Positive Definite Matrix

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Linear Algebra
Outline and Notation

- $x^T A x$: quadratic form
- $f(x)$: multi-variate function
- $\nabla f(x)$: gradient vector
- $H$: Hessian matrix
- $\sigma_i$: singular value
- $\Sigma$: singular value matrix
- $A = U \Sigma V^T$: singular value decomposition of $A$
- $A^+$: pseudo-inverse of $A$
Quadratic Function and Matrix
A quadratic function of variables \( x_1, \ldots, x_n \) is a linear combination of the second-order terms \( x_i^2 \) and \( x_i x_j \).

**Details.** Let \( c_{ij} \) be the coefficient of term \( x_i x_j \) of a quadratic function of \( n \) variables \( f(x_1, \ldots, x_n) \). Then

\[
f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j
\]
A quadratic function of $n$ variables can be represented by a symmetric matrix of order $n \times n$.

**Construction of matrix.** For $f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} c_{ij}x_ix_j$, define matrix $A$ with $a_{ij} = \frac{1}{2}(c_{ij} + c_{ji})$. Note $a_{ij} = a_{ji}$ so $A$ is symmetric. Furthermore, $a_{ij} + a_{ji} = c_{ij} + c_{ji}$, so

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} c_{ij}x_ix_j = \sum_{i,j=1}^{n} a_{ij}x_ix_j = \mathbf{x}^T A \mathbf{x}$$

**Example.** $f(x, y) = ax^2 + 2bxy + cy^2$ can be represented by $f(x) = \mathbf{x}^T A \mathbf{x}$ where

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$
The function $x^T A x$ is called the quadratic form of $A$.

**Positive definite.** Matrix $A$ is said to be positive definite if its quadratic form $x^T A x$ is positive for any $x \neq 0$. 
Positivity of Eigenvalues

Every eigenvalue of a positive definite matrix is positive.

Proof. Suppose $A$ is a positive definite matrix. Let $\lambda$ be an eigenvalue of $A$, and $s$ be an eigenvector of $A$ corresponding to $\lambda$. We have

$$As = \lambda s$$

It follows that

$$s^T As = \lambda (s^T s)$$

Hence

$$\lambda = \frac{s^T As}{s^T s} > 0$$
A matrix is positive definite if every eigenvalue of the matrix is positive.

**Proof.** Suppose every eigenvalue of \( A \) is positive. By spectral theorem, \( A \) has an eigenvalue decomposition \( A = Q \Lambda Q^T \). It follows that

\[
x^T A x = x^T Q \Lambda Q^T x = y^T \Lambda y = \sum_i \lambda_i y_i^2
\]

Hence, the quadratic form \( x^T A x \) is positive for any \( x \neq 0 \), and \( A \) is positive definite.
If a matrix is positive definite, then the determinant of every leading principal sub-matrix is positive.

\textbf{Proof.} Suppose } \mathbf{A} \text{ is positive definite. For every } k, \text{ consider } \mathbf{x}^T = \begin{bmatrix} \mathbf{x}_k^T & 0^T \end{bmatrix} \text{ with } \mathbf{x}_k \in \mathbb{R}^k. \text{ For a non-zero } \mathbf{x}_k, \text{ we have } \mathbf{x} \neq 0, \text{ and}

\[ \mathbf{x}^T \mathbf{A} \mathbf{x} = \begin{bmatrix} \mathbf{x}_k^T & 0^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_k & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{x}_k \\ 0 \end{bmatrix} = \mathbf{x}_k^T \mathbf{A}_k \mathbf{x}_k > 0 \]

So } \mathbf{A}_k, \text{ the leading principle sub-matrix of } \mathbf{A} \text{ of order } k \times k, \text{ is positive definite. Since the determinant of a matrix is the product of eigenvalues, and every eigenvalue of } \mathbf{A}_k \text{ is positive, } |\mathbf{A}_k| \text{ must be positive.
Positivity of Pivots

If the determinant of every leading principal sub-matrix of a matrix is positive, then the matrix has full positive pivots.

