathbb{C}^{\times n}$ is normal $\Leftrightarrow$ every eigenvector of A is also eigenvector of A^*.

(7) Show that if $A \in \mathbb{C}^{n\times n}$ is normal $\Rightarrow \|A\|_2 = \max_j |\lambda_j(A)|$, where $\lambda_j(A)$ are the eigenvalues of A. Is the converse also true?

*(8) Show that $\|A^*A\|_2 = \max_j(\lambda_j(A^*A))$. [*Hint*: Observe that the eigenvalues of A^*A are ≥ 0, then diagonalize A^*A with a unitary transformation.]*

Theorem 2.3.19 (Spectral norm for matrices) *Let $A \in \mathbb{C}^{n\times n}$ (or $\mathbb{R}^{n\times n}$). Then*

$$\|A\|_2 = \max_j (\lambda_j(A^*A))^{1/2}\ .$$

Pf. Recall that $\|A\|_2 = \max_{x:\|x\|_2=1} \|Ax\|_2$. Let x be an aribitrary unit vector, $\|x\|_2 = 1$. Then, we have

$$\|Ax\|_2^2 = \langle Ax, Ax\rangle = x^*A^*Ax = \langle A^*Ax, x\rangle \overset{Cauchy-Schwartz}{\leq} \|x\|_2\|A^*Ax\|_2$$

$\therefore$ $\|Ax\|_2^2 \leq \|A^*Ax\|_2 \Rightarrow \|A\|_2^2 \leq \|A^*A\|_2 = \max_j(\lambda_j(A^*A)) = \lambda_{\max}(A^*A)$, where we used Exercise 2.3.18-(8).

Now, let $x_{\max}$ be an eigenvector (of norm 1) such that $A^*Ax_{\max} = \lambda_{\max}x_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of A^*A. Then, $\langle A^*Ax_{\max}, x_{\max}\rangle = \lambda_{\max} = \|Ax_{\max}\|_2^2$. □

Exercises 2.3.20

(1) Given $A \in \mathbb{C}^{n\times n}$, normal, characterize rank(A), nullity(A), Ker(A), Im(A) in terms of A's eigenvalues/eigenvectors. If A is not normal, relate its rank to its eigenstructure.

(2) Show that $\|A\|_2 = \|A^*\|_2$ for any $A \in \mathbb{C}^{n\times n}$.

(3) Let $A \in \mathbb{C}^{n\times n}$, and λ_j, $j = 1,\dots,n$, be its eigenvalues. Show that $\|A\|_2 \geq \max_j |\lambda_j|$.

2.3.1 Projections from Eigenspaces

Let us now look at some implications of the spectral theorems for Hermitian (symmetric) matrices (Theorems 2.3.6 and 2.3.8).

What these theorems say is that $\mathbb{C}^n$ (or $\mathbb{R}^n$) can be decomposed as the direct sum of pairwise orthogonal eigenspaces of a Hermitian (symmetric) matrix A. That is, if A has distinct eigenvalues $\lambda_1,\dots,\lambda_p$, of multiplicity $n_1,\dots,n_p$, respectively, then:

$$\mathbb{C}^n = V^{(1)} \oplus V^{(2)} \oplus \cdots \oplus V^{(p)} \quad (\text{or } \mathbb{R}^n = V^{(1)} \oplus \cdots \oplus V^{(p)})$$

where, for each $j = 1,\dots,p$, $V^{(j)}$ is the subspace spanned by the n_j eigenvectors relative to the eigenvalue λ_j (and $\lambda_j \neq \lambda_i$, $i \neq j$, $i,j = 1,\dots,p$).

Therefore, every given $v \in \mathbb{C}^n$ (or in $\mathbb{R}^n$) can be written (uniquely) as

$$v = v^{(1)} + \cdots + v^{(p)}, \quad v^{(j)} \in V^{(j)} \Rightarrow Av = \lambda_1 v^{(1)} + \cdots + \lambda_p v^{(p)}\,.$$

Observe that each $v^{(j)}$ is a function of v. Indeed, we have $v^{(j)} = \sum_{k=1}^{n_j} \alpha_k^{(j)} v_k^{(j)}$, so

$$v = \sum_{k=1}^{n_1} \alpha_k^{(1)} v_k^{(1)} + \cdots + \sum_{k=1}^{n_p} \alpha_k^{(p)} v_k^{(p)}$$

and thus $\alpha_k^{(j)} = (v_k^{(j)})^* v$, from which

$$v = \sum_{k=1}^{n_1} (v_k^{(1)}(v_k^{(1)})^*)v + \sum_{k=1}^{n_2} (v_k^{(2)}(v_k^{(2)})^*)v + \cdots + \sum_{k=1}^{n_p} (v_k^{(p)}(v_k^{(p)})^*)v\,.$$

Now, let $P_j = \sum_{k=1}^{n_j} v_k^{(j)}(v_k^{(j)})^*$. Then we have $v = \left(\sum_{j=1}^{p} P_j\right) v$ and $Av = \left(\sum_{j=1}^{p} \lambda_j P_j\right) v$. Therefore (since v is arbitrary), these give $\begin{cases} I = \sum_{j=1}^{p} P_j \\ A = \sum_{j=1}^{p} \lambda_j P_j\,. \end{cases}$

The following properties are immediate.

(1) $P_jP_\ell = 0$, $j \neq \ell$, and $P_j^2 = P_j$. [Obvious, since $P_j = \sum_{k=1}^{n_j} v_k^{(j)}(v_k^{(j)})^*$ and $\{v_k^{(j)}\}_{\substack{k=1:p \\ j=1:n_j}}$ is an orthonormal set.]

(2) $P_j^* = P_j$ ($\Rightarrow P_j^*P_j = P_j^2 = P_j$). [Again obvious from the form of P_j].

Therefore, each P_j is *symmetric and idempotent*. These two properties characterize an *orthogonal projection* operator: Each P_j projects (orthogonally) a vector onto the eigenspace $V^{(j)}$ (i.e., onto the subspace spanned by the eigenvectors in $V^{(j)}$).

The formula

$$I = \sum_{j=1}^{p} P_j \text{ is called } \textit{resolution of the identity}$$

$$\text{and} \quad A = \sum_{j=1}^{p} \lambda_j P_j \text{ is called } \textit{spectral resolution of } A.$$

Note that the resolution of the identity holds no matter what was the matrix A from which we formed the eigenspaces and the projections, whereas in the spectral resolution of A we must use the projections formed by A itself.

Example 2.3.21 *Suppose $A = \sum_{j=1}^{p} \lambda_j P_j$ as above.*

(a) Then $A^2 = \left(\sum_{j=1}^{p} \lambda_j P_j\right)\left(\sum_{j=1}^{p} \lambda_j P_j\right) = \sum_{j=1}^{p} \lambda_j^2 P_j$ and inductively $A^m = \sum_{j=1}^{p} \lambda_j^m P_j$.

(b) $q(A) = \sum_{j=1}^{p} q(\lambda_j) P_j$ for any polynomial in A

(c) If $f(A) = e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} \Rightarrow e^A = \sum_{j=1}^{p} e^{\lambda_j} P_j$.

The most important consequence of having an orthogonal projection operator is the possibility to approximate, optimally, from within a certain subspace.

Example 2.3.22 (On Best Approximation) *Let us see one common use of* projections *formed from the spectral decomposition of a symmetric matrix: $A \in \mathbb{R}^{n\times n}$, $A = A^T$. Recall that*

$$\mathbb{R}^n : V^{(1)} \oplus V^{(2)} \oplus \cdots \oplus V^{(k)}\ , \quad \begin{cases} I = \sum_{j=1}^{k} P_j, & P_jP_\ell = 0,\ j \neq \ell \\ A = \sum_{j=1}^{k} \lambda_j P_j, & P_j^2 = P_j, P_j^T = P_j \end{cases}$$

$$\textit{and} \quad P_j = \sum_{i=1}^{n_j} v_i^{(j)}(v_i^{(j)})^T, \quad j = 1, \ldots, k,$$

are the orthogonal projections onto the subspaces $V^{(j)}$'s.

Now, suppose we want to solve this frequently arising problem: "Find $v \in V^{(j)}$ closest (in the 2-norm) to a given vector $b \in \mathbb{R}^n$." [Best approximation out of a subspace.]

That is, we want

$$\min_{v \in V^{(j)}} \|v - b\|_2^2 .$$

Since $v \in V^{(j)} \Rightarrow v = P_j v$ and we can write $b = I \cdot b = \sum_{j=1}^k P_j b$. So, we have $\|P_1 b + \cdots + P_j b + \cdots + P_k b - P_j v\|^2 = (P_1 b + \cdots + P_j(b - v) + \cdots + P_k b)^T (P_1 b + \cdots + P_j(b - v) + \cdots + P_k b) = \|P_1 b\|^2 + \cdots + \|P_j b - v\|^2 + \cdots + \|P_k b\|^2$ and the only part we can control is $\|P_j b - v\|$. As a consequence, the minimum is obtained by choosing $v = P_j b$. □

Definition 2.3.23 *The vector $v = P_j b$ is called the orthogonal projection of b onto $V^{(j)}$.*

The following consequence of the above is now immediate.

Theorem 2.3.24 *In the 2-norm, the orthogonal projection is the best approximation to b by a vector in $V^{(j)}$.*

Example 2.3.25 *Suppose we have the overdetermined system of equations $Ax = b$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $m > n$, and A is full rank. We seek $x \in \mathbb{R}^n$ such that $\|Ax - b\|_2^2$ is minimized. We recognize this as a standard least squares problem, and we know that if A is full rank, then the solution (via a simple calculation) is given by $x = (A^T A)^{-1} A^T b$ (in the notation of Example 2.3.22, it is $v = Ax$). Therefore, $A(A^T A)^{-1} A^T$ must be the orthogonal projection onto range of A. Let*

$$P = A(A^T A)^{-1} A^T ,$$

and let us verify that this $P \in \mathbb{R}^{m \times m}$ is the orthogonal projection onto $\mathcal{R}(A)$.

1) Obviously, $P = P^T$, and $P^2 = A(A^T A)^{-1} A^T A (A^T A)^{-1} A^T = P$.

2) Clearly $Px \in \mathcal{R}(A)$, $\forall\, x \in \mathbb{R}^m$.

2) Now, let $b \in \mathcal{R}(A) \Rightarrow b = Ay$, $y \in \mathbb{R}^m$, so we need to show that $\|b - z\|$ for $z \in \mathcal{R}(A)$ is minimized by $z = A(A^T A)^{-1} A^T b$. But this is exactly how we derived it: $\|b - z\|_2 = \min_{y \in \mathbb{R}^n} \|b - Ay\|_2 \Rightarrow y = (A^T A)^{-1} A^T b$.

To conclude this Example, we observe that there has to be a complementary orthogonal projection as well, which we find from the resolution of identity, which projects onto the orthogonal complement of $\mathcal{R}(A)$. We call it $P^{\perp} = I - P$ and $\therefore$ $P^{\perp} = I - A(A^T A)^{-1} A^T$ and clearly $P^{\perp} P = 0$. □

As it turns out, every projection has a very simple canonical form. Below we find this form working in $\mathbb{R}$, but same result holds in $\mathbb{C}$. To begin with, we have the following result.

Exercise 2.3.26 *Let P be the orthogonal projection onto a nontrivial subspace V of $\mathbb{R}^n$. Then: $\|P\|_2 = 1$.*

Solution. $\|P\|_2^2 = \lambda_{\max}(P^T P) = \lambda_{\max}(P^2) = \lambda_{\max}(P)$. Now, take x and write it as $x = x_P + x_{P^\perp} = Px + (I-P)x \Rightarrow \|Px\|_2 = \|Px_P\|_2 = \|x_P\|_2$.
$\therefore$ $\max_{x\in\mathbb{R}^n,\,\|x\|=1} \|Px\|_2^2 = \max_{x\in\mathcal{R}(P),\,\|x\|=1} \|Px\|_2^2 = \max_{x\in\mathcal{R}(P),\,\|x\|=1} x^T PPx = x^T x = 1$
$\therefore$ P has largest eigenvalue equal to 1. $\square$

Moreover, in Exercise 2.3.26 we have used that any $x \in \mathcal{R}(P)$ gives $Px = x$ and any $x \in \mathcal{R}(I-P)$ gives $Px = 0$. Because of the spectral resolution relative to P, this means that P has as many eigenvalues equal to 1 as $\operatorname{rank}(P)$, the rest being all 0's.

The following corollary is now an immediate consequence of symmetry, spectral theorem for symmetric matrices, and rank of an orthogonal projection onto a p-dimensional subspace.

Corollary 2.3.27 *After a change of coordinates, each orthogonal projection onto a p-dimensional subspace can be written as $P = \begin{bmatrix} I_p & 0 \\ 0 & 0 \end{bmatrix}$.* $\square$

Exercise 2.3.28 $A = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$. *Clearly A is a projection ($A^T = A$, $A^2 = A$) and has rank 1 and therefore A has one eigenvalue equal to 1 and $(n-1)$ eigenvalues equal to 0.*

$$\text{If } Q : Q^T A Q = \Lambda \Rightarrow \Lambda = \begin{bmatrix} 1 & & & 0 \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix} \Rightarrow A = q_1 q_1^T \ \text{(and } q_1 = \tfrac{1}{\sqrt{n}}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}\text{)}. \qquad \square$$

2.3.2 Positive definiteness and congruence

An important class of Hermitian (symmetric) matrices are those which have all positive eigenvalues.

Definition 2.3.29 *$A \in \mathbb{C}^{n\times n}$ (respectively, in $\mathbb{R}^{n\times n}$), $A^* = A$ (respectively, $A^T = A$), is called positive definite if $\langle x, Ax\rangle > 0$, $\forall\, x \neq 0$.*

- Similar definitions exist for nonnegative definite, negative definite, and nonpositive definite matrices. Be aware that often nonnegative definite matrices are called positive semi-definite, and nonpositive definite matrices are called negative semi-definite. Also, be aware that we only defined positive definite matrices which are Hermitian. (One could define positive definite non-Hermitian matrices as those matrices A for which $\langle x, Ax\rangle > 0$, $\forall\, x \neq 0$; for example, the real matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ would be positive definite).

Exercises 2.3.30

(1) *(a) Show that if A, B are positive definite $\Rightarrow A + B$ is too. (b) Show that if A is positive definite, and $\alpha \in \mathbb{R}$, $\alpha > 0$, then αA is also positive definite.* [These two properties say that the set of positive definite matrices is a *cone*.]

(2) *Show that if A is positive definite and S is any invertible matrix $\Rightarrow S^*AS$ is positive definite.* [This fact we will encounter again, it is a "congruence" relation.] *In particular, positive definiteness is preserved by unitary similarity.*

(3) *Show that A is positive definite $\Leftrightarrow$ all its eigenvalues are positive.*
As consequences of this fact, we have:
a) $\langle x, Ax\rangle \geq \lambda_{\min}\|x\|^2$, $\forall\, x$.
b) A is invertible and A^{-1} is also positive definite.

Let us now verify some important properties of positive definite matrices.

Lemma 2.3.31 (Positive definite square root.) *If A is positive definite, then $\exists!\ B$, positive definite, such that $B^2 = A$.*

Pf. ($\exists$) Take U such that $U^*AU = D$, diagonal and positive definite: $D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$, $d_i > 0$. Then, define the (diagonal and positive definite) matrix

$D^{1/2} = \begin{pmatrix} d_1^{1/2} & & 0 \\ & \ddots & \\ 0 & & d_n^{1/2} \end{pmatrix}$, so that $(D^{1/2})^2 = D$. Now take $B = UD^{1/2}U^*$, which is clearly positive definite and satisfies $B^2 = A$. Next, let us show uniqueness. We verify this, while also showing other things and introducing a useful technique.

Since $B^2 = A$ and $B = UD^{1/2}U^* \Rightarrow AB = BA$. Moreover, consider the following data set:
$\{(\lambda_1, \sqrt{\lambda_1}), (\lambda_2, \sqrt{\lambda_2}), \dots, (\lambda_n, \sqrt{\lambda_n})\}$ and further the subset given by the distinct points in this data set, call it (after possible relabeling) $\{(\lambda_1, \sqrt{\lambda_1}), (\lambda_2, \sqrt{\lambda_2}), \dots, (\lambda_p, \sqrt{\lambda_p})\}$. Now, let $p(t)$ by the unique interpolatory polynomial of degree $\le (p-1)$ through these points. Then, we have (see below for this computation)

$$p(D) = D^{1/2} \Rightarrow p(A) = p(UDU^*) = Up(D)U^* = UD^{1/2}U^* = B \ ,$$

so that *There is a polynomial $p(t)$ such that $p(A) = B$.*

Now, suppose that C is another positive definite square root of A: $C^2 = A \Rightarrow B = p(A) = p(C^2) \Rightarrow CB = Cp(C^2) = p(C^2)C = BC \Rightarrow B$ and C commute and being both Hermitian, then are simultaneously unitarily diagonalizable. That is: $\exists V : V^*BV = \Lambda_B$ and $V^*CV = \Lambda_C$, but since $C^2 = B^2 = A \Rightarrow \Lambda_c^2 = \Lambda_B^2$ but positive square root is unique and so $\Lambda_C = \Lambda_B \Rightarrow B = C$.

Finally, let us verify that $p(D) = D^{1/2}$. The interpolatory polynomial in Lagrange form is

$$p(t) = \sqrt{\lambda_1}L_1(t) + \cdots + \sqrt{\lambda_p}L_p(t) \ , \quad \text{where} \quad L_i(t) = \frac{\prod_{j=1, j\neq i}^{p}(t-\lambda_j)}{\prod_{j=1, j\neq i}^{p}(\lambda_i-\lambda_j)} \ , \ i = 1, \dots, p$$

and thus

$$p(D) = \sqrt{\lambda_1}L_1(D) + \cdots + \sqrt{\lambda_p}L_p(D) \ .$$

Now, examine the terms $L_1(D), \dots, L_p(D)$. We have

$$L_1(D) = \frac{1}{\prod_{j=2}^{p}(\lambda_1-\lambda_j)}(D-\lambda_2 I)\cdots(D-\lambda_p I) \ ,$$

and therefore a simple computation gives $L_1(D) = \mathrm{diag}(I_{n_1}, 0_{n_2}, \dots, 0_{n_p})$. Similarly for $L_2(D), \dots, L_p(D)$, fom which $p(D) = \sqrt{D}$. $\square$

Lemma 2.3.32 (Open cone.) *Consider the set $\mathcal{P} = \{A \in \mathbb{C}^{n\times n} : A^* = A$ and A is positive definite$\}$. Then $\mathcal{P}$ is an open set in the set of all Hermitian matrices. Similarly for real symmetric matrices.*

Pf. Let $A = A^*$ be positive definite and let $\lambda_m = \lambda_{\min}(A)$ (> 0). Let $B = B^*$ be such that $\|A - B\| < \lambda_m$. We claim that B is positive definite. (Notice that, then, it will follow that $\mathcal{P}$ is open).

Let $M = A - B$. We know that for any $x \neq 0$:

$$\begin{cases} \|Mx\| < \lambda_m \|x\| \\ \langle x, Ax \rangle \geq \lambda_m \|x\|^2 \ . \end{cases}$$

Now, for any $x \neq 0$, we have:

$$0 \leq |\langle x, Mx \rangle| \leq \|x\| \cdot \|Mx\| < \lambda_m \|x\|^2 \,,$$

and also

$$\begin{aligned} \langle x, Bx \rangle = \langle x, (A - M)x \rangle = \langle x, Ax \rangle - \langle x, Mx \rangle &\geq \lambda_m \|x\|^2 - \langle x, Mx \rangle \\ &> \lambda_m \|x\|^2 - \lambda_m \|x\|^2 = 0 \ . \end{aligned}$$

Therefore, $\langle x, Bx \rangle > 0$, $\forall\, x \neq 0$ and so B is positive definite. □

Lemma 2.3.33 (Boundary of Cone) *The boundary elements of the set of positive-definite matrices are the nonnegative definite matrices, which are not positive positive definite.*

Pf. We need to look at matrices which are limits of positive definite matrices:

$$B = \lim_{k \to \infty} A_k, \quad A_k = A_k^* \text{ positive definite.}$$

Now, $\forall\, x \neq 0 : \langle x, A_k x \rangle > 0 \therefore \lim_{k\to\infty} \langle x, A_k x \rangle \geq 0 \therefore \langle x, Bx \rangle \geq 0$, $\forall\, x \in \mathbb{C}^n \therefore$ B is nonnegative definite. But B cannot be positive definite, otherwise by Lemma 2.3.32 it could not be a boundary element. Next, we show that if B is a nonnegative definite matrix, then it is on the boundary. But, for this, it is enough to take the sequence of positive definite matrices given by $A_k = B + \frac{1}{k} I$. □

Lemmata 2.3.32 and 2.3.33 give us the geometrical description of the set of positive definite matrices, to which we simply add that the vertex of the cone is the origin (the 0-matrix).

- Several other useful characterizations of positive definite matrices exist as well. We recall a few below.

Theorem 2.3.34 *The following statements are equivalent:*
0) $A = A^*$ *is positive definite.*
1) $\exists\, C$ *nonsingular, such that* $A = CC^*$.
2) $\exists!\, L$, *lower triangular with positive diagonal such that* $A = LL^*$.

Statement (1) in Theorem 2.3.34 is another instance of a congruence relation (see below). Statement (2) gives the so-called Choleski factorization of A.

Exercises 2.3.35

(1) Show that if A is positive definite $\Rightarrow$ it has a unique positive definite k-th root, for any $k \in \mathbb{Z}$.
(2) Prove Theorem 2.3.34.
(3) Show that if $A = A^*$ is nonnegative definite, then it is has a unique nonnegative definite square root.
(4) [Signed Choleski] *Show that $A = A^*$ is positive definite if and only if there exists L, lower triangular and nonsingular, such that $A = LL^*$.*
(5) Let $A \in \mathbb{C}^{n\times n}$ be positive definite. How many Hermitian square roots of A are there? How many Hermitian cubic roots? And Hermitian k-th roots ($k = 4, \dots$)? Are there any non-Hermitian square roots (that is, matrices B such that $B^2 = A$)?
(6) Let $A \in \mathbb{C}^{2\times 2}$ be nonnegative definite. How many Hermitian square roots of A are there? How many Hermitian and non-Hermitian square roots?
(7) [Harder]. *Take any $A \in \mathbb{C}^{2\times 2}$, possibly not even Hermitian. Discuss when A has a square root B, that is a matrix $B \in \mathbb{C}^{2\times 2}$ such that $B^2 = A$, and discuss uniqueness of B. Further, discuss when/if B is a polynomial of A.*

- Another useful characterization of positive definiteness is in terms of the Gram-matrix.

Definition 2.3.36 *Given vectors $v_1, \dots, v_p$ in $\mathbb{C}^n$, the matrix G of entries $G_{ij} = \langle v_j, v_i\rangle$ is called Gram-matrix (or Gramian) of the given vectors.*

Observation 2.3.37 *Since $G_{ij} = \langle v_j, v_i\rangle = v_i^* v_j = \bar{v}_i^T v_j \Rightarrow \bar{G}_{ji} = (\overline{\bar{v}_j^T v_i}) = v_j^T \bar{v}_i = \bar{v}_i^T v_j = G_{ij} \Rightarrow G^* = G$. So, the Gramian is always Hermitian.*

Theorem 2.3.38 *For a Gram matrix G, the following hold.*
1) G is nonnegative.
2) The vectors $\{v_1, \dots, v_n\}$ are linearly independent $\iff$ G is positive definite.
3) Every positive definite matrix $A \in \mathbb{C}^{n\times n}$ is the Gram matrix of n vectors in $\mathbb{C}^n$.

Pf. For any vector x, we have

$$\langle x, Gx\rangle = x^* G^* x = x^* G x = \sum_{i,j=1}^{n} G_{ij}\bar{x}_i x_j = \sum_{i,j=1}^{n} \langle v_j, v_i\rangle \bar{x}_i x_j$$

$$= \sum_{i,j=1}^{n} \langle x_j v_j, x_i v_i \rangle = \left\langle \sum_{j=1}^{n} x_j v_j, \sum_{i=1}^{n} x_i v_i \right\rangle = \left\| \sum_{i=1}^{n} x_i v_i \right\|^2$$

and so 1) and 2) follow at once.

To get 3), let B be such that $B^2 = A$, $B = B^* \Rightarrow A = B^* B \Rightarrow A_{ij} = \langle b_j, b_i \rangle$ □

Exercises 2.3.39

(1) *Suppose we have two set of vectors $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ in $\mathbb{C}^n$ giving the same Gramian. What is the relation between the matrices $V = [v_1, \ldots, v_n]$ and $W = [w_1, \ldots, w_n]$?* (Hint: *You may want to first assume that the matrices are invertible. Then, it may be simpler to do this exercise after section 2.4.2.*)

(2) *How does the Gramian of a set of orthogonal vectors look like? Of an orthonormal set?*

Remark 2.3.40 *Gram matrices can be built over any inner product space. See Examples below.*

Example 2.3.41 [Hilbert matrix] *Consider $C([0,1], \mathbb{R})$ with the usual inner product $(f,g) = \int_0^1 f(t)g(t)dt$, and the Gram matrix of entries $G_{ij} = (f_j, f_i)$. Now, take $f_j = t^{j-1}$, so that $(f_j, f_i) = G_{ij} = \frac{1}{i+j-1}$. This gives the famous Hilbert matrix, which is therefore positive definite, since the functions $\{f_k = t^{k-1}\}$ are linearly independent. The Hilbert matrix is a standard example of a severely* ill-conditioned *matrix for large n. The so-called* condition number *of G, $\|G\|\|G^{-1}\|$, gets large for n large.*

Example 2.3.42 [Hankel matrix] *Given a positive function f, $f(t) > 0$ for $t \in [0,1]$, consider the* algebraic moments

$$a_k = \int_0^1 t^k f(t)dt, \quad k = 0, 1, 2, \ldots \ .$$

To these, we associate the quadratic form

$$\sum_{j,k=0}^{n} a_{j+k} x_j x_k = \sum_{j,k=0}^{n} \int_0^1 t^{j+k} x_j x_k f(t)dt \ ,$$

and the matrix $A \in \mathbb{R}^{n+1,n+1}$ *defined by* $(A)_{ij} = a_{i+j}$*. Obviously,* $A = A^T$ *and* A *is positive definite because*

$$x^T A x = \sum_{j,k=0}^{n} a_{j+k} x_j x_k = \int_0^1 \left(\sum_{k=0}^{n} x_k t^k \right)^2 f(x) dx \ .$$

Notice that the entries of A *are functions only of* $i + j$*. Every time this fact holds, we have a* Hankel matrix*.*

Exercise 2.3.43 [Toeplitz matrix] *Consider the trigonometric moments:*

$$a_k = \int_0^1 e^{ikt} f(t) dt, \quad k = 0, \pm 1, \pm 2, \ldots,$$

and the associated quadratic form: $\sum_{j,k=0}^{n} a_{j-k} z_j \bar{z}_k$*. Form the matrix* A *and verify that its entries only depend on* $i - j$*. These are called* Toeplitz matrices*. Show that if* $f > 0$*, then* A *is positive definite.*

- Statement (1) of Theorem 2.3.34 told us that $A = A^*$ is positive definite $\Leftrightarrow$ $A = CC^*$ for an invertible matrix C. By looking at this as $C^{-1} A C^{-*} = I$, then we'd have that A is *congruent* to the identity.

Definition 2.3.44 *Given* $A = A^*$ *and* S *invertible, then* $B = SAS^*$ *is called congruent to* A*. (Similarly in real case:* $B = SAS^T$*.)*

Observe that congruence is an equivalence relation. Of course, congruence is not the same as similarity (unless S is unitary), but it still allows for some simplifications in many problems.

Remark 2.3.45 *It is useful to think of similarity as a simplification of the "dynamics" of a problem. For example, consider the discrete dynamical system* $y^{(k+1)} = Ay^{(k)}$*. The change of variable* $v = V^{-1} y$ *gives* $v^{(k+1)} = (V^{-1} A V) v^{(k)}$ *and* V *should be chosen so that* $V^{-1} A V$ *is simpler (say, in Jordan form). As the example below will show, in some contexts, congruence arises naturally when we change the "internal" variables of a system.*

Example 2.3.46 *Let* f *be a twice continuously differentiable function in some domain* D *of* $\mathbb{R}^n$*, and consider the general 2nd order linear differential operator* L*:*

$$Lf(x) = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \ .$$

To this operator, we associate the matrix $A = (a_{ij})$, where the entries a_{ij} are functions of $x \in D$. Observe that A may not be symmetric, but since

$$\begin{aligned}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j} &= \sum_{i,j=1}^{n}\left[\frac{1}{2}a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{1}{2}a_{ji}(x)\frac{\partial^2 f}{\partial x_j \partial x_i}\right] \\ &= \sum_{i,j=1}^{n}\frac{1}{2}(a_{ij}(x)+a_{ji}(x))\frac{\partial^2 f}{\partial x_i \partial x_j},\end{aligned}$$

then L is impacted only by the symmetric part of A and so we may as well assume that A is symmetric.

*Now, suppose that we make a nonsingular change of variables from x to $s = (s_i)_{i=1}^n$, with $s_i = s_i(x_1, \ldots, x_n)$. Nonsingularity implies that the Jacobian matrix $S = \left(\frac{\partial s_i(x)}{\partial x_j}\right)_{i,j=1}^n$ is nonsingular (and also that the change of variable is invertible: $x = x(s)$, at least locally). Under this change of variables (**Exercise**: Verify (lengthy, but easy)), we get*

$$Lf = \sum_{i,j=1}^{n} \underbrace{\left(\sum_{p,q=1}^{n}\frac{\partial s_i}{\partial x_p}a_{pq}\frac{\partial s_j}{\partial x_q}\right)}_{b_{ij}(s)}\frac{\partial^2 f}{\partial s_i \partial s_j}\ .$$

Thus, if we let $B = (b_{ij})_{i,j=1}^n$, then $B = SAS^T$. Intuitively, we would expect that physical laws described by the original differential equation will not be impacted by this change of coordinates. □

Motivated by the above, we will now address the following questions:

(1) What are the invariants of congruence transformations?
(2) What kind of simplifications can we achieve by congruence?

The answers turn out to be surprisingly simple.

Definition 2.3.47 (Inertia) *Let $A = A^* \in \mathbb{C}^{n\times n}$ (or $A = A^T \in \mathbb{R}^{n\times n}$). We define* inertia *of A, and write it $i(A)$, to be the triplet $i(A) = (n_+(A), n_-(A), n_0(A))$ where*

$n_+(A) = \#$ *positive eigenvalues of A,*
$n_-(A) = \#$ *negative eigenvalues of A,*
$n_0(A) = \#$ *zero eigenvalues of A.*

Further, we define inertia matrix *of A the matrix* $I(A) = \begin{bmatrix} I_{n_+(A)} & & \\ & -I_{n_-(A)} & \\ & & 0_{n_0(A)} \end{bmatrix}.$

Remark 2.3.48 *The pair $(n_+(A), n_-(A))$ is often called* signature *of A. But, be aware that there is no consensus on this terminology.*

Theorem 2.3.49 (Inertia Matrix) *Let $A = A^* \in \mathbb{C}^{n\times n}$ (or $A = A^T \in \mathbb{R}^{n\times n}$), and let $i(A) = (n_+(A), n_-(A), n_0(A))$ be its inertia. Then, A is congruent to its own inertia matrix.*

Pf. Let U be such that $U^*AU = \Lambda$ where in Λ we first put all positive eigenvalues then all the negative ones, then all the 0 ones. Accordingly, taking the positive square root, define the positive definite matrix D:

$$D = \operatorname{diag}\left(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{n_+}}, \sqrt{-\lambda_{n_1+1}}, \dots, \sqrt{-\lambda_{n_++n_-}}, 1, \dots, 1\right)$$
$$\Rightarrow \Lambda = DI(A)D \Rightarrow A = (UD)I(A)(DU^*)\ .$$

Now, let $S = UD$. $\square$

Theorem 2.3.50 (Sylvester law of inertia) *Two symmetric (Hermitian) matrices are congruent iff they have same inertia, and therefore the same inertia matrix.*

Pf. ($\Leftarrow$) If $i(B) = i(A) \Rightarrow I(B) = I(A)$, and therefore $B = S_BI(B)S_B^T = S_BI(A)S_B^T = S_BS_A^{-1}S_AI(A)S_A^T(S_A^T)^{-1}S_B^T \Rightarrow B = (S_BS_A^{-1})A(S_BS_A^{-1})^T$
$\therefore$ A and B are congruent.

($\Rightarrow$) Let S invertible be such that $A = SBS^T$. Obviously, $\operatorname{rank}(A) = \operatorname{rank}(B)$ and (since A and B are symmetric) $\operatorname{rank}(A) = n - n_0(A) = n - n_0(B) = \operatorname{rank}(B)$, so that $n_+(A) + n_-(A) = n_+(B) + n_-(B)$. Now we show that $n_+(A) = n_+(B)$. Note:

$$I(A) = \begin{bmatrix} I_{n_+(A)} & & \\ & -I_{n_-(A)} & \\ & & 0_{n_0} \end{bmatrix}, \quad I(B) = \begin{bmatrix} I_{n_+(B)} & & \\ & -I_{n_+(B)} & \\ & & 0_{n_0} \end{bmatrix}$$

and since $A = S_AI(A)S_A^T$, $B = S_BI(B)S_B^T \Rightarrow$ we have $I(B) = T^TI(A)T$, with $T = S_B^TS^TS_A^{-T}$ invertible, and $\therefore I(A)$ and $I(B)$ are congruent.

So, we want to show that two congruent inertia matrices have the same inertia.

Suppose not. Then, without loss generality $n_+(A) < n_+(B)$ and so $n_-(A) > n_-(B)$. Consider the subspace $V = \left\{ v \in \mathbb{R}^n : V = \begin{bmatrix} 0 \\ v_- \\ 0 \end{bmatrix} \begin{matrix} \}n_+(A) \\ \}n_-(A) \\ \}n_0 \end{matrix} \right\}$ and notice that $\dim(V) = n_-(A)$ and $\langle v, I(A)v\rangle < 0$, $\forall\, v \neq 0$, $v \in V$. Consider also the subspace

$W = \left\{ w \in \mathbb{R}^n : w = \begin{bmatrix} w_+ \\ 0 \\ w_0 \end{bmatrix} \begin{matrix} \}n_+(B) \\ \}n_-(B) \\ \}n_0 \end{matrix} \right\}$ and notice that $\dim(W) = n_+(B) + n_0 = n - n_-(B) = \dim(TW)$, since T is invertible. Moreover, $0 \leq \langle w, I(B)w \rangle = \langle w, T^T I(A) T w \rangle = \langle Tw, I(A)Tw \rangle, \forall\, w \in W$.

Now: $\dim(V) + \dim(TW) = n_-(A) + n - n_-(B) > n \Rightarrow \exists\, x \neq 0,\, x \in V \cap TW$. But then for this x we would have

$$\begin{aligned} \langle x, I(A)x \rangle &< 0, \quad (x \in V) \\ \langle x, I(A)x \rangle &\geq 0, \quad (x \in TW) \end{aligned}$$

which is a contradiction and therefore $n_+(A) = n_+(B)$. □

Remark 2.3.51 *In other words, under congruence, we have a partitioning of the set of Hermitian (symmetric) matrices into equivalence classes, each having the same inertia, and hence same inertia matrix*

Exercise 2.3.52 *Given the set* $H = \{A \in \mathbb{C}^{n \times n}\ ,\ A^* = A\}$*. How many equivalence classes are there for* H *under congruence?*

- An interesting question, also with physical motivation, is to decide if/when two matrices can be simultaneously diagonalized by congruence. (Note: not necessarily to the same inertia matrix.)

