Linear Space Outline

0.1. Linear Spaces

In this outline we define norms and inner products for various types of spaces that occur frequently in applications of linear algebra. In this first section we define norms and inner products abstractly. In the next two sections we apply these definitions to various spaces. In particular, we will define norms for vector spaces, matrix spaces, and function spaces; also, we will define inner products for vector spaces and function spaces. Since all of these spaces are *linear spaces* we begin by discussing basic definitions and facts about abstract linear spaces.

0.1.1. Some Facts About Linear Spaces

We begin with a definition of a *linear space*. Although we will usually work with *real* linear spaces, we will include the definition of *complex* linear spaces for completeness. In this definition there are no differences in the expressions in the real and the complex cases. However, in this section there will be times where the expressions will be different and so all definitions will be stated carefully (with the complex case in parentheses) to be correct in both cases.

In order to keep our notation abstract, we will generally use the symbol P to refer to a space and Q and R to refer to subsets of this space. Elements in P will be denoted by p or by $p_1, p_2, \ldots, p_k$, and similarly for elements of Q and R. We are attempting to use symbols that do not generally refer to any particular type of space. Thus the space P might be a vector space, a space of matrices, a function space, or a space of operators, and the element p might be a vector, a matrix, a function, or an operator, respectively.

Definition 0.1. Let P be a nonempty set of elements on which the operations of addition and scalar multiplication are defined. That is, for any elements $p_1, p_2 \in P$, the operation of addition, i.e., $p_1 + p_2$, is defined, and $p_1 + p_2 \in P$. Similarly, for any $\alpha \in \mathbb{R}\ (\mathbb{C})$, the operation of scalar multiplication, i.e., αp_1, is defined, and $\alpha p_1 \in P$. The set P with these two operations is called a *real* (*complex*) *linear space* if the following axioms are satisfied:

(1) $p_1 + p_2 = p_2 + p_1$ for all $p_1, p_2 \in P$.

(2) $(p_1 + p_2) + p_3 = p_1 + (p_2 + p_3)$ for all $p_1, p_2, p_3 \in P$.

(3) There is a unique element in P, denoted by 0_P, such that $p + 0_P = p$ for all $p \in P$.

(4) For each $p \in P$ there is a unique element, denoted by $-p$, in P such that $p + (-p) = 0_P$.

(5) $\alpha(p_1 + p_2) = \alpha p_1 + \alpha p_2$ for all $p_1, p_2 \in P$ and $\alpha \in \mathbb{R}\ (\mathbb{C})$.

(6) $(\alpha + \beta)p = \alpha p + \beta p$ for all $p \in P$ and $\alpha, \beta \in \mathbb{R}\ (\mathbb{C})$.

(7) $\alpha(\beta p) = (\alpha\beta)p$ for all $p \in P$ and $\alpha, \beta \in \mathbb{R}\ (\mathbb{C})$.

(8) $1p = p$ for all $p \in P$.

Note that in this definition we differentiate the number 0 in $\mathbb{R}\ (\mathbb{C})$ from the zero element 0_P in P. We only do so in this section, but we feel that it helps distinguish “numbers” from “elements”. Most books denote the zero element by 0 or $\mathbf{0}$, which can be somewhat confusing.

As an example, the set of vectors in three-dimensional Euclidean space, i.e., $\mathbb{R}^3$, is a linear space with the usual rules of addition and scalar multiplication. That is, if

$$\mathtt{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathtt{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} ,$$

then

$$\mathtt{x} + \mathtt{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix} \quad \text{and} \quad \alpha\mathtt{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{pmatrix} .$$

Notation: As we will soon see, there are many more applications of linear spaces than sets of vectors in n-dimensional Euclidean space. However, the words "vector space" have become synonymous with "linear space". Additionally, the elements of a linear space are often called "vectors". This notation can be quite confusing since matrices are the elements of matrix spaces, functions are the elements of function spaces, and operators are the elements of operator spaces. Generically calling all these types of elements "vectors" is, in our opinion, a misuse of terminology (which we will avoid).

Here we will never be concerned with the eight axioms for a linear space. They are included only for completeness. The important properties about linear spaces, as far as we are concerned, are that

- if $p_1, p_2 \in P$ then $p_1 + p_2 \in P$ and
- if $\alpha \in \mathbb{R}\ (\mathbb{C})$ then $\alpha p_1 \in P$.

