Numerical Optimization
Mathematical Background (I)

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization
Sets

Definition
A set is a collection of objects satisfying certain property $P$.

Examples:
- A set of natural numbers, \{1, 2, 3, \ldots\}
- $\{x \in \mathbb{R} : 1 \leq x \leq 3\}$

Note: A set not containing any object is called the empty set and is denoted by $\emptyset$.

Let $A$ and $B$ be two sets.

- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- Difference: $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$
Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

If the intersection of two sets is empty, we say that the sets are \textit{disjoint}. That is, for two disjoint sets $A$ and $B$, $A \cap B = \emptyset$. 
Difference: $A \setminus B = \{x : x \in A \text{ and } x \not\in B\}$
Let $A$ and $B$ be two sets. If $A$ is a subset of $B$, that is, every member of $A$ is also a member of $B$, we write $A \subseteq B$. Further, if $A$ is a subset of $B$ and there exists $y \in B$ such that $y \notin A$, then we write $A \subset B$. 
Supremum and Infimum of a set

**Definition**

A set $A$ of real numbers is said to be *bounded above*, if there is a real number $y$ such that $x \leq y$ for every $x \in A$. The smallest possible real number $y$ satisfying $x \leq y$ for every $x \in A$ is called the *least upper bound* or *supremum* of $A$ and is denoted by $\sup\{x : x \in A\}$.

Similarly, one can define *greatest lower bound* or *infimum*, $\inf\{x : x \in A\}$.

**Example:** Consider the set, $A = \{x : 1 \leq x < 3\}$

- $\sup\{x : x \in A\} = 3 (\notin A)$
- $\inf\{x : x \in A\} = 1 (\in A)$
Vector Space
A nonempty set $S$ is called a *vector space* if

1. For any $x, y \in S$, $x + y$ is defined and is in $S$. Further,
   
   $x + y = y + x$ (commutativity)

   $x + (y + z) = (x + y) + z$ (associativity)

2. There exists an element in $S$, $0$, such that $x + 0 = 0 + x = x$ for all $x$.

3. For any $x \in S$, there exists $y \in S$ such that $x + y = 0$.

4. For any $x \in S$ and $\alpha \in \mathbb{R}$, $\alpha x$ is defined and is in $S$. Further, $1x = x$ for every $x$.

5. For any $x, y \in S$ and $\alpha, \beta \in \mathbb{R}$,

   $\alpha(x + y) = \alpha x + \alpha y$

   $(\alpha + \beta)x = \alpha x + \beta x$

   $\alpha(\beta x) = (\alpha \beta)x$

Elements in $S$ are called *vectors*
Notations

- \( \mathbb{R} \): Vector space of real numbers
- \( \mathbb{R}^n \): Vector space of real \( n \times 1 \) vectors
- \( n \)-vector \( \mathbf{x} \) is an array of \( n \) scalars, \( x_1, x_2, \ldots, x_n \)
  \[
  \mathbf{x} = \begin{pmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
  \end{pmatrix}
  \]
- \( \mathbf{x} \in \mathbb{R}^n, x_i \in \mathbb{R} \ \forall \ i \)
- \( \mathbf{x}^T = (x_1, x_2, \ldots, x_n) \)
- \( \mathbf{0}^T = (0, 0, \ldots, 0) \)
- \( \mathbf{1}^T = (1, 1, \ldots, 1) \) (We also use \( \mathbf{e} \) to denote this vector)
Mathematical Background

**Definition**
If $S$ and $T$ are vector spaces such that $S \subseteq T$, then $S$ is called a *subspace* of $T$.

*Question*: What are all possible subspaces of $\mathbb{R}^2$?
Mathematical Background

Spanning Set

**Definition**

A set of vectors \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \) is said to *span* the vector space \( S \) if any vector \( \mathbf{x} \in S \) can be represented as

\[
\mathbf{x} = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i
\]

for some real coefficients \( \alpha_i, \ i = 1, \ldots, k. \)
Example: The vectors, 
\[ a_1 = (1, 0)^T, \ a_2 = (1, 1)^T; \ a_3 = (0, 1)^T, \ a_4 = (-1, 0)^T \text{ and } \ a_5 = (1, -1)^T \] span \( \mathbb{R}^2 \).
Linear Independence