**Proof.** By assumption $|A| > 0$, so $A$ is non-singular. Let $A = LDU$ be the LDU decomposition of $A$. Explicitly

$$
\begin{bmatrix}
A_k & B \\
B^T & C
\end{bmatrix}
= \begin{bmatrix}
L_k & 0 \\
0 & * \\
0 & *
\end{bmatrix}
\begin{bmatrix}
D_k & 0 \\
0 & *
\end{bmatrix}
\begin{bmatrix}
U_k & * \\
0 & *
\end{bmatrix}
= \begin{bmatrix}
L_k D_k U_k & * \\
* & *
\end{bmatrix}
$$

So $A_k = L_k D_k U_k$ and $|A_k| = |D_k| = d_1 \ldots d_k$ where $d_i$ is a pivot. Thus

$$
d_k = \frac{|A_k|}{|A_{k-1}|} > 0, \quad k = 1, \ldots, n
$$
Positive Pivots

If a matrix has full positive pivots, then the matrix is positive definite.

**Proof.** By assumption, $A$ has full pivots, so it is non-singular. Let $A = LDU$ be the LDU decomposition of $A$. Since $A$ is symmetric, $A = A^T$ or $LDU = U^TDL^T$, so $U = L^T$. Thus

$$A = LDL^T = LD^{1/2}D^{1/2}L^T = R^TR$$

where $R = D^{1/2}L^T$ is non-singular. The quadratic form of $A$ is

$$x^TAx = x^TR^TRx = (Rx)^T(Rx) = \|Rx\|^2$$

which is positive for $x \neq 0$. Hence $A$ is positive definite.
Equivalent Statements for PDM

There are many ways to say a matrix is positive definite.

1. \(A\) is positive definite.
2. Every eigenvalue of \(A\) is positive.
3. The determinant of every leading principal sub-matrices of \(A\) is positive.
4. \(A\) has full positive pivots.

What we have shown in the previous slides are

\[1 \iff 2\]

and

\[1 \implies 3 \implies 4 \implies 1\]
Example

\[
A = \begin{bmatrix}
2 & -1 & 0 \\
-1 & 2 & -1 \\
0 & -1 & 2
\end{bmatrix}
\]

The quadratic form of \( A \) is

\[
x^T A x = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3
\]

\[
= 2 \left( x_1 - \frac{1}{2}x_2 \right)^2 + \frac{3}{2} \left( x_2 - \frac{2}{3}x_3 \right)^2 + \frac{4}{3}x_3^2
\]

The eigenvalues, the determinants, and the pivots are

\[
\text{spectrum}(A) = \{2, 2 \pm \sqrt{2}\}, \ |A_1| = 2, \ |A_2| = 3, \ |A_3| = 4
\]
Let $A$ be a positive definite matrix. Then the equation $x^T A x = 1$ is an **ellipsoid**.

**Explanation.** By spectral theorem $A = Q \Lambda Q^T$. Note that $\{q_1, \ldots, q_n\}$ is an orthonormal basis, and the representation of $x$ in this basis is $Q^T x$. By a change of basis, $x^T A x = 1$ can be converted to

$$x^T Q \Lambda Q^T x = y^T \Lambda y = \sum_i \lambda_i y_i^2 = 1$$

This is an ellipsoid with the axes of symmetry along $q_i$’s, with the intercepts of

$$y_i = \pm \left(\sqrt{\lambda_i}\right)^{-1}$$
Negative definite. Matrix $A$ is said to be negative definite if its quadratic form $x^T Ax$ is negative for any $x \neq 0$.

