Chapter 7
Norms

In this chapter, the field K is always $\mathbb{R}$ or $\mathbb{C}$ and E denotes K^n. The scalar (if $K = \mathbb{R}$) or Hermitian (if $K = \mathbb{C}$) product on E is denoted by $\langle x, y \rangle := \sum_j \bar{x}_j y_j$.

Definition 7.1 *If $A \in \mathbf{M}_n(K)$, the* spectral radius *of A, denoted by $\rho(A)$, is the largest modulus of the eigenvalues of A:*

$$\rho(A) = \max\{|\lambda|; \lambda \in \mathrm{Sp}(A)\}.$$

When $K = \mathbb{R}$, this takes into account the complex eigenvalues when computing $\rho(A)$.

7.1 A Brief Review

7.1.1 The $\ell^{\mathbf{p}}$ Norms

The vector space E is endowed with various norms, pairwise equivalent because E has finite dimension (Proposition 7.3 below). Among these, the most used norms are the l^p norms:

$$\|x\|_p = \left(\sum_j |x_j|^p\right)^{1/p}, \quad \|x\|_\infty = \max_j |x_j|.$$

Proposition 7.1 *For $1 \le p \le \infty$, the map $x \mapsto \|x\|_p$ is a norm on E. In particular, one has* Minkowski's inequality

$$\|x + y\|_p \le \|x\|_p + \|y\|_p. \tag{7.1}$$

Furthermore, one has Hölder's inequality

D. Serre, *Matrices,* Graduate Texts in Mathematics 216,
DOI 10.1007/978-1-4419-7683-3_7,

$$|\langle x,y\rangle| \le \|x\|_p \|y\|_{p'}, \quad \frac{1}{p}+\frac{1}{p'}=1. \tag{7.2}$$

The numbers p, p' are called *conjugate exponents*.

Proof. Everything is obvious except perhaps the Hölder and Minkowski inequalities. When $p=1$ or $p=\infty$, these inequalities are trivial. We thus assume that $1<p<\infty$.

Let us begin with (7.2). If x or y is null, it is obvious. Indeed, one can even assume, by decreasing the value of n, that none of the x_j, y_js is null. Likewise, because $|\langle x,y\rangle| \le \sum_j |x_j||y_j|$, one can also assume that the x_j, y_j are real and positive. Dividing by $\|x\|_p$ and by $\|y\|_{p'}$, one may restrict attention to the case where $\|x\|_p = \|y\|_{p'} = 1$. Hence, $x_j, y_j \in (0,1]$ for every j. Let us define

$$a_j = p\log x_j, \quad b_j = p'\log y_j.$$

Because the exponential function is convex,

$$\mathrm{e}^{a_j/p+b_j/p'} \le \frac{1}{p}\mathrm{e}^{a_j} + \frac{1}{p'}\mathrm{e}^{b_j};$$

that is,

$$x_j y_j \le \frac{1}{p}x_j^p + \frac{1}{p'}y_j^{p'}.$$

Summing over j, we obtain

$$\langle x,y\rangle \le \frac{1}{p}\|x\|_p^p + \frac{1}{p'}\|y\|_{p'}^{p'} = \frac{1}{p}+\frac{1}{p'} = 1,$$

which proves (7.2).

We now turn to (7.1). First, we have

$$\|x+y\|_p^p = \sum_k |x_k+y_k|^p \le \sum_k |x_k||x_k+y_k|^{p-1} + \sum_k |y_k||x_k+y_k|^{p-1}.$$

Let us apply Hölder's inequality to each of the two terms of the right-hand side. For example,

$$\sum_k |x_k||x_k+y_k|^{p-1} \le \|x\|_p \left(\sum_k |x_k+y_k|^{(p-1)p'}\right)^{1/p'},$$

which amounts to

$$\sum_k |x_k||x_k+y_k|^{p-1} \le \|x\|_p \|x+y\|_p^{p-1}.$$

Finally,

$$\|x+y\|_p^p \le (\|x\|_p + \|y\|_p)\|x+y\|_p^{p-1},$$

which gives (7.1). □

For $p = 2$, the norm $\|\cdot\|_2$ is given by an Hermitian form and thus satisfies the Cauchy–Schwarz inequality:

$$|\langle x, y\rangle| \leq \|x\|_2 \|y\|_2.$$

This is a particular case of Hölder's inequality.

Proposition 7.2 *For conjugate exponents p, p', one has*

$$\|x\|_p = \sup_{y \neq 0} \frac{\Re\langle x, y\rangle}{\|y\|_{p'}} = \sup_{y \neq 0} \frac{|\langle x, y\rangle|}{\|y\|_{p'}}.$$

Proof. The inequality $\geq$ is a consequence of Hölder's. The reverse inequality is obtained by taking $y_j = \bar{x}_j |x_j|^{p-2}$ if $p < \infty$. If $p = \infty$, choose $y_j = \bar{x}_j$ for an index j such that $|x_j| = \|x\|_\infty$. For $k \neq j$, take $y_k = 0$. □

7.1.2 Equivalent Norms

Definition 7.2 *Two norms N and N' on a (real or complex) vector space are said to be equivalent if there exist two numbers $c, c' \in \mathbb{R}$ such that*

$$N \leq cN', \quad N' \leq c'N.$$

The equivalence between norms is obviously an equivalence relation, as its name implies. As announced above, we have the following result.

Proposition 7.3 *All norms on $E = K^n$ are equivalent. For example,*

$$\|x\|_\infty \leq \|x\|_p \leq n^{1/p}\|x\|_\infty.$$

Proof. It is sufficient to show that every norm is equivalent to $\|\cdot\|_1$.

Let N be a norm on E. If $x \in E$, the triangle inequality gives

$$N(x) \leq \sum_i |x_i| N(\mathbf{e}^i),$$

where $(\mathbf{e}^1, \ldots, \mathbf{e}^n)$ is the canonical basis. One thus has $N \leq c\|\cdot\|_1$ for $c := \max_i N(\mathbf{e}^i)$. Observe that this first inequality expresses the fact that N is Lipschitz (hence continuous) on the metric space $X = (E, \|\cdot\|_1)$.

For the reverse inequality, we reduce *ad absurdum*. Let us assume that the supremum of $\|x\|_1/N(x)$ is infinite for $x \neq 0$. By homogeneity, there would then exist a sequence of vectors $(x^m)_{m\in\mathbb{N}}$ such that $\|x^m\|_1 = 1$ and $N(x^m) \to 0$ when $m \to +\infty$. The unit sphere of X is compact, thus one may assume (up to the extraction of a subsequence) that x^m converges to a vector x such that $\|x\|_1 = 1$. In particular, $x \neq 0$. Because N is continuous on X, one has also $N(x) = \lim_{m\to+\infty} N(x^m) = 0$ and because N is a norm, we deduce $x = 0$, a contradiction. □

7.1.3 Duality

Definition 7.3 *Given a norm $\|\cdot\|$ on K^n ($K = \mathbb{R}$ or $\mathbb{C}$), its dual norm on K^n is defined by*

$$\|x\|' := \sup_{y \neq 0} \frac{\Re\langle x, y\rangle}{\|y\|},$$

or equivalently

$$\|x\|' := \sup_{y \neq 0} \frac{|\langle x, y\rangle|}{\|y\|}.$$

The fact that $\|\cdot\|'$ is a norm is obvious. For every $x, y \in K^n$, one has