  A typical result reads as follows (for a proof, and similar statements, see [4]): "Let $A = A^*$, $B = B^*$ and A be nonsingular. Form $C = A^{-1}B$. Then, there exists invertible S such that SAS^* and SBS^* are both diagonal $\Leftrightarrow$ C is diagonalizable and has real eigenvalues."

2.3.3 More inner products and projections

One natural way in which congruence transformations arise is when we work with a non-Euclidean norm. The starting point is to realize that, whereas the identity matrix defines the Euclidean metric, any other positive definite matrix can be used to define a metric. In fact, any positive definite matrix can be used to define an inner product and hence a metric.

Definition 2.3.53 (Positive definite inner product) *Let* $G = G^* \in \mathbb{C}^{n \times n}$ *(or* $G = G^T \in \mathbb{R}^{n \times n}$*) be positive definite. The quantity*

$$\langle x, y \rangle_G \;=\; y^* G x$$

*is the inner product associated to G or just G-*inner product.

The G-norm of a vector $x \in \mathbb{C}^n$ is given by $\|x\|_G = (x^* G x)^{1/2}$.

Observe that –letting $G^{1/2}$ to be the unique positive definite square root of G– the G-inner product can be appreciated to be a standard weighted inner product $\langle x, y\rangle_G = \langle G^{1/2}x, G^{1/2}y\rangle_2$. Since $G^{1/2}$ is diagonalizable with a unitary transformation, and the diagonal is positive, in the appropriate system of coordinates, we are just assigning positive weights to the different coordinates. G-norms are also called ellipsoidal norms.

Example 2.3.54 *Consider $\mathbb{R}^3$, and let $G = \begin{bmatrix} 2.3104 & 1.2672 & -0.3840 \\ 1.2672 & 3.0496 & 0.2880 \\ -0.3840 & 0.2880 & 1.6499 \end{bmatrix}$. In figure 2.1 are visualizations of the sets $\{x \in \mathbb{R}^3 \ : \ \|x\|_2^2 = 1\}$ (the standard unit sphere) and $\{x \in \mathbb{R}^3 \ : \ \|x\|_G^2 = 1\}$. The latter is an* ellipsoid*, though both are unit spheres in their respective metric. In the coordinate system of its principal axes, the ellipsoid is simply $x_1^2 + 2x_2^2 + 4x_3^2 = 1$.*

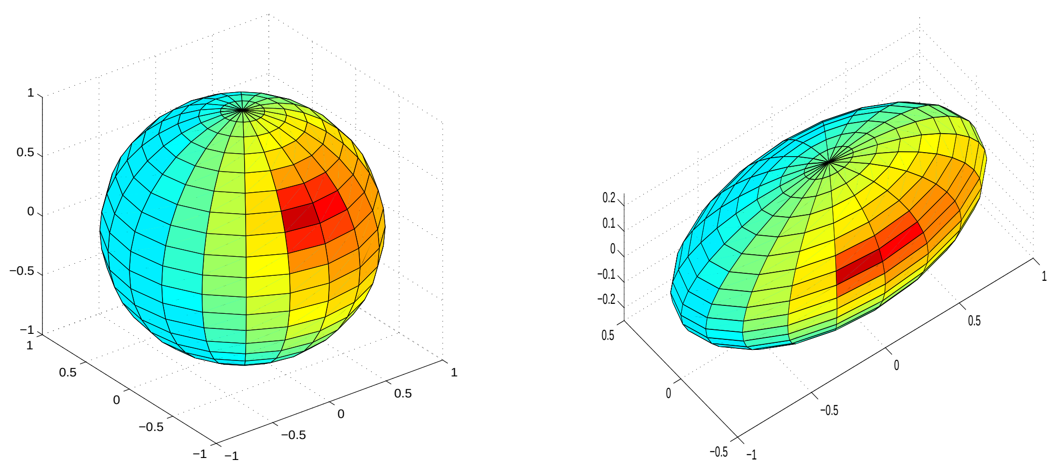


Figure 2.1: The unit sphere in the standard metric and in that induced by G.

It is a simple exercise to verify that the constructions we carried out with respect to the standard inner product can be performed with respect to a G-inner product. For example, the Cauchy-Schwartz inequality trivially generalizes:

$$|\langle x, y\rangle_G| \ \le \ \|x\|_G \|y\|_G \,,$$

and as a consequence we can also talk of angles between vectors from

$$\cos\theta = \frac{\langle x, y\rangle_G}{\|x\|_G\|y\|_G}\,,$$

and have an immediate definition of G-orthogonality:

$$\langle x, y\rangle_G = 0 \Longrightarrow x \text{ and } y \quad \text{are} \quad G-\text{orthogonal}\,.$$

With this, we can define vectors to be G-orthonormal, and extend the Gram-Schmidt process with respect to the G-inner product. Likewise, a matrix $U \in \mathbb{C}^{n\times n}$ will be called G-unitary if $U^*GU = I$. Schur's theorem and the like also have a natural extension to this new setting. As illustration, in the next result we look at the G-unitary version of the spectral theorem for Hermitian matrices. You will recognize in it a special congruence transformation.

Theorem 2.3.55 (G-unitary spectral theorem) *Let $A \in \mathbb{C}^{n\times n}$, $A^* = A$, and let $G \in \mathbb{C}^{n\times n}$ be positive definite. Then, there exist $U \in \mathbb{C}^{n\times n}$, G-unitary, such that $U^*AU = D$ and D is diagonal.*

Pf. Let $G^{1/2}$ be the unique positive definite square root of G, and define $B = G^{-1/2}AG^{-1/2}$, which is Hermitian. Let V, unitary, be such that $V^*BV = D$, diagonal. Now let $U = G^{-1/2}V$, so that $U^*GU = I$ and $U^*AU = D$. □

Remark 2.3.56 *Notice that we do not get the eigenvalues of A from a result like Theorem 2.3.55. All we can say is that the matrix D (which is real) has the same inertia as A (in fact, it is trivial to get $I(A)$ from D). Moreover, observe that the matrix U in Theorem 2.3.55 is the product of a positive definite matrix and a unitary one (this is a polar factorization, see Section 2.4).*

Once we have Theorem 2.3.55, we can define projections as well, though it is not obvious how to do it.

In the notation of Theorem 2.3.55, let $U^*AU = D$ be a G-unitary diagonalization of a Hermitian matrix A. Let $d_1, \dots, d_p$, be the distinct entries on the diagonal of D, of multiplicity $n_1, \dots, n_p$, respectively, and let $u_k^{(j)}$, $j = 1, \dots, p$, $k = 1, \dots, n_j$, be the columns of U corresponding to the values d_j in D, $j = 1, \dots, p$. So, we have the subspace decomposition of $\mathbb{C}^n$:

$$\begin{aligned} \mathbb{C}^n &= U^{(1)} \oplus U^{(2)} \oplus \cdots \oplus U^{(p)}\,, \\ \text{where} \quad U^{(j)} &= \operatorname{span}\big(u_1^{(j)}, \dots, u_{n_j}^{(j)}\big),\ j = 1, \dots, p. \end{aligned} \tag{2.3.2}$$

Now, we claim that the following are orthogonal projections in the usual sense:

$$P_j = \sum_{k=1}^{n_j} G^{1/2} u_k^{(j)} (u_k^{(j)})^* G^{1/2} \ , \ \ j = 1, \dots, p \ , \tag{2.3.3}$$

where $G^{1/2}$ is the unique positive definite square root of G. The verification that these are orthogonal projections is simple.

- That $P_j^* = P_j$ is obvious, since $G^{1/2}$ is positive definite.
- To show that $P_j P_j = P_j$ is also a direct verification using that the matrix U is G-unitary.

To define G-orthogonal projections onto a subspace $\mathcal{V}$ of dimension q (e.g., into one of the $U^{(j)}$, $j = 1, \dots, p$, above), we resort to the following (and see Exercise 2.3.57-(3)). Let $U \in \mathbb{R}^{n \times n}$ be a G-orthogonal matrix, partitioned as $U = [V, W]$, where $V \in \mathbb{R}^{n \times q}$ and the columns of V span $\mathcal{V}$. Then, we define the G-orthogonal projection onto $\mathcal{V}$ to be

$$P = V V^T G \ . \tag{2.3.4}$$

Note that $P^2 = P$, but P is not symmetric, at least not in the usual sense; this is actually a deep fact, since in the end symmetry is an inner-product dependent concept[1]. Also, note that in the case of $\mathcal{V} = U^{(j)}$, for some $j = 1, \dots, p$, then we have $P = G^{-1/2} P_j G^{1/2}$, with P_j as in (2.3.3).

Exercises 2.3.57

(1) *Find the analogous formulas to the "resolution of the identity" and "spectral resolution of A", according to the projections in* (2.3.3)*. Do it also for the G-orthogonal projections given by* (2.3.4)*, that is* $G^{-1/2} P_j G^{1/2}$.

(2) *Let* P *be a projection with respect to the* G*-norm. What are the eigenvalues of the projection* P*? Can you define the complementary projection?*

(3) *Let* $\mathcal{V}$ *be a subspace of* $\mathbb{C}^n$*, and let* $b \in \mathbb{C}^n$ *be given. Find a closed form expression for the solution of the problem*

$$\min_{x \in \mathcal{V}} \|x - b\|_G \ .$$

[In essence, this is the reason why we defined the G-projection as we did in (2.3.4).]

[1] Think about this statement.

(4) We are in $\mathbb{R}^3$*. Let* V *be the line of direction* $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$*, and let* $G = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 25 \end{pmatrix}$*.*

(a) Find a G*-orthogonal matrix whose first column spans* V*.*

(b) Solve the problem $\min_{x\in V} \|x-b\|_G$*, where* $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$*, and find the residual* $\|x-b\|_G$ *with respect to the optimal solution you found. Compare this solution to the standard least squares solution of this problem.*

2.4 Polar form and the SVD

In this section, we introduce one of the most useful decomposition of a general matrix, the Singular Value Decomposition (SVD). As a stepping stone, we will introduce the polar form of a matrix. To appreciate the latter, it will be useful to understand some important similarities between representations and transformations we are used to perform on complex numbers, and their matrix analogues. We already encountered one of them, the decomposition of a matrix $A \in \mathbb{C}^{n\times n}$ into Hermitian and anti-Hermitian parts, $A_s = (A + A^*)/2$ and $A_a = (A - A^*)/2$. Indeed, this is the analog of the representation of a complex number as sum of its real and imaginary parts:

$$z \in \mathbb{C}\ ,\ \ z = \mathrm{Re}\ z + i\mathrm{Im} z\ \ \Rightarrow A \in \mathbb{C}^{n\times n}\ ,\ \ A = A_s + A_a\ .$$

2.4.1 Polar Factorization

Let us begin with an example which extends to matrices an important transformation we do on complex numbers.

Exercise 2.4.1 (Möbius transform) *Let* $z \in \mathbb{C}$*,* $z = a + ib$*,* $a \geq 0$*, and let* $\zeta = \frac{1-z}{1+z} = \frac{(1-a)-ib}{(1+a)+ib}$*. Then,* $|\zeta| \leq 1$*. In other words, the mapping* $\frac{1-z}{1+z}$ *takes the RHP (right half plane) inside the unit disc. Moreover, it takes the imaginary axis onto unit circle.*

Solution. Rewrite $\zeta = \frac{(1-a^2-b^2)}{(1+a)^2+b^2} - i\frac{2b}{(1+a)^2+b^2}$. So, the claim is true if $(1-a^2-b^2)^2 + 4b^2 \leq ((1+a)^2+b^2)^2$, which boils down to having $a + 2a^2 + ab^2 + a^3 \geq 0$, which is true since $a \geq 0$. To verify that it takes the imaginary axis onto unit circle is equally simple:

$$a = 0 \Rightarrow \zeta = \frac{1-ib}{1+ib} \Rightarrow \zeta = \frac{1}{1+b^2}((1-b^2) - 2ib)$$

and $(1-b^2)^2 + 4b^2 = (1+b^2)^2$). $\square$

Exercise 2.4.2 *Similarly to the previous exercise, let $w = \alpha + i\beta$ with $\alpha > 0$ be a given complex number with positive real part, and consider the mapping $z \to \zeta = \frac{1-wz}{1+\bar{w}z}$. Show that this maps the RHP into unit disk $|\zeta| < 1$.*

- The last exercise has an important analog for matrices.

Theorem 2.4.3 (Generalized Cayley transform)

a) Let $w = \alpha + i\beta$ be a given complex number with $\alpha > 0$, and let $A \in \mathbb{C}^{n\times n}$ be such that A_s is positive definite. Then the matrix $B = (I - wA)(I + \bar{w}A)^{-1}$ is such that $\|B\| < 1$. Conversely, if $\|B\| < 1$, with $B = (I - wA)(I + \bar{w}A)^{-1}$, then A_s is positive definite. Here, the norm is the 2-norm.

b) In the special case of $w = 1$ and all eigenvalues of A_s being 0, then B is unitary[2]*.*

Pf. Let us prove (a). Observe that, since $A_s = A_s^*$ is positive definite $\Rightarrow \mathrm{Re}(\lambda) > 0$, where λ is any eigenvalue of A. Since all eigenvalues of $I + \bar{w}A$ are of the form $1 + \bar{w}\lambda$, with λ eigenvalue of A, they are non-zero and $I + \bar{w}A$ is invertible. So, B is well posed. Now, $\forall x$, let $y = (I + \bar{w}A)^{-1}x \to x = (I + \bar{w}A)y \Rightarrow Bx = (I - wA)(I + \bar{w}A)^{-1}x = (I - wA)y$. We want $\|Bx\|^2 < \|x\|^2$, $\forall x \neq 0$. This is true if and only if

$$\begin{aligned}
\|(I - wA)y\|^2 &< \|(I + \bar{w}A)y\|^2 \leftrightarrow y^*(I - wA)^*(I - wA)y \\
&< y^*(I + \bar{w}A)^*(I + \bar{w}A)y \leftrightarrow y^*y - \bar{w}y^*A^*y - wy^*Ay + |w|^2y^*A^*Ay \\
&< y^*y + wy^*A^*y + \bar{w}y^*Ay + |w|^2y^*A^*Ay \leftrightarrow 0 \\
&< \bar{w}y^*(A^* + A)y + wy^*(A + A^*)y \leftrightarrow 2y^*A_sy(w + \bar{w}) > 0 \\
&\leftrightarrow 4\alpha y^*A_sy > 0 \leftrightarrow y^*A_sy > 0
\end{aligned}$$

and therefore A_s is positive definite. Since all this chain of inequalities can be reversed, we also get that if $\|B\| < 1$ then A_s is positive definite.

[2] In this case, B is called just the Cayley transform of A

To prove (b), we observe that if $w = 1$ and all eigenvalues of A_s are 0, then (since $A_s = A_s^*$) $A_s = 0$, and we must have that all eigenvalues of A are on the imaginary axis and $A = A_a$. Once more B is well defined, since $I + A$ is invertible.

Now we only need to verify that $B^*B = I$. We have $B = (I - A)(I + A)^{-1} \Rightarrow B^* = (I + A^*)^{-1}(I - A)^* = (I - A)^{-1}(I + A)$. Moreover, $B = (I - A)(I + A)^{-1} = (I+A)^{-1}(I-A)$ since the two factors commute: $(I-A)(I+A)^{-1} = (I+A)^{-1}(I-A) \leftrightarrow (I + A)(I - A) = (I - A)(I + A)$ which is obvious. Therefore

$$B^*B = (I - A)^{-1}(I + A)(I + A)^{-1}(I - A) = I \ .$$

□

Remark 2.4.4 *With Theorem 2.4.3, we can map matrices with eigenvalues in the RHP into a matrix with eigenvalues in the unit disk. There is an immediate similar construction (with $B = (I + wA)(I - \bar{w}A)^{-1}$) to map matrices with eigenvalues in the LHP inside the unit disk.*

Exercises 2.4.5

(1) If $A = A^*$ is nonnegative definite, where are the eigenvalues of $B = (I - A)(I + A)^{-1}$?

(2) We have seen that: If λ is an eigenvalue of A, and $Av = \lambda v$, $v^*v = 1 \Rightarrow v^*A_sv =$ Reλ. *Find an example of a matrix $A \in \mathbb{R}^{2\times 2}$ whose eigenvalues have* positive real part, *but such that A_s is not positive definite. (*Hint*: can the matrix be normal?)*

• Next, we continue exploring another, deep, analogy between matrix factorizations and complex numbers mappings and representations. The next result is the analog of the representation of a complex number z in polar form: $z = \rho e^{i\phi}$.

Theorem 2.4.6 (Polar Factorization) *Let $A \in \mathbb{C}^{n\times m}$, $n \leq m$. Then A admits the factorization $A = PU$, where $P \in \mathbb{C}^{n\times n} : P^* = P$ and is nonnegative definite, and $U \in \mathbb{C}^{n\times m}$ has orthonormal rows: $UU^* = I_n$. If A is full rank (i.e., n), then P is positive definite. In all cases, P is uniquely determined as $P = (AA^*)^{1/2}$, the unique nonnegative definite square root of AA^*, and U is uniquely determined if A is full rank.*

Pf. First, suppose A is full rank. Then, AA^* is positive definite: $\langle AA^*x, x\rangle = \langle A^*x, A^*x\rangle = \|A^*x\|^2 > 0$, $\forall\, x \neq 0$. Then, $\exists!$ positive definite square root P

of AA^*, $P = P^* : P^2 = (AA^*)$. Define $U = P^{-1}A \Rightarrow U^* = A^*P^{-1}$ and so $UU^* = P^{-1}(AA^*)P^{-1} = P^{-1}P^2P^{-1} = I \therefore U$ has orthonormal rows and $A = PU$.

If A is not full rank $\Rightarrow AA^*$ is only nonnegative definite. Still, it has a unique nonnegative definite square root $P : (AA^*)^{1/2} = P$, $P^* = P \in \mathbb{C}^{n\times n}$. Let rank of P be q $(< n)$. Let $V \in \mathbb{C}^{n\times n}$ unitary such that $V^*PV = \begin{array}{@{}c@{}l@{}} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} & \begin{array}{l} \}q \\ \}n-q \end{array} \end{array}$ where D is diagonal with positive entries, so D^{-1} exists. Let $\hat{U} \in \mathbb{C}^{n\times m}$ be defined as $\hat{U} = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} V^*A$. Observe

$$\hat{U}\hat{U}^* = \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} V^*(AA^*)V \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ 0 & 0 \end{pmatrix}$$

$\therefore$ first q rows of $\hat{U}$ are orthonormal. Now, extend these to an orthonormal $\tilde{U} \in \mathbb{C}^{n\times m}$ (extension is obviously not unique) in such a way that first q rows of $\tilde{U}$ are those of $\hat{U}$. So: $\tilde{U}\tilde{U}^* = I_n$.

Now, take $U = V\tilde{U} \in \mathbb{C}^{n\times m}$ and clearly $UU^* = I_n$. Finally, let us verify that with this U we get $A = PU$. But, to have $A = PU$ is the same as

$$A = \left(V \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} V^*\right) V\tilde{U} \leftrightarrow V^*A = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \tilde{U}\,.$$

The first q rows of this relation are clearly satisfied, since the first q rows of $\tilde{U}$ are those of $\hat{U}$ and $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \hat{U} = V^*A$. For the last $(n-q)$ rows, we have to verify $v_j^*A = 0$, since the last $(n-q)$ rows of $\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \hat{U}$ are clearly 0. Recall that v_j, $j = p+1, \dots, n$, are the eigenvectors of P relative to the eigenvalue 0. Therefore, $v_j^*AA^*v_j = v_j^*P^2v_j = 0 \therefore \|A^*v_j\|^2 = 0 \Rightarrow A^*v_j = 0$, $j = q+1, \dots, n$. □

Remarks 2.4.7

(1) *If $A \in \mathbb{R}^{n\times m}$, $n \le m$, then Theorem 2.4.6 holds in the real field, unchanged. (Of course, now $P = P^T$ and U is real with orthonormal rows: $UU^T = I_n$).*

(2) *If $A \in \mathbb{C}^{n\times n}$ (square case), then $A = PU$, $U^*U = UU^* = I$ that is, U is unitary.*

(3) *Obviously, there is an analogous "polar factorization" for $A \in \mathbb{C}^{m\times n}$, $m \ge n : A = WR$, $W \in \mathbb{C}^{m\times n} : W^*W = I_n$ (orthonormal columns) and $R = R^*$ nonnegative definite. [Just apply Theorem 2.4.6 to A^*.]*

(4) *From above Remarks (2) and (3) we immediately get that a square matrix $A \in \mathbb{C}^{n\times n}$ can be factored as*

$$A = PU = WR, \quad U^*U = UU^* = W^*W = WW^* = I \quad (U \text{ and } W \text{ unitary})$$

and $P^* = P$, $R^* = R$, both nonnegative definite (positive definite if A is full rank). [These are often called right and left polar factorizations.*]*

Exercises 2.4.8 *Below, $\sigma(B)$ indicates the set of eigenvalues of a matrix B repeated according to their multiplicity. E.g., $\sigma\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) = \{1, 1\}$.*

(1) *Let $A \in \mathbb{C}^{n\times n}$. Show that $\sigma(AA^*) = \sigma(A^*A)$.*

(2) *Let $A \in \mathbb{C}^{m\times n}$, $m \geq n$. Show that*

$$\sigma(AA^*) = \sigma(A^*A) \cup \underbrace{\{0, \dots, 0\}}_{m-n} .$$

(3) *Let $A \in \mathbb{C}^{n\times n}$ and let $A = PU$, $A = WR$ be its right and left polar factorizations. Show:*

a) If A is nonsingular, then $U = W$;

b) If $P = R \Leftrightarrow A$ is normal.

[Hint*: do this problem after the SVD below is introduced.*]

(4) *Show that if $A \in \mathbb{C}^{n\times n}$ is normal, then its polar factors P and U commute.* [Hint*: How many nonnegative definite square roots does a nonnegative definite matrix have?*]

As a consequence of Exercise 2.4.8-(4), observe that if $PU = UP \Rightarrow AA^* = A^*A$. In fact, $A = PU \Rightarrow AA^* = P^2$ and $A^*A = U^*PPU = U^*PUP = U^*UP^2 = P^2$. That is, we have found one more characterization of normality:

"*$A \in \mathbb{C}^{n\times n}$ is normal if and only if $PU = UP$, where $A = PU$ is a polar factorization of A.*"

2.4.2 The SVD: Singular Value Decomposition

An immediate consequence of the polar factorization is one of the most useful matrix decompositions.

Theorem 2.4.9 (SVD) *Given $A \in \mathbb{C}^{m\times n}$ and assume $m \geq n$. Then A may be written as $A = U\Sigma V^*$, where $U \in \mathbb{C}^{m\times m}$ unitary, $V \in \mathbb{C}^{n\times n}$ unitary and $\Sigma \in \mathbb{R}^{m\times n}$ is "nonnegative diagonal," of the form*

$$\Sigma = \underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & \sigma_{k+1} & & \\ & & & & \ddots & \\ & & & & & \sigma_n \\ 0 & \cdots & & 0 & \cdots & 0 \end{bmatrix}}_{n} \}m-n,$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0$.

Pf. Consider the polar form $A = WR$, where $W \in \mathbb{C}^{m\times n}$: $W^*W = I_n$, and $R = R^* = (A^*A)^{1/2}$ is nonnegative of rank k. Let V be a unitary matrix: $VRV^* = \Lambda = \mathrm{diag}(\lambda_i, i = 1, \ldots, n)$, where $\lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} = \cdots = \lambda_n = 0$. Therefore, $A = (WV^*)\Lambda V = \hat{U}\Lambda V$ where $\hat{U} \in \mathbb{C}^{m\times n}$ has orthonormal rows: $\hat{U}^*\hat{U} = I_n$.

Now, complete $\hat{U}$ to an orthonormal basis for $\mathbb{C}^m$: $U \in \mathbb{C}^{m\times m}$ so that the first n columns of U are those of $\hat{U}$, and complete Λ to Σ: $\Sigma = \begin{bmatrix}\Lambda \\ 0\end{bmatrix}$. □

Remarks 2.4.10

(1) *The values $\sigma_1, \ldots, \sigma_n$ are called* singular values *of A. Clearly,* $\mathrm{rank}(A) = k$*, the number of nonzero singular values.*

Notice that since $\begin{cases} AA^* = U\Sigma\Sigma^*U^* \\ A^*A = V\Sigma^*\Sigma V^* \end{cases}$*, then the σ_i's are the positive square roots of the eigenvalues of A^*A. The columns of the unitary factor U are called* left singular vectors *of A (they are eigenvectors of the matrix AA^*). The columns of the unitary factor V are called* right singular vectors *of A (they are eigenvectors of A^*A).*

(2) *There is an immediate analogous result in real case: $A = U\Sigma V^T$, with $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ orthogonal.*

(3) *The ordering of the singular values is the usual way the result is given. In principle, we could order them differently: It is the set $\{\sigma_1, \ldots, \sigma_n\}$ which is*

uniquely determined. [At times, also the restriction of $\sigma_i \geq 0$ *is lifted; in this case, one obtains a so-called* signed SVD.]

(4) *Of course, there is no need to assume* $m \geq n$. *If we have* $m < n$, *then we can just apply the SVD to* A^*.

(5) *Notice that the SVD is a two-sided diagonalization by unitary matrices obtained by effetively piecing together the unitary facgtors of the two nonnegative definite matrices* AA^* *and* A^*A.

(6) *From the above proof, obviously the unitary factors* U *and* V *are not unique. In case* $k = n$, *and distinct singular values, then* V *is unique up to a diagonal phase matrix (see Homework 2):* $V\Phi$, $\Phi = \mathrm{diag}(e^{i\phi_j}, j = 1, \dots, n;\ \phi_j \in \mathbb{R})$. *If* $m > n$, *of course* U *is not unique. But if* $m = n = k$, *and* V *is fixed, then* U *is unique:* $U = AV\Sigma^{-1}$.

(7) *In the same notation of Theorem 2.4.9, one can also rewrite the result relatively to the so-called* reduced SVD, *which is obtained neglecting the last (m-n) rows of* Σ*:*

$$A = U\Sigma V^* \ , \ U \in \mathbb{C}^{m\times n} \ : U^*U = I_n \ , \ \ \Sigma \in \mathbb{R}^{n\times n} \ , \ \ V \in \mathbb{C}^{n\times n} \ : V^*V = I \, .$$

(8) *If* $A = A^*$, *then* $\sigma_i = |\lambda_i|$ *(the absolute values of the eigenvalues of* A*). Indeed, from the Schur theorem we have, for a unitary* U*:* $A = U\Lambda U^* = U|\Lambda|SU^* = U\Sigma V^*$, *where* $V = US$, *and* $S = \mathrm{diag}(\pm 1)$, *where we take the sign* ± 1 *corresponding to the positive/negative eigenvalues (and either one of* ± 1 *relatively to the* 0 *eigenvalues).*

• An interesting consequence of the SVD is that near any given matrix there is one with distinct singular values. Although we state and prove the theorem below for the 2-norm, the result holds in any norm, since all norms are equivalent.

Theorem 2.4.11 *Given* $A \in \mathbb{C}^{m\times n}$, $m \geq n$, *whose singular values are not all distinct. Then,* $\forall \varepsilon > 0$, *there exists* $A_\varepsilon \in \mathbb{C}^{m\times n}$ *such that all singular values of* A_ε *are distinct and* $\|A - A_\varepsilon\|_2 < \varepsilon$.

Pf. From $A = U\Sigma V^*$, let k be the rank of A, and write $\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}$. So, we have that $S = \mathrm{diag}(\sigma_i)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_n = 0$. Let us first assume that not all singular values are equal. Let $d = \min_{j:\ \sigma_j > \sigma_{j+1}} (\sigma_j - \sigma_{j+1})$ and let η: $0 < \eta \leq d$. Consider the matrix $\Sigma_\eta = \begin{bmatrix} S_\eta \\ 0 \end{bmatrix}$, where $S_\eta = \mathrm{diag}(\sigma_k + \eta/k,\ k = 1, 2, \dots, n)$. Form $A_\eta = U\Sigma_\eta V^*$, so that $\|A - A_\eta\| = \|\Sigma - \Sigma_\eta\| = \eta$; now take $\eta < \varepsilon$. If all singular values are equal, then just take $\eta = \varepsilon$ in the definition of Σ_η. $\square$

• There is also an interesting, and very useful, characterization of the SVD from an enlarged Hermitian problem. We show it in the next exercise.

Exercise 2.4.12 *Let $A \in \mathbb{C}^{m\times n}$, $m \geq n$. Define $M = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix} \in \mathbb{C}^{n+m,n+m}$. Clearly $M = M^*$ and* $\operatorname{rank}(M) \leq 2n$. *Now, if $A = U\Sigma V^*$ is an SVD of A, and we further partition U columnwise $U = [U_1\ U_2]$, where $U_1 \in \mathbb{C}^{m\times n}$, and $\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}$, then $M = \begin{bmatrix} 0 & U_1SV^* \\ VSU_1^* & 0 \end{bmatrix}$. Now, observe that this can be also rewritten as*

$$M = \begin{bmatrix} U_1/\sqrt{2} & -U_1/\sqrt{2} & U_2 \\ V/\sqrt{2} & V/\sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} S & 0 & 0 \\ 0 & -S & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^*/\sqrt{2} & V^*/\sqrt{2} \\ -U_1^*/\sqrt{2} & V^*/\sqrt{2} \\ U_2^* & 0 \end{bmatrix}$$

and that the matrix

$$Z = \begin{bmatrix} U_1/\sqrt{2} & -U_1/\sqrt{2} & U_2 \\ V/\sqrt{2} & V/\sqrt{2} & 0 \end{bmatrix}$$

is unitary, $ZZ^* = I_{n+m}$. In other words, the SVD of A gives a Schur form for M with all positive eigenvalues $\lambda_1, \ldots, \lambda_k$ first (ordered decreasingly), followed by the negative eigenvalues (ordered increasingly), and then the 0 eigenvalues; further, $\lambda_j = \sigma_j$, $j = 1, \ldots, k$. Likewise, from a Schur form of M (with eigenvalues ordered as above), we immediately get an SVD of A upon partitioning the unitary factor Z of the Schur form of M as above. □

The equivalence of Exercise 2.4.12 suggests that results that we can prove for Hermitian matrices should have a counterpart in results we can prove for the SVD. This is essentially correct; however, since the SVD applied to rectangular matrices, we need to carry out the details carefully –even when we have distinct singular values– and need to take care of the part relative to the zero eigenvalues of AA^*.

A nice byproduct of the SVD is that it reveals a lot of structure about norms as well as fundamental subspaces associated to A. We see some of these results below, in the form of exercises.

Exercise 2.4.13 *Use the SVD of a matrix $A \in \mathbb{C}^{m\times n}$ to find orthonormal bases for $\mathcal{R}(A)$, $\mathcal{N}(A)$, $\mathcal{R}(A^T)$, $\mathcal{N}(A^T)$ and to express orthogonal projection matrices onto these subspaces.*

Exercise 2.4.14 (2-norm) *Show that $\|A\|_2 = \sigma_1$ (largest singular value of A).*

Solution. The verification is immediate:

$$\|A\|_2^2 = \|U\Sigma V^*\|_2^2 = \lambda_{\max}(V\Sigma^T U^* U\Sigma V^*) = \lambda_{\max}(S^2) = \sigma_1^2 \ .$$

[An alternative proof of the result rests on the fact that the 2-norm is invariant with respect to unitary transformations on the right and left.] □

Exercise 2.4.15 (Frobenius norm) *Show that* $\|A\|_F = \left(\sum_{i=1}^{\mathrm{rank}(A)} \sigma_i^2\right)^{1/2}$.

Solution. Again a straightforward computation:

$$\|A\|_F^2 = \mathrm{tr}(A^*A) = \mathrm{tr}(V\Sigma^T U^* U\Sigma V^*) = \mathrm{tr}(\Sigma^T\Sigma) \ ,$$

where we have used that the trace is similarity invariant. □

A consequence of the last two exercises is that

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{\min(n,m)}\|A\|_2 \ .$$

The SVD also allows for writing the solution of the least squares problem in a very compact way, as we do in the example below.

Example 2.4.16 *Let us revisit the full rank least-squares problem: "Find $x \in \mathbb{R}^n$ such that $\|Ax-b\|_2$ is minimized, when $A \in \mathbb{R}^{m\times n}$, $m \geq n$, A full rank, and $b \in \mathbb{R}^m$." Of course, we know that $x = (A^TA)^{-1}A^Tb$. Let us rewrite this in terms of the SVD of A: $A = U\Sigma V^T$. Then, since $\Sigma = \begin{bmatrix} S \\ 0 \end{bmatrix}$ with S invertible, we can also write:*

$$x = V\big(S^{-1}, \ 0)U^Tb = V\Sigma^+U^Tb \,,$$

where we have set $\Sigma^+ = \big(S^{-1}, \ 0\big)$. Observe that $\Sigma^+\Sigma = I_n$. We call Σ^+ a left-inverse *for Σ. Moreover, observe that for this Σ^+ we also have the following four properties:*

$$\begin{cases} \Sigma^+\Sigma\Sigma^+ = \Sigma^+, \\ \Sigma\Sigma^+\Sigma = \Sigma, \\ \Sigma\Sigma^+ \quad \textit{and} \quad \Sigma^+\Sigma \quad \textit{are Hermitian.} \end{cases} \tag{2.4.1}$$

We also observe that if we had $m < n$, hence $\Sigma = (S, \ 0)$, then defining $\Sigma^+ = \begin{bmatrix} S^{-1} \\ 0 \end{bmatrix}$ such Σ^+ would satisfy properties (2.4.1); now Σ^+ would be a right-inverse *for Σ: $\Sigma\Sigma^+ = I_n$.* □

Let us generalize this last example. Consider $A = U\Sigma V^*$ (regardless of $m \lesseqgtr n$).

Definition 2.4.17 (Pseudo-inverse) *Let $A = U\Sigma V^*$ be an SVD of $A \in \mathbb{C}^{m\times n}$. Let $R \in \mathbb{R}^{m\times n}$ be the matrix obtained by replacing the positive singular values in Σ by their reciprocal values. Finally, let $\Sigma^+ = R^T$. Then, Σ^+ is called* pseudo-inverse *of Σ.*

Further, define A^+ as $A^+ = V\Sigma^+U^*$. Then, A^+ is called pseudo-inverse *of A, or also* Moore-Penrose generalized inverse*.*

Exercises 2.4.18

(1) *Verify that Σ^+ just defined satisfies properties (2.4.1).*

(2) *Verify that A^+ satisfies properties (2.4.1) (with A and A^+ replacing Σ, Σ^+ there).*

(3) *Verify that if A is full rank, and $m \geq n$, then $A^+ = (A^TA)^{-1}A^T$. What is the formula for $m < n$? What for $m = n$?*

• The pseudo-inverse of A allows to write the solution of the best-approximation problem in the rank deficient case in a very compact way.

Exercise 2.4.19 *Show that regardless of whether or not $A \in \mathbb{R}^{m\times n}$ is full rank, the minimum-norm solution of the least squares problem, $\min_x \|Ax - b\|_2$, is $x = A^+b$.*

- Arguably, the SVD is the most practically useful tool to solve a number of approximation problems. We are going to review some of them here below, in the form of Examples.

Example 2.4.20 (Nearness to singularity) *Given $A \in \mathbb{C}^{n\times n}$, invertible, we want to find a matrix B of minimal 2-norm such that $A + B$ is singular. [We could have also used the F-norm.]*

Solution. Observe that since $A+B = A(I+A^{-1}B) \Rightarrow$ if $\|A^{-1}B\| < 1 \Rightarrow I+A^{-1}B$ would be invertible[3], and so would be $A + B \therefore 1 \leq \|A^{-1}B\| \leq \|A^{-1}\|\|B\| \therefore B$ must satisfy $\|B\| \geq \frac{1}{\|A^{-1}\|}$. But $\frac{1}{\|A^{-1}\|} = \frac{1}{\|(U\Sigma V^T)^{-1}\|} = \frac{1}{\|\Sigma^{-1}\|} = \frac{1}{1/\sigma_n} = \sigma_n \therefore \|B\| \geq \sigma_n$.

