Inner Product and Normed Space

In all that follows, the $n \times n$ identity matrix is denoted by $I_n$, the $n \times n$ zero matrix by $Z_n$, and the zero vector by $\theta_n$.

An inner product is a generalization of the dot product. In a vector space, it is a way to multiply vectors together, with the result of this multiplication being a scalar.

More precisely, for a real vector space, an inner product $\langle \cdot, \cdot \rangle$ satisfies the following four properties.

Let $u$, $v$, and $w$ be vectors and $\alpha$ be a scalar, then:

**Linearity:**

$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.

$\langle \alpha v, w \rangle = \langle \alpha v, w \rangle$.

**Symmetry:**

$\langle v, w \rangle = \langle w, v \rangle$.

**Positive-definiteness:**

$\langle v, v \rangle \geq 0$ and equal if and only if $v = 0$.

A vector space together with an inner product on it is called an inner product space.

When given a complex vector space, the symmetry property above is usually replaced by

**Conjugate Symmetry:**

$\langle v, w \rangle = \overline{\langle w, v \rangle}$, where $\overline{\langle w, v \rangle}$ refers to complex conjugation.

With this property, the inner product is called a Hermitian inner product and a complex vector space with a Hermitian inner product is called a Hermitian inner product space.

An inner product space is a vector space with the additional inner product structure. An inner product naturally induces an associated norm,

$$\|u\| = \sqrt{\langle u, u \rangle}.$$  

Thus an inner product space is also a normed vector space.

**Examples of Inner Product Space.**

(a) The Euclidean inner product on $\mathbb{F}^n$, where $\mathbb{F}$ is the set of real or complex number is defined by

$$\langle (u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_n) \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$  

(b) An inner product can be defined on the vector space of smooth real-valued functions on the interval $[-1, 1]$ by

$$\langle f(x), g(x) \rangle = \int_{-1}^{1} f(x) g(x) \, dx.$$  

*California State University, East Bay*
(c) An inner product can be defined on the set of real polynomials by
\[ \langle p(x), q(x) \rangle = \int_{0}^{\infty} p(x) q(x) e^{-x} \, dx. \]

Examples of Vector Norms.
Here are three mostly used vector norms:
(a) $1$-norm: $\|v\|_1 = \sum_i |v_i|.$
(b) $2$-norm or Euclidean norm: $\|v\|_2 = \sqrt{\sum_i |v_i|^2}.$
(c) $\infty$-norm: $\|v\|_\infty = \max |x_i|.$

Also, note that if $\|\cdot\|$ is a norm and $M$ is any nonsingular square matrix, then $v \rightarrow \|Mv\|$ is also a norm. The case where $M$ is diagonal is particularly common in practice.

Note. Any Hermitian positive definite matrix $H$ induces an inner product $\langle u, v \rangle_H = u^* H v$; and thus a norm $\|u\|_H = \langle u, u \rangle_H$.

In MatLab, the $1$-norm, $2$-norm, and $\infty$-norm are invoked by the statements `norm(u, 1)`, `norm(u, 2)`, and `norm(u, inf)`, respectively. The $2$-norm is the default in MatLab. The statement `norm(u)` is interpreted as the $2$-norm of $u$ in MatLab.

Theorem. If $\|\cdot\|$ is a norm induce by an inner product, then
1. $\|u\| = 0$, if and only if $u = \theta$, otherwise $\|u\| > 0$;
2. $\|\lambda u\| = |\lambda| \|u\|$;
3. (Cauchy-Schwarz inequality) $|\langle u, v \rangle| \leq \|u\| \|v\|$;
4. (Triangle inequality) $\|u + v\| \leq \|u\| + \|v\|$.

Orthogonality
In an inner product vector space $V$, vectors $u$ and $v$ are orthogonal, if $\langle u, v \rangle = 0$. A subset $S$ of $V$ is orthogonal, if any two distinct vectors of $S$ are orthogonal. A vector $u$ is a unit vector, if $\|u\| = 1$. Finally a subset $S$ of $V$ is orthonormal, if $S$ is orthogonal and consists entirely of unit vectors.

If $v \neq \theta$, then the vector $u = \frac{v}{\|v\|}$ is a unit vector; we say that $v$ was normalized.

Note. The zero vector $\theta$ is orthogonal to every vector in $V$ and $\theta$ is the only vector in $V$ that is orthogonal to itself.