We will often be concerned with a subset of a linear space where the subset is also a linear space.

Definition 0.2. Let P be a real (complex) linear space and let Q be a non-empty subset of P (so that $Q \subset P$). Q is a *linear subspace of* P if:

(i) $\alpha q \in Q$ for all $q \in Q$ and for all $\alpha \in \mathbb{R}\ (\mathbb{C})$, and

(ii) $q_1 + q_2 \in Q$ for all $q_1, q_2 \in Q$.

Note that the smallest subspace of any linear space is the set consisting only of the zero element. That is, $\{\ 0_P\ \}$ is a subspace of any linear space.

One of the most important concepts in a linear space is that of linear independence and linear dependence.

Definition 0.3. A set of elements $\{\,p_1, p_2, \dots p_n\,\} \in P$ is *linearly independent* if the only solution to the equation

$$\sum_{j=1}^{n} c_j p_j = 0_P \,, \tag{0.1}$$

where $c_j \in \mathbb{R}\ (\mathbb{C})$ for all j, is the trivial (zero) solution, i.e., $c_j = 0$ for all j.
If there is a non-trivial solution, then the set is said to be *linearly dependent.*
A sum of the form

$$\sum_{j=1}^{n} c_j p_j$$

is called a *linear combination* of the set $\{\ p_j \mid 1 \le j \le n\ \}$.

Note: If the set of elements $\{\,p_1, p_2, \dots p_n\,\} \in P$ is linearly dependent, then there exists at least one nonzero coefficient, say c_ℓ. We can rewrite the linear combination as

$$p_\ell = -\frac{1}{c_\ell} \sum_{\substack{k=1 \\ k \ne \ell}}^{n} c_k p_k \,, \tag{0.2}$$

so that p_ℓ can be written as a linear combination of the other elements.

We often form a subspace of a linear space by using linear combinations of elements of the linear space.

Definition 0.4. Let $\{ p_1, p_2, \ldots, p_n \} \subset P$. Then the *span* of this set of elements, denoted by $\operatorname{span}\{ p_1, p_2, \ldots, p_n \}$, is the set of all linear combinations of these elements. That is,

$$\operatorname{span}\{ p_1, p_2, \ldots, p_n \} = \left\{ \sum_{j=1}^{n} c_j p_j \;\middle|\; c_j \in \mathbb{R}\ (\mathbb{C}) \right\} . \tag{0.3}$$

It is easy to show that $\operatorname{span}\{ p_1, p_2, \ldots, p_n \}$ is a subspace of P since, obviously, properties i and ii of Definition 0.2 are satisfied.

We close with the definition of a basis.

Definition 0.5. Let P be a real (complex) linear space and $Q \subset P$ be a subspace. The set of elements $\{ q_1, q_2, \ldots, q_n \} \in Q$ is said to be a *span* for Q if every vector in Q can be written as a linear combination of this set.
If $Q = \operatorname{span}\{ q_1, q_2, \ldots, q_n \}$ and the set of vectors $\{ q_1, q_2, \ldots, q_n \}$ is linearly independent, then this set is said to be a *basis* for Q.
The *dimension* of the subspace Q is the number of elements in a basis for Q and is denoted by $\dim Q$. (All bases for such a subspace have the same number of elements so this is well-defined.)

Note: Suppose that $Q = \operatorname{span}\{ p_1, p_2, \ldots, p_n \} \subset P$ but that the elements $\{ p_1, p_2, \ldots, p_n \}$ are not a basis. Then these elements are linearly dependent so there exists at least one nonzero coefficient, say c_ℓ, in eq. (0.1). Thus p_ℓ is some linear combination of the remaining elements so that

$$\operatorname{span}\{ p_1, p_2, \ldots, p_{\ell-1}, p_{\ell+1}, \ldots, p_n \} = \operatorname{span}\{ p_1, p_2, \ldots, p_n \} .$$

We can continue in this way removing elements from the spanning set which are linear combinations of the remaining elements until we are left with only linearly independent elements. This is how we turn a spanning set into a basis.

