6 Linear algebra

For the rest of the course, we are concerned with functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. In preparation for this material and to set some conventions, we must first review a bit of linear algebra. It is assumed that the reader has had a course in linear algebra and is conversant with matrix computations.

6 Linear algebra **85**

6.1 Linear transformations 85

6.2 All norms on $\mathbb{R}^n$ are equivalent 89

6.3 The metric space of linear transformations 92

6.4 Invertible matrices 95

6.1 Linear transformations

We view the elements of $\mathbb{R}^n$, for $n \in \mathbb{N}$, as being *column* vectors. Thus

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}.$$

This is simply a convention (we could instead take $\mathbb{R}^n$ to consist of row vectors), but this is the only convention compatible with matrix multiplication (without reversing the order or taking transposes constantly).

On this set we have two operations, addition and scalar multiplication, which are defined by

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \text{and} \quad c \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} cx_1 \\ \vdots \\ cx_n \end{pmatrix}$$

for vectors $x, y \in \mathbb{R}^n$ and scalars $c \in \mathbb{R}$. With these operations, $\mathbb{R}^n$ is a vector space.

We will not review the definition of vector spaces, with its many axioms.

Of more interest to us are the mappings between these spaces, in particular (initially), the linear ones.

Definition 6.1.1. The mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is *linear* (or, a *linear transformation*) if, for all vectors $x, y \in \mathbb{R}^n$ and all scalars $c \in \mathbb{R}$, we have

(i) $T(x+y) = T(x) + T(y)$ and

(ii) $T(cx) = cT(x)$.

The reader will note that we do note decorate our vectors by writing them as $\vec{x}$ or $\mathbf{x}$; whether a variable is a vector or a scalar (or something else) must be made clear from the context.

Note that we can simplify this definition somewhat; the mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if

$$T(cx+y) = cT(x) + T(y)$$

for all vectors $x, y \in \mathbb{R}^n$ and scalars $c \in \mathbb{R}$.

Example 6.1.2. The mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ x_1 + x_2 \end{pmatrix}$$

is linear.

Strictly speaking, we should denote the imagine of the vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ under the mapping T by $T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right)$, but we suppress this extra set of parentheses.

Example 6.1.3 (Scaling)**.** The mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \end{pmatrix}$$

is linear.

Example 6.1.4 (Rotation)**.** For any angle θ, the mapping $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by rotation by θ radians counterclockwise is linear.

Example 6.1.5 (Projection)**.** The mapping $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

is linear.

Example 6.1.6 (Inclusion)**.** The mapping $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}$$

is linear.

The mapping defined by

$$T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$$

is *not* linear. Why not?

As we assume the reader has already seen, linear transformations can all be expressed in terms of matrix multiplication and conversely, matrix multiplication always defines a linear transformation. We briefly revisit these concepts now.

Definition 6.1.7. Let $a_1, \ldots, a_n \in \mathbb{R}^m$ denote the columns of the $m \times n$ matrix A, so that

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}.$$

If $x \in \mathbb{R}^n$, then the product of A and x is

$$Ax = \sum_{k=1}^{n} a_k x_k \in \mathbb{R}^m.$$

Viewing the product Ax as a linear combination of the columns of A is frequently more convenient than the alternative dot product formulation.

In particular, when $e_k \in \mathbb{R}^n$ is the kth *standard basis vector*, namely the vector with a 1 in the kth position and 0 elsewhere, we see that $Ae_k = a_k$.

Proposition 6.1.8. *If A is an $m \times n$ matrix, then the mapping $L_A : \mathbb{R}^n \to \mathbb{R}^m$ defined by*

$$L_A(x) = Ax$$

is linear.