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|'. \tag{7.3}$$

Proposition 7.2 shows that the dual norm of $\|\cdot\|_p$ is $\|\cdot\|_q$ for $1/p + 1/q = 1$. Because $p \mapsto q$ is an involution, this suggests the following property:

Proposition 7.4 *The bi-dual (dual of the dual norm) of a norm is this norm itself:*

$$\left(\|\cdot\|'\right)' = \|\cdot\|.$$

Proof. From (7.3), one has $\left(\|\cdot\|'\right)' \leq \|\cdot\|$. The converse is a consequence of the Hahn–Banach theorem: the unit ball B of $\|\cdot\|$ is convex and compact. If x is a point of its boundary (i.e., $\|x\| = 1$), there exists an $\mathbb{R}$-affine (i.e., of the form constant plus $\mathbb{R}$-linear) function that vanishes at x and is nonpositive on B. Such a function can be written in the form $z \mapsto \Re\langle z, y\rangle + c$, where c is a constant, necessarily equal to $-\Re\langle x, y\rangle$. Without loss of generality, one may assume that $\langle y, x\rangle$ is real and nonnegative. Hence

$$\|y\|' = \sup_{\|z\|=1} \Re\langle y, z\rangle = \langle y, x\rangle.$$

One deduces

$$\left(\|x\|'\right)' \geq \frac{\langle y, x\rangle}{\|y\|'} = 1 = \|x\|.$$

By homogeneity, this is true for every $x \in \mathbb{C}^n$. □

7.1.4 Matrix Norms

Let us recall that $\mathbf{M}_n(K)$ can be identified with the set of endomorphisms of $E = K^n$ by

$$A \mapsto (x \mapsto Ax).$$

Definition 7.4 *If $\|\cdot\|$ is a norm on E and if $A \in \mathbf{M}_n(K)$, we define*

$$\|A\| := \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}.$$

Equivalently,

$$\|A\| = \sup_{\|x\|\leq 1} \|Ax\| = \max_{\|x\|\leq 1} \|Ax\|.$$

One verifies easily that $A \mapsto \|A\|$ is a norm on $\mathbf{M}_n(K)$. It is called the *norm induced* by that of E, or the norm *subordinated* to that of E. Although we adopted the same notation $\|\cdot\|$ for the two norms, that on E and that on $\mathbf{M}_n(K)$, these are, of course, distinct objects. In many places, one finds the notation $|||\cdot|||$ for the induced norm. When one does not wish to mention by which norm on E a given norm on $\mathbf{M}_n(K)$ is induced, one says that $A \mapsto \|A\|$ is a *matrix norm*. The main properties of matrix norms are

$$\|AB\| \leq \|A\|\,\|B\|, \quad \|I_n\| = 1.$$

These properties are those of any *algebra norm*. In particular, one has $\|A^k\| \leq \|A\|^k$ for every $k \in \mathbb{N}$.

Examples

Three l^p-matrix norms can be computed in closed form:

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{i=n} |a_{ij}|,$$

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{j=n} |a_{ij}|,$$

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

To prove these formulæ , we begin by proving the inequalities $\geq$, selecting a suitable vector x, and writing $\|A\|_p \geq \|Ax\|_p/\|x\|_p$. For $p = 1$ we choose an index j such that the maximum in the above formula is achieved. Then we let $x_j = 1$, and $x_k = 0$ otherwise. For $p = \infty$, we let $x_j = \bar{a}_{i_0 j}/|a_{i_0 j}|$, where i_0 achieves the maximum in the above formula. For $p = 2$ we choose an eigenvector of A^*A associated with an eigenvalue of maximal modulus. We thus obtain three inequalities. The reverse inequalities are direct consequences of the definitions. The similarity of the formulæ for $\|A\|_1$ and $\|A\|_\infty$, as well as the equality $\|A\|_2 = \|A^*\|_2$ illustrate the general formula

$$\|A^*\|_{p'} = \|A\|_p = \sup_{x\neq 0}\sup_{y\neq 0} \frac{\Re(y^*Ax)}{\|x\|_p \cdot \|y\|_{p'}} = \sup_{x\neq 0}\sup_{y\neq 0} \frac{|(y^*Ax)|}{\|x\|_p \cdot \|y\|_{p'}},$$

where again p and p' are conjugate exponents.

We point out that if H is Hermitian, then $\|H\|_2 = \rho(H^2)^{1/2} = \rho(H)$. Therefore the spectral radius is a norm over $\mathbf{H}_n$, although it is not over $\mathbf{M}_n(\mathbb{C})$. We already mentioned this fact, as a consequence of the Weyl inequalities.

Proposition 7.5 *For an induced norm, the condition $\|B\| < 1$ implies that $I_n - B$ is invertible, with the inverse given by the sum of the series*

$$\sum_{k=0}^{\infty} B^k.$$

Proof. The series $\sum_k B^k$ is normally convergent, because $\sum_k \|B^k\| \le \sum_k \|B\|^k$, where the latter series converges because $\|B\| < 1$. Because $\mathbf{M}_n(K)$ is complete, the series $\sum_k B^k$ converges. Furthermore, $(I_n - B)\sum_{k\le N} B^k = I_n - B^{N+1}$, which tends to I_n. The sum of the series is thus the inverse of $I_n - B$. One has, moreover,

$$\|(I_n - B)^{-1}\| \le \sum_k \|B\|^k = \frac{1}{1 - \|B\|}.$$

□

One can also deduce Proposition 7.5 from the following statement.

Proposition 7.6 *For every induced norm, one has*

$$\rho(A) \le \|A\|.$$

Proof. The case $K = \mathbb{C}$ is easy, because there exists an eigenvector $X \in E$ associated with an eigenvalue of modulus $\rho(A)$:

$$\rho(A)\|X\| = \|\lambda X\| = \|AX\| \le \|A\|\,\|X\|.$$

If $K = \mathbb{R}$, one needs a more involved trick.

Let us choose a norm on $\mathbb{C}^n$ and let us denote by N the induced norm on $\mathbf{M}_n(\mathbb{C})$. We still denote by N its restriction to $\mathbf{M}_n(\mathbb{R})$; it is a norm. This space has finite dimension, thus any two norms are equivalent: there exists $C > 0$ such that $N(B) \le C\|B\|$ for every B in $\mathbf{M}_n(\mathbb{R})$. Using the result already proved in the complex case, one has for every $m \in \mathbb{N}$ that

$$\rho(A)^m = \rho(A^m) \le N(A^m) \le C\|A^m\| \le C\|A\|^m.$$

Taking the mth root and letting m tend to infinity, and noticing that $C^{1/m}$ tends to 1, one obtains the announced inequality. □

In general, the equality does not hold. For example, if A is nilpotent although nonzero, one has $\rho(A) = 0 < \|A\|$ for every matrix norm.

Proposition 7.7 *Let $\|\cdot\|$ be a norm on K^n and $P \in \mathbf{GL}_n(K)$. Hence, $N(x) := \|Px\|$ defines a norm on K^n. Denoting still by $\|\cdot\|$ and N the induced norms on K^n, one has $N(A) = \|PAP^{-1}\|$.*

Proof. Using the change of dummy variable $y = Px$, we have

$$N(A) = \sup_{x \neq 0} \frac{\|PAx\|}{\|Px\|} = \sup_{y \neq 0} \frac{\|PAP^{-1}y\|}{\|y\|} = \|PAP^{-1}\|.$$

□

7.2 Householder's Theorem

Householder's theorem is a kind of converse of the inequality $\rho(B) \leq \|B\|$.