Semi-definite. Matrix $A$ is said to be positive semi-definite (resp. negative semi-definite) if its quadratic form $x^T Ax$ is non-negative (resp. non-positive) for any $x$. 
Approximation and Extremal Points
The first-order approximation to a multi-variate function $f(x)$ near $x_0$ is

$$f(x) \approx f(x_0) + \nabla f(x_0)^T(x - x_0)$$

where

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

Gradient. $\nabla f(x)$ is called the gradient of $f(x)$. 

Chen P  Positive Definite Matrix
The second-order approximation to \( f(x) \) near \( x_0 \) is

\[
f(x) \approx f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2} (x - x_0)^T H(x_0) (x - x_0)
\]

where

\[
H(x) = \begin{bmatrix}
\frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 f(x)}{\partial x_i \partial x_j} & \cdots & \frac{\partial^2 f(x)}{\partial x_j^2}
\end{bmatrix}
\]

Hessian. \( H(x) \) is called the Hessian of \( f(x) \).
$x_0$ is called a **stationary point** of $f(x)$ if $\nabla f(x_0) = 0$.

Near a stationary point. Suppose $x_0$ is a stationary point of $f(x)$. Near $x_0$, the second-order approximation to $f(x)$ is

$$f(x) \approx f(x_0) + \frac{1}{2}(x - x_0)^T H(x_0)(x - x_0)$$
Example

Find the second-order approximation near $x_0 = 0$ to

$$f(x) = 2x^2 + 4xy + y^2$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f(x)}{\partial x} \\ \frac{\partial f(x)}{\partial y} \end{bmatrix} = \begin{bmatrix} 4x + 4y \\ 4x + 2y \end{bmatrix}, \quad H(x) = \begin{bmatrix} \frac{\partial}{\partial x} \left( \frac{\partial f(x)}{\partial x} \right) & \frac{\partial}{\partial y} \left( \frac{\partial f(x)}{\partial x} \right) \\ \frac{\partial}{\partial x} \left( \frac{\partial f(x)}{\partial y} \right) & \frac{\partial}{\partial y} \left( \frac{\partial f(x)}{\partial y} \right) \end{bmatrix}$$

$$f(0) = 0, \quad \nabla f(0) = 0, \quad H(0) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$$

$$f(x) \approx f(0) + \frac{1}{2}(x - 0)^T H(0)(x - 0) = 2x^2 + 4xy + y^2$$
Example

Find the second-order approximation near $x_0 = 0$ to

$$F(x) = 7 + 2(x + y)^2 - y \sin y - x^3$$

\[
\begin{align*}
\nabla F(x) &= \left[\frac{\partial F(x)}{\partial x}, \frac{\partial F(x)}{\partial y}\right] = \begin{bmatrix} 4(x + y) - 3x^2 \\ 4(x + y) - \sin y - y \cos y \end{bmatrix} \\
H(x) &= \begin{bmatrix}
\frac{\partial}{\partial x} \left( \frac{\partial F(x)}{\partial x} \right) & \frac{\partial}{\partial y} \left( \frac{\partial F(x)}{\partial x} \right) \\
\frac{\partial}{\partial x} \left( \frac{\partial F(x)}{\partial y} \right) & \frac{\partial}{\partial y} \left( \frac{\partial F(x)}{\partial y} \right)
\end{bmatrix} = \begin{bmatrix} 4 - 6x & 4 \\ 4 & 4 - 2 \cos y + y \sin y \end{bmatrix}
\end{align*}
\]

$$F(0) = 7, \quad \nabla F(0) = 0, \quad H(0) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}$$

$$F(x) \approx F(0) + \frac{1}{2}(x - 0)^T H(0)(x - 0) = 7 + 2x^2 + 4xy + y^2$$
A point $x_0$ is called a **local minimum** of $f(x)$ if $f(x) \geq f(x_0)$ for every point in a small neighborhood of $x_0$.

Similarly, $x_0$ is called a **local maximum** if $f(x) \leq f(x_0)$ in a small neighborhood of $x_0$. 
Let $x_0$ be a stationary point of $f(x)$.

