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
- Positive definite matrix
- $f(x)$: 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 $f$. Then $f$ can be written by

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

**Matrix construction.** Consider $f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} c_{ij}x_ix_j$.

Define matrix $A = \{a_{ij}\}$ with $a_{ij} = \frac{1}{2}(c_{ij} + c_{ji})$. We have $a_{ij} = a_{ji}$ and $a_{ij} + a_{ji} = c_{ij} + c_{ji}$, and

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

**Example.** A quadratic function $f(x, y) = ax^2 + 2bxy + cy^2$ can be represented by $f(x) = x^TAx$ where

$$x = \begin{bmatrix} x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$
Matrix $A$ is said to be **positive definite** if $x^T A x$ is positive for any $x \neq 0$.

**Semi-definite.** Matrix $A$ is said to be positive semi-definite if

$$x^T A x \geq 0$$
Every eigenvalue of a positive definite matrix is positive.

**Proof.** Let $A$ be a positive definite matrix, $\lambda$ be an eigenvalue of $A$, and $x$ be an eigenvector of $A$ corresponding to $\lambda$. Then

$$Ax = \lambda x \Rightarrow x^T Ax = \lambda (x^T x)$$

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

**Proof.** By spectral theorem, a real symmetric matrix has an eigenvalue decomposition, say $A = Q \Lambda Q^T$. Suppose every eigenvalue of $A$ is positive. Then for any $x \neq 0$, the quadratic function is positive since

$$x^T A x = x^T Q \Lambda Q^T x = y^T \Lambda y = \sum_{i} \lambda_i y_i^2 > 0$$
If a matrix is positive definite, then the determinant of every leading principal sub-matrix is positive.

**Proof.** Let $A$ be positive definite. For $x_k \neq 0$, consider the quadratic function of $x^T = \begin{bmatrix} x_k^T & 0^T \end{bmatrix}$.

$$x^T Ax = \begin{bmatrix} x_k^T & 0^T \end{bmatrix} \begin{bmatrix} A_k & B \\ B^T & C \end{bmatrix} \begin{bmatrix} x_k \\ 0 \end{bmatrix} = x_k^T A_k x_k > 0$$

So $A_k$, a leading principle sub-matrix of $A$ of order $k \times k$, is positive definite. It follows that every eigenvalue of $A_k$ is positive, which implies

$$|A_k| = \lambda_{k,1} \times \cdots \times \lambda_{k,k} > 0$$
If the determinant of every leading principal sub-matrix of a matrix is positive, then every pivot of the matrix is positive.

**Proof.** $|A|$ is positive, so $A$ is non-singular, with the LDU decomposition of $A = LDU$. Explicitly

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

So $A_k = L_kD_kU_k$ and $|A_k| = |D_k| = d_1 \ldots d_k$. Thus we have

\[
d_k = \frac{|A_k|}{|A_{k-1}|} > 0, \quad k = 1, \ldots, n
\]
If every pivot of a matrix is positive, then the matrix is positive definite.

**Proof.** With positive pivots, $A$ is non-singular with the LDU decomposition of $A = LDU$, where $d_{ii} > 0$. Here we have $A^T = A$, or $LDU = U^TDL^T$. By the uniqueness of LDU decomposition, we have $U = L^T$. Thus

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

where $R = D^{1/2}L^T$ is non-singular. For $x \neq 0$

$$x^TAx = x^TR^TRx = (Rx)^T(Rx) = \|Rx\|^2 > 0$$

Thus $A$ is positive definite.
Example

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

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

\[
= 2 \left( x_1 - \frac{1}{2} x_2 \right)^2 + \frac{3}{2} \left( x_2 - \frac{2}{3} x_3 \right)^2 + \frac{4}{3} x_3^2 > 0
\]

\[
\lambda = 2, 2 \pm \sqrt{2}
\]

\[
|A_1| = 2, \; |A_2| = 3, \; |A_3| = 4
\]

\[
A = \begin{bmatrix}
1 & 0 & 0 \\
-\frac{1}{2} & 1 & 0 \\
0 & -\frac{2}{3} & 1 \\
\end{bmatrix} \begin{bmatrix}
2 \\
\frac{3}{2} \\
\frac{4}{3} \\
\end{bmatrix} \begin{bmatrix}
1 & -\frac{1}{2} & 0 \\
0 & 1 & -\frac{2}{3} \\
0 & 0 & 1 \\
\end{bmatrix}
\]
Let $A$ be a positive definite matrix of order $n \times n$. The equation

$$x^T Ax = 1$$

represents an **ellipsoid**.