0.1.2. Norms and Inner Products

Now we define a *norm.*

Definition 0.6. A real (complex) linear space P is called a real (complex) *normed linear space* if there is a function $\| \cdot \| : P \to \mathbb{R}$ which satisfies the following three properties:
(1) $\|p\| \geq 0$ for all $p \in P$, and $\|p\| = 0$ if and only if $p = 0_P$.
(2) $\|cp\| = |c|\,\|p\|$ for all $c \in \mathbb{R}\ (\mathbb{C})$ and $p \in P$.
(3) $\|p_1 + p_2\| \leq \|p_1\| + \|p_2\|$ for all $p_1, p_2 \in P$.
Any function, $\| \cdot \|$, which satisfies the above three properties is called a *norm.* Property 3 is called the *triangle inequality.*

Our final definition is an inner product.

Definition 0.7. The real (complex) normed linear space is called a real (complex) *inner product space* if there is a function $\langle \cdot, \cdot \rangle : P \times P \to \mathbb{R}\ (\mathbb{C})$ which satisfies the following three properties:
(1) $\langle p, p \rangle \geq 0$ for all $p \in P$, and $\langle p, p \rangle = 0$ if and only if $p = 0_P$.
(2) $\langle p_1, p_2 \rangle = \langle p_2, p_1 \rangle$ $\big(\langle p_1, p_2 \rangle = \langle p_2, p_1 \rangle^* \big)$ for all $p_1, p_2 \in P$.
(3) $\langle c_1 p_1 + c_2 p_2, p_3 \rangle = c_1 \langle p_1, p_3 \rangle + c_2 \langle p_2, p_3 \rangle$ for all $c_1, c_2 \in \mathbb{R}\ (\mathbb{C})$ and $p_1, p_2, p_3 \in P$.
(Note that $\langle p_1, c_1 p_2 + c_2 p_3 \rangle = c_1^* \langle p_1, p_2 \rangle + c_2^* \langle p_1, p_3 \rangle$ by properties 2 and 3). Any function, $\langle \cdot, \cdot \rangle$, which satisfies the above three properties is called an *inner product.*

It is important to realize that there are always an infinite number of different norms in a normed linear space and an infinite number of inner products in an inner product space (as we will see in the next two sections).

One of the most important concepts in an inner product space is that of orthogonality.

Definition 0.8. Let P be a real (complex) inner product space. We say that two elements $p_1, p_2 \in P$ are *orthogonal* if $\langle p_1, p_2 \rangle = 0$. This is often denoted by $p_1 \perp p_2$.
A subset $Q \subset P$ is called an *orthogonal set* if $q_1 \perp q_2 = 0$ for all $q_1, q_2 \in Q$. If, in addition, $\langle q, q \rangle = 1$ for all $q \in Q$, then Q is called an *orthonormal set.*
If $p \in P$ is orthogonal to every element of Q, i.e., $\langle p, q \rangle = 0$ for all $q \in Q$, we write $p \perp Q$.

Definition 0.9. We define the *angle*, $\theta \in [0, \pi]$, between two elements $p_1, p_2 \in P$ in a *real* inner product space from the formula

$$\cos\theta = \frac{\langle p_1, p_2 \rangle}{\sqrt{\langle p_1, p_1 \rangle}\,\sqrt{\langle p_2, p_2 \rangle}} \,. \tag{0.4}$$

This definition of the angle between vectors in two and three dimensions agrees with the "standard" definition (of measuring the angle with a protractor). The definition is then generalized to an arbitrary inner product space.

It is not yet clear that this definition of the angle between two elements is well-defined, because it is not clear that $\left|\langle p_1, p_2 \rangle\right| \le \sqrt{\langle p_1, p_1 \rangle}\,\sqrt{\langle p_2, p_2 \rangle}$ for all $p_1, p_2 \in P$. (We require this inequality so that $\cos\theta \in [-1, +1]$.) The proof of this is called the Cauchy-Schwarz inequality.

Theorem 0.1. *Cauchy-Schwarz Inequality.* Let P be a real (complex) inner product space. Then

$$\left|\langle p_1, p_2 \rangle\right| \le \sqrt{\langle p_1, p_1 \rangle}\,\sqrt{\langle p_2, p_2 \rangle}$$

for all $p_1, p_2 \in P$.

There are a few more definitions that will be useful.