Proof. Let $a_1, \ldots, a_n \in \mathbb{R}^m$ denote the columns of A and take $x, y \in \mathbb{R}^n$ and $c \in \mathbb{R}$. We have

$$L_A(cx + y) = A(cx + y) = \sum_{k=1}^{n} (cx_k + y_k) a_k.$$

Because $\mathbb{R}^m$ is a vector space, the summands here expand to $cx_k a_k + y_k a_k$, and thus our expression of $L_A(cx + y)$ becomes

$$c \sum_{k=1}^{n} x_k a_k + \sum_{k=1}^{n} y_k a_k = cAx + Ay = cL_A(x) + L_A(y),$$

establishing that L_A is indeed linear. □

Conversely, every linear transformation can be expressed as left multiplication by a matrix. We saw above that if A is an $m \times n$ matrix and e_k is the kth standard basis vector, then

$$L_A(e_k) = Ae_k = a_k.$$

This shows that if $T = L_A$, then the columns of A must be the images $T(e_k)$; our result below verifies that this matrix indeed represents T.

Proposition 6.1.9. *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then there exists a unique $m \times n$ matrix A such that $T = L_A$.*

In some instances (but not in what we will do), it is convenient to have a notation such as $[T]$ to denote the matrix representing the linear transformation T. (In fact sometimes one will use a notation like $[T]_\alpha^\beta$ where α and β are bases of the domain and codomain.)

Proof. We have already shown that if the matrix A represents T, then the columns of A must $T(e_k)$. This shows that there is at most one matrix that represents T, but we must show that this matrix does indeed represent it.

Let A denote the $m \times n$ matrix whose kth column is given by the vector $T(e_k) \in \mathbb{R}^m$, so

$$A = \begin{pmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{pmatrix}.$$

Given any column vector $x \in \mathbb{R}^n$, we have

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{k=1}^{n} x_k e_k.$$

Thus by the linearity of T, we have that

$$T(x) = \sum_{k=1}^{n} x_k T(e_k). = \sum_{k=1}^{n} x_k a_k = Ax,$$

as desired. □

We leave it to the reader to verify that composition of linear transformations corresponds to matrix multiplication.

Given the fact that linear transformations are equivalent to left-multiplying by matrices, one may wonder why we don't just work with matrices all the time. There are several reasons, but one is that expressing a linear transformation as a matrix requires us to fix bases for the domain and codomain.

Proposition 6.1.10. *Suppose that $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^p$ are both linear, and that the $m \times n$ matrix A and the $p \times m$ matrix B satisfy*

$$S = L_A \quad \textit{and} \quad T = L_B.$$

Then the mapping $T \circ S : \mathbb{R}^n \to \mathbb{R}^p$ satisfies

$$T \circ S = L_B \circ L_A = L_{BA},$$

where BA is the $p \times n$ matrix product of B and A.

Exercises

Exercise 6.1.1. Prove that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $T(0) = 0$.

In Exercise 6.1.1, the first 0 is the zero vector in $\mathbb{R}^n$ and the second 0 is the zero vector in $\mathbb{R}^m$.

Exercise 6.1.2. Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2x_1 - x_2 + 3x_3 \\ x_3 - x_1 \end{pmatrix}.$$

Find a matrix A such that $T = L_A$ and explain how doing so shows that T is linear.

Exercise 6.1.3. Suppose that $\| \cdot \|$ is a norm on $\mathbb{R}^m$ and that the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is injective. Prove that the function $\| \cdot \|_* : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\|x\|_* = \|T(x)\|$$

is a norm on $\mathbb{R}^n$.

6.2 All norms on $\mathbb{R}^n$ are equivalent

We saw in the previous section how norms induce metrics, and thus when we are working with a space that has a norm, this norm is in some sense more fundamental that the induced metric. The main result of this section is that all norms on n-dimensional real space are, for most purposes of analysis, the same. First we must formalize this notion.

Definition 6.2.1. The norm $\|\cdot\|_a$ is *equivalent* to the norm $\|\cdot\|_b$ if there exist constants $C \geq c > 0$ such that

$$c\|x\|_a \leq \|x\|_b \leq C\|x\|_a$$

for all vectors $x \in V$.

Exercise 6.2.1 asks the reader to verify that this is indeed an equivalence relation.

