CS724: Topics in Algorithms
Norms and Inner Products - II
Slide Set 5

Prof. Dan A. Simovici
1 Norms for Matrices
2 Inner Products
3 Hyperplanes in $\mathbb{R}^n$
4 Unitary and Orthogonal Matrices
5 Projection on Subspaces
6 Positive Definite and Positive Semidefinite Matrices
7 The Gram-Schmidt Orthogonalization Algorithm
8 The QR Factorization of Matrices
9 MATLAB Computations
The set $\mathbb{C}^{m \times n}$ is a linear space. Therefore, it is natural to consider norms defined on matrices. We discuss two basic methods for defining norms for matrices.

- The first approach treats matrices as vectors (through the vec mapping).
- The second, regards matrices as representations of linear operators, and defined norms for matrices starting from operator norms.
Definition

The \((m \times n)\)-vectorization mapping is the mapping \(\text{vec} : \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^{mn}\) defined by

\[
\text{vec}(A) = \begin{pmatrix}
a_{11} \\
\vdots \\
a_{m1} \\
\vdots \\
a_{1n} \\
\vdots \\
a_{mn}
\end{pmatrix},
\]

obtained by reading \(A\) column-wise.
The following equality is immediate for a matrix $A \in \mathbb{C}^{m \times n}$:

$$\text{vec}(A) = \begin{pmatrix} Ae_1 \\ Ae_2 \\ \vdots \\ Ae_n \end{pmatrix}.$$ 

The vectorization mapping vec is an isomorphism between the linear space $\mathbb{C}^{m \times n}$ and the linear space $\mathbb{C}^{mn}$, as can be easily verified.
Example

For the matrix $I_n$ we have

$$\text{vec}(I_n) = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}.$$
Definition

Let $\nu$ be a vector norm on the space $\mathbb{R}^{mn}$. The vectorial matrix norm $\mu^{(m,n)}$ on $\mathbb{R}^{m \times n}$ is the mapping $\mu^{(m,n)} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}_{\geq 0}$ defined by

$$\mu^{(m,n)}(A) = \nu(\text{vec}(A)),$$

for $A \in \mathbb{R}^{m \times n}$.

Vectorial norms of matrices are defined without regard for matrix products.
A **consistent family of matrix norms** is a family of functions 
\( \mu^{(m,n)} : \mathbb{C}^{m \times n} \rightarrow \mathbb{R}_{\geq 0} \), where \( m, n \in \mathbb{P} \) that satisfies the following conditions:

- \( \mu^{(m,n)}(A) = 0 \) if and only if \( A = O_{m,n} \);
- \( \mu^{(m,n)}(A + B) \leq \mu^{(m,n)}(A) + \mu^{(m,n)}(B) \) (the subadditivity property);
- \( \mu^{(m,n)}(aA) = |a|\mu^{(m,n)}(A) \);
- \( \mu^{(m,p)}(AB) \leq \mu^{(m,n)}(A)\mu^{(n,p)}(B) \) for every matrix \( A \in \mathbb{R}^{m \times n} \) and \( B \in \mathbb{R}^{n \times p} \) (the submultiplicative property).

If the format of the matrix \( A \) is clear from context or is irrelevant, then we shall write \( \mu(A) \) instead of \( \mu^{(m,n)}(A) \).
Example

Let $P \in \mathbb{C}^{n \times n}$ be an idempotent matrix, that is, a matrix $P$ such that $P^2 = P$. If $\mu$ is a matrix norm, then either $\mu(P) = 0$ or $\mu(P) \geq 1$. Indeed, since $P$ is idempotent we have $\mu(P) = \mu(P^2)$. By the submultiplicative property, $\mu(P^2) \leq (\mu(P))^2$, so $\mu(P) \leq (\mu(P))^2$. Consequently, if $\mu(P) \neq 0$, then $\mu(P) \geq 1$. 
Some vectorial matrix norms turn out to be actual matrix norms; others fail to be matrix norms. This point is illustrated by the next examples.
Example

Consider the vectorial matrix norm $\mu_1$ induced by the vector norm $\nu_1$. We have $\mu_1(A) = \sum_{i=1}^n \sum_{j=1}^m |a_{ij}|$ for $A \in \mathbb{R}^{m \times n}$. Actually, this is a matrix norm. To prove this fact consider the matrices $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{p \times n}$. We have:

$$
\mu_1(AB) = \sum_{i=1}^m \sum_{j=1}^n \left| \sum_{k=1}^p a_{ik} b_{kj} \right| \leq \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^p |a_{ik} b_{kj}|
$$

$$
\leq \sum_{i=1}^m \sum_{j=1}^n \sum_{k'=1}^p \sum_{k''=1}^p |a_{ik'}| |b_{k''j}|
$$

(because we added extra non-negative terms to the sums)

$$
= \left( \sum_{i=1}^m \sum_{k'=1}^p |a_{ik'}| \right) \cdot \left( \sum_{j=1}^n \sum_{k''=1}^p |b_{k''j}| \right)
$$

$$
= \mu_1(A) \mu_1(B).
$$

We denote this vectorial matrix norm by the same notation as the corresponding vector norm, that is, by $\| A \|_1$. 

Example

The vectorial norm of $A \in \mathbb{C}^{m \times n}$,

$$
\mu_2(A) = \left( \sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}},
$$
denoted also by $\| A \|_F$, is known as the **Frobenius norm**. For $A \in \mathbb{R}^{m \times n}$ we have

$$
\| A \|_F = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2}.
$$

It is easy to see that for real matrices we have

$$
\| A \|_F^2 = \text{trace}(AA') = \text{trace}(A'A).
$$

For complex matrices the corresponding equality is

$$
\| A \|_F^2 = \text{trace}(AA^H) = \text{trace}(A^HA).
$$

Note that $\| A^H \|_F^2 = \| A \|_F^2$ for every $A$. 
Example

The vectorial norm $\mu_\infty$ induced by the vector norm $\nu_\infty$ is denoted by $\| A \|_\infty$ and is given by

$$\| A \|_\infty = \max_{i,j} |a_{ij}|$$

for $A \in \mathbb{C}^{n \times n}$. This is not a matrix norm. Indeed, let $a, b$ be two positive numbers and consider the matrices

$$A = \begin{pmatrix} a & a \\ a & a \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b & b \\ b & b \end{pmatrix}.$$ 

We have $\| A \|_\infty = a$ and $\| B \|_\infty = b$. However, since

$$AB = \begin{pmatrix} 2ab & 2ab \\ 2ab & 2ab \end{pmatrix},$$

we have $\| AB \|_\infty = 2ab$ and the submultiplicative property of matrix norms is violated.
A technique that always produces matrix norms starting from vector norms is introduced in the next theorem.

**Definition**

Let $\nu_m$ be a norm on $\mathbb{C}^m$ and $\nu_n$ be a norm on $\mathbb{C}^n$ and let $A \in \mathbb{C}^{n \times m}$ be a matrix. The **operator norm** of $A$ is the number $\mu^{(n,m)}(A) = \mu^{(n,m)}(h_A)$, where $\mu^{(n,m)} = N(\nu_m, \nu_n)$. 
Theorem

Let $\mu$ be the matrix norm on $\mathbb{C}^{n \times n}$ induced by the vector norm $\nu$. We have $\nu(Au) \leq \mu(A)\nu(u)$ for every $u \in \mathbb{C}^n$.

Proof.

