Linear Algebra

Lecture 14 - Quadratic Forms

Oskar Kędzierski

22 January 2024

Quadratic Form

Definition

A function $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a **quadratic form** if

$$Q((x_1, \ldots, x_n)) = a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij}x_ix_j,$$

that is, it is a function given by a homogeneous polynomial of degree 2 in variables $x_1, \ldots, x_n$.

Quadratic Form

Definition

A function $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a **quadratic form** if

$$Q((x_1, \ldots, x_n)) = a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij}x_ix_j,$$

that is, it is a function given by a homogeneous polynomial of degree 2 in variables $x_1, \ldots, x_n$.

Examples

$$Q((x_1, x_2)) = x_1^2 - x_2^2$$

Quadratic Form

Definition

A function $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is called a **quadratic form** if

$$Q((x_1, \ldots, x_n)) = a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij}x_i x_j,$$

that is, it is a function given by a homogeneous polynomial of degree 2 in variables $x_1, \ldots, x_n$.

Examples

$$Q((x_1, x_2)) = x_1^2 - x_2^2$$

$$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3$$

Symmetric Matrix

Recall

Definition

Matrix $A = [a_{ij}] \in M(n \times n; \mathbb{R})$ is called **symmetric** if $A^{\intercal} = A$, i.e. $a_{ij} = a_{ji}$ for $i, j = 1, \dots, n$.

Symmetric Matrix

Recall

Definition

Matrix $A = [a_{ij}] \in M(n \times n; \mathbb{R})$ is called **symmetric** if $A^\mathsf{T} = A$, i.e. $a_{ij} = a_{ji}$ for $i, j = 1, \ldots, n$.

Example

$$\text{matrix} \begin{bmatrix} 0 & 2 & 5 \\ 2 & 4 & -3 \\ 5 & -3 & 1 \end{bmatrix} \text{ is symmetric}$$

Symmetric Matrix

Recall

Definition

Matrix $A = [a_{ij}] \in M(n \times n; \mathbb{R})$ is called **symmetric** if $A^{\intercal} = A$, i.e. $a_{ij} = a_{ji}$ for $i, j = 1, \ldots, n$.

Example

$$\text{matrix} \begin{bmatrix} 0 & 2 & 5 \\ 2 & 4 & -3 \\ 5 & -3 & 1 \end{bmatrix} \text{ is symmetric}$$

$$\text{matrix} \begin{bmatrix} 0 & 2 & 6 \\ 2 & 4 & -3 \\ 5 & -3 & 1 \end{bmatrix} \text{ is not symmetric}$$

Matrix of a Quadratic Form

Definition

Let $Q((x_1, \ldots, x_n)) = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij} x_i x_j$ be a quadratic form. The matrix of the quadratic form Q is a symmetric matrix $M = [b_{ij}] \in M(n \times n; \mathbb{R})$ such that $b_{ii} = a_{ii}$ and $b_{ij} = b_{ji} = \frac{1}{2} a_{ij}$ for $1 \leqslant i < j \leqslant n$.

Matrix of a Quadratic Form

Definition

Let $Q((x_1, \ldots, x_n)) = \sum_{i=1}^{n} a_{ii} x_i^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij} x_i x_j$ be a quadratic form. The matrix of the quadratic form Q is a symmetric matrix $M = [b_{ij}] \in M(n \times n; \mathbb{R})$ such that $b_{ii} = a_{ii}$ and $b_{ij} = b_{ji} = \frac{1}{2} a_{ij}$ for $1 \leqslant i < j \leqslant n$.

Example

The matrix of the form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is $M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Matrix of a Quadratic Form

Definition

Let $Q((x_1, \ldots, x_n)) = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{1 \leqslant i < j \leqslant n} a_{ij}x_ix_j$ be a quadratic form. The matrix of the quadratic form Q is a symmetric matrix $M = [b_{ij}] \in M(n \times n; \mathbb{R})$ such that $b_{ii} = a_{ii}$ and $b_{ij} = b_{ji} = \frac{1}{2}a_{ij}$ for $1 \leqslant i < j \leqslant n$.

Example

The matrix of the form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is $M = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

The matrix of the form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2$ $-4x_1x_3 + 8x_2x_3$ is

$$M = \begin{bmatrix} 1 & 1 & -2 \\ 1 & 2 & 4 \\ -2 & 4 & 5 \end{bmatrix}$$

Matrix of a Quadratic Form (continued)

Proposition

Let M be a matrix of the quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$. Then

$$Q((x_1, \ldots, x_n)) = x^\mathsf{T} M x,$$

where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

Matrix of a Quadratic Form (continued)

Proposition

Let M be a matrix of the quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$. Then

$$Q((x_1, \ldots, x_n)) = x^\intercal M x,$$

where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$.

Proof.

Entries of matrix M in the i-th row get multiplied by x_i and elements in the j-th column get multiplied by x_j. For every $i \neq j$ the monomial $x_i x_j$ comes from the entry in the i-th row, j-th column and from the entry in the j-th row, i-th column. □

Matrix of a Quadratic Form (continued)

Formal explanation

$$Q((x_1,\ldots,x_n)) = x^\intercal M x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \sum_{s=1}^n b_{1s}x_s \\ \sum_{s=1}^n b_{2s}x_s \\ \vdots \\ \sum_{s=1}^n b_{ns}x_s \end{bmatrix} =$$

$$= x_1 \sum_{s=1}^n b_{1s}x_s + x_2 \sum_{s=1}^n b_{2s}x_s + \ldots + x_n \sum_{s=1}^n b_{ns}x_s =$$

$$= \sum_{i,j=1}^n b_{ij}x_i x_j.$$

Positive/Negative Definite Quadratic Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive definite** if $Q(x) > 0$ (resp. $x^\mathsf{T} M x > 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.

Positive/Negative Definite Quadratic Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive definite** if $Q(x) > 0$ (resp. $x^\intercal Mx > 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.
Quadratic form Q (resp. symmetric matrix M) is **negative definite** if $Q(x) < 0$ (resp. $x^\intercal Mx < 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.

Positive/Negative Definite Quadratic Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive definite** if $Q(x) > 0$ (resp. $x^\mathsf{T} M x > 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.
Quadratic form Q (resp. symmetric matrix M) is **negative definite** if $Q(x) < 0$ (resp. $x^\mathsf{T} M x < 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.

Example

The quadratic form $\|\cdot\|^2\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is positive definite since $\|x\|^2 = x_1^2 + \ldots + x_n^2 > 0$ for $x \neq \mathbf{0}$.

Positive/Negative Definite Quadratic Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive definite** if $Q(x) > 0$ (resp. $x^\intercal M x > 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.
Quadratic form Q (resp. symmetric matrix M) is **negative definite** if $Q(x) < 0$ (resp. $x^\intercal M x < 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.

Example

The quadratic form $\|\cdot\|^2\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is positive definite since $\|x\|^2 = x_1^2 + \ldots + x_n^2 > 0$ for $x \neq \mathbf{0}$.
The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is not positive definite since $Q((0,1)) = -1 < 0$. It is not negative definite since $Q((1,0)) = 1 > 0$.

Positive/Negative Definite Quadratic Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive definite** if $Q(x) > 0$ (resp. $x^\mathsf{T} Mx > 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.
Quadratic form Q (resp. symmetric matrix M) is **negative definite** if $Q(x) < 0$ (resp. $x^\mathsf{T} Mx < 0$) for any $x \in \mathbb{R}^n,\ x \neq \mathbf{0}$.

Example

The quadratic form $\|\cdot\|^2\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is positive definite since $\|x\|^2 = x_1^2 + \ldots + x_n^2 > 0$ for $x \neq \mathbf{0}$.
The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is not positive definite since $Q((0,1)) = -1 < 0$. It is not negative definite since $Q((1,0)) = 1 > 0$.
The quadratic form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 -2x_1x_3 + 2x_2x_3 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$ is not positive definite since $Q((3,-2,1)) = 0$. It is not negative definite.

Recall

$$(a_1+a_2+\ldots+a_n)^2 = a_1^2+a_2^2+\ldots+a_n^2+2a_1a_2+2a_1a_3+\ldots+2a_1a_n+$$

$$+2a_2a_3+2a_2a_4+\ldots+2a_2a_n+2a_3a_4+\ldots+2a_3a_n+\ldots\ldots+2a_{n-1}a_n$$

Recall

$$(a_1 + a_2 + \ldots + a_n)^2 = a_1^2 + a_2^2 + \ldots + a_n^2 + 2a_1 a_2 + 2a_1 a_3 + \ldots + 2a_1 a_n +$$

$$+ 2a_2 a_3 + 2a_2 a_4 + \ldots + 2a_2 a_n + 2a_3 a_4 + \ldots + 2a_3 a_n + \ldots \ldots + 2a_{n-1} a_n$$

For example

$$(x_1 - 3x_2 + 2x_3)^2 =$$

Recall

$$(a_1+a_2+\ldots+a_n)^2 = a_1^2+a_2^2+\ldots+a_n^2+2a_1a_2+2a_1a_3+\ldots+2a_1a_n+$$

$$+2a_2a_3+2a_2a_4+\ldots+2a_2a_n+2a_3a_4+\ldots+2a_3a_n+\ldots\ldots+2a_{n-1}a_n$$

For example

$$(x_1-3x_2+2x_3)^2 =$$

$$= x_1^2+(-3)^2x_2^2+2^2x_3^2+2\cdot(-3)x_1x_2+2\cdot 2x_1x_3+2\cdot(-3)\cdot 2x_2x_3 =$$

$$= x_1^2+9x_2^2+4x_3^2-6x_1x_2+4x_1x_3-12x_2x_3$$

Recall

$$(a_1+a_2+\ldots+a_n)^2 = a_1^2+a_2^2+\ldots+a_n^2+2a_1a_2+2a_1a_3+\ldots+2a_1a_n+$$

$$+2a_2a_3+2a_2a_4+\ldots+2a_2a_n+2a_3a_4+\ldots+2a_3a_n+\ldots\ldots+2a_{n-1}a_n$$

For example

$$(x_1-3x_2+2x_3)^2 =$$

$$= x_1^2+(-3)^2x_2^2+2^2x_3^2+2\cdot(-3)x_1x_2+2\cdot 2x_1x_3+2\cdot(-3)\cdot 2x_2x_3 =$$
$$= x_1^2+9x_2^2+4x_3^2-6x_1x_2+4x_1x_3-12x_2x_3$$

Proposition

A quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ can expressed (possibly after a change of coordinates) as $Q((x_1,\ldots,x_n)) = \pm l_1^2 \pm l_2^2 \pm \ldots \pm l_n^2$ where $l_1,\ldots,l_n$ are linear functions such that $l_i,\ldots,l_n$ do not depend on the variables $x_1,\ldots,x_{i-1}$ for $i=2,\ldots,n$.

Recall

$$(a_1+a_2+\ldots+a_n)^2 = a_1^2+a_2^2+\ldots+a_n^2+2a_1a_2+2a_1a_3+\ldots+2a_1a_n+$$

$$+2a_2a_3+2a_2a_4+\ldots+2a_2a_n+2a_3a_4+\ldots+2a_3a_n+\ldots\ldots+2a_{n-1}a_n$$

For example

$$(x_1-3x_2+2x_3)^2 =$$

$$= x_1^2+(-3)^2x_2^2+2^2x_3^2+2\cdot(-3)x_1x_2+2\cdot 2x_1x_3+2\cdot(-3)\cdot 2x_2x_3 =$$
$$= x_1^2+9x_2^2+4x_3^2-6x_1x_2+4x_1x_3-12x_2x_3$$

Proposition

A quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ can expressed (possibly after a change of coordinates) as $Q((x_1,\ldots,x_n)) = \pm l_1^2 \pm l_2^2 \pm \ldots \pm l_n^2$ where $l_1,\ldots,l_n$ are linear functions such that $l_i,\ldots,l_n$ do not depend on the variables $x_1,\ldots,x_{i-1}$ for $i=2,\ldots,n$.