Theorem 7.1 *For every $B \in \mathbf{M}_n(\mathbb{C})$ and all $\varepsilon > 0$, there exists a norm on $\mathbb{C}^n$ such that for the induced norm,*

$$\|B\| \leq \rho(B) + \varepsilon.$$

In other words, $\rho(B)$ is the infimum of $\|B\|$, as $\|\cdot\|$ ranges over the set of matrix norms.

Proof. From Theorem 3.5 there exists $P \in \mathbf{GL}_n(\mathbb{C})$ such that $T := PBP^{-1}$ is upper-triangular. From Proposition 7.7, one has

$$\inf \|B\| = \inf \|PBP^{-1}\| = \inf \|T\|,$$

where the infimum is taken over the set of induced norms. Because B and T have the same spectra, hence the same spectral radius, it is enough to prove the theorem for upper-triangular matrices.

For such a matrix T, Proposition 7.7 still gives

$$\inf \|T\| \leq \inf\{\|QTQ^{-1}\|_2 ; Q \in \mathbf{GL}_n(\mathbb{C})\}.$$

Let us now take $Q(\mu) = \mathrm{diag}(1, \mu, \mu^2, \ldots, \mu^{n-1})$. The matrix $Q(\mu)TQ(\mu)^{-1}$ is upper-triangular, with the same diagonal as that of T. Indeed, the entry with indices (i, j) becomes $\mu^{i-j} t_{ij}$. Hence,

$$\lim_{\mu \to \infty} Q(\mu)TQ(\mu)^{-1}$$

is simply the matrix $D = \mathrm{diag}(t_{11}, \ldots, t_{nn})$. Because $\|\cdot\|_2$ is continuous (as is every norm), one deduces

$$\inf \|T\| \leq \lim_{\mu \to \infty} \|Q(\mu)TQ(\mu)^{-1}\|_2 = \|D\|_2 = \sqrt{\rho(D^* D)} = \max |t_{jj}| = \rho(T).$$

□

Remark

The theorem tells us that $\rho(A) = \Lambda(A)$, where

$$\Lambda(A) := \inf \|A\|,$$

the infimum being taken over the set of matrix norms. The first part of the proof tells us that ρ and Λ coincide on the set of diagonalizable matrices, which is a dense subset of $\mathbf{M}_n(\mathbb{C})$. But this is insufficient to conclude, since Λ is a priori only upper semicontinuous, as the infimum of continuous functions. The continuity of Λ is actually a consequence of the theorem.

An interesting consequence of Householder's theorem is the following link between matrix norms and the spectral radius.

Proposition 7.8 *If $A \in \mathbf{M}_n(k)$ (with $k = \mathbb{R}$ or $\mathbb{C}$), then*

$$\rho(A) = \lim_{m\to\infty} \|A^m\|^{1/m}$$

for every matrix norm.

Proof. Let $\|\cdot\|$ be a matrix norm over $\mathbf{M}_n(k)$. From Proposition 7.6 and the fact that

$$\mathrm{Sp}(A^m) = \{\lambda^m \,|\, \lambda \in \mathrm{Sp}A\},$$

we have

$$\rho(A) = \rho(A^m)^{1/m} \le \|A^m\|^{1/m}.$$

Passing to the limit, we have

$$\rho(A) \le \liminf_{m\to+\infty} \|A^m\|^{1/m}. \tag{7.4}$$

Conversely, let $\varepsilon > 0$ be given. From Theorem 7.1, there exists a matrix norm N over $\mathbf{M}_n(\mathbb{C})$ such that $N(A) < \rho(A) + \varepsilon$. By finite-dimensionality, the restriction of N over $\mathbf{M}_n(k)$ (just in case that $k = \mathbb{R}$) is equivalent to $\|\cdot\|$: there exists a constant c such that $\|\cdot\| \le cN$. If $m \ge 1$, we thus have

$$\|A^m\| \le cN(A^m) \le cN(A)^m < c(\rho(A) + \varepsilon)^m,$$

whence

$$\|A^m\|^{1/m} \le c^{1/m}(\rho(A) + \varepsilon).$$

Passing to the limit, we have

$$\limsup_{m\to+\infty} \|A^m\|^{1/m} \le \rho(A) + \varepsilon.$$

Letting ε tend to zero, there remains

$$\limsup_{m\to+\infty} \|A^m\|^{1/m} \le \rho(A).$$

Comparing with (7.4), we see that the sequence $\|A^m\|^{1/m}$ is convergent and its limit equals $\rho(A)$.

□

Comments

- In the first edition, we proved Proposition 7.8 by applying the Hadamard formula for the convergence radius of the series $\sum_{j\geq 0} z^n A^n$.
- A classical lemma in calculus tells us that the limit of $\|A^m\|^{1/m}$ is also its infimum over m, because the sequence $\|A^m\|$ is submultiplicative.

7.3 An Interpolation Inequality

Theorem 7.2 (case $K = \mathbb{C}$) *Let $\|\cdot\|_p$ be the norm on $\mathbf{M}_n(\mathbb{C})$ induced by the norm l^p on $\mathbb{C}^n$. The function*

$$\begin{aligned} 1/p &\mapsto \log \|A\|_p, \\ [0,1] &\to \mathbb{R}, \end{aligned}$$

is convex. In other words, if $1/r = \theta/p + (1-\theta)/q$ with $\theta \in (0,1)$, then

$$\|A\|_r \leq \|A\|_p^\theta \|A\|_q^{1-\theta}.$$

Remarks

1. The proof uses the fact that $K = \mathbb{C}$. However, the norms induced by the $\|\cdot\|_p$s on $\mathbf{M}_n(\mathbb{R})$ and $\mathbf{M}_n(\mathbb{C})$ take the same values on real matrices, even although their definitions are different (see Exercise 6). The statement is thus still true in $\mathbf{M}_n(\mathbb{R})$.
2. The case $(p,q,r) = (1,\infty,2)$ admits a direct proof. See the exercises.
3. The result still holds true in infinite dimension, at the expense of some functional analysis. One can even take different L^p norms at the source and target spaces. Here is an example.

   Theorem 7.3 (Riesz–Thorin) *Let Ω be an open set in $\mathbb{R}^D$ and ω an open set in $\mathbb{R}^d$. Let p_0, p_1, q_0, q_1 be four numbers in $[1,+\infty]$. Let $\theta \in [0,1]$ and p,q be defined by*

   $$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

   Consider a linear operator T defined on $L^{p_0} \cap L^{p_1}(\Omega)$, taking values in $L^{q_0} \cap L^{q_1}(\omega)$. Assume that T can be extended as a continuous operator from $L^{p_j}(\Omega)$ to $L^{q_j}(\omega)$, with norm M_j, $j = 1,2$:

$$M_j := \sup_{f \neq 0} \frac{\|Tf\|_{q_j}}{\|f\|_{p_j}}.$$

Then T can be extended as a continuous operator from $L^p(\Omega)$ to $L^q(\omega)$, and its norm is bounded above by

$$M_0^{1-\theta} M_1^{\theta}.$$

4. A fundamental application is the continuity of the Fourier transform from $L^p(\mathbb{R}^d)$ into its dual $L^{p'}(\mathbb{R}^d)$ when $1 \leq p \leq 2$. We have only to observe that

$$(p_0, p_1, q_0, q_1) = (1, 2, +\infty, 2)$$

is suitable. It can be proved by inspection that every pair (p,q) such that the Fourier transform is continuous from $L^p(\mathbb{R}^d)$ into $L^q(\mathbb{R}^d)$ has the form (p, p') with $1 \leq p \leq 2$.
5. One has analogous results for the Fourier series. Therein lies the origin of the Riesz–Thorin theorem.