The inequality is obviously satisfied when $u = 0_n$. Therefore, we may assume that $u \neq 0_n$ and let $x = \frac{1}{\nu(u)}u$. Clearly, $\nu(x) = 1$ and

$$\nu \left( A \frac{1}{\nu(u)}u \right) \leq \mu(A)$$

for every $u \in \mathbb{C}^n - \{0_n\}$. This implies immediately the desired inequality.
If $\mu$ is a matrix norm induced by a vector norm on $\mathbb{R}^n$, then

$$\mu(I_n) = \sup\{\nu(l_n x) \mid \nu(x) \leq 1\} = 1.$$ This necessary condition can be used for identifying matrix norms that are not induced by vector norms. The operator matrix norm induced by the vector norm $\| \cdot \|_p$ is denoted by $\| \cdot \|_p$. 
Example

To compute $\|A\|_1 = \sup\{\|Ax\|_1 : \|x\|_1 \leq 1\}$, where $A \in \mathbb{R}^{n \times n}$, suppose that the columns of $A$ are the vectors $a_1, \ldots, a_n$, that is

$$a_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$ 

Let $x \in \mathbb{R}^n$ be a vector whose components are $x_1, \ldots, x_n$. Then, $Ax = x_1a_1 + \cdots + x_na_n$, so

$$\|Ax\|_1 = \|x_1a_1 + \cdots + x_na_n\|_1 \\
\leq \sum_{j=1}^{n} |x_j| \|a_j\|_1 \\
\leq \max_j \|a_j\|_1 \sum_{j=1}^{n} |x_j| \\
= \max_j \|a_j\|_1 \cdot \|x\|_1.$$
Example cont’d

Let $e_j$ be the vector whose components are 0 with the exception of its $j^{th}$ component that is equal to 1. Clearly, we have $\|e_j\|_1 = 1$ and $a_j = Ae_j$. This, in turn implies $\|a_j\|_1 = \|Ae_j\|_1 \leq \|A\|_1$ for $1 \leq j \leq n$. Therefore, $\max_j \|a_j\|_1 \leq \|A\|_1$, so

$$\|A\|_1 = \max_j \|a_j\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|.$$ 

In other words, $\|A\|_1$ equals the maximum column sum of the absolute values.
Example

Consider now a matrix $A \in \mathbb{R}^{n \times n}$. We have

$$
\|Ax\|_\infty = \max_{1 \leq i \leq n} \left| \sum_{j=1}^{n} a_{ij}x_j \right|
$$

$$
\leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}x_j|
$$

$$
\leq \max_{1 \leq i \leq n} \|x\|_\infty \sum_{j=1}^{n} |a_{ij}|
$$

Consequently, if $\|x\|_\infty \leq 1$ we have $\|Ax\|_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$

Thus, $\|A\|_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|$. 
The converse inequality is immediate if $A = O_{n,n}$. Therefore, assume that
$A \neq O_{n \times n}$, and let $(a_{p1}, \ldots, a_{pn})$ be any row of $A$ that has at least one element
distinct from 0. Define the vector $z \in \mathbb{R}^n$ by

$$z_j = \begin{cases} \frac{|a_{pj}|}{a_{pj}} & \text{if } a_{pj} \neq 0, \\ 1 & \text{otherwise,} \end{cases}$$

for $1 \leq j \leq n$. It is clear that $z_j \in \{-1, 1\}$ for every $j$, $1 \leq j \leq n$ and, therefore,
$\|z\|_\infty = 1$. Moreover, we have $|a_{pj}| = a_{pj}z_j$ for $1 \leq j \leq n$. Therefore, we can write:

$$\sum_{j=1}^n |a_{pj}| = \sum_{j=1}^n a_{pj}z_j \leq \max_{1 \leq i \leq n} \left| \sum_{j=1}^n a_{ij}z_j \right| = \|Az\|_\infty \leq \max \{\|Ax\|_\infty : \|x\|_\infty \leq 1\} = \|A\|_\infty.$$
Example cont’d

Example

Since this holds for every row of $A$, it follows that

$$\max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}| \leq \|A\|_{\infty},$$

which proves that

$$\|A\|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|.$$ 

In other words, $\|A\|_{\infty}$ equals the maximum row sum of the absolute values.
Example

Let $D = \text{diag}(d_1, \ldots, d_n) \in \mathbb{C}^{n \times n}$ be a diagonal matrix. If $\mathbf{x} \in \mathbb{C}^n$ we have

$$D\mathbf{x} = \begin{pmatrix} d_1 x_1 \\ \vdots \\ d_n x_n \end{pmatrix} ,$$

so

$$\| D \|_2 = \max \{ \| D\mathbf{x} \|_2 \mid \| \mathbf{x} \|_2 = 1 \}$$

$$= \max \{ \sqrt{(d_1 x_1)^2 + \cdots + (d_n x_n)^2} \mid x_1^2 + \cdots + x_n^2 = 1 \}$$

$$= \max \{ |d_i| \mid 1 \leq i \leq n \} .$$
Certain norms are invariant with respect to multiplication by unitary matrices. We refer to these norms as \textit{unitarily invariant norms}.

\textbf{Theorem}

Let \( U \in \mathbb{C}^{n \times n} \) be a unitary matrix. The following statements hold:

- \( \| Ux \|_2 = \| x \|_2 \) for every \( x \in \mathbb{C}^n \);
- \( \| UA \|_2 = \| A \|_2 \) for every \( A \in \mathbb{C}^{n \times p} \);
- \( \| UA \|_F = \| A \|_F \) for every \( A \in \mathbb{C}^{n \times p} \).
Proof

For the first part of the theorem note that

\[ \| Ux \|_2^2 = (Ux)^H Ux = x^H U^H Ux = x^H x = \| x \|_2^2, \]

because \( U^H A = I_n \).

The second part of the theorem is shown next:

\[ \| UA \|_2 = \max \{ \| (UA)x \|_2 \mid x \|_2 = 1 \} \]
\[ = \max \{ \| U(Ax) \|_2 \mid x \|_2 = 1 \} \]
\[ = \max \{ \| Ax \|_2 \mid x \|_2 = 1 \} \]
\[ (\text{by Part (i)}) \]
\[ = \| A \|_2. \]
Proof cont’d

For the Frobenius norm note that

\[ \| UA \|_F = \sqrt{\text{trace}((UA)^HUA)} = \sqrt{\text{trace}(A^HU^HUA)} = \sqrt{\text{trace}(A^HA)} = \| A \|_F \]
Corollary

If $U \in \mathbb{C}^{n \times n}$ is a unitary matrix, then $\|U\|_2 = 1$.

Proof.

Since $\|U\|_2 = \sup\{\|Ux\|_2 \mid \|x\|_2 \leq 1\}$, we have

$$\|U\|_2 = \sup\{\|x\|_2 \mid \|x\|_2 \leq 1\} = 1.$$
Corollary

Let $A, U \in \mathbb{C}^{n \times n}$. If $U$ is an unitary matrix, then

$$\| U^H A U \|_F = \| A \|_F.$$ 

Proof.

Since $U$ is a unitary matrix, so is $U^H$. By a previous Theorem,

$$\| U^H A U \|_F = \| A U \|_F = \| U^H A^H \|_F^2 = \| A^H \|_F^2 = \| A \|_F^2,$$

which proves the corollary.
Example

Let \( S = \{ \mathbf{x} \in \mathbb{R}^n \mid \| \mathbf{x} \|_2 = 1 \} \) be the surface of the sphere in \( \mathbb{R}^n \). The image of \( S \) under the linear transformation \( h_U \) that corresponds to the unitary matrix \( U \) is \( S \) itself. Indeed, \( \| h_U(\mathbf{x}) \|_2 = \| \mathbf{x} \|_2 = 1 \), so \( h_U(\mathbf{x}) \in S \) for every \( \mathbf{x} \in S \). Also, note that \( h_U \) restricted to \( S \) is a bijection because \( h_U^H(h_U(\mathbf{x})) = \mathbf{x} \) for every \( \mathbf{x} \in \mathbb{R}^n \).
Theorem

Let $A \in \mathbb{R}^{n \times n}$. We have $\|A\|_2 \leq \|A\|_F$.

Proof.