Definition 0.10. Let P be a real (complex) inner product space with subspaces Q and R. We say that Q and R are *orthogonal*, denoted by $Q \perp R$, if $\langle q, r \rangle = 0$ for all $q \in Q$ and $r \in R$.
The sum of these two subspaces, denoted $Q + R$, is defined by

$$Q + R = \{\, q + r \mid q \in Q \ \text{ and } \ r \in R \,\} \,.$$

Note: We write $Q \oplus R$ instead of $Q + R$ if this representation is unique. That is, for each $p \in Q \oplus R$, there exists unique $q \in Q$ and $r \in R$ such that $p = q + r$.

Finally, the *orthogonal complement* of Q, denoted $Q^\perp$, is defined by

$$Q^\perp = \{\, p \in P \mid p \perp Q \,\} \,.$$

We now restate the above definitions in $P = \mathbb{R}^n$ to make their meaning clearer. Let Q and R be subspaces of $\mathbb{R}^n$, and let $\{\, \mathtt{q}_1, \mathtt{q}_2, \ldots, \mathtt{q}_k \,\}$ be a basis for Q and $\{\, \mathtt{r}_1, \mathtt{r}_2, \ldots, \mathtt{r}_l \,\}$ be a basis for R.
Q and R are orthogonal if each of the basis elements for Q is orthogonal to each basis element of R.
The subspace $Q + R$ can be written in terms of these basis elements as

$$Q + R = \operatorname{span}\{\, \mathtt{q}_1, \mathtt{q}_2, \ldots, \mathtt{q}_k, \mathtt{r}_1, \mathtt{r}_2, \ldots, \mathtt{r}_l \,\} \,.$$

The set of vectors in the span for $Q + R$ does not have to be linearly independent. If this set of vectors *is* linearly independent, then we write this subspace as $Q \oplus R$.
If $R = Q^\perp$ then each basis element for Q is orthogonal to each basis element for R and

$$Q \oplus R = \mathbb{R}^n = \operatorname{span}\{\, \mathtt{q}_1, \mathtt{q}_2, \ldots, \mathtt{q}_k, \mathtt{r}_1, \mathtt{r}_2, \ldots, \mathtt{r}_l \,\} \,.$$

0.1.3. Some Facts About Matrices And Linear Operators

$\mathbb{R}^{m \times n}$ $\left(\mathbb{C}^{m \times n}\right)$ is the set of all real (complex) m by n matrices.

Definition 0.11. Let $\mathtt{A} \in \mathbb{R}^{m\times n}\ \left(\mathbb{C}^{m\times n}\right)$. The *range* of $\mathtt{A}$ is defined by

$$R(\mathtt{A}) = \left\{\ \mathtt{b} \in \mathbb{R}^m\ \left(\mathbb{C}^m\right)\ \middle|\ \mathtt{b} = \mathtt{Ax}\ \text{ for some }\ \mathtt{x} \in \mathbb{R}^n\ \left(\mathbb{C}^n\right)\ \right\}\ .$$

The *null space* of $\mathtt{A}$ is defined by

$$N(\mathtt{A}) = \left\{\ \mathtt{x} \in \mathbb{R}^n\ \left(\mathbb{C}^n\right)\ \middle|\ \mathtt{Ax} = \mathbf{0}\ \right\}\ .$$

The *rank* of $\mathtt{A}$ is defined by $\operatorname{rank}(\mathtt{A}) = \dim R(\mathtt{A})$.

It can be shown that $\operatorname{rank}(\mathtt{A}) = \operatorname{rank}\left(\mathtt{A}^{\mathrm{T}}\right)$ so that the rank of a matrix $\mathtt{A}$ is the number of linearly independent rows *or* columns of $\mathtt{A}$.

Definition 0.12. Let P and $\widetilde{P}$ be real (complex) linear spaces. Also let L be a mapping from P to $\widetilde{P}$, denoted by $L : P \to \widetilde{P}$. L is a *linear operator* or a *linear function* if it satisfies the following two properties:

1) $L(\alpha p) = \alpha L(p)$ for all $\alpha \in \mathbb{R}\ (\mathbb{C})$ and $p \in P$.
2) $L(p_1 + p_2) = L(p_1) + L(p_2)$ for all $p_1, p_2 \in P$.

L is called a *linear functional* if $\widetilde{P} = \mathbb{R}\ (\mathbb{C})$.