Proof.

(sketch) Use the above formula. $\square$

Example

$$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3 =$$
$$(x_1 + x_2 - x_3)^2 + x_2^2 + 4x_2x_3 + 4x_3^2 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$$

Example

$$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3 = (x_1 + x_2 - x_3)^2 + x_2^2 + 4x_2x_3 + 4x_3^2 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$$

$$Q((x_1, x_2, x_3)) = x_1^2 - x_2^2 + x_3^2 + 2x_1x_2 - 4x_1x_3 = (x_1+x_2-2x_3)^2-2x_2^2+4x_2x_3-3x_3^2 = (x_1+x_2-2x_3)^2-2(x_2-x_3)^2-x_3^2$$

Example

$$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3 =$$
$$(x_1 + x_2 - x_3)^2 + x_2^2 + 4x_2x_3 + 4x_3^2 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$$

$$Q((x_1, x_2, x_3)) = x_1^2 - x_2^2 + x_3^2 + 2x_1x_2 - 4x_1x_3 =$$
$$(x_1+x_2-2x_3)^2 - 2x_2^2 + 4x_2x_3 - 3x_3^2 = (x_1+x_2-2x_3)^2 - 2(x_2-x_3)^2 - x_3^2$$

What to do if there is no square? Do the following substitution:

$$Q((x_1, x_2, x_3)) = x_1x_2 + 2x_1x_3 = \left\{\begin{smallmatrix} x_1 = y_1 - y_2 \\ x_2 = y_1 + y_2 \\ x_3 = y_3 \end{smallmatrix}\right\} =$$
$$(y_1 - y_2)(y_1 + y_2) + 2(y_1 - y_2)y_3 = y_1^2 - y_2^2 + 2y_1y_3 - 2y_2y_3 =$$
$$(y_1 + y_3)^2 - y_2^2 - y_3^2 - 2y_2y_3 = (y_1 + y_3)^2 - (y_2 + y_3)^2$$

Example (continued)

$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 \; -2x_1x_3 + 2x_2x_3 =$
$(x_1 + x_2 - x_3)^2 + x_2^2 + 4x_2x_3 + 4x_3^2 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$
Let

$$\begin{cases} y_1 & = & x_1 + x_2 - x_3 \\ y_2 & = & \phantom{x_1 +{}} x_2 + 2x_3 \\ y_3 & = & \phantom{x_1 + x_2 + 2{}} x_3 \end{cases},$$

then $Q((y_1, y_2, y_3)) = y_1^2 + y_2^2$, where

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = P \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \text{for} \quad P = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

In particular

$$y^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} y = (Px)^T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} Px = x^T \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix} x.$$

Sylvester's Criterion

Proposition

Let $M \in M(n \times n; \mathbb{R})$ *be a symmetric matrix. Let* W_i *denote the determinant of the upper-left* i*-by-*i *submatrix of* M*. Matrix* M *is positive definite if and only if* $W_i > 0$ *for* $i = 1, \ldots, n$*.*

Sylvester's Criterion

Proposition

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric matrix. Let W_i denote the determinant of the upper-left i-by-i submatrix of M. Matrix M is positive definite if and only if $W_i > 0$ for $i = 1, \ldots, n$.

Proof.

Omitted. $\square$

Remark

The determinants W_i are sometimes called **leading principal minors**.

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

and compute its leading principal minors

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

$W_3 = \det \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix} \overset{\substack{c_1+c_3 \\ c_2+c_3}}{=} \det \begin{bmatrix} 0 & 0 & -1 \\ 2 & 3 & 1 \\ 5 & 7 & 6 \end{bmatrix} = 1 > 0.$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

$W_3 = \det \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 6 \end{bmatrix} \overset{\substack{c_1+c_3 \\ c_2+c_3}}{=} \det \begin{bmatrix} 0 & 0 & -1 \\ 2 & 3 & 1 \\ 5 & 7 & 6 \end{bmatrix} = 1 > 0.$

By Sylverster's criterion the quadratic form $x_1^2 + 2x_2^2 + 6x_3^2 + 2x_1x_2$ $-2x_1x_3 + 2x_2x_3$ is positive definite.

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

$$W_3 = \det \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix} \overset{\substack{c_1+c_3 \\ c_2+c_3}}{=} \det \begin{bmatrix} 0 & 0 & -1 \\ 2 & 3 & 1 \\ 4 & 6 & 5 \end{bmatrix} = 0 \not> 0.$$

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors

$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0,$

$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0,$

$W_3 = \det \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix} \overset{\substack{c_1+c_3 \\ c_2+c_3}}{=} \det \begin{bmatrix} 0 & 0 & -1 \\ 2 & 3 & 1 \\ 4 & 6 & 5 \end{bmatrix} = 0 \not> 0.$

By Sylverster's criterion the quadratic form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3$ is not positive definite.

Another Example

Consider the symmetric matrix

$$M = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} 1 \end{bmatrix} = 1 > 0$,
$W_2 = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 > 0$,
$W_3 = \det \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ -1 & 1 & 5 \end{bmatrix} \overset{\substack{c_1+c_3 \\ c_2+c_3}}{=} \det \begin{bmatrix} 0 & 0 & -1 \\ 2 & 3 & 1 \\ 4 & 6 & 5 \end{bmatrix} = 0 \not> 0.$
By Sylverster's criterion the quadratic form
$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3$ is not positive definite. In fact, $Q((3, -2, 1)) = 0$.

Sylvester's Criterion (continued)

A quadratic form Q is positive definite if and only if $-Q$ is negative definite.

Sylvester's Criterion (continued)

A quadratic form Q is positive definite if and only if $-Q$ is negative definite.

Proposition

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric matrix. Let W_i denote the determinant of the upper-left i-by-i submatrix of M. Matrix M is negative definite if and only if

Sylvester's Criterion (continued)

A quadratic form Q is positive definite if and only if $-Q$ is negative definite.

Proposition

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric matrix. Let W_i denote the determinant of the upper-left i-by-i submatrix of M. Matrix M is negative definite if and only if

$$W_i < 0 \text{ for odd } i,$$

$$W_i > 0 \text{ for even } i,$$

for $i = 1, \ldots, n$.

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}$$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}$$

and compute its leading principal minors

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} -1 \end{bmatrix} = -1 < 0,$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} -1 \end{bmatrix} = -1 < 0,$
$W_2 = \det \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix} = 1 > 0,$

Example

Consider the symmetric matrix

$$M = \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix}$$

and compute its leading principal minors
$W_1 = \det \begin{bmatrix} -1 \end{bmatrix} = -1 < 0,$
$W_2 = \det \begin{bmatrix} -1 & -1 \\ -1 & -2 \end{bmatrix} = 1 > 0,$
The quadratic form $-x_1^2 - 2x_1x_2 - 2x_2^2 = -(x_1 + x_2)^2 - x_2^2$ is negative definite.

Sylvester's Criterion – Warning

It crucial that matrix A is **symmetric**. For example, let

$$M = \begin{bmatrix} -1 & -3 \\ 1 & 2 \end{bmatrix}.$$

Then

$$\varepsilon_1^\intercal M \varepsilon_1 = -1, \quad \varepsilon_2^\intercal M \varepsilon_2 = 2,$$

hence matrix M is indefinite, however

$$W_1 = -1, \quad W_2 = 1.$$

Postive/Negative Semidefinite Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive semidefinite** if $Q(x) \geqslant 0$ (resp. $x^\mathsf{T} M x \geqslant 0$) for any $x \in \mathbb{R}^n$.

Postive/Negative Semidefinite Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive semidefinite** if $Q(x) \geqslant 0$ (resp. $x^\intercal Mx \geqslant 0$) for any $x \in \mathbb{R}^n$.
Quadratic form Q (resp. symmetric matrix M) is **negative semidefinite** if $Q(x) \leqslant 0$ (resp. $x^\intercal Mx \leqslant 0$) for any $x \in \mathbb{R}^n$.

Postive/Negative Semidefinite Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive semidefinite** if $Q(x) \geqslant 0$ (resp. $x^\mathsf{T} M x \geqslant 0$) for any $x \in \mathbb{R}^n$.
Quadratic form Q (resp. symmetric matrix M) is **negative semidefinite** if $Q(x) \leqslant 0$ (resp. $x^\mathsf{T} M x \leqslant 0$) for any $x \in \mathbb{R}^n$.
Quadratic form Q (resp. symmetric matrix M) is **indefinite** if there exist $x, y \in \mathbb{R}^n$ such that $Q(x) > 0, Q(y) < 0$ (resp. $x^\mathsf{T} M x > 0, y^\mathsf{T} M y < 0$).

Postive/Negative Semidefinite Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive semidefinite** if $Q(x) \geqslant 0$ (resp. $x^\intercal M x \geqslant 0$) for any $x \in \mathbb{R}^n$.

Quadratic form Q (resp. symmetric matrix M) is **negative semidefinite** if $Q(x) \leqslant 0$ (resp. $x^\intercal M x \leqslant 0$) for any $x \in \mathbb{R}^n$.

Quadratic form Q (resp. symmetric matrix M) is **indefinite** if there exist $x, y \in \mathbb{R}^n$ such that $Q(x) > 0, Q(y) < 0$ (resp. $x^\intercal M x > 0, y^\intercal M y < 0$).

Remark

A positive (resp. negative) definite quadratic form is positive (resp. negative) semidefinite.

Postive/Negative Semidefinite Form

Definition

Quadratic form $Q\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ (resp. symmetric matrix $M \in M(n \times n; \mathbb{R})$) is **positive semidefinite** if $Q(x) \geqslant 0$ (resp. $x^\intercal Mx \geqslant 0$) for any $x \in \mathbb{R}^n$.
Quadratic form Q (resp. symmetric matrix M) is **negative semidefinite** if $Q(x) \leqslant 0$ (resp. $x^\intercal Mx \leqslant 0$) for any $x \in \mathbb{R}^n$.
Quadratic form Q (resp. symmetric matrix M) is **indefinite** if there exist $x, y \in \mathbb{R}^n$ such that $Q(x) > 0, Q(y) < 0$ (resp. $x^\intercal Mx > 0, y^\intercal My < 0$).

Remark

A positive (resp. negative) definite quadratic form is positive (resp. negative) semidefinite. A quadratic form is indefinite if and only if it is not positive semidefinite and it is not negative semidefinite.

Examples

The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is indefinite since $Q((1, 0)) > 0$ and $Q((0, 1)) < 0$.

Examples

The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is indefinite since $Q((1,0)) > 0$ and $Q((0,1)) < 0$.

The quadratic form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2$ $-2x_1x_3 + 2x_2x_3 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$ is positive semidefinite. It is not positive definite.

Examples

The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is indefinite since $Q((1,0)) > 0$ and $Q((0,1)) < 0$.

The quadratic form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2$ $-2x_1x_3 + 2x_2x_3 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$ is positive semidefinite. It is not positive definite.

The quadratic form
$Q((x_1, x_2)) = -x_1^2 - 2x_1x_2 - 2x_2^2 = -(x_1 + x_2)^2 - x_2^2$ is negative definite.

Examples

The quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is indefinite since $Q((1, 0)) > 0$ and $Q((0, 1)) < 0$.

The quadratic form $Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 5x_3^2 + 2x_1x_2 - 2x_1x_3 + 2x_2x_3 = (x_1 + x_2 - x_3)^2 + (x_2 + 2x_3)^2$ is positive semidefinite. It is not positive definite.

The quadratic form $Q((x_1, x_2)) = -x_1^2 - 2x_1x_2 - 2x_2^2 = -(x_1 + x_2)^2 - x_2^2$ is negative definite.

The quadratic form $Q((x_1, x_2, x_3)) = -x_1^2 - 2x_1x_2 - 2x_2^2 = -(x_1 + x_2)^2 - x_2^2$ is not negative definite since $Q((0, 0, 1)) = 0$. It is negative semidefinite.