- $x_0$ is **local minimum** if $H(x_0)$ is positive definite.
- $x_0$ is **local maximum** if $H(x_0)$ is negative definite.
- $x_0$ is a **saddle point** if it is neither local maximum nor local minimum.

For example, $0$ is a stationary point of $F(x)$, and it is a saddle point because $2x^2 + 4xy + y^2$ can be positive or negative as $x$ and $y$ vary.
Singular Value Decomposition
A **singular value** of a real matrix $A$ is the square root of a non-zero eigenvalue of $(A^TA)$.

It means to find the singular values of $A$, one needs to find the non-zero eigenvalues of $(A^TA)$.

**Singular vector.** If $\sigma$ is a singular value of $A$, then there exists $\mathbf{v} \neq 0$ such that

$$
(A^TA)\mathbf{v} = \sigma^2\mathbf{v}
$$

Such a $\mathbf{v}$ is called a right singular vector of $A$ with singular value $\sigma$. It is an eigenvector of $(A^TA)$ with eigenvalue $\sigma^2$. 
A singular value is always positive.

The matrix \((A^T A)\) is positive semi-definite:

\[ x^T (A^T A) x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0 \]

so the eigenvalues of \((A^T A)\) must be non-negative, and the non-zero eigenvalues must be positive. Hence a singular value is positive.
A matrix of rank \( r \) has exactly \( r \) singular values.

**Proof.** Note \((A^T A)x = 0 \Leftrightarrow x^T A^T A x = 0 \Leftrightarrow A x = 0\), so \(\mathcal{N}(A^T A) = \mathcal{N}(A)\). Let \( A \) be of order \( m \times n \) with rank \( r \). Then \( \dim \mathcal{N}(A) = n - r = \dim \mathcal{N}(A^T A) \). Matrix \((A^T A)\) is non-defective, so the algebraic multiplicity of eigenvalue 0 is \( (n - r) \). It follows that the total algebraic multiplicities of the non-zero eigenvalues of \((A^T A)\) is

\[
n - (n - r) = r
\]

**Notation.** Singular values are denoted by \( \sigma_1, \ldots, \sigma_r \).
A real matrix can be decomposed by its singular values and singular vectors. This is called singular value decomposition.

A matrix of order $m \times n$ has SVD

$$A = U \Sigma V^T$$

where $\Sigma$ is an $m \times n$ "diagonal" matrix with the singular values of $A$ as the leading diagonal elements, $U$ is an $m \times m$ orthogonal matrix with the eigenvectors of $(A A^T)$ as columns, and $V$ is an $n \times n$ orthogonal matrix with the eigenvectors of $(A^T A)$ as columns.
Proof of SVD 1

Let $r$ be the rank of $A$. Let $\sigma_1 \ldots \sigma_r$ be the singular values of $A$. Let $v_1 \ldots v_r$ be orthonormal eigenvectors of $(A^T A)$ with positive eigenvalues $\sigma_i^2$ and $u_1 \ldots u_r$ be defined by $u_i = \frac{Av_i}{\sigma_i}$. Note

$$(AA^T)u_i = \frac{AA^T Av_i}{\sigma_i} = \frac{A\sigma_i^2 v_i}{\sigma_i} = \sigma_i^2 u_i, \quad i = 1, \ldots, r$$

So $u_i$ is an eigenvector of $(AA^T)$ with the same eigenvalue $\sigma_i^2$. Let $v_{r+1} \ldots v_n$ be orthonormal eigenvectors of $(A^T A)$ with eigenvalue 0, and $u_{r+1} \ldots u_m$ be eigenvectors of $(AA^T)$ with eigenvalue 0. Construct matrices $U$ and $V$ by

$$U = \begin{bmatrix} u_1 & \ldots & u_m \end{bmatrix}, \quad V = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}$$
We show $\mathbf{U}^T \mathbf{A} \mathbf{V} = \Sigma$ which leads to SVD $\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$. For $j = 1 \ldots r$, we have $\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{\sigma_j}$, so $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$ and

$$(\mathbf{U}^T \mathbf{A} \mathbf{V})_{ij} = \mathbf{u}_i^T \mathbf{A} \mathbf{v}_j = \mathbf{u}_i^T (\sigma_j \mathbf{u}_j) = \sigma_j \delta_{ij}, \ i = 1, \ldots, m$$