Definition 0.13. Let P and $\widetilde{P}$ be linear spaces and let $L : P \to \widetilde{P}$ be a linear operator. The *image* of L is the subspace of $\widetilde{P}$ defined by

$$R(L) = \left\{\ \widetilde{p} \in \widetilde{P}\ \middle|\ \widetilde{p} = L\{p\}\ \text{ for some }\ p \in P\ \right\}\ .$$

The *kernel* of L is the subspace of P defined by

$$N(L) = \{p \in P \mid L\{p\} = 0_{\widetilde{P}}\}$$

where $0_{\widetilde{P}}$ is the zero element in $\widetilde{P}$.
The *rank* of L is defined by $\operatorname{rank}(L) = \dim R(L)$.

Note: If $L : \mathbb{R}^m \to \mathbb{R}^n\ \left(\mathbb{C}^m \to \mathbb{C}^n\right)$ is a linear operator, then there exists an $\mathtt{A} \in \mathbb{R}^{m\times n}\ \left(\mathbb{C}^{m\times n}\right)$ such that $L(\mathtt{x}) = \mathtt{Ax}$ for all $\mathtt{x}$. The kernel of L is the null space of $\mathtt{A}$ and the image of L is the range of $\mathtt{A}$.

0.2. Vector and Matrix Norms

0.2.1. Vector Norms

We need some way to measure the "size" of vectors. There are many possible functions that can be used, but they will all be norms.

The most frequently used family of vector norms are the p-norms,

$$\|\mathtt{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} \quad \text{for }\ p \in [1, \infty]\ . \tag{0.5}$$

These are often called the ℓ_p-norms. The most important of these, and the ones we will use in are

$$\|\mathtt{x}\|_1 = \sum_{j=1}^{n} |x_j| \tag{0.6−1}$$

$$\|\mathtt{x}\|_2 = \left(\sum_{j=1}^{n} |x_j|^2 \right)^{1/2} \tag{0.6−2}$$

and

$$\|\mathtt{x}\|_\infty = \lim_{p\to\infty} \|\mathtt{x}\|_p = \max_{1\le j\le n} |x_j| \,. \tag{0.6−∞}$$

$\|x\|_2$ is called the *Euclidean norm* and gives the usual length of a vector in Euclidean space. Note that $\|\mathtt{x}\|_2 = \sqrt{\mathtt{x}^{\mathrm{T}}\mathtt{x}}$ $\left(\|\mathtt{x}\|_2 = \sqrt{\mathtt{x}^{\mathrm{H}}\mathtt{x}}\right)$.

It might seem that there are an incredible number of norms to choose from. However, in $\mathbb{R}^n$ $\left(\mathbb{C}^n\right)$ all norms are equivalent.

Theorem 0.2. Every pair of matrices $\|\cdot\|$ and $[\![\cdot]\!]$ on $\mathbb{R}^n$ $\left(\mathbb{C}^n\right)$ are *equivalent*. That is, there are positive constants m and M (which depend on n) such that

$$m\,[\![\mathtt{x}]\!] \le \|\mathtt{x}\| \le M\,[\![\mathtt{x}]\!] \,. \tag{0.7}$$

for all vectors $\mathtt{x} \in \mathbb{R}^n$ $\left(\mathbb{C}^n\right)$.

We can also define an inner product for vectors. For vectors in $\mathbb{R}^n$ $\left(\mathbb{C}^n\right)$ the most common inner product is

$$\langle\,\mathtt{x},\mathtt{y}\,\rangle = \mathtt{x}^{\mathrm{T}}\mathtt{y} \quad (\langle\,\mathtt{x},\mathtt{y}\,\rangle = \mathtt{y}^{\mathrm{H}}\mathtt{x}) \tag{0.8}$$

or, equivalently,

$$\langle\,\mathtt{x},\mathtt{y}\,\rangle = \sum_{j=1}^{n} x_j y_j \quad \left(\langle\,\mathtt{x},\mathtt{y}\,\rangle = \sum_{j=1}^{n} x_j y_j^* \right) \,.$$

We say that the ℓ_2-norm is *induced* by this inner product because $\|\mathtt{x}\|_2 = \sqrt{\langle \mathtt{x},\mathtt{x}\rangle}$. In fact, ℓ_2 is the only ℓ_p space which is an inner product space.

