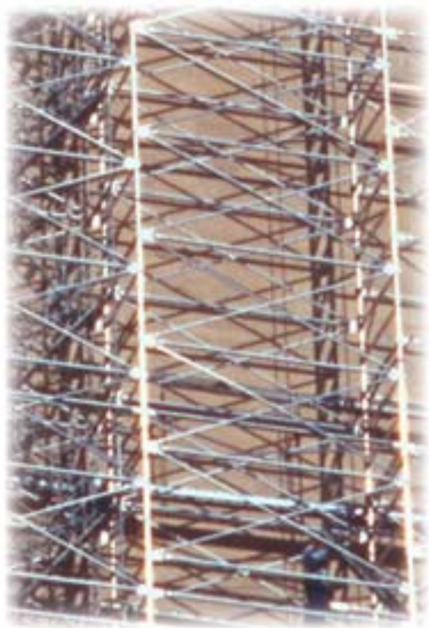


CHAPTER 5

Norms, Inner Products, and Orthogonality

5.1 VECTOR NORMS

A significant portion of linear algebra is in fact geometric in nature because much of the subject grew out of the need to generalize the basic geometry of $\Re^2$ and $\Re^3$ to nonvisual higher-dimensional spaces. The usual approach is to coordinatize geometric concepts in $\Re^2$ and $\Re^3$, and then extend statements concerning ordered pairs and triples to ordered n-tuples in $\Re^n$ and $\mathcal{C}^n$.

For example, the length of a vector $\mathbf{u} \in \Re^2$ or $\mathbf{v} \in \Re^3$ is obtained from the Pythagorean theorem by computing the length of the hypotenuse of a right triangle as shown in Figure 5.1.1.

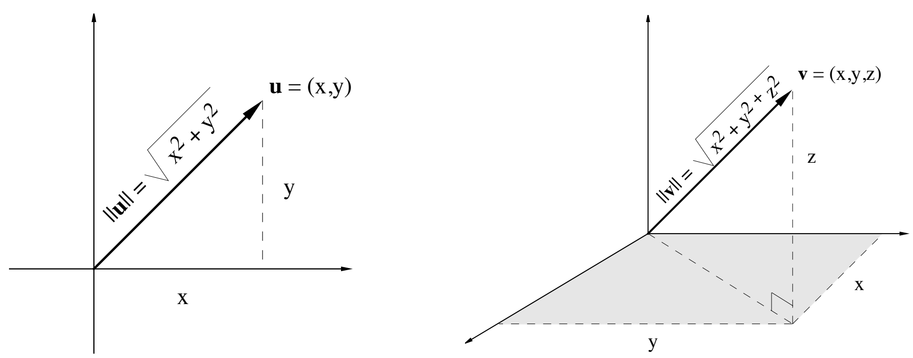


FIGURE 5.1.1

This measure of length,

$$\|\mathbf{u}\| = \sqrt{x^2 + y^2} \quad \text{and} \quad \|\mathbf{v}\| = \sqrt{x^2 + y^2 + z^2},$$

is called the *euclidean norm* in $\Re^2$ and $\Re^3$, and there is an obvious extension to higher dimensions.

Euclidean Vector Norm

For a vector $\mathbf{x}_{n\times 1}$, the ***euclidean norm*** of $\mathbf{x}$ is defined to be

- $\|\mathbf{x}\| = \left(\sum_{i=1}^n x_i^2\right)^{1/2} = \sqrt{\mathbf{x}^T\mathbf{x}}$ whenever $\mathbf{x} \in \Re^{n\times 1}$,
- $\|\mathbf{x}\| = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} = \sqrt{\mathbf{x}^*\mathbf{x}}$ whenever $\mathbf{x} \in \mathcal{C}^{n\times 1}$.

For example, if $\mathbf{u} = \begin{pmatrix} 0 \\ -1 \\ 2 \\ -2 \\ 4 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} \mathrm{i} \\ 2 \\ 1-\mathrm{i} \\ 0 \\ 1+\mathrm{i} \end{pmatrix}$, then

$$\|\mathbf{u}\| = \sqrt{\sum u_i^2} = \sqrt{\mathbf{u}^T\mathbf{u}} = \sqrt{0+1+4+4+16} = 5,$$

$$\|\mathbf{v}\| = \sqrt{\sum |v_i|^2} = \sqrt{\mathbf{v}^*\mathbf{v}} = \sqrt{1+4+2+0+2} = 3.$$

There are several points to note.[33]

- The complex version of $\|\mathbf{x}\|$ includes the real version as a special case because $|z|^2 = z^2$ whenever z is a real number. Recall that if $z = a + \mathrm{i}b$, then $\bar{z} = a - \mathrm{i}b$, and the magnitude of z is $|z| = \sqrt{\bar{z}z} = \sqrt{a^2+b^2}$. The fact that $|z|^2 = \bar{z}z = a^2 + b^2$ is a real number insures that $\|\mathbf{x}\|$ is real even if $\mathbf{x}$ has some complex components.
- The definition of euclidean norm guarantees that for all scalars α,

$$\|\mathbf{x}\| \geq 0, \quad \|\mathbf{x}\| = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}, \quad \text{and} \quad \|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|. \tag{5.1.1}$$

- Given a vector $\mathbf{x} \neq \mathbf{0}$, it's frequently convenient to have another vector that points in the same direction as $\mathbf{x}$ (i.e., is a positive multiple of $\mathbf{x}$) but has unit length. To construct such a vector, we ***normalize*** $\mathbf{x}$ by setting $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$. From (5.1.1), it's easy to see that

$$\|\mathbf{u}\| = \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = \frac{1}{\|\mathbf{x}\|}\|\mathbf{x}\| = 1. \tag{5.1.2}$$

---

[33] By convention, column vectors are used throughout this chapter. But there is nothing special about columns because, with the appropriate interpretation, all statements concerning columns will also hold for rows.