[3]See Homework 2

Now, take $B = U \begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & -\sigma_n \end{pmatrix} V^* = -\sigma_n u_n v_n^*$. Clearly B is of minimal norm and $A+B$ is singular. $\square$

Remark 2.4.21 *Observe that B in Example 2.4.20 is surely not unique if $\sigma_{n-1} = \sigma_n$. The minimal value of $\|B\|$ is uniquely determined.*

Exercise 2.4.24 below is very useful, somewhat easy to accept, but its verification is lengthy and nontrivial. It will be broken down in several steps. The following Lemma will be useful and it is of independent interest.

Lemma 2.4.22 (Eigenvalues of product) *Let $A, B \in \mathbb{R}^{m\times n}$, $m \geq n$. Consider $A^TB \in \mathbb{R}^{n\times n}$ and $BA^T \in \mathbb{R}^{m\times m}$. Then:*

$$\sigma(BA^T) = \sigma(A^TB) \cup \underbrace{\{0, \dots, 0\}}_{m-n} .$$

That is, the eigenvalues of BA^T are the same as those of A^TB –counting multiplicities– plus an additional $(m-n)$ zero eigenvalues.

Pf. One way to show it is the following. Consider the matrices in $\mathbb{R}^{m+n,m+n}$: $M = \begin{bmatrix} BA^T & 0 \\ A^T & 0 \end{bmatrix}$ and $N = \begin{bmatrix} 0 & 0 \\ A^T & A^TB \end{bmatrix}$. Observe that

$$M \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} = \begin{bmatrix} BA^T & BA^TB \\ A^T & A^TB \end{bmatrix}$$

$$\text{and} \quad \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} N = \begin{bmatrix} BA^T & BA^TB \\ A^T & A^TB \end{bmatrix} .$$

So,

$$M = \begin{bmatrix} BA^T & 0 \\ A^T & 0 \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & 0 \\ A^T & A^TB \end{bmatrix} \begin{bmatrix} I & -B \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & B \\ 0 & I \end{bmatrix} N \begin{bmatrix} I & B \\ 0 & I \end{bmatrix}^{-1}$$

and so M and N are similar and $\sigma(M) = \sigma(N)$. Given the block triangular structure of M and N, we thus have

$$\sigma(M) = \sigma(BA^T) \cup \underbrace{\{0, \dots, 0\}}_{n} , \quad \sigma(N) = \sigma(A^TB) \cup \underbrace{\{0, \dots, 0\}}_{m} ,$$

and the result follows. $\square$

Corollary 2.4.23 *With notation as in Lemma 2.4.22, we have* $\mathrm{tr}(A^TB) = \mathrm{tr}(BA^T)$.

Pf. This is because the trace of a matrix is the sum of its eigenvalues. $\square$

Example 2.4.24 (Nearest rank k to A) *Let $A \in \mathbb{R}^{m\times n}$, $m \geq n$, and* $\mathrm{rank}(A) \geq k$. *We want A_k of rank k : $\|A - A_k\|_F$ is minimized. [Here we are using the F-norm, though the result holds also for the 2-norm, in fact for any orthogonally invariant norm. Also, the same result –and much the same proof– holds for matrices having complex valued entries.]*

Solution. Let $A = U\Sigma V^T$ be a SVD of A. Then, take $\Sigma_k = \begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & 0 & & \end{bmatrix}$

and let $A_k = U\Sigma_k V^T$. Obviously A_k has rank k and $\|A - A_k\|_F^2 = \sum_{j=k+1}^{\mathrm{rank}(A)} \sigma_j^2$. It remains to show that for any matrix $B \in \mathbb{R}^{m\times n}$ of rank k (and therefore for which we have $\sigma_{k+1}(B) = \cdots = \sigma_n(B) = 0$) we have

$$\begin{aligned} &\|A - B\|_F \geq \|\Sigma(A) - \Sigma(B)\|_F \\ &= \|\mathrm{diag}(\sigma_1(A) - \sigma_1(B), \ldots, \sigma_k(A) - \sigma_k(B), \sigma_{k+1}(A), \ldots, \sigma_m(A))\|_F \end{aligned} \tag{2.4.2}$$

and the result will follow. This last fact is actually rather involved, and we show it in several steps.

1st step. $0 \leq \|A - B\|_F^2 = \mathrm{tr}((A-B)^T(A-B)) = \mathrm{tr}(A^TA - A^TB - B^TA + B^TB) = \mathrm{tr}(A^TA) + \mathrm{tr}(B^TB) - \mathrm{tr}(A^TB + B^TA)$. Now, $\mathrm{tr}(A^TB) = \mathrm{tr}(B^TA)$ since $B^TA = (A^TB)^T$ so they have same diagonal.

$\therefore \min\limits_{\substack{B\in\mathbb{R}^{m+n} \\ rank(B)=k}} \|A-B\|_F^2 = \mathrm{tr}(A^TA) + \mathrm{tr}(B^TB) - 2\mathrm{tr}(A^TB) = \sum_{j=1}^n \sigma_j^2(A) + \sum_{j=1}^{n(\text{or } k)} \sigma_j^2(B) - 2\mathrm{tr}(A^TB)$.

$\therefore$ Now, we need to find $B \in \mathbb{R}^{m\times n}$, of rank k, to maximize the above expression. Note: $A^TB \in \mathbb{R}^{n\times n}$. Next, we solve this problem by keeping A fixed and letting B be a matrix with given (but arbitrary) singular values $\sigma_1(B) \geq \cdots \geq \sigma_k(B) > 0$, all other singular values being 0. In other words, we are letting B to be of the form $B = U\Sigma_B V^T$, with U and V orthogonal of appropriate dimension, and we are going to maximize (over the set of orthogonal matrices) the expression $\mathrm{tr}(A^T U\Sigma_B V^T)$. If we can show that –for any given k-tuple of singular values of BD– we have $\mathrm{tr}(A^T B) \le \sum_{i=1}^n \sigma_i(A)\sigma_i(B)$, then the result will follow.

2nd step. Now, maximizing $\hat{U}$ and $\hat{V}$ for $\mathrm{tr}(A^T U\Sigma_B V^T)$ exist, since the orthogonal matrices form compact sets, though they are possibly not unique. In any case, for any orthogonal $V \in \mathbb{R}^{n\times n}$, we must have

$$\mathrm{tr}\big((A^T\hat{U}\Sigma_B)\hat{V}\big) \ge \mathrm{tr}\big((A^T\hat{U}\Sigma_B)V\big)\ .$$

Take the SVD of $A^T\hat{U}\Sigma_B$: $A^T\hat{U}\Sigma_B = Q\Sigma W^T$. Then,

$$\mathrm{tr}\big((A^T\hat{U}\Sigma_B)V\big) = \mathrm{tr}(Q\Sigma W^T V) = \mathrm{tr}(\Sigma Z)\ ,$$

where $Z = W^T V Q$ is orthogonal. So:

$$\mathrm{tr}\big((A^T\hat{U}\Sigma_B)V\big) \le \sum_{i=1}^n \sigma_i z_{ii} \le \sum_{i=1}^n \sigma_i\ ,$$

and so $Z = I$ surely maximizes. This means that $\hat{V} = WQ^T$, and so $(A^T\hat{U}\Sigma_B)\hat{V} = Q\Sigma Q^T$, and thus we can restrict our search for B such that $B^T A$ is symmetric and nonnegative definite.

3rd step. In light of Lemma 2.4.22 and Corollary 2.4.23, we can thus restrict our search to $B \in \mathbb{R}^{m\times n}$ such that we maximize $\mathrm{tr}(BA^T)$. But, as in Step 2, this implies that we can restrict our search to the case of BA^T being symmetric and nonnegative definite.

4th step. So, we are restricting our search to the case when both $A^T B$ and BA^T are nonnegative definite. In this case, we have:

Claim. $\mathrm{tr}(A^T B) = \mathrm{tr}(BA^T) = \sum_{i=1}^n \sigma_i(A)\sigma_{\pi(i)}(B)$, where $\pi(i)$ is a permutation of $\{1,\dots,n\}$ and the singular values of A and B are ordered.

Postponing (see Lemma 2.4.25 below) the proof of this claim, we can now conclude our argument.

5th step. The observation is that the identity is a maximizing permutation in the formula for $\mathrm{tr}(A^T B)$. In fact, suppose π is not the identity. Then, there are indices, $i_1, i_2 : 1 \le i_1 < i_2 \le n$ for which $\sigma_{\pi(i_1)}(B) \le \sigma_{\pi(i_2)}(B)$.

But then consider our sum, $S = \sum_{i=1}^n \sigma_i(A)\sigma_{\pi(i)}(B)$, and exchange the positions of these two singular values obtaining $\hat{S} = \sum_{i \ne i_1, i_2} \sigma_i(A)\sigma_{\pi(i)}(B) + \sigma_{i_1}(A)\sigma_{\pi(i_2)}(B) + \sigma_{i_2}(A)\sigma_{\pi(i_1)}(B)$. Now $\hat{S} - S = \sigma_{i_1}(A)\sigma_{\pi(i_2)}(B) - \sigma_{i_1}(A)\sigma_{\pi(i_2)}(B) + \sigma_{i_2}(A)\sigma_{\pi(i_1)}(B) - \sigma_{i_2}(A)\sigma_{\pi(i_2)}(B) =$
$\underbrace{(\sigma_{i_1}(A) - \sigma_{i_2}(A))}_{\ge 0}\underbrace{(\sigma_{\pi(i_2)}(B) - \sigma_{\pi(i_1)}(B))}_{\ge 0} \ge 0$, and so since the singular values of B (and of A) are ordered, the identity maximizes the sum.

$\therefore$ $\max(\mathrm{tr}(A^T B)) = \sum_{i=1}^n \sigma_i(A)\sigma_i(B)$.

$\therefore$ $\min_{\mathrm{rank}(B)=k} \|A - B\|_F^2 = \sum_{j=1}^n \sigma_j^2(A) + \sum_{j=1}^n \sigma_j^2(B) - 2\sum_{j=1}^n \sigma_j(A)\sigma_j(B) = \sum_{j=1}^n (\sigma_j(A) - \sigma_j(B))^2$ which is what we wanted in (2.4.2). $\square$

- Next, we are left to verify the Claim. We single this out as

Lemma 2.4.25 *Let $A, B \in \mathbb{R}^{m \times n}$, $m \ge n$, be such that $A^T B$ and BA^T are symmetric and nonnegative definite. Then:*

$$tr(A^T B) = \mathrm{tr}(BA^T) = \sum_{i=1}^n \sigma_i(A)\sigma_{\pi(i)}(B) \ ,$$

where π is a permutation of the indices $\{1, \dots, n\}$.

Pf. First, we observe that the proof is simpler if A and B are both symmetric and nonnegative definite (and so $m = n$). In this case, since $AB = A^T B = (AB)^T = BA^T = BA$, they would commute and thus can be simultaneously diagonalized by an orthogonal matrix U: $U^T A^T A = \Lambda_A$ and $U^T BU = \Lambda_B$. Therefore, $\mathrm{tr}(A^T B) = \mathrm{tr}(U^T A^T U U^T BU) = \mathrm{tr}(\Lambda_A \Lambda_B) = \sum_{i=1}^n \lambda_i(A)\lambda_i(B)$. In this case, the set of singular values of A, B concide with the set of their eigenvalues, of course. Note that we do not know that the eigenvalues of A and B are ordered by U, but nevertheless can surely conclude that $\sum_{i=1}^n \lambda_i(A)\lambda_i(B) = \sum_{i=1}^n \sigma_i(A)\sigma_{\pi(i)}(B)$, as desired.

So, the idea is to reduce ourselves to this case of A and B being symmetric.

We search for $W \in \mathbb{R}^{n \times m}$ and $V \in \mathbb{R}^{m \times m}$, $V^T V = I$ (V orthogonal) and $WW^T = I_n$ (W has orthonormal rows) such that defining

$$\hat{A}^T = W^T A^T V \in \mathbb{R}^{m \times m}, \quad \hat{B} = V^T BW \in \mathbb{R}^{m \times m} \tag{2.4.3}$$

then we have

a) $\hat{A}^T\hat{B} = \hat{B}\hat{A}^T$ and
b) $\hat{A}, \hat{B}$ both symmetric nonnegative definite.
If we can do this, then

$$\begin{aligned}\mathrm{tr}(A^TB) &= \mathrm{tr}((A^TBW)W^T) = \mathrm{tr}(W^T(A^TBW)) = \mathrm{tr}(W^TA^TVV^TBW) \\ &= \mathrm{tr}(\hat{A}^T\hat{B}) = \sum_{i=1}^{n}\lambda_i(\hat{A})\lambda_i(\hat{B}) = \sum_{i=1}^{n}\sigma_i(A)\sigma_{\pi(i)}(B),\end{aligned}$$

where the last equality is a consequence of the fact that $\hat{A}\hat{A}^T = V^TAWW^TA^TV = V^TAA^TV$ and $\hat{B}\hat{B}^T = V^TBWW^TB^TV = V^TBB^TV$ so that the set of singular values of A, respectively of B, coincide with the set of eigenvalues of $\hat{A}$, respectively of $\hat{B}$, and we would be done. So, let us build W and V as in (2.4.3). We do this in three steps.

(i) Since A^TB and BA^T are nonnegative definite, let orthogonal U and Z be such that $A^TB = U\Lambda U^T$ and $BA^T = Z\begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix}Z^T = \begin{bmatrix}Z_1 & Z_2\end{bmatrix}\begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}Z_1^T\\ Z_2^T\end{bmatrix} = Z_1\Lambda Z_1^T$. Let $Y = UZ_1^T$ ($\Rightarrow Y \in \mathbb{R}^{m\times m} : YY^T = I_n$). Note $Y^TA^TBY = Y^TU\Lambda U^TY = Z_1\Lambda Z_1^T = BA^T$ and $BYY^TA^T = BA^T$ $\therefore$ Y^TA^T and BY are in $\mathbb{R}^{m\times m}$ and commute and $(W^TA^T)(BW)$ and $(BW)(W^TA^T)$ are nonnegative definite.

(ii) So, without loss of generality we can restrict to the case of $A, B \in \mathbb{R}^{m\times m}$: $A^TB = BA^T$ both nonnegative definite, though the single factors A and B are not necessarily nonnegative definite. Let $U : U^T(A^TB)U = \Lambda = \mathrm{diag}(\lambda_kI_k, k = 1 : p)$, where I_k is of size m_k and $m_1 + \cdots + m_p = m$. Notice: $\begin{cases}U^T(A^TB)BU = U^TB(A^TB)U = (U^TBU)\Lambda\\ U^T(A^TB)A^TU = U^TA^T(A^TB)U = (U^TA^TU)\Lambda\end{cases}$
but also $\begin{cases}U^T(A^TB)BU = \Lambda(U^TBU)\\ U^TA^TBA^TU = \Lambda(U^TA^TU)\end{cases}$ and so $\begin{cases}(U^TBU)\Lambda = \Lambda(U^TBU)\\ (U^TA^TU)\Lambda = \Lambda(U^TA^TU)\end{cases}$
and easily we must then have $U^TA^TU = \mathrm{diag}(A_{jj}^T, j = 1 : p)$, and $U^TBU = \mathrm{diag}(B_{jj}, j = 1 : p)$.

$\therefore$ We can directly consider $A := U^TAU$ and $B := U^TBU$ both block-diagonal and thus $A_{jj}^TB_{jj} = B_{jj}A_{jj}^T = \lambda_jI_j$, $j = 1, \dots, p$.

(iii) So, without loss of generality, we have $A, B : A^TB = BA^T = \lambda I$, $\lambda \geq 0$. We only need to show that we can modify A, B and take $\hat{A} = \hat{A}^T$ and $\hat{B} = \hat{B}^T$

both nonnegative definite. Now, take the polar form of A^T : $A^T = PU$ and let $\hat{A}^T = A^T U^T = P$ which is symmetric nonnegative definite. Define $\hat{B} = UB \Rightarrow \hat{A}^T \hat{B} = A^T B = \lambda I = BA^T = UBA^T U^T = \hat{B}\hat{A} \Rightarrow P\hat{B} = \lambda I$. We have two more cases to consider.

(iii-a) $\lambda > 0 \Rightarrow A$ nonsingular and P is positive definite $\Rightarrow \hat{B} = \lambda P^{-1}$ is also positive definite and we are done.

(iii-b) $\lambda = 0 \Rightarrow A^T B = BA^T = 0$. Still, take $A^T = PU$ (polar) and $\hat{A}^T = A^T U^T = P$ which is nonnegative definite. Let $\bar{B} = UB$, then clearly $\hat{A}^T \bar{B} = A^T B = BA^T = 0 = UBA^T U^T = \bar{B}\hat{A}^T$. So $\hat{A}^T$ and $\bar{B}$ commute and $\hat{A}^T = P$ is (symmetric) nonnegative definite. Let W orthogonal be such that $W^T P W = D = \mathrm{diag}(d_j I_j, j = 1 : r)$ and $d_j \geq 0$. Then, from $0 = P\bar{B} = \bar{B}P \to D(W^T \bar{B} W) = (W^T \bar{B} W)D \Rightarrow W^T \bar{B} W = \mathrm{diag}(\bar{B}_{jj}, i = 1 : r) \therefore d_j \bar{B}_{jj} = 0$. If $d_j \neq 0 \Rightarrow \bar{B}_{jj} = 0$. If $d_j = 0 \Rightarrow \bar{B}_{jj}$ not necessarily 0; but, $\bar{B}_{jj} = V_{jj} P_{jj}$ (polar factorization) so, we take $\hat{B}_{jj} = V_{jj}^T \bar{B}_{jj}$ and the pair we take is $d_j I \cdot V_{jj}$ and $V_{jj}^T \bar{B}_{jj}$. $\square$

Exercises 2.4.26

(1) We have seen that given $A, B \in \mathbb{R}^{m\times n}$, $m \geq n$, for the set of eigenvalues we have $\sigma(BA^T) = \sigma(A^T B) \cup \{0, \dots, 0\}$. Show that such a result is not true for singular values.

(2) Let $A \in \mathbb{C}^{n\times n}$. Show that $\|A - U\|_F \geq \|\Sigma(A) - I\|_F$ for any unitary U and that the inequality is sharp. [*That is, for any given A, there is a unitary U such that the equality sign holds in the above bound.*] *From this, you have that the distance (in F-norm) from A to the compact set of unitary matrices is $\|\Sigma(A) - I\|_F$.*

(3) [Wielandt-Hoffman] *Show that if A, B are Hermitian in $\mathbb{C}^{n\times n}$, then*

$$\|A - B\|_F^2 \geq \sum_{i=1}^{n} (\lambda_i(A) - \lambda_i(B))^2 \,, \tag{2.4.4}$$

where the eigenvalues of A and B are ordered decreasingly.

In the next example, we look explicitly at an interesting problem which has applications in statistics of data set: We try to "rotate" a matrix B (a data set) as close as possible to another matrix A (a different data set). Note that in case $m = n$ and $B = I$, then we are seeking the closest unitary matrix to A. (See above Exercise (2).)

Example 2.4.27 (Procrustes Problem) [4] *This is a recurring problem where we try to modify the data to match another set of data. The problem is the following. "Given $A, B \in \mathbb{C}^{m\times n}$, we want to find $U \in \mathbb{C}^{m\times m}$ unitary such that $\|A - UB\|_F$ is minimized. Also, we want to find the expression of the error."*

Solution.

$$\begin{aligned}\|A - UB\|_F^2 &= \operatorname{tr}((A-UB)^*(A-UB)) = \operatorname{tr}(A^* - B^*U^*)(A-UB)\\ &= \operatorname{tr}(A^*A - B^*U^*A - A^*UB + B^*B)\\ &= \operatorname{tr}(A^*A) + \operatorname{tr}(B^*B) - \operatorname{tr}(A^*UB) - \operatorname{tr}(B^*U^*A)\ .\end{aligned}$$

Now, we have

$$\begin{aligned}\operatorname{tr}(A^*UB) + \operatorname{tr}(B^*U^*A) &= \operatorname{tr}(A^*UB) + \operatorname{tr}((A^*UB)^*)\\ &= 2\mathrm{Re}\ \operatorname{tr}(A^*UB)\end{aligned}$$

and therefore we are seeking U unitary to maximize $\mathrm{Re}\ \operatorname{tr}(A^*UB) = \mathrm{Re}\ \operatorname{tr}(UBA^*)$.

Let $BA^* = V\Sigma W^*$ be a SVD of BA^*. Then

$$\begin{aligned}\mathrm{Re}\ \operatorname{tr}(UBA^*) &= \mathrm{Re}\ \operatorname{tr}(UV\Sigma W^*)\\ &= \mathrm{Re}\ \operatorname{tr}(WW^*UV\Sigma W^*) = \mathrm{Re}\ \operatorname{tr}(W^*UV\Sigma) = \mathrm{Re}\ \operatorname{tr}(Z\Sigma)\end{aligned}$$

where $Z = W^*UV$ is unitary, $Z^*Z = ZZ^* = I_m$. So, we need to maximize $\mathrm{Re}\sum_{i=1}^m \sigma_i(BA^*)Z_{ii}$ which is obviously maximized if $Z_{ii} = 1 \Rightarrow Z = I \Rightarrow W^*UV = I \Rightarrow U = WV^*$. Therefore, the best solution is $U = WV^*$. Note, since $BA^* = V\Sigma W^* = V\Sigma V^*VW^* \Rightarrow AB^* = (WV^*)(V\Sigma V^*) = QP \Rightarrow$ the optimal U is the left unitary polar factor of AB^*.

The error is $\|A - UB\|_F^2 = \|A\|_F^2 + \|B\|_F^2 - 2\sum_{i=1}^m \sigma_i(BA^*)$. In particular, there is no error only if $\underbrace{\|A\|_F = \|B\|_F}_{\text{surely necessary}} = \sum_{i=1}^m \sigma_i(BA^*)$. In the special case of $B = I \Rightarrow \|A - U\|_F^2 = \sum_{i=1}^n \sigma_i^2(A) + n - 2\sum_{i=1}^n \sigma_i(A) = \sum_{i=1}^n(\sigma_i(A) - 1)^2$. □

2.5 Partial order and inequalities

Hermitian matrices allow for a natural partial order.

[4] A mythological Greek figure, who stretched/cut the visitors who could not "fit" into a bed he provided for their rest.

Definition 2.5.1 *$A, B \in \mathbb{C}^{n\times n}$, or $\mathbb{R}^{n\times n}$, be Hermitian, respectively symmetric. We say that A is greater than B and write $A \succ B$ if $A - B$ is positive definite. Similarly, we write $A \succeq B$ if $A - B$ is nonnegative definite. Similarly for $A \prec B$, $A \preceq B$.* □

Exercises 2.5.2

(1) Show that this is only a partial order (that is, there exist A and B Hermitian such that $A \not\succeq B$ nor $A \not\preceq B$).
(2) Show that if $A_1 \succeq B_1$ and $A_2 \succeq B_2 \Rightarrow A_1 + A_2 \succeq B_1 + B_2$.
(3) Show that $A \succeq I \Leftrightarrow$ all A's eigenvalues are ≥ 1.
(4) Show that partial order is invariant under congruence. That is, if S is invertible, $A \succeq B \Leftrightarrow S^*AS \succeq S^*BS$.
(5) Give an argument of why we may as well restrict partial order to nonnegative matrices. [Of course, this is in principle only, in practice a given matrix may be Hermitian without being definite.]

To investigate properties of the above partial order, the following results are handy.

Lemma 2.5.3 *If $A, B \in \mathbb{C}^{n\times n}$ are Hermitian and there exist a real number α such that $\alpha A + B$ is positive definite, then there exists $S \in \mathbb{C}^{n\times n}$, nonsingular, such that S^*AS and S^*BS are both diagonal.*

Pf. Let $P = \alpha A + B$ be positive definite. Since $B = P - \alpha A$, we set up to show that A and P are simultaneously diagonalizable by congruence. Since $P = P^*$ is positive definite $\Rightarrow \exists\, C$ such that $C^*PC = I$ (just take $C = P^{-1/2}$). Now take C^*AC which is Hermitian $\Rightarrow \exists\, U$ unitary, such that $U^*(C^*AC)U = D$ (diagonal). Take $S = CU$, then $U^*C^*PCU = I \quad \text{and} \quad U^*C^*ACU = D$. □

Corollary 2.5.4 *If A is positive definite and B is Hermitian, then $\exists\, S : S^*AS = I$ and $S^*BS = D$, diagonal.* □

We are now ready to give a classical result about ordering of positive definite matrices.

Theorem 2.5.5 *A, B be Hermitian and positive definite. Then*
(a) $A \succ B \Leftrightarrow \rho(A^{-1}B) < 1$. (Similarly, $A \succeq B \succ 0 \Leftrightarrow \rho(A^{-1}B) \leq 1$.)
(b) $A \succ B \Leftrightarrow B^{-1} \succ A^{-1}$.
(c) If $A \succeq B \Rightarrow \det A \geq \det B$ and $\operatorname{tr}(A) \geq \operatorname{tr}(B)$.

Pf.

(a) By Corollary 2.5.4 we can assume that $A = I$ and $B = D = \operatorname{diag}(d_i,\ i = 1, \dots, n)$, and we have to show that $I \succ D \Leftrightarrow \rho(D) < 1$. But this is obvious.

(b) Again, by Corollary 2.5.4 we can assume that $A = I$ and $B = D = \operatorname{diag}(d_i,\ i = 1, \dots, n)$, and we have to show that $I \succ D \Leftrightarrow D^{-1} \succ I$ which is again obvious.

(c) Since $A \succeq B$, then from part (a) $\rho(A^{-1}B) \le 1$ and therefore $\sigma(A^{-1}B) \in (0, 1]$. (For this last inference, we have used that $A^{-1}B$ has all positive eigenvalues, which follows from $A^{-1}B = A^{-1/2}\big(A^{-1/2}BA^{1/2}\big)A^{-1/2}$). Therefore, $\prod \lambda_i(A^{-1}B) \le 1$ and so $\det(A^{-1}B) \le 1$ from which $\det A^{-1} \det B \le 1$, or $\det B \le \det A$. For the trace, from Corollary 2.5.4 we can write $A = CC^*$, $B = CDC^*$ and $0 < d_i \le 1$, $i = 1, \dots, n$. So $\operatorname{tr}(A) = \sum_{i,j=1}^n |c_{ij}|^2$, $\operatorname{tr}(B) = \operatorname{tr}(CDC^*) = \operatorname{tr}(DCC^*) = \sum_{i,j=1}^n d_i |c_{ij}|^2 \le \operatorname{tr} A$. $\square$

- An interesting form of ordering is obtained for matrices depending on a real variable.

Definition 2.5.6 *Given $A : \mathbb{R} \to \mathbb{C}^{n\times n}$, and let α be a real parameter. We say that $A(\alpha)$ is a continuously differentiable function of α if its entries are, and we write $A \in C^1(\mathbb{R}, \mathbb{C}^{n\times n})$. For the derivative, we write $\frac{dA}{d\alpha}$ or also $\dot{A}(\alpha)$.*

Remark 2.5.7 *Observe that if $A(\alpha)$ is Hermitian for all α, then so is $\dot{A}(\alpha)$.*

Lemma 2.5.8 (Monotone function) *Suppose $A \in C^1(\mathbb{R}, \mathbb{C}^{n\times n})$ and Hermitian. If $\dot{A}(\alpha)$ is positive definite, then A is an increasing function; that is, $A(s) \prec A(t)$ for $s < t$.*

Pf. Let $x \in \mathbb{C}^n$, $x \neq 0$. Take $\frac{d}{d\alpha}(x, A(\alpha)x) = (x, \dot{A}x) > 0$. So, if we let the function $g(\alpha)$ be defined as $g(\alpha) = (x, A(\alpha)x)$, then g is increasing and therefore $(x, A(s)x) < (x, A(t)x)$, $s < t$, for any $x \neq 0$. $\square$

As we saw, the positive definite partial order is kept under addition (see Exercise 2.5.2-(2)). Unfortunately, the situation for the product is not as favorable. In fact, in general, product of Hermitian matrices may even fail to be Hermitian!

Exercise 2.5.9 *Give an example of 2×2 matrices A and B, to show that the product of two Hermitian matrices is not necessarily Hermitian. (Hint: take two symmetric matrices).*

Because of the above Exercise, it makes sense to introduce the so-called "*symmetrized product*" of two Hermitian matrices $A, B \in \mathbb{C}^{n\times n}$:

$$S = AB + BA\,.$$

Clearly $S = S^*$ and $(x, Sx) = (x, ABx) + (x, BAx) = (Ax, Bx) + (Bx, Ax)$. If $A, B \in \mathbb{R}^{n\times n}$ (symmetric) $\Rightarrow (x, Sx) = 2(Ax, Bx)$.

Remark 2.5.10 *Note that if A, B are positive definite, we cannot say that the symmetrized product S is positive definite. In fact, A, B positive definite means $(x, Ax) > 0, (x, Bx) > 0$, $\forall\, x$. So:*

$$\begin{aligned} 0 < x^*Ax = \|x\| \cdot \|Ax\| \cos\alpha \to -\frac{\pi}{2} < \alpha < \frac{\pi}{2} \\ 0 < x^*Bx = \|x\| \cdot \|Bx\| \cos\beta \to -\frac{\pi}{2} < \beta < \frac{\pi}{2} \end{aligned}$$

However, we cannot say much about the angle between Ax and Bx.

Exercise 2.5.11 *Give two positive definite matrices in $\mathbb{R}^{2\times 2}$ whose symmetrized product is not positive definite.*

In spite of the last Exercise, we have that if S and one of A or B are positive definite, then the other matrix is as well.

Theorem 2.5.12 (On symmetrized product) *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian and such that A is positive definite and that $S = AB + BA$ is also positive definite. Then B is positive definite.*

Pf. Take $B(\alpha) = \alpha A + B$, $\alpha \geq 0$. Consider

$$\begin{aligned} S(\alpha) &= AB(\alpha) + B(\alpha)A = AB + 2\alpha A^2 + BA \\ &= \underbrace{S}_{\succ 0} + 2\underbrace{\alpha A^2}_{\succeq 0} \Rightarrow S(\alpha) \succ 0, \quad \forall\, \alpha \geq 0\;. \end{aligned}$$

Now, consider $B(\alpha)$. Since A is positive definite and B is Hermitian, then by Corollary 2.5.4 there exists C nonsingular such that $C^*AC = D_A$ and $C^*BC = D_B$, both diagonal. Then, $C^*B(\alpha)C = \alpha D_A + D_B$ $\therefore$ for α sufficiently large, say $\alpha > \alpha_0$, we have that $\alpha D_A + D_B$ is positive definite and so is $B(\alpha)$. In particular, $\lambda(B(\alpha)) > 0$ for $\alpha > \alpha_0$, where $\lambda(B(\alpha))$ denote any of the eigenvalues of $B(\alpha)$. Next, suppose that $B(\alpha)$ is not positive definite for all $\alpha \geq 0$. Then, since $B(\alpha)$ is continuous and Hermitian, so are its eigenvalues[5], which of course are real for all values of α. Then there is a value $\hat{\alpha} \in [0, \alpha_0]$ such that $B(\hat{\alpha})$ has an eigenvalue equal to 0. Then, $\exists\, z \neq 0$ such that $B(\hat{\alpha})z = 0$. Take this z and consider $z^* S(\hat{\alpha}) z = z^* A \underbrace{B(\hat{\alpha})z}_{=0} + \underbrace{z^* B(\alpha)}_{=0} Az = 0$. But this contradicts that $S(\alpha) \succ 0$, $\forall\, \alpha \geq 0$. $\square$

- There is an interesting consequence of Theorem 2.5.12 and Lemma 2.5.8.

Theorem 2.5.13 (Square root ordering) *Let A, B be positive definite such that $A \succ B$. Then $\sqrt{A} \succ \sqrt{B}$, where we are taking the unique positive definite square root.*

Pf. Let $A(\alpha) = B + \alpha(A - B)$ for $\alpha \geq 0$. Since $A(\alpha)$ is positive definite, we let $S(\alpha) = \sqrt{A(\alpha)}$, $S^2(\alpha) = A(\alpha)$, and $S(\alpha)$ is positive definite. Now, **assume** that $S(\alpha)$ is differentiable in α (it is, see Chapter 4). Then

$$\dot{S}S + S\dot{S} = \dot{A}.$$

But $S \succ 0$ and $\dot{A} = A - B \succ 0$, then by Theorem 2.5.12 $\dot{S} \succ 0 \Rightarrow S$ is increasing. In particular $S(0) \prec S(1) \Rightarrow \sqrt{B} \prec \sqrt{A}$. $\square$

Remark 2.5.14 *The above Theorem can be extended to all so-called monotone functions of positive definite matrices; e.g., the negative inverse, $-A^{-1}$, the exponential, e^A, and the principal branch of the logarithm, $\log(A)$, are all monotone. [See [7] for a complete characterization of all monotone functions.]*

Exercises 2.5.15

(1) *Give an example of $0 \prec B \prec A$ for which it is not true that $B^2 \prec A^2$.*
(2) [**Harder**.] *Show that if $A(\alpha) = B + \alpha(A - B)$ with $A \succ B \succ 0$, $\alpha \geq 0$, then the unique positive definite square root $S(\alpha) = (A(\alpha))^{1/2}$ is differentiable in α.*
(3) *Show that if $0 \prec B \prec A \Rightarrow \log(B) \preceq \log(A)$. Here, $\log(A)$, where A is positive definite, can be defined from the Schur form of A. That is, if $A = U\Lambda U^*$, where U is unitary and the diagonal matrix $\Lambda \succ 0$, then $\log(A) = U\log(\Lambda)U^*$. It is important to note that $\log(A)$ is symmetric, and –as expected– $e^{\log(A)} = A$.* (Hint: To prove the stated result, you may want to **assume** that $\log(C(\alpha))$ is differentiable in α if the positive definite function $C(\alpha)$ is.)

[5]They are roots of the characteristic polynomial, hence they depend continuously on the polynomial coefficients, which in turn depend continuously on the entries of $B(\alpha)$.

2.5.1 Determinantal Inequalities

One of the few equalities involving determinant is of course that $\det(AB) = \det(A)\det(B)$. Also, for $A \succeq B \succeq 0$ we know that $\det(A) \geq \det B$. But it is harder to give good bounds on $\det(A)$ itself in terms of easily computed quantities, or on $\det(A+B)$ in terms of $\det A$ and $\det B$. This is our present goal.

Before presenting the celebrated Hadamard inequality, let us recall the arithmetic-geometric mean inequality:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad x_i \geq 0\ ,\ i = 1,\ldots,n\ . \tag{2.5.1}$$

Exercises 2.5.16

(1) *Show the arithmetic-geometric mean inequality (2.5.1) following these steps.*

   i) *Take any convex function f, that is, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$, $\alpha \in [0,1]$ and $x, y \in J$ (some interval of $\mathbb{R}^+$). By induction, generalize this to*

   $$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i), \quad n = 2, 3, \ldots$$

   where $\alpha_i \geq 0$, $i = 1:n$, $\sum_{i=1}^{n} \alpha_i = 1$ and $x_1, \ldots, x_n \in J$.

   ii) *Now use $f(x) = -\log(x)$, $J = (0,\infty)$, and obtain a generalization of (2.5.1).*

   iii) *Finally, choose the α_i's appropriately.*

(2) *Show that there is equality in* (2.5.1) *iff all x_i are equal. [Hint: $x + 1 \leq e^x$.]*

Theorem 2.5.17 (Hadamard Inequality) *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian and non-negative definite. Then*

$$\det(A) \leq \prod_{i=1}^{n} a_{ii}\,. \tag{2.5.2}$$

If A is positive definite, then we have equality iff A is diagonal.