0.2.2. Matrix Norms

For matrices in $\mathbb{R}^{m\times n}$ $\left(\mathbb{C}^{m\times n}\right)$ we can define matrix norms. One frequently used matrix norm is the *Frobenius norm*

$$\|\mathtt{A}\|_{\mathrm{F}} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2} \,. \tag{0.9}$$

The other frequently used norm is the family of norms, called the *matrix p-norms*,

$$\|\mathtt{A}\|_p = \max_{\mathbf{x}\ne\mathbf{0}} \frac{\|\mathtt{Ax}\|_p}{\|\mathtt{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|\mathtt{Ax}\|_p \,, \tag{0.10}$$

which are constructed from the ℓ_p-norms. For these norms it can be proven that the maximum exists. That is, for every $\mathtt{A}$ and for every $p \in [1,\infty]$ there exists one or more unit vectors $\mathtt{y}$ such that $\|\mathtt{A}\|_p = \|\mathtt{Ay}\|_p$. Both of these norms, the Frobenius norm and the ℓ_p norms, satisfy the following condition.

Definition 0.14. A matrix norm is *consistent* if, in addition to satisfying the three properties of a norm in Definition 0.6, it also satisfies the following property:

4) $\|\mathtt{AB}\| \leq \|\mathtt{A}\|\,\|\mathtt{B}\|$ for all $\mathtt{A} \in \mathbb{R}^{m\times n}\ \left(\mathbb{C}^{m\times n}\right),\ \mathtt{B} \in \mathbb{R}^{n\times q}\ \left(\mathbb{C}^{n\times q}\right).$

Note: We will only work with norms that satisfy property 4.

This condition is so important because it enables us to obtain useful bounds for norms. This condition eliminates many simple choices for matrix norms, such as $\|\mathtt{A}\| = \max_{i,j} |a_{ij}|$.

Note that this condition involves a relationship among norms on three different spaces, namely $\mathbb{R}^{m\times n}\ \left(\mathbb{C}^{m\times n}\right)$, $\mathbb{R}^{n\times q}\ \left(\mathbb{C}^{n\times q}\right)$, and $\mathbb{R}^{m\times q}\ \left(\mathbb{C}^{m\times q}\right)$.

There is another useful condition that is satisfied by the ℓ_p norms and by the Frobenius norm.

Definition 0.15. A matrix norm is *compatible* with a vector norm if they satisfy

$$\|\mathtt{Ax}\| \leq \|\mathtt{A}\|\,\|\mathtt{x}\| \quad \text{for all } \mathtt{A} \in \mathbb{R}^{m\times n}\ \left(\mathbb{C}^{m\times n}\right) \text{ and } \mathtt{x} \in \mathbb{R}^{n}\ \left(\mathbb{C}^{n}\right).$$

Note that this is actually a relationship between *three* norms: $\|\mathtt{A}\|$ is a matrix norm while $\|\mathtt{x}\|$ is a vector norm in $\mathbb{R}^n\ \left(\mathbb{C}^n\right)$ and $\|\mathtt{Ax}\|$ is a vector norm in $\mathbb{R}^m\ \left(\mathbb{C}^m\right)$.

As before, the most common matrix p-norms are $\|\mathtt{A}\|_1$, $\|\mathtt{A}\|_2$, and $\|\mathtt{A}\|_\infty$. Two of these are easy to calculate,

$$\|\mathtt{A}\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}| \tag{0.11–1}$$

and

$$\|\mathtt{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|\,, \tag{0.11–∞}$$

while the third, $\|\mathtt{A}\|_2$, which is induced by the Euclidean norm, requires the calculation of eigenvalues. To describe the induced norm we first need to define the spectral radius

Definition 0.16. The *spectral radius*, $\rho(\mathtt{B})$, of a matrix $\mathtt{B} \in \mathbb{R}^{n\times n}\ \left(\mathbb{C}^{n\times n}\right)$ is defined by

$$\rho(\mathtt{B}) = \max_{j} |\lambda_j|\,, \tag{0.11–ρ}$$

where $\left\{\ \lambda_i \ \middle|\ 1 \leq i \leq n\ \right\}$ is the set of eigenvalues of the matrix $\mathtt{B}$.

Lemma 0.1. The induced norm is

$$\|\mathtt{A}\|_2 = \sqrt{\rho(\mathtt{A}^{\mathrm{T}}\mathtt{A})} \quad \left(\|\mathtt{A}\|_2 = \sqrt{\rho(\mathtt{A}^{\mathrm{H}}\mathtt{A})}\right)\,, \tag{0.11–2}$$

which is commonly called the *spectral norm*.