Let $x \in \mathbb{R}^n$. We have

$$Ax = \begin{pmatrix} r_1 x \\ \vdots \\ r_n x \end{pmatrix},$$

where $r_1, \ldots, r_n$ are the rows of the matrix $A$. Thus,

$$\frac{\|Ax\|_2}{\|x\|_2} = \frac{\sqrt{\sum_{i=1}^n (r_i x)^2}}{\|x\|_2}.$$

By Cauchy-Schwarz inequality we have: $(r_i x)^2 \leq \|r_i\|_2^2 \|x\|_2^2$, so

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \sqrt{\sum_{i=1}^n \|r_i\|_2^2} = \|A\|_F.$$

This implies $\|A\|_2 \leq \|A\|_F$. □
Definition

Let \( L \) be a \( \mathbb{C} \)-linear space. An *inner product* on \( L \) is a function \( f : L \times L \rightarrow \mathbb{C} \) that has the following properties:

- \( f(ax + by, z) = af(x, z) + bf(y, z) \) (linearity in the first argument);
- \( f(x, y) = f(y, x) \) for \( y, x \in L \) (conjugate symmetry);
- if \( x \neq 0 \), then \( f(x, x) \) is a positive real number (positivity),
- \( f(x, x) = 0 \) if and only if \( x = 0 \) (definiteness),

for every \( x, y, z \in L \) and \( a, b \in \mathbb{C} \).

The pair \((L, f)\) is called an *inner product space*.

An alternative terminology for real inner product spaces is *Euclidean spaces*, and *Hermitian spaces* for complex inner product spaces.
For the second argument of an inner product we have the property of \textit{conjugate linearity}, that is,

\begin{equation*}
f(z, ax + by) = \bar{a}f(z, x) + \bar{b}f(z, y)
\end{equation*}

for every \(x, y, z \in L\) and \(a, b \in \mathbb{C}\). Indeed, by the conjugate symmetry property we can write

\begin{align*}
f(z, ax + by) &= f(ax + by, z) \\
&= af(x, z) + bf(y, z) \\
&= \bar{a}f(x, z) + \bar{b}f(y, z) \\
&= \bar{a}f(z, x) + \bar{b}f(z, y).
\end{align*}
Observe that conjugate symmetry property on inner products implies that for $x \in L$, $f(x, x)$ is a real number because $f(x, x) = \overline{f(x, x)}$. When $L$ is a real linear space the definition of the inner product becomes simpler because the conjugate of a real number $a$ is $a$ itself. Namely, for real linear spaces, the conjugate symmetry is replaced by the plain symmetry property,

$$f(x, y) = f(y, x),$$

for $x, y \in L$ and $f$ is linear in both arguments.
Let \( W = \{w_1, \ldots, w_n\} \) be a basis in the complex \( n \)-dimensional inner product space \( L \). If \( x = \sum_{i=1}^{n} x^i w_i \) and \( y = \sum_{j=1}^{n} y^j w_j \), then

\[
f(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} x^i y^j f(w_i, w_j),
\]
due to the bilinearity of the inner product. If we denote \( f(w_i, w_j) \) by \( g_{ij} \), then \( f(x, y) \) can be written as

\[
f(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} x^i y^j g_{ij}
\]

for \( x, y \in L \).

If \( L \) is a real inner product space \( L \), then

\[
f(x, y) = \sum_{i=1}^{n} \sum_{j=1}^{n} x^i y^j g_{ij}
\]

To simplify notations, if there is no risk of confusion, we denote the inner product \( f(u, v) \) as \((u, v)\).  

---

Prof. Dan A. Simovici  
CS724: Topics in Algorithms Norms and Inner Products - II Slide Set 5  
UMASS BOSTON
Two vectors $u, v \in \mathbb{C}^n$ are said to be orthogonal with respect to an inner product if $(u, v) = 0$. This is denoted by $x \perp y$.

An orthogonal set of vectors in an inner product space $L$ equipped with an inner product is a subset $W$ of $L$ such that for every $u, v \in W$ we have $u \perp v$. 
Theorem

Any inner product on a linear space $L$ generates a norm on that space defined by $\| x \| = \sqrt{(x, x)}$ for $x \in L$. 
Proof

Let $L$ be a $\mathbb{C}$-linear space. We need to verify that the norm satisfies the conditions of Definition. Applying the properties of the inner product we have

\[
\left\| \mathbf{x} + \mathbf{y} \right\|^2 = (\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) \\
= (\mathbf{x}, \mathbf{x}) + 2(\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{y}) \\
= \left\| \mathbf{x} \right\|^2 + 2(\mathbf{x}, \mathbf{y}) + \left\| \mathbf{y} \right\|^2 \\
\leq \left\| \mathbf{x} \right\|^2 + 2\left\| \mathbf{x} \right\| \left\| \mathbf{y} \right\| + \left\| \mathbf{y} \right\|^2 \\
= \left( \left\| \mathbf{x} \right\| + \left\| \mathbf{y} \right\| \right)^2.
\]

Because $\left\| \mathbf{x} \right\| \geq 0$ it follows that $\left\| \mathbf{x} + \mathbf{y} \right\| \leq \left\| \mathbf{x} \right\| + \left\| \mathbf{y} \right\|$, which is the subadditivity property.

If $a \in \mathbb{C}$, then

\[
\left\| a\mathbf{x} \right\| = \sqrt{(a\mathbf{x}, a\mathbf{x})} = \sqrt{a\overline{a}(\mathbf{x}, \mathbf{x})} = \sqrt{|a|^2(\mathbf{x}, \mathbf{x})} = |a|\sqrt{(\mathbf{x}, \mathbf{x})} = |a| \left\| \mathbf{x} \right\|.
\]

From the definiteness property of the inner product it follows that $\left\| \mathbf{x} \right\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$. 
The norm induced by the inner product $f(x, y) = x^i y^j g_{ij}$ is

$$\| x \|^2 = f(x, x) = x^i x^j g_{ij}.$$
**Theorem**

*If* $W$ *is a set of orthogonal vectors in a* $n$-*dimensional* $\mathbb{C}$-*linear space* $L$ *and* $0 \notin W$, *then* $W$ *is linearly independent.*

**Proof.**

Let $c = a^1 w_1 + \cdots + a^n w_n$ *a linear combination in* $L$ *such that* $a^1 w_1 + \cdots + a^n w_n = 0$. *Since* $(c, w_i) = a_i \| w_i \|^2 = 0$, *we have* $a_i = 0$ *because* $\| w_i \|^2 \neq 0$, *and this holds for every* $i$, *where* $1 \leq i \leq n$. *Thus,* $W$ *is linearly independent.*
Definition

An *orthonormal set of vectors* in an inner product space $L$ equipped with an inner product is an orthogonal subset $W$ of $L$ such that for every $u$ we have $\| u \| = 1$, where the norm is induced by the inner product.

Corollary

*If $W$ is an orthonormal set of vectors in an $n$-dimensional $\mathbb{C}$-linear space $L$ and $|W| = n$, then $W$ is a basis in $L$.***
If \( W = \{w_1, \ldots, w_n\} \) is an orthonormal basis in \( \mathbb{C}^n \) we have

\[
g_{ij} = (w_i, w_j) = \begin{cases} 
0 & \text{if } i \neq j, \\
1 & \text{if } i = j,
\end{cases}
\]

which means that the inner product of the vectors \( x = x^i w_i \) and \( y = y^j w_j \) is given by:

\[
(x, y) = x^i y^j (w_i, w_j) = x^i y^i. \tag{2}
\]

Consequently, \( \| x \|^2 = \sum_{i=1}^{n} |x^i|^2 \).

The inner product of \( x, y \in \mathbb{R}^n \) is

\[
(x, y) = x^i y^j (w_i, w_j) = x^i y^i. \tag{3}
\]
Not every norm can be induced by an inner product. A characterization of this type of norms in linear spaces is presented next. This equality shown in the next theorem is known as the parallelogram equality.

