Positive Definite Matrix

Chia-Ping Chen

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

Linear Algebra
Notation

- $x^T A x$: quadratic form
- $f(x)$: a multi-variate function
- $\nabla f(x)$: the gradient vector of $f(x)$
- $H(x)$: Hessian matrix of a multi-variate function
- $\sigma$: a singular value
- $\Sigma$: a singular value matrix
- $A = U \Sigma V^T$: the singular value decomposition of $A$
- $A^+$: pseudo-inverse of $A$
Quadratic Function and Quadratic Form
**Definition**

A function is **quadratic** if it is a sum of the second-order terms.

Let \( f(x_1, \ldots, x_n) \) be quadratic. Then

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

Let \( f(x_1, \ldots, x_n) \) be quadratic. Then

\[
f(x_1, \ldots, x_n) = \mathbf{x}^T \mathbf{A} \mathbf{x}
\]

where \( \mathbf{x} = (x_1, \ldots, x_n)^T \) and \( \mathbf{A} \) is a symmetric matrix.

Suppose

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

Define

\[
a_{ij} = \frac{1}{2} (c_{ij} + c_{ji})
\]

Then \( \mathbf{A} = \mathbf{A}^T \) and

\[
f(x_1, \ldots, x_n) = \mathbf{x}^T \mathbf{A} \mathbf{x}
\]
Quadratic Form

**Definition**

Let $A$ be a real symmetric matrix. The **quadratic form** of $A$ is $x^T Ax$.

The quadratic form of a symmetric matrix is a quadratic function. For example, the quadratic form of

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

is

\[
x^T Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = ax_1^2 + 2bx_1x_2 + cx_2^2
\]
Positive Definite Matrix

**Definition**

Let $A$ be a real symmetric matrix. Then $A$ is **positive definite** if for any $x \neq 0$, the quadratic form of $A$ is positive

$$x^T Ax > 0$$
**Proposition**

Let $A$ be positive definite. Then the eigenvalues of $A$ are positive.

**Proof.**

Let $\lambda$ be an eigenvalue of $A$ and $s$ be a corresponding eigenvector. Then

$$As = \lambda s$$

It follows that

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

Hence

$$\lambda = \frac{s^T As}{s^T s} > 0$$
Proposition

Let $A$ be a real symmetric matrix. If every eigenvalue of $A$ is positive, then $A$ is positive definite.

Proof.

By the spectral theorem, we have $A = Q\Lambda Q^T$ where $Q$ is orthogonal. Consider the quadratic form of $A$.

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

For $x \neq 0$, we have $y \neq 0$ and thus $x^T A x = \sum_i \lambda_i y_i^2 > 0$. Hence $A$ is positive definite.
Proposition

Let $A$ be positive definite. Then every leading principal sub-matrix of $A$ has a positive determinant.

Consider $x^T = [x_k^T \ 0^T]$ with $x_k \in \mathbb{R}^k$. For $x_k \neq 0$

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

So $A_k$ is positive definite, the eigenvalues of $A_k$ are positive, and

$$|A_k| = \prod_{i=1}^{k} \lambda_{k,i} > 0$$

where $\lambda_{k,i}$ is an eigenvalue of $A_k$. 

Chen P  Positive Definite Matrix
Proposition

Let $A$ be real symmetric. If every leading principal sub-matrix of $A$ has a positive determinant, then $A$ has full positive pivots.

Proof.

$A$ is non-singular since $|A| \neq 0$. Let $A = LDU$ be the LDU decomposition of $A$. Then

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

So $A_k = L_k D_k U_k$. Let $d_1, \ldots, d_n$ be the pivots of $A$. Then

$$
d_1 = a_{11} > 0, \quad d_k = \frac{|D_k|}{|D_{k-1}|} = \frac{|A_k|}{|A_{k-1}|} > 0, \quad k = 2, \ldots, n
$$
Positive Pivots

Proposition

Let $A$ be real symmetric. If $A$ has full positive pivots, then $A$ is positive definite.

