5

Norms on Linear Spaces

5.1 Introduction

In this chapter we study norms on real or complex linear linear spaces.

Definition 5.1. *Let $\mathcal{F} = (F, \{0, 1, +, -, \cdot\})$ be the real or the complex field. A* semi-norm *on an F-linear space V is a mapping $\nu : V \longrightarrow \mathbb{R}$ that satisfies the following conditions:*
(i) $\nu(\boldsymbol{x} + \boldsymbol{y}) \leq \nu(\boldsymbol{x}) + \nu(\boldsymbol{y})$ *(the subadditivity property of semi-norms), and*
(ii) $\nu(a\boldsymbol{x}) = |a|\nu(\boldsymbol{x})$*,*
for $\boldsymbol{x}, \boldsymbol{y} \in V$ and $a \in F$.

By taking $a = 0$ in the second condition of the definition we have $\nu(\mathbf{0}) = 0$ for every semi-norm on a real or complex space.

A semi-norm can be defined on every linear space. Indeed, if B is a basis of V, $B = \{\mathbf{v}_i \mid i \in I\}$, J is a finite subset of I, and $\mathbf{x} = \sum_{i\in I} x_i \mathbf{v}_i$, define $\nu_J(\mathbf{x})$ as

$$\nu_J(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{x} = \mathbf{0}, \\ \sum_{j\in J} |a_j| & \end{cases}$$

for $\mathbf{x} \in V$. We leave to the reader the verification of the fact that ν_J is indeed a seminorm.

Theorem 5.2. *If V is a real or complex linear space and $\nu : V \longrightarrow \mathbb{R}$ is a semi-norm on V, then*

$$\nu(\boldsymbol{x} - \boldsymbol{y}) \geq |\nu(\boldsymbol{x}) - \nu(\boldsymbol{y})|,$$

for $\boldsymbol{x}, \boldsymbol{y} \in V$.

Proof. We have $\nu(\mathbf{x}) \leq \nu(\mathbf{x} - \mathbf{y}) + \nu(\mathbf{y})$, so

$$\nu(\mathbf{x}) - \nu(\mathbf{y}) \leq \nu(\mathbf{x} - \mathbf{y}). \tag{5.1}$$

Since $\nu(\mathbf{x}-\mathbf{y}) = |-1|\nu(\mathbf{y}-\mathbf{x}) \geq \nu(\mathbf{y}) - \nu(\mathbf{x})$ we have

$$-(\nu(\mathbf{x}) - \nu(\mathbf{y})) \leq \nu(\mathbf{x}) - \nu(\mathbf{y}). \tag{5.2}$$

The Inequalities 5.1 and 5.2 give the desired inequality. □

Corollary 5.3. *If $p : V \longrightarrow \mathbb{R}$ is a semi-norm on V, then $p(\boldsymbol{x}) \geq 0$ for $\boldsymbol{x} \in V$.*

Proof. Choose $\mathbf{y} = \mathbf{0}$ in the inequality of Theorem 5.2 we have $\nu(\mathbf{x}) \geq |\nu(\mathbf{x})| \geq 0$. □

Definition 5.4. *Let $\mathcal{F} = (F, \{0, 1, +, -, \cdot\})$ be the real or the complex field. A* norm *on an F-linear space V is a semi-norm $\nu : V \longrightarrow \mathbb{R}$ such that $\nu(\boldsymbol{x}) = 0$ implies $\boldsymbol{z} = \boldsymbol{0}$ for $\boldsymbol{x} \in V$.*

The pair (V, ν) is referred to as a normed linear space*.*

5.2 Vector Norms on $\mathbb{R}^n$

Lemma 5.5. *Let $p, q \in \mathbb{R} - \{0, 1\}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then we have $p > 1$ if and only if $q > 1$. Furthermore, one of the numbers p, q belongs to the interval $(0, 1)$ if and only if the other number is negative.*

Proof. We leave to the reader the simple proof of this statement. □

Lemma 5.6. *Let $p, q \in \mathbb{R} - \{0, 1\}$ be two numbers such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$. Then, for every $a, b \in \mathbb{R}_{\geq 0}$, we have*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

where the equality holds if and only if $a = b^{-\frac{1}{1-p}}$.

Proof. By Lemma 5.5, we have $q > 1$. Consider the function $f(x) = \frac{x^p}{p} + \frac{1}{q} - x$ for $x \geq 0$. We have $f'(x) = x^{p-1} - 1$, so the minimum is achieved when $x = 1$ and $f(1) = 0$. Thus,

$$f\left(ab^{-\frac{1}{p-1}}\right) \geq f(1) = 0,$$

which amounts to

$$\frac{a^p b^{-\frac{p}{p-1}}}{p} + \frac{1}{q} - ab^{-\frac{1}{p-1}} \geq 0.$$

By multiplying both sides of this inequality by $b^{\frac{p}{p-1}}$, we obtain the desired inequality. □

Observe that if $\frac{1}{p} + \frac{1}{q} = 1$ and $p < 1$, then $q < 0$. In this case, we have the reverse inequality

$$ab \geq \frac{a^p}{p} + \frac{b^q}{q}. \tag{5.3}$$

which can be shown by observing that the function f has a maximum in $x = 1$. The same inequality holds when $q < 1$ and therefore $p < 0$.

Theorem 5.7 (The Hölder Inequality). *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ nonnegative numbers, and let p and q be two numbers such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$. We have*

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} b_i^q \right)^{\frac{1}{q}}.$$

Proof. Define the numbers

$$x_i = \frac{a_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}}} \text{ and } y_i = \frac{b_i}{\left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}}$$

for $1 \leq i \leq n$. Lemma 5.6 applied to x_i, y_i yields

$$\frac{a_i b_i}{\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q\right)^{\frac{1}{q}}} \leq \frac{1}{p} \frac{a_i^p}{\sum_{i=1}^{n} a_i^p} + \frac{1}{q} \frac{b_i^p}{\sum_{i=1}^{n} b_i^p}.$$

Adding these inequalities, we obtain

$$\sum_{i=1}^{n} a_i b_i \leq \left(\sum_{i=1}^{n} a_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} b_i^q \right)^{\frac{1}{q}}$$

because $\frac{1}{p} + \frac{1}{q} = 1$. □

The nonnegativity of the numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$ can be relaxed by using absolute values. Indeed, we can easily prove the following variant of Theorem 5.7.

Theorem 5.8. *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ numbers and let p and q be two numbers such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$. We have*

$$\left| \sum_{i=1}^{n} a_i b_i \right| \leq \left(\sum_{i=1}^{n} |a_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} |b_i|^q \right)^{\frac{1}{q}}.$$

Proof. By Theorem 5.7, we have

$$\sum_{i=1}^{n} |a_i||b_i| \leq \left(\sum_{i=1}^{n} |a_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} |b_i|^q \right)^{\frac{1}{q}}.$$

The needed equality follows from the fact that

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leq \sum_{i=1}^{n} |a_i||b_i|.$$

□

We obtain as a special case the Cauchy-Schwarz inequality that was proven in Theorem 7.3.

Corollary 5.9 (The Cauchy-Schwarz Inequality for $\mathbb{R}^n$). *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ numbers. We have*

$$\left|\sum_{i=1}^{n} a_i b_i\right| \leq \sqrt{\sum_{i=1}^{n} |a_i|^2} \cdot \sqrt{\sum_{i=1}^{n} |b_i|^2}.$$

Proof. The inequality follows immediately from Theorem 5.8 by taking $p = q = 2$. □

Theorem 5.10 (Minkowski's Inequality). *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be $2n$ nonnegative numbers. If $p \geq 1$, we have*

$$\left(\sum_{i=1}^{n} (a_i + b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}.$$

If $p < 1$, the inequality sign is reversed.