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$.

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite.

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite. Its matrix is

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

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite. Its matrix is

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

Compute the leading principal minors

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite. Its matrix is

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

Compute the leading principal minors
$W_1 = \det \begin{bmatrix} 0 \end{bmatrix} = 0 \geqslant 0,$

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite. Its matrix is

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

Compute the leading principal minors
$W_1 = \det \begin{bmatrix} 0 \end{bmatrix} = 0 \geqslant 0,$
$W_2 = \det \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} = 0 \geqslant 0,$

Warning

Consider the quadratic form $Q((x_1, x_2)) = -x_2^2$. It is negative semidefinite. Its matrix is

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

Compute the leading principal minors
$W_1 = \det \begin{bmatrix} 0 \end{bmatrix} = 0 \geqslant 0$,
$W_2 = \det \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} = 0 \geqslant 0$,
This shows **there is no** direct analogue of Sylvester's criterion for positive/negative semidefinite matrices.

Warning (continued)

Proposition

Let $M \in M(n \times n; \mathbb{R})$ *be a symmetric square matrix. Then matrix* M *is positive semidefinite if and only if for any* $J \subset \{1, \dots, n\}, J \neq \varnothing$

$$\det M_{J;J} \geqslant 0,$$

that is **all** *principal minors are non–negative.*

Warning (continued)

Proposition

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric square matrix. Then matrix M is positive semidefinite if and only if for any $J \subset \{1, \ldots, n\}, J \neq \varnothing$

$$\det M_{J;J} \geqslant 0,$$

that is **all** *principal minors are non–negative.*

Proof.

The proof uses spectral theorem and eigenvalue criterion.
($\Longrightarrow$) The restriction of M to the subspace $\operatorname{lin}(\{\varepsilon_i \mid i \in J\})$ is positive semidefinite and has matrix equal to $M_{J;J}$. Since $M_{J;J}$ is symmetric and positive semidefinite, by the eigenvalue criterion $\det M_{J;J}$ is equal to the product of eignevalues hence it is non–negative.

Warning (continued)

Proof.

($\Longleftarrow$) Proof by induction on n. Let $Q(x) = x^\intercal Mx$ and let $u_1, \ldots, u_n \in \mathbb{R}^n$ be an orthonormal basis such that $u_i^\intercal Mu_j = 0$ for $i \neq j$. Moreover assume, by rearranging u_i's, that $Q(u_1) \leqslant Q(u_2) \ldots \leqslant Q(u_n)$. It is enough to prove $Q(u_1) \geqslant 0$. If $u_1 \cdot \varepsilon_k = 0$ (i.e. the $k-$th component of u_1 vanishes) for some $k \in \{1, \ldots, n\}$ then $u_1 \in \operatorname{lin}(\varepsilon_1, \ldots, \varepsilon_{k-1}, \varepsilon_{k+1}, \ldots, \varepsilon_n)$ and $Q(u_1) \geqslant 0$, by the inductive assumption. Assume $u_1 \cdot \varepsilon_k \neq 0$ for any $k = 1, \ldots, n$.

Warning (continued)

Proof.

For $i \geqslant 2$ and some $k = 1, \ldots, n$ consider vector

$$v = (u_i \cdot \varepsilon_k)u_1 - (u_1 \cdot \varepsilon_k)u_i.$$

Since $v \cdot \varepsilon_k = 0$ by the inductive assumption

$$Q(v) = (u_i \cdot \varepsilon_k)^2 Q(u_1) + (u_1 \cdot \varepsilon_k)^2 Q(u_i) \geqslant 0.$$

If some $Q(u_i) = 0$ with k such that $u_i \cdot \varepsilon_k \neq 0$ (u_i needs to have a non–zero coordinate) then $Q(u_1) \geqslant 0$. Assume now $Q(u_2), \ldots, Q(u_n) > 0$. Then, by choosing $J = \{1, \ldots, n\}$ and using the eigenvalue criterion

$$Q(u_1)Q(u_2) \cdot \ldots \cdot Q(u_n) \geqslant 0,$$

that is $Q(u_1) \geqslant 0$. $\square$

Warning (continued)

Remark

Note that for a $n \times n$ *matrix there are* $2^n - 1$ *conditions to check, making this criterion impractical.*

Warning (continued)

Corollary

Let $A \in M(n \times n; \mathbb{R})$ be a symmetric square matrix. Then matrix A is negative semidefinite if and only if for any $J \subset \{1, \ldots, n\}, J \neq \varnothing$

$$\det A_{J;J} \geqslant 0, \quad \textit{when } \#J \textit{ is even},$$

$$\det A_{J;J} \leqslant 0, \quad \textit{when } \#J \textit{ is odd},$$

that is principal minors of M of even order are non–negative and principal minors of M of odd order are non–positive.

Proof.

Matrix M is positive semidefinite if and only if matrix $-M$ is negative semidefinite. $\square$

Warning (continued)

In particular, for

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

we have

$$\det M_{1;1} = \det \begin{bmatrix} 0 \end{bmatrix} = 0, \quad \det M_{2;2} = \det \begin{bmatrix} -1 \end{bmatrix} = -1 < 0,$$

$$\det M_{1,2;1,2} = \det M = 0,$$

therefore matrix M is not positive semidefinite. In fact, it is negative semidefinite.

Positive Definite Quadratic Form

Proposition

Let $A \in M(m \times n; \mathbb{R})$ *be a matrix. Then matrix* $M = A^{\intercal}A \in M(n \times n; \mathbb{R})$ *is symmetric and positive semidefinite. Moreover, the matrix* $A^{\intercal}A$ *is positive definite if and only if* $r(A) = n$ *(i.e. columns of matrix* A *are linearly independent).*

Positive Definite Quadratic Form

Proposition

Let $A \in M(m \times n; \mathbb{R})$ *be a matrix. Then matrix* $M = A^{\intercal}A \in M(n \times n; \mathbb{R})$ *is symmetric and positive semidefinite. Moreover, the matrix* $A^{\intercal}A$ *is positive definite if and only if* $r(A) = n$ *(i.e. columns of matrix* A *are linearly independent).*

Proof.

For any $x \in \mathbb{R}^n$

$$x^{\intercal}(A^{\intercal}A)x = (Ax)^{\intercal}(Ax) = \|Ax\|^2 \geqslant 0.$$

$(\Leftarrow)$ If $r(A) = n$ (i.e. the linear transformation $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $A = M(\varphi)_{st}^{st}$ is injective by the rank–nullity theorem) then

$$\|Ax\| = 0 \iff Ax = \mathbf{0} \iff x = \mathbf{0}.$$

Positive Definite Quadratic Form

Proposition

Let $A \in M(m \times n; \mathbb{R})$ be a matrix. Then matrix $M = A^\intercal A \in M(n \times n; \mathbb{R})$ is symmetric and positive semidefinite. Moreover, the matrix $A^\intercal A$ is positive definite if and only if $r(A) = n$ (i.e. columns of matrix A are linearly independent).

Proof.

For any $x \in \mathbb{R}^n$

$$x^\intercal (A^\intercal A) x = (Ax)^\intercal (Ax) = \|Ax\|^2 \geqslant 0.$$

($\Leftarrow$) If $r(A) = n$ (i.e. the linear transformation $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $A = M(\varphi)_{st}^{st}$ is injective by the rank–nullity theorem) then

$$\|Ax\| = 0 \iff Ax = \mathbf{0} \iff x = \mathbf{0}.$$

($\Rightarrow$) if $Ax = \mathbf{0} \Rightarrow x = \mathbf{0}$ then $\ker \varphi = \{\mathbf{0}\}$ which, by the rank–nullity theorem, gives $r(A) = n$. $\square$

Example

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} \in M(3 \times 2; \mathbb{R})$ where $r(A) = 2$. The matrix

$$A^\intercal A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 7 & 14 \end{bmatrix},$$

is positive definite and the matrix

$$(A^\intercal)^\intercal A^\intercal = AA^\intercal = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 4 & 7 \\ 4 & 5 & 5 \\ 7 & 5 & 10 \end{bmatrix},$$

is positive semidefinite and it is not positive definite (this will be justified later).

Eigenvalues and Positivity

Theorem (Spectral Theorem)

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ *is diagonalizable by an orthonormal eigenbasis.*

Eigenvalues and Positivity

Theorem (Spectral Theorem)

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ *is diagonalizable by an orthonormal eigenbasis.*

In particular, the characteristic polynomial $w_M(\lambda) = \det(M - \lambda I)$ has n real roots (=eigenvalues) counted with multiplicities .

Eigenvalues and Positivity

Theorem (Spectral Theorem)

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ is diagonalizable by an orthonormal eigenbasis.

In particular, the characteristic polynomial $w_M(\lambda) = \det(M - \lambda I)$ has n real roots (=eigenvalues) counted with multiplicities .

Theorem

Let $Q \colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be a quadratic form and let M be its matrix. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ be the roots of $w_M(\lambda)$. Then

i) *form Q is positive definite* $\iff \lambda_1, \ldots, \lambda_n > 0$,

ii) *form Q is positive semidefinite* $\iff \lambda_1, \ldots, \lambda_n \geqslant 0$,

iii) *form Q is negative definite* $\iff \lambda_1, \ldots, \lambda_n < 0$,

iv) *form Q is negative semidefinite* $\iff \lambda_1, \ldots, \lambda_n \leqslant 0$,

v) *form Q is indefinite* $\iff \lambda_i < 0, \lambda_j > 0$ *for some* $1 \leqslant i, j \leqslant n$.

Eigenvalues and Positivity (continued)

Proof.

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of M, that is

$$Mv_i = \lambda_i v_i \text{ for } i = 1, \ldots, n,$$

where $\lambda_i \in \mathbb{R}$ is an eigenvalue of M and $v_i = \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix} \in M(n \times 1; \mathbb{R})$ is taken to be a n-by-1 matrix.

Eigenvalues and Positivity (continued)

Proof.

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of M, that is

$$Mv_i = \lambda_i v_i \text{ for } i = 1, \ldots, n,$$

where $\lambda_i \in \mathbb{R}$ is an eigenvalue of M and $v_i = \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix} \in M(n \times 1; \mathbb{R})$ is taken to be a n-by-1 matrix. Let v_i, v_j be vectors such that $\lambda_i \neq \lambda_j$. Then

$$v_i^\mathsf{T} M v_j = v_i^\mathsf{T}(Mv_j) = v_i^\mathsf{T}(\lambda_j v_j) = \lambda_j (v_i \cdot v_j),$$

Eigenvalues and Positivity (continued)

Proof.

Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of M, that is

$$Mv_i = \lambda_i v_i \text{ for } i = 1, \ldots, n,$$

where $\lambda_i \in \mathbb{R}$ is an eigenvalue of M and $v_i = \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix} \in M(n \times 1; \mathbb{R})$ is taken to be a n-by-1 matrix. Let v_i, v_j be vectors such that $\lambda_i \neq \lambda_j$. Then

$$v_i^\mathsf{T} M v_j = v_i^\mathsf{T}(Mv_j) = v_i^\mathsf{T}(\lambda_j v_j) = \lambda_j(v_i \cdot v_j),$$
$$v_i^\mathsf{T} M v_j = (v_i^\mathsf{T} M^\mathsf{T})v_j = (Mv_i)^\mathsf{T} v_j = (\lambda_i v_i)^\mathsf{T} v_j = \lambda_i(v_i \cdot v_j).$$

This is possible only if $v_i \cdot v_j = 0$, i.e. vectors v_i, v_j are perpendicular.

Eigenvalues and Positivity (continued)

Proof.