**Theorem**

Let $L$ be a real linear space. A norm $\| \cdot \|$ is induced by an inner product if and only if

$$\| x + y \|^2 + \| x - y \|^2 = 2(\| x \|^2 + \| y \|^2),$$

for every $x, y \in L$. 
Proof

Suppose that the norm is induced by an inner product. In this case we can write for every $x$ and $y$:

$$\| x + y \|^2 = (x + y, x + y) = (x, x) + 2(x, y) + (y, y),$$
$$\| x - y \|^2 = (x - y, x - y) = (x, x) - 2(x, y) + (y, y).$$

Thus,

$$(x + y, x + y) + (x - y, x - y) = 2(x, x) + 2(y, y),$$

which can be written in terms of the norm generated as the inner product as

$$\| x + y \|^2 + \| x - y \|^2 = 2(\| x \|^2 + \| y \|^2).$$

The proof of the reverse implication is omitted.
Definition

Let \( \mathbf{w} \in \mathbb{R}^n - \{0\} \) and let \( a \in \mathbb{R} \). The hyperplane determined by \( \mathbf{w} \) and \( a \) is the set

\[
H_{\mathbf{w},a} = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{w}' \mathbf{x} = a \}.
\]
If $x_0 \in H_{\mathbf{w},a}$, then $\mathbf{w}'x_0 = a$, so $H_{\mathbf{w},a}$ is also described by the equality

$$H_{\mathbf{w},a} = \{ x \in \mathbb{R}^n \mid \mathbf{w}'(x - x_0) = 0 \}.$$  

Any hyperplane $H_{\mathbf{w},a}$ partitions $\mathbb{R}^n$ into three sets:

- $H_{\mathbf{w},a}^> = \{ x \in \mathbb{R}^n \mid \mathbf{w}'x > a \}$,
- $H_{\mathbf{w},a}^0 = H_{\mathbf{w},a}$,
- $H_{\mathbf{w},a}^< = \{ x \in \mathbb{R}^n \mid \mathbf{w}'x < a \}$.

The sets $H_{\mathbf{w},a}^>$ and $H_{\mathbf{w},a}^<$ are the positive and negative open half-spaces determined by $H_{\mathbf{w},a}$, respectively. The sets

- $H_{\mathbf{w},a}^\geq = \{ x \in \mathbb{R}^n \mid \mathbf{w}'x \geq a \}$,
- $H_{\mathbf{w},a}^\leq = \{ x \in \mathbb{R}^n \mid \mathbf{w}'x \leq a \}$.

are the positive and negative closed half-spaces determined by $H_{\mathbf{w},a}$, respectively.
If $x_1, x_2 \in H_{w,a}$ we say that the vector $x_1 - x_2$ is located in the hyperplane $H_{w,a}$. In this case $w \perp x_1 - x_2$. This justifies referring to $w$ as the normal to the hyperplane $H_{w,a}$. Observe that a hyperplane is fully determined by a vector $x_0 \in H_{w,a}$ and by $w$. 
Let $x_0 \in \mathbb{R}^n$ and let $H_{w,a}$ be a hyperplane. We seek $x \in H_{w,a}$ such that $\| x - x_0 \|_2$ is minimal. Finding $x$ amounts to minimizing the function $f(x) = \| x - x_0 \|_2^2 = \sum_{i=1}^{n} (x_i - x_{0i})^2$ subjected to the constraint $w_1 x_1 + \cdots + w_n x_n - a = 0$. Using the Lagrangian $\Lambda(x) = f(x) + \lambda (w'x - a)$ and the multiplier $\lambda$ we impose the conditions

$$\frac{\partial \Lambda}{\partial x_i} = 0 \text{ for } 1 \leq i \leq n$$

which amount to

$$\frac{\partial f}{\partial x_i} + \lambda w_i = 0$$

for $1 \leq i \leq n$. These equalities yield $2(x_i - x_{0i}) + \lambda w_i = 0$, so we have $x_i = x_{0i} - \frac{1}{2} \lambda w_i$. 
Consequently, we have $x = x_0 - \frac{1}{2} \lambda w$. Since $x \in H_{w,a}$ this implies

$$w'x = w'x_0 - \frac{1}{2} \lambda w'w = a.$$ 

Thus,

$$\lambda = 2 \frac{w'x_0 - a}{w'w} = 2 \frac{w'x_0 - a}{\|w\|^2_2}.$$ 

We conclude that the closest point in $H_{w,a}$ to $x_0$ is

$$x = x_0 - \frac{w'x_0 - a}{\|w\|^2_2} w.$$
The smallest distance between $\mathbf{x}_0$ and a point in the hyperplane $H_{\mathbf{w},a}$ is given by

$$
\| \mathbf{x}_0 - \mathbf{x} \| = \frac{\mathbf{w}'\mathbf{x}_0 - a}{\| \mathbf{w} \|_2}.
$$

If we define the distance $d(H_{\mathbf{w},a}, \mathbf{x}_0)$ between $\mathbf{x}_0$ and $H_{\mathbf{w},a}$ as this smallest distance we have

$$
d(H_{\mathbf{w},a}, \mathbf{x}_0) = \frac{\mathbf{w}'\mathbf{x}_0 - a}{\| \mathbf{w} \|_2}. \quad (4)
$$
Lemma

Let $A \in \mathbb{C}^{n \times n}$. If $x^H A x = 0$ for every $x \in \mathbb{C}^n$, then $A = O_{n,n}$. 
Proof

If $\mathbf{x} = \mathbf{e}_k$, then $\mathbf{x}^\mathsf{H} \mathbf{A} \mathbf{x} = a_{kk}$ for every $k, 1 \leq k \leq n$, so all diagonal entries of $\mathbf{A}$ equal 0. Choose now $\mathbf{x} = \mathbf{e}_k + \mathbf{e}_j$. Then,

$$(\mathbf{e}_k + \mathbf{e}_j)^\mathsf{H} \mathbf{A} (\mathbf{e}_k + \mathbf{e}_j)$$

$$= \mathbf{e}_k^\mathsf{H} \mathbf{A} \mathbf{e}_k + \mathbf{e}_k^\mathsf{H} \mathbf{A} \mathbf{e}_j + \mathbf{e}_j^\mathsf{H} \mathbf{A} \mathbf{e}_k + \mathbf{e}_j^\mathsf{H} \mathbf{A} \mathbf{e}_j$$

$$= \mathbf{e}_k^\mathsf{H} \mathbf{A} \mathbf{e}_j + \mathbf{e}_j^\mathsf{H} \mathbf{A} \mathbf{e}_k$$

$$= a_{kj} + a_{jk} = 0.$$
Similarly, if we choose $x = e_k + ie_j$ we obtain:

\[
(e_k + ie_j)^H A (e_k + ie_j) \\
= (e_k^H - ie_j^H) A (e_k + ie_j) \\
= e_k^H A e_k - ie_j^H A e_k + ie_k^H A e_j + e_j^H A e_j \\
= -ia_{jk} + ia_{kj} = 0.
\]

The equalities $a_{kj} + a_{jk} = 0$ and $-a_{jk} + a_{kj} = 0$ imply $a_{kj} = a_{jk} = 0$. Thus, all off-diagonal elements of $A$ are also 0, hence $A = O_{n,n}$. 
A matrix \( U \in \mathbb{C}^{n \times n} \) is unitary if \( \| Ux \|_2 = \| x \|_2 \) for every \( x \in \mathbb{C}^n \).
Proof

If $U$ is unitary we have

$$\| Ux \|_2^2 = (Ux)^H Ux = x^H U^H U x = \| x \|_2^2$$

because $U^H U = I_n$. Thus, $\| Ux \|_2 = \| x \|_2$.

Conversely, let $U$ be a matrix such that $\| Ux \|_2 = \| x \|_2$ for every $x \in \mathbb{C}^n$. This implies $x^H U^H U x = x^H x$, hence $x^H (U^H U - I_n) x = 0$ for $x \in \mathbb{C}^n$. This implies $U^H U = I_n$, so $U$ is a unitary matrix.
Corollary

The following statements that concern a matrix $U \in \mathbb{C}^{n \times n}$ are equivalent:

- $U$ is unitary;
- $\| Ux - Uy \|_2 = \| x - y \|_2$ for $x, y \in \mathbb{C}^n$;
- $(Ux, Uy) = (x, y)$ for $x, y \in \mathbb{C}^n$. 
The counterpart of unitary matrices in the set of real matrices are introduced next.

