1.4. Inner Products, Vector Norms, and Matrix Norms

We briefly review basic results for inner products, vector norms, and matrix norms. In §1.4.1 basic properties of norms and inner products are given. In §1.4.2, we discuss the properties of the norms that are based on the Euclidean norm.

1.4.1. Basic Definitions and Lemmas

First, we give the definition of an inner product.

**Definition 1.2.1** An inner product in \( \mathbb{R}^n \) is a function \((\cdot,\cdot)\) mapping \( \mathbb{R}^n \times \mathbb{R}^n \) into \( \mathbb{R} \) that satisfies the following four axioms

1. \( (x,x) \geq 0; \)  
   \( (x,x) = 0, \text{ if and only if } x = 0, \quad x \in \mathbb{R}^n; \)
2. \( (\alpha x, y) = \alpha (x, y), \quad x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}, \)
3. \( (x, y) = (y, x), \quad x, y \in \mathbb{R}^n, \)
4. \( (x + z, y) = (x, y) + (z, y), \quad x, y, z \in \mathbb{R}^n. \)

We note that if we define an inner product for complex vectors, the third axiom becomes

\[
(x, y) = \overline{(y, x)} \quad x, y \in \mathbb{C}^n
\]

where \( \overline{a} \) denotes the complex conjugate of \( a \).

The inner product that is used most often is the Euclidean dot product

\[
(x, y) = x^T y = \sum_{i=1}^{n} x_i y_i.
\]

Others are introduced in Chapter ??.

The Cauchy-Schwarz inequality given below is quite useful for all inner products.
Lemma 1.3 (Cauchy-Schwarz inequality) Let $(\cdot, \cdot)$ be an inner product in $\mathbb{R}^n$. Then for all $x, y \in \mathbb{R}^n$ we have

$$|\langle x, y \rangle| \leq \langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}.$$  \hfill (1.6)

Moreover, equality in (1.6) holds if and only if $x = \alpha y$ for some $\alpha \in \mathbb{R}$.

This inequality leads to the following definition of the angle between two vectors relative to an inner product.

Definition 1.3.1 The angle $\theta$ between the two nonzero vectors $x, y \in \mathbb{R}^n$ with respect to an inner product $(\cdot, \cdot)$ is given by

$$\cos \theta = \frac{\langle x, y \rangle}{\langle x, x \rangle^{1/2} \langle y, y \rangle^{1/2}}.$$  \hfill (1.7)

The next important definition is that of a vector norm.

Definition 1.3.2 A norm in $\mathbb{R}^n$ is a function $\| \cdot \|$ mapping $\mathbb{R}^n$ into $\mathbb{R}$ satisfying the following three axioms

1. $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$, $x \in \mathbb{R}^n$.
2. $\|\alpha x\| = |\alpha| \|x\|$, $x \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.
3. $\|x + y\| \leq \|x\| + \|y\|$, $x, y \in \mathbb{R}^n$.

Definition 1.3.2 and the Cauchy-Schwarz inequality give us that for any inner product $(\cdot, \cdot)_\alpha$ we can define a norm $\| \cdot \|_\alpha$ by

$$\|x\|_\alpha = (\langle x, x \rangle)^{1/2}.$$

(1.8)

The most well known of these norms is the Euclidean norm given by

$$\|x\|_2 = (x^T x)^{1/2} = \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$  \hfill (1.9)

It is also referred to as the two-norm. This is the most important vector norm in the book. The subscript “2” is explained below.

The two-norm is one of the class of $p$-norms. These are given by

$$\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \quad p \geq 1.$$  \hfill (1.10)
Clearly, \( p = 2 \) leads to the class (1.9). Except for the two-norm, the only norms we will use from this class are the one-norm given by

\[
\|x\|_1 = \sum_{i=1}^{n} |x_i|,
\]

and the \( \infty \)-norm given by

\[
\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| = \lim_{p \to \infty} \|x\|_p.
\]

The next lemma, the Hölder inequality, relates the \( p \)-norms to the Euclidean inner product.