Pythagorean Theorem. Suppose $u$ and $v$ are orthogonal vectors in $V$. Then
\[ \|u + v\|^2 = \|u\|^2 + \|v\|^2. \]

Proof. we have
\[ \|u + v\|^2 = |\langle u + v, u + v \rangle| = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle = \|u\|^2 + \|v\|^2 \]
as desired.
The **Cauchy-Schwarz inequality** states that for all vectors \( u \) and \( v \) of an inner product space:

\[
|\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle,
\]

where \( \langle \cdot, \cdot \rangle \) is the inner product.

Equivalently, by taking the square root of both sides, and referring to the norms of the vectors, the inequality is written as

\[
|\langle u, v \rangle| \leq \| u \| \| v \|.
\]

Moreover, the two sides are equal if and only if \( u \) and \( v \) are linearly dependent.

**Proof.** Let \( u, v \) be arbitrary vectors in a vector space \( V \) over \( \mathbb{F} \) with an inner product, where \( \mathbb{F} \) is the field of real or complex numbers.

If \( \langle u, v \rangle = 0 \), the theorem holds trivially.

If not, then \( u \neq 0, v \neq 0 \). For any \( \lambda \in \mathbb{F} \), we have

\[
0 \leq \| u - \lambda v \|^2 = \langle u - \lambda v, u - \lambda v \rangle = \langle u, u - \lambda v \rangle - \lambda \langle v, u - \lambda v \rangle
\]

\[
= \langle u, u \rangle - \lambda \langle v, u \rangle - \lambda \langle v, u \rangle + \lambda \overline{\lambda} \langle v, v \rangle.
\]

Choose \( \lambda = \frac{\langle u, v \rangle}{\langle v, v \rangle} \), then

\[
0 \leq \langle u, u \rangle - \overline{\lambda} \langle u, v \rangle - \lambda \langle v, u \rangle + \lambda \overline{\lambda} \langle v, v \rangle = \langle u, u \rangle - \frac{|\langle u, v \rangle|^2}{\langle v, v \rangle} = \| u \|^2 - \frac{|\langle u, v \rangle|^2}{\| v \|^2}.
\]

It follows that \( |\langle u, v \rangle|^2 \leq \langle u, u \rangle \langle v, v \rangle \) or \( |\langle u, v \rangle| \leq \| u \| \| v \| \).

**Examples of the Cauchy-Schwarz Inequality**

(a) If \( u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_n \) are real numbers, then

\[
|u_1 v_1 + u_2 v_2 + \cdots + u_n v_n|^2 \leq \left( u_1^2 + u_2^2 + \cdots + u_n^2 \right) \left( v_1^2 + v_2^2 + \cdots + v_n^2 \right).
\]

(b) If \( f(x) \) and \( g(x) \) are smooth real-valued functions, then

\[
\left| \int_{-1}^{1} f(x) g(x) \, dx \right| \leq \left( \int_{-1}^{1} f(x)^2 \, dx \right) \left( \int_{-1}^{1} g(x)^2 \, dx \right).
\]

**Triangle Inequality.** For any vectors \( u \) and \( v \),

\[
\| u + v \| \leq \| u \| + \| v \|.
\]

**Proof.** We have

\[
\| u + v \|^2 = \langle u + v, u + v \rangle = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle
\]

\[
= \| u \|^2 + 2 \Re \langle u, v \rangle + \| v \|^2
\]

\[
\leq \| u \|^2 + 2 \| u \| \| v \| + \| v \|^2
\]

\[
\leq \| u \|^2 + 2 \| u \| \| v \| + \| v \|^2
\]

\[
= \| u \|^2 + \| v \|^2.
\]
Gram-Schmidt Process

It is often more simpler to work in an orthogonal basis. Gram-Schmidt process is a method for obtaining orthonormal vectors from a linearly independent set of vectors in an inner product space, most commonly the Euclidean space \( \mathbb{R}^n \). \( S = \{v_1, v_2, v_3, \ldots, v_k\} \) and generates an orthogonal set \( S_{\perp G} = \{u_1, u_2, u_3, \ldots, u_k\} \) or orthonormal basis \( S_{\perp N} = \{e_1, e_2, e_3, \ldots, e_k\} \) that spans the same subspace of \( \mathbb{R}^n \) as the set \( S \).

We define the projection operator by \( \text{proj}_u(v) = \frac{\langle v, u \rangle}{\langle u, u \rangle} u \), where \( \langle v, u \rangle \) denotes the inner product of the vectors \( v \) and \( u \). This operator projects the vector \( v \) orthogonally onto the line spanned by vector \( u \).