Notation: Whenever we use the norm notation $\|\cdot\|_F$ or $\|\cdot\|_p$ for matrix spaces we will mean eqs. (0.9) or (0.10).

It is also true that all matrix p-norms are equivalent.

Theorem 0.3. For every pair of matrix p-norms $\|\cdot\|_p$ and $\|\cdot\|_q$, there are positive constants m and M such that

$$m\|\mathtt{A}\|_q \leq \|\mathtt{A}\|_p \leq M\,\|\mathtt{A}\|_q\,.$$

for all matrices $\mathtt{A} \in \mathbb{R}^{n_1\times n_2}\ \left(\mathbb{C}^{n_1\times n_2}\right)$.

We show some examples of these bounds below, and we also show that the Frobenius norm is equivalent to the matrix p-norms.

Lemma 0.2. Let $\mathtt{A} \in \mathbb{R}^{n_1 \times n_2}$ $\left(\mathbb{C}^{n_1 \times n_2}\right)$. Then the following inequalities hold:

$$\begin{aligned} \|\mathtt{A}\|_2 \le \|\mathtt{A}\|_F &\le \sqrt{n_2}\,\|\mathtt{A}\|_2 \\ \frac{1}{\sqrt{n_2}}\|\mathtt{A}\|_\infty \le \|\mathtt{A}\|_2 &\le \sqrt{n_1}\,\|\mathtt{A}\|_\infty \\ \frac{1}{\sqrt{n_1}}\|\mathtt{A}\|_1 \le \|\mathtt{A}\|_2 &\le \sqrt{n_2}\,\|\mathtt{A}\|_1 \,. \end{aligned}$$

0.2.3. Condition Numbers

Although the condition number of a square matrix $\mathtt{A}$ is normally used to determine the accuracy of a numerical calculation, it can also be used analytically to determine how stable the matrix is. That is, do small changes in the right-hand side $\mathtt{b}$ induce small changes in the solution $\mathtt{x}$? Thus, consider the two linear equations

$$\mathtt{Ax} = \mathtt{b} \quad \text{and} \quad \mathtt{A}(\mathtt{x} + \delta\mathtt{x}) = \mathtt{b} + \delta\mathtt{b}$$

where $\delta\mathtt{b}$ is “small”. The question is whether $\delta\mathtt{x}$ is “small” or not. A single expression which estimates this is

$$\frac{\dfrac{\|\delta\mathtt{x}\|}{\|\mathtt{x}\|}}{\dfrac{\|\delta\mathtt{b}\|}{\|\mathtt{b}\|}}\,,$$

which is the relative error in $\mathtt{x}$ compared to the relative error in $\mathtt{b}$? The optimal bounds for this expression are given by

$$\frac{1}{\kappa_p(\mathtt{A})} \le \frac{\dfrac{\|\delta\mathtt{x}\|}{\|\mathtt{x}\|}}{\dfrac{\|\delta\mathtt{b}\|}{\|\mathtt{b}\|}} \le \kappa_p(\mathtt{A})\,, \tag{0.12a}$$

where $\kappa_p(\mathtt{A})$, the *condition number* of $\mathtt{A}$ in the p-norm, is defined by

$$\kappa_p(\mathtt{A}) = \|\mathtt{A}\|_p\,\|\mathtt{A}^{-1}\|_p\,. \tag{0.12b}$$

(These bounds are called *optimal* because for every $\mathtt{A}$ there exists a $\mathtt{b}$ and a $\delta\mathtt{b}$ such that the right inequality is an equality; the same holds for the left inequality.)