**Lemma 1.4 (Hölder inequality)** Let \( \|\cdot\|_p \) and \( \|\cdot\|_q \) be norms in \( \mathbb{R}^n \) from the class (1.10) such that \( p^{-1} + q^{-1} = 1 \). Then for all \( x, y \in \mathbb{R}^n \) we have

\[
|x^T y| \leq \|x\|_p \|y\|_q.
\]

The interesting cases for us are \( p = q = 2 \), \( p = 1, q = \infty \), and \( p = \infty, q = 1 \). For \( p = q = 2 \), the Hölder inequality is just the Cauchy-Schwarz inequality.

The next inequality states that if we can bound \( \|x\|_\alpha \), for a given \( x \in \mathbb{R}^n \) in any norm \( \|\cdot\|_\alpha \), the quantity \( \|x\|_\beta \) can be bounded in any other norm \( \|\cdot\|_\beta \).

**Lemma 1.5** All norms on \( \mathbb{R}^n \) are uniformly equivalent, meaning that for any two norms \( \|\cdot\|_\alpha \) and \( \|\cdot\|_\beta \) there are constants \( c_1 \) and \( c_2 \) such that

\[
c_1 \|x\|_\beta \leq \|x\|_\alpha \leq c_2 \|x\|_\beta
\]

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

For the two-norm, the one-norm, and the \( \infty \)-norm the uniform equivalence relations are summarized by

\[
\frac{1}{\sqrt{n}} \|x\|_1 \leq \|x\|_2 \leq \|x\|_1,
\]

\[
\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty,
\]

\[
\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty.
\]

We will also need norms for matrices.

**Definition 1.5.1** A norm in \( \mathbb{R}^{m \times n} \) is a function \( \|\cdot\| \) mapping \( \mathbb{R}^{m \times n} \) into \( \mathbb{R} \) satisfying the following three axioms
1. \( \| X \| \geq 0; \)
\( \| X \| = 0 \) if and only if \( X = 0, \ X \in \mathbb{R}^{m \times n} \)
2. \( \| \alpha X \| = |\alpha| \| X \| \) \( X \in \mathbb{R}^{m \times n}, \alpha \in \mathbb{R} \)
3. \( \| X + Y \| \leq \| X \| + \| Y \| \) \( X, Y \in \mathbb{R}^{m \times n} \).

This definition is isomorphic to the definition of a vector norm on \( \mathbb{R}^{mn} \).
For example, the Frobenius norm defined by
\[
\| X \|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} x_{i,j}^2 \right)^{1/2}
\]
is isomorphic to the two-norm on \( \mathbb{R}^{mn} \).

For semantic reasons, we define a family of norms.

**Definition 1.5.2** Let \( \| \cdot \|_{\alpha, m, n} \) be a norm on \( \mathbb{R}^{m \times n} \) and let it be well-defined for every positive integer \( m \) and \( n \). Then the set
\[
N_\alpha = \{ \| \cdot \|_{\alpha, m, n} : m, n \text{ positive integers} \}
\]
is called a family of norms. For any positive integers \( m \) and \( n \), and any \( X \in \mathbb{R}^{m \times n} \), we denote \( \| X \|_{\alpha, m, n} \) by \( \| X \|_\alpha \). That is, for any matrix \( X \), the quantity \( \| X \|_\alpha \) is the appropriate member of \( N_\alpha \) applied to \( X \).

The set \( V_\alpha \subset N_\alpha \) given by
\[
V_\alpha = \{ \| \cdot \|_{\alpha, m, 1} : m \text{ positive integer} \}
\]
is the associated family of vector norms. For any \( x \in \mathbb{R}^m \), the quantity \( \| x \|_\alpha = \| (x) \|_\alpha \) is the appropriate member of \( V_\alpha \) applied to \( x \).