Proof.

Let $A = LDU$ be the LDU decomposition of $A$.

$$A = A^T \Rightarrow LDU = U^TDL^T \Rightarrow U = L^T$$

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 $x \neq 0$

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

Hence $A$ is positive definite.
Equivalent Statements for PDM

**Theorem**

Let $A$ be a real symmetric matrix. The following statements are equivalent.

1. $A$ is positive definite.
2. The eigenvalues of $A$ are positive.
3. The determinants of the leading principal sub-matrices of $A$ are positive.
4. The pivots of $A$ are positive.

What we have shown in the previous slides are

$$1 \iff 2$$

and

$$1 \Rightarrow 3 \Rightarrow 4 \Rightarrow 1$$
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

\[ \text{spectrum}(A) = \{2, 2 \pm \sqrt{2}\}, \quad |A_1| = 2, \quad |A_2| = 3, \quad |A_3| = 4 \]

\[ A = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3/2 & 0 \\ 0 & 4/3 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1/2 & 0 \\ 0 & 1 & -2/3 \\ 0 & 0 & 1 \end{bmatrix} \]
**Proposition**

Let $A$ be positive definite. Then the equation $x^T Ax = 1$ is an ellipsoid.

**Explanation.** By the spectral theorem, we have $A = Q \Lambda Q^T$. Thus $x^T Ax = 1$ can be converted to

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

Note that $Q = \{q_1, \ldots, q_n\}$ is an orthonormal basis, and $y = Q^T x$, or $y_i = q_i^T x$, is the representation of $x$ with $Q$. This is an ellipsoid with the axes of symmetry along $q_i$’s, with the intercepts of

$$y_i = \pm \left(\sqrt{\lambda_i}\right)^{-1}$$
Figure 6.2: The ellipse $x^TAx = 5u^2 + 8uv + 5v^2 = 1$ and its principal axes.
Let $A$ be a real symmetric matrix.

- $A$ is **negative definite** if $x^T Ax < 0$ for $x \neq 0$.
- $A$ is **positive semidefinite** if $x^T Ax \geq 0$ for any $x$.
- $A$ is **negative semidefinite** if $x^T Ax \leq 0$ for any $x$. 

**Chen P**

**Positive Definite Matrix**
Approximation and Extremal Points
Partial Derivatives

Definition

Let \( f(x_1, \ldots, x_n) \) be a multi-variate function. A \textbf{first-order partial derivative} of \( f \) is

\[
\frac{\partial f}{\partial x_i} = \frac{f_{x_i}}{
}
\]

A \textbf{second-order partial derivative} of \( f \) is

\[
f_{x_i x_j} = \frac{\partial f_{x_i}}{\partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}
\]

Note that

\[
f_{x_i x_j} = f_{x_j x_i}
\]
Gradient and Hessian

**Definition**

The **gradient** of $f(x_1, \ldots, x_n)$ is a vector of functions

$$\nabla f = \begin{bmatrix} f_{x_1} \\ \vdots \\ f_{x_n} \end{bmatrix}$$

The **Hessian** of $f(x_1, \ldots, x_n)$ is a matrix of functions

$$H = \begin{bmatrix} f_{x_1x_1} & \cdots & f_{x_1x_n} \\ \vdots & \ddots & \vdots \\ f_{x_nx_1} & \cdots & f_{x_nx_n} \end{bmatrix}, \quad H_{ij} = f_{x_ix_j}$$

**Chen P**

**Positive Definite Matrix**
The **first-order approximation** to $f(x)$ near a point $x_0$ is

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

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

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

**Definition**

Positive Definite Matrix
A point $x_0$ is a **stationary point** of $f(x)$ if

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

Let $x_0$ be a stationary point of $f(x)$. Then the second-order approximation to $f(x)$ near $x_0$ 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 \((0, 0)\) to

\[ f(x, y) = 2x^2 + 4xy + y^2 \]

\[ \nabla f = \begin{bmatrix} f_x \\ f_y \end{bmatrix} = \begin{bmatrix} 4x + 4y \\ 4x + 2y \end{bmatrix}, \quad H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 2 \end{bmatrix} \]