Definition

A set of vectors $x_1, x_2, \ldots, x_k$ is said to *linearly independent* if

$$
\sum_{i=1}^{k} \alpha_i x_i = 0 \Rightarrow \alpha_i = 0 \ \forall \ i.
$$

Otherwise, they are linearly dependent and one of them is a linear combination of the others.
Example: In $\mathbb{R}^2$, 

- $a_1 = (1, 0)$ and $a_2 = (1, 1)$ are linearly independent.
- $a_1 = (1, 0)$ and $a_4 = (-1, 0)$ are linearly dependent.
**Basis**

**Definition**

A set of vectors is said to be a *basis* for the vector space $S$ if it is linearly independent and spans $S$. 
**Example**: For $\mathbb{R}^2$,

- $a_1 = (1, 0)$ and $a_2 = (1, 1)$ form a basis
- $a_1 = (1, 0)$ and $a_3 = (0, 1)$ form a basis
Mathematical Background

- A vector space does not have a unique basis.
- If $x_1, x_2, \ldots, x_k$ is a basis for $S$, then any $x \in S$ can be uniquely represented using $x_1, x_2, \ldots, x_k$.
- Any two bases of a vector space have the same cardinality.
- The dimension of the vector space $S$ is the cardinality of a basis of $S$.
- The dimension of the vector space $\mathbb{R}^n$ is $n$.
- Let $e_i$ denote an $n$-dimensional vector whose $i$-th element is 1 and the remaining elements are 0’s. Then, the set $e_1, e_2, \ldots, e_n$ forms a standard basis for $\mathbb{R}^n$.
- A basis for the vector space $S$ is a maximal independent set of vectors which spans the space $S$.
- A basis for the vector space $S$ is a minimal spanning set of vectors which spans the space $S$. 
Functions

**Definition**
A function $f$ from a set $A$ to a set $B$ is a rule that assigns to each $x$ in $A$ a unique element $f(x)$ in $B$. This function can be represented by

$$f : A \rightarrow B.$$  

**Note:**
- $A$: Domain of $f$
- $\{y \in B : (\exists x)[y = f(x)]\}$: Range of $f$
- Range of $f \subseteq B$

**Examples:**
- $f : \mathbb{R} \rightarrow \mathbb{R}$ defined as $f(x) = x^2$
- $f : (-1, 1) \rightarrow \mathbb{R}$ defined as $f(x) = \frac{1}{|x|-1}$
Mathematical Background

Definition

A norm on $\mathbb{R}^n$ is a real-valued function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ which obeys

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

Shirish Shevade

Numerical Optimization
Mathematical Background

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

Some popular norms:

- $L_2$ or Euclidean norm

\[ \|x\|_2 = \left( \sum_{i=1}^{n} (x_i)^2 \right)^{\frac{1}{2}} \]

- $L_1$ norm

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

- $L_\infty$ norm

\[ \|x\|_\infty = \max_{i=1,...,n} |x_i| \]
Mathematical Background

Illustration of $L_2$ norm:

$$\|x\|_2 = \left( \sum_{i=1}^{n} (x_i)^2 \right)^{\frac{1}{2}}$$
\[ S = \{ \mathbf{x} \in \mathbb{R}^2 : \| \mathbf{x} \|_2 \leq r \} \]
Mathematical Background

\[ S = \{ x \in \mathbb{R}^2 : \| x \|_1 \leq 1 \} \]
$S = \{ \mathbf{x} \in \mathbb{R}^2 : \| \mathbf{x} \|_\infty \leq 1 \}$
Mathematical Background

- In general, the class of $L_p$ ($1 \leq p < \infty$) vector norms is defined as

$$
\|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}
$$

- **Question**: Does the convergence of a particular optimization algorithm depend on what norm its stopping criterion used?