Pf. If $0 \in \sigma(A) \Rightarrow \det A = 0$ and since $a_{ii} = e_i^T A e_i \geq 0$, there is nothing to prove. So, assume A invertible and so A is positive definite (since 0 is not an eigenvalue). Therefore, $a_{ii} > 0$. Now, take $d_j = \frac{1}{\sqrt{a_{jj}}}$, $D = \mathrm{diag}(d_j, j = 1, \ldots, n)$, and consider DAD. Now: DAD is a positive definite matrix (congruence of positive definite) with 1's on the diagonal, and since $\det DAD = (\det D)^2 \det A$, we have $\det DAD \leq 1 \leftrightarrow \det A \leq \prod_{i=1}^n a_{ii}$. So, without loss of generality we can assume that A is positive definite and has all diagonal entries equal to 1. Obviously $\det A = \prod_{i=1}^n \lambda_i$, where $\lambda_i > 0$, $i = 1, \ldots, n$. From (2.5.1), we obtain

$$\prod_{i=1}^n \lambda_i \leq \left(\frac{1}{n}\sum_{i=1}^n \lambda_i\right)^n = \left(\frac{1}{n}\mathrm{tr}(A)\right)^n = \left(\frac{1}{n}(1 + \cdots + 1)\right)^n = 1 \ ,$$

and so $\det A \leq 1$ is derived.

Finally, if $\det A = 1$ then $\prod_{i=1}^n \lambda_i = \frac{1}{n}\sum_{i=1}^n \lambda_i$ that is there is equality in the arithmetic-geometric mean inequality. But, this is possible (see Exercise 2.5.16-(2)) only if all λ_i's are equal and therefore all equal to 1. Then, A is Hermitian, positive-definite, with all eigenvalues equal to 1, that is $A = I$. $\square$

Corollary 2.5.18 (Hadamard) *Let C be any matrix in $\mathbb{C}^{n\times n}$. Write $C = [c_1, \ldots, c_n]$ (columnwise). Then $|\det C| \leq \prod_{j=1}^n \|c_j\|$. If C is nonsingular, we have equality iff the columns of C are orthogonal. Here, $\|\cdot\|$ is the 2-norm.*

Pf. If C is singular, there is nothing to prove. So, let C be invertible and consider $A = C^*C$ which is positive-definite $\therefore \det A \leq \prod_{i=1}^n a_{ii}$. But $a_{ii} = c_i^* c_i = \|c_i\|^2$ and $\det A = \det C^* \det C$
$\therefore |\det C|^2 \leq \left(\prod_{j=1}^n \|c_j\|\right)^2$. Finally, the columns of C are orthogonal precisely when A is diagonal and in this case we have equality in Hadamard theorem. $\square$

Remark 2.5.19 *Note that – if $C \in \mathbb{R}^{n\times n}$ – the above Corollary states that among all parallelepipeds with edges of length $\|c_j\|$, the one of largest volume is rectangular.*

- A final useful extension of Hadamard theorem is related to block partitioning of positive definite matrices. First, we have this simple Lemma.

Lemma 2.5.20 *Let $A \in \mathbb{C}^{n\times n}$ be positive definite. Write A in partitioned form as $A = (A_{ij})_{i,j=1}^p$, where $A_{ij} \in \mathbb{C}^{n_i \times n_j}$, $i, j = 1, \ldots, p$, and $n_1 + \cdots + n_p = n$. Then, the diagonal blocks A_{ii} are positive definite, $i = 1, \ldots, p$.*

Pf. Since $z^*Az > 0$, for all nonzero vectors z, it suffices to take vectors z partitioned conformally with A, as $z = \begin{bmatrix} z_1 \\ \vdots \\ z_p \end{bmatrix}$, letting $z_i \neq 0$, and $z_j = 0$, $j \neq i$, for each $i = 1, \dots, p$. $\square$

The next result is really a block-extension of the Hadamard inequality (2.5.2).

Theorem 2.5.21 (Fischer) *Let* $M = \begin{array}{c} \\ n\{ \\ m\{ \end{array} \!\!\begin{pmatrix} \overbrace{A}^{n} & \overbrace{B}^{m} \\ B^* & C \end{pmatrix} \in \mathbb{C}^{m+n,m+n}$ *be positive definite, with* $A \in \mathbb{C}^{n\times n}$*,* $B \in \mathbb{C}^{m+m}$*. Then:*

$$\det M \leq \det A \cdot \det C\,. \tag{2.5.3}$$

Pf. We consider a congruence with $\begin{pmatrix} I & X \\ 0 & I \end{pmatrix}$:

$$\begin{aligned}\det M &= \det \begin{pmatrix} I & 0 \\ X^* & I \end{pmatrix} \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \det \begin{pmatrix} A & B \\ X^*A_1B^* & X^*B + C \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \\ &= \det \begin{pmatrix} A & AX + B \\ X^*A + B^* & X^*AX + B^*X + X^*B + C \end{pmatrix}.\end{aligned}$$

Now, thanks to the previous Lemma, A is positive definite and thus so is A^{-1}. So, we choose $X = -A^{-1}B$ and obtain $\det M = \det \begin{pmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{pmatrix}$ $\therefore \det M = \det A \cdot \det(C - B^*A^{-1}B)$. Now we show that $\det(C - B^*A^{-1}B) \leq \det C$. To achieve this, it is sufficient to show that $C \succeq C - B^*A^{-1}B \succ 0$ (see Theorem 2.5.5-(c)). But $C \succeq C - B^*A^{-1}B \leftrightarrow C - (C - B^*A^{-1}B) \succeq 0 \leftrightarrow B^*A^{-1}B \succeq 0$. The latter relation is clearly true, since for any $z \in \mathbb{C}^n$, we have $z^*B^*A^{-1}Bz = (Bz)^*A^{-1}Bz \geq 0$. Finally, that $C - B^*A^{-1}B \succ 0$ is a consequence of the fact that we did a congruence transformation of M to obtain $\begin{pmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{pmatrix}$ and thus the sub-block $C - B^*A^{-1}B$ is positive definite. $\square$

Exercises 2.5.22

(1) *Deduce Hadamard inequality* (2.5.2) *from Fischer inequality* (2.5.3).
(2) *Prove Fischer inequality using the Choleski factorization of* M*.*

(3) Prove "Ostrowski" inequality: "Given $A \in C^{n\times n}$, consider $S = \frac{A+A^*}{2}$ and assume S is positive definite. Then $\det S \le |\det A|$*."* [Hint: $A = S+H$, $H = (A-A^*)/2$. So, show $|\det(I + S^{-1}H)| \ge 1$.]

There are also important and useful determinantal inequalities for the sum of positive-definite matrices.

Theorem 2.5.23 *Let $A, B \in \mathbb{C}^{n\times n}$ be positive definite. Then*

$$\det(\alpha A + (1-\alpha)B) \ge (\det A)^{\alpha}(\det B)^{1-\alpha} \ , \quad \forall\, \alpha \in [0,1].$$

Pf. $\det(\alpha A+(1-\alpha)B) = \det[B(\alpha B^{-1}A+(1-\alpha)I)] \overset{C=B^{-1}A}{=} \det B\cdot\det(\alpha C+(1-\alpha)I)$
$\therefore$ we need to show $\det(\alpha C + (1-\alpha)I) \ge (\det A)^{\alpha}(\det B)^{-\alpha} = (\det C)^{\alpha}$. Now, let λ_j be eigenvalues of C. So, we need to show

$$\prod_{j=1}^{n}(\alpha\lambda_j + (1-\alpha)) \ge \prod_{j=1}^{n}\lambda_j^{\alpha} \ .$$

Now, recall that $\lambda_j > 0$ because $C = B^{-1}A$ and A, B positive-definite (see Problem 2 of Homework Set # 3). So, the result follows if

$$\alpha x + (1-\alpha) \ge x^{\alpha}, \quad \forall\, x > 0 \text{ and } 0 \le \alpha \le 1 \ .$$

But this is a simple calculation [do it!]. $\square$

The last result we give is reminiscent of a similar inequality in real analysis.

Theorem 2.5.24 (Minkowski) *Let $A, B \in \mathbb{C}^{n\times n}$ be positive-definite. Then*

$$(\det(A+B))^{1/n} \ge (\det A)^{1/n} + (\det B)^{1/n}.$$

Pf. $(\det(A+B))^{1/n} \ge (\det A)^{1/n} + (\det B)^{1/n}$ if and only if
$(\det A^{-1/2})^{1/n}(\det(A+B))^{1/n}(\det A^{-1/2})^{1/n} \ge (\det A^{-1/2})^{1/n}[(\det A)^{1/n}+(\det B)^{1/n}](\det A^{-1/2})^{1/n}$
that is if and only if $(\det(I + A^{-1/2}BA^{-1/2}))^{1/n} \ge \det I + (\det(A^{-1/2}BA^{-1/2}))^{1/n}$. Observe that $A^{-1/2}BA^{-1/2}$ is positive definite and thus without loss of generality we show $(\det(I+B))^{1/n} \ge 1 + (\det B)^{1/n}$, where B is positive definite. But, letting λ_j's be the eigenvalues of B, then this is the same as

$$\left(\prod_{j=1}^{n}(1+\lambda_j)\right)^{1/n} \ge 1 + \left(\prod \lambda_j\right)^{1/n} \leftrightarrow \frac{1 + \left(\prod \lambda_j\right)^{1/n}}{\left(\prod_{j=1}^{n}(1+\lambda_j)\right)^{1/n}} \le \ 1 \ .$$

But
$$\frac{1+(\prod\lambda_j)^{1/n}}{\left(\prod_{j=1}^n(1+\lambda_j)\right)^{1/n}}=\left(\prod\frac{1}{1+\lambda_j}\right)^{1/n}+\left(\prod\frac{\lambda_j}{1+\lambda_j}\right)^{1/n}$$
and thanks to (2.5.1) this last quantity is lesser than or equal to
$$\frac{1}{n}\sum_j\frac{1}{1+\lambda_j}+\frac{1}{n}\sum_j\frac{\lambda_j}{1+\lambda_j}=\frac{1}{n}\sum_j\frac{1+\lambda_j}{1+\lambda_j}=1\,. \qquad \square$$

Exercises 2.5.25

(1) Let $A,B\in\mathbb{C}^{n\times n}$ be positive definite. Show: $\det(A+B)\ge\det A+\det B$.

(2) Let A be positive definite. Show that
$$\det A=\min\left\{\prod_{i=1}^n v_i^*Av_i : \{v_1,\dots,v_n\}\ \textit{is orthonormal set}\right\}\ .$$

2.5.2 Variational characterization of the eigenvalues and more inequalities

For a general matrix $A\in\mathbb{C}^{n\times n}$, it is hard to characterize the eigenvalues except as roots of the characteristic polynomial. However, if $A=A^*$ they can be characterized as solutions of optimization problems. Below, the norm is always the 2-norm.

So, let $A=A^*$ in what follows. For historical reasons, let eigenvalues (which are real) be labeled as
$$\lambda_{\min}=\lambda_1\le\lambda_2\le\cdots\le\lambda_n=\lambda_{\max}\ .$$

Theorem 2.5.26 (Rayleigh-Ritz) *We have*
$$\lambda_{\min}=\lambda_1=\min_{x\neq0}\frac{x^*Ax}{x^*x}=\min_{\|x\|=1}x^*Ax\ ,$$
$$\lambda_{\max}=\lambda_n=\max_{x\neq0}\frac{x^*Ax}{x^*x}=\max_{\|x\|=1}x^*Ax\ .$$

Pf. Let U unitary be such that $U^*AU=\Lambda=\operatorname{diag}(\lambda_1,\dots,\lambda_n)$. Let $x\in\mathbb{C}^n$, then $x^*Ax=x^*UU^*AUU^*x=(U^*x)^*\Lambda(U^*x)=y^*\Lambda y=\sum_{i=1}^n\lambda_i|y_i|^2$ where we have set $y=U^*x$. So, $\lambda_{\min}\sum_{i=1}^n|y_i|^2\le x^*Ax=\sum_{i=1}^n\lambda_i|y_i|^2\le\lambda_{\max}\sum_{i=1}^n|y_i|^2$. Since

$\sum_{i=1}^n |y_i|^2 = y^*y = x^*x$, then we get $\lambda_1 x^*x \leq x^*Ax \leq \lambda_n x^*x$. Observe that these estimates are sharp: if take $x = u_1 \Rightarrow$ get λ_1, take $x = u_n \Rightarrow$ get λ_n. Finally if $x \neq 0 \Rightarrow \lambda_1 \leq \frac{x^*Ax}{x^*x} \leq \lambda_n$. But, since these are attained, then

$$\lambda_n = \max_{x\neq 0} \frac{x^*Ax}{x^*x} = \max_{\|x\|=1} x^*Ax \ ,$$
$$\lambda_1 = \min_{x\neq 0} \frac{x^*Ax}{x^*x} = \min_{\|x\|=1} x^*Ax \ . \qquad \square$$

Notation: The quantity $\frac{x^*Ax}{x^*x}$ is called *Rayleigh quotient*.

Remarks 2.5.27

(1) Since unit ball is compact (we are in finite dimension), then λ_1 is min *value of x^*Ax as x ranges over unit sphere. (λ_n is* max *value of x^*Ax as x ranges over unit sphere.)*

(2) If $x \neq 0$ is any vector in $\mathbb{C}^n$, then since $\lambda_1 \leq \frac{x^*Ax}{x^*x} \leq \lambda_n$, by letting $\alpha = \frac{x^*Ax}{x^*x}$, then A has at least one eigenvalue in $[\alpha, \infty)$ and at least one eigenvalue in $(-\infty, \alpha]$.

- What can we say about the other eigenvalues? The idea is to realize that the extremal eigenvalues λ_1 and λ_n have been obtained as solutions of unconstrained minimazation/maximization problems. If we restrict our search to appropriate subspaces, we will manage to characterize also non-extremal eigenvalues. The next exercise gives the key insight.

Exercise 2.5.28 Let $x \in \mathbb{C}^n : x^*u_1 = 0$ and look for $\min\limits_{\substack{x\neq 0\\ x\perp u_1}} \frac{x^*Ax}{x^*x} = \min\limits_{\substack{\|x\|=1\\ x\perp u_1}} x^*Ax$. With $U : \ U^*AU = \Lambda$, we have

$$x^*Ax \underset{y=U^*x}{=} \sum_{i=2}^n \lambda_i |y_i|^2 \geq \lambda_2 \sum_{i=2}^n |y_i|^2 \underset{y_i=0}{=} \lambda_2 \sum_{i=1}^n |U^*x|^2 = \lambda_2 (x^*x)$$

$\therefore \min\limits_{\substack{\|x\|=1\\ x\perp u_1}} x^*Ax = \lambda_2$. In the same way, it is immediate to obtain $\max\limits_{\substack{\|x\|=1\\ x\perp u_n}} x^*Ax = \lambda_{n-1}$. And more generally

$$\begin{cases} \min\limits_{\substack{x\neq 0\\ x\perp u_1,u_2,\dots,u_{k-1}}} \frac{x^*Ax}{x^*x} = \lambda_k, & k = 2,3,\dots,n \\ \max\limits_{\substack{x\neq 0\\ x\perp u_n,\dots,u_{n-k+1}}} \frac{x^*Ax}{x^*x} = \lambda_{n-k}, & k = 1,2,\dots,n-1 \ . \end{cases}$$

$\square$

- Using this, we now show the celebrated

Theorem 2.5.29 (Courant-Fischer Mini-Max Theorem) *Let $A \in \mathbb{C}^{n\times n}$, $A = A^*$, with eigenvalues $\lambda_1 \le \cdots \le \lambda_n$, and let $k : 1 \le k \le n$. Then*

$$\min_{w_1,\dots,w_{n-k}} \max_{\substack{x\neq 0\\ x\perp w_1,\dots,w_{n-k}}} \frac{x^*Ax}{x^*x} = \lambda_k \quad (\min\text{-}\max)\ ,$$

$$\max_{w_1,\dots,w_{k-1}} \min_{\substack{x\neq 0\\ x\perp w_1,\dots,w_{k-1}}} \frac{x^*Ax}{x^*x} = \lambda_k \quad (\max\text{-}\min)\ .$$

Pf. Take $x \neq 0 \Rightarrow \frac{x^*Ax}{x^*x} = \frac{x^*U\Lambda U^*x}{x^*UU^*x} = \frac{y^*\Lambda y}{y^*y}$. Show (min-max)

$$\sup_{\substack{x\neq 0\\ x\perp w_1,\dots w_{n-k}}} \frac{x^*Ax}{x^*x} = \sup_{\substack{y\neq 0\\ y\perp U^*w_1,\dots,U^*w_{n-k}}} \frac{y^*\Lambda y}{y^*y} = \sup_{\substack{\|y\|=1\\ y\perp U^*w_1,\dots,U^*w_{n-k}}} \sum_{i=1}^n \lambda_i |y_i|^2$$

$$\ge \sup_{\substack{\|y\|=1\\ y\perp U^*w_1,\dots,U^*w_{n-k}\\ y_1=\cdots=y_{k-1}=0}} \sum_{i=1}^n \lambda_i |y_i|^2 = \sup_{\substack{\sum_{j=k}^n |y_j|^2=1\\ y\perp U^*w_1,\dots,U^*w_{n-k}}} \sum_{i=k}^n \lambda_i |y_i|^2 \ge \lambda_k$$

$\therefore$ $\sup_{x\perp w_1,\dots,w_{n-k}} \frac{x^*Ax}{x^*x} \ge \lambda_k$. But, if we choose $w_1 = u_n, \dots, w_{n-k} = u_{k+1}$, then have equality, and therefore $\inf_{w_1,\dots,w_{n-k}} \sup_{\substack{x\neq 0\\ x\perp w_1,\dots,w_{n-k}}} \frac{x^*Ax}{x^*x} = \lambda_k$ and since the extremum is attained we can use min and max. To show (max-min) is similar: use "inf" instead of "sup" to obtain

$$\inf_{\substack{x\neq 0\\ x\perp w_1,\dots w_{k-1}}} \frac{x^*Ax}{x^*x} \le \inf_{\substack{\|y\|=1\\ y\perp U^*w_1,\dots,U^*w_{k-1}\\ y_n=\cdots=y_{k+1}=0}} \sum_{i=1}^k \lambda_i |y_i|^2 \le \lambda_k$$

then, choose $w_1 = u_1, \dots, w_{k-1} = u_{k-1}$. $\square$

Remark 2.5.30 *The mini-max theorem can be also formulated in an equivalent subspace notation. That is, with S denoting a subspace of $\mathbb{C}^n$ of the stated dimension, one has:*

$$\min_{\dim(S)=k} \max_{\substack{x\neq 0\\ x\in S}} \frac{x^*Ax}{x^*x} = \lambda_k\ ,$$

$$\max_{\dim(S)=n-k+1} \min_{\substack{x\neq 0\\ x\in S}} \frac{x^*Ax}{x^*x} = \lambda_k\ .$$

- If A is not Hermitian, there is no nice mini-max theorem for the eigenvalues. However, something can still be said.

Exercise 2.5.31 *Let $A \in \mathbb{C}^{n\times n}$, not necessarily Hermitian. Show*

$$\min_{\|x\|=1} |x^*Ax| \leq |\lambda_i| \leq \max_{\|x\|=1} |x^*Ax| \ .$$

[Observation: We cannot claim that these bounds are attained at eigenvectors, unlike the case of A Hermitian. For example, if $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \lambda_1 = \lambda_2 = 1$, x is the only eigenvector, and $x^*Ax = 1$. But if we take

$$x = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow x^*Ax = \frac{\sqrt{2}}{2}[1,1]\frac{\sqrt{2}}{2}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{3}{2} > 1.] \qquad \square$$

- Also, even if A is not Hermitian, a variational characterization can always be given for the singular values.

Corollary 2.5.32 (min-max for singular values) *$A \in \mathbb{C}^{m\times n}$, $m \geq n$, $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ be singular values, and let k: $1 \leq k \leq n$. Then*

$$\min_{w_1,\dots,w_{k-1}} \ \max_{\substack{x\neq 0 \\ x\perp w_1,\dots,w_{k-1}}} \frac{\|Ax\|}{\|x\|} = \sigma_k \ ,$$

$$\max_{w_1,\dots,w_{n-k}} \ \min_{\substack{x\neq 0 \\ x\perp w_1,\dots,w_{n-k}}} \frac{\|Ax\|}{\|x\|} = \sigma_k \ .$$

Pf. Since $\sigma_1^2 \geq \cdots \geq \sigma_n^2$ are the eigenvalues of A^*A, we only need to take care of the "opposite" ordering of eigenvalues with respect to singular values and just apply the mini-max theorem 2.5.29. $\square$

Among the most important consequences of the min-max theorem 2.5.29 are those results which allow us to give bounds on eigenvalues of sum of matrices, or of bordered matrices. We see these next.

Theorem 2.5.33 (Weyl) *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian. Let $\lambda_1(A+B) \leq \cdots \leq \lambda_n(A+B)$ be the ordered eigenvalues of $A+B$ and $\lambda_1(A) \leq \cdots \leq \lambda_n(A)$ and $\lambda_1(B) \leq \cdots \leq \lambda_n(B)$ be the ordered eigenvalues of A and B. Then, for any $k = 1,\dots,n$, we have*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B) \ .$$

.

Corollary 2.5.34 (Monotonicity) *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian, with usual ordering of eigenvalues.*

(1) If $B \succeq 0$, then $\lambda_k(A+B) \geq \lambda_k(A)$, $k = 1, \ldots, n$. (Eigenvalues cannot decrease if we add a nonnegative definite matrix.)

(2) If $A \succeq B \succeq 0$, then $\lambda_k(A) \geq \lambda_k(B)$, $k = 1, \ldots, n$. □

Exercise 2.5.35 *Prove Theorem 2.5.33 and Corollary 2.5.34.*

- In case B of Corollary 2.5.34 is of rank 1, then we get a very useful interlacing result.
  Note that if $B \in \mathbb{C}^{n\times n}$ is Hermitian and of rank 1, then it is of the form $B = \pm zz^*$.

Indeed, it is enough to take the Schur form of B : $B = U \begin{pmatrix} \lambda_1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix} U^* = u_1\lambda_1 u_1^* = \pm u_1|\lambda_1|^{1/2}|\lambda_1|^{1/2}u_1^*$. In one case $B \succeq 0$, in the other case $0 \succeq B$.

Theorem 2.5.36 (Interlacing Theorem) *Let A be Hermitian and let $B = zz^*$ ($z \neq 0$). Then $\lambda_1(A) \leq \lambda_1(A + zz^*) \leq \lambda_2(A) \leq \lambda_2(A + zz^*) \leq \cdots \leq \lambda_n(A) \leq \lambda_n(A + zz^*)$.*

Pf. Since $B \succeq 0$, we only need to show that $\lambda_k(A+zz^*) \leq \lambda_{k+1}(A)$, $k = 1, 2, \ldots, n-1$. We know:

$$\begin{aligned}
\lambda_k(A + zz^*) &= \max_{w_1,\ldots,w_{k-1}} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_{k-1}}} \frac{x^*(A + zz^*)x}{x^*x} \leq \text{(search for min on a smaller set)}\\
&\leq \max_{w_1,\ldots,w_{k-1}} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_{k-1}\\ x\perp z}} \frac{x^*(A + zz^*)x}{x^*x}\\
&= \max_{\substack{w_1,\ldots,w_{k-1}\\ w_k = z}} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots,w_{k-1},w_k}} \frac{x^*Ax}{x^*x} \leq \text{(search for max on a larger set)}\\
&\leq \max_{w_1,\ldots,w_k} \min_{\substack{x\neq 0\\ x\perp w_1,\ldots w_k}} \frac{x^*Ax}{x^*x} = \lambda_{k+1}(A)\ .
\end{aligned}$$

□

Remark 2.5.37 *Of course, there is an immediate analog for $A - zz^*$:*

$$\lambda_1(A - zz^*) \le \lambda_1(A) \le \cdots \le \lambda_n(A - zz^*) \le \lambda_n(A) \ .$$

(Just let $A \leftarrow A - zz^*$ in Theorem 2.5.36.)

Exercise 2.5.38 *Prove or disprove*

$$\lambda_k(A + zz^*) \le \lambda_{k+1}(A - zz^*), \quad k = 1, 2, \ldots, n-1.$$

- The next result gives a different type of interlacing, when we add a row/column to a given matrix. It is very useful in so called updating techniques.

Theorem 2.5.39 (Bordered matrix interlacing) *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. Let $z \in \mathbb{C}^n$, $b \in \mathbb{R}$, and form $B \in \mathbb{C}^{n\times 1, n+1}$, Hermitian, as $B = \begin{bmatrix} A & z \\ z^* & b \end{bmatrix}$. With usual ordering of the eigenvalues, $\lambda_1(A) \le \cdots \le \lambda_n(A)$ and $\lambda_1(B) \le \cdots \le \lambda_{n+1}(B)$, we have $\lambda_1(B) \le \lambda_1(A) \le \lambda_2(B) \le \lambda_2(A) \le \cdots \le \lambda_n(B) \le \lambda_n(A) \le \lambda_{n+1}(B)$.*

Pf. We will show that $\lambda_k(B) \le \lambda_k(A) \le \lambda_{k+1}(B)$, for any $k = 1, 2, \ldots, n$. From mini-max theorem, we have

$$\begin{aligned}\lambda_{k+1}(B) &= \min_{y_1,\ldots,y_{n+1-(k+1)} \in \mathbb{C}^{n+1}} \ \max_{\substack{\hat{x}\neq 0, \hat{x}\in\mathbb{C}^{n+1} \\ \hat{x}\perp y_1,\ldots,y_{n+1-(k+1)}}} \frac{\hat{x}^* B \hat{x}}{\hat{x}^*\hat{x}} \ge \text{(max on a smaller set)} \\ &\ge \min_{y_1,\ldots,y_{n-k} \in \mathbb{C}^{n+1}} \ \max_{\substack{\hat{x}\neq 0, \hat{x}\in\mathbb{C}^{n+1} \\ \hat{x}\perp y_1,\ldots,y_{n-k} \\ \hat{x}\perp e_{n+1}}} \frac{\hat{x}^* B \hat{x}}{\hat{x}^*\hat{x}}.\end{aligned}$$

Now, let $\hat{x} = \begin{bmatrix} x \\ \xi \end{bmatrix}$, $x \in \mathbb{C}^n$ and $y_i = \begin{bmatrix} w_i \\ \eta_i \end{bmatrix}$, $\eta_i \in \mathbb{C}$, $w_i \in \mathbb{C}^n$, $i = 1, \ldots, n-k$. If $\hat{x} \perp y_1, \ldots, y_{n-k}$ and $\hat{x} \perp e_{n+1} \Rightarrow \xi = 0$ and $x \perp w_1, \ldots, w_{n-k}$ and the values of η_i are immaterial. Further $\hat{x}^* B \hat{x} = x^* A x$. Therefore:

$$\lambda_{k+1}(B) \ge \min_{w_1,\ldots,w_{n-k} \in \mathbb{C}^n} \ \max_{\substack{x\neq 0, x\in\mathbb{C}^n \\ x\perp w_1,\ldots,w_{n-k}}} \frac{x^* A x}{x^* x} = \lambda_k(A) \ .$$

Similarly:

$$\lambda_k(B) = \max_{y_1,\ldots,y_{k-1} \in \mathbb{C}^{n+1}} \ \min_{\substack{\hat{x}\neq 0, \hat{x}\in\mathbb{C}^{n+1} \\ \hat{x}\perp y_1,\ldots,y_{k-1}}} \frac{\hat{x}^* B \hat{x}}{\hat{x}^*\hat{x}} \le \text{(min on a smaller set)}$$

$$\leq \max_{y_1,\ldots,y_{k-1}} \min_{\substack{\hat{x}\neq 0 \\ \hat{x}\perp y_1,\ldots,y_{k-1} \\ \hat{x}\perp e_{n+1}}} \frac{\hat{x}^* B\hat{x}}{\hat{x}^*\hat{x}}$$

$$= \max_{w_1,\ldots,w_{k-1}\in\mathbb{C}^n} \min_{\substack{x\neq 0, x\in\mathbb{C}^n \\ x\perp w_1,\ldots,w_{k-1}}} \frac{x^*Ax}{x^*x} = \lambda_k(A)\ . \qquad \square$$

The matrix B in Theorem 2.5.39 is called a "bordering" of A.

Example 2.5.40 *Let* $A = [a] \in \mathbb{R}, z \in \mathbb{C}, b \in \mathbb{R} \Rightarrow B = \begin{bmatrix} a & z \\ \bar{z} & b \end{bmatrix}$. *We just said that* $\lambda_1(B) \leq a \leq \lambda_2(B)$. *Since eigenvalues of* B *satisfy* $(a-\lambda)(b-\lambda) - |z|^2 = 0 \to \lambda_{1,2} = \frac{a+b}{2} \mp \frac{\sqrt{(a-b)^2+4|z|^2}}{2}$, *we are saying simply that*

$$\frac{a+b}{2} - \frac{\sqrt{(a-b)^2+4|z|^2}}{2} \leq a \leq \frac{a+b}{2} + \frac{\sqrt{(a-b)^2+4|z|^2}}{2}\ . \qquad \square$$

Remarks 2.5.41

(1) It is often useful to interpret Theorem 2.5.39 the other way around. That is, when we remove the last row and column from given Hermitian matrix.

(2) It is also useful to observe that there is nothing special about the column and row being the last one. We could have inserted (or deleted) any pair of row and column with same index. (This is because we can use similarity by permutation.) In Example 2.5.40, we are simply saying that $\lambda_1(B) \leq b \leq \lambda_2(B)$.

Taking the last remark to its logical conclusion, we could also consider what happens when we delete several rows and columns (with same index) from a given Hermitian matrix. Obviously, we are left with a submatrix of A which is a *principal sub-matrix of A*. Here, the word "principal" means precisely this: obtained from deleting any number of rows and columns of same index. Naturally, every row/column deletion leads to an interlacing of the eigenvalues, progressively coarser.

Exercise 2.5.42 *Prove Theorem 2.5.43.* [Hint: Make repeated application of the interlacing property.]

Theorem 2.5.43 (Inclusion Principle) *If* A_p *is a* $(p\times p)$ *principal submatrix of* A, *for* $p = 1,\ldots,n$, *for the usual ordering of the eigenvalues* $(\lambda_1 \leq \lambda_2 \leq \cdots)$ *we have*

$$\lambda_k(A) \leq \lambda_k(A_p) \leq \lambda_{k+n-p}(A), \quad 1 \leq k \leq p.$$

An important consequence of the inclusion principle is the following result, which is useful when we have information on the inner products $u_i^* A u_j$ with orthonormal vectors, but no explicit knowledge of A nor of its eigenspace, as it is the case in quantum mechanics.

Theorem 2.5.44 (Poincaré separation) *Suppose $A \in \mathbb{C}^{n\times n}$ is Hermitian and let $p : 1 \le p \le n$. Let $u_1, \dots, u_p$ be given orthonormal vectors. Let $B = [u_i^* A u_j] \in \mathbb{C}^{p\times p}$. Let $\lambda_1(A) \le \cdots \le \lambda_n(A)$ and $\lambda_1(B) \le \cdots \le \lambda_p(B)$ be the eigenvalues of A, respectively B. Then:*

$$\lambda_k(A) \le \lambda_k(B) \le \lambda_{k+n-p}(A), \quad k = 1, 2, \dots, p.$$

Pf. Extend $u_1, \dots, u_p$ to an orthonormal basis for $\mathbb{C}^n : u_1, \dots, u_p, u_{p+1}, \dots, u_n$. Form $U = [u_1, \dots, u_n]$ and take U^*AU which is unitarily similar to A. Now B is the leading (p, p) principal submatrix of U^*AU and we can use the inclusion principle. $\square$

Exercises 2.5.45

(1) Let $A \in \mathbb{C}^{m\times n}$, $m \ge n$ with singular values $\sigma_1(A) \ge \cdots \ge \sigma_n(A) \ge 0$. Let B be matrix obtained deleting one column of A: $B \in \mathbb{C}^{m\times(n-1)}$ and let $\sigma_1(B) \ge \cdots \ge \sigma_{n-1}(B) \ge 0$ be its singular values. Show

$$\sigma_1(A) \ge \sigma_1(B) \ge \sigma_2(A) \ge \cdots \ge \sigma_{n-1}(B) \ge \sigma_n(A).$$

(2) Let $A, B \in \mathbb{C}^{m\times n}$, $m \ge n$, and let $\sigma_i(A+B)$, $\sigma_i(A)$, $\sigma_i(B)$ be the ordered singular values, $\sigma_1 \ge \cdots \ge \sigma_n$. Show

$$\sigma_{i+j-1}(A+B) \le \sigma_i(A) + \sigma_j(B), \quad 1 \le i, j \le n, \quad i + j \le n + 1.$$

(Hint: Use $M = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ and Exercise (3) below.)

[Note that this result implies $\sigma_k(A+B) \le \min(\sigma_k(A) + \sigma_1(B), \sigma_1(A) + \sigma_k(B))$, $k = 1 : n$.]

(3) [Harder] Let $A, B \in \mathbb{C}^{n\times n}$, Hermitian, and let the eigenvalues of $A, B, A+B$, be arranged as usual: $\lambda_1 \le \cdots \le \lambda_n$. Show that

$$\lambda_{i+j-n}(A+B) \le \lambda_i(A) + \lambda_j(B), \quad 1 \le i, j \le n, \quad i + j \ge n + 1\ ,$$

and

$$\lambda_{i+j-1}(A+B) \ge \lambda_i(A) + \lambda_j(B), \quad 1 \le i, j \le n, \quad i + j \le n + 1\ .$$

We conclude this section with a simple but useful result and a consequence of Weyl theorem.

Theorem 2.5.46 *Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian and $\lambda_1(A) \le \cdots \le \lambda_n(A)$, $\lambda_1(B) \le \cdots \le \lambda_n(B)$ be their ordered eigenvalues. Then:*

$$|\lambda_j(A) - \lambda_j(B)| \le \|A - B\|, \quad j = 1, 2, \ldots, n.$$

Pf. Recall that if $A \succeq C \Rightarrow \lambda_i(A) \ge \lambda_i(C)$, $i = 1, 2, \ldots, n$. Now, let $C = B + \|A - B\|\ I$, $D = B - \|A - B\|\ I$.

We claim that $D \preceq A \preceq C$. [Indeed, $C - A \succeq 0 \leftrightarrow (B - A) + \|A - B\|\ I \succeq 0$. But this matrix has eigenvalues ≥ 0 since $\|A - B\| = |\lambda_{\max}(A - B)|$, so $C - A \succeq 0$. Similarly $A - D \succeq 0$.]