The notions of convergence and continuity are the same on equivalent norms. Accordingly, we may freely move between equivalent norms for many purposes of analysis. This is frequently useful, because depending on what we are trying to do, one norm may be easier to work with than others.

For example, in $\mathbb{R}^n$, we have that

$$\|x\|_2^2 = \sum_{k=1}^{n} x_k^2 \leq \left(\sum_{k=1}^{n} |x_k|\right)^2 = \|x\|_1^2,$$

so $\|x\|_2 \leq \|x\|_1$ for every vector $x \in \mathbb{R}^n$. In the other direction, it follows direction from the Cauchy–Schwarz inequality that

$$\|x\|_1 = \sum_{k=1}^{n} |x_k| = \sum_{k=1}^{n} 1 \cdot |x_k| \leq \left\| \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\|_2 \left\| \begin{pmatrix} |x_1| \\ \vdots \\ |x_n| \end{pmatrix} \right\|_2 = \sqrt{n}\,\|x\|_2.$$

Therefore the norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on the vector space $\mathbb{R}^n$ are equivalent, because

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$$

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

We are thinking of n as fixed, so the $\sqrt{n}$ here is a constant.

For our proof we need one more definition. In $\mathbb{R}^n$ with the Euclidean metric $\|\cdot\|_2$, the *unit sphere* is the set

$$S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}.$$

Note that the unit sphere in $\mathbb{R}^n$ is denoted by S^{n-1}, not S^n. This is because the sphere itself is $(n-1)$-dimensional. For example, when you walk around on the surface of the Earth (that is, when you walk on S^2), it is virtually indistinguishable from walking on the plane $\mathbb{R}^2$; hence, *Flat Earthers*.

Because S^{n-1} is closed and bounded in $(\mathbb{R}^n, d_2)$, it is compact set in this metric space by the Heine–Borel theorem.

Theorem 6.2.2. *All norms on $\mathbb{R}^n$ are equivalent.*

One consequence of Theorem 6.2.2 is that if $\|\cdot\|$ is a norm on $\mathbb{R}^n$ and d is the induced metric, then the closed and bounded sets in $(\mathbb{R}^n, d)$ are the same as those in the Euclidean space $(\mathbb{R}^n, d_2)$.

Proof. The strategy is to show that every norm $\|\cdot\|$ on $\mathbb{R}^n$ is equivalent to $\|\cdot\|_2$. To this end, let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$ and define

$$M = \max\{\|e_1\|, \ldots, \|e_n\|\},$$

where e_k is the kth standard basis vector of $\mathbb{R}^n$.

For any $x \in \mathbb{R}^n$, writing $x = \sum x_k e_k$, we see that

$$\begin{aligned} \|x\| &= \left\| \sum_{k=1}^{n} x_k e_k \right\| \\ &\leq \sum_{k=1}^{n} \|x_k e_k\| \\ &= \sum_{k=1}^{n} |x_k| \, \|e_k\| \\ &\leq \sum_{k=1}^{n} M \, |x_k| \\ &= M \, \|x\|_1 \\ &\leq (M\sqrt{n}) \, \|x\|_2. \end{aligned}$$

The definition of M shows that our proof is dependent upon the fact that $\mathbb{R}^n$ is finite dimensional. Indeed, there are examples of inequivalent norms on infinite dimensional vector spaces; see Exercise 6.2.2.

The above gives us one of the inequalities we need. Obtaining the other is more difficult. We begin by defining the function $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = \|x\|$.