- The distance between vectors in $\Re^3$ can be visualized with the aid of the parallelogram law as shown in Figure 5.1.2, so for vectors in $\Re^n$ and $\mathcal{C}^n$, the ***distance*** between $\mathbf{u}$ and $\mathbf{v}$ is naturally defined to be $\|\mathbf{u}-\mathbf{v}\|$.

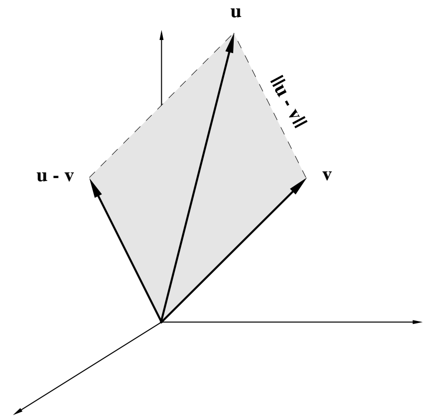


FIGURE 5.1.2

Standard Inner Product

The scalar terms defined by

$$\mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i \in \Re \quad \text{and} \quad \mathbf{x}^*\mathbf{y} = \sum_{i=1}^{n} \bar{x}_i y_i \in \mathcal{C}$$

are called the ***standard inner products*** for $\Re^n$ and $\mathcal{C}^n$, respectively.

The Cauchy–Bunyakovskii–Schwarz (CBS) inequality [34] is one of the most important inequalities in mathematics. It relates inner product to norm.

---

[34] The Cauchy–Bunyakovskii–Schwarz inequality is named in honor of the three men who played a role in its development. The basic inequality for real numbers is attributed to Cauchy in 1821, whereas Schwarz and Bunyakovskii contributed by later formulating useful generalizations of the inequality involving integrals of functions.

Augustin-Louis Cauchy (1789–1857) was a French mathematician who is generally regarded as being the founder of mathematical analysis—including the theory of complex functions. Although deeply embroiled in political turmoil for much of his life (he was a partisan of the Bourbons), Cauchy emerged as one of the most prolific mathematicians of all time. He authored at least 789 mathematical papers, and his collected works fill 27 volumes—this is on a par with Cayley and second only to Euler. It is said that more theorems, concepts, and methods bear Cauchy's name than any other mathematician.

Victor Bunyakovskii (1804–1889) was a Russian professor of mathematics at St. Petersburg, and in 1859 he extended Cauchy's inequality for discrete sums to integrals of continuous functions. His contribution was overlooked by western mathematicians for many years, and his name is often omitted in classical texts that simply refer to the *Cauchy–Schwarz inequality.*

Hermann Amandus Schwarz (1843–1921) was a student and successor of the famous German mathematician Karl Weierstrass at the University of Berlin. Schwarz independently generalized Cauchy's inequality just as Bunyakovskii had done earlier.

Cauchy–Bunyakovskii–Schwarz (CBS) Inequality

$$|\mathbf{x}^*\mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\| \quad \text{for all } \mathbf{x}, \mathbf{y} \in \mathcal{C}^{n\times 1}. \tag{5.1.3}$$

Equality holds if and only if $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha = \mathbf{x}^*\mathbf{y}/\mathbf{x}^*\mathbf{x}$.

Proof. Set $\alpha = \mathbf{x}^*\mathbf{y}/\mathbf{x}^*\mathbf{x} = \mathbf{x}^*\mathbf{y}/\|\mathbf{x}\|^2$ (assume $\mathbf{x} \neq \mathbf{0}$ because there is nothing to prove if $\mathbf{x} = \mathbf{0}$) and observe that $\mathbf{x}^*(\alpha\mathbf{x} - \mathbf{y}) = 0$, so

$$\begin{aligned} 0 \leq \|\alpha\mathbf{x} - \mathbf{y}\|^2 &= (\alpha\mathbf{x} - \mathbf{y})^*(\alpha\mathbf{x} - \mathbf{y}) = \bar{\alpha}\mathbf{x}^*(\alpha\mathbf{x} - \mathbf{y}) - \mathbf{y}^*(\alpha\mathbf{x} - \mathbf{y}) \\ &= -\mathbf{y}^*(\alpha\mathbf{x} - \mathbf{y}) = \mathbf{y}^*\mathbf{y} - \alpha\mathbf{y}^*\mathbf{x} = \frac{\|\mathbf{y}\|^2\|\mathbf{x}\|^2 - (\mathbf{x}^*\mathbf{y})(\mathbf{y}^*\mathbf{x})}{\|\mathbf{x}\|^2}. \end{aligned} \tag{5.1.4}$$

Since $\mathbf{y}^*\mathbf{x} = \overline{\mathbf{x}^*\mathbf{y}}$, it follows that $(\mathbf{x}^*\mathbf{y})(\mathbf{y}^*\mathbf{x}) = |\mathbf{x}^*\mathbf{y}|^2$, so

$$0 \leq \frac{\|\mathbf{y}\|^2\|\mathbf{x}\|^2 - |\mathbf{x}^*\mathbf{y}|^2}{\|\mathbf{x}\|^2}.$$

Now, $0 < \|\mathbf{x}\|^2$ implies $0 \leq \|\mathbf{y}\|^2\|\mathbf{x}\|^2 - |\mathbf{x}^*\mathbf{y}|^2$, and thus the CBS inequality is obtained. Establishing the conditions for equality is Exercise 5.1.9. ∎

One reason that the CBS inequality is important is because it helps to establish that the geometry in higher-dimensional spaces is consistent with the geometry in the visual spaces $\Re^2$ and $\Re^3$. In particular, consider the situation depicted in Figure 5.1.3.