Proof. (Due to Riesz)
Let us fix x and y in K^n. We have to bound

$$|\langle y, Ax \rangle| = \left| \sum_{j,k} a_{jk} x_j \bar{y}_k \right|.$$

Let B be the strip in the complex plane defined by $\Re z \in [0,1]$. Given $z \in B$, define "conjugate" exponents $r(z)$ and $r'(z)$ by

$$\frac{1}{r(z)} = \frac{z}{p} + \frac{1-z}{q}, \qquad \frac{1}{r'(z)} = \frac{z}{p'} + \frac{1-z}{q'}.$$

Set

$$X_j(z) := |x_j|^{-1+r/r(z)} x_j = x_j \exp\left(\left(\frac{r}{r(z)} - 1\right) \log |x_j|\right),$$
$$Y_j(z) := |y_j|^{-1+r'/r'(\bar{z})} y_j.$$

We then have

$$\|X(z)\|_{r(\Re z)} = \|x\|_r^{r/r(\Re z)}, \quad \|Y(z)\|_{r'(\Re z)} = \|y\|_{r'}^{r'/r'(\Re z)}.$$

Next, define a holomorphic map in the strip B by $f(z) := \langle Y(z), AX(z) \rangle$. It is bounded, because the numbers $X_j(z)$ and $Y_k(z)$ are. For example,

$$|X_j(z)| = |x_j|^{r/r(\Re z)}$$

lies between $|x_j|^{r/p}$ and $|x_j|^{r/q}$.

Let us set $M(\theta) = \sup\{|f(z)|; \Re z = \theta\}$. Hadamard's *three-line lemma* (see [33], Chapter 12, Exercise 8) tells us that, f being bounded and holomorphic in the strip,

$$\theta \mapsto \log M(\theta)$$

is convex on $(0,1)$. However, $r(0) = q$, $r(1) = p$, $r'(0) = q'$, $r'(1) = p'$, $r(\theta) = r$, $r'(\theta) = r'$, $X(\theta) = x$, and $Y(\theta) = y$. Hence

$$|\langle y, Ax\rangle| = |f(\theta)| \leq M(\theta) \leq M(1)^{\theta} M(0)^{1-\theta}.$$

Now we have

$$\begin{aligned} M(1) &= \sup\{|f(z)|; \Re z = 1\} \\ &\leq \sup\{\|AX(z)\|_{r(1)} \|Y(z)\|_{r(1)'}; \Re z = 1\} \\ &= \sup\{\|AX(z)\|_p \|Y(z)\|_{p'}; \Re z = 1\} \\ &\leq \|A\|_p \sup\{\|X(z)\|_p \|Y(z)\|_{p'}; \Re z = 1\} \\ &= \|A\|_p \|x\|_r^{r/p} \|y\|_{r'}^{r'/p'}. \end{aligned}$$

Likewise, $M(0) \leq \|A\|_q \|x\|_r^{r/q} \|y\|_{r'}^{r'/q'}$. Hence

$$\begin{aligned} |\langle y, Ax\rangle| &\leq \|A\|_p^{\theta} \|A\|_q^{1-\theta} \|x\|_r^{r(\theta/p + (1-\theta)/q)} \|y\|_{r'}^{r'(\theta/p' + (1-\theta)/q')} \\ &= \|A\|_p^{\theta} \|A\|_q^{1-\theta} \|x\|_r \|y\|_{r'}. \end{aligned}$$

Finally,

$$\|Ax\|_r = \sup_{y \neq 0} \frac{|\langle y, Ax\rangle|}{\|y\|_{r'}} \leq \|A\|_p^{\theta} \|A\|_q^{1-\theta} \|x\|_r,$$

which proves the theorem. □

7.4 Von Neumann's Inequality

We say that a matrix $M \in \mathbf{M}_n(\mathbb{C})$ is a *contraction* if $\|Mx\|_2 \leq \|x\|_2$ for every vector x, or equivalently $\|M\|_2 \leq 1$. Developing in the form $x^*M^*Mx \leq x^*x$, this translates as $M^*M \leq I_n$ in the sense of Hermitian matrices. Inasmuch as $\|M^*\|_2 = \|M\|_2$, the Hermitian conjugate of a contraction is a contraction. When U and V are unitary matrices, U^*MV is a contraction if and only if M is a contraction.

The following statement is due to von Neumann.

Theorem 7.4 *Let $M \in \mathbf{M}_n(\mathbb{C})$ be a contraction and $P \in \mathbb{C}[X]$ be a polynomial. Then*

$$\|P(M)\|_2 \leq \sup\{|P(z)| \,|\, z \in \mathbb{C}, |z| = 1\}.$$

Proof. We begin with the easy case, where M is normal. Then $M = U^*DU$ with $U \in \mathbf{U}_n$ and D is diagonal (Theorem 5.4). Because U is unitary, D is a contraction, meaning that its diagonal entries d_j belong to the unit disk. We have $P(M) = U^*P(D)U$, whence

$$\|P(M)\|_2 = \|P(D)\|_2 = \max_j |P(d_j)| \leq \sup\{|P(z)|\,|\, z \in \mathbb{C},\, |z| = 1\},$$

using the maximum principle.

We now turn to the general case of a nonnormal contraction. Using the square root of nonnegative Hermitian matrices, and thanks to $M^*M \leq I_n$ and $MM^* \leq I_n$, we denote

$$S = \sqrt{I_n - M^*M}, \qquad T := \sqrt{I_n - MM^*}.$$

Let us choose a real polynomial $Q \in \mathbb{R}[X]$ with the interpolation property that $Q(t) = \sqrt{1-t}$ over the spectrum of M^*M (which is a finite set and equals the spectrum of MM^*). Then $S = Q(M^*M)$ and $T = Q(MM^*)$. Because $M(M^*M)^r = (MM^*)^r M$ for every $r \in \mathbb{N}$, we infer $MS = TM$. Likewise, we have $SM^* = M^*T$.

Given an integer $k \geq 1$ and $\ell = 2k+1$, we define a matrix $V_k \in \mathbf{M}_{\ell n}(\mathbb{C})$ blockwise:

$$V_k = \begin{pmatrix}
 & \ddots & & & & & & & \\
 & & I_n & & & & & & \\
 & & & I_n & & & & & \\
 & & & & S & -M^* & & & \\
 & & & & M & T & & & \\
 & & & & & & I_n & & \\
 & & & & & & & I_n & \\
 & & & & & & & & \ddots \\
I_n & & & & & & & &
\end{pmatrix},$$

where the dots represent blocks I_n. The column and row indices range from $-k$ to k. The central block indexed by $(0,0)$ is M. All the missing (apart from dots) entries are blocks 0_n. In particular, the diagonal blocks are equal to 0_n, except for the central one.

Computing the product $V_k^* V_k$, and using $SM^* = M^*T$ and $MS = TM$, we find $I_{\ell n}$. In other words, V_k is a unitary matrix. This is a special case of normal contraction, therefore the first step above thus applies:

$$\|P(V_k)\|_2 \leq \sup\{|P(z)|\,|\, z \in \mathbb{C},\, |z| = 1\}. \tag{7.5}$$

Let us now observe that in the qth power of V_k, the central block is M^k provided that $q \leq 2k$. This because V_k is block-triangular[1] up to the lower-left block I_n. Therefore the central block of $p(V_k)$ for a polynomial p with $\mathrm{d}^o p \leq 2k$ is precisely $p(M)$.