We want to show that f is continuous (actually, uniformly continuous), where its domain $\mathbb{R}^n$ is considered with the Euclidean metric d_2 and the codomain $\mathbb{R}$ is considered with the usual metric. For any pair of points $x, y \in \mathbb{R}^n$, we have by the norm axioms that

$$\|x\| - \|y\| = \|x - y + y\| - \|y\| \leq \|x - y\| + \|y\| - \|y\| = \|x - y\|,$$

and by interchanging the roles of x and y,

$$\|y\| - \|x\| = \|y - x + x\| - \|x\| \leq \|y - x\| + \|x\| - \|x\| = \|y - x\|.$$

It follows that

$$|\|x\| - \|y\|| \leq \|x - y\|$$

for any pair of points $x, y \in \mathbb{R}^n$. Now (to establish uniform continuity), let $\epsilon > 0$ be arbitrary. For any $x, y \in \mathbb{R}^n$ with

$$d_2(x, y) = \|x - y\|_2 < \frac{\epsilon}{M\sqrt{n}},$$

we have

$$|f(x) - f(y)| = |\|x\| - \|y\|| \leq \|x - y\| \leq M\sqrt{n} \, \|x - y\|_2 < \epsilon.$$

Therefore f is (uniformly) continuous with as a mapping from $(\mathbb{R}^n, d_2)$ to $(\mathbb{R}, d)$.

From this, it follows that the restriction $f|_{S^{n-1}}$ is a continuous function on a compact set, so it achieves its minimum, say $c \geq 0$. We must further have $c > 0$ because norms can only be 0 on the zero vector, and the zero vector is not in the unit sphere. Thus for any $x \in S^{n-1}$, we have

$$f(x) = \|x\| \geq c.$$

If $y \neq 0$, then $x = y/\|y\|_2 \in S^{n-1}$ so by the properties of norms it follows that

$$\|y\| = \left\| \|y\|_2 \frac{y}{\|y\|_2} \right\| = \|y\|_2 \, \|x\| \geq c\|y\|_2,$$

and the proof is complete. □

Exercises

Exercise 6.2.1. Prove that the notion of equivalence of norms really is an equivalence relation. That is, if $\|\cdot\|_a$, $\|\cdot\|_b$, and $\|\cdot\|_c$ are norms on some vector space V, prove that

(i) $\|\cdot\|_a$ is equivalent to itself;
(ii) if $\|\cdot\|_a$ is equivalent to $\|\cdot\|_b$, then $\|\cdot\|_b$ is equivalent to $\|\cdot\|_a$; and
(iii) if $\|\cdot\|_a$ is equivalent to $\|\cdot\|_b$ and $\|\cdot\|_b$ is equivalent to $\|\cdot\|_c$, then $\|\cdot\|_a$ is equivalent to $\|\cdot\|_c$.

Exercise 6.2.2. Consider the vector space $\mathcal{C}([0,1])$ with the norms

$$\|f\|_2 = \sqrt{\int_0^1 |f(x)|^2} \quad \text{and} \quad \|f\|_\infty = \max\{|f(x)| : 0 \leq x \leq 1\}.$$

For $n \in \mathbb{N}$, define the function $f_n : [0,1] \to [0,1]$ by $f_n(x) = x^n$ and show that

$$\lim_{n\to\infty} \|f_n\|_2 = 0 \quad \text{whereas} \quad \|f_n\|_\infty = 1 \text{ for all } n \in \mathbb{N}.$$

Conclude that the norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are not equivalent on $\mathcal{C}([0,1])$.

6.3 The metric space of linear transformations

Let $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ denote the set of linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. In Section 6.1, we described the canonical identification of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with the set

$$M(m, n) = \{X : X \text{ is an } m \times n \text{ matrix with entries from } \mathbb{R}\}.$$

In fact, this identification is an isomorphism between $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and $M(m, n)$, both viewed as vector spaces. Moreover, these vector spaces are isomorphic to $\mathbb{R}^{mn}$.

Since we have norms on $\mathbb{R}^{mn}$, we can apply these norms to linear transformations. In particular, we can give $M(m, n)$ the Euclidean norm by defining

$$\|X\|_2 = \sqrt{\sum_{k,\ell} x_{k,\ell}^2}$$

for all $X = (x_{k,\ell}) \in M(m, n)$. When viewed as a norm on linear transformations, this is often called the *Frobenius norm*. However, there is another norm on linear transformations that, for most purposes, is more natural and easier to work with, and our goal in this section is to define and study it. We begin with a result that says that linear transformation can only scale vectors by a certain amount, no matter the transformation or the norms used.