**Definition**

A matrix $A \in \mathbb{R}^{n \times n}$ is **orthogonal** or **orthonormal** if it is unitary.

In other words, a real matrix $A \in \mathbb{R}^{n \times n}$ is orthogonal if and only if $A'A = AA' = I_n$. Clearly, $A$ is orthogonal if and only if $A'$ is orthogonal.
Theorem

If $A \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then $\det(A) \in \{-1, 1\}$.

Proof.

By a previous Corollary, $|\det(A)| = 1$. Since $\det(A)$ is a real number, it follows that $\det(A) \in \{-1, 1\}$.
Corollary

Let $A$ be a matrix in $\mathbb{R}^{n \times n}$. The following statements are equivalent:

- $A$ is orthogonal;
- $A$ is invertible and $A^{-1} = A'$;
- $A'$ is invertible and $(A')^{-1} = A$;
- $A'$ is orthogonal.

Thus, a matrix $A$ is orthogonal if and only if it preserves the length of vectors.
Definition

A **rotation matrix** is an orthogonal matrix $R \in \mathbb{R}^{n \times n}$ such that $\det(R) = 1$. A **reflection matrix** is an orthogonal matrix $R \in \mathbb{R}^{n \times n}$ such that $\det(R) = -1$. 
Example

In the 2-dimensional case, \( n = 2 \), a rotation is a matrix \( R \in \mathbb{R}^{2 \times 2} \),

\[
R = \begin{pmatrix}
  r_{11} & r_{12} \\
  r_{21} & r_{22}
\end{pmatrix}
\]

such that

\[
r_{11}^2 + r_{21}^2 = 1,
\]
\[
r_{12}^2 + r_{22}^2 = 1,
\]
\[
r_{11}r_{12} + r_{21}r_{22} = 0
\]

and

\[
r_{11}r_{22} - r_{12}r_{21} = 1.
\]

This implies

\[
r_{22}(r_{11}r_{12} + r_{21}r_{22}) - r_{12}(r_{11}r_{22} - r_{12}r_{21}) = -r_{12},
\]
or
Since $r_{11}^2 \leq 1$ it follows that there exists $\theta$ such that $r_{11} = \cos \theta$. This implies that $R$ has the form

$$R = \begin{pmatrix} \cos \theta & - \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Its effect on a vector $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ is to produce the vector $y = Rx$, where

$$y = \begin{pmatrix} x_1 \cos \theta - x_2 \sin \theta \\ x_1 \sin \theta + x_2 \cos \theta \end{pmatrix},$$

which is obtained from $x$ by a counterclockwise rotation by the angle $\theta$. 

---

Prof. Dan A. Simovici  
CS724: Topics in Algorithms  
Norms and Inner Products - II  
Slide Set 5
It is easy to see that \( \det(R) = 1 \), so the term “rotation matrix” is clearly justified for \( R \). To mark the dependency of \( R \) on \( \theta \) we will use the notation

\[
R(\theta) = \begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix}.
\]
If the angle of the vector \( \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \) with the \( x_1 \) axis is \( \alpha \) and \( \mathbf{x} \) is rotated counterclockwise by \( \theta \) to yield the vector \( \mathbf{y} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2 \), then \( x_1 = r \cos \alpha, \ x_2 = r \sin \alpha, \) and

\[
\begin{align*}
y_1 & = r \cos(\alpha + \theta) = r \cos \alpha \cos \theta - r \sin \alpha \sin \theta = x_1 \cos \theta - x_2 \sin \theta, \\
y_2 & = r \sin(\alpha + \theta) = r \sin \alpha \cos \theta + r \cos \alpha \sin \theta = x_1 \sin \theta + x_2 \cos \theta,
\end{align*}
\]

which are the formulas that describe the transformation of \( \mathbf{x} \) into \( \mathbf{y} \).
Definition

Let $U$ be an $m$-dimensional subspace of $\mathbb{C}^n$ and let $\{u_1, \ldots, u_m\}$ be an orthonormal basis of this subspace. The orthogonal projection of the vector $x \in \mathbb{C}^n$ on the subspace $U$ is the vector $u = (x, u_1)u_1 + \cdots + (x, u_m)u_m$. 
Theorem

Let $U$ be an $m$-dimensional subspace of $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$. The vector $y = x - \text{proj}_U(x)$ belongs to the subspace $U^\perp$.

Proof.

Let $B_U = \{u_1, \ldots, u_m\}$ be an orthonormal basis of $U$. Note that

$$
(y, u_j) = (x, u_j) - \left( \sum_{i=1}^{m} (x, u_i)u_i, u_j \right)
$$

$$
= (x, u_j) - \sum_{i=1}^{m} (x, u_i)(u_i, u_j) = 0,
$$

due to the orthogonality of the basis $B_U$. Therefore, $y$ is orthogonal on every linear combination of $B_U$, that is on the subspace $U$. $\square$
Theorem

Let $U$ be an $m$-dimensional subspace of $\mathbb{C}^n$ having the orthonormal basis
{$u_1, \ldots, u_m$}. The orthogonal projection $\text{proj}_U$ is given by
$\text{proj}_U(x) = B_U B_U^H x$ for $x \in \mathbb{C}^n$, where $B_U \in \mathbb{R}^{n \times m}$ is the matrix
$B_U = (u_1 \cdots u_m) \in \mathbb{C}^{n \times m}$.

Proof.

We can write

$$\text{proj}_U(x) = \sum_{i=1}^{m} u_i (u_i^H x) = (u_1 \cdots u_m) \begin{pmatrix} u_1^H \\ \vdots \\ u_m^H \end{pmatrix} x = B_U B_U^H x.$$
Since the basis \( \{ u_1, \ldots, u_m \} \) is orthonormal, we have \( B_U^H B_U = I_m \). Observe that the matrix \( B_U B_U^H \in \mathbb{C}^{n \times n} \) is symmetric and idempotent because

\[
(B_U B_U^H)(B_U B_U^H) = B_U (B_U^H B_U) B_U^H = B_U B_U^H.
\]

For an \( m \)-dimensional subspace \( U \) of \( \mathbb{C}^n \) we denote by \( P_U = B_U B_U^H \in \mathbb{C}^{n \times n} \), where \( B_U \) is a matrix of an orthonormal basis of \( U \) as defined before. \( P_U \) is the projection matrix of the subspace \( U \).
Corollary

For every non-zero subspace $U$, the matrix $P_U$ is a Hermitian matrix, and therefore, a self-adjoint matrix.

Proof.

Since $P_U = B_U B_U^H$ where $B_U$ is a matrix of an orthonormal basis of the subspace $S$, it is immediate that $P_U^H = P_U$.

The self-adjointness of $P_U$ means that $(x, P_U y) = (P_U x, y)$ for every $x, y \in \mathbb{C}^n$. 
Corollary

Let $U$ be an $m$-dimensional subspace of $\mathbb{C}^n$ having the orthonormal basis $\{u_1, \ldots, u_m\}$. If $B_U = (u_1 \cdots u_m) \in \mathbb{C}^{n \times m}$, then for every $x \in \mathbb{C}$ we have the decomposition $x = P_U x + Q_U x$, where $P_U = B_U B_U^H$ and $Q_U = I_n - P_U$, $P_U x \in U$ and $Q_U x \in U^\perp$. 
Observe that

\[ Q_U^2 = (I_n - P_UP_H)(I_n - P_UP_H) = I_n - P_UP_H - P_UP_H + P_UP_HP_UP_HP_UP_H = Q_U, \]

so \( Q_U \) is an idempotent matrix. The matrix \( Q_U \) is the projection matrix on the subspace \( U \perp \). Clearly, we have

\[ P_U = Q_U = I_n - P_U. \]  \hspace{1cm} (5)

It is possible to give a direct argument for the independence of the projection matrix \( P_U \) relative to the choice of orthonormal basis in \( U \).
It is possible to give a direct argument for the independence of the projection matrix \( P_U \) relative to the choice of orthonormal basis in \( U \).