The Gram-Schmidt process then works as follows:

\[
\begin{align*}
  u_1 &= v_1, \\
  u_2 &= v_2 - \text{proj}_{u_1}(v_2), \\
  u_3 &= v_3 - \text{proj}_{u_1}(v_3) - \text{proj}_{u_2}(v_3), \\
  u_4 &= v_4 - \text{proj}_{u_1}(v_4) - \text{proj}_{u_2}(v_4) - \text{proj}_{u_3}(v_4), \\
  & \vdots \\
  u_k &= v_k - \sum_{j=1}^{k-1} \text{proj}_{u_j}(v_k),
\end{align*}
\]

The sequence \( \{u_1, u_2, \ldots, u_k\} \) is the required system of orthogonal vectors, and the normalized vectors \( \{e_1, e_2, \ldots, e_k\} \) form an orthonormal set. The calculation of the sequence \( u_1, u_2, \ldots, u_k \) is known as Gram-Schmidt orthogonalization, while the calculation of the sequence \( e_1, e_2, \ldots, e_k \) is known as Gram-Schmidt orthonormalization as the vectors are normalized.

Example. Consider the basis \( B = \left\{ v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \right\} \). Then use Gram-Schmidt to find the orthogonal basis \( B_{\perp G} = \{u_1, u_2, u_3\} \) and the orthonormal basis \( B_{\perp N} = \{e_1, e_2, e_3\} \) associated with \( B \).

\[
\begin{align*}
  u_1 &= v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}; \\
  u_2 &= v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} - \frac{2}{3} egin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}; \\
  u_3 &= v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} - \frac{2}{3} egin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{3} egin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}; \\
  e_1 &= \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}; \\
  e_2 &= \frac{u_2}{\|u_2\|} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}; \\
  e_3 &= \frac{u_3}{\|u_3\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.
\end{align*}
\]

Hence

\[
B_{\perp G} = \left\{ u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\} \quad \text{and} \quad B_{\perp N} = \left\{ e_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad e_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad e_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}.
\]
Matrix Norms

A matrix norm is an extension of the notion of a vector norm to matrices. It is a vector norm on $\mathbb{F}^{m \times n}$. That is, if $\|A\|$ denotes the norm of the matrix $A$, then,

1. $\|A\| \geq 0$;
2. $\|A\| = 0$ if and only if $A = 0$;
3. $\|\alpha A\| = |\alpha| \|A\|$ for all $\forall \alpha \in \mathbb{F}$ and $\forall A \in \mathbb{F}^{m \times n}$;
4. $\forall A, B \in \mathbb{F}^{m \times n}, \|A + B\| \leq \|A\| + \|B\|$.

Additionally, in the case of square matrices (thus, $m = n$), some (but not all) matrix norms satisfy the following condition, which is related to the fact that matrices are more than just vectors:

5. $\|AB\| \leq \|A\|\|B\|$ for all matrices $A$ and $B$ in $\mathbb{F}^{m \times n}$.

A matrix norm that satisfies this additional property is called a sub-multiplicative norm.

Operator Norm

If vector norms on $\mathbb{F}^m$ and $\mathbb{F}^n$ are given ($\mathbb{F}$ is the field of real or complex numbers), then one defines the corresponding operator norm or induced norm or, on the space of $m$-by-$n$ matrices as the following maxima:

$$
\|A\| = \sup\{\|Ax\| : x \in \mathbb{F}^n, \|x\| = 1\} = \sup \left\{ \frac{\|Ax\|}{\|x\|} : x \in \mathbb{F}^n, x \neq \theta \right\}.
$$

The operator norm corresponding to the p-norm for vectors is:

$$
\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.
$$

Note. All operator norms are sub-multiplicative.

Here are some example matrix norms:

**Column norm or 1-norm:** $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|$, which is simply the maximum absolute column sum of the matrix.

**Row norm or infinity-norm:** $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$, which is simply the maximum absolute row sum of the matrix.

The 2-norm induced by the euclidean vector norm is $\|A\| = \lambda_{max}$, where $\lambda_{max}$ is the largest eigenvalue of the matrix $A^*A$.

Although the Frobenius norm, defined as:

$$
\|A\|_F = \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{n} |a_{ij}|^2} = \text{trace} (A^*A)
$$

California State University, East Bay
is not an induced norm but it is a sub-multiplicative norm. In MatLab, the Frobenius norm is invoked by the statement \( \text{norm}(A, 'fro') \).

The max norm \( \| A \|_{\text{max}} = \max |a_{ij}| \) is neither an induced norm nor a sub-multiplicative norm.