[1] Compare with Exercise 12 of Chapter 5, which states that normal matrices are far from triangular.

We now choose $k \geq (\mathrm{d}^o P)/2$, so that $P(M)$ is a block of $P(V_k)$. We have $P(M) = \Pi P(V_k)\Pi*$ where

$$\Pi = (\dots, 0_n, I_n, 0_n, \dots).$$

If $x \in \mathbb{C}^n$, let $X \in \mathbb{C}^{\ell n}$ denote the vector $\Pi^* x$. We have $\|X\|_2 = \|x\|_2$. Because Π is an orthogonal projection, we find

$$\begin{aligned}\|P(M)x\|_2 = \|\Pi P(V_k)X\|_2 \leq \|P(V_k)X\|_2 &\leq \sup\{|P(z)|\,|\,z \in \mathbb{C}, |z| = 1\}\|X\|_2 \\ &= \sup\{|P(z)|\,|\,z \in \mathbb{C}, |z| = 1\}\|x\|_2,\end{aligned}$$

where we have used (7.5). This is exactly

$$\|P(M)\|_2 \leq \sup\{|P(z)|\,|\,z \in \mathbb{C}, |z| = 1\}.$$

□

Exercises

1. Under what conditions on the vectors $a, b \in \mathbb{C}^n$ does the matrix M defined by $m_{ij} = a_i b_j$ satisfy $\|M\|_p = 1$ for every $p \in [1, \infty]$?
2. Under what conditions on x, y, and p does the equality in (7.2) or (7.1) hold?
3. Show that
$$\lim_{p \to +\infty} \|x\|_p = \|x\|_\infty, \quad \forall x \in E.$$
4. A norm on K^n is a *strictly convex* norm if $\|x\| = \|y\| = 1$, $x \neq y$, and $0 < \theta < 1$ imply $\|\theta x + (1-\theta)y\| < 1$.
   a. Show that $\|\cdot\|_p$ is strictly convex for $1 < p < \infty$, but is not so for $p = 1, \infty$.
   b. Deduce from Corollary 8.1 that the induced norm $\|\cdot\|_p$ is not strictly convex on $\mathbf{M}_n(\mathbb{R})$.
5. Let N be a norm on $\mathbb{R}^n$.
   a. For $x \in \mathbb{C}^n$, define
   $$N_1(x) := \inf\left\{\sum_\ell |\alpha_\ell| N(x^\ell)\right\},$$
   where the infimum is taken over the set of decompositions $x = \sum_\ell \alpha_\ell x^\ell$ with $\alpha_\ell \in \mathbb{C}$ and $x^\ell \in \mathbb{R}^n$. Show that N_1 is a norm on $\mathbb{C}^n$ (as a $\mathbb{C}$-vector space) whose restriction to $\mathbb{R}^n$ is N. **Note:** N_1 is called the *complexification* of N.
   b. Same question as above for N_2, defined by
   $$N_2(x) := \frac{1}{2\pi}\int_0^{2\pi} [\mathrm{e}^{\mathrm{i}\theta} x] d\theta,$$

where

$$[x] := \sqrt{N(\Re x)^2 + N(\Im x)^2}.$$

c. Show that $N_2 \leq N_1$.
d. If $N(x) = \|x\|_1$, show that $N_1(x) = \|x\|_1$. Considering then the vector

$$x = \begin{pmatrix} 1 \\ i \end{pmatrix},$$

show that $N_2 \neq N_1$.

6. (Continuation of Exercise 5)
The norms N (on $\mathbb{R}^n$) and N_1 (on $\mathbb{C}^n$) lead to induced norms on $\mathbf{M}_n(\mathbb{R})$ and $\mathbf{M}_n(\mathbb{C})$, respectively. Show that if $M \in \mathbf{M}_n(\mathbb{R})$, then $N(M) = N_1(M)$. Deduce that Theorem 7.2 holds true in $\mathbf{M}_n(\mathbb{R})$.
7. Let $\|\cdot\|$ be an algebra norm on $\mathbf{M}_n(K)$ ($K = \mathbb{R}$ or $\mathbb{C}$), that is, a norm satisfying $\|AB\| \leq \|A\| \cdot \|B\|$. Show that $\rho(A) \leq \|A\|$ for every $A \in \mathbf{M}_n(K)$.
8. Let $B \in \mathbf{M}_n(\mathbb{C})$ be given. Assume that there exists an induced norm such that $\|B\| = \rho(B)$. Let λ be an eigenvalue of maximal modulus and X a corresponding eigenvector. Show that X does not belong to the range of $B - \lambda I_n$. Deduce that the Jordan block associated with λ is diagonal (Jordan reduction is presented in Chapter 9).
9. (Continuation of Exercise 8)
Conversely, show that if the Jordan blocks of B associated with the eigenvalues of maximal modulus of B are diagonal, then there exists a norm on $\mathbb{C}^n$ such that, using the induced norm, $\rho(B) = \|B\|$.
10. Here is another proof of Theorem 7.1. Let $K = \mathbb{R}$ or $\mathbb{C}$, $A \in \mathbf{M}_n(K)$, and let N be a norm on K^n. If $\varepsilon > 0$, we define for all $x \in K^n$

$$\|x\| := \sum_{k \in \mathbb{N}} (\rho(A) + \varepsilon)^{-k} N(A^k x).$$

   a. Show that this series is convergent (use Proposition 7.8).
   b. Show that $\|\cdot\|$ is a norm on K^n.
   c. Show that for the induced norm, $\|A\| \leq \rho(A) + \varepsilon$.

11. A norm $\|\cdot\|$ on $\mathbf{M}_n(\mathbb{C})$ is said to be *unitarily invariant* if $\|UAV\| = \|A\|$ for every $A \in \mathbf{M}_n(\mathbb{C})$ and all unitary matrices U, V.

   a. Find, among the most classical norms, two examples of unitarily invariant norms.
   b. Given a unitarily invariant norm, show that there exists a norm N on $\mathbb{R}^n$ such that

$$\|A\| = N(s_1(A), \ldots, s_n(A)),$$

   where the $s_j(A)$s, the eigenvalues of H in the polar decomposition $A = QH$ (see Section 11.4 for this notion), are called the *singular values* of A.

12. (Bhatia [5]) Suppose we are given a norm $\|\cdot\|$ on $\mathbf{M}_n(\mathbb{C})$ that is unitarily invariant (see the previous exercise). If $A \in \mathbf{M}_n(\mathbb{C})$, we denote by $D(A)$ the diagonal matrix obtained by keeping only the a_{jj} and setting all the other entries to zero. If σ is a permutation, we denote by A^σ the matrix whose entry of index (j,k) equals a_{jk} if $k = \sigma(j)$, and zero otherwise. For example, $A^{id} = D(A)$, where id is the identity permutation. If r is an integer between $1-n$ and $n-1$, we denote by $D_r(A)$ the matrix whose entry of index (j,k) equals a_{jk} if $k-j=r$, and zero otherwise. For example, $D_0(A) = D(A)$.

