Positive Definite Matrix

Chia-Ping Chen

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

Linear Algebra
Focus on **real symmetric matrices** in this chapter

- Important in probability theory and function approximation
- Real eigenvalues
- Instrumental to the decomposition of real rectangular matrices
Quadratic Function and Matrix
By definition, a multi-variate quadratic function of variables $x_1, \ldots, x_n$ is

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j$$

where $c_{ij}$’s are coefficients.
A multi-variate quadratic function of \( n \) variables can be represented by a symmetric matrix of size \( n \times n \).

The matrix is decided by the coefficients of the quadratic function:

Given

\[
f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_i x_j
\]

define

\[
a_{ij} = \frac{1}{2} (c_{ij} + c_{ji}), \quad A = \{a_{ij}\}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}
\]

then

\[
f(x) = x^T A x
\]
A bi-variate quadratic function

\[ f(x, y) = ax^2 + 2bxy + cy^2 \]

can be represented as

\[ f(x) = x^T Ax \]

where

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix}, \quad
\begin{bmatrix}
  a & b \\
  b & c
\end{bmatrix}
\]
By definition, a real symmetric matrix $A$ is **positive definite** if the quadratic function defined by $A$, i.e. $x^T Ax$, is always positive, unless $x = 0$. That is

$$x^T Ax > 0, \ \forall x \neq 0$$

A real symmetric matrix $A$ is **positive semi-definite** if

$$x^T Ax \geq 0, \ \forall x$$
The following statements are equivalent.

a. A is a positive definite matrix

\[ x^T A x > 0, \ \forall x \neq 0 \]

b. The eigenvalues of A are positive.

c. The determinants of the leading principal sub-matrices of A are positive.

d. The pivots of A are positive.
If a real symmetric matrix is positive definite, then the eigenvalues of the matrix are positive.

Let $\lambda$ be an eigenvalue of $A$, and $x$ be an eigenvector with eigenvalue $\lambda$. Then

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

$$\Rightarrow \lambda = \frac{x^T Ax}{x^T x} > 0$$
If the eigenvalues of a real symmetric matrix are positive, then the matrix is positive definite.

By the spectral theorem, a real symmetric matrix has an eigenvalue decomposition, so

\[ A = Q \Lambda Q^T \]

For the quadratic function defined by \( A \)

\[ x^T A x = x^T Q \Lambda Q^T x = y^T \Lambda y = \sum_{i} \lambda_i y_i^2 > 0 \]
If a real symmetric matrix is positive definite, then the determinants of the leading principal sub-matrices are positive.

For \( x^T = \begin{bmatrix} x_k^T & 0^T \end{bmatrix} \) with \( x_k \neq 0 \), we have

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

So \( A_k \), a leading principle sub-matrix of \( A \) of size \( k \times k \), is positive definite. It follows that all eigenvalues of \( A_k \) are positive. Thus

\[
|A_k| = \lambda_{k,1} \times \cdots \times \lambda_{k,k} > 0
\]
If the determinants of the leading principal sub-matrices of a real symmetric matrix are positive, then the pivots of the matrix are positive.

\[ |A| \text{ is positive, so } A \text{ is non-singular and } \]

\[ A = LDU \text{ for some } L, D, U \]

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

\[ \Rightarrow A_k = L_k D_k U_k \Rightarrow |A_k| = |D_k| = d_1 \ldots d_k \]

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

Suppose the LDU decomposition of a real symmetric matrix \( A \) is \( A = LDU \). Since \( A^T = A \), the LDU decomposition of \( A^T \) is the same, so

\[
A = LDU = A^T = U^TDL^T \Rightarrow U = L^T
\]

If \( \forall i \ d_{ii} > 0 \),

\[
A = LDL^T = LD^{1/2}D^{1/2}L^T = R^TR, \quad R = D^{1/2}L^T
\]

Thus

\[
x^T Ax = x^T R^T Rx = (Rx)^T(Rx) = \|Rx\|^2 > 0
\]
Example

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

\[ x^T A x = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 \]
\[ = 2 \left( x_1 - \frac{1}{2} x_2 \right)^2 + \frac{3}{2} \left( x_2 - \frac{2}{3} x_3 \right)^2 + \frac{4}{3} x_3^2 > 0 \]
\[ \lambda = 2, 2 \pm \sqrt{2} \]
\[ |A_1| = 2, \ |A_2| = 3, \ |A_3| = 4 \]

\[ A = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 \\ \frac{3}{2} \\ \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} \]
Ellipsoids in $n$ Dimensions

For a positive definite matrix $A$ of size $n \times n$, the equation

$$x^T A x = 1$$

is an ellipsoid in $n$ dimensions.

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

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

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

$$y_i = \pm \left(\sqrt{\lambda_i}\right)^{-1}, \quad y_j = 0 \ \forall \ j \neq i$$
By definition, a real symmetric matrix is negative definite if

\[ x^T A x < 0, \quad \forall x \neq 0 \]

The following statements are equivalent.

- A is negative definite

\[ x^T A x < 0, \quad \forall x \neq 0 \]

- The eigenvalues of A are negative.

- The determinants of the leading principal sub-matrices of A alternate between negative and positive.