Let $v_1, \dots, v_n \in \mathbb{R}^n$ be a basis of $\mathbb{R}^n$ consisting of eigenvectors of M, that is

$$Mv_i = \lambda_i v_i \text{ for } i = 1, \dots, n,$$

where $\lambda_i \in \mathbb{R}$ is an eigenvalue of M and $v_i = \begin{bmatrix} * \\ \vdots \\ * \end{bmatrix} \in M(n \times 1; \mathbb{R})$ is taken to be a n-by-1 matrix. Let v_i, v_j be vectors such that $\lambda_i \neq \lambda_j$. Then

$$v_i^\intercal M v_j = v_i^\intercal (M v_j) = v_i^\intercal (\lambda_j v_j) = \lambda_j (v_i \cdot v_j),$$
$$v_i^\intercal M v_j = (v_i^\intercal M^\intercal) v_j = (M v_i)^\intercal v_j = (\lambda_i v_i)^\intercal v_j = \lambda_i (v_i \cdot v_j).$$

This is possible only if $v_i \cdot v_j = 0$, i.e. vectors v_i, v_j are perpendicular. Using Gram-Schmidt process for eigenspaces $V_{(\lambda_i)}$ one can assume the basis $v_1, \dots, v_n$ is orthonormal.

Eigenvalues and Positivity (continued)

Proof.

That is

$$v_i \cdot v_j = v_i^\mathsf{T} v_j = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}.$$

For any $v \in \mathbb{R}^n$ there exist unique $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n.$$

Eigenvalues and Positivity (continued)

Proof.

That is

$$v_i \cdot v_j = v_i^\mathsf{T} v_j = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}.$$

For any $v \in \mathbb{R}^n$ there exist unique $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n.$$

Now

$$Q(v) = v^\mathsf{T} M v = v^\mathsf{T} M(\alpha_1 v_1 + \ldots + \alpha_n v_n) =$$

$$= (\alpha_1 v_1 + \ldots + \alpha_n v_n)^\mathsf{T}(\lambda_1 \alpha_1 v_1 + \ldots + \lambda_n \alpha_n v_n) = \lambda_1 \alpha_1^2 + \ldots + \lambda_n\, \alpha_n^2.$$

Eigenvalues and Positivity (continued)

Proof.

That is

$$v_i \cdot v_j = v_i^\intercal v_j = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j \end{cases}.$$

For any $v \in \mathbb{R}^n$ there exist unique $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n.$$

Now

$$Q(v) = v^\intercal M v = v^\intercal M(\alpha_1 v_1 + \ldots + \alpha_n v_n) =$$

$$= (\alpha_1 v_1 + \ldots + \alpha_n v_n)^\intercal (\lambda_1 \alpha_1 v_1 + \ldots + \lambda_n \alpha_n v_n) = \lambda_1 \alpha_1^2 + \ldots + \lambda_n \, \alpha_n^2.$$

In particular

$$Q(v_i) = v_i^\intercal M v_i = \lambda_i,$$

$$Q(v) > 0 \text{ for any } v \neq 0 \iff \lambda_1, \ldots, \lambda_n > 0,$$

$$Q(v) \geqslant 0 \text{ for any } v \iff \lambda_1, \ldots, \lambda_n \geqslant 0.$$

Example

Let

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

The eigenvalues are $\lambda_1 = 1 > 0, \lambda_2 = -1 < 0$ therefore the quadratic form $Q((x_1, x_2)) = x_1^2 - x_2^2$ is indefinite.

Example

Let

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

Example

Let

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix}.$$

The characteristic polynomial
$w_M(\lambda) = (1-\lambda)((2-\lambda)^2 - 4) = \lambda(1-\lambda)(\lambda - 4)$ has
non-negative roots $\lambda_1 = 0, \lambda_2 = 1, \lambda_3 = 4,\ \lambda_1, \lambda_2, \lambda_3 \geqslant 0$.
Therefore the quadratic form
$Q((x_1, x_2, x_3)) = x_1^2 + 2x_2^2 + 2x_3^2 + 4x_2x_3 = x_1^2 + 2(x_2 + x_3)^2$ is
positive semidefinite.

Example (continued)

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} \in M(3 \times 2; \mathbb{R})$ where $r(A) = 2$. The matrix

$$A^{\mathsf{T}} A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 4 & 7 \\ 4 & 5 & 5 \\ 7 & 5 & 10 \end{bmatrix},$$

is positive semidefinite and it is not positive definite.

$$\det \begin{bmatrix} 5-\lambda & 4 & 7 \\ 4 & 5-\lambda & 5 \\ 7 & 5 & 10-\lambda \end{bmatrix} \overset{c_3 - 2c_2}{=} \det \begin{bmatrix} 5-\lambda & 4 & -1 \\ 4 & 5-\lambda & 2\lambda - 5 \\ 7 & 5 & -\lambda \end{bmatrix} \overset{\substack{c_1 + (5-\lambda)c_3 \\ c_2 + 4c_3}}{=}$$

$$= \det \begin{bmatrix} 0 & 0 & -1 \\ -2\lambda^2 + 15\lambda - 21 & 7\lambda - 15 & 2\lambda - 5 \\ \lambda^2 - 5\lambda + 7 & 5 - 4\lambda & -\lambda \end{bmatrix} =$$

$$= -\det \begin{bmatrix} -2\lambda^2 + 15\lambda - 21 & 7\lambda - 15 \\ \lambda^2 - 5\lambda + 7 & 5 - 4\lambda \end{bmatrix} =$$

$$= -\lambda(\lambda^2 - 20\lambda + 35).$$

Example (continued)

Therefore one eigenvalue of $A^\intercal A$ is equal to 0, and, by the Viete's formulas,

$$\lambda_1 + \lambda_2 = 20 > 0,$$

$$\lambda_1 \lambda_2 = 35 > 0,$$

the other two eigenvalues are non–negative. In fact

$$\begin{bmatrix} 5 & -1 & -3 \end{bmatrix} \begin{bmatrix} 5 & 4 & 7 \\ 4 & 5 & 5 \\ 7 & 5 & 10 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ -3 \end{bmatrix} = 0.$$

Spectral Theorem

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ be an endomorphism. If there exists an orthonormal basis $\mathcal{A}$ of $\mathbb{R}^n$ such that $M(\varphi)_{\mathcal{A}}^{\mathcal{A}}$ is symmetric then for any orthonormal basis $\mathcal{B}$ of $\mathbb{R}^n$ matrix $M(\varphi)_{\mathcal{B}}^{\mathcal{B}}$ is symmetric.

Spectral Theorem

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ be an endomorphism. If there exists an orthonormal basis $\mathcal{A}$ of $\mathbb{R}^n$ such that $M(\varphi)_{\mathcal{A}}^{\mathcal{A}}$ is symmetric then for any orthonormal basis $\mathcal{B}$ of $\mathbb{R}^n$ matrix $M(\varphi)_{\mathcal{B}}^{\mathcal{B}}$ is symmetric.

Proof.

Let $M = M(\varphi)_{\mathcal{A}}^{\mathcal{A}}, N = M(\varphi)_{\mathcal{B}}^{\mathcal{B}}$. Matrix

$$Q = M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}} = M(\mathrm{id})_{st}^{\mathcal{A}} M(\mathrm{id})_{\mathcal{B}}^{st},$$

is orthogonal, i.e. $Q^{-1} = Q^{\mathsf{T}}$.

Spectral Theorem

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ *be an endomorphism. If there exists an orthonormal basis* $\mathcal{A}$ *of* $\mathbb{R}^n$ *such that* $M(\varphi)_{\mathcal{A}}^{\mathcal{A}}$ *is symmetric then for any orthonormal basis* $\mathcal{B}$ *of* $\mathbb{R}^n$ *matrix* $M(\varphi)_{\mathcal{B}}^{\mathcal{B}}$ *is symmetric.*

Proof.

Let $M = M(\varphi)_{\mathcal{A}}^{\mathcal{A}}, N = M(\varphi)_{\mathcal{B}}^{\mathcal{B}}$. Matrix

$$Q = M(\mathrm{id})_{\mathcal{B}}^{\mathcal{A}} = M(\mathrm{id})_{st}^{\mathcal{A}} M(\mathrm{id})_{\mathcal{B}}^{st},$$

is orthogonal, i.e. $Q^{-1} = Q^{\mathsf{T}}$. Because

$$N = Q^{\mathsf{T}} M Q,$$

we have

$$N^{\mathsf{T}} = Q^{\mathsf{T}} M^{\mathsf{T}} (Q^{\mathsf{T}})^{\mathsf{T}} = N.$$

$\square$

Spectral Theorem (continued)

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ *be an endomorphism such that matrix* $M(\varphi)_{st}^{st}$ *is symmetric. Then there exists* $\mu \in \mathbb{R}$ *and* $v \in \mathbb{R}^n, v \neq \mathbf{0}$ *such that* $\varphi(v) = \mu v$*, i.e* μ *is an eigenvalue of* φ *and* v *is an eigenvector for* μ*.*

Spectral Theorem (continued)

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ be an endomorphism such that matrix $M(\varphi)_{st}^{st}$ is symmetric. Then there exists $\mu \in \mathbb{R}$ and $v \in \mathbb{R}^n, v \neq \mathbf{0}$ such that $\varphi(v) = \mu v$, i.e μ is an eigenvalue of φ and v is an eigenvector for μ.

Proof.

Let $M = M(\varphi)_{st}^{st}$ and let $w_M(\lambda) = \det(M - \lambda I_n)$ be the characteristic polynomial of φ. By the fundamental theorem of algebra there exists a complex root $\mu \in \mathbb{C}$ of w_M and a complex eigenvector $v \in \mathbb{C}^n$, i.e. $w_M(\mu) = 0$ and $Mv = \mu v$. Matrix M is real therefore $M\overline{v} = \overline{\mu v}$. Moreover

$$\overline{v}^{\mathsf{T}} M v = (M\overline{v})^{\mathsf{T}}\, v = \overline{\mu}\overline{v}^{\mathsf{T}} v = \overline{\mu}\|v\|,$$

$$\overline{v}^{\mathsf{T}} M v = \overline{v}^{\mathsf{T}}\,(Mv) = \overline{v}^{\mathsf{T}}\,(\mu v) = \mu\|v\|.$$

This implies $\mu \in \mathbb{R}$ and since $V_{(\mu)}$ is given a system of linear equations with real coefficients one can choose $v \in \mathbb{R}^n$. □

Spectral Theorem (continued)

Proposition

Let $\varphi \colon \mathbb{R}^n \to \mathbb{R}^n$ be an endomorphism such that matrix $M(\varphi)_{st}^{st}$ is symmetric. Then for any subspace $W \subset \mathbb{R}^n$ such that $\varphi(W) \subset W$,

$$\varphi(W^{\perp}) \subset W^{\perp}.$$

Spectral Theorem (continued)

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ *be an endomorphism such that matrix* $M(\varphi)_{st}^{st}$ *is symmetric. Then for any subspace* $W \subset \mathbb{R}^n$ *such that* $\varphi(W) \subset W$*,*

$$\varphi(W^\perp) \subset W^\perp.$$

Let $M = M(\varphi)_{st}^{st}$ and let $w \in W^\perp$ be any vector. We need to check that $v^\mathsf{T}(Mw) = 0$ for a any $v \in W$.

Spectral Theorem (continued)

Proposition

Let $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ be an endomorphism such that matrix $M(\varphi)_{st}^{st}$ is symmetric. Then for any subspace $W \subset \mathbb{R}^n$ such that $\varphi(W) \subset W$,

$$\varphi(W^\perp) \subset W^\perp.$$

Let $M = M(\varphi)_{st}^{st}$ and let $w \in W^\perp$ be any vector. We need to check that $v^\mathsf{T}(Mw) = 0$ for a any $v \in W$. In fact

$$v^\mathsf{T} M w = (Mv)^\mathsf{T} w = 0.$$

Spectral Theorem

Theorem

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ *is diagonalizable.*

Spectral Theorem

Theorem

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ *is diagonalizable.*

Proof.

Let φ be an endomorphism given by $M = M(\varphi)_{st}^{st}$. Assume $W \subset \mathbb{R}^n$ is a subspace spanned by pairwise perpendicular eigenvectors of φ. Let $V = W^{\perp}$.