Proof. For $p = 1$, the inequality is immediate. Therefore, we can assume that $p > 1$. Note that

$$\sum_{i=1}^{n} (a_i + b_i)^p = \sum_{i=1}^{n} a_i (a_i + b_i)^{p-1} + \sum_{i=1}^{n} b_i (a_i + b_i)^{p-1}.$$

By Hölder's inequality for p, q such that $p > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\begin{aligned}\sum_{i=1}^{n} a_i (a_i + b_i)^{p-1} &\leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} (a_i + b_i)^{(p-1)q}\right)^{\frac{1}{q}} \\ &= \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} (a_i + b_i)^{p}\right)^{\frac{1}{q}}.\end{aligned}$$

Similarly, we can write

$$\sum_{i=1}^{n} b_i (a_i + b_i)^{p-1} \leq \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} (a_i + b_i)^{p}\right)^{\frac{1}{q}}.$$

Adding the last two inequalities yields

$$\sum_{i=1}^{n}(a_i+b_i)^p \leq \left(\left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}\right)\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{q}},$$

which is equivalent to inequality

$$\left(\sum_{i=1}^{n}(a_i+b_i)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} b_i^p\right)^{\frac{1}{p}}.$$

□

Corollary 5.11. *For $p \geq 1$, the function $\nu_p : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}$ defined by*

$$\nu_p(x_1,\ldots,x_n) = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}},$$

where $\boldsymbol{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$, is a norm on the linear space $(\mathbb{R}^n, +, \cdot)$.

Proof. We must prove that ν_p satisfies the conditions of Definition 5.1 and that $\nu_p(\mathbf{x}) = 0$ implies $\mathbf{x} = \mathbf{0}$.

Let $\mathbf{x} = (x_1,\ldots,x_n)$, $\mathbf{y} = (y_1,\ldots,y_n) \in \mathbb{R}^n$. Minkowski's inequality applied to the nonnegative numbers $a_i = |x_i|$ and $b_i = |y_i|$ amounts to

$$\left(\sum_{i=1}^{n}(|x_i|+|y_i|)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}.$$

Since $|x_i + y_i| \leq |x_i| + |y_i|$ for every i, we have

$$\left(\sum_{i=1}^{n}(|x_i+y_i|)^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}},$$

that is, $\nu_p(\mathbf{x}+\mathbf{y}) \leq \nu_p(\mathbf{x}) + \nu_p(\mathbf{y})$. We leave to the reader the verification of the remaining conditions. Thus, ν_p is a norm on $\mathbb{R}^n$. □

Example 5.12. Consider the mappings $\nu_1, \nu_\infty : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$\nu_1(\mathbf{x}) = |x_1| + |x_2| + \cdots + |x_n|,$$
$$\nu_\infty(\mathbf{x}) = \max\{|x_1|, |x_2|, \ldots, |x_n|\},$$

for every $\mathbf{x} = (x_1,\ldots,x_n) \in \mathbb{R}^n$. Both ν_1 and ν_∞ are norms on $\mathbb{R}^n$.

We will frequently use the alternative notation $\| \mathbf{x} \|_p$ for $\nu_p(\mathbf{x})$.

A special norm on $\mathbb{R}^n$ is the function $\nu_\infty : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}$ given by

$$\nu_\infty(\mathbf{x}) = \max\{|x_i| \mid 1 \leq i \leq n\} \tag{5.4}$$

for $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

We verify here that ν_∞ satisfies the first condition of Definition 5.1. We start from the inequality

$$|x_i + y_i| \leq |x_i| + |y_i| \leq \nu_\infty(\mathbf{x}) + \nu_\infty(\mathbf{y})$$

for every i, $1 \leq i \leq n$. This in turn implies

$$\nu_\infty(\mathbf{x} + \mathbf{y}) = \max\{|x_i + y_i| \mid 1 \leq i \leq n\} \leq \nu_\infty(\mathbf{x}) + \nu_\infty(\mathbf{y}),$$

which gives the desired inequality.

This norm can be regarded as a limit case of the norms ν_p. Indeed, let $\mathbf{x} \in \mathbb{R}^n$ and let $M = \max\{|x_i| \mid 1 \leq i \leq n\} = |x_{\ell_1}| = \cdots = |x_{\ell_k}|$ for some $\ell_1, \ldots, \ell_k$, where $1 \leq \ell_1, \ldots, \ell_k \leq n$. Here $x_{\ell_1}, \ldots, x_{\ell_k}$ are the components of $\mathbf{x}$ that have the maximal absolute value and $k \geq 1$. We can write

$$\lim_{p \to \infty} \nu_p(\mathbf{x}) = \lim_{p \to \infty} M \left(\sum_{i=1}^{n} \left(\frac{|x_i|}{M} \right)^p \right)^{\frac{1}{p}} = \lim_{p \to \infty} M(k)^{\frac{1}{p}} = M,$$

which justifies the notation ν_∞.

Example 5.13. For $p \geq 1$, let ℓ_p be the set that consists of sequences of real numbers $\mathbf{x} = (x_0, x_1, \ldots)$ such that the series $\sum_{i=0}^{\infty} |x_i|^p$ is convergent. We can show that ℓ_p is a linear space.

Let $\mathbf{x}, \mathbf{y} \in \ell_p$ be two sequences in ℓ_p. Using Minkowski's inequality we have

$$\sum_{i=0}^{n} |x_i + y_i|^p \leq \sum_{i=0}^{n} (|x_i| + |y_i|)^p \leq \sum_{i=0}^{n} |x_i|^p + \sum_{i=0}^{n} |y_i|^p,$$

which shows that $\mathbf{x} + \mathbf{y} \in \ell_p$. It is immediate that $\mathbf{x} \in \ell_p$ implies $a\mathbf{x} \in \ell_p$ for every $a \in \mathbb{R}$ and $\mathbf{x} \in \ell_p$.

Recall that a *metric* on a set M is a function $d : M \times M \longrightarrow \mathbb{R}_{\geq 0}$ that satisfies the following conditions:

(i) $d(x, y) = 0$ if and only if $x = y$;
(ii) $d(x, y) = d(y, x)$;
(iii) $d(x, y) \leq d(x, z) + d(z, y)$,

for all $x, y, z \in M$. If the first condition is replaced by $d(x, x) = 0$ for every $x \in M$, then d is a *semi-metric* on M.

The following statement shows that any norm defined on a linear space generates a metric on the space.

Theorem 5.14. *Each norm $\nu : V \longrightarrow \mathbb{R}_{\geq 0}$ on a real linear space V generates a metric on the set V defined by $d_\nu(\boldsymbol{x}, \boldsymbol{y}) = \| \boldsymbol{x} - \boldsymbol{y} \|$ for $\boldsymbol{x}, \boldsymbol{y} \in V$.*

Proof. Note that if $d_\nu(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \| = 0$, it follows that $\mathbf{x} - \mathbf{y} = \mathbf{0}$; that is, $\mathbf{x} = \mathbf{y}$.

The symmetry of d_ν is obvious and so we need to verify only the triangular axiom. Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in L$. Applying the third property of Definition **??**, we have

$$\nu(\mathbf{x} - \mathbf{z}) = \nu(\mathbf{x} - \mathbf{y} + \mathbf{y} - \mathbf{z}) \leq \nu(\mathbf{x} - \mathbf{y}) + \nu(\mathbf{y} - \mathbf{z})$$

or, equivalently, $d_\nu(\mathbf{x}, \mathbf{z}) \leq d_\nu(\mathbf{x}, \mathbf{y}) + d_\nu(\mathbf{y}, \mathbf{z})$, for every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in L$, which concludes the argument. □

We refer to d_ν as the *metric induced by the norm ν on the linear space V*.

For $p \geq 1$, then d_p denotes the metric d_{ν_p} induced by the norm ν_p on the linear space $\mathbb{R}^n$ known as the *Minkowski metric* on $\mathbb{R}^n$.

If $p = 2$, we have the *Euclidean metric* on $\mathbb{R}^n$ given by

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

For $p = 1$, we have

$$d_1(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|.$$

A representation of these metrics can be seen in Figure 5.1 for the special case of $\mathbb{R}^2$. If $\mathbf{x} = (x_0, x_1)$ and $\mathbf{y} = (y_0, y_1)$, then $d_2(\mathbf{x}, \mathbf{y})$ is the length of the hypotenuse of the right triangle and $d_1(\mathbf{x}, \mathbf{y})$ is the sum of the lengths of the two legs of the triangle.