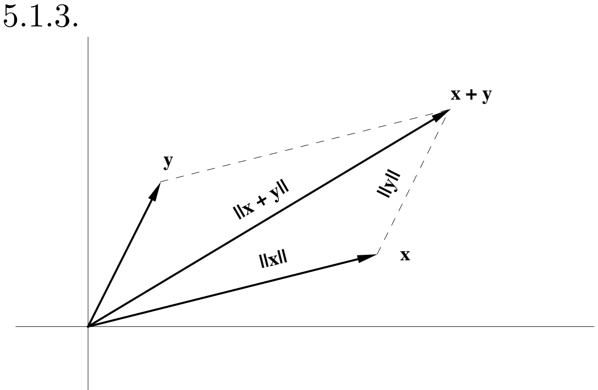


FIGURE 5.1.3

Imagine traveling from the origin to the point $\mathbf{x}$ and then moving from $\mathbf{x}$ to the point $\mathbf{x}+\mathbf{y}$. Clearly, you have traveled a distance that is at least as great as the direct distance from the origin to $\mathbf{x}+\mathbf{y}$ along the diagonal of the parallelogram. In other words, it's visually evident that $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. This observation is known as the ***triangle inequality.*** In higher-dimensional spaces we do not have the luxury of visualizing the geometry with our eyes, and the question of whether or not the triangle inequality remains valid has no obvious answer. The CBS inequality is precisely what is required to prove that, in this respect, the geometry of higher dimensions is no different than that of the visual spaces.

Triangle Inequality

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad \text{for every } \mathbf{x}, \mathbf{y} \in \mathcal{C}^n.$$

Proof. Consider $\mathbf{x}$ and $\mathbf{y}$ to be column vectors, and write

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})^*(\mathbf{x}+\mathbf{y}) = \mathbf{x}^*\mathbf{x} + \mathbf{x}^*\mathbf{y} + \mathbf{y}^*\mathbf{x} + \mathbf{y}^*\mathbf{y} \\ &= \|\mathbf{x}\|^2 + \mathbf{x}^*\mathbf{y} + \mathbf{y}^*\mathbf{x} + \|\mathbf{y}\|^2. \end{aligned} \tag{5.1.5}$$

Recall that if $z = a + ib$, then $z + \bar{z} = 2a = 2\,\mathrm{Re}\,(z)$ and $|z|^2 = a^2 + b^2 \geq a^2$, so that $|z| \geq \mathrm{Re}\,(z)$. Using the fact that $\mathbf{y}^*\mathbf{x} = \overline{\mathbf{x}^*\mathbf{y}}$ together with the CBS inequality yields

$$\mathbf{x}^*\mathbf{y} + \mathbf{y}^*\mathbf{x} = 2\,\mathrm{Re}\,(\mathbf{x}^*\mathbf{y}) \leq 2\,|\mathbf{x}^*\mathbf{y}| \leq 2\,\|\mathbf{x}\|\,\|\mathbf{y}\|.$$

Consequently, we may infer from (5.1.5) that

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2. \quad \blacksquare$$

It's not difficult to see that the triangle inequality can be extended to any number of vectors in the sense that $\left\|\sum_i \mathbf{x}_i\right\| \leq \sum_i \|\mathbf{x}_i\|$. Furthermore, it follows as a corollary that for real or complex numbers, $\left|\sum_i \alpha_i\right| \leq \sum_i |\alpha_i|$ (the triangle inequality for scalars).

Example 5.1.1

Backward Triangle Inequality. The triangle inequality produces an upper bound for a sum, but it also yields the following lower bound for a difference:

$$\big|\, \|\mathbf{x}\| - \|\mathbf{y}\| \,\big| \leq \|\mathbf{x}-\mathbf{y}\|. \tag{5.1.6}$$

This is a consequence of the triangle inequality because

$$\|\mathbf{x}\| = \|\mathbf{x}-\mathbf{y}+\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}\| \implies \|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x}-\mathbf{y}\|$$

and

$$\|\mathbf{y}\| = \|\mathbf{x}-\mathbf{y}-\mathbf{x}\| \leq \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{x}\| \implies -(\|\mathbf{x}\| - \|\mathbf{y}\|) \leq \|\mathbf{x}-\mathbf{y}\|.$$

There are notions of length other than the euclidean measure. For example, urban dwellers navigate on a grid of city blocks with one-way streets, so they are prone to measure distances in the city not as the crow flies but rather in terms of lengths on a directed grid. For example, instead of than saying that "it's a one-half mile straight-line (euclidean) trip from here to there," they are more apt to describe the length of the trip by saying, "it's two blocks north on Dan Allen Drive, four blocks west on Hillsborough Street, and five blocks south on Gorman Street." In other words, the length of the trip is $2+|-4|+|-5|=11$ blocks—absolute value is used to insure that southerly and westerly movement does not cancel the effect of northerly and easterly movement, respectively. This "grid norm" is better known as the 1-norm because it is a special case of a more general class of norms defined below.

p-Norms

For $p \geq 1$, the ***p-norm*** of $\mathbf{x} \in \mathcal{C}^n$ is defined as $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$.