Note that all norms on linear transformations can be viewed as norms on $\mathbb{R}^{mn}$, and so they are all equivalent by the results of the previous section.

Proposition 6.3.1. *Suppose that* $\|\cdot\|_{\mathbb{R}^n}$ *and* $\|\cdot\|_{\mathbb{R}^m}$ *are norms on* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*, respectively. If* $T : \mathbb{R}^n \to \mathbb{R}^m$ *is linear, then there is a constant* $C > 0$ *such that*

$$\|T(x)\|_{\mathbb{R}^m} \leq C\|x\|_{\mathbb{R}^n}$$

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

This result shows that T is uniformly continuous when viewed as a mapping from the space $\mathbb{R}^n$ with the metric induced by $\|\cdot\|_{\mathbb{R}^n}$ to the space $\mathbb{R}^m$ with the metric induced by $\|\cdot\|_{\mathbb{R}^m}$.

Proof. Define

$$M = \max\{\|T(e_1)\|_{\mathbb{R}^m}, \ldots, \|T(e_n)\|_{\mathbb{R}^m}\},$$

where $e_1, \ldots, e_n$ are the standard basis vectors for $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$ be arbitrary. Writing $x = \sum x_k e_k$ in the usual way, we have the bound

$$\begin{aligned}
\|T(x)\|_{\mathbb{R}^m} &= \left\| \sum_{k=1}^n x_k \, T(e_k) \right\|_{\mathbb{R}^m} \\
&\leq \sum_{k=1}^n |x_k| \, \|T(e_k)\|_{\mathbb{R}^m} \\
&\leq \sum_{k=1}^n M \, |x_k| \\
&= M \, \|x\|_1.
\end{aligned}$$

This is almost what we want, but it is in terms of the 1-norm, not the arbitrary norm $\|\cdot\|_{\mathbb{R}^n}$. However, the results of the previous section show that $\|\cdot\|_{\mathbb{R}^n}$ is equivalent to the 1-norm, and so there is some constant $K > 0$ such that

$$\|y\|_1 \leq K\|y\|_{\mathbb{R}^n}$$

for all $y \in \mathbb{R}^n$. Combining this with our previous inequality shows that

$$\|T(x)\|_{\mathbb{R}^m} \leq MK\,\|x\|_{\mathbb{R}^n},$$

which proves the result, with $C = MK$. □

Note that if $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear and invertible, then Proposition 6.3.1 also applies to T^{-1}.

The quantity $C > 0$ in the previous result is a bound on the amount that the transformation T can stretch a vector (measured in terms of whichever norms we are using on its domain and codomain).

We now specialize to only consider the 2*-norm*, and so our previous result shows that there is a constant $C > 0$ such that

$$\|T(x)\|_2 \leq C\|x\|_2$$

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

The two 2-norms in this inequality are, strictly speaking, different—the one on the left is defined on $\mathbb{R}^m$, while the one on the right is defined on $\mathbb{R}^n$.

We define the *operator norm* (sometimes called the *matrix norm*) on linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ by

$$\|T\|_{\text{op}} = \inf\{C : \|T(x)\|_2 \leq C\|x\|_2 \text{ for all } x \in \mathbb{R}^n\}.$$

By our previous result, the set in the above definition is nonempty, and it is trivially bounded below by 0, so this infimum indeed exists. We leave the proof that $\|\cdot\|_{\text{op}}$ is actually a norm to the reader.

Recall that the norm axioms require that for all vectors $x, y \in V$ and all scalars $c \in \mathbb{R}$:

(i) $\|x\| \geq 0$;
(ii) $\|x\| = 0$ if and only if $x = 0$;
(iii) $\|cx\| = |c|\,\|x\|$; and
(iv) $\|x + y\| \leq \|x\| + \|y\|$.

Proposition 6.3.2. *The operator norm $\|\cdot\|_{\text{op}}$ is a norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.*

The proof of the above result is made somewhat easier by the following reformulation of the operator norm.