**Theorem**

Let \( U \) be an \( m \)-dimensional subspace of \( \mathbb{C}^n \) having the orthonormal bases \( \{ u_1, \ldots, u_m \} \) and \( \{ v_1, \ldots, v_m \} \) and let \( B_U = (u_1 \cdots u_m) \in \mathbb{C}^{n \times m} \) and \( \tilde{B}_U = (v_1 \cdots v_m) \in \mathbb{C}^{n \times m} \). The matrix \( B_U^H \tilde{B}_U \in \mathbb{C}^{m \times m} \) is unitary and \( \tilde{B}_U B_U^H = B_U B_U^H \).
Proof

Since the both sets of columns of $B_U$ and $\tilde{B}_U$ are bases for $U$, there exists a unique square matrix $Q \in \mathbb{C}^{m \times m}$ such that $B_U = \tilde{B}_U Q$. The orthonormality of $B_U$ and $\tilde{B}_U$ implies $B_U^H B_U = \tilde{B}_U^H \tilde{B}_U = I_m$. Thus, we can write

$$I_m = B_U^H B_U = Q^H \tilde{B}_U^H \tilde{B}_U Q = Q^H Q,$$

which shows that $Q$ is unitary. Furthermore, $B_U^H \tilde{B}_U = Q^H \tilde{B}_U^H \tilde{B}_U = Q^H$ is unitary and

$$B_U B_U^H = \tilde{B}_U QQ^H \tilde{B}_U^H = \tilde{B}_U \tilde{B}_U^H.$$
Definition

A matrix $A \in \mathbb{C}^{n \times n}$ is positive definite if $x^H A x$ is a real positive number for every $x \in \mathbb{C}^n - \{0\}$. 
Theorem

If $A \in \mathbb{C}^{n \times n}$ is positive definite, then $A$ is Hermitian.

Proof.

Let $A \in \mathbb{C}^{n \times n}$ be a matrix. Since $x^H A x$ is a real number it follows that it equals its conjugate, so $x^H A x = x^H A^H x$ for every $x \in \mathbb{C}^n$. Therefore, there exists a unique pair of Hermitian matrices $H_1$ and $H_2$ such that $A = H_1 + iH_2$, which implies $A^H = H_1^H - iH_2^H$. Thus, we have

$$x^H (H_1 + iH_2) x = x^H (H_1^H - iH_2^H) x = x^H (H_1 - iH_2) x,$$

because $H_1$ and $H_2$ are Hermitian. This implies $x^H H_2 x = 0$ for every $x \in \mathbb{C}^n$, which, in turn, implies $H_2 = O_{n,n}$. Consequently, $A = H_1$, so $A$ is indeed Hermitian.
A matrix \( A \in \mathbb{C}^{n \times n} \) is positive semidefinite if \( x^H A x \) is a non-negative real number for every \( x \in \mathbb{C}^n - \{0\} \).

Positive definiteness (positive semidefiniteness) is denoted by \( A \succ 0 \) (\( A \succeq 0 \), respectively).
The definition of positive definite (semidefinite) matrix can be specialized for real matrices as follows.

**Definition**

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is **positive definite** if $\mathbf{x}'A\mathbf{x} > 0$ for every $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$.

If $A$ satisfies the weaker inequality $\mathbf{x}'A\mathbf{x} \geq 0$ for every $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$, then we say that $A$ is **positive semidefinite**.

$A \succ 0$ denotes that $A$ is positive definite and $A \succeq 0$ means that $A$ is positive semidefinite.
Note that in the case of real-valued matrices we need to require explicitly the symmetry of the matrix because, unlike the complex case, the inequality $\mathbf{x}'A\mathbf{x} > 0$ for $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}_n\}$ does not imply the symmetry of $A$. For example, consider the matrix

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

where $a, b \in \mathbb{R}$ and $a > 0$. We have

$$\mathbf{x}'A\mathbf{x} = (x_1 \ x_2) \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a(x_1^2 + x_2^2) > 0$$

if $\mathbf{x} \neq \mathbf{0}_2$. 
Example

The symmetric real matrix

\[ A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \]

is positive definite if and only if \( a > 0 \) and \( b^2 - ac < 0 \). Indeed, we have \( x'Ax > 0 \) for every \( x \in \mathbb{R}^2 - \{0\} \) if and only if \( ax_1^2 + 2bx_1x_2 + cx_2^2 > 0 \), where \( x' = (x_1 \ x_2) \); elementary algebra considerations lead to \( a > 0 \) and \( b^2 - ac < 0 \).
A positive definite matrix is non-singular. Indeed, if \( A\mathbf{x} = \mathbf{0} \), where \( A \in \mathbb{R}^{n \times n} \) is positive definite, then \( \mathbf{x}^H A \mathbf{x} = 0 \), so \( \mathbf{x} = \mathbf{0} \). Therefore, \( A \) is non-singular.

**Example**

If \( A \in \mathbb{C}^{m \times n} \), then the matrices \( A^H A \in \mathbb{C}^{n \times n} \) and \( AA^H \in \mathbb{C}^{m \times m} \) are positive semidefinite. For \( \mathbf{x} \in \mathbb{C}^n \) we have

\[
\mathbf{x}^H (A^H A) \mathbf{x} = (\mathbf{x}^H A^H) (A \mathbf{x}) = (A \mathbf{x})^H (A \mathbf{x}) = \| A \mathbf{x} \|_2^2 \geq 0.
\]

The argument for \( AA^H \) is similar.

If \( \text{rank}(A) = n \), then the matrix \( A^H A \) is positive definite because \( \mathbf{x}^H (A^H A) \mathbf{x} = 0 \) implies \( A \mathbf{x} = \mathbf{0} \), which, in turn, implies \( \mathbf{x} = \mathbf{0} \).
Theorem

If \( A \in \mathbb{C}^{n \times n} \) is a positive definite matrix, then any principal submatrix \( B = A \begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix} \) is a positive definite matrix.

Proof.

Let \( x \in \mathbb{C}^n - \{0\} \) be a vector such that all components located on positions other than \( i_1, \ldots, i_k \) equal 0 and let \( y = x \begin{bmatrix} i_1 & \cdots & i_k \\ 1 \end{bmatrix} \in \mathbb{C}^k \) be the vector obtained from \( x \) by retaining only the components located on positions \( i_1, \ldots, i_k \). Since \( y^H B y = x^H A x > 0 \) it follows that \( B \succ 0 \). \( \square \)
Corollary

If $A \in \mathbb{C}^{n \times n}$ is a positive definite matrix, then any diagonal element $a_{ii}$ is a real positive number for $1 \leq i \leq n$. 
**Theorem**

If $A, B \in \mathbb{C}^{n \times n}$ are two positive semidefinite matrices and $a, b$ are two non-negative numbers, then $aA + bB \succeq 0$.

**Proof.**

The statement holds because $\mathbf{x}^\mathsf{H}(aA + bB)\mathbf{x} = a\mathbf{x}^\mathsf{H}A\mathbf{x} + b\mathbf{x}^\mathsf{H}B\mathbf{x} \succeq 0$, due to the fact that $A$ and $B$ are positive semidefinite.
Definition

Let $L = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$ be a sequence of vectors in $\mathbb{R}^n$. The **Gram matrix of** $L$ is the matrix $G_L = (g_{ij}) \in \mathbb{R}^{m \times m}$ defined by $g_{ij} = \mathbf{v}_i' \mathbf{v}_j$ for $1 \leq i, j \leq m$.