For $j = r + 1 \ldots n$, we have $(\mathbf{A}^T \mathbf{A}) \mathbf{v}_j = 0$, so $\mathbf{A} \mathbf{v}_j = 0$ and

$$(\mathbf{U}^T \mathbf{A} \mathbf{V})_{ij} = \mathbf{u}_i^T \mathbf{A} \mathbf{v}_j = 0, \ i = 1, \ldots, m$$

Combining the results, we get

$$\mathbf{U}^T \mathbf{A} \mathbf{V} = \Sigma$$

Hence

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^T$$
Matrices in SVD

For a matrix of order $m \times n$ with SVD $A = U\Sigma V^T$, the column vectors of $U$ (resp. $V$) is an orthonormal basis of $\mathbb{R}^m$ (resp. $\mathbb{R}^n$).

$U$ must be an eigenvector matrix of $AA^T$.

$$AA^T = (U\Sigma V^T)(V\Sigma^T U^T) = U \begin{pmatrix} \Sigma \Sigma^T \end{pmatrix} U^T$$

Similarly, $V$ must be an eigenvector matrix of $A^T A$.

$$A^T A = (V\Sigma^T U^T)(U\Sigma V^T) = V \begin{pmatrix} \Sigma^T \Sigma \end{pmatrix} V^T$$
The right (resp. left) singular vectors in an SVD of a matrix form an orthonormal basis of the row space (resp. column space) of the matrix.

**Row space.** \( \{v_{r+1}, \ldots, v_n\} \) contains eigenvectors of \( (A^T A) \) with eigenvalue 0, so it is a basis of \( N(A^T A) = N(A) \). This implies \( \{v_1, \ldots, v_r\} \) is a basis of the orthogonal complement of \( N(A) \), i.e. the row space of \( A \).

**Column space.** We have \( AV = U\Sigma \). The first \( r \) columns are

\[
Av_i = \sigma_i u_i, \quad i = 1, \ldots, r
\]

So \( \{u_1, \ldots, u_r\} \) is a linearly independent set in the column space of \( A \). Hence, it is a basis of \( C(A) \).
Example

Find SVD of

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

The eigenvalues of $$A^T A$$ are

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

are $$\lambda_1 = 3, \lambda_2 = 1, \lambda_3 = 0$$. Hence the singular values of $$A$$ are

$$\sigma_1 = \sqrt{3}, \sigma_2 = 1$$
Orthonormal eigenvectors of \( (A^T A) \) are

\[
\mathbf{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\]

The corresponding left singular vectors of \( A \) are

\[
\mathbf{u}_1 = \frac{A \mathbf{v}_1}{\sigma_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{A \mathbf{v}_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}
\]

So

\[
A = U \Sigma V^T = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}
\]
Every real matrix of rank $r$ is the sum of $r$ real matrices of rank 1 based on singular values and singular vectors.

By SVD

$$A = U \Sigma V^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

$$= A_1 + \cdots + A_r$$

**Image approximation.** For an image of size $1000 \times 1000$, a compression rate of 90% is achieved if 50 terms are used.

*Data Compression with SVD*
Let $A = UΣV^T$ be an SVD of $A$. For a rectangular system of linear equations $Ax = b$, the least-squares solution with the minimum length is $x^+ = VΣ^+U^Tb$.

**Pseudo-inverse.** The minimum-length least-squares solution can be written as $x^+ = A^+b$, where $A^+ = VΣ^+U^T$. $A^+$ is called the **pseudo-inverse** of $A$. 