   a. Let $\omega = \exp(2i\pi/n)$ and let U be the diagonal matrix whose diagonal entries are the roots of unity $1, \omega, \dots, \omega^{n-1}$. Show that
   $$D(A) = \frac{1}{n}\sum_{j=0}^{n-1} U^{*j} A U^j.$$
   Deduce that $\|D(A)\| \le \|A\|$.
   b. Show that $\|A^\sigma\| \le \|A\|$ for every $\sigma \in S_n$. Observe that $\|P\| = \|I_n\|$ for every permutation matrix P. Show that $\|M\| \le \|I_n\|$ for every bistochastic matrix M (see Section 8.5 for this notion).
   c. If $\theta \in \mathbb{R}$, let us denote by U_θ the diagonal matrix, whose kth diagonal term equals $\exp(ik\theta)$. Show that
   $$D_r(A) = \frac{1}{2\pi}\int_0^{2\pi} e^{ir\theta} U_\theta A U_\theta^* d\theta.$$
   d. Deduce that $\|D_r(A)\| \le \|A\|$.
   e. Let p be an integer between zero and $n-1$, and set $r = 2p+1$. Let us denote by $T_r(A)$ the matrix whose entry of index (j,k) equals a_{jk} if $|k-j| \le p$, and zero otherwise. For example, $T_3(A)$ is a tridiagonal matrix. Show that
   $$T_r(A) = \frac{1}{2\pi}\int_0^{2\pi} d_p(\theta) U_\theta A U_\theta^* d\theta,$$
   where
   $$d_p(\theta) = \sum_{-p}^{p} e^{ik\theta}$$
   is the *Dirichlet kernel*.
   f. Deduce that $\|T_r(A)\| \le L_p \|A\|$, where
   $$L_p = \frac{1}{2\pi}\int_0^{2\pi} |d_p(\theta)| d\theta$$
   is the *Lebesgue constant* (**Note:** $L_p = 4\pi^{-2}\log p + O(1)$).
   g. Let $\Delta(A)$ be the upper-triangular matrix whose entries above the diagonal coincide with those of A. Using the matrix

$$B = \begin{pmatrix} 0_n & \Delta(A)^* \\ \Delta(A) & 0_n \end{pmatrix},$$

show that $\|\Delta(A)\|_2 \le L_n \|A\|_2$ (observe that $\|B\|_2 = \|\Delta(A)\|_2$).

h. What inequality do we obtain for $\Delta_0(A)$, the strictly upper-triangular matrix whose entries lying strictly above the diagonal coincide with those of A?

13. We endow $\mathbb{C}^n$ with the usual Hermitian structure, so that $\mathbf{M}_n(\mathbb{C})$ is equipped with the norm $\|A\|_2 = \rho(A^*A)^{1/2}$.
Suppose we are given a sequence of matrices $(A_j)_{j\in\mathbb{Z}}$ in $\mathbf{M}_n(\mathbb{C})$ and a summable sequence $\gamma \in l^1(\mathbb{Z})$ of positive real numbers. Assume, finally, that for every pair $(j,k) \in \mathbb{Z}\times\mathbb{Z}$,

$$\|A_j^* A_k\|_2 \le \gamma(j-k)^2, \quad \|A_j A_k^*\|_2 \le \gamma(j-k)^2.$$

a. Let F be a finite subset of $\mathbb{Z}$. Let B_F denote the sum of the A_js as j runs over F. Show that

$$\|(B_F^* B_F)^{2m}\|_2 \le \text{ card } F \, \|\gamma\|_1^{2m}, \quad \forall m \in \mathbb{N}.$$

b. Deduce that $\|B_F\|_2 \le \|\gamma\|_1$.

c. Show (*Cotlar's lemma*) that for every $x, y \in \mathbb{C}^n$, the series

$$y^T \sum_{j\in\mathbb{Z}} A_j x$$

is convergent, and that its sum $y^T Ax$ defines a matrix $A \in \mathbf{M}_n(\mathbb{C})$ that satisfies

$$\|A\| \le \sum_{j\in\mathbb{Z}} \gamma(j).$$

Hint: For a sequence $(u_j)_{j\in\mathbb{Z}}$ of real numbers, the series $\sum_j u_j$ is absolutely convergent if and only if there exists $M < +\infty$ such that $\sum_{j\in F} |u_j| \le M$ for every finite subset F.

d. Deduce that the series $\sum_j A_j$ converges in $\mathbf{M}_n(\mathbb{C})$. May one conclude that it converges normally?

14. Let $\|\cdot\|$ be an induced norm on $\mathbf{M}_n(\mathbb{R})$. We wish to characterize the matrices $B \in \mathbf{M}_n(\mathbb{R})$ such that there exist $\varepsilon_0 > 0$ and $\omega > 0$ with

$$(0 < \varepsilon < \varepsilon_0) \Longrightarrow (\|I_n - \varepsilon B\| \le 1 - \omega\varepsilon).$$

a. For the norm $\|\cdot\|_\infty$, it is equivalent that B be strictly diagonally dominant.

b. What is the characterization for the norm $\|\cdot\|_1$?

c. For the norm $\|\cdot\|_2$, it is equivalent that $B^T + B$ is positive-definite.

15. Let $B \in \mathbf{M}_n(\mathbb{C})$ be given.

a. Returning to the proof of Theorem 7.1, show that for every $\varepsilon > 0$ there exists on $\mathbb{C}^n$ an Hermitian norm $\|\cdot\|$ such that for the induced norm $\|B\| \leq \rho(B) + \varepsilon$.
b. Deduce that $\rho(B) < 1$ holds if and only if there exists a matrix $A \in \mathbf{HPD}_n$ such that $A - B^*AB \in \mathbf{HPD}_n$.

16. Let $A \in \mathbf{M}_n(\mathbb{C})$ be a diagonalizable matrix: $A = S\,\mathrm{diag}(d_1, \dots, d_n)S^{-1}$. Let $\|\cdot\|$ be an induced norm for which $\|D\| = \max_j |d_j|$ holds, where

$$D := \mathrm{diag}(d_1, \dots, d_n).$$

Show that for every $E \in \mathbf{M}_n(\mathbb{C})$ and for every eigenvalue λ of $A + E$, there exists an index j such that

$$|\lambda - d_j| \leq \|S\| \cdot \|S^{-1}\| \cdot \|E\|.$$

17. Let $A \in \mathbf{M}_n(K)$, with $K = \mathbb{R}$ or $\mathbb{C}$. Give another proof, using the Cauchy–Schwarz inequality, of the following particular case of Theorem 7.2:

$$\|A\|_2 \leq \|A\|_1^{1/2} \|A\|_\infty^{1/2}.$$

18. Show that if $A \in \mathbf{M}_n(\mathbb{C})$ is normal, then $\rho(A) = \|A\|_2$. Deduce that if A and B are normal, $\rho(AB) \leq \rho(A)\rho(B)$.
19. Let N_1 and N_2 be two norms on $\mathbb{C}^n$. Denote by $\mathcal{N}_1$ and $\mathcal{N}_2$ the induced norms on $\mathbf{M}_n(\mathbb{C})$. Let us define

$$R := \max_{x \neq 0} \frac{N_1(x)}{N_2(x)}, \quad S := \max_{x \neq 0} \frac{N_2(x)}{N_1(x)}.$$

a. Show that
$$\max_{A \neq 0} \frac{\mathcal{N}_1(A)}{\mathcal{N}_2(A)} = RS = \max_{A \neq 0} \frac{\mathcal{N}_2(A)}{\mathcal{N}_1(A)}.$$
b. Deduce that if $\mathcal{N}_1 = \mathcal{N}_2$, then N_2/N_1 is constant.
c. Show that if $\mathcal{N}_1 \leq \mathcal{N}_2$, then N_2/N_1 is constant and therefore $\mathcal{N}_2 = \mathcal{N}_1$.