Therefore, $\lambda_j(B) - \|A - B\| \le \lambda_j(A) \le \lambda_j(B) + \|A - B\|$. $\square$

Remark 2.5.47 *The novelty in Theorem 2.5.46 is that the inequality holds for all the ordered differences of the eigenvalues of A and B, not for the eigenvalues of their difference. [The latter is a simple property of the norm.]*

- Finally, let us recall Weyl's theorem 2.5.33 (A and B are Hermitian with eigenvalues ordered as usual):

  $$\lambda_k(A) + \lambda_1(B) \le \lambda_k(A + B) \le \lambda_k(A) + \lambda_n(B)\,.$$

  Therefore, we immediately get

  $$\alpha\lambda_k(A)+(1-\alpha)\lambda_1(B) \le \lambda_k(\alpha A+(1-\alpha)B) \le \alpha\lambda_k(A)+(1-\alpha)\lambda_n(B)\,, \quad 0 \le \alpha \le 1\,.$$

  In particular, using $k = 1$ and $k = n$, respectively, we get

  $$\lambda_1(\alpha A + (1 - \alpha)B) \ge \alpha\lambda_1(A) + (1 - \alpha)\lambda_1(B)\ .$$

  Wo, we have obtained that

  "*The smallest eigenvalue is a* concave function *on the vector space of Hermitian matrices.*"

  Similarly,

  $$\lambda_n(\alpha A + (1 - \alpha)B) \le \alpha\lambda_n(A) + (1 - \alpha)\lambda_n(B)\ .$$

That is, we found that

"*The largest eigenvalue is a* convex function *on the vector space of Hermitian matrices.*"

Somewhere in between the smallest and largest eigenvalue there is a change in convexity.

Chapter 3

Positive (Nonnegative) Matrices: Perron Frobenius Theory

To prepare these lectures, have used [4], [9] and [1].

3.1 Preliminaries

Here we consider matrices $A \in \mathbb{R}^{m \times n}$ which have positive (or nonnegative) entries. Be careful that these are quite different than positive (or nonnegative) definite matrices. In fact, now matrices are not even required to be symmetric.

- We write $A > 0$, or $A \geq 0$, if the entries $a_{ij} > 0$, or ≥ 0. Similarly, $A > B$ if $A - B > 0$ entrywise.
- If $A \in \mathbb{C}^{m \times n}$ or $\mathbb{R}^{m \times n}$, we also often consider $|A| = (|a_{ij}|)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$.

A number of very simple properties of nonnegative matrices are handy. For example, these are useful (and perhaps not immediately obvious).

Exercises 3.1.1 *Below, matrices are square as needed.*

(a) $|A| \leq |B| \Rightarrow \|A\|_F \leq \|B\|_F$.

(b) $|A^m| \leq |A|^m$.

(c) $0 \leq A \leq B$, $0 \leq C \leq D \Rightarrow 0 \leq AC \leq BD$. [Note: A, B can be in $\mathbb{R}^{m \times n}$ and $C, D \in \mathbb{R}^{n \times p}$].

(d) $0 \leq A \leq B \Rightarrow 0 \leq A^m \leq B^m$, $m = 1, 2, \dots$.

(e) $A > 0$, $x \geq 0$ *but* $x \neq 0 \Rightarrow Ax > 0$.

Our interest in positive/nonnegative matrices is given by the large number of applications where they arise quite naturally. For example, in numerical analysis, in Markov-chains, in Economics (Leontiev input-output models), in theory of games, in dynamical systems.

The next two examples highlight the kind of questions we will try to answer.

Example 3.1.2 (Population Dynamics) *Consider three (hypothetical) species that reside in regions A, B, C. Every month, the entire population of each region splits evenly in the two other regions. Initially, the population of each group is 400, 600, 800, respectively. We want to describe the population distribution after 1 month, 2 months, ..., k months and as $k \to \infty$. Let* $p^{(0)} = \begin{bmatrix} \alpha^{(0)} \\ \beta^{(0)} \\ \gamma^{(0)} \end{bmatrix}$ *be initial population:* $p^{(0)} = \begin{bmatrix} 400 \\ 600 \\ 800 \end{bmatrix}$ *and* $p^{(k)} = \begin{bmatrix} \alpha^{(k)} \\ \beta^{(k)} \\ \gamma^{(k)} \end{bmatrix}$ *be the distribution after k months. We represent this situation as* $p^{(1)} = Ap^{(0)} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} p^{(0)} = \begin{bmatrix} 700 \\ 600 \\ 500 \end{bmatrix}$. *We observe that* $A \geq 0$, $A = A^T$, *and* $\sum_{i=1}^3 a_{ij} = 1 = \sum_{j=1}^3 a_{ij}$, *a fact that will be referred as saying that A is double stochastic. Next,* $p^{(2)} = Ap^{(1)} = A^2 p^{(0)} = \begin{bmatrix} 550 \\ 600 \\ 650 \end{bmatrix}$. *Note that* $A^2 = \frac{1}{4} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ *and we have $A^2 > 0$. Inductively, we get* $p^{(k)} = Ap^{(k-1)} = A^k p^{(0)}$ *and* $A^k = \begin{bmatrix} t_k & t_{k+1} & t_{k+1} \\ t_{k+1} & t_k & t_{k+1} \\ t_{k+1} & t_{k+1} & t_k \end{bmatrix}$; $t_j = \frac{1}{3}\left[1 + \frac{(-1)^j}{2^{j-1}}\right]$, $j = k, k+1$ *(the formulas for t_k are simple to verify by induction). Therefore, we have* $\lim_{k \to \infty} A^k = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ *and so* $\lim_{k \to \infty} p^{(k)} = \begin{bmatrix} 600 \\ 600 \\ 600 \end{bmatrix}$ *is the stationary distribution.*

Let us make a note of these facts: (a) $A = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} \geq 0$. *(b) For the spectral radies, we have* $\rho(A) = 1 = \lambda_{\max}(A)$ *with eigenvector* $e = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$*:* $Ae = e$ *(and note that* $e > 0$*). Here, since* $A = A^T$ *also* $A^T e = e \to e^T A = e^T$*, so that* e *is both a left and right eigenvector of* A*. (c) Also, observe the very interesting fact that* $\lim_{k\to\infty} A^k = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{3}ee^T$*. In particular, if we normalize the positive right and left eigenvectors* x *and* y *such that* $y^T x = 1$*, then as* $k \to \infty$ $A^k \to = xy^T (= yx^T)$*.* □

Example 3.1.3 (Random Walk) *This is a model for a "Random Walk" of a particle in a finite 1-d lattice. There are n states,* $s_1, \dots, s_n$*, and a particle moves from a state to an adjacent one with a certain probability. Namely, if the particle is in state* s_k $(k \neq 1, n)$*, then it moves to the Right (i.e., to* s_{k+1}*) with probability* p_k : $0 < p_k < 1$*, and to the Left (that is, to the state* s_{k-1}*) with probability* $q_k = 1 - p_k$*. If it is in state* s_1*, it goes to the Right with probability 1. If it is in state* s_n*, it goes to the Left with probability 1.*

We will model this situation with a transition from an initial probability distribution to another. That is, if $p^{(0)} = \begin{bmatrix} p_1^{(0)} \\ \vdots \\ p_n^{(0)} \end{bmatrix}$ *is the initial probability distribution* $(p_i^{(0)} \geq 0, \sum_{i=1}^n p_i^{(0)} = 1)$ *then we seek* $p^{(1)} = Ap^{(0)}$*. To get* A*, we reason as follows. Suppose that the initial probability is concentrated all at the initial state* s_1*:* $p^{(0)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$*, then* $p^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ *and so* $Ap^{(0)} = p^{(1)}$ *is the first column of* A*. Similarly, by progressively taking the initial probability concentrated at the* s_i*'s,* $i = 2, \dots, n$*,* *that is taking $p^{(0)} = e_i$, $i = 1 : n$, then we get all of A's columns. This gives*

$$A = \begin{bmatrix} 0 & q_2 & & & & \\ 1 & 0 & q_3 & O & & \\ 0 & p_2 & 0 & \ddots & & \\ \vdots & & p_3 & \ddots & \ddots & \\ & & & & q_{n-1} & \\ 0 & & & p_{n-2} & 0 & 1 \\ 0 & O & & & p_{n-1} & 0 \end{bmatrix}$$

Again, $A \geq 0$. Now $A \neq A^T$. Still: $\sum_{i=1}^n a_{ij} = 1$, $\forall\, j = 1, \ldots, n$: A is a column-stochastic *matrix. Again we observe that $\rho(A) = 1$. However, now it is not easy at all to answer questions such as "What is the asymptotic probability distribution"? In fact, "is there one"? In this example, these questions are harder, because –and this is not obvious at all– $A^m \not> 0$, $\forall\, m]$; in fact, and even this is not obvious, $\rho(A) = 1$ is not attained just when $\lambda = 1$ is an eigenvalue of A.* □

3.2 Perron-Frobenius theory

The understanding of the asymptotic behavior of positive (nonnegative) matrices is the essence of Perron-Frobenius theory. Perron developed the theory for $A > 0$, then Frobenius extended it to the case of $A \geq 0$.

The simple observation is that if $A \geq 0 \Rightarrow A^k \geq 0$, for all $k \Rightarrow$ iterates of a positive vector remain positive; in other words, the positive orthant is invariant. We expect the dominant behavior of an iteration process with A to depend on $\rho(A)$. So, to understand the eventual behavior of iterating with A will lead us to study the spectral radius of A. And, it will be important to see if/when the right/left eigenvectors associated to $\rho(A)$ are in the positive orthant. These are the key tools the theory. Indeed, the Perron-Frobenius theory will tell us that/when spectral radius is also itself an eigenvalue and if/when is of multiplicity one. Moreover, it will also give us information on the associated right and left eigenvectors and on the limiting behavior of A^k.

- Let us recall formulas for the sup-norm (or ∞-norm) and 1-norm of matrices. Here, we can take $A \in \mathbb{C}^{m \times n}$. Recall that (just as for any norm induced by a vector norm) $\|A\|_{\infty,1} = \max_{\|x\|_{\infty,1}=1} \|Ax\|_{\infty,1}$, and that for vectors $\|x\|_\infty = \max_i |x_i|$ and $\|x\|_1 = \sum_i |x_i|$. With this, we have:

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$$
$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}| . \tag{3.2.1}$$

Exercise 3.2.1 *Prove* (3.2.1).

We are now ready to explore properties of the spectral radius of nonnegative matrices.

Facts 3.2.2 *All matrices below are in* $\mathbb{R}^{n\times n}$.

(1) *Let* $|A| \le B$. *Then,* $\rho(A) \le \rho(|A|) \le \rho(B)$.
Pf. Since $A \le |A| \le B \Rightarrow A^m \le |A|^m \le B^m$ and thus $\|A^m\|_F \le \||A|^m\|_F \le \|B^m\|_F \Rightarrow (\|A^m\|_F^{1/m}) \le \||A|^m\|_F^{1/m} \le \|B^m\|_F^{1/m}$. But we know that $\rho(A) = \lim_{m\to\infty} \|A^m\|^{1/m}$ (we did this for 2-norm, but in fact result and proof is true for any norm), and the claim follows. □

(2) *If* $0 \le A \le B$, *then* $\rho(A) \le \rho(B)$.
Pf. Obvious from (1). □

(3) *If* $A \ge 0$ *and* A_p *is a* $p\times p$ *principal submatrix of* A, *then* $\rho(A_p) \le \rho(A)$. *In particular:* $a_{ii} \le \rho(A)$, $i = 1,\dots,n$.
Pf. Take $B \in \mathbb{R}^{n\times n}$ being given by original A with rows/columns of 0's corresponding to the rows/columns of A we deleted to get A_p. Then $0 \le B \le A \Rightarrow \rho(B) \le \rho(A)$. But B is similar via permutation to $\begin{bmatrix} A_p & 0 \\ 0 & 0 \end{bmatrix}$ and so $\rho(B) = \rho(A_p)$. □

(4) *Let* $A \ge 0$. *Then:*

(a) *If* $\sum_{j=1}^n a_{ij}$ *is constant for all* $i = 1,\dots,n$, *then* $\rho(A) = \|A\|_\infty$.
(b) *If* $\sum_{i=1}^n a_{ij}$ *is constant for all* $j = 1,\dots,n$, *then* $\|A\|_1 = \rho(A)$.

Pf. We know that $|\lambda| \le \|A\|$ for any norm. Now, if $\sum_j a_{ij}$ is constant for all $i = 1, \dots, n$, then $e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is an eigenvector: $Ae = \left(\sum_j a_{ij}\right) e \Rightarrow \rho(A) = \|A\|_\infty$ and past (a) follows. For part (b), use the 1-norm. $\square$

(5) *Let $A \geq 0$. Then:*

(a) $\min_i \sum_j a_{ij} \leq \rho(A) \leq \max_i \sum_j a_{ij}$, *and*

(b) $\min_i \sum_j a_{ij} \leq \rho(A) \leq \max_j \sum_i a_{ij}$.

Pf. Let us show (a). Since $\max_i \sum_j a_{ij} = \|A\|_\infty$, then $\rho(A) \leq \max_i \sum_j a_{ij}$ is obvious. To get the other bound, let $c = \min_i \sum_j a_{ij}$. Build $B \geq 0$ such that $\sum_j b_{ij} = c$, $i = 1, \dots, n$, as follows. If $c = 0 \Rightarrow B = 0$, if $c \neq 0$, let $b_{ij} = c \frac{a_{ij}}{\sum_j a_{ij}}$. From this, we have $A \geq B \geq 0$ and $\rho(B) = \min_i \sum_j a_{ij} \leq \rho(A)$, by (2) above. For part (b), use A^T and part (a). $\square$

(6) *Let $A \geq 0$. Let $x \in \mathbb{R}^n$, $x > 0$. Then:*

$$\min_i \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j \leq \rho(A) \leq \max_i \frac{1}{x_i} \sum_j a_{ij} x_j \ .$$

Pf. Build $D = \begin{bmatrix} x_1 & & & O \\ & x_2 & & \\ & & \ddots & \\ O & & & x_n \end{bmatrix}$, so D is invertible and $(D^{-1}AD)_{ij} = \frac{a_{ij}x_j}{x_i}$. Now the result follows from Fact (5) above. $\square$

(7) *If $A \geq 0$, and there is a vector $x \in \mathbb{R}^n$, $x > 0$, such that $\alpha x \leq Ax \leq \beta x$ for some $\alpha, \beta \geq 0$, then $\alpha \leq \rho(A) \leq \beta$. Further, if $\alpha x < Ax \Rightarrow \alpha < \rho(A)$, and if $Ax < \beta x \Rightarrow \rho(A) < \beta$.*

Pf. This follows from (6), because if $\alpha x \leq Ax \Rightarrow \alpha \leq \min_i \frac{1}{x_i} \sum_j a_{ij} x_j$. $\square$

We are now ready to give Perron's theorem, which deals with the case of $A > 0$.

Theorem 3.2.3 (Perron) *Let $A \in \mathbb{R}^{n \times n}$, $A > 0$. Then:*

(a) $\rho(A) > 0$;

(b) $\rho(A)$ *is an eigenvalue of* A *and it is simple (i.e., it has algebraic multiplicity one);*
(c) $\exists\ x \in \mathbb{R}^n$, $x > 0$, *such that* $Ax = \rho(A)x$*;*
(d) $|\lambda| < \rho(A)$, $\forall\ \lambda \neq \rho(A)$ *eigenvalue of* A *(i.e.,* $\rho(A)$ *is the only eigenvalue of largest modulus);*
(e) $\left(\frac{A}{\rho(A)}\right)^k \underset{k\to\infty}{\longrightarrow} L$, $L = xy^T$, $Ax = \rho(A)x$, $A^Ty = \rho(A)y$, $x > 0$, $y > 0$, $x^Ty = 1$.

- The proof will be somewhat long. Before embarking on it, let us make a few observations.

Remarks 3.2.4
(1) Parts (a)–(d) characterize the dominant eigenstructure of A*. Part (e) is about asymptotic behavior and it states that – after rescaling by* $\rho(A)$ *– asymptotically the problem is effectively that of a positive matrix of rank 1.*
(2) The vectors x *and* y *are called right, and left, Perron vectors. To be precise, Perron vectors are typically normalized so that* $\Sigma x_i = \Sigma y_i = 1$*.*

Proof of Theorem 3.2.3. Part (a) is obvious, from Fact 3.2.2-(5).

Let us show the first half of (b) and point (c). Suppose $Ax = \lambda x$, $x \neq 0$ and $|\lambda| = \rho(A)$. Then $|\lambda||x| = \rho(A)|x| = |\lambda x| = |Ax| \leq |A||x| = A|x| \therefore \rho(A)|x| \leq A|x|$. Let $y = A|x| - \rho(A)|x|$, so that $y \geq 0$. Notice that if $y = 0$, then since $A|x| > 0$ (since $x \neq 0$ and $A > 0$) $\Rightarrow |x| = \frac{1}{\rho(A)}A|x| > 0$. So, if $y = 0$ we get (c). Now, if $y \neq 0$, then $0 < Ay = A(A|x| - \rho(A)|x|) \overset{z=A|x|>0}{=} (A - \rho(A)I)A|x| = Az - \rho(A)z \Rightarrow Az > \rho(A)z \overset{\text{Fact 3.2.2}-(7)}{\Longrightarrow} \rho(A) > \rho(A)$ and so $y = 0$ and $A|x| = \rho(A)|x| \therefore \rho(A)$ is an eigenvalue of A with an eigenvector with positive entries and in fact we have seen that if $Ax = \lambda x$, $|\lambda| = \rho(A) \Rightarrow |x| > 0$, and so (c) and the first half of (b) are proven.

We now move to show (d): that $\rho(A)$ is the only eigenvalue of largest modulus. First, we show that if $Ax = \lambda x$, $|\lambda| = \rho(A) \Rightarrow x = e^{i\theta}|x|$, where $|x| : A|x| = \rho(A)|x|$. Reason as follows. Suppose $Ax = \lambda x$, $x \neq 0$ and $|\lambda| = \rho(A) \Rightarrow |Ax| = |\lambda x| = \rho(A)|x|$. But we know that $A|x| = \rho(A)|x|$, $|x| > 0 \Rightarrow \rho(A)|x_k| = |\lambda x_k| = \left|\sum_{j=1}^n a_{kj}x_j\right| \leq \sum_{j=1}^n a_{kj}|x_j| = \rho(A)|x_k|$, for all $k = 1, 2, \ldots, n$. So, we must have $\left|\sum_{j=1}^n a_{kj}x_j\right| = \sum_{j=1}^n a_{kj}|x_j|$. But each entry $x_j = \rho_j e^{i\phi_j} \Rightarrow a_{kj}x_j = (a_{kj}\rho_j)e^{i\phi_j}$ and since $\left|\sum_j a_{kj}x_j\right| = \sum_j a_{kj}|x_j| \Rightarrow$ all phases ϕ_j must be same, $j = 1, \ldots, n$, call it θ. Therefore, $|x| = e^{-i\theta}x > 0$. So, we have shown that if $Ax = \lambda x$, $|\lambda| = \rho(A) \Rightarrow x = e^{i\theta}|x|$, where $|x| : A|x| = \rho(A)|x|$.

Now, suppose λ eigenvalue of A, $\lambda \neq \rho(A)$, $Ax = \lambda x$, and suppose $|\lambda| = \rho(A) \Rightarrow Aw = \lambda w$ and $w = e^{-i\theta}x > 0 \Rightarrow \lambda > 0$ (since $Aw > 0$) $\Rightarrow \lambda = \rho(A)$. So, we have proved (d).

We now show that the *geometric multiplicity* of $\rho(A)$ is 1. Let w, z such that $Aw = \rho(A)w$ and $Az = \rho(A)z$. Then, we know that there $\exists$ vectors $p > 0$, $q > 0$, $w = e^{i\theta_1}p$ and $z = e^{i\theta_2}q$, or $p = e^{-i\theta_1}w$, $q = e^{-i\theta_2}z$. Define $\alpha = \min_{1\leq i\leq n} q_i/p_i$ and take the vector $r = q - \alpha p$. Since each entry of r is $r_j = q_j - (\min_{1\leq i\leq n} q_i/p_i)\, p_j$, then $r \geq 0$ and at least one entry of r is 0. Now, $Ar = Aq - \alpha Ap = \rho(A)q - \alpha\rho(A)p = \rho(A)r$. Therefore, if $r \neq 0 \Rightarrow$ we'd have $r \geq 0$, $A > 0$, $r \neq 0 \Rightarrow Ar > 0 \Rightarrow r > 0$ which is a contradiction since one entry of r is 0. So, $r = 0 \Rightarrow q = \alpha p \Rightarrow w = cz$, and the geometric multiplicity of $\rho(A)$ is 1.

So, there is a unique eigenvector (*right Perron vector*) associated to $\rho(A)$, call it x, and $x > 0$; we can normalize it so that $\sum_i x_i = 1$. Of course, there is also a unique *left Perron vector* $z : A^T z = \rho(A)z$, $z > 0$, normalized so that $\sum_i z_i = 1$.

Next, we show that if we let $y = \frac{1}{x^T z}z$, and hence $x^T y = 1$, then $\lim_{k\to\infty}\left(\frac{A}{\rho(A)}\right)^k = L$, $L = xy^T$. To prove this fact, we make the following observations.

(i) $Lx = x$, $y^T L = y^T$ (this is obvious);

(ii) $L^k = L$, $k = 1, 2, \ldots$, (this is also obvious);

(iii) $A^k L = LA^k = \rho(A)^k L$, $k = 1, 2, \ldots$; [$A^k L = A^{k-1}Axy^k = A^{k-1}\rho(A)xy^k = \cdots = \rho(A)^k L$; similarly for LA^k.]

(iv) $(A - \rho(A)L)^k = A^k - \rho(A)^k L$. [By induction. Obviously true for $k = 1$. Write $(A-\rho(A)L)^k = (A-\rho(A)L)^{k-1}(A-\rho(A)L) =$ (induction hypothesis) $= (A^{k-1} - \rho(A)^{k-1}L)(A - \rho(A)L) = A^k - \rho(A)^{k-1}LA - \rho(A)A^{k-1}L + \rho(A)^k L^2 =$ (use (ii)–(iii)) $= A^k - \rho(A)^k L - \rho(A)^k L + \rho(A)^k L$.]

(v) If $\lambda \neq 0$ is an eigenvalue of $A - \rho(A)L \Rightarrow \lambda$ is also an eigenvalue of A. [Suppose $\lambda \neq 0$ is such that $(A - \rho(A)L)v = \lambda v \Rightarrow (LA - \rho(A)L^2)v = \lambda Lv \overset{\text{(iii)}}{\Rightarrow} (\rho(A)L - \rho(A)L)v = \lambda Lv \Rightarrow Lv = 0 \Rightarrow (A - \rho(A)L)v = \lambda v \leftrightarrow Av = \lambda v$.]

Therefore, to show that $\left(\frac{A}{\rho(A)}\right)^k - L \underset{k\to\infty}{\longrightarrow} 0$ is the same as to show that $\left(\frac{A}{\rho(A)} - L\right)^k \underset{k\to\infty}{\longrightarrow} 0$, and to show this is enough to show that $\rho(A - \rho(A)L) < \rho(A)$ which is further guaranteed (because of (v)) if $\rho(A)$ is not an eigenvalue of $A - \rho(A)L$. By contradiction, suppose that $\rho(A)$ is an eigenvalue of $A - \rho(A)L$. Then, there $\exists\ v \neq 0 :$ $(A - \rho(A)L)v = \rho(A)v$. Then, since $\rho(A)$ has geometric multiplicity 1 as eigenvalue of A, $v = \alpha x$ (see proof of (v)), $\alpha \neq 0 \Rightarrow \alpha Ax - \alpha\rho(A)xy^T x = 0 \Rightarrow \alpha\rho(A)v = 0$ which is a contradiction. Thus, $\rho(A)$ is not an eigenvalue of $A - \rho(A)L$ and so $\rho(A - \rho(A)L) < \rho(A) \therefore \rho\left(\frac{A}{\rho(A)} - L\right) < 1$ and $\left(\frac{A}{\rho(A)} - L\right)^k \underset{k\to\infty}{\longrightarrow} 0$, as desired. That is, (e) is proven.

Last thing left to show is the second part of (a): the algebraic multiplicity of $\rho(A)$ is 1, that is $\rho(A)$ is a simple eigenvalue. Suppose it has multiplicity $p \geq 1$. Take the Schur form of $\frac{A}{\rho(A)}$:

$$\frac{U^*AU}{\rho(A)} = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}, \; R_{11} = \begin{bmatrix} 1 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & & \\ & 0 & & 1 \end{bmatrix}, \quad R_{22} = \begin{bmatrix} \lambda_{p+1}/\rho(A) & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & & & \\ & & & \lambda_n/\rho(A) \end{bmatrix}.$$

Therefore,

$$U^*LU = \lim_{k\to\infty} U^* \left(\frac{A}{\rho(A)}\right)^k U = \begin{bmatrix} 1 & x\cdots x & x & \cdots & x \\ & \ddots & \vdots & & \vdots \\ 0 & & 1 & x & x \\ & & & 0 & x\cdots x \\ & & & & \ddots \; \vdots \\ & 0 & & & 0 \end{bmatrix} =: B\,.$$

But $\text{rank}(B) \geq p$ while $\text{rank}(U^*LU) = 1 \therefore$ we must have $p = 1$. $\square$

Exercises 3.2.5

(1) If $A > 0$, describe all possible asymptotic behaviors of A^m. [Characterize and analyze the three possible cases.]
(2) Let $A > 0$ and let $x > 0$ be the right Perron vector. Verify that $\rho(A) = \sum_{i,j=1}^n a_{ij}x_j$.
(3) Show that if $A \in \mathbb{R}^{n\times n} > 0$, $n > 1$, is invertible, then A^{-1} cannot be ≥ 0.
(4) [Harder] *Suppose $A(\alpha)$, $\alpha \in \mathbb{R}$, are positive: $A(\alpha) > 0$, $\forall\, \alpha \in [a,b]$, and that A depends smoothly on α (i.e., it is continuously differentiable in α). Show that $\rho(A(\alpha))$ also depends smoothly on α, and so do the right/left Perron vectors.* [Hint: first argue for continuous dependence.]

Probably, the most important aspect of Perron Theorem is that

$$\lim_{k\to\infty} \left(\frac{A}{\rho(A)}\right)^k = L = xy^T \quad (Ax = \rho(A)x, A^Ty = \rho(A)y, x^Ty = 1). \tag{3.2.2}$$

The importance of (3.2.2) of course is that it tells us that there is a projection taking place: for k large, the rescaled matrix $(A/\rho(A))^k$ is well approximated by a rank one positive matrix. Although a similar approximation property is true for other matrices (e.g., symmetric matrices with one dominant eigenvalue associated to the spectral radius), it is specifically for positive matrices that this property provides the mathematical justification for search engines like Google.

Unfortunately, to obtain Perron Theorem, the assumption $A > 0$ was essential. In particular, it gave us that $\rho(A)$ is the only eigenvalue of max-modulus. What can we still say if we only have $A \geq 0$?

It is natural to think of perturbing $A \geq 0$ into a positive matrix and then take the limit as the perturbation goes to 0. By doing so, this is what we can say.

Theorem 3.2.6 *If $A \in \mathbb{R}^{n\times n}$, $A \geq 0$, then $\rho(A)$ is an eigenvalue of A and there is $x \geq 0$, $x \neq 0$: $Ax = \rho(A)x$.*

Pf. Take $A(\varepsilon)$: $(A(\varepsilon))_{ij} = a_{ij} + \varepsilon$, for $\varepsilon > 0$. So $A(\varepsilon) > 0$ and $\exists\ x(\varepsilon)$ (Perron vector): $A(\varepsilon)x(\varepsilon) = \rho(A(\varepsilon))x(\varepsilon)$, $\sum_{j=1}^n x_j(\varepsilon) = 1$, $x(\varepsilon) > 0$, $\forall\ \varepsilon > 0$. Now, the set of Perron vectors is inside a compact set (since $\sum_j x_j(\varepsilon) = 1$, $x(\varepsilon) > 0$), and so there exist a sequence $\{\varepsilon_k\}$, monotonically decreasing, $\lim_{k\to\infty}\varepsilon_k = 0$, such that $x(\varepsilon_k)$ converges as $k \to \infty$ to a vector x; further, since $\sum_j x_j = \lim_{x\to\infty}\sum_j x_j(\varepsilon_k) = 1 \Rightarrow x \neq 0$ and $x \geq 0$. Also, consider $\rho(A(\varepsilon_k))$. Observe that $\rho(A(\varepsilon_k)) \geq \rho(A(\varepsilon_{k+1})) \geq \cdots \geq \rho(A)$ (since $A(\varepsilon_k) > A(\varepsilon_{k+1})$, etc.) $\therefore$ the sequence $\rho_k = \rho(A(\varepsilon_k))$ is non-increasing $\therefore\ \exists\ \lim_{k\to\infty}\rho_k = \rho$ and $\rho \geq \rho(A)$. But $Ax = \lim_{k\to\infty} A(\varepsilon_k)x(\varepsilon_k) = \lim_{k\to\infty}\rho(A(\varepsilon_k))x(\varepsilon_k) = \rho x$, $x \neq 0$. So, ρ is an eigenvalue of A, but then $\rho \leq \rho(A) \therefore \rho = \rho(A)$. $\square$

With the help of Theorem 3.2.6, we can now show that in important circumstances $\rho(A)$ is a simple eigenvalue of A, thus also helping to clarify the behavior we observed in Example 3.1.2.

Theorem 3.2.7 *If $A \geq 0$ and $A^k > 0$, for some $k \geq 1$, then $\rho(A)$ is a simple eigenvalue of A.*

Pf. The eigenvalues of A^k are $\lambda_1^k, \ldots, \lambda_n^k$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. We know that $\rho(A)$ is an eigenvalue of A (see Theorem 3.2.6), so $(\rho(A))^k = \rho(A^k)$ is an eigenvalue of A^k. So, if $\rho(A)$ was a multiple eigenvalue of A, then $(\rho(A))^k$ would be a multiple eigenvalue of A^k, but this is not possible since $A^k > 0$ and by Perron Theorem 3.2.3, $\rho(A^k)$ is a simple eigenvalue of A^k. $\square$

- Without additional assumptions on $A \geq 0$, not much more can be said.
- To make some progress, we restrict consideration to *irreducible* matrices.

Definition 3.2.8 *An (n,n) matrix A is called* reducible *if there exists a permutation P such that*

$$PAP^T = \begin{array}{c} \\ r\{ \\ n-r\{ \end{array} \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \end{array} \quad \text{for some } r \geq 1.$$

If no such P exist, then A is called irreducible. □

Remark 3.2.9 *The meaning of reducibility is the following. Suppose we have to solve $Ax = b$ and A is reducible. Then:*

$$PAP^T \underbrace{Px}_{=y} = \underbrace{Pb}_{=c} \rightarrow \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

and the system size has reduced. □

- Obviously, if $A > 0$, then A is irreducible.
- There is a useful -and interesting- connection to a directed graph associated to A.

Definition 3.2.10 *Let $A = (a_{ij})_{i,j=1}^n$ be given. Take n distinct points (vertices, or nodes) in the plane: $P_1, P_2, \dots, P_n$. For $i \neq j$, connect P_i to P_j by a directed arc $\overrightarrow{P_iP_j}$ iff $a_{ij} \neq 0$. The resulting collection of vertices and directed arcs form the directed graph, $\Gamma(A)$, associated to A.*

Example 3.2.11

(a) The matrix of Example 3.1.2: $A = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ *has $\Gamma(A)$ as in Figure 3.1.*

(b) $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ *and (c)* $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ *have the same $\Gamma(A)$; see Figure 3.1.*

(d) $A = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}$ *has $\Gamma(A)$ as in Figure 3.2.*

(e) The matrix of Example 3.1.3: $A = \begin{bmatrix} 0 & q_2 & 0 \\ 1 & 0 & 1 \\ 0 & p_2 & 0 \end{bmatrix}$, $q_2, p_2 > 0$, $q_2 + p_2 = 1$, *has $\Gamma(A)$ as in Figure 3.2.*

Figure 3.1: Directed graphs associated to cases (a) and (b), (c).

Figure 3.2: Directed graphs associated to cases (d) and (e).

Definition 3.2.12 *We say that the directed graph $\Gamma(A)$ is strongly connected if given any two ordered vertices, P_i, P_j ($1 \le i, j \le n$), there is a directed path from P_i to P_j.* $\square$

Notice that there may be more than one directed path between two nodes. E.g., in Example (a) we can go from P_1 to P_2 as $\overrightarrow{P_1P_2}$ or $\overrightarrow{P_1P_3}$, $\overrightarrow{P_3P_2}$.

In Definition 3.2.12, we did require $i \neq j$; but this is irrelevant, since (even by allowing $P_i = P_j$) if $\Gamma(A)$ is strongly connected we can surely come back to P_i through a directed path.

The above Examples (a), (b), (c) and (e) give strongly connected graphs $\Gamma(A)$, example (d) does not. Also Example (e) with $q_2 = 0, 1$, does not give a strongly connected $\Gamma(A)$.

- Our goal is to show that $\Gamma(A)$ is strongly connected if and only if A is irreducible. We will do this in three steps. (In the general case, (1) and (2) below hold if we replace A with $|A|$).

Facts 3.2.13 *Let $A \in \mathbb{R}^{n\times n}$, $A \geq 0$. Then, the following facts hold.*

(1) *There is a directed path of length m in $\Gamma(A)$ from P_i to P_j iff $(A^m)_{ij} > 0$. In particular, there is a directed path of length m from any P_i to any P_j iff $A^m > 0$.*

Pf. By induction. For $m = 1$ is obvious. Now, let $m = 2$, then $(A^2)_{ij} = \sum_{k=1}^{n} a_{ik}a_{kj}$ $\therefore$ $(A^2)_{ij} > 0$ iff for at least one k both a_{ik} and a_{kj} are $\neq 0$, which is true $\Leftrightarrow$ there is a directed path of length 2 between P_i and P_j. In general: $(A^{q+1})_{ij} = \sum_{k=1}^{n}(A^q)_{ik}a_{kj} > 0 \Leftrightarrow$ for at least one k we have $(A^q)_{ik}$ and a_{kj} both positive. By induction hypothesis this is true iff $\exists$ a path of length q from P_i to P_k and a path of length 1 from P_k to P_j $\therefore$ there is a path of length $q+1$ from P_i to P_j. $\square$

(2) *$\Gamma(A)$ is strongly connected $\Leftrightarrow (I + A)^{n-1} > 0$.*

Pf. $(I+A)^{n-1} = I + \binom{n-1}{1}A + \binom{n-1}{2}A^2 + \cdots + A^{n-1} > 0$ if and only if for each (i, j) at least one of $A, A^2, \ldots, A^{n-1}$ has (i, j)-th entry > 0. Now use 3.2.13-(1) above. $\square$

(3) *A is irreducible $\Leftrightarrow (I + A)^{n-1} > 0$.*

Pf. ($\Rightarrow$) We show that $\forall\, x \neq 0$, $x \geq 0 \Rightarrow (I+A)^{n-1}x > 0$. (The result will follow since x is arbitrary; so, we can take $x = e_1, e_2, \ldots, e_n$.) So, let $x \geq 0$, $x \neq 0$, be given, and consider the sequence $\begin{cases} x^{(0)} = x \\ x^{(k+1)} = (I + A)x^{(k)}, \quad k = 0, 1, \ldots, n-2. \end{cases}$

Let $\zeta(x)$ denote the number of 0-components in x. We show that $x^{(k+1)}$ has fewer 0-components than $x^{(k)}$. Obviously, since $x^{(k+1)} = x^{(k)} + Ax^{(k)} \Rightarrow \zeta(x^{(k+1)}) \leq \zeta(x^{(k)})$. Suppose $\zeta(x^{(k+1)}) = \zeta(x^{(k)}) \Rightarrow$ they must be in same position (since $Ax^{(k)} \geq 0) \Rightarrow \exists$ permutation P: $Px^{(k+1)} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \begin{matrix} \}m \\ \}n-m \end{matrix}$, $Px^{(n)} \begin{pmatrix} \beta \\ 0 \end{pmatrix} \begin{matrix} \}m \\ \}n-m \end{matrix}$, $\alpha > 0$, $\beta > 0$. Then $\begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix} + PAP^T \begin{pmatrix} \beta \\ 0 \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \beta \\ 0 \end{pmatrix}$ $\Rightarrow 0 = 0 + A_{21}\beta$ and since $\beta > 0 \Rightarrow$ if $A_{21} \geq 0 \to A_{21}\beta > 0$ $\therefore$ $A_{21} = 0$ $\therefore$ A is reducible, which is a contradiction $\therefore$ $\zeta(x^{(k+1)}) < \zeta(x^{(k)})$ $\therefore$ $\zeta(x^{(n-1)}) = 0$ $\therefore$ $x^{(n-1)} > 0$, but $x^{(n-1)} = (I + A)^{n-1}x^{(0)}$ $\therefore$ $(I + A)^{n-1} > 0$, since $x^{(0)} \geq 0$ was arbitrary.