**Result**

If $\| \cdot \|_p$ and $\| \cdot \|_q$ are any two norms on $\mathbb{R}^n$, then there exist positive constants $\alpha$ and $\beta$ such that

$$
\alpha \|x\|_p \leq \|x\|_q \leq \beta \|x\|_p
$$

for any $x \in \mathbb{R}^n$. 
**Inner Product**

**Definition**

Let \( x, y \in \mathbb{R}^n \) and \( x \neq 0 \neq y \). The *inner* or *dot* product of \( x \) and \( y \) is defined as

\[
x \cdot y \equiv x^T y = \sum_{i=1}^{n} x_i \cdot y_i = \|x\| \cdot \|y\| \cos \theta
\]

where \( \theta \) is the angle between \( x \) and \( y \).

**Note:**

- \( x^T x = \|x\|^2 \).
- \( x^T y = y^T x \)
- \( |x \cdot y| \leq \|x\| \cdot \|y\| \) \hspace{1cm} (*Cauchy-Schwartz inequality*)
Orthogonality

Suppose $\mathbf{x}$ and $\mathbf{y}$ are perpendicular to each other.

Using Pythagoras formula,

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{y}\|^2,$$

which gives $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\mathbf{x}^T\mathbf{y}$. That is, $\mathbf{x}^T\mathbf{y} = 0$
Orthogonality

**Definition**

Let \( \mathbf{x} \in \mathbb{R}^n \) and \( \mathbf{y} \in \mathbb{R}^n \). \( \mathbf{x} \) and \( \mathbf{y} \) are said to be perpendicular or orthogonal to each other if \( \mathbf{x}^T \mathbf{y} = 0 \).

**Definition**

Two subspaces \( S \) and \( T \) of the same vector space \( \mathbb{R}^n \) are orthogonal if every vector \( \mathbf{x} \in S \) is orthogonal to every vector \( \mathbf{y} \in T \), i.e. \( \mathbf{x}^T \mathbf{y} = 0 \ \forall \mathbf{x} \in S, \mathbf{y} \in T \).
Definition

Given a subspace $S$ of $\mathbb{R}^n$, the space of all vectors orthogonal to $S$ is called the *orthogonal complement* of $S$. 
Mutual Orthogonality

Definition

Vectors \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n \) are said to be \textit{mutually orthogonal} if \( \mathbf{x}_i \cdot \mathbf{x}_j = 0 \) for all \( i \neq j \). If, in addition, \( ||\mathbf{x}_i|| = 1 \) for every \( i \), the set \( \{ \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \} \) is said to be \textit{orthonormal}.
Mutual Orthogonality

Is the set of mutually orthogonal vectors linearly independent?
Result

If \( \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \) are mutually orthogonal nonzero vectors, then they are linearly independent.

We need to show that

\[
\sum_{i=1}^{k} \alpha_i \mathbf{x}_i = 0 \Rightarrow \alpha_i = 0 \quad \forall \ i.
\]

Proof.

Let \( \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \ldots + \alpha_k \mathbf{x}_k = 0. \)

Therefore, \( (\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \ldots + \alpha_k \mathbf{x}_k)^T \mathbf{x}_1 = 0, \) or,

\[
\sum_{i=1}^{k} \alpha_i \mathbf{x}_i^T \mathbf{x}_1 = 0.
\]

This gives \( \alpha_1 \mathbf{x}_1^T \mathbf{x}_1 = 0 \) which implies \( \alpha_1 = 0. \)

Similarly we can show that each \( \alpha_i \) is zero.

Therefore, the mutually orthogonal vectors are linearly independent.
Suppose $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthonormal.
Given any vector $\mathbf{x}$, we can write $\mathbf{x} = (\mathbf{x}^T \mathbf{x}_1) \mathbf{x}_1 + (\mathbf{x}^T \mathbf{x}_2) \mathbf{x}_2$.
We require orthonormality of given set of vectors.
**Mathematical Background**

*Question*: How to produce an orthonormal basis starting with a given basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$?
Mathematical Background

**Gram-Schmidt Procedure**

- Given $x_1, x_2, x_3$, a basis in $\mathbb{R}^3$
- To produce an orthonormal basis $y_1, y_2, y_3$.
- Without loss of generality, set $y_1 = \frac{x_1}{\|x_1\|}$
- Consider $x_2$ and remove its component in the $y_1$ direction.