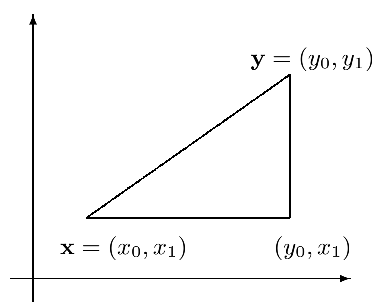

Fig. 5.1. The distances $d_1(\mathbf{x}, \mathbf{y})$ and $d_2(\mathbf{x}, \mathbf{y})$.

Theorem 5.16 to follow allows us to compare the norms ν_p (and the metrics of the form d_p) that were introduced on $\mathbb{R}^n$. We begin with a preliminary result.

Lemma 5.15. *Let $a_1, \ldots, a_n$ be n positive numbers. If p and q are two positive numbers such that $p \leq q$, then*

$$(a_1^p + \cdots + a_n^p)^{\frac{1}{p}} \geq (a_1^q + \cdots + a_n^q)^{\frac{1}{q}}.$$

Proof. Let $f : \mathbb{R}^{>0} \longrightarrow \mathbb{R}$ be the function defined by

$$f(r) = (a_1^r + \cdots + a_n^r)^{\frac{1}{r}}.$$

Since

$$\ln f(r) = \frac{\ln (a_1^r + \cdots + a_n^r)}{r},$$

it follows that

$$\frac{f'(r)}{f(r)} = -\frac{1}{r^2}(a_1^r + \cdots + a_n^r) + \frac{1}{r} \cdot \frac{a_1^r \ln a_1 + \cdots + a_n^r \ln a_r}{a_1^r + \cdots + a_n^r}.$$

To prove that $f'(r) < 0$, it suffices to show that

$$\frac{a_1^r \ln a_1 + \cdots + a_n^r \ln a_r}{a_1^r + \cdots + a_n^r} \leq \frac{\ln (a_1^r + \cdots + a_n^r)}{r}.$$

This last inequality is easily seen to be equivalent to

$$\sum_{i=1}^{n} \frac{a_i^r}{a_1^r + \cdots + a_n^r} \ln \frac{a_i^r}{a_1^r + \cdots + a_n^r} \leq 0,$$

which holds because

$$\frac{a_i^r}{a_1^r + \cdots + a_n^r} \leq 1$$

for $1 \leq i \leq n$. □

Theorem 5.16. *Let p and q be two positive numbers such that $p \leq q$. For every $\boldsymbol{u} \in \mathbb{R}^n$, we have $\| \boldsymbol{u} \|_p \geq \| \boldsymbol{u} \|_q$.*

Proof. This statement follows immediately from Lemma 5.15. □

Corollary 5.17. *Let p, q be two positive numbers such that $p \leq q$. For every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have $d_p(\boldsymbol{x}, \boldsymbol{y}) \geq d_q(\boldsymbol{x}, \boldsymbol{y})$.*

Proof. This statement follows immediately from Theorem 5.16. □

Example 5.18. For $p = 1$ and $q = 2$ the inequality of Theorem 5.16 becomes

$$\sum_{i=1}^{n} |u_i| \leq \sqrt{\sum_{i=1}^{n} |u_i|^2},$$

which is equivalent to

$$\frac{\sum_{i=1}^{n} |u_i|}{n} \leq \sqrt{\frac{\sum_{i=1}^{n} |u_i|^2}{n}}. \tag{5.5}$$

Theorem 5.19. *Let $p \geq 1$. For every $\boldsymbol{x} \in \mathbb{R}^n$ we have*

$$\| \boldsymbol{x} \|_\infty \leq \| \boldsymbol{x} \|_p \leq n \| \boldsymbol{x} \|_\infty .$$

Proof. The first inequality is an immediate consequence of Theorem 5.16. The second inequality follows by observing that

$$\| \mathbf{x} \|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \leq n \max_{1 \leq i \leq n} |x_i| = n \| \mathbf{x} \|_\infty .$$

□

Corollary 5.20. *Let p and q be two numbers such that $p, q \geq 1$. There exist two constants $c, d \in \mathbb{R}_{>0}$ such that*

$$c \| \boldsymbol{x} \|_q \leq \| \boldsymbol{x} \|_p \leq d \| \boldsymbol{x} \|_q$$

for $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. Since $\| \mathbf{x} \|_\infty \leq \| \mathbf{x} \|_p$ and $\| x \|_q \leq n \| \mathbf{x} \|_\infty$, it follows that $\| x \|_q \leq n \| \mathbf{x} \|_p$. Exchanging the roles of p and q, we have $\| x \|_p \leq n \| \mathbf{x} \|_q$, so

$$\frac{1}{n} \| \mathbf{x} \|_q \leq \| \mathbf{x} \|_p \leq n \| \mathbf{x} \|_q$$

for every $\mathbf{x} \in \mathbb{R}^n$. □

Corollary 5.21. *For every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $p \geq 1$, we have $d_\infty(\boldsymbol{x}, \boldsymbol{y}) \leq d_p(\boldsymbol{x}, \boldsymbol{y}) \leq n d_\infty(\boldsymbol{x}, \boldsymbol{y})$. Further, for $p, q > 1$, there exist $c, d \in \mathbb{R}_{>0}$ such that*

$$c d_q(\boldsymbol{x}, \boldsymbol{y}) \leq d_p(\boldsymbol{x}, \boldsymbol{y}) \leq c d_q(\boldsymbol{x}, \boldsymbol{y})$$

for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

Proof. This follows from Theorem 5.19 and from Corollary 5.21. □

Corollary 5.17 implies that if $p \leq q$, then the closed sphere $B_{d_p}(\mathbf{x}, r)$ is included in the closed sphere $B_{d_q}(\mathbf{x}, r)$. For example, we have

$$B_{d_1}(\mathbf{0}, 1) \subseteq B_{d_2}(\mathbf{0}, 1) \subseteq B_{d_\infty}(\mathbf{0}, 1).$$

In Figures 5.2 (a) - (c) we represent the closed spheres $B_{d_1}(\mathbf{0}, 1)$, $B_{d_2}(\mathbf{0}, 1)$, and $B_{d_\infty}(\mathbf{0}, 1)$.

An useful consequence of Theorem 5.7 is the following statement:

Theorem 5.22. *Let $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$ be $2m$ nonnegative numbers such that $\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i = 1$ and let p and q be two positive numbers such that $\frac{1}{p} + \frac{1}{q} = 1$. We have*

$$\sum_{j=1}^{m} x_j^{\frac{1}{p}} y_j^{\frac{1}{q}} \leq 1.$$

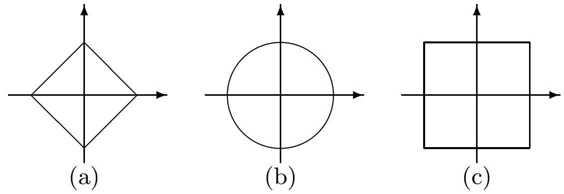

Fig. 5.2. Spheres $B_{d_p}(\mathbf{0}, 1)$ for $p = 1, 2, \infty$.

Proof. The Hölder inequality applied to $x_1^{\frac{1}{p}}, \ldots, x_m^{\frac{1}{p}}$ and $y_1^{\frac{1}{q}}, \ldots, y_m^{\frac{1}{q}}$ yields the needed inequality

$$\sum_{j=1}^{m} x_j^{\frac{1}{p}} y_j^{\frac{1}{q}} \leq \sum_{j=1}^{m} x_j \sum_{j=1}^{m} y_j = 1$$

□

Theorem 5.22 allows the formulation of a generalization of the Hölder Inequality.

Theorem 5.23. *Let A be an $n \times m$ matrix, $A = (a_{ij})$, having positive entries such that $\sum_{j=1}^{m} a_{ij} = 1$ for $1 \leq i \leq n$. If $\mathbf{p} = (p_1, \ldots, p_n)$ is an n-tuple of positive numbers such that $\sum_{i=1}^{n} p_i = 1$, then*

$$\sum_{j=1}^{m} \prod_{i=1}^{n} a_{ij}^{p_i} \leq 1.$$

Proof. The argument is by induction on $n \geq 2$. The basis case, $n = 2$ follows immediately from Theorem 5.22 by choosing $p = \frac{1}{p_1}$, $q = \frac{1}{p_2}$, $x_j = a_{1j}$, and $y_j = a_{2j}$ for $1 \leq j \leq m$.