Spectral Theorem

Theorem

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ *is diagonalizable.*

Proof.

Let φ be an endomorphism given by $M = M(\varphi)_{st}^{st}$. Assume $W \subset \mathbb{R}^n$ is a subspace spanned by pairwise perpendicular eigenvectors of φ. Let $V = W^{\perp}$. Matrix of φ relative to some orthonormal basis of V is symmetric and $\varphi(V) \subset V$. Therefore there exists $\mu \in \mathbb{R}$ and a non–zero vector $v \in V$ such that $\varphi(v) = \mu v$.

Spectral Theorem

Theorem

Symmetric matrix $M \in M(n \times n; \mathbb{R})$ is diagonalizable.

Proof.

Let φ be an endomorphism given by $M = M(\varphi)_{st}^{st}$. Assume $W \subset \mathbb{R}^n$ is a subspace spanned by pairwise perpendicular eigenvectors of φ. Let $V = W^{\perp}$. Matrix of φ relative to some orthonormal basis of V is symmetric and $\varphi(V) \subset V$. Therefore there exists $\mu \in \mathbb{R}$ and a non–zero vector $v \in V$ such that $\varphi(v) = \mu v$. By continuing this process we obtain an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors of φ. □

Spectral Theorem (continued)

Corollary

For any symmetric matrix $M \in M(n \times n; \mathbb{R})$ *there exists matrix* $Q \in M(n \times n; \mathbb{R})$ *such that* $Q^{-1} = Q^{\mathsf{T}}$ *and the matrix*

$$D = Q^{\mathsf{T}} M Q,$$

is diagonal.

Spectral Theorem (continued)

Corollary

For any symmetric matrix $M \in M(n \times n; \mathbb{R})$ *there exists matrix* $Q \in M(n \times n; \mathbb{R})$ *such that* $Q^{-1} = Q^{\intercal}$ *and the matrix*

$$D = Q^{\intercal} M Q,$$

is diagonal.

Proof.

Let $\mathcal{A} = (v_1, \dots, v_n)$ be an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of M. If $Q = M(\mathrm{id})^{st}_{\mathcal{A}}$ then the matrix $Q^{-1} M Q$ is diagonal and $Q^{\intercal} Q = I_n$, i.e. $Q^{-1} = Q^{\intercal}$. □

Characterization of Real Symmetric Matrices

Corollary

Let $M \in M(n \times n; \mathbb{R})$ be a real matrix. Then $M = M^\intercal$ if and only if there exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ (i.e. $Q^\intercal Q = I$) such that the matrix

$$D = Q^\intercal M Q \in M(n \times n; \mathbb{R}),$$

is diagonal.

$(\Rightarrow)$ previous corollary

Characterization of Real Symmetric Matrices

Corollary

Let $M \in M(n \times n; \mathbb{R})$ be a real matrix. Then $M = M^\intercal$ if and only if there exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ (i.e. $Q^\intercal Q = I$) such that the matrix

$$D = Q^\intercal M Q \in M(n \times n; \mathbb{R}),$$

is diagonal.

$(\Rightarrow)$ previous corollary

$(\Leftarrow)$ If $D = Q^\intercal M Q$ then $M = QDQ^\intercal$ and since $D^\intercal = D$

$$M^\intercal = (Q^\intercal)^\intercal D^\intercal Q^\intercal = QDQ^\intercal = M.$$

Bilinear Form

Definition

Let V be a vector space. A function

$$B\colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v + u, w) = B(v, w) + B(u, w)$ for any $u, v, w \in V$,

Bilinear Form

Definition

Let V be a vector space. A function

$$B\colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v+u,w) = B(v,w) + B(u,w)$ for any $u, v, w \in V$,

ii) $B(v,u+w) = B(v,u) + B(v,w)$ for any $u, v, w \in V$,

Bilinear Form

Definition

Let V be a vector space. A function

$$B\colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v+u, w) = B(v, w) + B(u, w)$ for any $u, v, w \in V$,

ii) $B(v, u+w) = B(v, u) + B(v, w)$ for any $u, v, w \in V$,

iii) $B(\alpha v, w) = \alpha B(v, w)$ for any $v, w \in V, \alpha \in \mathbb{R}$,

Bilinear Form

Definition

Let V be a vector space. A function

$$B\colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v + u, w) = B(v, w) + B(u, w)$ for any $u, v, w \in V$,

ii) $B(v, u + w) = B(v, u) + B(v, w)$ for any $u, v, w \in V$,

iii) $B(\alpha v, w) = \alpha B(v, w)$ for any $v, w \in V, \alpha \in \mathbb{R}$,

iv) $B(v, \beta w) = \beta B(v, w)$ for any $v, w \in V, \beta \in \mathbb{R}$.

Bilinear Form

Definition

Let V be a vector space. A function

$$B\colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v+u,w) = B(v,w) + B(u,w)$ for any $u,v,w \in V$,

ii) $B(v,u+w) = B(v,u) + B(v,w)$ for any $u,v,w \in V$,

iii) $B(\alpha v,w) = \alpha B(v,w)$ for any $v,w \in V, \alpha \in \mathbb{R}$,

iv) $B(v,\beta w) = \beta B(v,w)$ for any $v,w \in V, \beta \in \mathbb{R}$.

Bilinear Form

Definition

Let V be a vector space. A function

$$B \colon V \times V \to \mathbb{R}$$

is called a **bilinear form** if

i) $B(v+u, w) = B(v, w) + B(u, w)$ for any $u, v, w \in V$,
ii) $B(v, u+w) = B(v, u) + B(v, w)$ for any $u, v, w \in V$,
iii) $B(\alpha v, w) = \alpha B(v, w)$ for any $v, w \in V, \alpha \in \mathbb{R}$,
iv) $B(v, \beta w) = \beta B(v, w)$ for any $v, w \in V, \beta \in \mathbb{R}$.

Bilinear form B is called **symmetric** if moreover

v) $B(v, w) = B(w, v)$ for any $v, w \in V$.

Bilinear Forms (continued)

Definition

If $B\colon V \times V \to \mathbb{R}$ is a bilinear form and $\mathcal{A} = (v_1, \ldots, v_n)$ is a basis of V then the **matrix of bilinear form** B relative to the basis $\mathcal{A}$ is equal to

$$M(B)_{\mathcal{A}} = [m_{ij}] \in M(n \times n; \mathbb{R}),$$

where $m_{ij} = B(v_i, v_j)$, i.e

$$M(B)_{\mathcal{A}} = \begin{bmatrix} B(v_1, v_1) & B(v_1, v_2) & \ldots & B(v_1, v_n) \\ B(v_2, v_1) & B(v_2, v_2) & \ldots & B(v_2, v_n) \\ \vdots & \vdots & \ddots & \vdots \\ B(v_n, v_1) & B(v_n, v_2) & \ldots & B(v_n, v_n) \end{bmatrix}.$$

Bilinear Forms (continued)

Proposition

For any $v, w \in V$ and any basis $\mathcal{A}$

$$B(v, w) = [v]_{\mathcal{A}}^{\intercal} M(B)_{\mathcal{A}} [w]_{\mathcal{A}} .$$

Bilinear Forms (continued)

Proposition

For any $v, w \in V$ and any basis $\mathcal{A}$

$$B(v,w) = \left[v\right]_{\mathcal{A}}^{\mathsf{T}} M(B)_{\mathcal{A}} \left[w\right]_{\mathcal{A}}.$$

Proof.

Let $\mathcal{A} = (v_1, \ldots, v_n)$ and let
$v = \alpha_1 v_1 + \ldots + \alpha_n v_n$, $w = \beta_1 v_1 + \ldots + \beta_n v_n$.

$$B(v,w) = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} B(v_i, w_j) \beta_j \right) =$$

$$= \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} m_{ij} \beta_j \right),$$

where $M(B)_{\mathcal{A}} = [m_{ij}]$. $\square$

Bilinear Forms (continued)

Corollary

If $\mathcal{A}, \mathcal{B}$ *are bases of* V *then for* $C = M(\mathrm{id})_{\mathcal{A}}^{\mathcal{B}}$

$$M(B)_{\mathcal{A}} = C^{\intercal} M(B)_{\mathcal{B}} C.$$

Bilinear Forms (continued)

Corollary

If $\mathcal{A}, \mathcal{B}$ are bases of V then for $C = M(\mathrm{id})^{\mathcal{B}}_{\mathcal{A}}$

$$M(B)_{\mathcal{A}} = C^{\intercal} M(B)_{\mathcal{B}} C.$$

Proof.

Let $M(B)_{\mathcal{A}} = M = [m_{ij}]$ and let $\mathcal{A} = (v_1, \ldots, v_n)$. By the previous proposition

$$\varepsilon_i^{\intercal}\left(C^{\intercal} M(B)_{\mathcal{B}} C\right)\varepsilon_j = \left(C\varepsilon_j\right)^{\intercal} M(B)_{\mathcal{B}}\left(C\varepsilon_j\right) = \left[v_i\right]_{\mathcal{B}}^{\intercal} M(B)_{\mathcal{B}}\left[v_j\right]_{\mathcal{B}} =$$

$$= B(v_i, v_j).$$

On the other hand, for any $M = [m_{ij}]$

$$\varepsilon_i^{\intercal} M \varepsilon_j = m_{ij},$$

hence $C^{\intercal} M(B)_{\mathcal{B}} C = M = M(B)_{\mathcal{A}}$. $\square$

Bilinear Forms (continued)

Let $B\colon V \times V \to \mathbb{R}$.

Corollary

If B is a symmetric bilinear form then for any basis $\mathcal{A}$ of V

$$M(B)_{\mathcal{A}}^{\intercal} = M(B)_{\mathcal{A}}.$$

Bilinear Forms (continued)

Let $B\colon V \times V \to \mathbb{R}$.

Corollary

If B is a symmetric bilinear form then for any basis $\mathcal{A}$ of V

$$M(B)_{\mathcal{A}}^{\intercal} = M(B)_{\mathcal{A}}.$$

If B is a bilinear form and for some basis $\mathcal{A}$ of V

$$M(B)_{\mathcal{A}}^{\intercal} = M(B)_{\mathcal{A}},$$

then B is symmetric bilinear form.

Bilinear Forms (continued)

Let $B\colon V \times V \to \mathbb{R}$.

Corollary

If B is a symmetric bilinear form then for any basis $\mathcal{A}$ of V

$$M(B)_{\mathcal{A}}^{\intercal} = M(B)_{\mathcal{A}}.$$

If B is a bilinear form and for some basis $\mathcal{A}$ of V

$$M(B)_{\mathcal{A}}^{\intercal} = M(B)_{\mathcal{A}},$$

then B is symmetric bilinear form.

Proof.

For the second claim, let $M = [m_{ij}] = M(B)_{\mathcal{A}}$ be symmetric, i.e. $M^{\intercal} = M$. Then for any $v, w \in \mathbb{R}^n$

$$B(v,w) = [v]_{\mathcal{A}}^{\intercal} M [w]_{\mathcal{A}} = \left([v]_{\mathcal{A}}^{\intercal} M [w]_{\mathcal{A}}\right)^{\intercal} = [w]_{\mathcal{A}}^{\intercal} M [v]_{\mathcal{A}} = B(w,v),$$

i.e. B is symmetric. $\square$

Quadratic Forms

Let V be a vector space.

Definition

Function $Q\colon V \to \mathbb{R}$ is a **quadratic form** if there exist a bilinear form $B\colon V \times V \to \mathbb{R}$ such that $Q(v) = B(v, v)$ for any $v \in V$.

Quadratic Forms

Let V be a vector space.

Definition

Function $Q \colon V \to \mathbb{R}$ is a **quadratic form** if there exist a bilinear form $B \colon V \times V \to \mathbb{R}$ such that $Q(v) = B(v, v)$ for any $v \in V$.