$$z_2 = x_2 - (x_2^T y_1) y_1$$

- $z_2$ is orthogonal to $y_1$
- Set $y_2 = \frac{z_2}{\|z_2\|}$
- Start with $x_3$ and remove its components in the $y_1$ and $y_2$ directions.

$$z_3 = x_3 - (x_3^T y_1) y_1 - (x_3^T y_2) y_2$$

- $z_3$ is orthogonal to $y_1$ and $y_2$
- Set $y_3 = \frac{z_3}{\|z_3\|}$
- Easy to extend this procedure to a basis in $\mathbb{R}^n$
Mathematical Background

Matrices

- $A \in \mathbb{R}^{m \times n}$. $A$ is a matrix of size $m \times n$.

$$A = \begin{pmatrix}
A_{11} & A_{12} & \ldots & A_{1n} \\
A_{21} & A_{22} & \ldots & A_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
A_{m1} & A_{m2} & \ldots & A_{mn}
\end{pmatrix}$$

- $A_{ij}$ denotes $(i,j)$-element of $A$.

- $A = (a_1, a_2, \ldots, a_n)$ where $a_i \in \mathbb{R}^m$, $i = 1, \ldots, n$

- The transpose of $A$, denoted by $A^T$ is the $n \times m$ matrix whose $(i,j)$-element is $A_{ji}$.

$$A^T = \begin{pmatrix}
a_1^T \\
a_2^T \\
\vdots \\
a_n^T
\end{pmatrix}$$
Mathematical Background

Matrices

- Diagonal Matrix: A square matrix $\Lambda$ such that $\Lambda_{ij} = 0$, $i \neq j$

\[
\Lambda = \begin{pmatrix}
\lambda_1 & 0 & \ldots & 0 \\
0 & \lambda_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda_n \\
\end{pmatrix}
\]

- Identity Matrix ($I$): A diagonal matrix such that $I_{ii} = 1 \ \forall \ i$

\[
I = \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1 \\
\end{pmatrix}
\]

- Lower Triangular Matrix ($L$): A square matrix such that $L_{ij} = 0$, $i < j$
Mathematical Background

Matrices

- Let $A \in \mathbb{R}^{m \times n}$

**Definition**

The subspace of $\mathbb{R}^m$, spanned by the column vectors of $A$ is called the *column space* of $A$. The subspace of $\mathbb{R}^n$, spanned by the row vectors of $A$ is called the *row space* of $A$.

**Definition**

*Column Rank*: The dimension of the column space
*Row Rank*: The dimension of the row space

**Definition**

The column rank of a matrix $A$ equals its row rank, and this common value is called the *rank* of $A$. 

Shirish Shevade | Numerical Optimization
Let $A = \begin{pmatrix} 1 & 3 & -2 & 4 \\ -1 & -3 & 1 & -2 \end{pmatrix}$. $\text{rank}(A) = 2$

The rank of a matrix is 0 if and only if it is a zero matrix.

Matrices with the smallest rank - Rank one matrices

*Example*:

\[
\begin{pmatrix} 3 & 1 & -1 \\ -3 & -1 & 1 \\ 6 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} (3 \ 1 \ -1) = uv^T
\]

Every matrix of rank one has the simplest form, $A = uv^T$. 
Mathematical Background

Matrices

Definition

A square matrix $A$ is said to be invertible if there exists a matrix $B$ such that $AB = BA = I$. There is at most one such $B$ and is denoted by $A^{-1}$.

Easy to verify that,

- \[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{if} \quad (ad - bc) \neq 0.
\]

- \[
\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 \\ 0 & 1/\lambda_2 \end{pmatrix} \quad \text{if} \quad \lambda_1, \lambda_2 \neq 0.
\]
Mathematical Background

Matrices

A product of invertible matrices is invertible and

$$(AB)^{-1} = B^{-1}A^{-1}$$

- We denote the determinant of a matrix $A$ by $\text{det}(A)$.

If $\text{det}(A) \neq 0$, then $A$ is invertible.