Since \( X \) represents a linear operator from \( \mathbb{R}^n \) to \( \mathbb{R}^m \), it is appropriate to define the induced norm \( \| \cdot \|_\alpha \) on \( \mathbb{R}^{m \times n} \) by
\[
\| X \|_\alpha = \sup_{y \neq 0} \frac{\| X y \|_\alpha}{\| y \|_\alpha}.
\]
(1.19)

It is a simple matter to show that
\[
\| X \|_\alpha = \max_{\| y \|_\alpha = 1} \| X y \|_\alpha.
\]
(1.20)

Note that the maximum is taken over a closed, bounded set, thus we have that
\[
\| X \|_\alpha = \| X y^* \|_\alpha
\]
(1.21)
1.4. INNER PRODUCTS, VECTOR NORMS, AND MATRIX NORMS

for some \( y^* \) such that \( \|y^*\|_\alpha = 1 \). The above definition leads to the very useful bound

\[
\|Xy\|_\alpha \leq \|X\|_\alpha \|y\|_\alpha
\]

(1.22)

where equality occurs for every vector of the form \( \gamma y^* \), \( \gamma \in \mathbb{R} \).

If \( X \in \mathbb{R}^{m \times 1} \) then (1.22) appears to give a second definition for \( \|X\|_\alpha \). However, it is easily verified that if \( X = (x) \), for some \( x \in \mathbb{R}^m \), then \( \|X\|_\alpha \), defined from the matrix norm (1.22), and \( \|x\|_\alpha \), defined from the vector norm, are the same value. Thus the notation \( \| \cdot \|_\alpha \) is unambiguous for all \( m \) and \( n \).

For any induced norm \( \| \cdot \| \), the identity matrix \( I_n \) for \( \mathbb{R}^{n \times n} \) satisfies

\[
\|I_n\| = 1.
\]

(1.23)

However, for the Frobenius norm

\[
\|I_n\|_F = \sqrt{n},
\]

thus it is not an induced norm for any vector norm.

For the one-norm and the \( \infty \)-norm there are formulas for the corresponding matrix norms and for a vector \( y^* \) satisfying (1.21). The one-norm formula is

\[
\|X\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |x_{ij}|.
\]

(1.24)

If \( j_{\text{max}} \) is the index of a column such that

\[
\|X\|_1 = \sum_{i=1}^{m} |x_{i,j_{\text{max}}}| \]

then \( y^* = e_{j_{\text{max}}} \), the corresponding column of the identity matrix.

The \( \infty \)-norm formula is

\[
\|X\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |x_{ij}|.
\]

(1.25)

If \( i_{\text{max}} \) is the index of a row such that

\[
\|X\|_\infty = \sum_{j=1}^{n} |x_{i_{\text{max}},j}| \]

then the vector \( y^* = (y_1^*, \ldots, y_n^*)^T \) with components

\[
y_j^* = \text{sign}(x_{i_{\text{max}},j})
\]

satisfies (1.21). Note that \( \|X\|_\infty = \|X^T\|_1 \).
The matrix two-norm does not have a formula like (1.24) or (1.25) and all other formulations are really equivalent to (1.19). Moreover, computing the vector \( \mathbf{y}^* \) in (1.21) is a nontrivial task that we will discuss in Chapters ?? and ??.

The induced norms have a convenient property that is important in understanding matrix computations. For \( X \in \mathbb{R}^{m \times n} \) and \( Y \in \mathbb{R}^{n \times s} \) consider \( \|XY\|_\alpha \). We have that

\[
\|XY\|_\alpha = \max_{\|z\|_\alpha = 1} \|XZ\|_\alpha \leq \max_{\|z\|_\alpha = 1} \|X\|_\alpha \|YZ\|_\alpha = \|X\|_\alpha \|Y\|_\alpha.
\]

Thus

\[
\|XY\|_\alpha \leq \|X\|_\alpha \|Y\|_\alpha. \tag{1.26}
\]

A norm \( \| \cdot \|_\alpha \) (or really family of norms) that satisfies the property (1.26) is said to be consistent. Since they are induced norms the two-norm, one-norm, and the \( \infty \)-norm are all consistent. The Frobenius norm also satisfies (1.26). An example of a matrix norm that is not consistent is given below.

**Example 1.1** Consider the norm \( \| \cdot \|_\beta \) on \( \mathbb{R}^{m \times n} \) given by

\[
\|X\|_\beta = \max_{(i,j)} |x_{ij}|.
\]

This is simply the \( \infty \)-norm applied to \( X \) written out as vector in \( \mathbb{R}^{mn} \). For \( m = n = 2 \), consider

\[
X = Y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
\]

Note that

\[
XY = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}
\]

and thus \( \|XY\|_\beta = 2 > \|X\|_\beta \|Y\|_\beta = 1 \). Clearly, \( \| \cdot \|_\beta \) norm is not consistent.