Note that if $A_L = (\mathbf{v}_1 \cdots \mathbf{v}_m) \in \mathbb{R}^{n \times m}$, then $G_L = A_L' A_L$. Also, note that $G_L$ is a symmetric matrix.
Example

Let

\[ \mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}. \]

The Gram matrix of the set \( L = \{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \} \) is

\[ G_L = \begin{pmatrix} 2 & -1 & 2 \\ -1 & 9 & 4 \\ 2 & 4 & 5 \end{pmatrix}. \]

Note that \( \det(G_L) = 1 \).
**Theorem**

Let \( L = (\mathbf{v}_1, \ldots, \mathbf{v}_m) \) be a sequence of \( m \) vectors in \( \mathbb{R}^n \), where \( m \leq n \). If \( L \) is linearly independent, then the Gram matrix \( G_L \) is positive definite.

**Proof.**

Suppose that \( L \) is linearly independent. Let \( \mathbf{x} \in \mathbb{R}^m \). We have
\[
\mathbf{x}' G_L \mathbf{x} = \mathbf{x}' A_L' A_L \mathbf{x} = (A_L \mathbf{x})' A_L \mathbf{x} = \| A_L \mathbf{x} \|_2^2.
\]
Therefore, if \( \mathbf{x}' G_L \mathbf{x} = 0 \), we have \( A_L \mathbf{x} = \mathbf{0} \), which is equivalent to \( x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n = 0 \). Since \( \{\mathbf{v}_1, \ldots, \mathbf{v}_m\} \) is linearly independent it follows that \( x_1 = \cdots = x_m = 0 \), so \( \mathbf{x} = \mathbf{0} \). Thus, \( A \) is indeed, positive definite.
The Gram matrix of an arbitrary sequence of vectors is positive semidefinite, as the reader can easily verify.

**Definition**

Let \( L = (\mathbf{v}_1, \ldots, \mathbf{v}_m) \) be a sequence of \( m \) vectors in \( \mathbb{R}^n \), where \( m \leq n \). The **Gramian** of \( L \) is the number \( \det(G_L) \).
Theorem

If \( L = (\mathbf{v}_1, \ldots, \mathbf{v}_m) \) is a sequence of \( m \) vectors in \( \mathbb{R}^n \). Then, \( L \) is linearly independent if and only if \( \det(G_L) \neq 0 \).

Proof.

Suppose that \( \det(G_L) \neq 0 \) and that \( L \) is not linearly independent. In other words, the numbers \( a_1, \ldots, a_m \) exists such that at least one of them is not 0 and \( a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m = \mathbf{0} \). This implies the equalities

\[
a_1(\mathbf{x}_1, \mathbf{x}_j) + \cdots + a_m(\mathbf{x}_m, \mathbf{x}_j) = \mathbf{0},
\]

for \( 1 \leq j \leq m \), so the system \( G_La = \mathbf{0} \) has a non-trivial solution in \( a_1, \ldots, a_m \). This implies \( \det(G_L) = 0 \), which contradicts the initial assumption.
Conversely, suppose that $L$ is linearly independent and $\det(G_L) = 0$. Then, the linear system

$$a_1(x_1, x_j) + \cdots + a_m(x_m, x_j) = 0,$$

for $1 \leq j \leq m$, has a non-trivial solution in $a_1, \ldots, a_m$. If $w = a_1x_1 + \cdots a_mx_m$, this amounts to $(w,x_i) = 0$ for $1 \leq i \leq n$. This, in turn, implies $(w,w) = \|w\|_2^2 = 0$, so $w = 0$, which contradicts the linear independence of $L$. 
The Gram-Schmidt algorithm constructs an orthonormal basis for a subspace $U$ of $\mathbb{C}^n$, starting from an arbitrary basis of $\{u_1, \ldots, u_m\}$ of $U$. The orthonormal basis is constructed sequentially such that $\langle w_1, \ldots, w_k \rangle = \langle u_1, \ldots, u_k \rangle$ for $1 \leq k \leq m$. 
Gram-Schmidt Orthogonalization Algorithm

**Data:** A basis \( \{u_1, \ldots, u_m\} \) for a subspace \( U \) of \( \mathbb{C}^n \)

**Result:** An orthonormal basis \( \{w_1, \ldots, w_m\} \) for \( U \)

\[
W = O_{n,m} \\
W(:, 1) = W(:, 1) + \frac{1}{\|U(:, 1)\|_2} U(:, 1) \\
\text{For } (k = 2 \text{ to } m) \{ \\
\quad P = I_n - W(:, 1 : (k - 1)) W(:, 1 : (k - 1))^H \\
\quad W(:, k) = W(:, k) + \frac{1}{\|P U(:, k)\|_2} P U(:, k) \\
\} \text{ return } W = (w_1 \cdots w_m)
\]
Theorem

Let \((\mathbf{w}_1, \ldots, \mathbf{w}_m)\) be the sequence of vectors constructed by the Gram-Schmidt algorithm starting from the basis \(\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}\) of an \(m\)-dimensional subspace \(U\) of \(\mathbb{C}^n\). The set \(\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}\) is an orthogonal basis of \(U\) and \(\langle \mathbf{w}_1, \ldots, \mathbf{w}_k \rangle = \langle \mathbf{u}_1, \ldots, \mathbf{u}_k \rangle\) for \(1 \leq k \leq m\).
Proof

In the algorithm the matrix $W$ is initialized as $O_{n,m}$. Its columns will contain eventually the vectors of the orthonormal basis $w_1, \ldots, w_m$. The argument is by induction on $k \geq 1$. The base case, $k = 1$, is immediate. Suppose that the statement of the theorem holds for $k$, that is, the set \{ $w_1, \ldots, w_k$ \} is an orthonormal basis for $U_k = \langle u_1, \ldots, u_k \rangle$ and constitutes the set of the initial $k$ columns of the matrix $W$, that is, $W_k = W(:, 1 : k)$. Then,

$$P_k = I_n - W_k W_k^H$$

is the projection matrix on the subspace $U_k^\perp$, so $P_k u_k$ is orthogonal on every $w_i$, where $1 \leq i \leq k$. Therefore, $w_{k+1} = W(:, (k + 1))$ is a unit vector orthogonal on all its predecessors $w_1, \ldots, w_k$, so \{ $w_1, \ldots, w_m$ \} is an orthonormal set.
The equality \( \langle u_1, \ldots, u_k \rangle = \langle w_1, \ldots, w_k \rangle \) clearly holds for \( k = 1 \). Suppose that it holds for \( k \). Then, we have

\[
    w_{k+1} = \frac{1}{\| P_k u_{k+1} \|_2} (u_{k+1} - W_k W_k^H u_{k+1})
\]

\[
    = \frac{1}{\| P_k u_{k+1} \|_2} (u_{k+1} - (w_1 \cdots w_k) W_k^H u_{k+1}).
\]

Since \( w_1, \ldots, w_k \) belong to the subspace \( \langle u_1, \ldots, u_k \rangle \) (by inductive hypothesis), it follows that \( w_{k+1} \in \langle u_1, \ldots, u_k, u_{k+1} \rangle \), so

\[
\langle w_1, \ldots, w_{k+1} \rangle \subseteq \langle u_1, \ldots, u_k \rangle.
\]
For the converse inclusion, since

$$u_{k+1} = \| P_k u_{k+1} \|_2 w_{k+1} + (w_1 \cdots w_k) W_k^H u_{k+1},$$

it follows that $u_{k+1} \in \langle \langle \langle w_1, \ldots, w_k, w_{k+1} \rangle \rangle \rangle$. Thus,

$$\langle u_1, \ldots, u_k, u_{k+1} \rangle \subseteq \langle w_1, \ldots, w_k, w_{k+1} \rangle.$$
Example

Let $A \in \mathbb{R}^{3 \times 2}$ be the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 3 \end{pmatrix}.$$ 