**Explanation.** By the spectral theorem, $A = Q\Lambda Q^T$, so

$$x^T Ax = x^T Q\Lambda Q^T x = y^T \Lambda y$$

where $y_i = q_i^T x$. Note $y_i q_i$ is the projection of $x$ to eigenvector $q_i$. The equation $y^T \Lambda y = 1$ is an ellipsoid, with axes aligning with the eigenvectors of $A$. Along the direction of $q_i$, the end points are

$$y_i = \pm \left(\sqrt{\lambda_i}\right)^{-1}, \quad y_j = 0 \ \forall \ j \neq i$$
A matrix is said to be **negative definite** if $x^T A x < 0$ for all $x \neq 0$.

Equivalent statements.

- $A$ is negative definite.
- Every eigenvalue of $A$ is negative.
- The determinants of the leading principal sub-matrices of $A$ alternate between negative and positive.
- Every pivot of $A$ is negative.
Approximation and Extremal Points
First-order Approximation

The first-order approximation to a multi-variate function $f(x)$ near $x_0$ is

$$f(x_0 + dx) \approx f(x_0) + \sum_i \frac{\partial f}{\partial x_i}(x_0) dx_i = f(x_0) + \nabla f(x_0)^T dx$$

where

$$dx = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}, \quad \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Gradient. $\nabla f$ is called the gradient of $f$. 
Second-order Approximation

The second-order approximation to $f(x)$ near $x_0$ is

$$f(x_0 + dx) \approx f(x_0) + \sum_i \frac{\partial f}{\partial x_i}(x_0) dx_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) dx_i dx_j$$

$$= f(x_0) + \nabla f(x_0)^T dx + \frac{1}{2} dx^T H(x_0) dx$$

where

$$dx = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}, \quad \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \quad H = \left\{ \frac{\partial^2 f}{\partial x_i \partial x_j} \right\}$$

Hessian. The matrix $H$ is called the Hessian of $f$. 
x₀ is called a **stationary point** of \( f(x) \) if \( \nabla f(x₀) = 0 \).

**Approximation near a stationary point.** Suppose \( x₀ \) is a stationary point of \( f \). Letting \( x = x₀ + dx \), the second-order approximation to \( f \) near \( x₀ \) is

\[
f(x) \approx f(x₀) + \frac{1}{2}(x - x₀)^T H(x₀)(x - x₀)
\]
Example

Find the second-order approximation near \( x_0 = 0 \) to

\[
\begin{align*}
f(x) &= 2x^2 + 4xy + y^2 \quad \text{and} \quad F(x) = 7 + 2(x + y)^2 - y \sin y - x^3
\end{align*}
\]

\[
f(0) = 0, \quad \nabla f(0) = 0, \quad H_f(0) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}
\]

\[
F(0) = 7, \quad \nabla F(0) = 0, \quad H_F(0) = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix}
\]

So

\[
f(x) \approx f(0) + \frac{1}{2} x^T H_f(0) x = 2x^2 + 4xy + y^2
\]

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

Similarly, $x_0$ is a **local maximum** if $f(x_0) \geq f(x)$ for every point 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 saddle point of $F(x)$. 
Singular Value Decomposition
A **singular value** of a real matrix $A$ is the square root of a non-zero eigenvalue of $(A^T A)$.

Let $\sigma$ be a singular value of $A$. Then there exists $v \neq 0$ such that

$$(A^T A) v = \sigma^2 v$$

A **singular vector** of $A$ corresponding to singular value $\sigma$ is an eigenvector of $(A^T A)$ corresponding to eigenvalue $\sigma^2$. 
The matrix \((A^T A)\) is positive semi-definite:

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

Therefore all of the eigenvalues of \((A^T A)\) are non-negative. It follows that every non-zero eigenvalue of \((A^T A)\) is positive. Thus, every singular value of \(A\) is positive.
A real matrix of rank $r$ has exactly $r$ singular values.

**Proof.** Note $A^T A x = 0 \iff x^T A^T A x = 0 \iff A x = 0$, so $\mathcal{N} (A^T A) = \mathcal{N}(A)$. Let $A$ be of order $m \times n$ with rank $r$. Then the dimension of $\mathcal{N}(A)$ is $n - r$, as well as $\mathcal{N} (A^T A)$. Since $(A^T A)$ is $n \times n$, the sum of the algebraic multiplicities of the eigenvalues of $(A^T A)$ is $n$. Among the eigenvalues, 0 is an eigenvalue with algebraic multiplicity of $n - r$. Thus, the sum of the algebraic multiplicities of the other (positive) eigenvalues is

$$n - (n - r) = r$$

Singular values are denoted by $\sigma_1, \ldots, \sigma_r$. 
Let $\mathbf{A}$ be a matrix of order $m \times n$. Since $(\mathbf{A}^T \mathbf{A})$ is real symmetric, an orthonormal eigenbasis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $(\mathbf{A}^T \mathbf{A})$ exists. We can let $\mathbf{v}_1 \ldots \mathbf{v}_r$ be singular vectors of $\mathbf{A}$ and let $\mathbf{v}_{r+1} \ldots \mathbf{v}_n$ be eigenvectors of $(\mathbf{A}^T \mathbf{A})$ with eigenvalue 0. Constructing $\mathbf{V}$ with $\mathbf{v}_1 \ldots \mathbf{v}_n$, we have

$$
(\mathbf{A}^T \mathbf{A}) = \mathbf{V} \text{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0) \mathbf{V}^T
$$
\((AA^T)\) and \((A^TA)\) have important things in common.