Henceforth, we use only consistent families of norms.

The above defined matrix norms satisfy the following useful inequalities for all \( X \in \mathbb{R}^{m \times n} \) and \( Y \in \mathbb{R}^{n \times s} \):

\[
\hat{+} \left( \|X\|_1 \|X\|_\infty \right)^{1/2} \leq \|X\|_2 \leq \left( \|X\|_1 \|X\|_\infty \right)^{1/2}, \tag{1.27}
\]

\[
\max \left\{ \frac{1}{\sqrt{m}}, \frac{1}{\sqrt{n}} \right\} \|X\|_F \leq \|X\|_2 \leq \|X\|_F, \tag{1.28}
\]

\[
\|XY\|_F \leq \|X\|_2 \|Y\|_F, \tag{1.29}
\]

\[
\|XY\|_F \leq \|X\|_F \|Y\|_2. \tag{1.30}
\]
For \( X \neq 0 \), the lower bound in (1.28) can be tightened into

\[
\frac{1}{\sqrt{\text{rank}(X)}}\|X\|_F \leq \|X\|_2. \tag{1.31}
\]

Since \text{rank}(X) is often more expensive to compute than a good approximation to \( \|X\|_2 \), the practical use of (1.31) is limited.

For any diagonal matrix \( A = \text{diag}(\lambda) \) and any \( p \)-norm \( \| \cdot \|_p \)

\[
\|A\|_p = \|\lambda\|_\infty = \max_{1 \leq i \leq n} |\lambda_i| \quad 1 \leq p \leq \infty.
\]

For any matrix \( X \in \mathbb{R}^{m \times n} \) we have the following relations for the matrix \( |X| \):

\[
\|X\|_2 \leq \| |X| \|_2 \leq \sqrt{\text{rank}(X)} \|X\|_2 \leq \min\{\sqrt{m}, \sqrt{n}\} \|X\|_2, \\
\|X\|_1 = \| |X| \|_1, \quad \|X\|_\infty = \| |X| \|_\infty, \\
\|X\|_F = \| |X| \|_F.
\]

In the context of linear least squares, our interest will be in the two-norm or in the Frobenius norm. The other norms will be used to bound them. Some special properties of these norms are given in the next section.

\subsection*{1.4.2. The Two-Norm, the Frobenius Norm, and Orthogonality}  

We begin by defining orthogonality. We then relate orthogonality to the matrix two-norm and Frobenius norm.

\textbf{Definition 1.5.3} Two vectors \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \) are orthogonal with respect to an inner product \( (\cdot, \cdot) \) if \( (\mathbf{x}, \mathbf{y}) = 0 \). The set \( S_1 \subseteq \mathbb{R}^n \) is orthogonal to the set \( S_2 \subseteq \mathbb{R}^n \), if for each \( \mathbf{x} \in S_1 \) and \( \mathbf{y} \in S_2 \), we have \( (\mathbf{x}, \mathbf{y}) = 0 \).

We write

\[
\mathbf{x} - \mathbf{y}, \quad S_1 - S_2
\]

to mean “\( \mathbf{x} \) is orthogonal to \( \mathbf{y} \)” and “\( S_1 \) is orthogonal to \( S_2 \).”

The zero vector is orthogonal to any vector and the set \( \{0\} \) is orthogonal to any set.

An orthogonal set of vectors is defined as follows.

\textbf{Definition 1.5.4} A set of \( k \) vectors \( \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \), where each \( \mathbf{x}_i \in \mathbb{R}^n \), is said to be an orthogonal with respect to the inner product \( (\cdot, \cdot) \) if \( (\mathbf{x}_i, \mathbf{x}_j) = 0 \) for \( i \neq j \). The set is said to be orthonormal if it is orthogonal and \( (\mathbf{x}_i, \mathbf{x}_i) = 1 \) for \( i = 1, 2, \ldots, k \).

The definition of an orthogonal matrix is related to the definition for vectors.