- The matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if

  $$\text{det} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$$

  i.e. $ad - bc \neq 0$

- The matrix $Q$ is orthogonal if $Q^{-1} = Q^T$. 
Matrix-vector multiplication, $Ax$

- $A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix}$ and $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.
- $Ax = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$
Matrix-vector multiplication, \( Ax \)

- \( A = \begin{pmatrix} 3 & 2 \\ 2 & 0 \end{pmatrix} \) and \( x = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \).
- \( Ax = \begin{pmatrix} 8 \\ 4 \end{pmatrix} = 4 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 4x \)
Eigenvalues and Eigenvectors

Definition

Let $A \in \mathbb{R}^{n \times n}$. The eigenvalues and eigenvectors of $A$ are the real or complex scalars $\lambda$ and $n$-dimensional vectors $x$ such that

$$Ax = \lambda x, \ x \neq 0.$$

- $Ax = \lambda x \Rightarrow (A - \lambda I)x = 0$
- $\lambda$ is an eigenvalue of $A$ if and only if
  $$\det(A - \lambda I) = 0 \quad (\text{characteristic equation of } A)$$
- This equation has $n$ roots and are called the eigenvalues of $A$. 
Eigenvalues and Eigenvectors

- Let $A = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$.
- Characteristic equation:

$$\det \begin{pmatrix} 4 - \lambda & -5 \\ 2 & -3 - \lambda \end{pmatrix} = 0$$

$$\Rightarrow (\lambda^2 - \lambda - 2) = 0$$

$$\Rightarrow \lambda = 2 \text{ or } \lambda = -1$$

- $\lambda_1 = 2$, $(A - \lambda_1 I)x_1 = 0$ gives $x_1$ to be a multiple of $(5, 2)^T$.
- $\lambda_2 = -1$, $(A - \lambda_2 I)x_2 = 0$ gives $x_2$ to be a multiple of $(1, 1)^T$.
- Eigenvalues of $A$ : 2 and $-1$
- The corresponding eigenvectors of $A$ : $(5, 2)^T$ and $(1, 1)^T$
Mathematical Background

Symmetric Matrices

**Definition**

Let $A \in \mathbb{R}^{n \times n}$. The matrix $A$ is said to be *symmetric* if $A^T = A$.

- Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then,
  - $A$ has $n$ real eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, and
  - a corresponding set of eigenvectors $\{x_1, x_2, \ldots, x_n\}$ can be chosen to be orthonormal.
  - $S = (x_1, x_2, \ldots, x_n)$ is an orthogonal matrix ($S^{-1} = S^T$).
  - $S^T A S = \begin{pmatrix}
\lambda_1 & 0 & \ldots & 0 \\
0 & \lambda_2 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots \\
0 & 0 & \ldots & \lambda_n
\end{pmatrix} = \Lambda$
Mathematical Background

Quadratic Form

- Let \( A \in \mathbb{R}^{n \times n} \) be a symmetric matrix
- Consider \( f(x) = x^T A x \), a pure quadratic form

<table>
<thead>
<tr>
<th>A is said to be</th>
<th>if</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive definite (pd)</td>
<td>( x^T A x &gt; 0 ) for every nonzero ( x \in \mathbb{R}^n )</td>
</tr>
<tr>
<td>positive semi-definite (psd)</td>
<td>( x^T A x \geq 0 ) for every ( x \in \mathbb{R}^n )</td>
</tr>
<tr>
<td>negative definite (nd)</td>
<td>( x^T A x &lt; 0 ) for every nonzero ( x \in \mathbb{R}^n )</td>
</tr>
<tr>
<td>negative semi-definite (nsd)</td>
<td>( x^T A x \leq 0 ) for every ( x \in \mathbb{R}^n )</td>
</tr>
<tr>
<td>indefinite</td>
<td>( A ) is neither positive definite\nor negative definite</td>
</tr>
</tbody>
</table>

- **Question**: How to numerically check the positive definiteness of \( A \)?
Mathematical Background

Quadratic Form

- Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix
- Consider $f(x) = x^T Ax$, a pure quadratic form
- Eigenvalues of $A$ : $\lambda_1, \lambda_2, \ldots, \lambda_n$
- Orthonormal Eigenvectors of $A$ : $x_1, x_2, \ldots, x_n$
- $S = (x_1, x_2, \ldots, x_n)$

\[
x^T Ax = x^T S \Lambda S^T x = y^T \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2
\]

Therefore, $\lambda_i > 0 \ \forall \ i \ \Rightarrow \ x^T Ax > 0$
Mathematical Background

To prove that $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \Rightarrow$ Every eigen value of $\mathbf{A}$ is positive.