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

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

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

\[ F(x,y) = 7 + 2(x+y)^2 - y \sin y - x^3 \]

\[ \nabla F = \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} 4(x+y) - 3x^2 \\ 4(x+y) - \sin y - y \cos y \end{bmatrix} \]

\[ H = \begin{bmatrix} F_{xx} & F_{xy} \\ F_{yx} & F_{yy} \end{bmatrix} = \begin{bmatrix} 4 - 6x & 4 \\ 4 & 4 - 2 \cos y + y \sin y \end{bmatrix} \]

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

\[ F(x) \approx F(0) + \frac{1}{2}(x - 0)^T H(0)(x - 0) = 7 + 2x^2 + 4xy + y^2 \]
Local Minimum and Local Maximum

Definition

- A point \( x_0 \) is a **local minimum** of \( f(x) \) if
  \[
  f(x) \geq f(x_0)
  \]
  for any \( x \) in a neighborhood of \( x_0 \).

- A point \( x_0 \) is a **local maximum** of \( f(x) \) if
  \[
  f(x) \leq f(x_0)
  \]
  for any \( x \) in a neighborhood of \( x_0 \).
Theorem

Let $x_0$ be a stationary point of $f(x)$, and $H$ be the Hessian of $f(x)$ at $x_0$.

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

For example, $(0, 0)$ is a saddle point of $F(x, y)$. 
Figure 6.1: A bowl and a saddle: Definite $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and indefinite $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. 
Singular Value Decomposition
Proposition

Let $A$ be a real matrix. Then $A$ can be expressed by the eigenvalues of $(A^T A)$ and the eigenvectors of $(A^T A)$ and $(AA^T)$.

The above expression of $A$ is singular value decomposition, which is an extension of eigenvalue decomposition.
Proposition

Let $A$ be a real matrix.

- $(A^T A)$ and $(AA^T)$ are positive semi-definite.
- Their eigenvalues are real and non-negative.

\[
\begin{align*}
x^T (A^T A) x &= (Ax)^T (Ax) = \|Ax\|^2 \geq 0 \\
y^T (AA^T) y &= (A^T y)^T (A^T y) = \|A^T y\|^2 \geq 0 \\
(A^T A)^T &= A^T (A^T)^T = (A^T A)
\end{align*}
\]
Singular Values

Definition
Let $A$ be a real matrix. A **singular value** of $A$ is the square root of a positive eigenvalue of $(A^T A)$.

Let $\sigma$ be a singular value of $A$. Then $\sigma > 0$ and $\sigma^2$ is an eigenvalue of $(A^T A)$. That is, there exists $\mathbf{v} \neq \mathbf{0}$ such that

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

Let \( A \) be a real matrix and \( \sigma \) be a singular value of \( A \). A **right singular vector** of \( A \) is a vector \( v \neq 0 \) such that

\[
(A^T A) v = \sigma^2 v
\]

A **left singular vector** of \( A \) is a vector \( u \neq 0 \) such that

\[
(A A^T) u = \sigma^2 u
\]
Definition

Let $A$ be a real matrix and $\sigma$ be a singular value of $A$. The right singular space of $A$ corresponding to $\sigma$ is

$$ \mathbb{R}_\sigma(A) = \{ v | (A^T A) v = \sigma^2 v \} $$

The left singular space of $A$ corresponding to $\sigma$ is

$$ \mathbb{L}_\sigma(A) = \{ u | (A A^T) u = \sigma^2 u \} $$

The dimension of $\mathbb{R}_\sigma(A)$ is the multiplicity of the eigenvalue $\sigma^2$ of $A^T A$. 
Lemma

Let $A$ be a real matrix of rank $r$. There exists a linearly independent set containing $r$ right singular vectors of $A$.