Proposition 6.3.3. *For all linear transformations $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, we have*

$$\|T\|_{\text{op}} = \sup\{\|T(x)\|_2 : \|x\|_2 = 1\},$$

and this supremum is attained.

Proof. By our previous results, the mapping $f : S^{n-1} \to \mathbb{R}$ defined by $f(x) = \|T(x)\|_2$ is continuous as a mapping from the metric space (S^{n-1}, d_2) to the metric space $(\mathbb{R}, d)$. Because S^{n-1} is a compact set in $(\mathbb{R}^n, d_2)$, the function f attains its supremum at some point. Let $x_0 \in S^{n-1}$ and $M \in \mathbb{R}$ be such that

$$f(x_0) = \sup\{\|T(x)\|_2 : x \in S^{n-1}\} = M.$$

For every nonzero $y \in \mathbb{R}^n$, we have (by the linearity of T and the axioms of norms) that

$$\|T(y)\|_2 = \|y\|_2 \left\| T\left(\frac{y}{\|y\|_2}\right) \right\|_2 \leq M\,\|y\|_2.$$

This inequality also holds for the zero vector because

$$\|T(0)\|_2 = 0 = M\|0\|_2.$$

This implies that the infimum in the definition of $\|T\|_{\text{op}}$ is at most M.

In the other direction, if $\|T(x)\|_2 \leq C\|x\|_2$ for all $x \in \mathbb{R}^n$, then $C \geq M = \|T(x_0)\|_2$, so the infimum in the definition of $\|T\|_{\text{op}}$ is at least M, and this completes the proof. □

As a consequence of our result above, we see that for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we have

$$\|T(x)\|_2 \leq \|T\|_{\text{op}}\|x\|_2$$

for all vectors $x \in \mathbb{R}^n$, and equality is attained for at least some vectors.

Exercises

Exercise 6.3.1. Prove Proposition 6.3.2, which states that the operator norm is in fact a norm on the space $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

Exercise 6.3.2. Given a vector $y \in \mathbb{R}^n$, define the linear transformation $f_y : \mathbb{R}^n \to \mathbb{R}$ by $f_y(x) = \langle x, y \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^n$. Compute $\|f\|_{\text{op}}$.

6.4 Invertible matrices

In the previous section, we introduced the operator norm for linear transformations and proved that it can be defined as

$$\|T\|_{\text{op}} = \sup\{\|T(x)\|_2 : \|x\|_2 = 1\}.$$

As linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ are equivalent to $m \times n$ real matrices, we can extend this norm to the context of real matrices. Given a matrix $A \in M(m,n)$, we define

$$\|A\|_{\text{op}} = \|L_A\|_{\text{op}}$$

where $L_A : \mathbb{R}^n \to \mathbb{R}^m$ is the linear transformation defined by $L_A(x) = Ax$ for all $x \in \mathbb{R}^n$. All of our previous results about the operator norm translate directly to the matrix context; in particular, we have

$$\|Ax\|_2 \leq \|A\|_{\text{op}}\|x\|_2$$

for all vectors x of the appropriate size, and equality is achieved for at least one such vector. From this inequality we see that given matrices A and B of the appropriate sizes and any appropriately sized vector x, we have

$$\|ABx\|_2 \leq \|A\|_{\text{op}}\|Bx\|_2 \leq \|A\|_{\text{op}}\|B\|_{\text{op}}\|x\|_2,$$

from which it follows that

$$\|AB\|_{\text{op}} \leq \|A\|_{\text{op}}\|B\|_{\text{op}}.$$

This shows that the operator norm is *submultiplicative* on matrices.

If A is square (so that it *has* eigenvalues) and its entries are all real (as in our context), then this inequality can be used to show that the operator norm of A is equal to the absolute value of the largest eigenvalue of A (this is called the *spectral radius* of A). If A is not square, one can instead look at its singular values. We utilize neither approach here.