Proposition

If $Q \colon V \to \mathbb{R}$ is a quadratic form then

$$B_s(v, w) = \frac{1}{2}\left(Q(v + w, v + w) - Q(v, v) - Q(w, w)\right),$$

is a **symmetric** *bilinear form such that $Q(v) = B_s(v, v)$.*

Quadratic Forms

Let V be a vector space.

Definition

Function $Q\colon V \to \mathbb{R}$ is a **quadratic form** if there exist a bilinear form $B\colon V \times V \to \mathbb{R}$ such that $Q(v) = B(v, v)$ for any $v \in V$.

Proposition

If $Q\colon V \to \mathbb{R}$ is a quadratic form then

$$B_s(v, w) = \frac{1}{2}\left(Q(v + w, v + w) - Q(v, v) - Q(w, w)\right),$$

is a **symmetric** *bilinear form such that $Q(v) = B_s(v, v)$.*

Proof.

$$B_s(v, w) = \frac{1}{2}\left(B(v, w) + B(w, v)\right).$$

$\square$

Sylvester's Criterion

Proposition

Let $Q \colon V \to \mathbb{R}$ be a quadratic form such that $Q(v) = B(v, v)$ where B is a symmetric bilinear form. Let $\mathcal{A} = (v_1, \ldots, v_n)$ be a basis of V. Then Q is positive definite if and only if

$$\det M(B)_{\mathcal{A}_i} > 0,$$

for $i = 1, \ldots, n$ where $\mathcal{A}_i = (v_1, \ldots, v_i)$.

Sylvester's Criterion

Proposition

Let $Q\colon V \to \mathbb{R}$ be a quadratic form such that $Q(v) = B(v, v)$ where B is a symmetric bilinear form. Let $\mathcal{A} = (v_1, \ldots, v_n)$ be a basis of V. Then Q is positive definite if and only if

$$\det M(B)_{\mathcal{A}_i} > 0,$$

for $i = 1, \ldots, n$ where $\mathcal{A}_i = (v_1, \ldots, v_i)$.

Proof.

$(\Rightarrow)$ The quadratic form Q restricted to $\mathrm{lin}(v_1, \ldots, v_i)$ is positive hence the matrix $M(B)_{\mathcal{A}_i}$ is symmetric diagonalizable and by the eigenvalue criterion its all eigenvalues $\lambda_1, \ldots, \lambda_i > 0$ are positive. Therefore

$$\det M(B)_{\mathcal{A}_i} = \lambda_1 \cdot \ldots \cdot \lambda_i > 0.$$

Note that eigenvalues depend on i.

Sylvester's Criterion (continued)

Proof.

$(\Leftarrow)$ let $V_k = \mathrm{lin}(v_1, \ldots, v_k)$. By induction on k we prove the claim

„the quadratic form $Q|_{V_k}$ is positive definite",

which for $k = n$ is the assertion of the Theorem.

Sylvester's Criterion (continued)

Proof.

($\Leftarrow$) let $V_k = \mathrm{lin}(v_1, \ldots, v_k)$. By induction on k we prove the claim

„the quadratic form $Q|_{V_k}$ is positive definite",

which for $k = n$ is the assertion of the Theorem.

For $k = 1$ the claim holds since $\det M(B)_{\mathcal{A}_1} = B(v_1, v_1) > 0$.

Sylvester's Criterion (continued)

Proof.

($\Leftarrow$) let $V_k = \mathrm{lin}(v_1, \ldots, v_k)$. By induction on k we prove the claim

$$\text{„the quadratic form } Q|_{V_k} \text{ is positive definite''},$$

which for $k = n$ is the assertion of the Theorem.

For $k = 1$ the claim holds since $\det M(B)_{\mathcal{A}_1} = B(v_1, v_1) > 0$.

For $k = 2$ let $\lambda_1, \lambda_2 \in \mathbb{R}$ are eigenvalues of $M(B)_{\mathcal{A}_2}$.

Sylvester's Criterion (continued)

Proof.

($\Leftarrow$) let $V_k = \text{lin}(v_1, \ldots, v_k)$. By induction on k we prove the claim

„the quadratic form $Q|_{V_k}$ is positive definite",

which for $k = n$ is the assertion of the Theorem.

For $k = 1$ the claim holds since $\det M(B)_{\mathcal{A}_1} = B(v_1, v_1) > 0$.

For $k = 2$ let $\lambda_1, \lambda_2 \in \mathbb{R}$ are eigenvalues of $M(B)_{\mathcal{A}_2}$. By Viete's formulas

$$\begin{cases} \lambda_1 + \lambda_2 & = & B(v_1, v_1) + B(v_2, v_2), \\ \lambda_1 \lambda_2 & = & B(v_1, v_1)B(v_2, v_2) - B(v_1, v_2)^2. \end{cases}$$

Because $\lambda_1 \lambda_2 = \det M(B)_{\mathcal{A}_2} > 0$ either $B(v_1, v_1) < 0, B(v_2, v_2) < 0$ or $B(v_1, v_1) > 0, B(v_2, v_2) > 0$. Since $B(v_1, v_1) = \det M(B)_{\mathcal{A}_1} > 0$ the latter holds, hence $\lambda_1, \lambda_2 > 0$.

Sylvester's Criterion (continued)

Proof.

Assume that $k \geqslant 3$ and $\det M(B)_{\mathcal{A}_i} > 0$, for $i = 1, \ldots, k$ (i.e. $Q|_{V_{k-1}}$ is positive definite) but $Q|_{V_k}$ is not positive definite.

Sylvester's Criterion (continued)

Proof.

Assume that $k \geqslant 3$ and $\det M(B)_{\mathcal{A}_i} > 0$, for $i = 1, \ldots, k$ (i.e. $Q|_{V_{k-1}}$ is positive definite) but $Q|_{V_k}$ is not positive definite.

Therefore $M(B)_{\mathcal{A}_k}$ has at least two negative eigenvalues $\lambda_1, \lambda_2 < 0$ or a negative eigenvalue $\lambda < 0$ of multiplicity at least 2 ($\det M(B)_{\mathcal{A}_k} > 0$ is equal to the product of eigenvalues).

Sylvester's Criterion (continued)

Proof.

Assume that $k \geqslant 3$ and $\det M(B)_{\mathcal{A}_i} > 0$, for $i = 1, \ldots, k$ (i.e. $Q|_{V_{k-1}}$ is positive definite) but $Q|_{V_k}$ is not positive definite.

Therefore $M(B)_{\mathcal{A}_k}$ has at least two negative eigenvalues $\lambda_1, \lambda_2 < 0$ or a negative eigenvalue $\lambda < 0$ of multiplicity at least 2 ($\det M(B)_{\mathcal{A}_k} > 0$ is equal to the product of eigenvalues).

In both cases there exist eigenvectors $w_1, w_2 \in V_k$ of $M(B)_{\mathcal{A}_k}$, that is

$$M(B)_{\mathcal{A}_k}\big[w_i\big]_{\mathcal{A}_k} = \lambda_i\big[w_i\big]_{\mathcal{A}_k} \text{ for } i = 1, 2,$$

and $\big[w_1\big]^{\mathsf{T}}_{\mathcal{A}_k}\big[w_2\big]_{\mathcal{A}_k} = 0$ (including the case $\lambda_1 = \lambda_2 = \lambda$).

Sylvester's Criterion (continued)

Proof.

Assume that $k \geqslant 3$ and $\det M(B)_{\mathcal{A}_i} > 0$, for $i = 1, \ldots, k$ (i.e. $Q|_{V_{k-1}}$ is positive definite) but $Q|_{V_k}$ is not positive definite.

Therefore $M(B)_{\mathcal{A}_k}$ has at least two negative eigenvalues $\lambda_1, \lambda_2 < 0$ or a negative eigenvalue $\lambda < 0$ of multiplicity at least 2 ($\det M(B)_{\mathcal{A}_k} > 0$ is equal to the product of eigenvalues).

In both cases there exist eigenvectors $w_1, w_2 \in V_k$ of $M(B)_{\mathcal{A}_k}$, that is

$$M(B)_{\mathcal{A}_k}\big[w_i\big]_{\mathcal{A}_k} = \lambda_i\big[w_i\big]_{\mathcal{A}_k} \text{ for } i = 1, 2,$$

and $\big[w_1\big]^{\mathsf{T}}_{\mathcal{A}_k}\big[w_2\big]_{\mathcal{A}_k} = 0$ (including the case $\lambda_1 = \lambda_2 = \lambda$). Note that $w_1, w_2 \notin V_{k-1}$.

Sylvester's Criterion (continued)

Proof.

Let $w_1 = \alpha_1 v_1 + \ldots + \alpha_k v_k, w_2 = \beta_1 v_1 + \ldots + \beta_k v_k$ and let $v = \gamma_1 w_1 + \gamma_2 w_2 \in V_{k-1}$ where $\gamma_1 = \beta_k, \gamma_2 = -\alpha_k$. Then $\gamma_1, \gamma_2 \neq 0$ since $w_1, w_2 \notin V_{k-1}$. Vectors w_1, w_2 are perpendicular (i.e. linearly independent), therefore $v \neq \mathbf{0}$. Hence

$$\left[v\right]_{\mathcal{A}_k}^{\mathsf{T}} M(B)_{\mathcal{A}_k} \left[v\right]_{\mathcal{A}_k} =$$

$$= \left(\gamma_1 \left[w_1\right]_{\mathcal{A}_k} + \gamma_2 \left[w_2\right]_{\mathcal{A}_k}\right)^{\mathsf{T}} M(B)_{\mathcal{A}_k} \left(\gamma_1 \left[w_1\right]_{\mathcal{A}_k} + \gamma_2 \left[w_2\right]_{\mathcal{A}_k}\right) =$$

$$= \lambda_1 \gamma_1^2 \left\| \left[w_1\right]_{\mathcal{A}_k} \right\|^2 + \lambda_2 \gamma_2^2 \left\| \left[w_2\right]_{\mathcal{A}_k} \right\|^2 < 0.$$

On the other hand

$$\left[v\right]_{\mathcal{A}_k}^{\mathsf{T}} M(B)_{\mathcal{A}_k} \left[v\right]_{\mathcal{A}_k} = Q(v) > 0,$$

because $Q|_{V_{k-1}}$ is positive definite, which yields a contradiction.

Summary

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric matrix. The following are equivalent

i) $x^\mathsf{T} M x \geqslant 0$ (matrix M is positive semidefinite),

ii) $\min\{\lambda \mid \lambda \text{ is an eigenvalue of } M\} \geqslant 0$,

iii) $\min\{x^\mathsf{T} M x \in \mathbb{R} \mid \|x\| = 1\} \geqslant 0$,

iv) all principal minors of M are non–negative,

v) there exists a matrix $N \in M(n \times n; \mathbb{R})$ such that $M = N^\mathsf{T} N$.

Summary (continued)

Let $M \in M(n \times n; \mathbb{R})$ be a symmetric matrix. The following are equivalent

i) $x^{\mathsf{T}} M x > 0$ for $x \neq \mathbf{0}$ (matrix M is positive definite),

ii) $\min\{\lambda \mid \lambda \text{ is an eigenvalue of } M\} > 0$,

iii) $\min\{x^{\mathsf{T}} M x \in \mathbb{R} \mid \|x\| = 1\} > 0$,

iv) all leading principal minors of M are positive,

v) there exists a non–singular (i.e. $\det N \neq 0$) matrix $N \in M(n \times n; \mathbb{R})$ such that $M = N^{\mathsf{T}} N$.

Interlacing Eigenvalues

Theorem

If $M \in M(n \times n; \mathbb{R})$ is a symmetric matrix, i.e. $M = M^{\intercal}$. Let M_i denote the top left $i{-}by{-}i$ submatrix of M. Fix $m < n$. Let $\lambda_1, \ldots, \lambda_m$ denote the eigenvalues of M_m and $\mu_1, \ldots, \mu_{m+1}$ denote the eigenvalues of M_{m+1}. Then

$$\mu_1 \leqslant \lambda_1 \leqslant \mu_2 \leqslant \lambda_2 \leqslant \mu_3 \leqslant \ldots \leqslant \lambda_m \leqslant \mu_{m+1}.$$

Proof.