- They are both positive semi-definite.
- They have the same rank: From \(\mathcal{N}(AA^T) = \mathcal{N}(A^T)\), the dimension of \(\mathcal{N}(AA^T)\) is \(m - r\), so the rank of \((AA^T)\) is \(r\).
- They have the same number of positive eigenvalues.
- They have the same positive eigenvalues.

\[
(AA^T) = U \text{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0) U^T
\]
Singular Value Decomposition

A real matrix $A$ of order $m \times n$ can be decomposed by its singular values and singular vectors, called SVD.

An SVD decomposition of $A$ is

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

where $\Sigma$ is an $m \times n$ matrix with the singular values of $A$ as the leading diagonal elements and 0 otherwise, $U$ is an $m \times m$ orthogonal matrix, and $V$ is an $n \times n$ orthogonal matrix.
Properties

\( \mathbf{U} \) is an eigenvector matrix of \( \mathbf{A} \mathbf{A}^T \), since

\[
\mathbf{A} \mathbf{A}^T = (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) (\mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^T) = \mathbf{U} (\mathbf{\Sigma} \mathbf{\Sigma}^T) \mathbf{U}^T
\]

Similarly, \( \mathbf{V} \) is an eigenvector matrix of \( \mathbf{A}^T \mathbf{A} \), since

\[
\mathbf{A}^T \mathbf{A} = (\mathbf{V} \mathbf{\Sigma}^T \mathbf{U}^T) (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) = \mathbf{V} (\mathbf{\Sigma}^T \mathbf{\Sigma}) \mathbf{V}^T
\]

Since \( \mathbf{A} \mathbf{V} = \mathbf{U} \mathbf{\Sigma} \), we also have

\[
\mathbf{A} \mathbf{v}_i = \sigma_i \mathbf{u}_i, \quad i = 1, \ldots, r
\]
Proof of SVD 1

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 $\lambda_i = \sigma_i^2$ and $u_1 \ldots u_r$ be defined by $u_i = \frac{Av_i}{\sigma_i}$. Let $v_{r+1} \ldots v_n$ be eigenvectors of $(A^T A)$ with eigenvalue 0, and $u_{r+1} \ldots u_m$ be eigenvectors of $(AA^T)$ with eigenvalue 0.

Note $u_j$ is an eigenvector of $(AA^T)$ with eigenvalue $\lambda_j$:

$$(AA^T) u_j = \frac{AA^T Av_j}{\sigma_j} = \frac{A\lambda_j v_j}{\sigma_j} = \lambda_j u_j, \; j = 1, \ldots, r$$

Construct matrices $U$ and $V$ by

$$U = \begin{bmatrix} u_1 & \ldots & u_m \end{bmatrix}, \quad V = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}$$
Proof of SVD 2

Next we will show $U^T AV = \Sigma$ which implies $A = U \Sigma V^T$.

For $j = 1 \ldots r$, we have $Av_j = \sigma_j u_j$ from $u_j = \frac{Av_j}{\sigma_j}$, and

$$(U^T AV)_{ij} = u_i^T Av_j = u_i^T (\sigma_j u_j) = \sigma_j \delta_{ij}, \ i = 1, \ldots, m$$

For $j = r + 1 \ldots n$, we have $Av_j = 0$ from $\left( A^T A \right) v_j = 0$, and

$$(U^T AV)_{ij} = u_i^T Av_j = 0, \ i = 1, \ldots, m$$

Thus

$U^T AV = \Sigma$

That is

$A = U \Sigma V^T$
Example

Compute an SVD of

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

The eigenvalues of

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

are \( \lambda = 3, 1, 0 \), so the singular values of \( A \) are

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

$$v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad v_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Orthonormal eigenvectors of $(AA^T)$ are

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad u_2 = \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So an SVD of $A$ is

$$A = U\Sigma V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$
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.
Let $A = U \Sigma V^T$ be an SVD of $A$. For a rectangular system of linear equations $Ax = b$, the minimum-length least-squares solution is $x^+ = V \Sigma^+ U^T b$. 

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