Another characterization of the condition number might also be helpful. It can be related to the set of singular matrices (for which the condition number is infinite) by

$$\kappa(\mathtt{A}) = \max_{\substack{\mathtt{E} \in \mathbb{R}^{n\times n} \\ \mathtt{A}+\mathtt{E}\ \text{singular}}} \frac{\|\mathtt{A}\|}{\|\mathtt{E}\|}\,, \tag{0.13}$$

which indicates that $1/\kappa_p(\mathtt{A})$ measures the p-norm distance from $\mathtt{A}$ to the set of singular matrices. For $p = 2$ the condition number is particularly simple, namely

$$\kappa_2(\mathtt{A}) = \sqrt{\frac{\max\limits_j |\lambda_j|}{\min\limits_j |\lambda_j|}} \tag{0.14}$$

where $\{\,\lambda_j\,\}$ is the set of eigenvalues of $\mathtt{A}^{\mathrm{T}}\mathtt{A}$ $\left(\mathtt{A}^{\mathrm{H}}\mathtt{A}\right)$. Also, if you know what the singular value decomposition of a matrix is, then $\kappa_2(\mathtt{A}) = s_1/s_n$ where $s_1 \ge s_2 \ge \cdots \ge s_n \ge 0$ are the singular values of $\mathtt{A}$.

0.3. Function Norms

0.3.1. Norms and Inner Products

Just as we can measure the "sizes" of vectors and matrices, we can also measure the "size" of functions. For example, suppose we approximate e^x by a finite Taylor series $f_N(x) = \sum_{k=0}^{N} x^k/k!$ Then we might like to know the "difference" between e^x and $f_N(x)$ as a function of N, or, equivalently, we would like to know the "size" of $e^x - f_N(x)$. We will generally work with *real* functions, but, again, we will show any differences that occur with complex functions.

The most frequently used family of function norms are the p-norms,

$$\|f\|_p = \left(\int_a^b \bigl| f(x) \bigr|^p \, dx \right)^{1/p} \quad \text{for } p \in [1, \infty] \,. \tag{0.15}$$

These are often called the L_p norms. The most important of these, and the ones we will use, are

$$\|f\|_1 = \int_a^b \bigl| f(x) \bigr| \, dx \tag{0.16–1}$$

$$\|f\|_2 = \left(\int_a^b \bigl| f(x) \bigr|^2 \, dx \right)^{1/2} \tag{0.16–2}$$

and

$$\|f\|_\infty = \max_{x \in [a,b]} \bigl| f(x) \bigr| \tag{0.16–∞}$$

.

If f is a continuous function then $\|f\|_\infty$ is the limit of the norms for finite p, just as for vector spaces. That is, $\|f\|_\infty = \lim_{p \to \infty} \|f\|_p$.

It is important to realize that these three function norms are not all equivalent, as was the case for vector norms. That is, Theorem 0.2 does not hold anymore. To show this consider the set of functions $f_k : [0,1] \to [0,1]$ defined by $f_k(x) = x^k$ for all integers $k \ge 1$. Then,

$$\|f_k\|_p = \left(\frac{1}{pk+1} \right)^{1/p} \quad \text{for } p < \infty$$

but

$$\|f_k\|_\infty = 1 \,.$$

Thus

$$\lim_{k \to \infty} \|f_k\|_p = \begin{cases} 0 & \text{if } p < \infty \\ 1 & \text{if } p = \infty \,, \end{cases}$$

so we cannot find positive constants m and M (independent of k) such that

$$m \, \| \, f_k\|_p \le \|f_k\|_\infty \le M \, \|f_k\|_p$$

for all k. For this example, Theorem 0.2 still holds for $\| \cdot \|_p$ when $p < \infty$ but it does not hold when $p = \infty$.

To see this, suppose f and g are two functions in $\mathrm{C}[a,b]$ which are "close". In the L_2 norm f and g can differ by a "large" amount on a "small" interval and we can still have $\|f - g\|_2 \ll 1$. However, if $\|f - g\|_\infty \ll 1$, then f and g must be "close" on the entire interval $[a,b]$. For example, let

$$f(x) = \begin{cases} 0 & \text{if } -1 \le x < -\epsilon \\ 1 + x/\epsilon & \text{if } -\epsilon \le x < 0 \\ 1 - x/\epsilon & \text{if } 0 \le x < \epsilon \\ 0 & \text{if } \epsilon \le x \le +1 \,, \end{cases}$$

where the plot is shown at the right. Then $\|f\|_2 = \sqrt{2/3\,\epsilon}$ while $\|f\|_\infty = 1$.

The most common inner product is

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx \quad \left(\langle f, g\rangle = \int_a^b f(x)g^*(x)\,dx\right) . \tag{0.17}$$

We say that the L_2 norm is *induced* by this inner product because $\|f\|_2 = \sqrt{\langle f, f\rangle}$.