Suppose that the statement holds for n, let A be an $(n+1) \times m$-matrix having positive entries such that $\sum_{j=1}^{m} a_{ij} = 1$ for $1 \leq i \leq n+1$, and let $\mathbf{p} = (p_1, \ldots, p_n, p_{n+1})$ be such that $p_1 + \cdots + p_n + p_{n+1} = 1$.

It is easy to see that

$$\sum_{j=1}^{m} \prod_{i=1}^{n+1} a_{ij}^{p_i} \leq \sum_{j=1}^{m} a_{1j}^{p_1} a_{n-1\,j}^{p_{n-1}} (a_{nj} + a_{n+1\,j})^{p_n + p_{n+1}}.$$

By applying the inductive hypothesis, we have

$$\sum_{j=1}^{m} \prod_{i=1}^{n+1} a_{ij}^{p_i} \leq 1.$$

□

A more general form of Theorem 5.23 is given next.

Theorem 5.24. *Let A be an $n \times m$ matrix, $A = (a_{ij})$, having positive entries. If $\boldsymbol{p} = (p_1, \ldots, p_n)$ is an n-tuple of positive numbers such that $\sum_{i=1}^n p_i = 1$, then*

$$\sum_{j=1}^{m} \prod_{i=1}^{n} a_{ij}^{p_i} \leq \prod_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} \right)^{p_i}.$$

Proof. Let $B = (b_{ij})$ be the matrix defined by

$$b_{ij} = \frac{a_{ij}}{\sum_{j=1}^m a_{ij}}$$

for $1 \leq i \leq n$ and $1 \leq j \leq m$. Since $\sum_{j=1}^m b_{ij} = 1$, we can apply Theorem 5.23 to this matrix. Thus, we can write

$$\begin{aligned}
\sum_{j=1}^{m} \prod_{i=1}^{n} b_{ij}^{p_i} &= \sum_{j=1}^{m} \prod_{i=1}^{n} \left(\frac{a_{ij}}{\sum_{j=1}^m a_{ij}} \right)^{p_i} \\
&= \sum_{j=1}^{m} \prod_{i=1}^{n} \frac{a_{ij}^{p_i}}{\left(\sum_{j=1}^m a_{ij}\right)^{p_i}} \\
&= \frac{\sum_{j=1}^m \prod_{i=1}^n a_{ij}^{p_i}}{\prod_{i=1}^n \left(\sum_{j=1}^m a_{ij}\right)^{p_i}} \leq 1.
\end{aligned}$$

□

We now give a generalization of Minkowski's inequality (Theorem 5.10). First, we need a preliminary result.

Lemma 5.25. *If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are positive numbers and $r < 0$, then*

$$\sum_{i=1}^{n} a_i^r b_i^{1-r} \geq \left(\sum_{i=1}^{n} a_i \right)^{r} \cdot \left(\sum_{i=1}^{n} b_i \right)^{1-r}.$$

Proof. Let $c_1, \ldots, c_n, d_1, \ldots, d_n$ be $2n$ positive numbers such that $\sum_{i=1}^n c_i = \sum_{i=1}^n d_i = 1$. Inequality (5.3) applied to the numbers $a = c_i^{\frac{1}{p}}$ and $b = d_i^{\frac{1}{q}}$ yields:

$$c_i^{\frac{1}{p}} d_i^{\frac{1}{q}} \geq \frac{c_i}{p} + \frac{d_i}{q}.$$

Summing these inequalities produces the inequality

$$\sum_{i=1}^{n} c_i^{\frac{1}{p}} d_i^{\frac{1}{q}} \geq 1,$$

or

$$\sum_{i=1}^{n} c_i^r d_i^{1-r} \geq 1,$$

where $r = \frac{1}{p} < 0$. Choosing $c_i = \frac{a_i}{\sum_{i=1}^n a_i}$ and $d_i = \frac{b_i}{\sum_{i=1}^n b_i}$, we obtain the desired inequality. □

Theorem 5.26. *Let A be an $n \times m$ matrix, $A = (a_{ij})$, having positive entries, and let p and q be two numbers such that $p > q$ and $p \neq 0$, $q \neq 0$. We have*

$$\left(\sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ij}^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \geq \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij}^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}}.$$

Proof. Define

$$E = \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ij}^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}},$$

$$F = \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij}^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}},$$

and $u_i = \sum_{j=1}^m a_{ij}^q$ for $1 \leq i \leq n$.

There are three distinct cases to consider related to the position of 0 relative to p and q.

Suppose initially that $p > q > 0$. We have

$$\begin{aligned} F^p &= \textstyle\sum_{i=1}^n u_i^{\frac{p}{q}} &&= \textstyle\sum_{i=1}^n u_i u_i^{\frac{p}{q}-1} \\ &= \textstyle\sum_{i=1}^n \sum_{j=1}^m a_{ij}^q u_i^{\frac{p}{q}-1} &&= \textstyle\sum_{j=1}^m \sum_{i=1}^n a_{ij}^q u_i^{\frac{p}{q}-1}. \end{aligned}$$

By applying the Hölder inequality, we have

$$\begin{aligned} \sum_{i=1}^{n} a_{ij}^q u_i^{\frac{p}{q}-1} &\leq \left(\sum_{i=1}^{n} (a_{ij}^q)^{\frac{p}{q}} \right)^{\frac{q}{p}} \cdot \left(\sum_{i=1}^{n} (u_i^{\frac{p}{q}-1})^{\frac{p}{p-q}} \right)^{1-\frac{q}{p}} \qquad (5.6) \\ &= \left(\sum_{i=1}^{n} a_{ij}^p \right)^{\frac{q}{p}} \cdot \left(\sum_{i=1}^{n} u_i^{\frac{p}{q}} \right)^{1-\frac{q}{p}}, \end{aligned}$$

which implies $F^p \leq E^q F^{p-q}$. This, in turn, gives $F^q \leq E^q$, which implies the generalized Minkowski inequality.

Suppose now that $0 > p > q$, so $0 < -p < -q$. Applying the generalized Minkowski inequality to the positive numbers $b_{ij} = \frac{1}{a_{ij}}$ gives the inequality

$$\left(\sum_{j=1}^{m}\left(\sum_{i=1}^{n} b_{ij}^{-q}\right)^{\frac{p}{q}}\right)^{-\frac{1}{p}} \geq \left(\sum_{i=1}^{n}\left(\sum_{j=1}^{m} b_{ij}^{-p}\right)^{\frac{q}{p}}\right)^{-\frac{1}{q}},$$

which is equivalent to

$$\left(\sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{q}\right)^{\frac{p}{q}}\right)^{-\frac{1}{p}} \geq \left(\sum_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij}^{p}\right)^{\frac{q}{p}}\right)^{-\frac{1}{q}}.$$

A last transformation gives

$$\left(\sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ij}^{q}\right)^{\frac{p}{q}}\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij}^{p}\right)^{\frac{q}{p}}\right)^{\frac{1}{q}},$$

which is the inequality to be proven.

Finally, suppose that $p > 0 > q$. Since $\frac{q}{p} < 0$, Inequality (5.7) is replaced by the opposite inequality through the application of Lemma 5.25:

$$\sum_{i=1}^{n} a_{ij}^{q} u_i^{\frac{p}{q}-1} \geq \left(\sum_{i=1}^{n} a_{ij}^{p}\right)^{\frac{q}{p}} \cdot \left(\sum_{i=1}^{n} u_i^{\frac{p}{q}}\right)^{1-\frac{q}{p}}.$$

This leads to $F^p \geq E^q F^{p-q}$ or $F^q \geq E^q$. Since $q < 0$, this implies $F \leq E$. □

5.3 Matrix Norms

In Chapter 3 we saw that the set $F^{m\times n}$ is a linear space. Therefore, it is natural to consider norms defined on matrices. In the case of matrices we need to consider a supplementary condition that defines the interaction between norms and matrix multiplication.