($\Leftarrow$) Suppose A is reducible $\Rightarrow \exists$ P such that $PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$ and so

(since $P^TP = I$)

$$(I+A)^{n-1} = P(I+PAP^T)^{n-1}P^T = P\begin{pmatrix} I+A_{11} & A_{12} \\ 0 & I+A_{22}\end{pmatrix}^{n-1} P^T =$$

$$P\begin{pmatrix}(I+A_{11})^{n-1} & * \\ 0 & (I+A_{22})^{n-1}\end{pmatrix} P^T$$

$\therefore$ $(I+A)^{n-1}$ is reducible $\therefore$ it cannot be that $(I+A)^{n-1} > 0$. □

Exercises 3.2.14

(1) *Show that an irreducible matrix can have at most $(n-1)$ entries equal to 0 in each row or column.*

(2) *Let $A \geq 0$ be diagonally dominant: $a_{ii} \geq \sum_{j=1}^n a_{ij}$, for all $i = 1, \dots, n$. Show that $\mathrm{Re}(\lambda_i) \geq 0$, for all eigenvalues of A.*

(3) *Show that for a general matrix $A \in \mathbb{R}^{n\times n}$ we have $\rho(I+A) \leq 1+\rho(A)$ and give an example where strict inequality is achieved. Then, show that if $A \geq 0 \Rightarrow \rho(I+A) = 1+\rho(A)$.*

- We are now ready for Frobenius result, which is an extension of Perron's theorem to the case of $A \geq 0$.

Theorem 3.2.15 (Frobenius) *Let $A \in \mathbb{R}^{n\times n}$, $A \geq 0$, and irreducible. Then,*
a) $\rho(A) > 0$ is an eigenvalue of A;
b) $\exists\, x > 0 : Ax = \rho(A)x$;
c) $\rho(A)$ is a simple eigenvalue of A.

Pf. From Theorem 3.2.6, we know that $\rho(A)$ is an eigenvalue of A with eigenvector $x \geq 0$, $x \neq 0$. Further, since A is irreducible $\Rightarrow \sum_j a_{ij} > 0$, for all $i = 1 : n$. Therefore, since $\min_i \sum_j a_{ij} \leq \rho(A) \leq \max_i \sum_j a_{ij}$, then $\rho(A) > 0$.

Now, let $x \geq 0$, $x \neq 0 : Ax = \rho(A)x \Rightarrow (A+I)x = (\rho(A)+1)x$ (we have used Exercise 3.2.14-(3)). Then $(A+I)^{n-1}x = (\rho(A)+I)^{n-1}x$, but $(A+I)^{n-1} > 0 \Rightarrow (A+I)^{n-1}x > 0 \Rightarrow \rho(A)x > 0 \Rightarrow x > 0$. Finally, if $\rho(A)$ is a multiple eigenvalue of $A \Rightarrow 1+\rho(A)$ is a multiple eigenvalue of $I+A$, but $(I+A)^{n-1} > 0$, then $\rho(A)+1$ must be a simple eigenvalue of A. □

Remark 3.2.16 *The vector $x > 0$, when normalized as $\sum_{i=1}^n x_i = 1$, is still called the right Perron vector. Of course, since A^T is irreducible if A is irreducible, then there is also a left Perron vector y.*

So, if $A \geq 0$ and irreducible, we have right/left Perron vectors x and z and we can choose $y = z/(x^T z)$ so that $x^T y = 1$ and let $L = xy^T$. But, do we have that $\lim_{k\to\infty} \left(\frac{A}{\rho(A)}\right)^k = L$? In general, no.

Example 3.2.17 *$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow A \geq 0$ and irreducible, $\rho(A) = 1$ and it is a simple eigenvalue. Perron vectors are $x = \binom{1/2}{1/2} = z$, and $y = \binom{1}{1}$. So, $L = xy^T = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. However, $(A^k) \not\xrightarrow[k\to\infty]{} L$. Note that here $A^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $A^3 = A$, etc., so A does not converge. Also, note that $\rho(A)$ is also obtained at the eigenvalue -1. Nevertheless, there is a convergence "on average", in the sense that the average of two consecutive terms in the sequence is constant and equal to $L/2$; see Exercise 3.2.20-(3). Finally, observe that this Example is a special case of Example 3.1.3.* $\square$

In order to obtain that, for $A \geq 0$, one has $\left(\frac{A}{\rho(A)}\right)^k \xrightarrow[k\to\infty]{} L$, Frobenius realized that this is true if A has only one eigenvalue of max-modulus, $\rho(A)$ itself. He called such matrices *primitive*.

Definition 3.2.18 (Frobenius, 1912) *Consider a matrix $A \geq 0$, irreducible. Let p be the number of eigenvalues of A of modulus $\rho(A)$. Then:*
i) If $p = 1$, then A is called primitive;
ii) If $p > 1$, then A is called cyclic of order p. $\square$

With this, the following is immediate.

Theorem 3.2.19 (Frobenius) *If $A \geq 0$ is primitive, then $\lim_{k\to\infty} \left(\frac{A}{\rho(A)}\right)^k = L = xy^T$ with $x^T y = 1$ and $Ax = \rho(A)x$, $A^T y = \rho(A)y$, $x > 0$, $y > 0$.* $\square$

Exercises 3.2.20

(1) *Prove Theorem 3.2.19.*

(2) *We know that when $A \geq 0$, irreducible, we cannot claim that $\lim_{k\to\infty}(A/\rho(A))^k$ exists. However, give an example to show that there are matrices $A \geq 0$, irreducible, not primitive, for which the limit does exist.*

(3) Let $A \geq 0$ be irreducible, and let x, y, be normalized Perron vectors associated to $\rho(A)$: $x^T y = 1$, and let $L = xy^T$. Show that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \big(A/\rho(A)\big)^j \;=\; L\,. \tag{3.2.3}$$

The next obvious question is: Which matrices are primitive?

Remark 3.2.21 *Clearly, if $A > 0 \Rightarrow A$ is primitive. Indeed, since $A > 0$, then A is irreducible. Moreover, Perron's theorem tells us that $\rho(A)$ is a simple eigenvalue and it is the unique one of max-modulus.*

Bust, beside positive matrices, what other matrices are primitive? In order to give some useful criteria to check whether $A \geq 0$ is primitive, the following preliminary lemmata are useful.

Lemma 3.2.22 *If A is primitive, then A^m is primitive for all $m \geq 1$.*

Pf. Since $\rho(A)$ is the only eigenvalue of max-modulus of A (and, since A is irreducible, it is simple by Theorem 3.2.15), then $(\rho(A))^m$ is the only eigenvalue of max-modulus of A^m, and it is simple. Further, since obviously $A^m \geq 0$ for all m, then we just need to show that A^m is irreducible (we know that A is). Suppose A^m is reducible, for some $m > 1$. Then, $\exists$ permutation P such that $PA^mP^T = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$. Since $Ax = \rho(A)x$, with $x > 0$, then $PA^mP^TPx = \rho(A)^m Px \overset{\hat{x}=Px>0}{\longrightarrow} \begin{bmatrix} B & C \\ 0 & D \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \rho(A)^m \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}$, $\Rightarrow D\hat{x}_2 = \rho(A)^m \hat{x}_2 \therefore \rho(A)^m$ is an eigenvalue of $D \geq 0$ with $\hat{x}_2 > 0$. Now, we likewise have that $A^T \geq 0$ is irreducible, and so $A^T y = \rho(A)y$, $y > 0$. Then $P(A^T)^m P^T Py = \rho(A)^m Py \overset{\hat{y}=Py}{\longrightarrow} \begin{bmatrix} B^T & 0 \\ C^T & D^T \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \rho(A)^m \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} \Rightarrow B^T \hat{y}_1 = \rho(A)^m \hat{y}_1$ and $\hat{y}_1 > 0$. Then, $\rho(A)^m$ is an eigenvalue of B^T, hence of B. But then $\rho(A)^m$ is eigenvalue of B and D $\therefore$ it is not a simple eigenvalue of A^m. Contradiction. $\square$

Lemma 3.2.23 *If $A \geq 0$ is irreducible and $a_{ii} > 0$, $i = 1, \dots, n$, then $A^{n-1} > 0$.*

Pf. By Fact 3.2.13-(3), $A \geq 0$ is irreducible $\Leftrightarrow (I + A)^{n-1} > 0$. Moreover, recall that if $A \geq B \geq 0 \Rightarrow A^m \geq B^m$, for any positive integer m. So, we are going to build a $B \geq 0$ such that $A \geq \mu(I+B)$ with $\mu > 0$ and B irreducible, so that $A^{n-1} \geq \mu^{n-1}(I+B)^{n-1} > 0$.

To build B reason as follows. Take $A \geq 0$ and write

$$A = D \begin{pmatrix} 1 & a_{12}/a_{11} & \cdots & a_{1n}/a_{11} \\ a_{21}/a_{22} & 1 & \cdots & a_{2n}/a_{22} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}/a_{nn} & \cdots & \cdots & 1 \end{pmatrix} =: D(I+B) \ , \ \ D = D = \begin{pmatrix} a_{11} & & O \\ & \ddots & \\ O & & a_{nn} \end{pmatrix}$$

$$\Rightarrow A \geq \min_{1\leq i\leq n}(a_{ii})(I+B) =: \mu(I+B) \ , \quad \mu = \min_{1\leq i\leq n}(a_{ii}) \ .$$

Now, just observe that $B \geq 0$ and irreducible (if not $\Rightarrow PBP^T = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix} \Rightarrow$ $P(I+B^T)P^T = \begin{pmatrix} I+B_{11} & B_{12} \\ 0 & I+B_{22} \end{pmatrix}$ and A would be reducible: $PAP^T = PDP^TP(I+B)P^T$). $\square$

We are now ready for the following result, which fully characterizes primitive matrices, and completely clarifies Example 3.1.2.

Theorem 3.2.24 *Let $A \in \mathbb{R}^{n\times n}$, $A \geq 0$. Then $A^m > 0$ for some $m \geq 1 \Leftrightarrow A$ is primitive.*

Pf. ($\Rightarrow$) Since $A^m > 0 \Rightarrow A^m$ is primitive. If A was not primitive, then it would have more than the eigenvalue $\rho(A)$ of max-modulus, but then also A^m would have more than the eigenvalue $(\rho(A))^m$ of max-modulus, which is a contradiction.

($\Leftarrow$) Since A is primitive, then by Lemma 3.2.22 A^m is primitive (for all $m \geq 1$), in particular it is irreducible, as is A. Now, since A is irreducible, there is a path in $\Gamma(A)$ from P_1 back to P_1, say of length k_1. But then $(A^{k_1})_{1,1} > 0$ and A^{k_1} is primitive (power of a primitive matrix). Now, for $\Gamma(A^{k_1})$ there is a path from P_2 to P_2, say of length k_2. Then $((A^{k_1})^{k_2})_{2,2} > 0$ as well of course as $((A^{k_1})^{k_2})_{1,1} > 0$. Continuing this way we get that $A^{k_1 k_2 \cdots k_n}$ is primitive with positive diagonal entries and so by Lemma 3.2.23 $(A^{k_1\cdots k_n})^{n-1} > 0$. $\square$

It is interesting to obtain good bounds on the smallest integer $\gamma(A)$ such that (if A is primitive) $A^{\gamma(A)} > 0$. The value $\gamma(A)$ is called *index of primitivity* of A. Good bounds for $\gamma(A)$ are useful, because in a practical situation we may have a matrix $A \geq 0$, but do not know if it is primitive. Suppose we start taking powers of A: A^m, $m = 1, 2, \ldots$. Can we stop at some value m and declare that if $A^m \not> 0$, then A is not primitive?

It is not hard to argue that $\gamma(A) \leq (n-1)n^n$ (i.e., that each k_i in the proof of Theorem 3.2.24 satisfies $k_i \leq n$). (Verify that!) Also, by Lemma 3.2.23, we know that $\gamma(A) \leq (n-1)$ if $a_{ii} > 0, i = 1, \ldots, n$. The following result of Wielandt gives the optimal upper bound (see [9]):

Let $A \geq 0$. Then A is primitive $\iff$ $A^{\gamma(A)} > 0$, where

$$\gamma(A) \leq (n-1)^2 + 1 \ . \tag{3.2.4}$$

Exercise 3.2.25 *Prove this statement or provide a counterexample: "If A and B are primitive, then AB is primitive".*

3.2.1 Stochastic matrices

We conclude our discussion of nonnegative matrices with a few considerations on stochastic matrices.

Example 3.2.26 *Suppose $A \geq 0$, $A \neq 0$, and it has a positive eigenvector x. Then, we know that $Ax = \rho(A)x$. [Simply because if $Ax = \lambda x \Rightarrow \lambda \geq 0$ and so $\lambda x \leq Ax \leq \lambda x \Rightarrow \rho(A) = \lambda > 0$; that $\lambda > 0$ is a consequence of $Ax > 0$.] Now, take $D = \begin{bmatrix} x_1 & & O \\ & \ddots & \\ O & & x_n \end{bmatrix} \Rightarrow$ since $Ax = \rho(A)x \rightarrow ADe = \rho(A)De$, where $e = (1, \ldots, 1)^T \therefore (D^{-1}AD)e = \rho(A)e$. But this means that each row of $\frac{D^{-1}AD}{\rho(A)}$ has nonnegative entries adding up to 1. In other words, each row of A is akin to a vector of probabilities.* □

Motivated by Example 3.2.26, the following definition is natural.

Definition 3.2.27 *If $A \geq 0$ is such that $\sum_{j=1}^n a_{ij} = 1$ for all $i = 1, \ldots, n$. then A is called (row) stochastic. If $\sum_{i=1}^n a_{ij} = 1$ for all $j = 1, \ldots, n$, then A is called (column) stochastic. If A is both row and column stochastic then it is called doubly stochastic.* □

Example 3.2.28
(1) $A \geq 0$ is doubly stochastic iff $Ae = e$ and $A^Te = e$.

(2) *Any* $A \geq 0$, $A \neq 0$, *with positive eigenvector* $x > 0$ *is such that* $\frac{A}{\rho(A)}$ *is similar to a stochastic matrix.*
(3) *Permutations are always doubly stochastic.*
(4) *If* U *is unitary* $\Rightarrow$ *the matrix* $A = |U|^2 = (|u_{ij}|^2)_{i,j=1}^n$ *is doubly stochastic. (Examples (3) and (4) are called ortho-stochastic).*
(5) *Any* $A \geq 0$*, irreducible, is similar to a stochastic matrix.*

Exercises 3.2.29

(1) *Verify the statements in Example 3.2.28.*
(2) *Show that if* $A \in \mathbb{R}^{2\times 2}$, $A \geq 0$, *and doubly stochastic, then it is* $A = A^T$ *and* $a_{11} = a_{22}$.
(3) *Show that if* A, B *are stochastic (or doubly stochastic), then* AB *is too. That is, the set of stochastic (doubly stochastic) matrices is a semigroup with respect to matrix multiplication. By example, show that it is not a group. It is easy to show that arithmetic mean of stochastic (or doubly stochastic) matrices is also stochastic (doubly stochastic). Generalize this to the convex sum* $\alpha_1 A_1 + \cdots + \alpha_m A_m$, *where* $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = 1$. *[That is, they form a convex set.]*
(4) *Show that if* $A \geq 0$ *is doubly stochastic and reducible* $\Rightarrow \ \exists\ P : PAP^T = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$ *where* A_{11} *and* A_{22} *are doubly stochastic.*

This next theorem is the celebrated Birkhoff-von Neumann theorem and fully characterizes doubly stochastic matrices. The proof we give is from [3].

Theorem 3.2.30 (Birkhoff doubly stochastic) $A \geq 0$ *is doubly stochastic iff there exist a finite value* m *such that* $A = \alpha_1 P_1 + \cdots + \alpha_m P_m$, *where* $\alpha_i \geq 0$, $\sum_{i=1}^m \alpha_i = 1$ *and* $P_1, \ldots, P_m$ *are permutation matrices.* $\square$

Pf. ($\Leftarrow$) This follows from Exercise 3.2.29-(3).

($\Rightarrow$) We will show that the set of doubly stochastic matrices form a polytope whose vertices correspond to permutation matrices, at which point we will use that every polytope is the convex hull of its extreme points (its vertices).

As customary, through the natural identification of a matrix in $\mathbb{R}^{n\times n}$ with a vector in $\mathbb{R}^{n^2}$ obtained upon writing the columns of A one after the other, we can think of the set of DS (doubly stochastic) matrices as points in $\mathbb{R}^{n^2}$.

So, let $A = (a_{ij})$ be a DS matrix. The constraints of the system on the a_{ij}'s are

$$a_{ij} \geq 0,\ \forall i, j, \quad \sum_{j=1}^n a_{ij} = 1,\ \forall i = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^n a_{ij} = 1,\ \forall j = 1, \ldots, n.$$

This defines a polyhedron Q which is actually a polytope since the linear constraints imply that each $0 \leq a_{ij} \leq 1$, and so Q is bounded. Now we show that every extreme point of Q is integral (has integer entries), by showing that any nonintegral point of Q is the center of some line segment inside Q. This will tell us that the vertices of the polytpe are integral, and therefore are permutation matrices.

So, suppose that A is DS and not a permutation, that is the associated vector $x \in Q$ is not integral, and let $0 < x_{i_1,j_1} < 1$. Since $\sum_{j=1}^{n} x_{i_1,j} = 1$, there must be some j_2: $0 < x_{i_1,j_2} < 1$. Similarly, since $\sum_{i=1}^{n} x_{i,j_2} = 1$, there must be some i_2 such that $0 < x_{i_2,j_2} < 1$. This process can be iterated, and we will stop the first time some index (i,j) is repeated. Consider the sequence just obtained, without repeating the first (and final) point, and therefore the sequence must have even length (because we alternated choosing row and column indices). Let $\mu = x_{i',j'}$ be the smallest entry in this sequence. Consider now the point $y \in \mathbb{R}^{n^2}$ associated to a matrix Y built as follows: Y has a 1 in the same position as the first entry of the above sequence, $x_{i,j}$, then -1 corresponding to the second entry, then $+1$ corresponding to the third entry, etc., all other entries of Y being 0. So, y is a vector with entries $1, -1$ or 0, and Y's rows/columns add to 1. Now take the two points $x^+ = x + \mu y$ and $x^- = x - \mu y$. By minimality of μ, x^+ and x^- are in Q, and therefore the associated matrices A^+ and A^- are DS. But, by construction, $x = 1/2(x^+ + x^-)$ and thus $A = 1/2(A^+ + A^-)$ and obviously $A \neq A^+, A^-$, and so x (which is not integral) is not a vertex of Q, and A is not an extreme point of the set of DS matrices. Therefore, every extreme point of Q is integral, it corresponds to a permutation matrix and every DS matrix is a convex combination of permutation matrices. $\square$

Remark 3.2.31 *The above proof is not constructive; constructive proofs can be given and used to provide insight into the best general upper bound for* m *in Theorem 3.2.30. It is interesting (see references in [4]) that the best general upper bound for* m *in Theorem 3.2.30 is* $(n-1)^2 + 1$ *(cfr. with* $\gamma(A)$ *in* (3.2.4)*).*

Exercise 3.2.32 *Use Theorem 3.2.30 to show the Wielandt-Hoffmann inequality* (2.4.4).

Chapter 4

Matrices depending on parameters

The purpose of this chapter is two-fold: to discuss perturbation and smoothness results for eigenvalues (and eigenvectors), and to give general smoothness results for bases of key subspaces, such as the kernel of a matrix.

The default setting will be to have $A \in \mathbb{C}^{n\times n}$, unless otherwise stated. Also, with $\sigma(A)$ we indicate the set of eigenvalues of A (repeated by multiplicity).

First, we investigate variation of eigenvalues (and eigenvectors) on the entries of the underlying matrix. We will discuss two types of results: (i) general bounds when A is perturbed, (ii) what can we say when A depends smoothly on one or more parameters.

Of course, we already encountered several perturbation results for eigenvalues, chiefly for Hermitian matrices (e.g., see the results in Section 2.5.2, but also results like (2.4.4)). What follows complement for general matrices these earlier results.

4.1 General Perturbation Results for Eigenvalues and Eigenvectors

The first result is comforting and simple: "The eigenvalues depend continuously on the entries of A". More precisely, we have the following result.

Theorem 4.1.1 *If $\{A_k\}$ is a sequence of matrices converging to A, then $\sigma(A_k) \xrightarrow[k]{} \sigma(A)$. That is, $\forall\, \varepsilon > 0$, $\exists\, k_\varepsilon$ such that if $k > k_\varepsilon$, then all eigenvalues of A_k are contained in disks of radius ε centered at the eigenvalues of A.*

Pf. This is simply because the roots of polynomials of (exact) degree n depend continuously on the coefficients of the polynomials. So, we can use this fact for the characteristic polynomials of A_k, whose coefficients approach (as $k \to \infty$) those of the characteristic polynomial of A, and the result follows. $\square$

To obtain easily computable localization results (bounds) for the eigenvalues, the following is a classical and useful result.

Theorem 4.1.2 (Gerschgorin Discs) *Given $A \in \mathbb{C}^{n\times n}$, then all eigenvalues of A are located in the union of n closed discs:*

$$\bigcup_{i=1}^{n} \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}| \right\} =: G.$$

Furthermore, if the union of p of these discs form a connected region R, disjoint from the remaining $(n-p)$ discs, then in R there are p eigenvalues.

Pf. Suppose λ is an eigenvalue, so $Ax = \lambda x$, $x \neq 0$. Let $|x_m| = \max_{1\leq i\leq n} |x_i| \Rightarrow \lambda x_m = \sum_{j=1}^{n} a_{mj}x_j \Rightarrow (\lambda - a_{mm})x_m = \sum_{\substack{j=1 \\ j\neq m}}^{n} a_{mj}x_j \Rightarrow (\lambda - a_{mm}) = \sum_{j\neq m} a_{mj}\frac{x_j}{x_m} \Rightarrow |\lambda - a_{mm}| \leq \sum_{j\neq m} |a_{mj}|$. However, we do not know m, so can only conclude that λ is in the union of these discs.

To show the second statement, we resort to a simple homotopy argument. Let $D = \operatorname{diag}(a_{ii}, i = 1, \dots, n)$ and let $B(t) = (1-t)D + tA$, $0 \leq t \leq 1$. So, the statement is surely true for $B(0) = D$. Now, notice that the diagonal entries of $B(t)$ are the same as those of A, so the center of the Gerschgorin discs of $B(t)$ and A are the same for all $t \in [0, 1]$. However, the radii of the discs for $B(t)$ are tR_i, where $R_i = \sum_{j\neq i} |a_{ij}|$ are the radii for the discs relative to A, and $0 \leq t \leq 1$. So, if p discs of A are disjoint from $n-p$ discs of A, for sure also the corresponding p discs of $B(t)$ are disjoints from the corresponding $(n-p)$ discs of $B(t)$. Since discs are closed, the distance between the unions of the two collections of discs for A is $d > 0$. Next, let $d(t)$ be the distance from any eigenvalue $\lambda(t)$ of $B(t)$ in the union of the p-discs of $B(t)$ from the remaining $(n-p)$ discs. Since eigenvalues are continuous, and so is $d(t)$, then $0 < d \leq d(t)$ for all $t \in [0, 1]$, and in particular $d(0) \geq d$. Now, if $\lambda(1)$ happened to be in the union of the $(n-p)$ discs of A, then $d(1) = 0$, and so at some value $0 < t_0 < 1$, we had to have $d(t_0) < d$, which is a contradiction. $\square$

Exercise 4.1.3 *Formulate and prove a Gerschgorin theorem with discs $|z - a_{ii}| \leq \sum_{j\neq i} |a_{ji}|$.*

With this, we can get our first perturbation result.

Theorem 4.1.4 *Let $A \in \mathbb{C}^{n\times n}$ be diagonalizable: $V^{-1}AV = \Lambda = diag(\lambda_i, i = 1, \ldots, n)$. Let $E \in \mathbb{C}^{n\times n}$. For any eigenvalue μ of $A+E$, there is an eigenvalue λ of A such that*

$$|\mu - \lambda| \le (\|V\|_\infty \cdot \|V^{-1}\|_\infty)\|E\|_\infty.$$

Pf. $A+E$ is similar to $V^{-1}(A+E)V = \Lambda + V^{-1}EV$, so we will show that $|\lambda - \mu| \le \|V^{-1}EV\|_\infty$ $(\le \|V^{-1}\|_\infty \cdot \|V\|_\infty \cdot \|E\|_\infty)$. Call $F = V^{-1}EV$, so that $V^{-1}(A+E)V = \Lambda + F$. By Gerschgorin theorem 4.1.2, the eigenvalues of $\Lambda + F$ are in union of discs $\bigcup_{i=1}^n \left\{ |z - \lambda_i - F_{ii}| \le \sum_{\substack{j=1\\ j\ne i}}^n |F_{ij}| \right\}$. But each of these discs is contained in the discs $|z-\lambda_i| \le \sum_{j=1}^n |F_{ij}|$. So, if $\mu \in \sigma(\Lambda + F)$ then there is some $\lambda \in \sigma(\Lambda)$ such that $|\mu - \lambda| \le \|F\|_\infty$. □

Remarks 4.1.5
(1) $\|V\|_\infty \cdot \|V^{-1}\|_\infty$ is the condition number of V in the sup-norm. We write it as $\mathrm{cond}_\infty(A)$ *or simply* $\mathrm{cond}(A)$ *when the norm is clear from the context.*
(2) If $\|E\|$ is small, Theorem 4.1.4 is effectively a continuity result.

Actually, a result like Theorem 4.1.4 holds for any norm for which $\|AB\| \le \|A\| \cdot \|B\|$ and for which $\|D\| = \max_{1\le i\le n} |D_{ii}|$ when D is diagonal. Let us call these norms "diagonally consistent". For such norms (e.g., the 2-norm or the 1-norm, but not the F-norm) we have much the same result, but need a different proof.

Theorem 4.1.6 *Let $A \in \mathbb{C}^{n\times n}$ be diagonalizable: $V^{-1}AV = \Lambda = diag(\lambda_i, i = 1, \ldots, n)$, and let $E \in \mathbb{C}^{n\times n}$. Let $\|\cdot\|$ be a diagonally consistent norm. Then, for any eigenvalue μ of $A+E$, there is an eigenvalue λ of A such that*

$$|\mu - \lambda| \le \|V^{-1}EV\| \le cond(V)\|E\| \ .$$

Pf. As before, we look at $\Lambda + F$, $F = V^{-1}EV$. Let $\mu \in \sigma(\Lambda + F) \Rightarrow \mu I - (\Lambda + F)$ is singular. If $\mu I - \Lambda$ is singular $\Rightarrow \mu$ is eigenvalue of $\Lambda \Rightarrow$ obviously the result is true. So, assume that $\mu I - \Lambda$ is invertible.

Since $\mu I - (\Lambda + F)$ is singular $\Rightarrow (\mu I - \Lambda)^{-1}[\mu I - \Lambda - F] = I - (\mu I - \Lambda)^{-1}F$ is singular. Then, $\|(\mu I - \Lambda)^{-1}F\| \ge 1$ (if $\|I - B\| < 1 \Rightarrow B$ invertible; in fact, $B^{-1} = \sum_{k=0}^\infty (I-B)^k$). So:

$$1 \le \|(\mu I - \Lambda)^{-1}F\| \le \|(\mu I - \Lambda)^{-1}\| \cdot \|F\| = \max_{n\le i\le n} |(\mu - \lambda_i)^{-1}| \cdot \|F\|$$

$$= \frac{\|F\|}{\min_{1\le i\le n}|\mu-\lambda_i|} \Rightarrow \min_{1\le i\le n}|\lambda_i-\mu| \le \|F\|.$$

$\square$

The term $\|V^{-1}\|\cdot\|V\| = \text{cond}(V)$ is unpleasant, since it can be large. Of course, the best situation is when this is as small as possible. Now, observe that since $1 = \|I\| = \|V^{-1}V\| \le \|V^{-1}\|\cdot\|V\|$, then $\text{cond}(V) \ge 1$. So, the best situation is when $\text{cond}(V) = 1$, which is surely guaranteed (in the 2-norm) if V is unitary, since then $\|V^{-1}\| = \|V\| = 1$.

Corollary 4.1.7 *If A is normal, and μ is an eigenvalue of $A+E$, then $\exists\,\lambda$ eigenvalue of A: $|\lambda-\mu| \le \|E\|_2$. A fortiori, also $|\lambda-\mu| \le \|E\|_F$ holds.* $\square$

- As a final result of the above type, we consider a so-called "a posteriori" estimation problem. The problem is the following.

Exercise 4.1.8 *Suppose $A \in \mathbb{C}^{n\times n}$ is diagonalizable by V, and let $\|\cdot\|$ be a diagonally consistent norm. Suppose we have found an approximation to an eigenpair: $(\hat\lambda,\hat x)$, $\hat x \neq 0$ (approximation means that $A\hat x \approx \hat\lambda\hat x$, but $\hat\lambda$ is not equal to any eigenvalue of A).*

Question: *How good an approximation is $\hat\lambda$ to an exact eigenvalue of A?*

Answer: *In general, it depends on cond(V). Let us see the details.*

We form the residual: $r = A\hat x - \hat\lambda\hat x$ (r is computable, and not 0). Thus, $r = V(\Lambda-\hat\lambda I)V^{-1}\hat x \to \hat x = V(\Lambda-\hat\lambda I)^{-1}V^{-1}r$ and so $\|\hat x\| \le \|V(\Lambda-\hat\lambda I)^{-1}V^{-1}\|\cdot\|r\| \le \text{cond}(V)\|r\|\frac{1}{\min_i|\lambda_i-\hat\lambda|}$ and finally $\min_{1\le i\le n}|\lambda_i-\hat\lambda| \le \frac{\|r\|}{\|\hat x\|}\text{cond(V)}$. $\square$

Example 4.1.9 *In particular, from Exercise 4.1.8, if A is normal $\Rightarrow$ cond$_2(V)$ is 1 and if $\|r\|$ is small then we can trust the approximate eigenvalue in the sense that there is always an eigenvalue λ of A close to $\hat\lambda$.*

Nevertheless, not even in this normal case, there is an analogously simple result for the eigenvectors.

For example, take $\hat A = \begin{pmatrix}1 & 0\\ 0 & 1+2\eta\end{pmatrix}$ and $E = \begin{pmatrix}0 & \varepsilon\\ \varepsilon & 0\end{pmatrix}$, $\eta,\varepsilon > 0$ and small, and let $A = \hat A + E$. Now, for the eigenvalues of $\hat A$ and A we have $\sigma(\hat A) = \{\hat\lambda_1 = 1, \hat\lambda_2 = 1+2\eta\}$, $\sigma(A) = \{\lambda_1 = 1+\eta-s, 1+\eta+s\}$, $s = (\varepsilon^2+\eta^2)^{1/2}$, which are indeed close to one another for any ratio of ε and η. Further, as (normalized) *eigenvectors of* $\hat{A}$ *we hae* $\hat{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $\hat{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$*, and for the residuals we have* $r_{1,2} = A\hat{x}_{1,2} - \hat{\lambda}_{1,2}\hat{x}_{1,2}$ *which gives* $r_1 = \begin{pmatrix} 0 \\ \varepsilon \end{pmatrix}$ *and* $r_2 = \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}$*, so that* $\|r_{1,2}\|_2 = \varepsilon$*, which is small, as expected. However, the true (normalized) eigenvectors of* A *are* $\frac{1}{\sqrt{2s(s-\eta)}} \begin{pmatrix} \varepsilon \\ \eta - s \end{pmatrix}$ *and* $\frac{1}{\sqrt{2s(s-\eta)}} \begin{pmatrix} -\eta + s \\ \varepsilon \end{pmatrix}$*, whose limiting behavior depends on the ratio of* ε *and* η*.* $\square$

Exercises 4.1.10

(1) Suppose $\lambda, \mu \in \sigma(A)$*,* $\lambda \neq \mu$*. Let* $v, w \neq 0 \;:\; Av = \lambda v$ *and* $A^* w = \bar{\mu} w$ *(that is,* w *is left eigenvector corresponding to* μ*). Show that* $v^* w = 0$*.*

(2) Let $A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ $(a \in \mathbb{R})$*, and* $E = \begin{bmatrix} \eta & \varepsilon \\ 0 & 0 \end{bmatrix}$*. We know that if* $\mu \in \sigma(A+E)$*, then* $\exists\ \lambda \in \sigma(A) : |\lambda - \mu| \leq \|E\|_2$*. Find* $\|E\|_2$*. Also, discuss the behavior of the eigenvectors of* $A+E$ *and contrast this to the eigenvectors of* A *for different ratios of* η *and* ε*.* [Hint: Different ratios means to consider different curves in the (η, ε)-plane along which we go to the origin. E.g., $\varepsilon = \eta^p$, where $p \in \mathbb{Z}$.]

4.2 Smoothness results

This next set of results is relative to the case of a matrix valued function $A(t)$, $t \in \mathbb{R}$ (could also have t in some interval, the half-line, etc.). Our interest in this case is when A has some good smoothness properties (often, A is analytic in t) and we want to know if/when/how the eigenvalues and eigenvectors have too. Henceforth, we will write $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$ when A is k-times continuously differentiable, as appropriate.

4.2.1 Simple Eigenvalues

- The first results are concerned with the case of a simple eigenvalue. Recall that a *simple eigenvalue* means that that it has algebraic multiplicity 1.

Theorem 4.2.1 *Let $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$, $k \geq 1$. Suppose that at a value t_0, λ_0 is a simple eigenvalue of $A(t_0)$. Then, for $|t - t_0|$ sufficiently small, there is a unique eigenvalue $\lambda(t)$ of $A(t)$ which is a $\mathcal{C}^k$ function of t and equal to λ_0 at t_0.*

Pf. There are many ways to show this result. Here, we will use the implicit function theorem (IFT).

Take the characteristic polynomial of $A(t) = \underbrace{\det(\lambda I - A(t))}_{p(\lambda,t)} = 0$. Since λ_0 is a simple eigenvalue of $A(t_0)$, then $\begin{cases} p(\lambda_0, t_0) = 0 \\ \frac{\partial}{\partial\lambda} p(\lambda, t_0)\big|_{\lambda_0} \neq 0 \end{cases}$.

But then the IFT guarantees that the equation $p(\lambda, t) = 0$ has a unique $\mathcal{C}^k$ solution (a branch) $\lambda(t)$ in a neighborhood of t_0 such that $\lambda(t_0) = \lambda_0$. $\square$

Remarks 4.2.2

(1) Theorem 4.2.1 says that -locally, near t_0- there is a smooth eigenvalue parametrizable in t passing through λ_0 at t_0. If we view this as a curve, then it has a well defined tangent, which we can also get from the IFT. Indeed, for $|t - t_0|$ sufficiently smally, we have $p(\lambda(t), t) = 0$, and thus $\frac{d}{dt} p(\lambda(t), t) = 0$, and so $\frac{\partial}{\partial\lambda} p(\lambda, t)\ \dot{\lambda} + \frac{d\lambda}{dt} p(\lambda, t)\ 1 = 0$. Now, since $\frac{\partial}{\partial\lambda} p(\lambda, t_0)\big|_{\lambda=\lambda_0} \neq 0 \Rightarrow$ what we get is that indeed there is a well defined tangent to the curve $\lambda(t)$ at t_0: $\frac{d\lambda}{dt}\big|_{t_0} = -\left[\frac{p_t(\lambda,t)}{p_\lambda(\lambda,t)}\right]_{(\lambda_0,t_0)}$.