It can be proven that the following properties of the euclidean norm are in fact valid for all p-norms:

$$\begin{aligned} &\|\mathbf{x}\|_p \geq 0 \quad \text{and} \quad \|\mathbf{x}\|_p = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}, \\ &\|\alpha\mathbf{x}\|_p = |\alpha|\ \|\mathbf{x}\|_p \quad \text{for all scalars } \alpha, \\ &\|\mathbf{x}+\mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \qquad \text{(see Exercise 5.1.13).} \end{aligned} \tag{5.1.7}$$

The generalized version of the CBS inequality (5.1.3) for p-norms is ***Hölder's inequality*** (developed in Exercise 5.1.12), which states that if $p>1$ and $q>1$ are real numbers such that $1/p+1/q=1$, then

$$|\mathbf{x}^*\mathbf{y}| \leq \|\mathbf{x}\|_p\ \|\mathbf{y}\|_q\,. \tag{5.1.8}$$

In practice, only three of the p-norms are used, and they are

$$\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \quad \text{(the grid norm)}, \quad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \text{(the euclidean norm)},$$

and

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p = \lim_{p\to\infty}\left(\sum_{i=1}^n |x_i|^p\right)^{1/p} = \max_i |x_i| \qquad \text{(the max norm).}$$

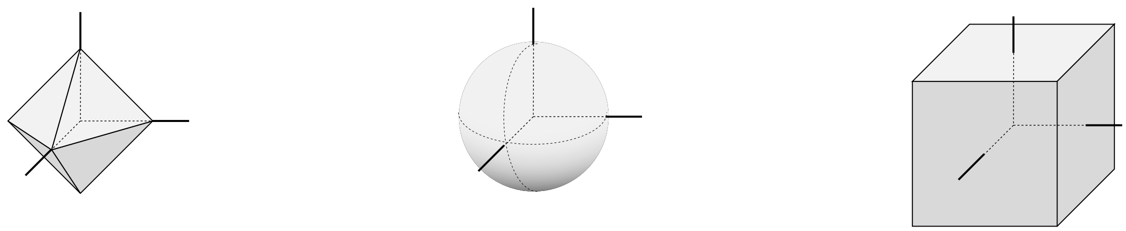
For example, if $\mathbf{x} = (3,\ 4-3\mathrm{i},\ 1)$, then $\|\mathbf{x}\|_1 = 9$, $\|\mathbf{x}\|_2 = \sqrt{35}$, and $\|\mathbf{x}\|_\infty = 5$.

To see that $\lim_{p\to\infty} \|\mathbf{x}\|_p = \max_i |x_i|$, proceed as follows. Relabel the entries of $\mathbf{x}$ by setting $\tilde{x}_1 = \max_i |x_i|$, and if there are other entries with this same maximal magnitude, label them $\tilde{x}_2, \ldots, \tilde{x}_k$. Label any remaining coordinates as $\tilde{x}_{k+1} \cdots \tilde{x}_n$. Consequently, $|\tilde{x}_i/\tilde{x}_1| < 1$ for $i = k+1, \ldots, n$, so, as $p \to \infty$,

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |\tilde{x}_i|^p\right)^{1/p} = |\tilde{x}_1| \left(k + \left|\frac{\tilde{x}_{k+1}}{\tilde{x}_1}\right|^p + \cdots + \left|\frac{\tilde{x}_n}{\tilde{x}_1}\right|^p\right)^{1/p} \to |\tilde{x}_1|.$$

Example 5.1.2

To get a feel for the 1-, 2-, and ∞-norms, it helps to know the shapes and relative sizes of the ***unit p-spheres*** $\mathcal{S}_p = \{\mathbf{x} \mid \|\mathbf{x}\|_p = 1\}$ for $p = 1, 2, \infty$. As illustrated in Figure 5.1.4, the unit 1-, 2-, and ∞-spheres in $\Re^3$ are an octahedron, a ball, and a cube, respectively, and it's visually evident that $\mathcal{S}_1$ fits inside $\mathcal{S}_2$, which in turn fits inside $\mathcal{S}_\infty$. This means that $\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2 \geq \|\mathbf{x}\|_\infty$ for all $\mathbf{x} \in \Re^3$. In general, this is true in $\Re^n$ (Exercise 5.1.8).



$\mathcal{S}_1$ $\mathcal{S}_2$ $\mathcal{S}_\infty$

FIGURE 5.1.4

Because the p-norms are defined in terms of coordinates, their use is limited to coordinate spaces. But it's desirable to have a general notion of norm that works for *all* vector spaces. In other words, we need a coordinate-free definition of norm that includes the standard p-norms as a special case. Since all of the p-norms satisfy the properties (5.1.7), it's natural to use these properties to extend the concept of norm to general vector spaces.

General Vector Norms

A ***norm*** for a real or complex vector space $\mathcal{V}$ is a function $\|\star\|$ mapping $\mathcal{V}$ into $\Re$ that satisfies the following conditions.

$$\begin{aligned} &\|\mathbf{x}\| \geq 0 \quad \text{and} \quad \|\mathbf{x}\| = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}, \\ &\|\alpha\mathbf{x}\| = |\alpha|\ \|\mathbf{x}\| \quad \text{for all scalars } \alpha, \\ &\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|. \end{aligned} \tag{5.1.9}$$

Example 5.1.3

Equivalent Norms. Vector norms are basic tools for defining and analyzing limiting behavior in vector spaces $\mathcal{V}$. A sequence $\{\mathbf{x}_k\} \subset \mathcal{V}$ is said to converge to $\mathbf{x}$ (write $\mathbf{x}_k \to \mathbf{x}$) if $\|\mathbf{x}_k - \mathbf{x}\| \to 0$. This depends on the choice of the norm, so, ostensibly, we might have $\mathbf{x}_k \to \mathbf{x}$ with one norm but not with another. Fortunately, this is impossible in finite-dimensional spaces because all norms are *equivalent* in the following sense.