Definition 5.27. *A* matrix norm *is a family of functions* $\mu^{(m,n)} : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}_{\geq 0}$*, where* $m, n \in \mathbb{P}$ *that satisfies the following conditions:*
(i) $\mu^{(m,n)}(A) = 0$ *if and only if* $A = O_{m,n}$*;*
(ii) $\mu^{(m,n)}(A+B) \leq \mu^{(m,n)}(A) + \mu^{(m,n)}(B)$ *(the* subadditivity property*), and*
(iii) $\mu^{(m,n)}(aA) = |a|\mu^{(m,n)}(A)$*,*
(iv) $\mu^{(m,p)}(AB) \leq \mu^{(m,n)}(A)\mu^{(n,p)}(B)$ *for every matrix* $A \in \mathbb{R}^{m\times n}$ *and* $B \in \mathbb{R}^{n\times p}$ *(the* submultiplicative property*).*

If the format of the matrix A is clear from context or is irrelevant, then we shall write $\mu(A)$ instead of $\mu^{(m,n)}(A)$ or just $|||A|||$.

If the last condition of Definition 5.27 is dropped we obtain generalized matrix norms which are, essentially, vector norms obtained by transforming matrices into vectors and, then, using vector norms.

Definition 5.28. *The* $(m \times n)$-vectorization mapping *is the mapping* $\mathsf{vec} : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^{mn}$ *defined by*

$$\mathsf{vec}(A) = (a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn}),$$

obtained by reading A *column-wise.*

The vectorization mapping vec is an isomorphism between the linear space $F^{m\times n}$ and the linear space F^{mn}, as the reader can easily verify.

Definition 5.29. *Let* ν *be a vector norm on the space* $\mathbb{R}^{mn}$*. The* generalized matrix norm $\mu^{(m,n)}$ on $\mathbb{R}^{m\times n}$ *is the mapping* $\mu^{(m,n)} : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}_{\geq 0}$ *defined by*

$$\mu^{(m,n)}(A) = \nu(\mathsf{vec}(A)),$$

for $A \in \mathbb{R}^{m\times n}$*.*

Some generalized matrix norms turn out to be actual matrix norms; others fail to be matrix norms. This point is illustrated by the next two examples.

Example 5.30. Consider the generalized matrix norm μ_1 induced by the vector norm ν_1. We have $\mu_1(A) = \sum_{i=1}^{n}\sum_{j=1}^{m} |a_{ij}|$ for $A \in \mathbb{R}^{m\times n}$. Actually, this is a matrix norm. To prove this fact consider the matrices $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$. We have

$$\begin{aligned}
\mu_1(AB) &= \sum_{i=1}^{m}\sum_{j=1}^{n}\left|\sum_{k=1}^{p} a_{ik}b_{kj}\right| \leq \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p} |a_{ik}b_{kj}| \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k'=1}^{p}\sum_{k''=1}^{p} |a_{ik'}||b_{k''j}| \\
&\quad \text{(because we added extra non-negative terms to the sums)} \\
&= \left(\sum_{i=1}^{m}\sum_{k'=1}^{p} |a_{ik'}|\right)\cdot\left(\sum_{j=1}^{n}\sum_{k''=1}^{p} |b_{k''j}|\right) \\
&= \mu_1(A)\mu_1(B).
\end{aligned}$$

We will denote this matrix norm by the same notation as the corresponding vector norm, that is, by $\| A \|_1$.

The generalized norm μ_2 induced by the vector norm ν_2 is also a matrix norm. Indeed, using the notations as above we have:

$$\begin{aligned}
(\mu_2(AB))^2 &= \sum_{i=1}^{m}\sum_{j=1}^{n}\left|\sum_{k=1}^{p} a_{ik}b_{kj}\right|^2 \\
&\leq \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\sum_{k=1}^{p} |a_{ik}|^2\right)\left(\sum_{l=1}^{p} |b_{lj}|^2\right) \\
&\quad \text{(by Cauchy-Schwarz inequality)} \\
&\leq (\mu_2(A))^2(\mu_2(B))^2.
\end{aligned}$$

The matrix norm μ_2, denoted by $\| A \|_2$, is known as the *Frobenius norm.* It is easy to see that

$$\| A \|_2 = \mathsf{trace}(A'A). \tag{5.7}$$

Example 5.31. The generalized norm μ_∞ induced by the vector norm ν_∞, denoted by $\| A \|_\infty$ is not a matrix norm. To see that, let a, b be two positive numbers and consider the matrices

$$A = \begin{pmatrix} a & a \\ a & a \end{pmatrix} \text{ and } B = \begin{pmatrix} b & b \\ b & b \end{pmatrix}.$$

Clearly, we have $\| A \|_\infty = a$ and $\| B \|_2 = b$. However, since

$$AB = \begin{pmatrix} 2ab & 2ab \\ 2ab & 2ab \end{pmatrix},$$

we have $\| AB \|_\infty = 2ab$ and the submultiplicative propery of matrix norms is violated.

A technique that always produces matrix norms starting from vector norms is introduced in the next theorem.

Theorem 5.32. *Let ν be a vector norm on $\mathbb{R}^n$. Then, the the function $\mu : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}_{\geq 0}$ defined as*

$$\mu(A) = \sup\{\nu(A\boldsymbol{x}) \mid \nu(\boldsymbol{x}) \leq 1\},$$

for $A \in \mathbb{R}^{m\times n}$, is a matrix norm.

Proof. We need to verify that μ satisfies the conditions of Definition 5.27.

It is easy to see that if we exclude the zero vector from the definition of $\mu(A)$ the value of $\mu(A)$ remains the same; in other words, $\mu(A) = \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1, \mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}\}$.

If $\mu(A) = 0$, this means that $\nu(A\mathbf{x}) = 0$ for every $\mathbf{x}$ such that $\nu(\mathbf{x}) \leq 1$; In other words, we have $A\mathbf{x} = \mathbf{0}$ for every $\mathbf{x} \in \mathbb{R}^n$. and this implies $A = O_{m,n}$. Since $\mu(O_{m,n}) = 0$, the first condition is satisfied.

If $A, B \in \mathbb{R}^{m\times n}$, then

$$\begin{aligned}
\mu(A + B) &= \sup\{\nu(A + B)\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\} \\
&= \sup\{\nu(A\mathbf{x} + B\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} \\
&\leq \sup\{\nu(A\mathbf{x}) + \nu(B\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} \\
&\leq \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} + \sup\{\nu(B\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} \\
&= \mu(A) + \mu(B),
\end{aligned}$$

which shows that the second condition of Definition 5.27 is satisfied.

Let $a \in \mathbb{R}$. We can write

$$\begin{aligned}\mu(aA) &= \sup\{\nu(aA)\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\}\\ &= \sup\{|a|\nu(A\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\}\\ &= |a| \sup\{\nu(A\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\} = |a|\mu(A),\end{aligned}$$

which means that the third condition is satisfied.

Finally, by the first part of Supplement 4 we have:

$$\begin{aligned}\mu(AB) &= \sup\{\nu((AB)\mathbf{x}) \mid \nu(\mathbf{x}) = 1\}\\ &= \sup\{\nu(A(B\mathbf{x})) \mid \nu(\mathbf{x}) = 1\}\\ &= \sup\left\{\nu\left(A\frac{B\mathbf{x}}{\nu(B\mathbf{x})}\right)\nu(B\mathbf{x}) \middle| \nu(\mathbf{x}) = 1\right\}\\ &\leq \sup \mu(A)\{\nu(B\mathbf{x}) \middle| \nu(\mathbf{x}) = 1\}\\ &= \mu(A)\mu(B),\end{aligned}$$

because

$$\nu\left(\frac{B\mathbf{x}}{\nu(B\mathbf{x})}\right) = 1.$$

□

By Corollary 2.153, we can replace sup by max in the definition of the matrix norm induced by a vector norm, $\mu(A) = \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\}$, because the closed sphere $B(\mathbf{0}, 1)$ is a compact set in $\mathbb{R}^n$. Thus, we have

$$\mu(A) = \max\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\}. \tag{5.8}$$

An equivalent definition of the matrix norm induced by a vector norm is given next.

Theorem 5.33. *Let ν be a vector norm on $\mathbb{R}^n$ and let μ be the matrix norm induced by ν, as in Theorem 5.32. Then,*

$$\mu(A) = \max\{\nu(A\boldsymbol{x}) \mid \nu(\boldsymbol{x}) = 1\},$$

for $A \in \mathbb{R}^{m\times n}$.