- Given, $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for every $\mathbf{x} \neq 0$
- Therefore, $\mathbf{x}_i^T \mathbf{A} \mathbf{x}_i > 0$ for every eigen vector $\mathbf{x}_i$
- That is, $\lambda_i \mathbf{x}_i^T \mathbf{x}_i > 0$ for every eigen vector $\mathbf{x}_i$
- Thus, $\lambda_i > 0$ for every eigen vector $\mathbf{x}_i$. 
Mathematical Background

Let \( A \in \mathbb{R}^{n \times n} \) be symmetric. Then,

<table>
<thead>
<tr>
<th>( A ) is said to be</th>
<th>if and only if, all the eigenvalues of ( A ) are</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive definite (pd)</td>
<td>positive</td>
</tr>
<tr>
<td>positive semi-definite (psd)</td>
<td>non-negative</td>
</tr>
<tr>
<td>negative definite (nd)</td>
<td>negative</td>
</tr>
<tr>
<td>negative semi-definite (nsd)</td>
<td>non-positive</td>
</tr>
</tbody>
</table>

- \( A \) is indefinite if and only if, it has both positive and negative eigenvalues.
Some other ways of checking positive definiteness

Let \( A \in \mathbb{R}^{n \times n} \) be symmetric.

- **Sylvester’s criterion:** \( A \) is positive definite if all its leading principal minors are positive.

\[
\begin{pmatrix}
  a & b & c \\
  b & e & f \\
  c & f & g
\end{pmatrix},
\begin{pmatrix}
  a & b & c \\
  b & e & f \\
  c & f & g
\end{pmatrix},
\begin{pmatrix}
  a & b & c \\
  b & e & f \\
  c & f & g
\end{pmatrix}
\]

- \( A \) is positive definite if there exists a unique lower triangular matrix \( L \in \mathbb{R}^{n \times n} \) with positive diagonal components such that \( A = LL^T \) (Cholesky Decomposition).
Examples

- \[
\begin{pmatrix}
2 & -1 & 0 \\
-1 & 2 & -1 \\
0 & -1 & 2
\end{pmatrix}
\]
is positive definite

(The eigenvalues are \(2 - \sqrt{2}, 2 + \sqrt{2}\) and 2).

- \[
\begin{pmatrix}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 2
\end{pmatrix}
\]
is positive semi-definite

- \[
\begin{pmatrix}
1 & -2 & 4 \\
-2 & 2 & 0 \\
4 & 0 & -7
\end{pmatrix}
\]
is indefinite
Solution of $Ax = b$

- Let $A \in \mathbb{R}^{n \times n}$, symmetric and positive definite
- Solution of $Ax = b$ is $x^* = A^{-1}b$
- Instead, use Cholesky decomposition of $A$, $A = LL^T$
- The given system of equations is $LL^T x = b$
- Solve the *triangular* system, $Ly = b$ using *forward substitution* to get $y$.
- Solve the *triangular* $L^T x = y$ using *backward substitution* to get $x^*$.
- Cholesky decomposition is a *numerically stable* procedure
Mathematical Background

Solution of $Ax = b$

- $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 0 \\ 4 \\ -4 \end{pmatrix}$

- Cholesky decomposition of $A = LL^T$ gives

$$L = \begin{pmatrix} 1.4142 & 0 & 0 \\ -0.7071 & 1.2247 & 0 \\ 0 & -0.8165 & 1.1547 \end{pmatrix}$$

- Solution of $Ly = b$ gives $y = \begin{pmatrix} 0 \\ 3.2660 \\ -1.1547 \end{pmatrix}$

- Solution of $L^Tx = y$ results in

$$x^* = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$
Some References