(2) Also, it is important to stress that -as long as $\frac{\partial}{\partial\lambda} p(\lambda, t)\big|_{\lambda(t)} \neq 0$, then the above argument can be continued and $\lambda(t)$ continues to exist as a smooth function. In other words, as long as the eigenvalue $\lambda(t)$ remains simple, it is a globally $\mathcal{C}^k$ function.

- Next, we see that –relatively to a simple eigenvalue– also the corresponding eigenvector can be chosen smoothly.

We will give three different arguments for this result, each of which offers a different insight, uses different tools, and has different strenghts and weaknesses. The first argument is from [7], the second argument will give more explicit information on the derivative, and the third argument is more original, and it will give some extra information as well.

Simple Eigenvalue: Smooth Eigenvector, Part 1

- The following preliminary Lemmata are useful and of independent interest.

Lemma 4.2.3 (On the derivative of $\det(A(t))$**)** *Let* $A \in C^1(\mathbb{R}, \mathbb{C}^{n\times n})$*, and write* A *in partitioned form column-wise:* $A(t) = [a_1(t), a_2(t), \dots, a_n(t)]$*, for all* $t \in \mathbb{R}$*. Then*

$$\frac{d(\det A(t))}{dt} = \det(\dot a_1, a_2, \dots, a_n) + \det(a_1, \dot a_2, \dots, a_n) + \cdots + \det(a_1, a_2, \dots, \dot a_n)\ .$$

Pf. The key observation is that $\det(A)$ is a multilinear function in $a_1, a_2, \dots, a_n$, that is, it is separately linear in each of the arguments $a_1, a_2, \dots, a_n$. Therefore, we have

$$\begin{aligned}
&\det A(t+h) - \det A(t) = \det(a_1(t+h), \dots, a_n(t+h)) - \det(a_1(t), \dots, a_n(t)) \\
&= \det(a_1(t+h) - a_1(t), a_2(t+h), \dots, a_n(t+h)) + \det(a_1(t), a_2(t+h), \dots, a_n(t+h)) \\
&- \det(a_1(t), \dots, a_n(t)) = \det(a_1(t+h) - a_1(t), a_2(t+h), a_3(t+h), \dots, a_n(t+h)) \\
&\quad + \det(a_1(t), a_2(t+h) - a_2(t), a_3(t+h), \dots, a_n(t+h)) \\
&\quad + \det(a_1(t), a_2(t), \dots, a_n(t+h)) - \det(a_1(t), \dots, a_n(t)) = \cdots = \\
&\det(a_1(t+h) - a_1(t), a_2(t+h), \dots, a_n(t+h)) + \det(a_1(t), a_2(t+h) - a_2(t), \dots, a_n(t+h)) \\
&\quad + \cdots + \det(a_1(t), a_2(t), \dots a_n(t+h) - a_n(t)) + \cancel{\det(a_1(t), \dots, a_n(t)) - \det(a_1(t), \dots, a_n(t))}
\end{aligned}$$

Now the result is obtained taking $\lim_{h\to 0} \frac{\det(A(t+h)) - \det(A(t)}{h}$. $\square$

Lemma 4.2.4 *Suppose* $A \in \mathbb{C}^{n\times n}$ *has a simple eigenvalue* $\hat\lambda$*. Then, at least one of the principal minors of* $A - \hat\lambda I$ *is nonsingular.*

Pf. Since $\hat\lambda$ is simple, for the characteristic polynomial $p(\lambda)$ we have $p(\hat\lambda) = 0$ and $\frac{d}{d\lambda}p(\lambda)|_{\hat\lambda} \neq 0$. Let us compute $\frac{d}{d\lambda}p(\lambda)|_{\hat\lambda}$. We have $p(\lambda) = \det(\lambda I - A) = (\lambda e_1 - a_1, \lambda e_2 - a_2, \dots, \lambda e_n - a_n)$. From Lemma 4.2.3, $\frac{d}{d\lambda}p(\lambda)|_{\hat\lambda} = \det(e_1, \hat\lambda e_2 - a_2, \dots, \hat\lambda e_n - a_n) + \cdots + \det(\hat\lambda e_1 - a_1, \dots, e_n)$. Now, observe that $\det(e_1, \hat\lambda e_2 - a_2, \dots, \hat\lambda e_n - a_n)$ is the determinant of the first principal minor of $A - \hat\lambda I$, $\det(\hat\lambda e_1 - a_1, e_2, \dots, \hat\lambda e_n - a_n)$ is the determinant of the 2nd principal minor of $A - \hat\lambda I$, etc.. So, at least one of the determinants of a principal minor of $A - \hat\lambda I$ is not zero. $\square$

Exercise 4.2.5 *Let* $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$ *be invertible for all* t*, and let* $b \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^n)$*. Then, the solution* x *of* $Ax = b$ *is also in* $\mathcal{C}^k(\mathbb{R}, \mathbb{C}^n)$*.*

We are now ready for the anticipated result.

Theorem 4.2.6 *Let* $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$*,* $k \geq 1$*, and let* λ_0 *be a simple eigenvalue of* $A(t_0)$*. For* $|t - t_0|$ *sufficiently small, say* $t \in I_\lambda = (t_0 - a, t_0 + a)$*, let* $\lambda(t)$ *be the*

$\mathcal{C}^k$ eigenvalue going through λ_0 at t_0. Then, there is an associated eigenvector $v(t)$, also a $\mathcal{C}^k$ function of t, for $t \in I_v \subseteq I_\lambda$.

Pf. We have an eigenvalue $\lambda(t)$ such that $(A(t) - \lambda(t)I)$ has a 1-d null space for $t \in I_\lambda$, and $\lambda(t_0) = \lambda_0$. Therefore, from Lemma 4.2.4, one of the principal minors of $\lambda_0 I - A(t_0)$ is invertible, suppose it is the i-th minor, call it $\lambda_0 I_{n-1} - A_i(t_0)$. So, we have $(\lambda_0 I - A(t_0))v_0 = 0$, $v_0 \neq 0$, and $\lambda_0 I_{n-1} - A_i(t_0)$ nonsingular. Now, observe that the i-th component of v_0 is nonzero, $(v_0)_i \neq 0$ (if not $\Rightarrow (\lambda_0 I - A(t_0))v_0 = 0$ with $v_0 \neq 0$, but since $(\lambda_0 I_{n-1} - A_i(t_0))\hat{v}_0 = 0$ has the only solution $\hat{v}_0 = 0$, then $v_0 = 0$, where $\hat{v}_0$ is the result of removing the i-th component from v_0). Therefore, $(v_0)_i \neq 0$ and so without loss of generality can take it to be 1. But then $(\lambda_0 I - A(t_0))v_0 = 0 \to (\lambda_0 I_{n-1} - A_i(t_0))\hat{v}_0 = \hat{a}_i(t_0)$, where $\hat{a}_i(t_0)$ is the i-th column of A with i-th entry removed. Therefore, we can get $\hat{v}_0 = (\lambda_0 I_{n-1} - A_i(t_0))^{-1}\hat{a}_i(t_0)$. However, the matrix $\lambda_0 I_{n-1} - A_i(t)$ must remain invertible for $|t - t_0|$ small, since it has no 0 eigenvalue at t_0, and the eigenvalues are continuous in t. So, for t in some interval I_v centered at t_0, we can define $v(t) :\ v(t_0) = v_0,\ (v(t))_i = 1$, and $\hat{v}(t) = (\lambda(t)I_{n-1} - A_i(t))^{-1}\hat{a}_i(t)$. From Exercise 4.2.5, $\hat{v}$ is $\mathcal{C}^k$, and so is v. $\square$

There are a few shortcomings of the above proof. The first is that it is a local result, ultimately because it relies on the implicit function theorem. The second shortcoming is that it does not extend nicely, since it hinges on selecting a certain principal minor. This is bothersome, because even if $\lambda(t)$, the eigenvalue branch through λ_0 at t_0, remains simple for all t, we cannot say that the i-th principal minor remains invertible for all t, as the next example clarifies.

Example 4.2.7 *Let $A(t) = \begin{pmatrix} \sin t & \sin t \\ \cos t & \cos t \end{pmatrix}$, $t \in (-\pi/4 + \varepsilon, 3\pi/4 - \varepsilon)$, $\varepsilon > 0$ a given small number. The eigenvalues are $\lambda = 0$, $\lambda = \sin t + \cos t$, both simple for all our values of t. Consider $\lambda = 0$, so that we have two principal minors of $A(t) - \lambda I$: $A_1 = \cos t$ and $A_2 = \sin t$, neither of which remains invertible $\forall\ t$ in our interval.* $\square$

One final general unpleasant aspect of the previous argument is that it does not give nice formulas for the derivatives, $\dot{\lambda}$ and $\dot{v}$ at t_0, since we have used the characteristic polynomial. Also, unfortunately we still do not have much insight into the variation of the eigenvector. This is because, as we will see, the behavior of the eigenvector depends also on the complementary subspace to the eigenvector v itself.

Simple Eigenvalue: Smooth Eigenvector, Part 2

Below, we are going to derive an important, and beautiful, formula for the derivative of the eigenvalue, and further show that there must be a smooth eigenvector, without resorting to the principal minors of $\lambda(t)I - A(t)$.

We want to find $v(t)$, smooth, such that $A(t)v(t) = \lambda(t)v(t)$, where $\lambda(t)$ is a simple eigenvalue of $A(t)$, for all $t \in \mathbb{R}$ (or, at least, near t_0). Note that –just as there is a smooth (right) eigenvector v, for t near t_0: $Av = \lambda v$– there is also a smooth left eigenvector w (again, at least near t_0) such that $w^*A = w^*\lambda$, or $A^*w = \bar{\lambda}w$.

Exercise 4.2.8 *Let $A \in \mathbb{C}^{n\times n}$. Let v be a right eigenvector associated to a simple eigenvalue $\lambda \in \sigma(A)$ and let w be an associated left eigenvector. Show that $v^*w \neq 0$.*

Now, reason as follows. Differentiate $Av = \lambda v \Rightarrow \dot{A}v + A\dot{v} = \dot{v}\lambda + v\dot{\lambda}$. Let $w : w^*A = w^*\lambda \Rightarrow w^*\dot{A}v + v^*A\dot{v} = (w^*\dot{v})\lambda + (w^*v)\dot{\lambda}$, or $w^*\dot{A}v = \dot{\lambda}(w^*v)$. But (see Exercise 4.2.8) $w^*v \neq 0$ and so $\dot{\lambda} = \frac{w^*\dot{A}v}{w^*v}$ which is a nice expression for $\dot{\lambda}$, repeated here for later reference and understood to be valid at least in an interval near t_0:

$$\dot{\lambda} \;=\; \frac{w^*\dot{A}v}{w^*v}\,. \tag{4.2.1}$$

With this, we can write $(\dot{A} - \dot{\lambda}I)v + (A - \lambda I)\dot{v} = 0$, and we have found the equation which must be satisfied by $\dot{v}$:

$$(A - \lambda I)\dot{v} = -(\dot{A} - \dot{\lambda}I)v\,. \tag{4.2.2}$$

Of course, on the LHS we have a singular matrix, but notice that a solution exists (and it will be smooth), since the kernel of $(A - \lambda I)$ remains 1-dimensional. In fact, $(A - \lambda I)\dot{v} \in \mathcal{R}(A - \lambda I) = (\mathcal{N}(A^* - \bar{\lambda}I))^{\perp}$ $\therefore$ $\dot{v}$ exists $\Leftrightarrow$ $z^*(\dot{A} - \dot{\lambda}I)v = 0$, $\forall\, z \in \mathcal{N}(A^* - \bar{\lambda}I)$; but (aside from normalization) there is only one such z, which is w. So, a solution exists precisely when $w^*(\dot{A} - \dot{\lambda}I)v = 0$, that is when $\dot{\lambda} = \frac{w^*\dot{A}v}{w^*v}$, which is exactly how we chose it in (4.2.1).

This still leaves open the problem of how to constructively find an expression for $\dot{v}$, now that we know that it must be such that

$$\begin{cases} \dot{\lambda} = (w^*\dot{A}v)/w^*v \\ (A - \lambda I)\dot{v} = -(\dot{A} - \dot{\lambda}I)v\ . \end{cases} \tag{4.2.3}$$

The approach below is essentially in [8].

One possibility is to impose the normalization condition

$$w(t_0)^* v(t) = \text{constant}, \quad \text{say } w_0^* v(t) = 1 \ .$$

[Note that we are using the "reference" left eigenvector at t_0, $w(t_0)$ $\therefore$ this normalization is guaranteed to be valid only locally.]

With this, we have $w_0^* \dot{v} = 0 \Rightarrow w_0 w_0^* \dot{v} = 0$. So, the expression $(A - \lambda I)\dot{v} = -(\dot{A} - \dot{\lambda} I)v$ can be rewritten as

$$(A - \lambda I + w_0 w_0^*)\dot{v} = -(\dot{A} - \dot{\lambda} I)v \ .$$

Let us evaluate the above expression at t_0:

$$(A(t_0) - \lambda_0 I + w_0 w_0^*)\dot{v}(t_0) = (\dot{\lambda}(t_0) I - \dot{A}(t_0))v(t_0),$$

and now we claim that the matrix $G(t_0) := (A(t_0) - \lambda_0 I + w_0 w_0^*)$ is nonsingular. Let us verify this last fact.

Suppose there exists $z \neq 0$, such that $(A_0 - \lambda_0 I + w_0 w_0^*)z = 0$. Then, $(A - \lambda I)z = -(w_0^* z)w_0$ and thus $w_0^*(A_0 - \lambda_0 I)z = -(w_0^* z)w_0^* w_0$, from which we would get that $w_0^* z = 0$. Therefore, $(A_0 - \lambda_0 I)z = 0$ and z must then be a multiple of the right eigenvector v_0, and thus we would have $w_0^* v_0 = 0$. But this is a contradiction, since the inner product of right and left eigenvectors associated to a simple eigenvalue is not 0 (see Exercise 4.2.8).

Therefore, the matrix $G(t_0)$ is invertible and we are ready to summarize the above in the following theorem.

Theorem 4.2.9 *Let $A \in \mathcal{C}^1(\mathbb{R}, \mathbb{C}^{n\times n})$, let λ_0 be a simple eigenvalue of $A(t_0)$, and let v_0, w_0, be the right, left, associated eigenvectors. Let $\lambda(t)$, $v(t)$ and $w(t)$ be the smooth branches of eigenvalue and right/left eigenvectors passing through λ_0, v_0, w_0. Then, for the derivatives of the eigenvalue and eigenvector v at t_0, we have:*

$$\begin{cases} \dot{v}(t_0) = G^{-1}(t_0)(\dot{\lambda}(t_0) I - \dot{A}(t_0))v(t_0) \\ \dot{\lambda}(t_0) = w^*(t_0)\dot{A}(t_0)v(t_0) \\ w^*(t_0)v(t_0) = 1 \ , \end{cases} \tag{4.2.4}$$

where $G(t_0) = A(t_0) - \lambda_0 I + w_0 w_0^*$. □

Clearly, (4.2.4) provides an explicit (and nice) formula for the derivative -at t_0- of $\dot{\lambda}$ and $\dot{v}$. At this stage, the one drawback left is that we do not yet know how to properly continue $\dot{v}$ in t, which we know must be possible as long as $\lambda(t)$ stays simple.

Nevertheless, the approach just taken is very useful and it can be further expanded to obtain also higher derivatives (if A is sufficiently differentiable of course). Let us see how things go for the 2nd derivative. Below, all derivatives are at t_0. From

$$(A-\lambda I)v = 0 \to (\ddot{A}-\ddot{\lambda}I)v + 2(\dot{A}-\dot{\lambda}I)\dot{v} + (A-\lambda I)\ddot{v} = 0 \qquad (4.2.5)$$

$\therefore (A-\lambda I)\ddot{v} \in \mathcal{R}(A-\lambda I) = (\mathcal{N}(A^*-\bar{\lambda}I))^\perp \therefore$ (4.2.5) has a solution iff $z^*[(\ddot{A}-\ddot{\lambda}I)v + 2(\dot{A}-\dot{\lambda}I)\dot{v}] = 0$, $\forall\ z \in \mathcal{N}(A^*-\bar{\lambda}I)$ which is however 1-dimensional, and spanned by w. Thus, (4.2.5) has a solution iff $w^*(A-\lambda I)\ddot{v} + w^*[(\ddot{A}-\ddot{\lambda}I)v + 2(\dot{A}-\dot{\lambda}I)\dot{v}] = 0$ or $\ddot{\lambda}w^*v = w^*\ddot{A}v + 2w^*(\dot{A}-\dot{\lambda}I)\dot{v}$, and thus

$$\ddot{\lambda} = \frac{w^*\ddot{A}v + 2w^*\dot{A}\dot{v} - 2\dot{\lambda}w^*\dot{v}}{w^*v} \quad \text{at} \quad t = t_0\,. \qquad (4.2.6)$$

To get an expression for $\ddot{v}$, use -as before- $w_0^*v = 1 \Rightarrow w_0^*\ddot{v} = 0 \Rightarrow w_0w_0^*\ddot{v} = 0$. Then, from (4.2.5), we get $(A_0 - \lambda_0 I + w_0w_0^*)\ddot{v}_0 = -(\ddot{A}_0 - \ddot{\lambda}_0 I)v_0 - 2(\dot{A}_0 - \dot{\lambda}_0 I)\dot{v}_0$, or

$$\ddot{v}_0 = G^{-1}(t_0)[(\ddot{\lambda}_0 I - \ddot{A}_0)v_0 + 2(\dot{\lambda}_0 I - \dot{A}_0)\dot{v}_0]\,. \qquad (4.2.7)$$

In principle, we could continue these type of calculations and obtain an expansion for $\lambda(t)$ and $v(t)$, which is valid near t_0. These are "regular" Taylor-like expansions. They are useful especially if $A(t)$ is a simple function, for example linear, say $A(t) = A_0 + tE$, because in this case $\ddot{A} = 0$, $\dot{A} = E$.

Exercises 4.2.10

(1) *Show formulas similar to the ones we have derived for the case of* $A(t,s) \in \mathcal{C}^k(\mathbb{R}\times\mathbb{R}, \mathbb{C}^{n\times n})$. *That is, assuming that* $P_0 \equiv (t_0, s_0)$ *is a point where* $A(t,s)$ *has a simple eigenvalue* λ_0, *derive expressions for* $\frac{\partial\lambda}{\partial t}\big|_{P_0}, \frac{\partial\lambda}{\partial s}\big|_{P_0}$, *as well as* $\frac{\partial^2\lambda}{\partial t^2}\big|_{P_0}, \frac{\partial^2\lambda}{\partial t\partial s}\big|_{P_0}$, *and* $\frac{\partial^2\lambda}{\partial s^2}\big|_{P_0}$.

(2) *Find the expressions for* $\frac{\partial v}{\partial s}\big|_{P_0}$ *as well as* $\frac{\partial^2 v}{\partial s\partial t}\big|_{P_0}$ *and* $\frac{\partial^2 v}{\partial t^2}\big|_{P_0}$, *where* v *is the eigenvector associated to the simple eigenvalue.*

Simple Eigenvalue: Smooth Eigenvector, Part 3

So, we know that if $\lambda(t)$ is a simple eigenvalue for all t (or at least near t_0), it stays smooth and so does the associated eigenvector. Still, we are missing a constructive procedure which defines the derivative of the eigenvector, except locally, near a point t_0. Indeed, in one of our approaches to obtain the derivative of the eigenvector at t_0, we neeeded to assume a certain principal minor to be non-singular, and there is no guarantee that the same prncipal minor remains nonsingular as we evolve the simple eigenvalue in t; in the other approach, we used a normalization of the (right) eigenvector v with respect to a fixed left eigenvector w_0 at t_0, and again there is no guarantee that this normalization is valid except near t_0. Then, the idea is to use a moving normalization, which gets automatically updated in t. The difficulty in doing so is that the evolution of an eigenvector now will depend on the (generalized) eigenspace complementary to the one spanned by the eigenvector itself, in other words we will need to bring into play the eigenspace complementary to v.

We now present an approach which resolves this need. The idea is to derive a differential equation whose solution describes the evolution of the eigenvector. It is similar in spirit to a constructive version of the IFT applied to the eigendecomposition of A.

So, we are being more ambitious, and seek a "similarity transformation" $V(\cdot)$ such that $A(t)V(t) = V(t)\begin{bmatrix}\lambda(t) & 0\\ 0 & B(t)\end{bmatrix}$, with V being a smooth invertible function, and $\lambda \notin \sigma(B)$, $\forall\ t$. We will derive differential equations for V, show that they are well defined, and this will imply that V exists and is smooth.

Formally differentiating the relation $AV = V\begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix}$ we get $\dot{A}V + A\dot{V} = \dot{V}\begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix} = V\begin{pmatrix}\dot{\lambda} & 0\\ 0 & \dot{B}\end{pmatrix}$ from which we get $V^{-1}\dot{A}V + V^{-1}AVV^{-1}\dot{V} = V^{-1}\dot{V}\begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix} + \begin{pmatrix}\dot{\lambda} & 0\\ 0 & \dot{B}\end{pmatrix}$.

Now, let $T = V^{-1}\dot{V}$ and use $V^{-1}AV = \begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix}$ (surely true at $t = 0$, for some reference $V(0)$). From this, we get the rewriting

$$V^{-1}\dot{A}V + \begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix} T = T\begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix} + \begin{pmatrix}\dot{\lambda} & 0\\ 0 & \dot{B}\end{pmatrix}.$$

Next, partition $V^{-1}\dot{A}V$ and T as $\begin{pmatrix}\lambda & 0\\ 0 & B\end{pmatrix}$ is; that is,

$$T = \begin{array}{r} \\ 1\} \\ n-1\} \end{array} \begin{array}{cc} \overbrace{\phantom{T_{11}}}^{1} & \overbrace{\phantom{T_{12}}}^{n-1} \\ \multicolumn{2}{c}{\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}} \end{array}, etc. .$$

So, we get:

$$\begin{pmatrix}(V^{-1}\dot{A}V)_{11} & (V^{-1}\dot{A}V)_{12}\\ (V^{-1}\dot{A}V)_{21} & (V^{-1}\dot{A}V)_{22}\end{pmatrix} + \begin{pmatrix}\lambda T_{11} & \lambda T_{12}\\ BT_{21} & BT_{22}\end{pmatrix} = \begin{pmatrix}T_{11}\lambda & T_{12}B\\ T_{21}\lambda & T_{22}B\end{pmatrix} + \begin{pmatrix}\dot{\lambda} & 0\\ 0 & \dot{B}\end{pmatrix},$$

$\therefore\ \dot{\lambda} = (V^{-1}\dot{A}V)_{11} + \lambda T_{11} - T_{11}\lambda$, and so

$$\dot{\lambda} = (V^{-1}\dot{A}V)_{11}\,. \tag{4.2.8}$$

Observe that T_{11} is not uniquely determined. Also, observe that (4.2.8) was already obtained in (4.2.4) (when we chose the normalization $w^*v = 1$).

Of course, we also have the relation $\dot{B} = (V^{-1}\dot{A}V)_{22} + BT_{22} - T_{22}B$, and we note that T_{22} also is not uniquely determined from this relation.

Finally, using the 0-structure of $\begin{pmatrix}\dot{\lambda} & 0\\ 0 & \dot{B}\end{pmatrix}$, we get

$$\begin{cases}0 = \lambda T_{12} - T_{12}B + (V^{-1}\dot{A}V)_{12}\\ 0 = BT_{21} - T_{21}\lambda + (V^{-1}\dot{A}V)_{21}\end{cases}$$

and since $\lambda \notin \sigma(B)$, then smooth T_{12}, T_{21} are uniquely determined as solutions of this linear system.

Therefore, from the relation $V^{-1}\dot{V} = T$, we can obtain the differential equation for V: $\dot{V} = VT$, $T = \begin{pmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{pmatrix}$ with T_{12} and T_{21} uniquely determined as above and T_{11}, T_{22} not yet specified.

Indeed, we are free to choose T_{11} and T_{22}, smooth, and have a differential equation for V, well defined, and giving the desired block eigendecomposition for A. $\square$

Example 4.2.11 *Here we discuss some constructive choices for T_{11}, T_{22}. Recall that $T = V^{-1}\dot{V} = \begin{pmatrix}T_{11} & T_{12}\\ T_{21} & T_{22}\end{pmatrix}$.*

(a) $T_{11} = 0$, $T_{22} = 0 \Rightarrow \dot{V} = V\begin{pmatrix}0 & T_{12}\\ T_{21} & 0\end{pmatrix}$ and if $V = (v, W)$, then $\dot{v} = WT_{21}$, $\dot{W} = v_1 T_{12}$ making it clear that evolution of $v(t)$ depends on its complementary eigen-space.

(b) $\begin{cases}v^*v = 1\\ W^*W = I_{n-1}\end{cases}$. Then $V^*V = \begin{pmatrix}v^*\\ W^*\end{pmatrix}(v, W) = \begin{pmatrix}1 & v^*W\\ W^*v & I\end{pmatrix}$. Now

$$\frac{d}{dt}(V^*V) = \begin{pmatrix}0 & \frac{d}{dt}(v^*W)\\ \frac{d}{dt}(W^*v) & 0\end{pmatrix} = \dot{V}^*V + V^*\dot{V}$$

$$= \dot{V}^* V^{-*} V^* V + V^* V V^{-1} \dot{V} = T^*(V^*V) + (V^*V)T$$
$$\Rightarrow \begin{pmatrix} T_{11}^* & T_{21}^* \\ T_{12}^* & T_{22}^* \end{pmatrix} \begin{pmatrix} 1 & v^*W \\ W^*v & I \end{pmatrix} + \begin{pmatrix} 1 & v^*W \\ W^*v & I \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}$$
$$= \begin{pmatrix} 0 & (v^*W)_t \\ (W^*v)_t & 0 \end{pmatrix}$$

$$\therefore \begin{cases} T_{11}^* + T_{21}^* W^* v + T_{11} + v^* W T_{21} = 0 \\ T_{22}^* + T_{12}^* v^* W + W^* v T_{12} + T_{22} = 0 \end{cases} \Rightarrow \begin{cases} T_{11}^* + T_{11} = -(v^* W T_{21} + T_{21}^* W^* v) \\ T_{22}^* + T_{22} = -(W^* v T_{12} + T_{12}^* v^* W) \end{cases}$$

and so we can uniquely determine the Hermitian part of T_{11} and T_{22}. For instance, we can choose T_{11} and T_{22} to be Hermitian. □

The above construction can be nicely generalized to obtain

Theorem 4.2.12 *Let $A \in C^k(\mathbb{R}, \mathbb{C}^{n\times n})$ have eigenvalues which can be clustered in p groups $\sigma_1, \dots, \sigma_p$, which remain disjoint for all t. Then, there exist $V \in \mathcal{C}^k$, invertible, such that* $V^{-1}AV = \begin{pmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_p \end{pmatrix}$, *where $\sigma(D_j) = \sigma_j$, $\forall\ t$, $j = 1, \dots, p$.*

□

Exercise 4.2.13 *Prove Theorem 4.2.12, and discuss different normalizations for V, as in Example 4.2.11.*

Hermitian case: simple eigenvalues

Let us specialize the construction we just presented to the case of a function $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$, which is Hermitian, $A = A^*$, $\forall t$, and with all eigenvalues simple. Now we seek a complete Schur decomposition of $A : U^*AU = \Lambda$ as smooth as A, with U unitary: $U^*U = I$.

Remark 4.2.14 *In this Hermitian case, left and right eigenvectors are of course equal.*

We proceed similarly to the general case.

Let U_0, Λ_0 be the Schur factors of an initial decomposition at $t_0 : A(t_0) = U_0\Lambda_0 U_0^*$. Differentiate the relation $AU = U\Lambda : \dot{A}U + A\dot{U} = \dot{U}\Lambda + U\dot{\Lambda}$, from which we get $U^*\dot{A}U + (U^*AU)U^*\dot{U} = U^*\dot{U}\Lambda + \dot{\Lambda}$. Now, let $H = U^*\dot{U}$, so that we have

$$\dot{\Lambda} = (U^*\dot{A}U) + \Lambda H - H\Lambda \ ,$$

that is

$$\begin{pmatrix} \dot{\lambda}_1 & & 0 \\ & \ddots & \\ 0 & & \dot{\lambda}_n \end{pmatrix} = (U^*\dot{A}U) + \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} H - H \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} . \tag{4.2.9}$$

Now, since $U^*U = I \Rightarrow U^*\dot{U} + \dot{U}^*U = 0 \Rightarrow U^*\dot{U} = -\dot{U}^*U \therefore H^* = -H$ that is H is skew-Hermitian.

So, using the 0-structure from the LHS of (4.2.9), we must have (for $i \neq j$)

$$0 = (U^*\dot{A}U)_{ij} + \lambda_i H_{ij} - H_{ij}\lambda_j \Rightarrow H_{ij} = \frac{(U^*\dot{A}U)_{ij}}{\lambda_j - \lambda_i} \ , \quad i \neq j,$$

which is well defined, since $\lambda_i \neq \lambda_j$, and notice that $H_{ji} = -H_{ij}$ since $\dot{A}$ is Hermitian. This allows us to determine H_{ij}, for $i \neq j$. The diagonal entries H_{jj} are *not* uniquely determined and a simple normalization choice is to set them to 0. So doing, we get the differential system defining a smooth Schur decomposition:

$$\begin{cases} \dot{\lambda}_j = (U^*\dot{A}U)_{jj} = u_j^*\dot{A}u_j, & j = 1, \dots, n, \\ \dot{U} = UH, H_{ij} = \frac{(U^*\dot{A}U)_{ij}}{\lambda_j - \lambda_i} & i \neq j \text{ and } H_{jj} = 0, j = 1 : n. \end{cases} \tag{4.2.10}$$

Remark 4.2.15 *Recalling Exercise 2.2.5, we know that any other smooth Schur decomposition must be of the type $V(t) = U(t)\Phi(t)$, where $\Phi(t) = diag(e^{i\phi_j(t)}, j = 1 : n)$ and ϕ_i are real valued smooth functions. This is exactly reflected in the freedom we have in choosing the entries H_{jj} above.*

4.2.2 Multiple eigenvalues

When we consider non-simple eigenvalues, things become considerably more complicated. As a general rule of thumb, the eigenspace behaves more singularly than the eigenvalues. This is true also in the Hermitian case, where there is also a distinct difference between the cases of A being an analytic function versus a (arbitrarily) smooth function. [Recall that a real analytic function is one which admits a convergent power (Taylor) series at any point.]

Example 4.2.16

(a) *This is a classical example due to Rellich of a Hermitian function, arbitrarily smooth but not analytic:*

$$A(t) = e^{-1/t^2} \begin{pmatrix} \cos\frac{2}{t} & \sin\frac{2}{t} \\ \sin\frac{2}{t} & -\cos\frac{2}{t} \end{pmatrix}, \quad A(0) = 0.$$

Note that $A \in C^\infty(\mathbb{R}, \mathbb{R}^{2\times 2})$ and $A = A^T$, $\forall\ t$. Since all derivatives vanish at 0, but A is not identically 0, the function is not analytic. The eigenvalues are $\lambda_{1,2}(t) = \pm e^{-1/t^2}$, $t \neq 0$ and $\lambda_{1,2}(0) = 0$, so the eigenvalues coalesce at the origin. Still, observe that $\lambda_{1,2} \in C^\infty$ as well. Now, the unit eigenvectors are $\begin{pmatrix} \cos\frac{1}{t} \\ \sin\frac{1}{t} \end{pmatrix}$ and $\begin{pmatrix} \sin\frac{1}{t} \\ -\cos\frac{1}{t} \end{pmatrix}$ for $t \neq 0$ and are in fact C^∞ on any interval not containing $t = 0$. However, they cannot *be continued as* continuous *functions at $t = 0$. Geometrically, the problem is that as $t \to 0$ each eigenvector points in any given direction infinitely often! Finally, let us notice that the function A is surely diagonalizable everywhere, just not continuously so.*

(b) *Even more dramatic is the situation in which the eigenspace changes dimension discontinuously. A nontrivial example is the following one from Kato. Take this (non-symmetric) function $A(t) = \begin{pmatrix} 0 & t & 0 \\ 0 & 0 & t \\ t & 0 & 1 \end{pmatrix}$, $t \in \mathbb{R}$. Obviously, $A(t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + t \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ is an analytic function. The eigenvalues satisfy $\lambda^3 - \lambda^2 - t^3 = 0$, and we claim (see Exercise 4.2.17) that the characteristic polynomial has 3 distinct roots, except when $t = 0$ or $t^3 = -4/27$. Now, for $t = 0$, $A(t)$ is diagonalizable even if it has a double eigenvalue $\lambda = 0$, but for $t = -\frac{1}{3}\sqrt[3]{4}$ the matrix is not diagonalizable.*

Exercise 4.2.17 *Verify the claims made in Example 4.2.16-(b) above. That is, that $\lambda^3 - \lambda^2 - t^3 = 0$ has three distinct roots for $t \neq 0$, $t \neq -\frac{1}{3}\sqrt[3]{4}$, and that A is not diagonalizable at $t = -\frac{1}{3}\sqrt[3]{4}$.*

To see what differences exist with respect to a simple eigenvalue, let us examine the situation of a double eigenvalue with one eigenvector only.

First, consider this example:

$$A(t) = \begin{pmatrix} 0 & 1 \\ t & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + t \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} , \tag{4.2.11}$$

so A is obviously an analytic function of $t \in \mathbb{R}$. The eigenvalues are $\pm\sqrt{t}$ and so they are:

$t < 0$,	$t = 0$,	$t > 0$,
purely imaginary,	double,	both real.

At $t = 0$, we have one eigenvector only and $\lambda(t)$ clearly is not differentiable at $t = 0$. What kind of perturbation result can we get for λ? The key observation is the following. From the characteristic polynomial $\lambda^2 - t = 0$, we know that the non-simple eigenvalue at $t = 0$ splits into two simple eigenvalues $\pm\sqrt{t}$.

Therefore, generalizing this idea, we next consider a sufficiently smooth matrix valued function A taking values in $\in \mathbb{R}^{n\times n}$ with a double real eigenvalue λ_0 at $t = 0$ and only one eigenvector. (If $\lambda_0 \in \mathbb{C}$, much the same contruction below still holds.) We anticipate a local expansion of the eigenpair as (*Newton–Puiseux series*)

$$\begin{cases} \lambda(t) = \lambda_0 + t^{1/2}\lambda_1 + t\lambda_2 + t^{3/2}\lambda_3 + \cdots \\ v(t) = v_0 + t^{1/2}v_1 + tv_2 + t^{3/2}v_3 + \cdots . \end{cases}$$

[Note that these become complex valued around $t = 0$, even if $\lambda_0 \in \mathbb{R}$.]