Proof. Since $\{\mathbf{x} \mid \nu(\mathbf{x}) = 1\} \subseteq \{\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\}$ it follows that

$$\max\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\} \leq \max\{\nu A\mathbf{x} \mid \nu(\mathbf{x}) \leq 1\} = \mu(A).$$

There exist $\mathbf{z} \in \mathbb{R}^n - \{\mathbf{0}\}$ such that $\nu(\mathbf{z}) \leq 1$ and $\nu(A\mathbf{z}) = \mu(A)$, which implies

$$\mu(A) = \nu(\mathbf{z})\nu\left(A\frac{\mathbf{z}}{\nu(\mathbf{z})}\right) \leq \nu\left(A\frac{\mathbf{z}}{\nu(\mathbf{z})}\right).$$

Since $\nu\left(\frac{\mathbf{z}}{\nu(\mathbf{z})}\right) = 1$ it follows that $\mu(A) \leq \max\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\}$. This yields the desired conclusion. □

If μ is a matrix norm induced by a vector norm on $\mathbb{R}^n$, then $\mu(I_n) = \sup\{\nu(I_n\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} = 1$. This necessary condition can be used for identifying matrix norms that are not induced by vector norms.

The matrix norm induced by the vector norm $\|\cdot\|_p$ will be denoted by $|||\cdot|||_p$.

Example 5.34. To compute the matrix norm $|||A|||_1 = \sup\{\| A\mathbf{x} \|_1 \mid \| \mathbf{x} \|_1 \leq 1\}$, where $A \in \mathbb{R}^{n\times n}$, suppose that the columns of A are the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Let $\mathbf{x} \in \mathbb{R}^n$ be a vector whose components are $x_1, \ldots, x_n$. Then, $\mathbf{Ax} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$, so

$$\begin{aligned}\| A\mathbf{x} \|_1 &= \| x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n \|_1\\ &\leq \sum_{i=1}^{n} |x_i| \, \| \mathbf{a}_i \|_1\\ &\leq \max_i \| \mathbf{a}_i \|_1 \sum_{i=1}^{n} |x_i|\\ &= \max_i \| \mathbf{a}_i \|_1 \cdot \| \mathbf{x} \|_1 .\end{aligned}$$

Thus, $|||A|||_1 \leq \max_i \| \mathbf{a}_i \|_1$.

Let $\mathbf{e}_i$ be the vector whose components are 0 with the exception of its i^{th} component that is equal to 1. Clearly, we have $\| \mathbf{e}_i \|_1 = 1$ and $\mathbf{a}_i = A\mathbf{e}_i$. This, in turn implies $\| \mathbf{a}_i \|_1 = \| A\mathbf{e}_i \|_1 \leq |||A|||_1$ for $1 \leq i \leq n$. Therefore, $\max_i \| \mathbf{a}_i \|_1 \leq |||A|||_1$, so

$$|||A|||_1 = \max_i \| \mathbf{a}_i \|_1 = \max_i \sum_{j=1}^{n} |a_{ij}|.$$

In other words, $|||A|||_1$ equals the maximum column sum of the absolute values.

Example 5.35. Consider now a matrix $A \in \mathbb{R}^{n\times n}$.

$$|||A|||_\infty = \sup\{\| A\mathbf{x} \|_\infty \mid \| \mathbf{x} \|_\infty \leq 1\}.$$

We have

$$\begin{aligned}\| A\mathbf{x} \|_\infty &= \max_{1\leq i\leq n} \left| \sum_{j=1}^{n} a_{ij}x_j \right|\\ &\leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}x_j|\\ &\leq \max_{1\leq i\leq n} \| \mathbf{x} \|_\infty \sum_{j=1}^{n} |a_{ij}|.\end{aligned}$$

Consequently, if $\| \mathbf{x} \|_\infty \leq 1$ we have $\| A\mathbf{x} \|_\infty \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$. Thus, $\|\!| A |\!\|_\infty \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$.

The converse inequality is immediate if $A = O_{n,n}$. Therefore, assume that $A \neq O_{n\times n}$, and let $(a_{p1}, \ldots, a_{pn})$ be any row of A that has at least one element distinct from 0. Define the vector $\mathbf{z} \in \mathbb{R}^n$ by

$$z_j = \begin{cases} \frac{|a_{pj}|}{a_{pj}} & \text{if } a_{pj} \neq 0, \\ 1 & \text{otherwise,} \end{cases}$$

for $1 \leq j \leq n$. It is clear that $z_j \in \{-1, 1\}$ for every j, $1 \leq j \leq n$ and, therefore, $\| \mathbf{z} \|_\infty = 1$. Moreover, we have $|a_{pj}| = a_{pj} z_j$ for $1 \leq j \leq n$. Therefore, we can write:

$$\begin{aligned} \sum_{j=1}^{n} |a_{pj}| = \sum_{j=1}^{n} a_{pj} z_j &= \left| \sum_{j=1}^{n} a_{pj} z_j \right| \\ &\leq \max_{1\leq i\leq n} \left| \sum_{j=1}^{n} a_{ij} z_j \right| \\ &= \| A\mathbf{z} \|_\infty \leq \sup\{\| A\mathbf{x} \|_\infty \mid \| \mathbf{x} \|_\infty \leq 1\} \leq \|\!| A |\!\|_\infty. \end{aligned}$$

Since this holds for every row of A, it follows that $\max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \leq \|\!| A |\!\|_\infty$, which proves that

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

In other words, $\|\!| A |\!\|_\infty$ equals the maximum row sum of the absolute values.

5.4 The Topology of Normed Linear Spaces

We saw that every norm $\| \cdot \|$ defined on a linear space V generates a metric $d : V^2 \longrightarrow \mathbb{R}_{\geq 0}$ given by $d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \|$. Therefore, by Theorem 2.93, any normed space can be equipped with the topology of a metric space, using the metric defined by the norm. Since this topology is induced by a metric. any normed space is a Hausdorff space by Theorem 2.112. Further, if $\mathbf{v} \in V$, then the collection of subsets $\{C(\mathbf{v}, r) \mid r > 0\}$ is a fundamental system of neighborhoods for $\mathbf{v}$.

A sequence $(\mathbf{x}_0, \mathbf{x}_1, \ldots)$ of elements of V converges to $\mathbf{x}$ if for every $\epsilon > 0$ there exists $n_\epsilon \in \mathbb{N}$ such that $n \geq n_\epsilon$ implies $\| \mathbf{x}_n - \mathbf{x} \| < \epsilon$.

Theorem 5.36. *In a normed linear space $(V, \| \cdot \|)$, the norm, the multiplication by scalars and the vector addition are continuous functions.*

Proof. By Theorem 5.2, we have $\parallel \mathbf{x} - \mathbf{y} \parallel \geq | \parallel \mathbf{x} \parallel - \parallel \mathbf{y} \parallel |$ for every $\mathbf{x}, \mathbf{y} \in V$. Therefore, if $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$, we have $\parallel \mathbf{x}_n - \mathbf{x} \parallel \geq | \parallel \mathbf{x}_n \parallel - \parallel \mathbf{x} \parallel |$, which implies $\lim_{n\to\infty} \parallel \mathbf{x}_n \parallel = \parallel \mathbf{x} \parallel$. Thus, by Theorem 2.128, the norm is continuous.

Suppose now that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$, where (a_n) is a sequence of scalars. Since the sequence $(\mathbf{x}_n)$ is bounded, we have

$$\begin{aligned}\parallel a\mathbf{x} - a_n\mathbf{x}_n \parallel &\leq \parallel a\mathbf{x} - a_n\mathbf{x} \parallel + \parallel a_n\mathbf{x} - a_n\mathbf{x}_n \parallel \\ &\leq |a - a_n| \parallel \mathbf{x} \parallel + a_n \parallel \mathbf{x} - \mathbf{x}_n \parallel,\end{aligned}$$

which implies that $\lim_{n\to\infty} a_n\mathbf{x}_n = a\mathbf{x}$. This shows that the multiplication by scalars is a continuous function.