It is easy to see that $\text{rank}(A) = 2$. We have $\{u_1, u_2\} \subseteq \mathbb{R}^3$ and we construct an orthogonal basis for the subspace generated by these columns. The matrix $W$ is initialized to $O_{3,2}$. 
Example cont’d

we begin by defining

\[ \mathbf{w}_1 = \frac{1}{\| \mathbf{u}_1 \|_2} \mathbf{u}_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \]

so

\[ \mathbf{W} = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 \\ \frac{\sqrt{2}}{2} & 0 \end{pmatrix}, \]

The projection matrix is

\[ P = I_3 - \mathbf{W}(\cdot, 1)\mathbf{W}(\cdot, 1)' = I_3 - \mathbf{w}_1 \mathbf{w}_1' = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}. \]
The projection of $u_2$ is

$$P u_2 = \begin{pmatrix} -1 \\
0 \\
1 \end{pmatrix}$$

and the second column of $W$ becomes

$$w_k = W(:, 2) = \frac{\|P u_2\|_2}{P} u_2 = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\
0 \\
\frac{\sqrt{2}}{2} \end{pmatrix}.$$
Thus, the orthonormal basis we are seeking consists of the vectors

\[
\begin{pmatrix}
\frac{\sqrt{2}}{2} \\
0 \\
\frac{\sqrt{2}}{2}
\end{pmatrix}
\quad \text{and} \quad 
\begin{pmatrix}
-\frac{\sqrt{2}}{2} \\
0 \\
\frac{\sqrt{2}}{2}
\end{pmatrix}.
\]
We describe a factorization algorithm for rectangular matrices which allows us to express a matrix as a product of a rectangular matrix with orthogonal columns and un upper triangular invertible matrix (the thin QR factorization).
(The Thin QR Factorization Theorem) Let \( A \in \mathbb{C}^{m \times n} \) be a full-rank matrix such that \( m \geq n \). Then, \( A \) can be factored as \( A = QR \), where \( Q \in \mathbb{C}^{m \times n} \), \( R \in \mathbb{C}^{n \times n} \) such that

- the columns of \( Q \) constitute an orthonormal basis for \( \text{range}(A) \), and
- \( R = (r_{ij}) \) is an upper triangular invertible matrix such that its diagonal elements are real non-negative numbers, that is, \( r_{ii} \geq 0 \) for \( 1 \leq i \leq n \).
Let \( u_1, \ldots, u_n \) be the columns of \( A \). Since \( \text{rank}(A) = n \), these columns constitute a basis for \( \text{range}(A) \). Starting from this set of columns construct an orthonormal basis \( w_1, \ldots, w_n \) for the subspace \( \text{range}(A) \) using the Gram-Schmidt algorithm. Define \( Q \) as the orthogonal matrix

\[
Q = (w_1 \cdots w_n).
\]

By the properties of the Gram-Schmidt algorithm we have 
\[
\langle u_1, \ldots, u_k \rangle = \langle w_1, \ldots, w_k \rangle
\]
for \( 1 \leq k \leq n \), so it is possible to write

\[
u_k = r_{1k} w_1 + \cdots + r_{kk} w_k
\]

\[
= (w_1 \cdots w_n) \begin{pmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = Q \begin{pmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\]

We can assume that \( r_{kk} \geq 0 \); otherwise, that is, if \( r_{kk} < 0 \), replace \( w_k \) by \(-w_k\). Clearly, this does not affect the orthonormality of the set \( \{w_1, \ldots, w_n\} \).
Example

Let us determine a QR factorization for the matrix

\[
A = \begin{pmatrix}
1 & 1 \\
0 & 0 \\
1 & 3 \\
\end{pmatrix}
\]

which has rank 2. We constructed an orthonormal basis for \(\text{range}(A)\) that consists of the vectors

\[
\begin{align*}
\mathbf{w}_1 &= \begin{pmatrix}
\frac{1}{\sqrt{2}} \\
0 \\
\frac{1}{\sqrt{2}} \\
\end{pmatrix} \\
\mathbf{w}_2 &= \begin{pmatrix}
-\frac{1}{\sqrt{2}} \\
0 \\
\frac{1}{\sqrt{2}} \\
\end{pmatrix}
\end{align*}
\]
Example cont’d

Thus, the orthogonal matrix $Q$ is

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$ 

To compute $R$ we need to express $u_1$ and $u_2$ as linear combinations of $w_1$ and $w_2$. Since

$$u_1 = \sqrt{2}w_1$$
$$u_2 = 2\sqrt{2}w_1 + \sqrt{2}w_2$$

the matrix $R$ is

$$R = \begin{pmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & \sqrt{2} \end{pmatrix}. $$
Vector norms can be computed using the function \texttt{norm} which comes in two signatures: \texttt{norm(v)} and \texttt{norm(v,p)}. The first variant computes \( \| v \|_2 \); the second computes \( \| v \|_p \) for any \( p, 1 \leq p \leq \infty \). In addition, \texttt{norm(v,\text{inf})} computes \( \| v \|_\infty = \max\{|v_i| \mid 1 \leq i \leq n\} \), where \( v \in \mathbb{R}^n \). If one uses \(-\infty\) as the second parameter, then \texttt{norm(v,-\text{inf})} returns \( \min\{|v_i| \mid 1 \leq i \leq n\} \).

**Example**

For the vector

\[ v = [2 \  -3 \  5 \  -4] \]

the computation

\[ \text{norms} = [\text{norm}(v,1),\text{norm}(v,2),\text{norm}(v,2.5),\text{norm}(v,\text{inf}),\text{norm}(v,-\text{inf})] \]

returns

\[ \text{norms} = [14.0000 \ 7.3485 \ 6.5344 \ 5.0000 \ 2.0000] \]
For matrices whose norm is expensive to compute, an approximative estimation of $\| A \|_2$ can be performed using the function `normest(A)`, or `normest(A,r)`, where $r$ is the relative error; the default for $r$ is $10^{-6}$. The following function implements the Gram-Schmidt algorithm.

```matlab
function [W] = gram(U)
%GRAM implements the classical Gram-Schmidt algorithm
[n,m] = size(U);
W = zeros(n,m);
W(:,1)= (1/norm(U(:,1)))*U(:,1);
for k = 2:1:m
    P = eye(n) - W*W';
    W(:,k) = W(:,k) + (1/norm(P*U(:,k)))* P*U(:,k);
end
end
```
The Cholesky decomposition of a Hermitian positive definite matrix is computed in MATLAB using the function \texttt{chol}. The function call \texttt{R = chol(A)} returns an upper triangular matrix \( R \), satisfying the equation \( R^H R = A \). If \( A \) is not positive definite an error message is generated. The matrix \( R \) is computed using the diagonal and the upper triangle of \( A \) and the computation makes sense only if \( A \) is Hermitian.
Example

Let $A$ be the symmetric positive definite matrix

$$A = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$ 

Then, $R = \text{chol}(A)$ yields

$$R =
\begin{pmatrix}
1.7321 & 0 & 1.1547 \\
0 & 1.4142 & 0.7071 \\
0 & 0 & 0.4082
\end{pmatrix}.$$
The call \( L = \text{chol}(A, 'lower') \) returns a lower triangular matrix \( L \) from the diagonal and lower triangle of matrix \( A \), satisfying the equation \( LL^H = A \). When \( A \) is sparse, this syntax of \text{chol} is faster.

**Example**

For the same matrix \( A \) \( L = \text{chol}(A, 'lower') \) returns

\[
L =
\begin{bmatrix}
1.7321 & 0 & 0 \\
0 & 1.4142 & 0 \\
1.1547 & 0.7071 & 0.4082
\end{bmatrix}
\]

For added flexibility, \([R,p] = \text{chol}(A)\) and \([L,p] = \text{chol}(A,'lower')\) set \( p \) to 0 if \( A \) is positive definite and to a positive number, otherwise, without returning an error message.
The thin QR decomposition of a matrix $A \in \mathbb{C}^{m \times n}$ is obtained using the function $qr$ as in

$$[Q \ R] = qr(A)$$

To obtain the full decomposition we write

$$[Q \ R] = qr(A,0)$$