Problem: For each pair of norms, $\|\star\|_a$, $\|\star\|_b$, on an n-dimensional space $\mathcal{V}$, exhibit positive constants α and β (depending only on the norms) such that

$$\alpha \leq \frac{\|\mathbf{x}\|_a}{\|\mathbf{x}\|_b} \leq \beta \quad \text{for all nonzero vectors in } \mathcal{V}. \tag{5.1.10}$$

Solution: For $\mathcal{S}_b = \{\mathbf{y} \mid \|\mathbf{y}\|_b = 1\}$, let $\mu = \min_{\mathbf{y}\in\mathcal{S}_b} \|\mathbf{y}\|_a > 0$, [35] and write

$$\frac{\mathbf{x}}{\|\mathbf{x}\|_b} \in \mathcal{S}_b \implies \|\mathbf{x}\|_a = \|\mathbf{x}\|_b \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_b} \right\|_a \geq \|\mathbf{x}\|_b \min_{\mathbf{y}\in\mathcal{S}_b} \|\mathbf{y}\|_a = \|\mathbf{x}\|_b \, \mu.$$

The same argument shows there is a $\nu > 0$ such that $\|\mathbf{x}\|_b \geq \nu \|\mathbf{x}\|_a$, so (5.1.10) is produced with $\alpha = \mu$ and $\beta = 1/\nu$. Note that (5.1.10) insures that $\|\mathbf{x}_k - \mathbf{x}\|_a \to 0$ if and only if $\|\mathbf{x}_k - \mathbf{x}\|_b \to 0$. Specific values for α and β are given in Exercises 5.1.8 and 5.12.3.

Exercises for section 5.1

5.1.1. Find the 1-, 2-, and ∞-norms of $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \\ -4 \\ -2 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} 1+\mathrm{i} \\ 1-\mathrm{i} \\ 1 \\ 4\mathrm{i} \end{pmatrix}$.

5.1.2. Consider the euclidean norm with $\mathbf{u} = \begin{pmatrix} 2 \\ 1 \\ -4 \\ -2 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$.

(a) Determine the distance between $\mathbf{u}$ and $\mathbf{v}$.
(b) Verify that the triangle inequality holds for $\mathbf{u}$ and $\mathbf{v}$.
(c) Verify that the CBS inequality holds for $\mathbf{u}$ and $\mathbf{v}$.

5.1.3. Show that $(\alpha_1 + \alpha_2 + \cdots + \alpha_n)^2 \leq n\left(\alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2\right)$ for $\alpha_i \in \Re$.

[35] An important theorem from analysis states that a continuous function mapping a closed and bounded subset $\mathcal{K} \subset \mathcal{V}$ into $\Re$ attains a minimum and maximum value at points in $\mathcal{K}$. Unit spheres in finite-dimensional spaces are closed and bounded, and every norm on $\mathcal{V}$ is continuous (Exercise 5.1.7), so this minimum is guaranteed to exist.

5.1.4. (a) Using the euclidean norm, describe the solid ball in $\Re^n$ centered at the origin with unit radius. (b) Describe a solid ball centered at the point $\mathbf{c} = (\xi_1 \quad \xi_2 \quad \cdots \quad \xi_n)$ with radius ρ.

5.1.5. If $\mathbf{x}, \mathbf{y} \in \Re^n$ such that $\|\mathbf{x} - \mathbf{y}\|_2 = \|\mathbf{x} + \mathbf{y}\|_2$, what is $\mathbf{x}^T\mathbf{y}$?

5.1.6. Explain why $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$ is true for all norms.

5.1.7. For every vector norm on $\mathcal{C}^n$, prove that $\|\mathbf{v}\|$ depends continuously on the components of $\mathbf{v}$ in the sense that for each $\epsilon > 0$, there corresponds a $\delta > 0$ such that $\big|\, \|\mathbf{x}\| - \|\mathbf{y}\| \,\big| < \epsilon$ whenever $|x_i - y_i| < \delta$ for each i.

5.1.8. (a) For $\mathbf{x} \in \mathcal{C}^{n \times 1}$, explain why $\|\mathbf{x}\|_1 \geq \|\mathbf{x}\|_2 \geq \|\mathbf{x}\|_\infty$.

(b) For $\mathbf{x} \in \mathcal{C}^{n \times 1}$, show that $\|\mathbf{x}\|_i \leq \alpha \|\mathbf{x}\|_j$, where α is the (i, j)-entry in the following matrix. (See Exercise 5.12.3 for a similar statement regarding matrix norms.)

$$\begin{array}{c} \\ 1 \\ 2 \\ \infty \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & \infty \end{array} \\ \begin{pmatrix} * & \sqrt{n} & n \\ 1 & * & \sqrt{n} \\ 1 & 1 & * \end{pmatrix} \end{array}.$$

5.1.9. For $\mathbf{x}, \mathbf{y} \in \mathcal{C}^n$, $\mathbf{x} \neq \mathbf{0}$, explain why equality holds in the CBS inequality if and only if $\mathbf{y} = \alpha\mathbf{x}$, where $\alpha = \mathbf{x}^*\mathbf{y}/\mathbf{x}^*\mathbf{x}$. **Hint:** Use (5.1.4).

5.1.10. For nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathcal{C}^n$ with the euclidean norm, prove that equality holds in the triangle inequality if and only if $\mathbf{y} = \alpha\mathbf{x}$, where α is real and positive. **Hint:** Make use of Exercise 5.1.9.