Now, if A is sufficiently smooth, then $A(t) = A_0 + tA_1 + t^2A_2 + \cdots$, where $A_0 = A(0)$, $A_1 = \dot{A}(0)$, $\ldots$, and so from $A(t)v(t) = v(t)\lambda(t)$, we get

$$\begin{aligned} &(A_0 + tA_1 + \cdots)(v_0 + t^{1/2}v_1 + tv_2 + t^{3/2}v_3 + \cdots) = \\ &(v_0 + t^{1/2}v_1 + tv_2 + t^{3/2}v_3 + \cdots)(\lambda_0 + t^{1/2}\lambda_1 + t\lambda_2 + \cdots) \end{aligned}$$

and equating same powers of t we get

$$\begin{aligned} A_0v_0 &= \lambda_0 v_0 \quad (O(1)) \\ A_0v_1 &= \lambda_0v_1 + \lambda_1v_0 \quad (O(t^{1/2})) \\ A_0v_2 + A_1v_0 &= \lambda_0v_2 + \lambda_1v_1 + \lambda_2v_0 \quad (O(t)) \\ \ldots &= \ldots \end{aligned} \tag{4.2.12}$$

To be able to solve this, we need to impose some normalization. The standard construction goes through consideration of the associated Jordan "chains" of length 2 at $t = 0$ (i.e., relatively to λ_0). We have a right Jordan chain:

$$A_0V = VJ_0, \quad V = [u_0, u_1] \to \begin{cases} A_0u_0 = \lambda_0u_0 \\ A_0u_1 = u_0 + \lambda_0u_1 \end{cases} , J_0 = \begin{bmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{bmatrix} ,$$

(in the case of our previous example (4.2.11), $\lambda_0 = 0$ and A_0 is the Jordan block $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$), as well as a left Jordan chain $A_0^T W = W J_0$, or $W^T A_0 = J_0^T W^T$, with $W = [w_0, w_1]$, which gives $\begin{cases} w_0^T A_0 = \lambda_0 w_0^T \\ w_1^T A_0 = w_0^T + \lambda_0 w_1^T \end{cases}$.

Of course, the right Jordan chain is not unique. But, once it is fixed, we will make the left Jordan chain unique according to the following normalization. The standard normalization conditions are:

$$w_0^T u_1 = 1, \quad w_1^T u_1 = 0 \; . \tag{4.2.13}$$

Now, with these, we have

$$w_0^T A_0 u_1 = w_0^T u_0 + \lambda_0 w_0^T u_1 \to \lambda_0 w_0^T u_1 = w_0^T u_0 + \lambda_0 w_0^T u_1 \to w_0^T u_0 = 0 \, ,$$

and

$$w_1^T A_0 u_1 = w_1^T u_0 + \lambda_0 w_1^T u_1 \to w_1^T A_0 u_1 = w_1^T u_0 \, ,$$

but also

$$w_1^T A_0 u_1 = w_0^T u_1 + \lambda_0 w_1^T u_1 \to w_1^T A_0 u_1 = w_0^T u_1 \, ,$$

from which we get $w_1^T u_0 = 1$. That is, the normalization (4.2.13) gives also

$$w_0^T u_0 = 0, \quad w_1^T u_0 = 1 \; . \tag{4.2.14}$$

Going back to (4.2.12), we will impose the standard choice $w_1^T v(t) = 1$, where w_1 is the left generalized eigenvector associated to the double eigenvalue. Now, because of (4.2.13) and (4.2.14), this gives (note that $v_0 = u_0$ in (4.2.13) and (4.2.14))

$$w_1^T v_0 = 1 \quad \text{and} \quad w_1^T v_i = 0, \quad i = 1, 2, \dots \; .$$

With these, we can now solve (4.2.12). Obviously, we have $\lambda_0 = w_1^T A_0 v_0$. Also, we formally get $\lambda_1 = w_1^T A_0 v_1$, though we do not have v_1 yet. To get an expression for v_1, we compare $\begin{cases} A_0 v_1 = \lambda_0 v_1 + \lambda_1 v_0 \\ A_0 u_1 = \lambda_0 u_1 + v_0 \end{cases} \Rightarrow \begin{cases} (A_0 - \lambda_0 \; I) v_1 = \lambda_1 v_0 \\ (A_0 - \lambda_0 \; I) u_1 = v_0 \end{cases} \Rightarrow v_1 = \lambda_1 u_1$.
Now we can use this expression in the third relation of (4.2.12):

$$(A_0 - \lambda_0 \; I) v_2 = -A_1 v_0 + \lambda_1 (\lambda_1 u_1) + \lambda_2 v_0 \; .$$

Obviously $A_0 - \lambda_0 \; I$ is singular, but $w_0^T (A_0 - \lambda_0 \; I) v_2 = 0 \; \therefore$ we must have

$$w_0^T (\lambda_1^2 u_1 + \lambda_2 v_0 - A_1 v_0) = 0 \to \lambda_1^2 = w_0^T A_1 v_0 \; ,$$

which will give us (in case $w_0^T A_1 v_0 \neq 0$) two values for the two branches of the eigenvalue and from these two values, we will get the associated two branches v_1 of the (bifurcating) eigenvectors.

This process can be continued, though it is a bit tedious. The main message we want to retain is that multiple eigenvalues typically lead to an expansion in fractional powers.

Remark 4.2.18 *We observe that the fact that there was only one eigenvector associated to λ_0 lead us to consider Jordan forms. But, the expansion of the eigenvalue in fractional powers did not directly depend on having just one eigenvector. Ultimately, this has to do with the behavior of roots of the characteristic polynomial: In general, a double root at $t = 0$ will split into a pair of roots whose locally leading term is $\pm t^{1/2}$. We also observe that if we had an eigenvalue with algebraic multiplicity m and geometric multiplicity 1, we should expect a local expansion of the eigenvalue and eigenvector in powers of $t^{1/m}$.*

Example 4.2.19 (Hermitian Analytic) *Here we will see a very important fact, whose explanation can be found in [6] in full details, though we will follow the more informal explanation from [7]. The end result is to show that* Eigenvalues of Hermitian analytic functions are analytic*.*

So, take a function $A \in \mathcal{C}^\omega(\mathbb{R}, \mathbb{C}^{n\times n})$, which is Hermitian, $A = A^*$, $\forall t$. [Recall that $\mathcal{C}^\omega$ means that A is real analytic.] So, we have $A(t) = \sum_{k=0}^\infty t^k A_k$, where $A_k^* = A_k$ for all k. Since A is analytic, also the characteristic polynomial $p(\lambda, t)$ will have coefficients which are analytic functions of t.

So, if $t = 0$ (say) is a value where the eigenvalues are simple, then the roots of the characteristic polynomial will also be analytic function of t. That is, the eigenvalues will have an expansion like $\lambda(t) = \sum_{k=0}^\infty t^k c_k$. On the other hand, if –say at $t = 0$– we have a multiple eigenvalue of algebraic multiplicity m, then that root of p will have an algebraic singularity and will be expressable as a Newton-Puiseux series around $t = 0$: $\lambda(t) = \sum_{k=0}^\infty t^{k/m} b_k$.

But*, for all $t \in \mathbb{R}$, the eigenvalues of $A(t)$ must be real since A is Hermitian, and fractional powers of t become complex valued for t near 0. This means that in the Newton–Puiseux series we can only have integer powers. That is, $\lambda(t)$ is an analytic function of t.*

Now, if A is real valued, and symmetric, then a similar argument (since the eigenvectors must remain real valued) tells us that the eigenvectors also admit a regular epxansion in powers of t. For the general Hermitian case, in [6] the same conclusion about the eigenvectors is also given, so that altogether one has:

Theorem 4.2.20 *An analytic Hermitian function admits an analytic eigendecomposition: $A(t)U(t) = U(t)\Lambda(t)$, for all $t \in \mathbb{R}$, with U unitary and analytic and Λ diagonal, and analytic.* $\square$

4.2.3 Multiple eigenvalues, several parameters: Hermitian case

Here we consider the symmetric (Hermitian) eigenproblem when A depends on two parameters.

To begin with, we need to realize that things can go very awry.

Example 4.2.21 *Let $(x, y) \in \mathbb{R}^2$, and consider the function of two parameters A below:*

$$A(x, y) = \begin{bmatrix} x & y \\ y & -x \end{bmatrix} .$$

Obviously, the function A is analytic in x and y (it is linear). However, the eigenvalues are $\pm\sqrt{x^2 + y^2}$ which are not even differentiable at $(0, 0)$. The problem is the lack of global differentiability at the origin, where both eigenvalues are 0. We notice that viewing $A(x, y)$ as a function of one parameter (holding the other frozen), renders analytic eigenvalues (see Example 4.2.19 and Theorem 4.2.20).

The above example notwithstanding, we now try to understand when/how a double eigenvalue of a (smooth, even analytic) Hermitian function of two parameters persist as a double eigenvalue.

So, we are given a symmetric matrix valued function A of two parameters (x, y) such that at the point $\xi_0 \equiv (x_0, y_0)$ A has a double eigenvalue λ_0. We will want to understand when/how this eigenvalue persists –as double eigenvalue– along a curve passing through ξ_0. We will assume that A depends analytically on x and y.

Let u_1, u_2, be orthonormal eigenvectors associated to λ_0 (for $A_0 \equiv A(\xi_0)$).

As we know, u_1 and u_2 are not unique. The degree of nonuniqueness is given by all possibilities: $[u_1, u_2]R$, where R is a $(2, 2)$ orthogonal matrix. However, for a given pair $[u_1, u_2]$, the left eigenvectors $[v_1, v_2]$ of A_0 such that $v_i^T u_j = \delta_{ij}$ are uniquely determined, and naturally are given by $[u_1, u_2]$.

Take a planar curve $\gamma(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}$, depending analytically on t, such that $\gamma(0) = \xi_0$. Let $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \frac{d\gamma}{dt}|_{t=0}$. Consider the restriction of A to this curve, and call

$A(t), t \geq 0$, the analytic function of t giving this restriction. So, we can write (where all matrices $A_1, A_x, A_y, \dots$, are evaluated at $t = 0$, that is at ξ_0):

$$A(t) = A_0 + tA_1 + \cdots \ , \quad A_1 = A_x d_1 + A_y d_2 \ , \cdots \ , \tag{4.2.15}$$

and all matrices are symmetric. Because of analyticity of $A(t)$, we have (see Theorem 4.2.20) that the eigenvalues and eigenvectors of A are analytic function of t. In other words, along the curve γ we have:

$$\lambda(t) = \lambda_0 + t\lambda_1 + \cdots \ , \quad u(t) = v_0 + tv_1 + \cdots \ . \tag{4.2.16}$$

Observe that v_0 is the limit as $t \to 0$ of $u(t)$ and it is not known ahead of time. All we can say is that v_0 will need to be a combination of u_1 and u_2.

From the eigenvalue relation $Au = \lambda u$, using the expansions (4.2.15-4.2.16), and equating equal powers of t, we must have:

$$\begin{aligned} &A_0 v_0 = \lambda_0 v_0 \implies v_0 = c_1 u_1 + c_2 u_2 \ , \\ &A_0 v_1 + A_1 v_0 = \lambda_0 v_1 + \lambda_1 v_0 \ . \end{aligned} \tag{4.2.17}$$

Multiplying the second relation by u_1^T on the left, and using the eigenvalue relation $A_0 u_1 = \lambda_0 u_1$ and the form of v_0, we get

$$(u_1^T A_1 u_1)c_1 + (u_1^T A_1 u_2)c_2 = \lambda_1 c_1 \, .$$

Similarly, multiplying the second relation in (4.2.17) by u_2^T on the left we get

$$(u_2^T A_1 u_1)c_1 + (u_2^T A_1 u_2)c_2 = \lambda_1 c_2 \, .$$

Therefore, we must have

$$M \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \lambda_1 \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \ , \quad M = \begin{bmatrix} u_1^T A_1 u_1 & u_1^T A_1 u_2 \\ u_2^T A_1 u_1 & u_2^T A_1 u_2 \end{bmatrix} \ . \tag{4.2.18}$$

So, for λ_0 to persist as double eigenvalue in some direction d, we must have that λ_1 is a double eigenvalue of M for that d. Since M is symmetric, the requirement that λ_1 is a double eigenvalue of M means that $M_{11} = M_{22}$ and $M_{12} = 0$. Recalling that $A_1 = A_x d_1 + A_y d_2$, this translates into the requirement

$$\begin{cases} (u_1^T A_x u_1)d_1 + (u_1^T A_y u_1)d_2 = (u_2^T A_x u_2)d_1 + (u_2^T A_y u_2)d_2 \\ \qquad\qquad (u_1^T A_x u_2)d_1 + (u_1^T A_y u_2)d_2 = 0 \end{cases} ,$$

which can be further rewritten as the system

$$N \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = 0 \ , \quad N = \begin{bmatrix} (u_1^T A_x u_1) - (u_2^T A_x u_2) & (u_1^T A_y u_1) - (u_2^T A_y u_2) \\ u_1^T A_x u_2 & u_1^T A_y u_2 \end{bmatrix} . \tag{4.2.19}$$

So, persistence as double eigenvalue requires non-trivial solutions of (4.2.19), that is $\det(N) = 0$. To further elucidate what this means, let

$$B = \begin{bmatrix} u_1^T A_x u_1 & u_1^T A_x u_2 \\ u_1^T A_x u_2 & u_2^T A_x u_2 \end{bmatrix} \ , \quad C = \begin{bmatrix} u_1^T A_y u_1 & u_1^T A_y u_2 \\ u_1^T A_y u_2 & u_2^T A_y u_2 \end{bmatrix} \ ,$$

so that $N = \begin{bmatrix} b_{11} - b_{22} & c_{11} - c_{22} \\ b_{12} & c_{12} \end{bmatrix}$ and so $\det(N) = 0$ is the same as the requirement $b_{11}c_{12} - b_{22}c_{12} - b_{12}c_{11} + b_{12}c_{22} = 0$. But this latter requirement is equivalent to the requirement $BC = CB$, as it is easily verified.

• **Conclusion**. λ_0 persists as a double eigenvalue –in some direction $d = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$– only if $BC = CB$, with $B = U^T A_x U$, $C = U^T A_y U$, $U = [u_1, u_2]$.

Some remarks are in order.

(a) We expect that, if N is singular, it is of rank 1, that is there is only one curve along which the eigenvalue through λ_0 stays double. If N has rank 0, then the double eigenvalue would persist along any direction. This requires that $B = bI$, $C = cI$.

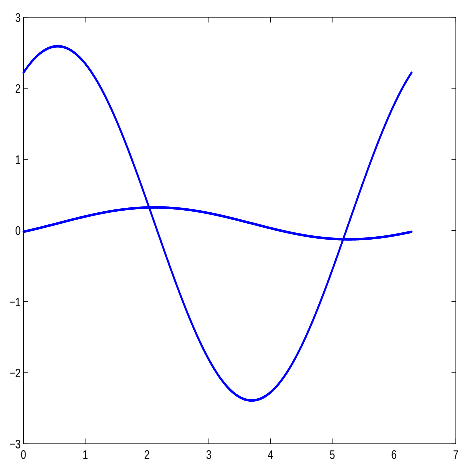
(b) In case λ_0 persists as a double eigenvalue, then λ_1 is a double eigenvalue of M in (4.2.18). Therefore, $u_1^T A_1 u_1 = u_2^T A_1 u_2$ and $u_1^T A_1 u_2 = 0$, $\lambda_1 = u_1^T A_1 u_1$, and $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ is any unit vector. This means that –at first order at least– the limiting value v_0 of $v(t)$ is not determined: any unit vector in the plane spanned by u_1, u_2, would be a possible limit. This means that, once we fixed u_1 and u_2, we will demand that v_0 be the same as u_1 or u_2. This should be contrasted to the case when λ_0 does not persist as double eigenvalue. Then, there are two distinct eigenvalues λ_1 of M, and two independent associated eigenvectors: Each of these would give a (unique, up to sign) pair of unit vectors $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ and a well defined limit, in general distinct from u_1 or u_2. In other words, if the double eigenvalue splits, then there are well defined eigenvectors paths. If the eigenvalue stays double, any eigenvector path in the plane spanned by u_1, u_2, could be retrieved.

(c) It is worth pointing out that the above construction and conclusions are independent of the choice of eigenvectors u_1, u_2, done at the beginning. In other words, replacing U with UR, R any $(2,2)$ orthogonal matrix, does not change anything.

Example 4.2.22 *Consider the following function*

$$A = \lambda_0 I + xB + yC \ , \quad B = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} \ , \ C = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \ .$$

At $(0,0)$ the eigenvalue λ_0 is double. Here, $BC = CB$, and neither B or C is diagonal. In this case (A is linear in x and y), we have an entire line along which the eigenvalue remains double. It is a simple computation to verify that the direction d of this line is $d = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$. In the figure we show the two eigenvalues of $A(x,y)$ computed along the circle $x = 1/2\cos(\theta),\ y = 1/2\sin(\theta)$, $\theta \in [0, 2\pi]$, clearly showing the double eigenvalue at the two values where the line intersects the circle ($\theta :\ \tan(\theta) = -2$).



Bibliography

[1] A. Berman and R. Plemmons, Nonnegative matrices in the mathematical sciences.

[2] I.N. Herstein, Topics in Algebra.

[3] G. Hurlbert, *A short proof of the Birkhoff-von Neumann Theorem* (2008).

[4] R. Horn and C. Johnson, Matrix Analysis.

[5] R. Horn and C. Johnson, Topics in Matrix Analysis.

[6] T. Kato, Perturbation Theory for Linear Operators.

[7] P. Lax, Linear Algebra and its Applications.

[8] A.P. Seyranian and A.A. Mailybaev, Multiparameter Stability Theory with Mechanical Applications.

[9] R.S. Varga, Matrix Iterative Analysis,

Chapter 5

Homework Assignments

Homework 1 Problems.

(1) [*20 points.*] Given the block upper triangular matrix $R \in \mathbb{F}^{n\times n}$ of the form

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1p} \\ 0 & R_{22} & \cdots & R_{2p} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & R_{pp} \end{bmatrix},$$

where each block $R_{jj} \in \mathbb{F}^{n_j \times n_j}$ is upper triangular, with constant diagonal given by λ_j, $j = 1, \dots, p$, $n_1 + \cdots + n_p = n$, and $\lambda_j \neq \lambda_k$ for $j \neq k$. Show that there exist a similarity transformation matrix $V \in \mathbb{F}^{n\times n}$ such that $V^{-1}RV = \operatorname{diag}(R_{jj},\ j = 1, \dots, p)$.

(2) [*20 points.*] Suppose that $A \in \mathbb{F}^{n\times n}$ has Jordan form $\operatorname{diag}(J_{n_j}(\lambda_j),\ j = 1, \dots, k)$, $n_1 + \cdots + n_k = n$. Assume that the field $\mathbb{F}$ has characteristic 0.

(a) If A is nonsingular, show that A^2 has Jordan form $\operatorname{diag}(J_{n_j}(\lambda_j^2),\ j = 1, \dots, k)$.

(b) Is the result of part (a) true if A is singular? Justify your answer.

(c) Is the result of part (a) true for all powers A^m, $m \geq 2$? Justify your answer.

(3) [*20 points.*] For parts (a) and (b), let $R \in \mathbb{F}^{n\times n}$ be upper triangular.

(a) Show that the eigenvalues of R are the diagonal entries of R, and only these.

(b) Suppose that $R = \lambda I + N$, where N is the strictly upper triangular part of R, and $\lambda \neq 0$. Give an explicit formula for R^{-1}.

(c) Now, let $A \in \mathbb{F}^{n\times n}$ be nilpotent of index k: $A^k = 0$, but $A^{k-1} \neq 0$. Show that $I + A$ is invertible, and give a formula for $(I + A)^{-1}$.

(4) [*10 points.*] Let $T \in Hom(V, V)$ where V is a vector space over $\mathbb{F}$ of dimension n. Suppose $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ are distinct eigenvalues of T with associated eigenvectors $v_1, \ldots, v_k$. Show that $v_1, \ldots, v_k$, are linearly independent.

(5) [*15 points.*] Let $T \in Hom(V, V)$ as in problem (4) and consider its minimal polynomial.

(a) Show that the minimal polynomial is unique.

(b) Show that T is invertible if and only if the constant term of the minimal polynomial is not 0.

(6) [*20 points.*] Suppose $A, B \in \mathbb{F}^{n\times n}$ are diagonalizable (via matrices in $\mathbb{F}^{n\times n}$). Show that they can be simultaneously diagonalized, that is there is $V \in \mathbb{F}^{n\times n}$ such that $V^{-1}AV$ and $V^{-1}BV$ are both diagonal, if and only if $AB = BA$.

(7) [*10 points.*] Suppose that $A \in \mathbb{C}^{n\times n}$ is such that $A^3 = I$. What are the possible Jordan forms of A?

(BONUS) [*25 points.*] (*This problem is IMPORTANT*).

Suppose that $A, B \in \mathbb{F}^{n\times n}$ are two different matrix representation of the same linear transformation $T \in Hom(V, V)$. Show that A and B are similar matrices. That is, that there is an invertible matrix $S \in \mathbb{F}^{n\times n}$ such that $A = S^{-1}BS$.

Homework 2 Problems.

(1) [*10 points.*] Let $A \in \mathbb{C}^{n\times n}$ and let A^T denote its transpose. Use the Jordan form of A to show that A and A^T are similar.

[*Observation: This result, and technique of proof, remain true for $A \in \mathbb{F}^{n\times n}$.*]

(2) [*10 points.*] Let $A \in \mathbb{C}^{n\times n}$. Let J be the Jordan form of A: $J = \mathrm{diag}(J_k(\lambda_k, 1),\ k = n_1, \dots, n_p)$, with $n_1 \geq \cdots \geq n_p$, $n_1 + \cdots + n_p = n$. Here, we have used the notation $J_k(\lambda_k, 1)$ for the standard Jordan blocks to signify that the eigenvalue is λ_k and the super-diagonal is made of 1's.

Now, let $\varepsilon > 0$ be a given (small) value. Show that J, and hence A, is similar to the matrix $J_\varepsilon = \mathrm{diag}(J_k(\lambda_k, \varepsilon),\ k = n_1, \dots, n_p)$, that is the 1's in the super-diagonal entries have been replaced by ε's.

(3) [*10 points.*] Let $A \in \mathbb{F}^{n\times n}$ be invertible, and let $B \in \mathbb{F}^{n\times n}$ be such that

$$\|A - B\| \ < \ 1/\|A^{-1}\| \ .$$

Show that B is invertible. [Here, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and the norm is the 2-norm.]

[*Observation: This very useful result also holds for an infinite dimensional version.*]

(4) [*25 points.*] [On Spectral Radius.] Let $A \in \mathbb{C}^{n\times n}$. Define the *spectral radius* of A as the quantity

$$\rho(A) := \max_{1\leq i\leq n} |\lambda_i| \,,$$

where λ_i's are A's eigenvalues. You have to show that:

$$\rho(A) \ = \ \lim_{k\to\infty} \|A^k\|^{1/k} \,.$$

The norm is the usual 2-norm. (*Hint*: The construction of Exercise 2 is useful.)

(5) [*10 points.*] [On Pfaffian.] Let $A \in \mathbb{R}^{n\times n}$ be anti-symmetric ($A^T = -A$), and suppose that n is an even number. You are asked to compute the determinant of A in the special case of A tridiagonal. Your formula must be expressed in terms of the entries of A.

[*Observation: It is known –though we have not seen it– that every anti-symmetric matrix is orthogonally similar to a tridiagonal one.*]

(6) [*20 points.*] (This is about uniqueness of reduction to Schur form.) Do exercises (1) and (2) p.23 of the notes I gave you.

(7) [*15 points.*] Do Exercise (1), bottom of p.30 of the notes.

Homework 3 Problems.

Notation. In the exercises below, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, indifferently.

(1) [*5 points each.*] [A collection of counterexamples.] By giving counterexamples, show that the following statements are generally false.

(a) Let $\Sigma(A)$ denote the collection of singular values of a matrix A, and let $B, C \in \mathbb{F}^{m\times n}, m \geq n$. Then: $\Sigma(CB^T) = \Sigma(B^TC) \cup \underbrace{\{0, \dots, 0\}}_{m-n}$.

(b) The product of two Hermitian matrices in $\mathbb{F}^{n\times n}$ is Hermitian.

(c) The symmetrized product of two positive definite matrices in $\mathbb{F}^{n\times n}$ is positive. [Hint: See Problem (2) below.]

(d) Let A and B be Hermitian and positive definite, and suppose $0 \prec B \prec A$. Then $0 \prec B^2 \prec A^2$.

(2) [*20 points.*] Let $A \in \mathbb{F}^{n\times n}$ be positive definite, and $B \in \mathbb{F}^{n\times n}$ be Hermitian. Show that AB is diagonalizable and it has all real eigenvalues, with the same number of positive, negative and zero eigenvalues as B. In particular, conclude that if B is positive definite, then AB is positive definite if and only if it is Hermitian.

(3) [*25 points.*] Let $A, B \in \mathbb{F}^{n\times n}$ be Hermitian and positive definite and assume that $S = AB$ is also positive definite.

(i) Show that for the unique positive definite square roots of A, B, S, we have

$$\sqrt{S} = \sqrt{AB} = \sqrt{A}\sqrt{B}\,.$$

(ii) Consider the positive definite matrix valued function $P(t) = B + tA$, $t \geq 0$. Show that the unique positive definite square root $\sqrt{P(t)}$ is a smooth function of $t \geq 0$. [Recall that a matrix valued function is smooth if its entries are.]

(4) [*20 points.*] [Projections from SVD.] Use the SVD of a matrix $A \in \mathbb{R}^{m\times n}$ to find orthonormal bases for $\mathcal{R}(A)$, $\mathcal{N}(A)$, $\mathcal{R}(A^T)$, $\mathcal{N}(A^T)$ and to express orthogonal projection matrices onto these subspaces.

(5) [*15 points.*] [On commutators.] Let $A, B \in \mathbb{F}^{n\times n}$. Define the *commutator* of A, B as

$$[A, B] \;=\; AB - BA\,.$$

It is easy to observe that if a matrix X is the commutator of two matrices A, B, then $\mathrm{tr}(X) = 0$. You need to show that:
"Any given matrix X whose diagonal entries are 0 can be represented as the commutator of two matrices A, B." [*Hint*: You are free to choose A and B.]
Bonus [*10 points.*] Show that a (2×2) matrix whose trace is 0 is similar to one whose diagonal entries are 0. This result is actually true for $(n \times n)$ matrices, but it is enough for you to do it for (2×2) matrices.
[Note that by putting together this result, and the previous one, we have obtained that: "*A matrix X is the commutator of A, B, if and only if* $\mathrm{tr}(X) = 0$."]

Homework 4 Problems.

(1) [*10 points each.*]

(a) [Lecture notes, p. 56] Prove the AGM inequality:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad x_i \geq 0\,. \tag{5.0.1}$$

(b) Prove that there is equality in (5.0.1) if and only if all x_i's are equal.

(c) [Lecture notes, p. 59] Let A, B be positive definite. Show $\det(A+B) \geq \det(A) + \det(B)$.

(2) [*20 points.*] Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, and $B \in \mathbb{C}^{n\times n}$ be positive definite. Consider the minization problem:

$$\min_{x\neq 0} \frac{x^*Ax}{x^*Bx}\,. \tag{5.0.2}$$

(i) Show that there is a (nonzero) vector $v \in \mathbb{C}^n$ which gives the minimum value in (5.0.2), call if μ. Show that the pair (v, μ) solves the equation $Av = \mu Bv$. [Generalized Eigenproblem.]

(ii) Show that the constrained minimization problem

$$\min_{\substack{x\neq 0\\ x\perp Bv}} \frac{x^*Ax}{x^*Bx}\,,$$

where v is the vector from point (i), has a solution ν, with associated vector w, which also solves the generalized eigenproblem: $Aw = \nu Bw$.

(3) [*15 points.*] [Lecture notes, p. 65] Let $A, B \in \mathbb{C}^{m\times n}$, $m \geq n$, and let $\sigma_i(A+B)$, $\sigma_i(A)$, $\sigma_i(B)$ be the ordered singular values: $\sigma_1 \geq \cdots \geq \sigma_n$. Show

$$\sigma_{i+j-1}(A+B) \leq \sigma_i(A) + \sigma_j(B), \quad 1 \leq i, j \leq n, \quad i+j \leq n+1.$$

(4) We have seen in class that if $A \in \mathbb{C}^{n\times n}$ is such that its symmetric part is negative definite,

$$A + A^* \prec 0 \ ,$$

then the eigenvalues of A have negative real part. You have to show the following generalization.

(a) [*15 points.*] If $B \in \mathbb{C}^{n\times n}$ is positive definite, and we have

$$BA + A^*B \prec 0 \ ,$$

then the eigenvalues of A have negative real part.

This part (a) is only half of a beautiful Theorem of Lyapunov. The remaining half states that

If A has eigenvalues with negative real part, then there exists a Hermitian, positive definite matrix B, such that $BA + A^*B \prec 0$.

(b) [*20 points.*] Assume that e^{At} converges to 0 (it does), take

$$B = \int_0^\infty e^{A^*t} e^{At} dt \, ,$$

and show that B is Hermitian positive definite and $BA + A^*B \prec 0$. In other words, you have proved Lyapunov theorem.

(c) [**Bonus**] [*10 points.*] Show that e^{At} converges to 0 as $t \to \infty$.

Homework 5 Problems.

(1) Recall that for any square matrix, say $A \in \mathbb{C}^{n\times n}$, the matrix exponential is given by $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$.

(a) [*10 points.*] Give an explicit counterexample to show that –in general– $e^A e^{A^*} \neq e^{A+A^*}$.

(b) [*5 points.*] If A is a normal matrix, show that $e^A e^{A^*} = e^{A+A^*}$.

(2) [*20 points.*] [Lecture notes, p. 84] Let $A, E \in \mathbb{R}^{2\times 2}$ of the form $A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ and $E = \begin{bmatrix} \eta & \varepsilon \\ 0 & 0 \end{bmatrix}$. We know that if $\mu \in \sigma(A+E)$, then $\exists\ \lambda \in \sigma(A) : |\lambda - \mu| \leq \|E\|_2$. Find $\|E\|_2$. Also, discuss the behavior of the eigenvectors of $A+E$ and contrast this to the eigenvectors of A for different ratios of η and ε.

(3) [*15 points.*] [Lecture notes, p. 80] Let $A \in \mathbb{R}^{n\times n}$, $A \geq 0$. Show that if A is doubly stochastic and reducible, then there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

where $A_{11} \in \mathbb{R}^{n_1\times n_1}$ and $A_{22} \in \mathbb{R}^{n_2\times n_2}$, $n_1 \geq 1, n_2 \geq 1, n_1 + n_2 = n$, are doubly stochastic.

(4) [*15 points.*] Let $A \in \mathbb{R}^{n\times n}$, $A > 0$, and A be (column) stochastic. Let $x \in \mathbb{R}^n$, $x \geq 0$, and $x \neq 0$. Show that

$$\lim_{m\to\infty} A^m x \;=\; cz\,,$$

where c is some positive constant, and z is the eigenvector associated to the spectral radius of A.

(5) This is an exercise about a special class of matrices, the so–called *symplectic* matrices, which are of key importance in Hamiltonian mechanics. These are matrices of even size, defined as follows:
"A matrix $S \in \mathbb{R}^{2n\times 2n}$ is called symplectic (or J-orthogonal) if

$$S^T JS \;=\; J\,, \qquad \text{where} \qquad J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}.$$

The matrix J is called symplectic identity and it is easy to see that it satisfies $J^T = -J = J^{-1}$."

Below, $S \in \mathbb{R}^{2n\times 2n}$ is symplectic. The exercises are arranged in such a way that you will need the previous one to solve the ones after it. If you cannot solve one of them, you may still use the result to solve the exercises that come after it; however, you cannot use the later results to show one of the preceeding ones.

(a) [*5 points.*] Show that S is invertible and S^{-1} is similar to S.

(b) [*5 points.*] Show that if λ is an eigenvalue of S, then also $1/\lambda$ is. (Note that we could have $\lambda = 1/\lambda$).

(c) [*10 points.*] Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Show that the matrix S defined by $S = e^{JA}$ is symplectic.
[Incidentally, matrices of the form JA with A symmetric are called *Hamiltonian*.]

(d) [*10 points.*] Show that if $\lambda = -1$ is an eigenvalue of S, then it cannot be a simple eigenvalue of S.
[The result is true also relatively to the eigenvalue $\lambda = 1$, but it is enough for you to show it for $\lambda = -1$. Moreover, it is also true that $\lambda = \pm 1$ cannot be eigenvalues of odd algebraic multiplicity of S; you do not have to show this last fact, though you can assume it for part (e).]

(e) [*5 points.*] Show that $\det(S) = 1$.

(f) [**Bonus**. *10 points.*] Show that if $\lambda = -1$ is an eigenvalue of S, then it cannot be an eigenvalue of odd multiplicity of S.

Chapter 6

Exams

Math 6112. Fall 2010. Exam 1.
October 6, 2010.

Problems.

(1) Consider a matrix $A \in \mathbb{R}^{n\times n}$ and its transpose A^T. Let λ be a (real) simple eigenvalue of A, hence also of A^T. Let v be an eigenvector of A associated to λ and let w be an eigenvector of A^T associated to λ. **Show** that $v^T w \neq 0$.
[*Notation*: The eigenvalue being simple means that it has algebraic multiplicity 1.]

(2) Let $A \in \mathbb{C}^{2\times 2}$ be a Hermitian positive definite matrix. Show that it admits a unique factorization $A = LL^*$, where L is lower triangular with positive (real) diagonal.

(3) Let $A \in \mathbb{C}^{n\times n}$ and let $p(z)$ be a polynomial. We know that if λ is an eigenvalue of A, then $p(\lambda)$ is an eigenvalue of $p(A)$. You need to **show** that: "*Every eigenvalue of $p(A)$ is of the form $p(\lambda)$, where λ is an eigenvalue of A.*"

(4) Show that a matrix $A \in \mathbb{C}^{n\times n}$ is normal if and only if it satisfies $A = A^*U$, where U is unitary.

(5) Let $A \in \mathbb{C}^{n\times n}$ be Hermitian negative definite. Consider the matrix

$$B = (I + A)(I - A)^{-1}\,.$$

Show that B is well defined and determine where are the eigenvalues of B in the complex plane.
[*Hint*: Use the Schur form of A to obtain a Schur form for B.]

(6) Let $A \in \mathbb{C}^{m\times n}$, $m \geq n$. Show that the nonzero eigenvalues of AA^* and of A^*A are the same, counted with their multiplicities.

Math 6112. Fall 2010. Exam 2.
November 29, 2010.

Problems.

(1) Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian.

(a) (*7 points*) Let $S \in \mathbb{C}^{n\times n}$ be invertible, Show that:

$$A \succeq B \Leftrightarrow S^*AS \succeq S^*BS\,.$$

(b) (*3 points*) Show that $A \succeq I \Leftrightarrow$ all A's eigenvalues are ≥ 1.

(2) (*10 points*) Let $A, B \in \mathbb{C}^{n\times n}$ be Hermitian. Show that if $A \succeq B \succ 0$, then all eigenvalues of $A^{-1}B$ are in $(0, 1]$.

(3) (*10 points*) Let $A \in \mathbb{R}^{n\times n}$, $A > 0$. Show that if $Ax = \lambda x$, with $x \in \mathbb{R}^n$, $x > 0$, then $\lambda = \rho(A)$.

(4) Let $A \in \mathbb{R}^{n\times n}$.

(a) (*2 points*) Show by example that we may have $\rho(I + A) < 1 + \rho(A)$.

(b) (*8 points*) Show that if $A \geq 0 \Rightarrow \rho(I + A) = 1 + \rho(A)$.

(5) (*10 points*) Suppose $\lambda, \mu \in \sigma(A)$, $\lambda \neq \mu$. Let $v, w : Av = \lambda v$ and $A^*w = \bar{\mu}w$ (that is, w is left eigenvector corresponding to μ). Show that $v^*w = 0$.

(6) (*10 points*) Let $A \in \mathbb{C}^{n\times n}$ be Hermitian ($n > 2$). Prove that if B is a $((n-2)\times(n-2))$ principal submatrix of A, and the eigenvalues of A and B are ordered in increasing fashion, we have

$$\lambda_k(A) \leq \lambda_k(B) \leq \lambda_{k+2}(A), \quad 1 \leq k \leq n-1\ .$$