Omitted. $\square$

Hessian Matrix

Definition

Let $f\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^k$, be a function of class $\mathcal{C}^2$ on the open set U. **Hessian matrix** at $x_0 \in U$ is the symmetric matrix $H_f(x_0) = H(x_0) \in M(k \times k; \mathbb{R})$ given by

$$H_f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k}(x_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_k}(x_0) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(x_0) & \frac{\partial^3 f}{\partial x_3^3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_3 \partial x_k}(x_0) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_k \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_k \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_k \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_k^2}(x_0) \end{bmatrix}.$$

Hessian Matrix

Definition

Let $f\colon U \to \mathbb{R},\ U \subset \mathbb{R}^k$, be a function of class $\mathcal{C}^2$ on the open set U. **Hessian matrix** at $x_0 \in U$ is the symmetric matrix $H_f(x_0) = H(x_0) \in M(k \times k; \mathbb{R})$ given by

$$H_f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k}(x_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_2^2}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_k}(x_0) \\ \frac{\partial^2 f}{\partial x_3 \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_3 \partial x_2}(x_0) & \frac{\partial^3 f}{\partial x_3^3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_3 \partial x_k}(x_0) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_k \partial x_1}(x_0) & \frac{\partial^2 f}{\partial x_k \partial x_2}(x_0) & \frac{\partial^2 f}{\partial x_k \partial x_3}(x_0) & \cdots & \frac{\partial^2 f}{\partial x_k^2}(x_0) \end{bmatrix}.$$

Remark

If f is not of class $\mathcal{C}^2$ the matrix $H_f(x_0)$ may not be symmetric.

Local Minima or Maxima of a Multivariate Function

Theorem

Let $f : U \to \mathbb{R},\ U \subset \mathbb{R}^k$ be a function of class $\mathcal{C}^2$ on the open set U. If $x_0 \in U$ is a **critical point** *of function f, i.e.*

$$f'(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \dots, \frac{\partial f}{\partial x_k}(x_0)\right) = \mathbf{0},$$

and the Hessian matrix $H(x_0)$ is negative (respectively, positive) definite, then f has strict local maximum (respectively strict local minimum) at the point $x_0 \in U$.

Local Minima or Maxima of a Multivariate Function

Theorem

Let $f : U \to \mathbb{R}$, $U \subset \mathbb{R}^k$ be a function of class $\mathcal{C}^2$ on the open set U. If $x_0 \in U$ is a **critical point** *of function f, i.e.*

$$f'(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_k}(x_0) \right) = \mathbf{0},$$

and the Hessian matrix $H(x_0)$ is negative (respectively, positive) definite, then f has strict local maximum (respectively strict local minimum) at the point $x_0 \in U$.

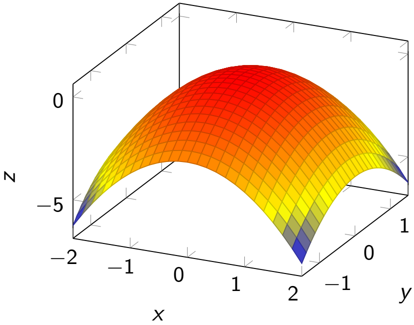
If the matrix $H(x_0)$ is indefinite then f has no local extremum at x_0 (the point x_0 is so called **saddle point***).*

Proof.

Analysis course (use multivariate Taylor formula). $\square$

Example – Local Maximum

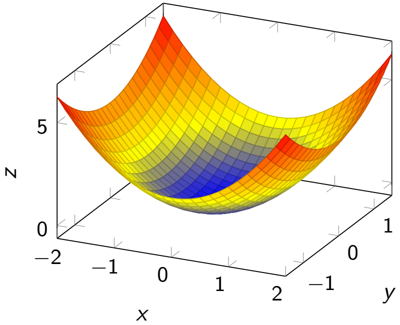
graph of the function $f(x, y) = -x^2 - y^2$



$$H_f(0,0) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} \text{ negative definite}$$

Example – Local Minimum

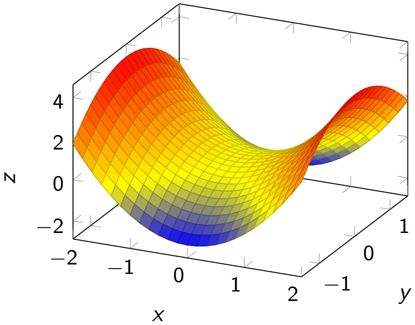
graph of the function $f(x, y) = x^2 + y^2$



$$H_f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ positive definite}$$

Example – Saddle Point – No Local Extremum

graph of the function $f(x, y) = x^2 - y^2$



$$H_f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \text{ indefinite}$$

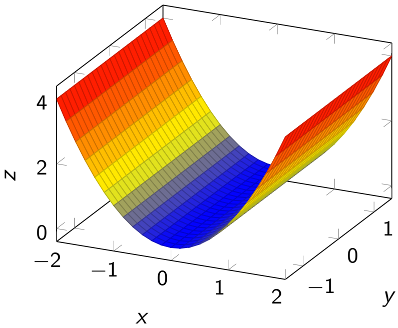
Local Minima or Maxima of a Multivariate Function (continued)

Remark

If the matrix $H(x_0)$ *is positive semidefinite or negative semidefinite then the function* f *has at* x_0 *local minimum or local maximum or a saddle point (the criterion is indecisive).*

Example – Hessian Matrix Positive Semidefinite – Weak Local Minimum

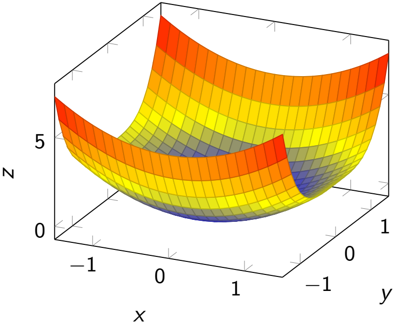
graph of the function $f(x, y) = x^2$



$$H_f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ positive semidefinite}$$

Example – Hessian Matrix Positive Semidefinite – Strict Local Minimum

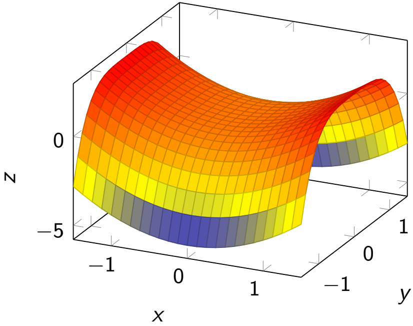
graph of the function $f(x, y) = x^2 + y^4$



$$H_f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ positive semidefinite}$$

Example – Hessian Matrix Positive Semidefinite – Saddle Point

graph of the function $f(x, y) = x^2 - y^4$



$$H_f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \text{ positive semidefinite}$$

Square Root of a Positive Semidefinite Matrix

Find a matrix $X \in M(2 \times 2; \mathbb{R})$ such that

$$X^2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} = A.$$

Square Root of a Positive Semidefinite Matrix

Find a matrix $X \in M(2 \times 2; \mathbb{R})$ such that

$$X^2 = \begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix} = A.$$

It can be checked that

$$A = Q^\intercal \begin{bmatrix} 1 & 0 \\ 0 & 9 \end{bmatrix} Q,$$

where

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Square Root of a Positive Semidefinite Matrix (continued)

$$X_1 = Q^\mathsf{T} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} Q = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

$$X_2 = Q^\mathsf{T} \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} Q = \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix},$$

$$X_3 = Q^\mathsf{T} \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix} Q = \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix},$$

$$X_4 = Q^\mathsf{T} \begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix} Q = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Multivariate Gaussian Distribution

The probability density function of multivariate n-dimensional Gaussian distribution is given by

$$p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{(\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^\intercal \Sigma^{-1}(x-\mu)},$$

where $x \in \mathbb{R}^n$ for some fixed $\mu \in \mathbb{R}^n$ and $\Sigma \in M(n \times n; \mathbb{R})$ a symmetric positive definite matrix.

Multivariate Gaussian Distribution

The probability density function of multivariate n-dimensional Gaussian distribution is given by

$$p(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{(\det \Sigma)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^\intercal \Sigma^{-1}(x-\mu)},$$

where $x \in \mathbb{R}^n$ for some fixed $\mu \in \mathbb{R}^n$ and $\Sigma \in M(n \times n; \mathbb{R})$ a symmetric positive definite matrix.There exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ (i.e. $QQ^\intercal = Q^\intercal Q = I$) such that

$$Q^\intercal \Sigma Q = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix},$$

where $\sigma_1, \ldots, \sigma_n > 0$, $Q = [v_1 \ v_2 \ \cdots v_n]$ and $\mathcal{B} = (v_1, v_2, \ldots, v_n)$ is an orthonormal basis of $\mathbb{R}^n$.

Multivariate Gaussian Distribution (continued)

Then if

$$x = \sum_{i=1}^{n} x_i v_i,$$

(i.e. $[x]_{\mathcal{B}} = [x_1 \; x_2 \; \cdots x_n]^\mathsf{T}$) and

$$\mu = (\mu_1, \mu_2, \ldots, \mu_n),$$

then

$$p(x \mid \mu, \Sigma) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma_i^2)^{\frac{1}{2}}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}},$$

i.e., it is a product of one–dimensional Gaussian probability density functions.

Multivariate Gaussian Distribution – Example

Let

$$\Sigma = \begin{bmatrix} \frac{5}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{5}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^{\intercal} \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix},$$

$$\Sigma^{-1} = \begin{bmatrix} \frac{5}{8} & \frac{3}{8} \\ \frac{3}{8} & \frac{5}{8} \end{bmatrix},$$

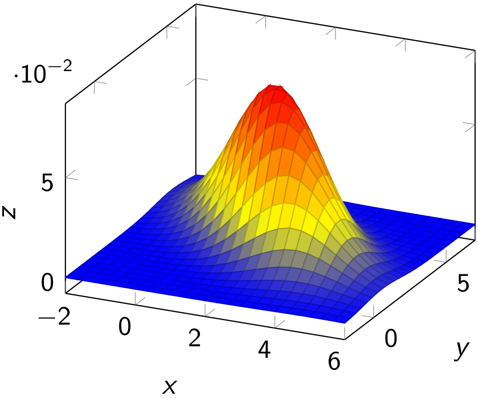
$$\mu = (2,3), \quad v_1 = \frac{1}{\sqrt{2}}(1,1), \quad v_2 = \frac{1}{\sqrt{2}}(1,-1),$$

then

$$p((x_1,x_2) \mid \mu,\Sigma) = \frac{1}{2\pi}\frac{1}{2}e^{-\frac{1}{16}(5(x_1-2)^2+6(x_1-2)(x_2-3)+5(x_2-3)^2)},$$

$$p(x_1 v_1 + x_2 v_2 \mid \mu,\Sigma) = \frac{1}{(2\pi)^{\frac{1}{2}}}e^{-\frac{(x_1-2)^2}{2}}\frac{1}{2(2\pi)^{\frac{1}{2}}}e^{-\frac{(x_2-3)^2}{8}},$$

Multivariate Gaussian Distribution – Example



probability density functions for $\Sigma = \frac{1}{2}\begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}, \ \mu = (2,3)$

Inner product

Definition

Inner product space V is a vector space V over $\mathbb{C}$ or $\mathbb{R}$, with a function

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C},$$

such that

i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$,

Inner product

Definition

Inner product space V is a vector space V over $\mathbb{C}$ or $\mathbb{R}$, with a function

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C},$$

such that

i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$,

ii) for any $v \in V$ the function $\langle v, \cdot \rangle \colon V \to \mathbb{C}$ is linear, i.e. for any $w, w' \in V$ and $\alpha \in \mathbb{C}$

Inner product

Definition

Inner product space V is a vector space V over $\mathbb{C}$ or $\mathbb{R}$, with a function

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C},$$

such that

i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$,

ii) for any $v \in V$ the function $\langle v, \cdot \rangle \colon V \to \mathbb{C}$ is linear, i.e. for any $w, w' \in V$ and $\alpha \in \mathbb{C}$

   a) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$,

Inner product

Definition

Inner product space V is a vector space V over $\mathbb{C}$ or $\mathbb{R}$, with a function

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C},$$

such that

i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$,

ii) for any $v \in V$ the function $\langle v, \cdot \rangle \colon V \to \mathbb{C}$ is linear, i.e. for any $w, w' \in V$ and $\alpha \in \mathbb{C}$

   a) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$,

   b) $\langle v, \alpha w \rangle = \alpha \langle v, w \rangle$

Inner product

Definition

Inner product space V is a vector space V over $\mathbb{C}$ or $\mathbb{R}$, with a function

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C},$$

such that

i) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ for any $v, w \in V$,

ii) for any $v \in V$ the function $\langle v, \cdot \rangle \colon V \to \mathbb{C}$ is linear, i.e. for any $w, w' \in V$ and $\alpha \in \mathbb{C}$

   a) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$,

   b) $\langle v, \alpha w \rangle = \alpha \langle v, w \rangle$

iii) $\langle v, v \rangle > 0$ for any $v \neq \mathbf{0}$.