We now consider inverses of matrices (or equivalently, inverses of linear transformations between Euclidean spaces). Let A be an $m \times n$ real matrix, so $L_A : \mathbb{R}^n \to \mathbb{R}^m$. If A is not square, then neither A nor the mapping L_A can be invertible—if $m < n$, then L_A cannot be one-to-one, while if $m > n$, then L_A cannot be onto.

On the other hand, if A is square, then we have the following result from linear algebra, which we state without proof.

Proposition 6.4.1. *For an $n \times n$ real matrix A, the following are equivalent:*

(a) A is invertible;

(b) L_A is one-one;

(c) L_A is onto;

(d) there exists an $n \times n$ matrix B such that $BA = I_n$ (and in this case $B = A^{-1}$);

(e) there exists an $n \times n$ matrix C such that $AC = I_n$ (and in this case $C = A^{-1}$);

(f) the null space of A is $\{0\}$; and

(g) $\det(A) \neq 0$.

To establish the main result of this section, we need the following computation.

Proposition 6.4.2. *If A is an $n \times n$ matrix and $\|A\|_{\text{op}} < 1$, then $I_n - A$ is invertible. Moreover,*

$$\|(I_n - A)^{-1}\|_{\text{op}} \leq \frac{1}{1 - \|A\|_{\text{op}}}.$$

Proof. First observe that for $x \in \mathbb{R}^n$,

$$\|Ax\|_2 \leq \|A\|_{\text{op}} \|x\|_2$$

by the definition of the operator norm. Therefore (using the triangle inequality on $\|\cdot\|_2$ for the first inequality) we have

$$\begin{aligned} \|x\|_2 &= \|(I_n - A)x + Ax\|_2 \\ &\leq \|(I_n - A)x\|_2 + \|Ax\|_2 \\ &\leq \|(I_n - A)x\|_2 + \|A\|_{\text{op}} \|x\|_2. \end{aligned}$$

This implies that

$$\|(I_n - A)x\|_2 \geq (1 - \|A\|_{\text{op}}) \|x\|_2.$$

In particular, if $x \neq 0$, then $(I_n - A)x \neq 0$, so $I_n - A$ is invertible.

Given any $x \in \mathbb{R}^n$, the inequality above shows us that

$$\begin{aligned} \|x\|_2 &= \left\| (I_n - A) \left((I_n - A)^{-1} x \right) \right\|_2 \\ &\geq \left(1 - \|A\|_{\text{op}}\right) \|(I_n - A)^{-1} x\|_2. \end{aligned}$$

Therefore

$$\|(I_n - A)^{-1} x\|_2 \leq \frac{1}{1 - \|A\|_{\text{op}}} \|x\|_2.$$

It follows that $\|(I_n - A)^{-1}\|_{\text{op}} \leq (1 - \|A\|_{\text{op}})^{-1}$. ☐

The result below is the first in which we explicitly consider the metric space of matrices under the metric induced by the operator norm. As usual, this metric is defined by

$$d_{\text{op}}(A, B) = \|A - B\|_{\text{op}}.$$

Proposition 6.4.3. *For every n, the set $\mathcal{I}_n$ of invertible $n \times n$ matrices is an open subset of the metric space $(M(n, n), d_{\text{op}})$. Moreover, the mapping $f : \mathcal{I}_n \to \mathcal{I}_n$ defined by*

$$F(A) = A^{-1}$$

is continuous with respect to the d_{op} metric.

By the equivalence of norms, Proposition 6.4.3 holds not just for the d_{op} metric on $M(m, n)$, but also for all other metrics induced by a norm, such as the Frobenius norm.

Proof. Let $A \in \mathcal{I}_n$ be arbitrary. We want to show that there is some $\eta > 0$ such that if $d_{\text{op}}(A, B) < \eta$, then B is also invertible. Setting $H = B - A$, since $d_{\text{op}}(A, B) = \|A - B\|_{\text{op}} = \|H\|_{\text{op}}$, it suffices to show that $A + H$ is invertible.