Proof. Let $A$ be of order $m \times n$ with rank $r$. Note

$\left( A^T A \right) x = 0 \Rightarrow x^T A^T A x = 0 \Rightarrow A x = 0 \Rightarrow \left( A^T A \right) x = 0$

Thus $\dim \mathbb{N} \left( A^T A \right) = \dim \mathbb{N} (A) = n - r$, and the multiplicity of eigenvalue 0 of $\left( A^T A \right)$ is $(n - r)$. The sum of the multiplicities of the non-zero eigenvalues of $\left( A^T A \right)$ is

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

So there are $r$ linearly independent right singular vectors of $A$. 
Singular Value Decomposition

Theorem (singular value decomposition)

Let \( A \) be a real matrix of order \( m \times n \). Then

\[
A = U \Sigma V^T
\]

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 \( (AA^T) \) as columns, and \( V \) is an \( n \times n \) orthogonal matrix with the eigenvectors of \( (A^T A) \) as columns.
Construction of SVD

Let \( r \) be the rank of \( A \) and \( v_1 \ldots v_r \) be linearly independent right singular vectors of \( A \) with singular values \( \sigma_1 \ldots \sigma_r \). Define

\[
    u_i = \frac{Av_i}{\sigma_i} \quad \text{for} \quad i = 1, \ldots, r.
\]

Note

\[
    \left( AA^T \right) u_i = \frac{AA^T Av_i}{\sigma_i} = \frac{A\sigma_i^2 v_i}{\sigma_i} = \sigma_i^2 u_i
\]

so \( u_i \) is a left singular vector of \( A \). 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

\[
    U = \begin{bmatrix} u_1 & \ldots & u_m \end{bmatrix}, \quad V = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}
\]
We show $U^TAV = \Sigma$, which then implies $A = U\Sigma V^T$. For $j = 1 \ldots r$, we have $u_j = \frac{Av_j}{\sigma_j}$ or $Av_j = \sigma_ju_j$, so

$$(U^TAV)_{ij} = u_i^TAv_j = u_i^T(\sigma_ju_j) = \sigma_j\delta_{ij}, \quad i = 1 \ldots m$$

For $j = r + 1 \ldots n$, we have $(A^TA)v_j = 0$ or $Av_j = 0$, so

$$(U^TAV)_{ij} = u_i^TAv_j = 0, \quad i = 1 \ldots m$$

Combining the results, we get

$$U^TAV = \Sigma$$

Hence

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

Let $A$ be a real matrix with SVD $A = U \Sigma V^T$. Then the column vectors in $U$ form an eigenbasis of $(AA^T)$, as is easily seen by

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

Similarly, the column vectors in $V$ form an eigenbasis of $(A^T A)$

\[
A^T A = (V \Sigma^T U^T)(U \Sigma V^T) = V \begin{pmatrix} \Sigma^T \Sigma \end{pmatrix} V^T
\]
Let $A$ be a real matrix with SVD $A = U\Sigma V^T$. Then the right singular vectors of $A$ in $V$ form an orthonormal basis of $\mathbb{C}(A^T)$.

Let $A$ be of order $m \times n$ and rank $r$. $v_{r+1}, \ldots, v_n$ are eigenvectors of $(A^T A)$ with eigenvalue 0 so

$$(A^T A) v_j = 0 \quad v_j = 0$$

and they form a basis of $\mathbb{N}(A^T A) = \mathbb{N}(A)$. It follows that $v_1, \ldots, v_r$ form a basis of the orthogonal complement of $\mathbb{N}(A)$, i.e. $\mathbb{C}(A^T)$. 

Chen P

Positive Definite Matrix
Let $A$ be a real matrix with SVD $A = U\Sigma V^T$. Then the left singular vectors of $A$ in $U$ form an orthonormal basis of $\mathbb{C}(A)$.