The largest eigenvalue of a square matrix \( A \) in modulus, denoted by \( \rho(A) \) and called the spectral radius of \( A \), never exceeds its operator norm:

\[
\rho(A) \leq \| A \|_p.
\]

Given two normed vector spaces \( V \) and \( W \) (over the same base field, either the real numbers or the complex numbers), a linear map \( A : V \to W \) is continuous if and only if there exists a real number \( \lambda \) such that for all \( v \in V \),

\[
\| Av \| \leq \lambda \| v \|.
\]

**Condition Number**

The condition number associated with the linear equation \( Ax = b \), gives a bound on how inaccurate the solution \( x \) will be after approximation. Note that this is before the effects of round-off error are taken into account; conditioning is a property of the matrix, not the algorithm or floating point accuracy of the computer used to solve the corresponding system.

Let \( x_0 \) be the solution of the linear equation \( Ax = b \) and let \( x' \) be the calculated solution of \( Ax = b \), so the error in the solution is \( e = x' - x_0 \), and the relative error in the solution is \( \| e \| = \frac{\| x' - x_0 \|}{\| x_0 \|} \).

Generally, we can’t verify the error or relative error (we don’t know \( x_0 \)), so a possible way to check this is by testing the accuracy back in the system \( Ax = b \):

\[
\text{residual } r = Ax_0 - Ax' = b - b',
\]

and relative residual is \( \frac{\| Ax_0 - Ax' \|}{\| Ax \|} = \frac{\| b - b' \|}{\| b \|} = \frac{\| r \|}{\| b \|} \).

A matrix \( A \) is said to be **ill-conditioned**, if relatively small changes in the input (in the matrix \( A \)) can cause large change in the output (the solution of \( Ax = b \)), i.e.; the solution is not very accurate if input is rounded. Otherwise it is **well-conditioned**.

The condition number of an invertible matrix \( A \) is defined to be \( \kappa(A) = \| AA^{-1} \| \). This quantity is always greater than or equal to 1.

The condition number is defined more precisely to be the maximum ratio of the relative error in \( x \) divided by the relative error in \( b \).

If \( A \) is nonsingular, then by knowing the eigenvalues of \( A^*A \) (\( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0 \)), we can compute the condition number using the 2-norm:

\[
\kappa_2(A) = \| A_2 A^{-1} \|_2 = \frac{\lambda_1}{\lambda_n}.
\]

*California State University, East Bay*
The smallest nonzero eigenvalue $\lambda_n$ of $A^*A$ is a measure of how close the matrix is to being singular. If $\lambda_n$ is small, then $\kappa_2(A)$ is large. Thus, the closer the matrix is to being singular, the more ill-conditioned it is.

**Consistent Norms**

A matrix norm $\| \cdot \|_{ab}$ on $\mathbb{F}^{m\times n}$ is called consistent with a vector norm $\| \cdot \|_a$ on $\mathbb{F}^n$ and a vector norm $\| \cdot \|_b$ on $\mathbb{F}^m$, if:

$$\|Ax\|_b \leq \|A\|_{ab} \|u\|_a, \text{ for all } A \in \mathbb{F}^{m\times n}, u \in \mathbb{F}^n.$$  

All induced norms are consistent by definition.

**Compatible Norms**

A matrix norm $\| \cdot \|_b$ on $\mathbb{F}^{n\times n}$ is called compatible with a vector norm $\| \cdot \|_a$ on $\mathbb{F}^n$ if:

$$\|Ax\|_a \leq \|A\|_b \|x\|_a, \text{ for all } A \in \mathbb{F}^{m\times n}, x \in \mathbb{F}^n.$$  

All induced norms are compatible by definition.

**Equivalence of Norms**

Two matrix norms $\| \cdot \|_\alpha$ and $\| \cdot \|_\beta$ are said to be equivalent, if for all matrices $A \in \mathbb{F}^{m\times n}$, there exist some positive numbers $r$ and $s$ such that

$$r \|A\|_\alpha \leq \|A\|_\beta \leq s \|A\|_\alpha.$$  

**Examples of Norm Equivalence**

For matrix $A \in \mathbb{R}^{n\times m}$ of rank $r$, the following inequalities hold:

- $\|A\|_2 \leq \|A\|_F \leq \sqrt{r}\|A\|_2$.
- $\|A\|_F \leq \|A\|_* \leq \sqrt{r}\|A\|_F$.
- $\|A\|_{\text{max}} \leq \|A\|_2 \leq \sqrt{mn}\|A\|_{\text{max}}$.
- $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\|A\|_\infty$.
- $\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$.

Here, $\|A\|_p$ refers to the matrix norm induced by the vector p-norm.

Another useful inequality between matrix norms is:

- $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$.