To prove that the vector addition is continuous, let $(\mathbf{x}_n)$ and $(\mathbf{y}_n)$ be two sequences in V such that $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$ and $\lim_{n\to\infty} \mathbf{y}_n = \mathbf{y}$. Note that

$$\parallel (\mathbf{x} + \mathbf{y}) - (\mathbf{x}_n + \mathbf{y}_n) \parallel \leq \parallel \mathbf{x} - \mathbf{x}_n \parallel + \parallel \mathbf{y} - \mathbf{y}_n \parallel,$$

which implies that $\lim_{n\to\infty}(\mathbf{x}_n + \mathbf{y}_n) = \mathbf{x} + \mathbf{y}$. Thus, the vector addition is continuous. □

Definition 5.37. *Two norms ν_0 and ν_1 on a linear space V are* equivalent *if they generate the same topology.*

Theorem 5.38. *Let V be a linear space and let $\nu_0 : V \longrightarrow \mathbb{R}_{\geq 0}$ and $\nu_1 : V \longrightarrow \mathbb{R}_{\geq 0}$ be two norms on V that generate the topologies $\mathcal{O}_0$ and $\mathcal{O}_1$ on V, respectively.*

The topology $\mathcal{O}_1$ is finer than the topology $\mathcal{O}_0$ (that is, $\mathcal{O}_0 \subseteq \mathcal{O}_1$) if and only if there exists $c \in \mathbb{R}_{>0}$ such that $\nu_0(\boldsymbol{v}) \leq c\nu_1(\boldsymbol{v})$ for every $\boldsymbol{v} \in V$.

Proof. Suppose that $\mathcal{O}_0 \subseteq \mathcal{O}_1$. Then, any open sphere $C_0(\mathbf{0}, r_0) = \{\mathbf{x} \in V \mid \nu_0(\mathbf{x}) < r_0\}$ (in $\mathcal{O}_0$) must be an open set in $\mathcal{O}_1$. Therefore, there exists an open sphere $C_1(\mathbf{0}, r_1)$ such that $C_1(\mathbf{0}, r_1) \subseteq C_0(\mathbf{0}, r_0)$. This means that for $r_0 \in \mathbb{R}_{\geq 0}$ and $\mathbf{v} \in V$ there exists $r_1 \in \mathbb{R}_{\geq 0}$ such that $\nu_1(\mathbf{v}) < r_1$ implies $\nu_0(\mathbf{v}) < r_0$ for every $\mathbf{u} \in V$. In particular, for $r_0 = 1$, there is $k > 0$ such that $\nu_1(\mathbf{v}) < k$ implies $\nu_0(\mathbf{v}) < 1$, which is equivalent to

$$c\nu_1(\mathbf{v}) < 1 \text{ implies } \nu_0(\mathbf{v}) < 1,$$

for every $\mathbf{v} \in V$ and $c = \frac{1}{k}$.

For $\mathbf{w} = \frac{1}{c+\epsilon}\frac{\mathbf{v}}{\nu_1(\mathbf{v})}$, where $\epsilon > 0$ it follows that

$$c\nu_1(\mathbf{w}) = c\nu_1\left(\frac{1}{c+\epsilon}\frac{\mathbf{v}}{\nu_1(\mathbf{v})}\right) = \frac{c}{c+\epsilon} < 1,$$

so

$$\nu_0(\mathbf{w}) = \nu_0\left(\frac{1}{c+\epsilon}\frac{\mathbf{v}}{\nu_1(\mathbf{v})}\right) = \frac{1}{c+\epsilon}\frac{\nu_0(\mathbf{v})}{\nu_1(\mathbf{v})} < 1.$$

Since this inequality holds for every $\epsilon > 0$ it follows that $\nu_0(\mathbf{v}) \leq c\nu_1(\mathbf{v})$.

Conversely, suppose that there exists $c \in \mathbb{R}_{>0}$ such that $\nu_0(\mathbf{v}) \leq c\nu_1(\mathbf{v})$ for every $\mathbf{v} \in V$. Since

$$\left\{\mathbf{v} \mid \nu_1(\mathbf{v}) \leq \frac{r}{c}\right\} \subseteq \{\mathbf{v} \mid \nu_0(\mathbf{v}) \leq r\},$$

for $\mathbf{v} \in V$ and $r > 0$ it follows that $\mathcal{O}_0 \subseteq \mathcal{O}_1$. □

Corollary 5.39. *Let V be a linear space and let $\nu_0 : V \longrightarrow \mathbb{R}_{\geq 0}$ and $\nu_1 : V \longrightarrow \mathbb{R}_{\geq 0}$ be two norms on V. Then, ν_0 and ν_1 are equivalent norms if and only if there exist $a, b \in \mathbb{R}_{>0}$ such that $a\nu_0(\boldsymbol{v}) \leq \nu_1(\boldsymbol{v}) \leq b\nu_0(\boldsymbol{v})$.*

Proof. This statement follows directly from Theorem 5.38. □

Example 5.40. By Corollary 5.20 any two norms ν_p and ν_q, on $\mathbb{R}^n$ (with $p, q \geq 1$) are equivalent.

Continuous linear functions between normed spaces have a simple characterization.

Theorem 5.41. *Let (V_1, ν_1) and (V_2, ν_2) be two normed F-linear spaces where F is either $\mathbb{R}$ or $\mathbb{C}$. A linear function $f : V_1 \longrightarrow V_2$ is continuous if and only if there exists $M \in \mathbb{R}_{>0}$ such that $\nu_2(f(\boldsymbol{x})) \leq M\nu_1(\boldsymbol{x})$ for every $\boldsymbol{x} \in V_1$.*

Proof. Suppose that $f : V_1 \longrightarrow V_2$ satisfies the condition of the theorem. Then,

$$f\left(C\left(\mathbf{0}_1, \frac{r}{M}\right)\right) \subseteq C(\mathbf{0}_2, r),$$

for every $r > 0$, which means that f is continuous in $\mathbf{0}_1$ and, therefore, it is continuous everywhere (by Theorem 2.161).

Conversely, suppose that f is continuous. Then, there exists $\delta > 0$ such that $f(C(\mathbf{0}_1, \delta)) \subseteq C(f(x), 1)$, which is equivalent to saying that $\nu_1(\mathbf{x}) < \delta$ implies $\nu_2(f(\mathbf{x})) < 1$. Let $\epsilon > 0$ and let $\mathbf{z} \in V_1$ be defined by

$$\mathbf{z} = \frac{\delta}{\nu_1(\mathbf{x}) + \epsilon}\mathbf{x}.$$

We have $\nu_1(\mathbf{z}) = \frac{\delta\nu_1(\mathbf{x})}{\nu_1(\mathbf{x})+\epsilon} < \delta$. This implies $\nu_2(f(\mathbf{z})) < 1$, which is equivalent to

$$\frac{\delta}{\nu_1(\mathbf{x}) + \epsilon}\nu_2(f(\mathbf{x})) < 1$$

because of the linearity of f. This means that

$$\nu_2(f(\mathbf{x})) < \frac{\nu_1(\mathbf{x}) + \epsilon}{\delta}$$

for every $\epsilon > 0$, so $\nu_2(f(\mathbf{x})) \leq \frac{1}{\delta}\nu_1(\mathbf{x})$. □

Definition 5.42. *A* Banach space *is a normed linear space that is a complete metric space with respect to the metric induced by the norm.*

5.5 Matrix Sequences and Matrix Series

The set of matrices $\mathbb{C}^{m\times p}$ is a $\mathbb{C}$-linear space, and the set of matrices $\mathbb{R}^{m\times p}$ is an $\mathbb{R}$-linear spaces. Using vector norms or matrix form, these spaces can be equipped with a topological structure, as we indicated above.

We focus now on the normed linear space $(\mathbb{R}^{p\times p}, \|\!| \cdot \|\!|)$, where $\|\!| \cdot \|\!|)$ is a matrix norm.