Let $A$ be of order $m \times n$ and rank $r$. The first $r$ columns of $AV = U\Sigma$ means

$$Av_i = \sigma_i u_i$$

Thus $u_1, \ldots, u_r$ are vectors in $\mathbb{C}(A)$. Since they are linearly independent, they form a basis of $\mathbb{C}(A)$. 
Example

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

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

\[
\begin{align*}
v_1 &= \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad v_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
\end{align*}
\]

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

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

So

\[
A = U \Sigma V^T = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix}
\]
Let \( A \) be a real matrix of rank \( r \). Then \( A \) can be expressed as 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*
Definition

The least-squares solution to $Ax = b$ with the minimum length is

$$x^+ = A^+ b$$

$A^+$ is called the **pseudo-inverse** of $A$.

Let $A$ be a real matrix with SVD $A = U\Sigma V^T$. Then

$$A^+ = V \Sigma^+ U^T$$
Figure 3.4: The true action $Ax = A(x_{\text{row}} + x_{\text{null}})$ of any $m$ by $n$ matrix.

Figure 6.3: The pseudoinverse $A^+$ inverts $A$ where it can on the column space.
Minimum Principles *
Let $A$ be positive definite. Then $x_0$ achieves the minimum of

$$P(x) = \frac{1}{2} x^T A x - x^T b$$

if and only if $Ax_0 = b$.

Note

$$\nabla \left( x^T A x \right) = 2Ax, \quad \nabla \left( x^T b \right) = b$$

so

$$\nabla P(x) = Ax - b$$
Figure 6.4: The graph of a positive quadratic $P(x)$ is a parabolic bowl.
Over-determined System

Let $\mathcal{L} : Ax = b$ be an over-determined system of linear equations. The sum of squared errors as a function of $x$ is

$$E^2(x) = \|Ax - b\|^2$$

$$= (Ax - b)^T (Ax - b)$$

$$= x^T A^T A x - 2x^T A^T b + b^T b$$

An $x_0$ that achieves the minimum of $E^2(x)$ satisfies

$$\nabla E^2(x) \bigg|_{x=x_0} = 0$$

That is

$$A^T A x_0 = A^T b$$
Definition

Let $A$ be a real symmetric matrix. The **Rayleigh quotient** of $A$ is

$$R(x) = \frac{x^T Ax}{x^T x}$$

Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of $A$. Then

$$\lambda_1 \leq R(x) \leq \lambda_n$$

Furthermore, the lower bound $\lambda_1$ is achieved by $s_1 \in \mathbb{E}_{\lambda_1}$, and the upper bound $\lambda_n$ is achieved by $s_n \in \mathbb{E}_{\lambda_n}$.
When $x = e_i$, the Rayleigh quotient is $R(x) = a_{ii}$. So

$$\lambda_1 \leq a_{ii} \leq \lambda_n$$

which implies

$$\frac{1}{\sqrt{\lambda_n}} \leq \frac{1}{\sqrt{a_{ii}}} \leq \frac{1}{\sqrt{\lambda_1}}$$

**Figure 6.6:** The farthest $x = x_1/\sqrt{\lambda_1}$ and the closest $x = x_n/\sqrt{\lambda_n}$ both give $x^T Ax = x^T \lambda x = 1$. These are the major axes of the ellipse.
The minimum of \( R(x) \) subject to \( x^T s_1 = 0 \) is \( \lambda_2 \). For any \( v \), the minimum of \( R(x) \) subject to \( x^T v = 0 \) cannot be above \( \lambda_2 \). That is

\[
\lambda_2 \geq \min_{x^T v = 0} R(x)
\]

Thus, the maximin principle for \( \lambda_2 \) is

\[
\lambda_2 = \max_v \left[ \min_{x^T v = 0} R(x) \right]
\]

More generally, let \( S_j \) be a subspace of dimension \( j \), then

\[
\lambda_{j+1} = \max_{S_j} \left[ \min_{x \perp S_j} R(x) \right], \quad \lambda_{n-j} = \min_{S_j} \left[ \max_{x \perp S_j} R(x) \right]
\]