20. (Continuation of Exercise 19)
Let $\|\cdot\|$ be an algebra norm on $\mathbf{M}_n(\mathbb{C})$. If $y \in \mathbb{C}^n$ is nonzero, we define $\|x\|_y := \|xy^*\|$.

a. Show that $\|\cdot\|_y$ is a norm on $\mathbb{C}^n$ for every $y \neq 0$.
b. Let $\mathcal{N}_y$ be the norm induced by $\|\cdot\|_y$. Show that $\mathcal{N}_y \leq \|\cdot\|$.
c. We say that $\|\cdot\|$ is *minimal* if there exists no other algebra norm less than or equal to $\|\cdot\|$. Show that the following assertions are equivalent.
   i. $\|\cdot\|$ is an induced norm on $\mathbf{M}_n(\mathbb{C})$.
   ii. $\|\cdot\|$ is a minimal norm on $\mathbf{M}_n(\mathbb{C})$.
   iii. For all $y \neq 0$, one has $\|\cdot\| = \mathcal{N}_y$.

21. (Continuation of Exercise 20)

Let $\|\cdot\|$ be an induced norm on $\mathbf{M}_n(\mathbb{C})$.

a. Let $y, z \neq 0$ be two vectors in $\mathbb{C}^n$. Show that (with the notation of the previous exercise) $\|\cdot\|_y / \|\cdot\|_z$ is constant.
b. Prove the equality

$$\|xy^*\| \cdot \|zt^*\| = \|xt^*\| \cdot \|zy^*\|.$$

22. Let $M \in \mathbf{M}_n(\mathbb{C})$ and $H \in \mathbf{HPD}_n$ be given. Show that

$$\|HMH\|_2 \leq \frac{1}{2}\|H^2M + MH^2\|_2.$$

23. We endow $\mathbb{R}^2$ with the Euclidean norm $\|\cdot\|_2$, and $\mathbf{M}_2(\mathbb{R})$ with the induced norm, also denoted by $\|\cdot\|_2$. We denote by Σ the unit sphere of $\mathbf{M}_2(\mathbb{R})$: $M \in \Sigma$ is equivalent to $\|M\|_2 = 1$, that is, to $\rho(M^TM) = 1$. Likewise, B denotes the unit ball of $\mathbf{M}_2(\mathbb{R})$.
Recall that if C is a convex set and if $P \in C$, then P is called an *extremal* point if $P \in [Q, R]$ and $Q, R \in C$ imply either $Q = P$ or $R = P$.

a. Show that the set of extremal points of B is equal to $\mathbf{O}_2(\mathbb{R})$.
b. Show that $M \in \Sigma$ if and only if there exist two matrices $P, Q \in \mathbf{O}_2(\mathbb{R})$ and a number $a \in [0, 1]$ such that

$$M = P\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}Q.$$

c. We denote by $\mathscr{R} = \mathbf{SO}_2(\mathbb{R})$ the set of rotation matrices, and by $\mathscr{S}$ that of matrices of planar symmetry. Recall that $\mathbf{O}_2(\mathbb{R})$ is the disjoint union of $\mathscr{R}$ and $\mathscr{S}$. Show that Σ is the union of the segments $[r, s]$ as r runs over $\mathscr{R}$ and s runs over $\mathscr{S}$.
d. Show that two such "open" segments (r, s) and (r', s') are either disjoint or equal.
e. Let $M, N \in \Sigma$. Show that $\|M - N\|_2 = 2$ (i.e., (M, N) is a diameter of B) if and only if there exists a segment $[r, s]$ ($r \in \mathscr{R}$ and $s \in \mathscr{S}$) such that $M \in [r, s]$ and $N \in [-r, -s]$.

24. (The Banach–Mazur distance)

a. Let N and N' be two norms on k^n ($k = \mathbb{R}$ or $\mathbb{C}$). If $A \in \mathbf{GL}_n(k)$, we may define norms

$$\|A\|_{\rightarrow} := \sup_{x \neq 0} \frac{N'(Ax)}{N(x)}, \quad \|A^{-1}\|_{\leftarrow} := \sup_{x \neq 0} \frac{N(A^{-1}x)}{N'(x)}.$$

Show that $A \mapsto \|A\|_{\rightarrow}\|A^{-1}\|_{\leftarrow}$ achieves its upper bound. We denote by $\delta(N, N')$ the minimum value. Verify

$$0 \leq \log \delta(N, N'') \leq \log \delta(N, N') + \log \delta(N', N'').$$

When $N = \|\cdot\|_p$, we write ℓ^p instead. If in addition $N' = \|\cdot\|_q$, we write $\|\cdot\|_{p,q}$ for $\|\cdot\|_{\to}$.

b. In the set $\mathscr{N}$ of norms on k^n, let us consider the following equivalence relation: $N \sim N'$ if and only if there exists an $A \in \mathbf{GL}_n(k)$ such that $N' = N \circ A$. Show that $\log \delta$ induces a metric d on the quotient set $\mathbf{Norm} := \mathscr{N}/\sim$. This metric is called the *Banach–Mazur distance*. How many classes of Hermitian norms are there ?

c. Compute $\|I_n\|_{p,q}$ for $1 \leq p, q \leq n$ (there are two cases, depending on the sign of $q - p$). Deduce that

$$\delta(\ell^p, \ell^q) \leq n^\kappa, \qquad \kappa := \left| \frac{1}{p} - \frac{1}{q} \right|.$$

d. Show that $\delta(\ell^p, \ell^q) = \delta(\ell^{p'}, \ell^{q'})$, where p', q' are the conjugate exponents.

e. i. When $H \in \mathbf{H}_n$ is positive-semidefinite, find that the average of $x^* H x$, as x runs over the set defined by $|x_j| = 1$ for all js, is $\operatorname{Tr} H$ (the measure is the product of n copies of the normalized Lebesgue measure on the unit disk). Deduce that

$$\sqrt{\operatorname{Tr} M^* M} \leq \|M\|_{\infty,2} := \sup_{x \neq 0} \frac{\|Mx\|_2}{\|x\|_\infty}$$

for every $M \in \mathbf{M}_n(k)$.

ii. Prove also that

$$\|A\|_{p,\infty} = \max_{1 \leq i \leq n} \|A^{(i)}\|_{p'},$$

where $A^{(i)}$ denotes the ith row vector of A.

iii. Deduce that $\delta(\ell^2, \ell^\infty) = \sqrt{n}$.

iv. Using the triangle inequality for $\log \delta$, deduce that

$$\delta(\ell^p, \ell^q) = n^\kappa$$

whenever $p, q \geq 2$, and then for every p, q such that $(p-2)(q-2) \geq 0$. **Note:** The exact value of $\delta(\ell^p, \ell^q)$ is not known when $(p-2)(q-2) < 0$.

v. Remark that the "curves" $\{\ell^p \,|\, 2 \leq p \leq \infty\}$ and $\{\ell^p \,|\, 1 \leq p \leq 2\}$ are geodesics, in the sense that the restrictions of the Banach–Mazur distance to these curves satisfy the triangular *equality*.

f. When $n = 2$, prove that $\delta(\ell^1, \ell^\infty) = 1$. On the contrary, if $n \geq 3$, then prove $\delta(\ell^1, \ell^\infty) > 1$.

g. A theorem proven by John states that the diameter of $(\mathbf{Norm}, d)$ is precisely $\frac{1}{2}\log n$. Show that this metric space is compact. **Note:** One may consider the norm whose unit ball is an m-agon in $\mathbb{R}^2$, with m even. Denote its class by N_m. It seems that $d(\ell^1, N_m) = \frac{1}{2}\log 2$ when $8|m$.