5.1.11. Use Hölder's inequality (5.1.8) to prove that if the components of $\mathbf{x} \in \Re^{n \times 1}$ sum to zero (i.e., $\mathbf{x}^T\mathbf{e} = 0$ for $\mathbf{e}^T = (1, 1, \ldots, 1)$), then

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_1 \left(\frac{y_{\max} - y_{\min}}{2} \right) \quad \text{for all } \mathbf{y} \in \Re^{n \times 1}.$$

Note: For "zero sum" vectors $\mathbf{x}$, this is at least as sharp and usually it's sharper than (5.1.8) because $(y_{\max} - y_{\min})/2 \leq \max_i |y_i| = \|\mathbf{y}\|_\infty$.

5.1.12. The classical form of ***Hölder's inequality***[36] states that if $p > 1$ and $q > 1$ are real numbers such that $1/p + 1/q = 1$, then

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

Derive this inequality by executing the following steps:

(a) By considering the function $f(t) = (1-\lambda) + \lambda t - t^\lambda$ for $0 < \lambda < 1$, establish the inequality

$$\alpha^\lambda \beta^{1-\lambda} \leq \lambda\alpha + (1-\lambda)\beta$$

for nonnegative real numbers α and β.

(b) Let $\hat{\mathbf{x}} = \mathbf{x}/\left\|\mathbf{x}\right\|_p$ and $\hat{\mathbf{y}} = \mathbf{y}/\left\|\mathbf{y}\right\|_q$, and apply the inequality of part (a) to obtain

$$\sum_{i=1}^{n} |\hat{x}_i \hat{y}_i| \leq \frac{1}{p}\sum_{i=1}^{n} |\hat{x}_i|^p + \frac{1}{q}\sum_{i=1}^{n} |\hat{y}_i|^q = 1.$$

(c) Deduce the classical form of Hölder's inequality, and then explain why this means that

$$|\mathbf{x}^*\mathbf{y}| \leq \left\|\mathbf{x}\right\|_p \left\|\mathbf{y}\right\|_q.$$

5.1.13. The triangle inequality $\left\|\mathbf{x}+\mathbf{y}\right\|_p \leq \left\|\mathbf{x}\right\|_p + \left\|\mathbf{y}\right\|_p$ for a general p-norm is really the classical ***Minkowski inequality,***[37] which states that for $p \geq 1$,

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

Derive Minkowski's inequality. **Hint:** For $p > 1$, let q be the number such that $1/q = 1 - 1/p$. Verify that for scalars α and β,

$$|\alpha+\beta|^p = |\alpha+\beta|\,|\alpha+\beta|^{p/q} \leq |\alpha|\,|\alpha+\beta|^{p/q} + |\beta|\,|\alpha+\beta|^{p/q},$$

and make use of Hölder's inequality in Exercise 5.1.12.

---

[36] Ludwig Otto Hölder (1859–1937) was a German mathematician who studied at Göttingen and lived in Leipzig. Although he made several contributions to analysis as well as algebra, he is primarily known for the development of the inequality that now bears his name.

[37] Hermann Minkowski (1864–1909) was born in Russia, but spent most of his life in Germany as a mathematician and professor at Königsberg and Göttingen. In addition to the inequality that now bears his name, he is known for providing a mathematical basis for the special theory of relativity. He died suddenly from a ruptured appendix at the age of 44.

Solutions for Chapter 5

Solutions for exercises in section 5. 1

5.1.1. (a) $\|\mathbf{x}\|_1 = 9, \quad \|\mathbf{x}\|_2 = 5, \quad \|\mathbf{x}\|_\infty = 4$

(b) $\|\mathbf{x}\|_1 = 5 + 2\sqrt{2}, \quad \|\mathbf{x}\|_2 = \sqrt{21}, \quad \|\mathbf{x}\|_\infty = 4$

5.1.2. (a) $\|\mathbf{u} - \mathbf{v}\| = \sqrt{31}$ (b) $\|\mathbf{u} + \mathbf{v}\| = \sqrt{27} \le 7 = \|\mathbf{u}\| + \|\mathbf{v}\|$

(c) $|\mathbf{u}^T\mathbf{v}| = 1 \le 10 = \|\mathbf{u}\|\ \|\mathbf{v}\|$

5.1.3. Use the CBS inequality with $\mathbf{x} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$.

5.1.4. (a) $\{\mathbf{x} \in \Re^n \mid \|\mathbf{x}\|_2 \le 1\}$ (b) $\{\mathbf{x} \in \Re^n \mid \|\mathbf{x} - \mathbf{c}\|_2 \le \rho\}$

5.1.5. $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2 \implies -2\mathbf{x}^T\mathbf{y} = 2\mathbf{x}^T\mathbf{y} \implies \mathbf{x}^T\mathbf{y} = 0.$

5.1.6. $\|\mathbf{x} - \mathbf{y}\| = \|(-1)(\mathbf{y} - \mathbf{x})\| = |(-1)|\ \|\mathbf{y} - \mathbf{x}\| = \|\mathbf{y} - \mathbf{x}\|$

5.1.7. $\mathbf{x} - \mathbf{y} = \sum_{i=1}^n (x_i - y_i)\mathbf{e}_i \implies \|\mathbf{x} - \mathbf{y}\| \le \sum_{i=1}^n |x_i - y_i|\ \|\mathbf{e}_i\| \le \nu \sum_{i=1}^n |x_i - y_i|$, where $\nu = \max_i \|\mathbf{e}_i\|$. For each $\epsilon > 0$, set $\delta = \epsilon/n\nu$. If $|x_i - y_i| < \delta$ for each i, then, using (5.1.6), $\big|\, \|\mathbf{x}\| - \|\mathbf{y}\|\, \big| \le \|\mathbf{x} - \mathbf{y}\| < \nu n \delta = \epsilon$.