To this end, set $\eta = 1/2\|A^{-1}\|_{\text{op}}$ and suppose that $\|H\|_{\text{op}} < \eta$. We have

$$A + H = A(I_n + A^{-1}H).$$

By our choice of H, we also have

$$\| - A^{-1}H\|_{\text{op}} \leq \|A^{-1}\|_{\text{op}} \, \|H\|_{\text{op}} < \frac{1}{2}.$$

Therefore $I_n + A^{-1}H$ is invertible by our previous result. Consequently, $B = A + H = A(I_n + A^{-1}H)$ is invertible, proving that the η-neighborhood of A lies in $\mathcal{I}_n$, and thus that $\mathcal{I}_n$ is open in the space $(M(n, n), d_{\text{op}})$.

To prove the second part of the result, first note that our previous result shows that

$$\begin{aligned}
\|(A + H)^{-1}\|_{\text{op}} &= \left\| \Big(A(I_n + A^{-1}H) \Big)^{-1} \right\|_{\text{op}} \\
&= \left\| (I_n + A^{-1}H))^{-1} A^{-1} \right\|_{\text{op}} \\
&\leq \|A^{-1}\|_{\text{op}} \, \|(I_n + A^{-1}H))^{-1}\|_{\text{op}} \\
&\leq \|A^{-1}\|_{\text{op}} \, \frac{1}{1 - \|A^{-1}H\|_{\text{op}}} \\
&\leq 2\|A^{-1}\|_{\text{op}}.
\end{aligned}$$

To see that F is continuous, again suppose $\|H\|_{\text{op}} < \eta$ and note that

$$\begin{aligned}
\|F(A + H) - F(A)\|_{\text{op}} &= \|(A + H)^{-1} \, (A - (A + H)) \, A^{-1}\|_{\text{op}} \\
&\leq \|A + H\|_{\text{op}}^{-1} \, \|H\|_{\text{op}} \, \|A^{-1}\|_{\text{op}} \\
&\leq 2\|A^{-1}\|_{\text{op}}^{2} \, \|H\|_{\text{op}}.
\end{aligned}$$

To complete the proof, let $\epsilon > 0$ be arbitrary and choose $0 < \delta \leq \eta$ such that $\delta < \epsilon/2\|A^{-1}\|_{\text{op}}$. ☐

Exercises

Exercise 6.4.1. Suppose that D is an $n \times n$ diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Prove that

$$\|D\|_{\text{op}} = \max\{|\lambda_j| : 1 \leq j \leq n\}.$$

Then prove that D is invertible if and only if all the diagonal entries are nonzero.

Exercise 6.4.2. For a fixed $\lambda > 0$, define the matrix

$$A = \begin{pmatrix} 0 & \lambda & 0 \\ 0 & 0 & \lambda \\ 0 & 0 & 0 \end{pmatrix}.$$

Show that $\|A\|_{\text{op}} = \lambda$. Also show that $I_3 - A$ is invertible even if $\lambda \geq 1$. Compare this result with Proposition 6.4.2.

Exercise 6.4.3. Suppose that $A \in M(n,n)$, $B \in M(m,n)$, and $C \in M(m,m)$, and define

$$X = \begin{pmatrix} A & 0 \\ B & C \end{pmatrix}.$$

Prove that X is invertible if and only if both A and B are.

Exercise 6.4.4. Complete the outline of the following alternate proof of Proposition 6.4.2. Suppose that A is an $n \times n$ matrix with $\|A\|_{\text{op}} < 1$. Show that the series

$$\sum_{k=0}^{\infty} A^k$$

is Cauchy in $(M(n,n), d_{\text{op}})$. Therefore this series converges to some matrix $A_* \in M(n,n)$, and moreover, we have

$$\|A_*\|_{\text{op}} \leq \sum_{k=0}^{\infty} \|A\|_{\text{op}}^k = \frac{1}{1 - \|A\|_{\text{op}}}.$$

Show further that

$$(I_n - A) \sum_{k=0}^{m} A^k = I_n - A^{m+1},$$

and that the righthand side converges to 0 as $m \to \infty$. Conclude that $I_n - A$ and A_* are inverses.