25. Given three matrices $A \in \mathbf{M}_{p\times q}(k)$, $B \in \mathbf{M}_{p\times s}(k)$, and $C \in \mathbf{M}_{r\times q}(k)$, we consider the affine set $\mathscr{W}$ of matrices $W \in \mathbf{M}_{n\times m}(k)$ of the form

$$W = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where D runs over $\mathbf{M}_{r\times s}(k)$. Thus $n = p+r$ and $m = q+s$.
Denoting

$$P = \begin{pmatrix} I \\ 0 \end{pmatrix}, \quad Q = (\, I \; 0 \,)$$

the projection matrices, we are going to prove (Parrott's lemma) that

$$\min\{\|W\|_2 \,|\, W \in \mathscr{W}\} = \max\{\|QW\|_2, \|WP\|_2\}, \tag{7.6}$$

where the right-hand side does not depend on D:

$$WP = \begin{pmatrix} A \\ C \end{pmatrix}, \quad QW = (\, A \; B \,)$$

a. Check the inequality

$$\inf\{\|W\|_2 \,|\, W \in \mathscr{W}\} \geq \max\{\|QW\|_2, \|WP\|_2\}.$$

b. Denote $\mu(D) := \|W\|_2$. Show that the infimum of μ on $\mathscr{W}$ is attained.
c. Show that it is sufficient to prove (7.6) when $s = 1$.
d. From now on, we assume that $s = 1$, and we consider a matrix $D_0 \in \mathbf{M}_{r\times 1}(k)$ such that μ is minimal at D_0. We denote by W_0 the associated matrix. Let us introduce a function $D \mapsto \eta(D) = \mu(D)^2$. Recall that η is the largest eigenvalue of W^*W. We denote f_0 its multiplicity when $D = D_0$.
   i. If $f_0 \geq 2$, show that $W_0^*W_0$ has an eigenvector v with $v_m = 0$. Deduce that $\mu(D_0) \leq \|WP\|_2$. Conclude in this case.
   ii. From now on, we suppose $f_0 = 1$. Show that $\eta(D)$ is a simple eigenvalue for every D in a small neighbourhood of D_0. Show that $D \mapsto \eta(D)$ is differentiable at D_0, and that its differential is given by

$$\Delta \mapsto \frac{2}{\|y\|_2^2}\Re\left[(QW_0y)^*\Delta Qy\right],$$

   where y is an associated eigenvector:

$$W_0^*W_0y = \eta(D_0)y.$$

   iii. Deduce that either $Qy = 0$ or $QW_0y = 0$.
   iv. In the case where $Qy = 0$, show that $\mu(D_0) \leq \|WP\|_2$ and conclude.
   v. In the case where $QW_0y = 0$, prove that $\mu(D_0) \leq \|QW\|_2$ and conclude.

26. Let k be $\mathbb{R}$ or $\mathbb{C}$. Given a bounded subset F of $\mathbf{M}_n(k)$, let us denote by F_k the set of all possible products of k elements in F. Given a matrix norm $\|\cdot\|$, we denote $\|F_k\|$ the supremum of the norms of elements of F_k.

    a. Show that $\|F_{k+l}\| \leq \|F_k\| \cdot \|F_\ell\|$.
    b. Deduce that the sequence $\|F_k\|^{1/k}$ converges, and that its limit is the infimum of the sequence.
    c. Prove that this limit does not depend on the choice of the matrix norm. This limit is called the *joint spectral radius* of the family F, and denoted $\rho(F)$. This notion is due to Rota and Strang.
    d. Let $\hat{\rho}(F)$ denote the infimum of $\|F\|$ when $\|\cdot\|$ runs over all matrix norms. Show that $\rho(F) \leq \hat{\rho}(F)$.
    e. Given a norm N on k^n and a number $\varepsilon > 0$, we define for every $x \in k^n$

$$\|x\| := \sum_{l=0}^{\infty} (\rho(F) + \varepsilon)^{-l} \max\{N(Bx) \,|\, B \in F_\ell\}.$$

        i. Show that the series converges, and that it defines a norm on k^n.
        ii. For the matrix norm associated with $\|\cdot\|$, show that $\|A\| \leq \rho(F) + \varepsilon$ for every $A \in F$.
        iii. Deduce that actually $\rho(F) = \hat{\rho}(F)$. Compare with Householder's theorem.

27. (Rota & Strang.) Let k be $\mathbb{R}$ or $\mathbb{C}$. Given a subset F of $\mathbf{M}_n(k)$, we consider the semi-group $\mathscr{F}$ generated by F. It is the union of sets F_k defined in the previous exercise, as k runs over $\mathbb{N}$. We have $F_0 = \{I_n\}$, $F_1 = F$, $F_2 = F \cdot F, \ldots$
    If $\mathscr{F}$ is bounded, prove that there exists a matrix norm $\|\cdot\|$ such that $\|A\| \leq 1$ for every $A \in F$. **Hint:** In the previous exercise, take a sup instead of a series.
28. Let $A \in \mathbf{M}_n(\mathbb{C})$ be given. Let $\sigma(A)$ be the spectrum of A and $\rho(A)$ its complement (the *resolvent set*). For $\varepsilon > 0$, we define the ε-pseudospectrum of A as

$$\sigma_\varepsilon(A) := \sigma(A) \cup \left\{ z \in \rho(A)\,;\, \|(z-A)^{-1}\|_2 \geq \frac{1}{\varepsilon} \right\}.$$

    a. Prove that

$$\sigma_\varepsilon(A) = \bigcup_{\|B\|_2 \leq \varepsilon} \sigma(A+B).$$

    b. Prove also that

$$\sigma_\varepsilon(A) \subset \{z \in \mathbb{C}\,;\, \mathrm{dist}(z; \mathscr{H}(A)) \leq \varepsilon\},$$

    where $\mathscr{H}(A)$ is the numerical range of A.

Note: the notion of pseudo-spectrum is fundamental in several scientific domains, including dynamical systems, numerical analysis, and quantum mechanics (in semiclassical analysis, one speaks of *quasi-modes*). The reader interested in this subject should consult the book by Trefethen and Embree [38].

29. We recall that the numerical radius of $A \in \mathbf{M}_n(\mathbb{C})$ is defined by

$$w(A) := \sup\{|z|\,;\, z \in \mathscr{H}(A)\} = \sup\{|x^*Ax|\,;\, x \in \mathbb{C}^n,\ \|x\|_2 = 1\}.$$

Prove that

$$w(A) \le \|A\|_2 \le 2w(A).$$

Hint: Use the polarization principle to prove the second inequality.

30. Let $A \in \mathbf{M}_n(\mathbb{C})$ be a nilpotent matrix of order two: $A^2 = 0_n$.

a. Using standard properties of the norm $\|\cdot\|_2$, verify that $\|M\|_2^2 \le \|MM^* + M^*M\|_2$ for every $M \in \mathbf{M}_n(\mathbb{C})$.

b. When k is a positive integer, compute $(AA^* + A^*A)^k$ in closed form. Deduce that

$$\|AA^* + A^*A\|_2 \le 2^{1/k}\|A\|_2^2.$$

c. Passing to the limit as $k \to +\infty$, prove that

$$\|A\|_2 = \|AA^* + A^*A\|_2^{1/2}. \tag{7.7}$$