5.1.8. To show that $\|\mathbf{x}\|_1 \le \sqrt{n}\, \|\mathbf{x}\|_2$, apply the CBS inequality to the standard inner product of a vector of all 1's with a vector whose components are the $|x_i|$'s.

5.1.9. If $\mathbf{y} = \alpha\mathbf{x}$, then $|\mathbf{x}^*\mathbf{y}| = |\alpha|\, \|\mathbf{x}\|^2 = \|\mathbf{x}\|\ \|\mathbf{y}\|$. Conversely, if $|\mathbf{x}^*\mathbf{y}| = \|\mathbf{x}\|\ \|\mathbf{y}\|$, then (5.1.4) implies that $\|\alpha\mathbf{x} - \mathbf{y}\| = 0$, and hence $\alpha\mathbf{x} - \mathbf{y} = \mathbf{0}$—recall (5.1.1).

5.1.10. If $\mathbf{y} = \alpha\mathbf{x}$ for $\alpha > 0$, then $\|\mathbf{x} + \mathbf{y}\| = \|(1 + \alpha)\mathbf{x}\| = (1 + \alpha)\, \|\mathbf{x}\| = \|\mathbf{x}\| + \|\mathbf{y}\|$. Conversely, $\|\mathbf{x} + \mathbf{y}\| = \|\mathbf{x}\| + \|\mathbf{y}\| \implies (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 = \|\mathbf{x} + \mathbf{y}\|^2 \implies$

$$\begin{aligned} \|\mathbf{x}\|^2 + 2\,\|\mathbf{x}\|\ \|\mathbf{y}\| + \|\mathbf{y}\|^2 &= (\mathbf{x}^* + \mathbf{y}^*)\,(\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x}^*\mathbf{x} + \mathbf{x}^*\mathbf{y} + \mathbf{y}^*\mathbf{x} + \mathbf{y}^*\mathbf{y} \\ &= \|\mathbf{x}\|^2 + 2\,\mathrm{Re}(\mathbf{x}^*\mathbf{y}) + \|\mathbf{y}\|^2, \end{aligned}$$

and hence $\|\mathbf{x}\|\ \|\mathbf{y}\| = \mathrm{Re}\,(\mathbf{x}^*\mathbf{y})$. But it's always true that $\mathrm{Re}\,(\mathbf{x}^*\mathbf{y}) \le \big|\mathbf{x}^*\mathbf{y}\big|$, so the CBS inequality yields

$$\|\mathbf{x}\|\ \|\mathbf{y}\| = \mathrm{Re}\,(\mathbf{x}^*\mathbf{y}) \le \big|\mathbf{x}^*\mathbf{y}\big| \le \|\mathbf{x}\|\ \|\mathbf{y}\|.$$

In other words, $\big|\mathbf{x}^*\mathbf{y}\big| = \|\mathbf{x}\|\ \|\mathbf{y}\|$. We know from Exercise 5.1.9 that equality in the CBS inequality implies $\mathbf{y} = \alpha\mathbf{x}$, where $\alpha = \mathbf{x}^*\mathbf{y}/\mathbf{x}^*\mathbf{x}$. We now need to show that this α is real and positive. Using $\mathbf{y} = \alpha\mathbf{x}$ in the equality $\|\mathbf{x} + \mathbf{y}\| =$

$\|\mathbf{x}\| + \|\mathbf{y}\|$ produces $|1+\alpha| = 1 + |\alpha|$, or $|1+\alpha|^2 = (1+|\alpha|)^2$. Expanding this yields

$$\begin{aligned}(1+\bar{\alpha})(1+\alpha) &= 1 + 2|\alpha| + |\alpha|^2 \\ \Longrightarrow \quad 1 + 2\,\mathrm{Re}(\alpha) + \bar{\alpha}\alpha &= 1 + 2|\alpha| + \bar{\alpha}\alpha \\ \Longrightarrow \quad \mathrm{Re}(\alpha) &= |\alpha|,\end{aligned}$$

which implies that α must be real. Furthermore, $\alpha = \mathrm{Re}\,(\alpha) = |\alpha| \geq 0$. Since $\mathbf{y} = \alpha\mathbf{x}$ and $\mathbf{y} \neq \mathbf{0}$, it follows that $\alpha \neq 0$, and therefore $\alpha > 0$.

5.1.11. This is a consequence of Hölder's inequality because

$$|\mathbf{x}^T\mathbf{y}| = |\mathbf{x}^T(\mathbf{y} - \alpha\mathbf{e})| \leq \|\mathbf{x}\|_1 \, \|\mathbf{y} - \alpha\mathbf{e}\|_\infty$$

for all α, and $\min_\alpha \|\mathbf{y} - \alpha\mathbf{e}\|_\infty = (y_{\max} - y_{\min})/2$ (with the minimum being attained at $\alpha = (y_{\max} + y_{\min})/2$).

5.1.12. (a) It's not difficult to see that $f'(t) < 0$ for $t < 1$, and $f'(t) > 0$ for $t > 1$, so we can conclude that $f(t) > f(1) = 0$ for $t \neq 1$. The desired inequality follows by setting $t = \alpha/\beta$.