Let $A \in \mathbb{R}^{p\times p}$. We can show by induction on n that

$$\|\!| A^n \|\!| \leq (\|\!| A \|\!|)^n. \tag{5.9}$$

The base step, $n = 0$, is immediate. Suppose that the inequality holds for n. We have

$$\begin{aligned}
\|\!| A^{n+1} \|\!| &= \|\!| A^n A \|\!| \\
&\leq \|\!| A^n \|\!| \|\!| A \|\!| \\
&\quad \text{(because } \|\!| \cdot \|\!| \text{ is a matrix norm)} \\
&\leq (\|\!| A \|\!|)^n \|\!| A \|\!| \\
&\quad \text{(by the inductive hypothesis)} \\
&= (\|\!| A \|\!|)^{n+1},
\end{aligned}$$

which concludes our argument.

If $\|\!| A \|\!| < 1$ the sequence of matrices $(A, A^2, \ldots, A^n, \ldots)$ converges towards the zero matrix $O_{p,p}$. Indeed, $\lim_{n\to\infty} \|\!| A^n - O_{p,p} \|\!| = \lim_{n\to\infty} \|\!| A^n \|\!| \leq \lim_{n\to\infty} (\|\!| A \|\!|)^n = 0$, which shows that $\lim_{n\to\infty} A^n = O_{p,p}$.

Definition 5.43. *Let* $\boldsymbol{A} = (A_0, A_1, \ldots, A_n, \ldots)$ *be a sequence of matrices in* $\mathbb{R}^{p\times p}$*. A* matrix series *having* $\boldsymbol{A}$ *as its sequence of terms is the sequence of matrices* $(S_0, S_1, \ldots, S_n, \ldots)$*, where* $S_i = \sum k = 0^i A_k$*.*

The series $(S_0, S_1, \ldots, S_n, \ldots)$ *will be denoted also by* $A_0 + A_1 + \cdots + A_n + \cdots$*.*

We say that the series $A_0 + A_1 + \cdots + A_n + \cdots$ *converges to a matrix* S *if* $\lim_{n\to\infty} S_n = S$*. This is also denoted by* $A_0 + A_1 + \cdots + A_n + \cdots = S$*.*

The subadditivity property of the norm can be generalized to series of matrices. Namely, if the series $A_0 + A_1 + \cdots + A_n + \cdots$ converges to S, then

$$\|\!| S \|\!| \leq \sum_{i=0}^{\infty} \|\!| A_i \|\!|.$$

Indeed, by the usual subadditivity property

$$\|\!| A_0 + A_1 + \cdots + A_n \|\!| \geq \sum_{i=0}^{n} \|\!| A_i \|\!|.$$

This implies

$$|||A_0 + A_1 + \cdots + A_n||| \geq \sum_{i=0}^{\infty} |||A_i|||,$$

for every $n \in \mathbb{N}$, so $|||S||| \geq \sum_{i=0}^{\infty} |||A_i|||$, due to the continuity of the matrix norm.

Example 5.44. Let $A \in \mathbb{R}^{p \times p}$ be a matrix such that $|||A||| < 1$. We claim that the matrix $I - A$ is invertible and $A^0 + A^1 + \cdots + A^n + \cdots = (I - A)^{-1}$.

Suppose that $I - A$ is not invertible. Then, the system $(I - A)\mathbf{x} = \mathbf{0}$ has a non-trivial solution. This implies $\mathbf{x} = A\mathbf{x}$, so $|||A||| \geq 1$, which contradicts the hypothesis. Thus, $I - A$ is an invertible matrix.

Observe that

$$(A^0 + A^1 + \cdots + A^n)(I - A)^{-1} = I - A^{n+1}.$$

Therefore,

$$\left(\lim_{n \to \infty} (A^0 + A^1 + \cdots + A^n)\right)(I - A)^{-1} = \lim_{n \to \infty} (I - A^{n+1}) = I,$$

since $\lim_{n \to \infty} A^{n+1} = O$. This shows that the series $A^0 + A^1 + \cdots + A^n + \cdots$ converges to the inverse of the matrix $I - A$, so

$$(I - A)^{-1} = \sum_{i=0}^{\infty} A^i.$$

Moreover, we have

$$\begin{aligned} |||(I - A)^{-1}||| &= |||\sum_{i=0}^{\infty} A^i||| \\ &\leq \sum_{i=0}^{\infty} |||A^i||| \\ &= \sum_{i=0}^{\infty} (|||A|||)^i \\ &= \frac{1}{1 - |||A|||}, \end{aligned}$$

because $|||A||| \leq 1$.

Exercises and Supplements

1. Prove that if $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are three vectors in a real vector space V and ν is a norm on V, then $\nu\mathbf{x} - \mathbf{y} \leq \nu\mathbf{x} - \mathbf{z} + \nu\mathbf{z} - \mathbf{y}$.

2. Let $A \in \mathbb{R}^{n\times n}$ (or in $\mathbb{C}^{n\times n}$) be a non-singular matrix. Prove that if ν is a norm on $\mathbb{R}^n$ (on $\mathbb{C}^n$, respectively), then ν_A defined by $\nu_A(\mathbf{x}) = \nu(A\mathbf{x})$ is a norm on $\mathbb{R}^n$ (on $\mathbb{C}^n$).
3. Two vector norms ν_p and ν_q on $\mathbb{R}^n$ are *conjugate* if $\frac{1}{p} + \frac{1}{q} = 1$. Prove that
$$|\mathbf{x}'\mathbf{y}| \leq \nu_p(\mathbf{x}) \cdot \nu_q(\mathbf{y})$$
for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
4. Let $A \in \mathbb{R}^{m\times n}$ and let ν be a vector norm on $\mathbb{R}^n$. Prove that if $A \in \mathbb{R}^{m\times n}$, then we have the following equalities:
$$\begin{aligned}\mu(A) &= \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\}\\ &= \sup\left\{\frac{\nu(A\mathbf{x})}{\nu(\mathbf{x})}\middle| \mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}\right\}\\ &= \inf\{k \mid \nu(A\mathbf{x}) \leq k\nu(\mathbf{x}),\ \text{for every } \mathbf{x} \in \mathbb{R}^n\}.\end{aligned}$$

Solution: To prove the first equality note that
$$\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\} \subseteq \{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\}.$$
This implies
$$\sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\} \leq \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} = \mu(A).$$
On the other hand, let $\mathbf{x}$ be a vector such that $\nu(\mathbf{x}) \leq 1$. We have $\mathbf{x} = \mathbf{0}$ if and only if $\nu(\mathbf{x}) = 0$ because ν is a norm. Otherwise, $\mathbf{x} \neq 0$, and for $\mathbf{y} = \frac{1}{\nu(\mathbf{x})}\mathbf{x}$ we have $\nu(\mathbf{y}) = 1$ and $\nu(A\mathbf{x}) = \nu(A(\nu(\mathbf{x})\mathbf{y}) = \nu(\mathbf{x})\nu(A\mathbf{y}) \leq \nu(A\mathbf{y})$. Therefore, in either case we have
$$\nu(A\mathbf{x}) \leq \sup\{\nu(A\mathbf{y}) \mid \nu(\mathbf{y}) = 1\}$$
for $\nu(\mathbf{x}) \leq 1$. Thus, we have the reverse inequality,
$$\sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) \leq 1\} \leq \sup\{\nu(A\mathbf{y}) \mid \nu(\mathbf{y}) = 1\},$$
so
$$\mu(A) = \sup\{\nu(A\mathbf{x}) \mid \nu(\mathbf{x}) = 1\}.$$
To prove the second equality observe that
$$\begin{aligned}\sup\left\{\frac{\nu(A\mathbf{x})}{\nu(\mathbf{x})}\middle| \mathbf{x} \neq \mathbf{0}\right\} &= \sup\left\{\nu(A\left(\frac{\mathbf{x}}{\nu(\mathbf{x})}\right)\middle| \mathbf{x} \neq \mathbf{0}\right\}\\ &= \sup\{\nu(A\mathbf{y}) \mid \nu(\mathbf{y}) = 1\} = \mu(A),\end{aligned}$$
because $\nu\left(\frac{\mathbf{x}}{\nu(\mathbf{x})}\right) = 1$ and every vector $\mathbf{y}$ with $\nu(\mathbf{y}) = 1$ can be written as $\mathbf{y} = \frac{1}{\nu(\mathbf{x})}\mathbf{x}$ for some $\mathbf{x} \neq \mathbf{0}$.
We leave the third equality to the reader.

Bibliographical Comments