- The pivots of A are negative.
Approximation and Extremal Points
The first-order approximation to a multi-variate function $f(x)$ is

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

where

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

$\nabla f(x)$ is called the **gradient** of $f(x)$. 
The second-order approximation to a multi-variate function \( f(x) \) is

\[
\begin{align*}
  f(x + dx) & \approx f(x) + \sum_i \frac{\partial f(x)}{\partial x_i} dx_i + \frac{1}{2} \sum_i \sum_j \frac{\partial^2 f(x)}{\partial x_i \partial x_j} dx_i dx_j \\
  & = f(x) + \nabla f(x)^T dx + \frac{1}{2} dx^T H(x) dx
\end{align*}
\]

The matrix

\[
H(x) = \left\{ \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right\}
\]

is called the **Hessian** of \( f(x) \).
By definition, a point $x_0$ is a **stationary point** of $f(x)$ if

$$\nabla f(x_0) = 0$$

Near a stationary point $x_0$, the first-order approximation of $f(x)$ is

$$f(x) \approx f(x_0)$$

and the second-order approximation of $f(x)$ 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 \( x_0 = 0 \) for

\[
f(x) = 2x^2 + 4xy + y^2 \text{ and } F(x) = 7 + 2(x + y)^2 - y \sin y - x^3
\]

so

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

so

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

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

so

\[
F(x) \approx F(0) + \frac{1}{2} x^T H(0) x = 7 + 2x^2 + 4xy + y^2
\]
Consider a stationary point $x_0$ of function $f(x)$. If

$$\exists \delta \text{ s.t. } \forall \|x - x_0\| < \delta \text{ we have } f(x_0) \leq f(x)$$

then $x_0$ is a local minimum. If

$$\exists \delta \text{ s.t. } \forall \|x - x_0\| < \delta \text{ we have } f(x_0) \geq f(x)$$

then $x_0$ is a local maximum.
Consider a stationary point $x_0$ of function $f(x)$. The "optimality" of $x_0$ is decided by the Hessian at $x_0$.

- If $H(x_0)$ is positive definite, $x_0$ is a **local minimum**.
- If $H(x_0)$ is negative definite, $x_0$ is a **local maximum**.
- If $H(x_0)$ is indefinite, $x_0$ is a **saddle point** (maximum in one direction, and minimum in another direction).

For example, 0 is a saddle point of $F(x)$. 
Singular Value Decomposition
By definition, the **singular values** of a real matrix $A$ are the square roots of the non-zero eigenvalues of $(A^T A)$.

Singular values are positive. The non-zero eigenvalues of $(A^T A)$ are positive since the eigenvalues of $(A^T A)$ are non-negative

$$x^T (A^T A) x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0$$
A matrix with rank $r$ has $r$ (possibly repeated) singular values.

Let $A$ be $m \times n$. Then $(A^T A)$ is $n \times n$.

$$A^T A x = 0 \iff x^T A^T A x = 0 \iff A x = 0$$

So

$$\mathcal{N}(A^T A) = \mathcal{N}(A), \quad \dim \mathcal{N}(A^T A) = \dim \mathcal{N}(A) = n - r$$

Thus $0$ is an eigenvalue of $(A^T A)$ with multiplicity $n - r$, and the sum of the multiplicities of the positive eigenvalues of $(A^T A)$ is

$$n - (n - r) = r$$
An SVD of a real matrix $A$ of size $m \times n$ is

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

where $U$ is an $m \times m$ orthogonal eigenvector matrix of $(AA^T)$

$$
(AA^T) = U(\Sigma\Sigma^T)U^T
$$

$V$ is an $n \times n$ orthogonal eigenvector matrix of $(A^TA)$

$$
(A^TA) = V(\Sigma^T\Sigma)V^T
$$

and $\Sigma$ is an $m \times n$ matrix with the singular values of $A$ on the leading diagonal positions, and 0s elsewhere.
Proof

Let

\[ V = \begin{bmatrix} v_1 & \ldots & v_r & v_{r+1} & \ldots & v_n \end{bmatrix} \]

where \( v_1, \ldots, v_r \) are orthonormal eigenvectors of \( (A^T A) \) with non-zero eigenvalues, and \( v_{r+1}, \ldots, v_n \) are eigenvectors of \( (A^T A) \) with eigenvalue 0. Let

\[ U = \begin{bmatrix} u_1 & \ldots & u_r & u_{r+1} & \ldots & u_m \end{bmatrix} \]

where \( u_j = \frac{Av_j}{\sigma_j}, \ j = 1, \ldots, r \) and \( u_{r+1}, \ldots, u_m \) are eigenvectors of \( (AA^T) \) of eigenvalue 0. Note \( u_1, \ldots, u_r \) are eigenvectors of \( (AA^T) \)

\[ (AA^T) u_j = \frac{AA^T Av_j}{\sigma_j} = \frac{A\lambda_j v_j}{\sigma_j} = \lambda_j u_j \]
Consider $U^T AV$ with elements

$$(U^T AV)_{ij} = u_i^T Av_j$$

For $j = 1, \ldots, r$

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

For $j = r + 1, \ldots, n$

$$(A^T A) v_j = 0 \Rightarrow Av_j = 0 \Rightarrow u_i^T Av_j = 0, \ i = 1, \ldots, m$$

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 \((A A^T)\) are

\[ u_1 = \frac{A v_1}{\sigma_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad u_2 = \frac{A v_2}{\sigma_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \]

So an SVD of \(A\) is

\[ A = U \Sigma V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \]
A matrix of rank $r$ can be expressed as the sum of $r$ matrices of rank 1.

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

$$= \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T$$

$$= A_1 + \cdots + A_r$$

For an image of size $1000 \times 1000$, compression of 90% can be achieved when it is approximated by the sum of 50 matrices of rank 1.

*Data Compression with SVD*