(b) This inequality follows from the inequality of part (a) by setting

$$\alpha = |\hat{x}_i|^p, \quad \beta = |\hat{y}_i|^q, \quad \lambda = 1/p, \quad \text{and} \quad (1-\lambda) = 1/q.$$

(c) Hölder's inequality results from part (b) by setting $\hat{x}_i = x_i / \|\mathbf{x}\|_p$ and $\hat{y}_i = y_i / \|\mathbf{y}\|_q$. To obtain the "vector form" of the inequality, use the triangle inequality for complex numbers to write

$$\begin{aligned}|\mathbf{x}^*\mathbf{y}| &= \left|\sum_{i=1}^n \overline{x_i} y_i\right| \leq \sum_{i=1}^n |\overline{x_i}|\,|y_i| = \sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q} \\ &= \|\mathbf{x}\|_p \, \|\mathbf{y}\|_q .\end{aligned}$$

5.1.13. For $p = 1$, Minkowski's inequality is a consequence of the triangle inequality for scalars. The inequality in the hint follows from the fact that $p = 1 + p/q$ together with the scalar triangle inequality, and it implies that

$$\sum_{i=1}^n |x_i + y_i|^p = \sum_{i=1}^n |x_i + y_i|\,|x_i + y_i|^{p/q} \leq \sum_{i=1}^n |x_i|\,|x_i + y_i|^{p/q} + \sum_{i=1}^n |y_i|\,|x_i + y_i|^{p/q}.$$

Application of Hölder's inequality produces

$$\begin{aligned}\sum_{i=1}^n |x_i|\,|x_i + y_i|^{p/q} &\leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i + y_i|^p\right)^{1/q} \\ &= \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |x_i + y_i|^p\right)^{(p-1)/p} \\ &= \|\mathbf{x}\|_p \, \|\mathbf{x} + \mathbf{y}\|_p^{p-1} .\end{aligned}$$

Similarly, $\sum_{i=1}^{n} |y_i|\,|x_i + y_i|^{p/q} \le \|\mathbf{y}\|_p \, \|\mathbf{x}+\mathbf{y}\|_p^{p-1}$, and therefore

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

Solutions for exercises in section 5. 2

5.2.1. $\|\mathbf{A}\|_F = \left[\sum_{i,j} |a_{ij}|^2\right]^{1/2} = [\mathit{trace}\,(\mathbf{A}^*\mathbf{A})]^{1/2} = \sqrt{10}$,

$\|\mathbf{B}\|_F = \sqrt{3}$, and $\|\mathbf{C}\|_F = \sqrt{9}$.

5.2.2. (a) $\|\mathbf{A}\|_1$ = max absolute column sum = 4, and $\|\mathbf{A}\|_\infty$ = max absolute row sum = 3. $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest value of λ for which $\mathbf{A}^T\mathbf{A} - \lambda\mathbf{I}$ is singular. Determine these λ's by row reduction.

$$\mathbf{A}^T\mathbf{A} - \lambda\mathbf{I} = \begin{pmatrix} 2 - -\lambda & -4 \\ -4 & 8-\lambda \end{pmatrix} \longrightarrow \begin{pmatrix} -4 & 8-\lambda \\ 2-\lambda & -4 \end{pmatrix}$$
$$\longrightarrow \begin{pmatrix} -4 & 8-\lambda \\ 0 & -4 + \frac{2-\lambda}{4}(8-\lambda) \end{pmatrix}$$

This matrix is singular if and only if the second pivot is zero, so we must have $(2-\lambda)(8-\lambda) - 16 = 0 \implies \lambda^2 - 10\lambda = 0 \implies \lambda = 0,\ \lambda = 10$, and therefore $\|\mathbf{A}\|_2 = \sqrt{10}$.

(b) Use the same technique to get $\|\mathbf{B}\|_1 = \|\mathbf{B}\|_2 = \|\mathbf{B}\|_\infty = 1$, and

(c) $\|\mathbf{C}\|_1 = \|\mathbf{C}\|_\infty = 10$ and $\|\mathbf{C}\|_2 = 9$.

5.2.3. (a) $\|\mathbf{I}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{I}\mathbf{x}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{x}\| = 1$.

(b) $\|\mathbf{I}_{n\times n}\|_F = \left[\mathit{trace}\left(\mathbf{I}^T\mathbf{I}\right)\right]^{1/2} = \sqrt{n}$.

5.2.4. Use the fact that $\mathit{trace}\,(\mathbf{AB}) = \mathit{trace}\,(\mathbf{BA})$ (recall Example 3.6.5) to write

$$\|\mathbf{A}\|_F^2 = \mathit{trace}\,(\mathbf{A}^*\mathbf{A}) = \mathit{trace}\,(\mathbf{A}\mathbf{A}^*) = \|\mathbf{A}^*\|_F^2\,.$$

5.2.5. (a) For $\mathbf{x} = \mathbf{0}$, the statement is trivial. For $\mathbf{x} \ne \mathbf{0}$, we have $\|(\mathbf{x}/\|\mathbf{x}\|)\| = 1$, so for any particular $\mathbf{x}_0 \ne \mathbf{0}$,

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = \max_{\mathbf{x}\ne\mathbf{0}} \left\|\mathbf{A}\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\| \ge \frac{\|\mathbf{A}\mathbf{x}_0\|}{\|\mathbf{x}_0\|} \implies \|\mathbf{A}\mathbf{x}_0\| \le \|\mathbf{A}\|\,\|\mathbf{x}_0\|\,.$$

(b) Let $\mathbf{x}_0$ be a vector such that $\|\mathbf{x}_0\| = 1$ and

$$\|\mathbf{AB}\mathbf{x}_0\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{AB}\mathbf{x}\| = \|\mathbf{AB}\|\,.$$

Make use of the result of part (a) to write

$$\|\mathbf{AB}\| = \|\mathbf{AB}\mathbf{x}_0\| \le \|\mathbf{A}\|\,\|\mathbf{B}\mathbf{x}_0\| \le \|\mathbf{A}\|\,\|\mathbf{B}\|\,\|\mathbf{x}_0\| = \|\mathbf{A}\|\,\|\mathbf{B}\|\,.$$