Example

Example

The vector space $V = \mathbb{C}^n$ with

$$\langle v, w \rangle = \sum_{j=1}^{n} \overline{v_j} w_j = \overline{v}^\intercal w = v^* w,$$

for any $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in \mathbb{C}^n$ is the standard inner product space

Example

Example

The vector space $V = \mathbb{C}^n$ with

$$\langle v, w \rangle = \sum_{j=1}^{n} \overline{v_j} w_j = \overline{v}^\mathsf{T} w = v^* w,$$

for any $v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in \mathbb{C}^n$ is the standard inner product space

Example

The vector space $V = \mathcal{C}([a, b]; \mathbb{C})$ of continuous functions

$$\langle f, g \rangle = \int_a^b w(x) \overline{f(x)} g(x) \, \mathrm{d}x \,,$$

where w is a fixed **weight function** $w \in V$ such that $w(x) \in \mathbb{R}, w(x) > 0$ for $x \in (a, b)$ is an inner product space.

Norm

Let V be an inner product space.

Definition

Norm of vector $v \in V$ is equal to

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

Norm

Let V be an inner product space.

Definition

Norm of vector $v \in V$ is equal to

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

If V is complete as a metric space induced by the norm it is called a Hilbert space.

Adjoint Transformation

Proposition

Let V and W be inner product spaces. For any linear transformation $\varphi\colon V \to W$ there exists a unique linear transformation $\varphi^*\colon W \to V$ such that

$$\langle \varphi(v), w \rangle_W = \langle v, \varphi^*(w) \rangle_V.$$

Adjoint Transformation (continued)

Proof.

The inner products induce isomorphisms $V \simeq V^*$ and $W \simeq W^*$ because the linear transformations are monomorphisms hence isomorphisms (since the product is positive definite),

$$V \ni v \mapsto \langle v, \cdot \rangle \in V^*,$$

$$W \ni w \mapsto \langle w, \cdot \rangle \in W^*.$$

These isomorphisms induce an isomorphism

$$\mathrm{Hom}(V, W) = V^* \otimes W \simeq V \otimes W^* \simeq W^* \otimes V = \mathrm{Hom}(W, V),$$

and φ^* is the image of φ under this isomorphism. $\square$

Adjoint Transformation (continued)

Proof.

Let

$$\varphi = \alpha \otimes t,$$

where $\alpha \in V^*, t \in W$. Let $s_\alpha \in V$ be a vector such that

$$\alpha(\cdot) = \langle s_\alpha, \cdot \rangle,$$

(i.e. vector corresponding to α under isomorphism $V \simeq V^*$). By definition

$$\alpha^* = \langle t, \cdot \rangle \otimes s_\alpha.$$

Then for any $v \in V, w \in W$

$$\langle \varphi(v), w \rangle = \langle \alpha(v) t, w \rangle = \overline{\alpha(v)} \langle t, w \rangle.$$

On the other hand

$$\langle v, \varphi^*(w) \rangle = \langle v, \langle t, w \rangle s_\alpha \rangle = \langle t, w \rangle \langle v, s_\alpha \rangle = \langle t, w \rangle \overline{\alpha(v)}.$$

Adjoint Transformation (continued)

Proposition

Let $\varphi \colon \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation where $\mathbb{C}^n$ is a standard inner product space (domain and codomain). If $A = M_{st}^{st}(\varphi)$ then

$$A^* = \overline{A}^{\mathsf{T}} = M_{st}^{st}(\varphi^*),$$

where

$$A^* = \overline{A}^{\mathsf{T}} = \overline{A^{\mathsf{T}}}.$$

Adjoint Transformation (continued)

Proposition

Let $\varphi \colon \mathbb{C}^n \to \mathbb{C}^n$ *be a linear transformation where* $\mathbb{C}^n$ *is a standard inner product space (domain and codomain). If* $A = M^{st}_{st}(\varphi)$ *then*

$$A^* = \overline{A}^\intercal = M^{st}_{st}(\varphi^*),$$

where

$$A^* = \overline{A}^\intercal = \overline{A^\intercal}.$$

Proof.

For any $v = (v_1, \dots, v_n), w = (w_1, \dots, w_n) \in \mathbb{C}^n$

$$\langle Av, w\rangle = (\overline{A}\overline{v})^\intercal w = \overline{v}^\intercal(\overline{A}^\intercal w) = \overline{v}^\intercal A^* w = \langle v, A^* w\rangle.$$

$\square$

Normal, Unitary, Hermitian and Skew–Hermitian Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{C})$ is **normal** if

$$A^*A = AA^*.$$

Normal, Unitary, Hermitian and Skew–Hermitian Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{C})$ is **normal** if

$$A^* A = AA^*.$$

Definition

Matrix $U \in M(n \times n; \mathbb{C})$ is **unitary** if

$$U^* U = UU^* = I_n.$$

Normal, Unitary, Hermitian and Skew–Hermitian Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{C})$ is **normal** if

$$A^* A = AA^*.$$

Definition

Matrix $U \in M(n \times n; \mathbb{C})$ is **unitary** if

$$U^* U = UU^* = I_n.$$

Definition

Matrix $H \in M(n \times n; \mathbb{C})$ is **Hermitian** if

$$H = H^*.$$

Matrix $H \in M(n \times n; \mathbb{C})$ is **skew–Hermitian** if

$$H = -H^*.$$

Normal, Unitary, Hermitian and Skew–Hermitian Matrix (continued)

Proposition

Unitary, Hermitian and skew–Hermitian matrices are normal.

Normal Transformation

Let V be an inner product space.

Definition

Endomorphism (linear transformation)

$$\varphi \colon V \to V,$$

is **normal** if

$$\varphi \circ \varphi^* = \varphi^* \circ \varphi.$$

Normal Transformation

Let V be an inner product space.

Definition

Endomorphism (linear transformation)

$$\varphi \colon V \to V,$$

is **normal** if

$$\varphi \circ \varphi^* = \varphi^* \circ \varphi.$$

Proposition

Endomorphism φ is normal if and only if the matrix $M_{\mathcal{A}}^{\mathcal{A}}(\varphi)$ is normal for any (some) orthonormal basis $\mathcal{A}$ of V.

Normal Transformation

Let V be an inner product space.

Definition

Endomorphism (linear transformation)

$$\varphi \colon V \to V,$$

is **normal** if

$$\varphi \circ \varphi^* = \varphi^* \circ \varphi.$$

Proposition

Endomorphism φ is normal if and only if the matrix $M_{\mathcal{A}}^{\mathcal{A}}(\varphi)$ is normal for any (some) orthonormal basis $\mathcal{A}$ of V.

Proof.

Exercise. $\square$

Normal Matrix is Unitary Diagonalizable

Proposition

Let $A \in M(n \times n; \mathbb{C})$ *be normal matrix. Then there exists a unitary matrix* $U \in M(n \times n; \mathbb{C})$ *such that the matrix*

$$U^* A U = U^{-1} A U,$$

is diagonal.

Normal Matrix is Unitary Diagonalizable

Proposition

Let $A \in M(n \times n; \mathbb{C})$ be normal matrix. Then there exists a unitary matrix $U \in M(n \times n; \mathbb{C})$ such that the matrix

$$U^* AU = U^{-1} AU,$$

is diagonal.

Proof.

Let $\lambda \in \mathbb{C}, v \in \mathbb{C}^n$. Then

$$\|Av - \lambda v\|^2 = \langle Av, Av \rangle - \langle \lambda v, Av \rangle - \langle Av, \lambda v \rangle + \langle \lambda v, \lambda v \rangle =$$

$$= \langle v, A^* Av \rangle - \langle A^* v, \overline{\lambda} v \rangle - \langle \overline{\lambda} v, A^* v \rangle + \langle \overline{\lambda} v, \overline{\lambda} v \rangle =$$

$$= \langle v, AA^* v \rangle - \langle A^* v, \overline{\lambda} v \rangle - \langle \overline{\lambda} v, A^* v \rangle + \langle \overline{\lambda} v, \overline{\lambda} v \rangle =$$

$$= \left\| A^* v - \overline{\lambda} v \right\|^2.$$

Therefore $v \in \mathbb{C}^n$ is an eigenvector of A if and only if it is an eigenvector of A^* (and the corresponding eigenvalues are conjugated).

Normal Matrix is Unitary Diagonalizable (continued)

Proof.

Let $\lambda \in \mathbb{C}$ be an eigenvalue of A and let $v \in \mathbb{C}^n$ be a corresponding eigenvector of norm 1. Let

$$V = \mathrm{lin}(v)^{\perp} = \{w \in W \mid \langle v, w \rangle = 0\}.$$

Then

$$AV \subset V,$$

since for $w \in V$

$$\langle v, Aw \rangle = \langle A^* v, w \rangle = \langle \overline{\lambda} v, w \rangle = 0.$$

The endomorphism $A|_V$ is normal (since $(\varphi|_V)^* = (\varphi^*)|_V$) and by the induction the theorem holds. The unitary matrix $U \in M(n \times n; \mathbb{C})$ has in columns normalized (i.e. of length 1) eigenvectors obtained by the above procedure. $\square$

Characterization of Complex Normal, Unitary, Hermitian and Skew–Hermitian Matrices

Let $A \in M(n \times n; \mathbb{C})$ be a matrix with (possibly repeating) eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. Let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in M(n \times n; \mathbb{C})$ be a diagonal matrix with complex numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ on the diagonal.

Proposition

Then

$$A \text{ is normal} \Leftrightarrow$$

$$\Leftrightarrow \text{ there exists unitary matrix } U \in M(n \times n; \mathbb{C}) \text{ such that } U^*AU = D.$$

Moreover

i) *matrix A is unitary $\Leftrightarrow$ $|\lambda_i| = 1$ for $j = 1, \ldots, n$,*

ii) *matrix A is Hermitian $\Leftrightarrow$ $\lambda_i \in \mathbb{R}$ for $j = 1, \ldots, n$,*

iii) *matrix A is skew–Hermitian $\Leftrightarrow$ $\lambda_i \in \sqrt{-1}\mathbb{R}$ for $j = 1, \ldots, n$,*

Characterization of Complex Normal, Unitary, Hermitian and Skew–Hermitian Matrices (continued)

Proof.

Easy exercise. Respectively, one has

i) $D^* = D^{-1}$,

ii) $D^* = D$,

iii) $D^* = -D$.

$\square$

Normal, Orthogonal, Symmetric and Skew–Symmetric Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{R})$ is **normal** if

$$A^\intercal A = AA^\intercal.$$

Normal, Orthogonal, Symmetric and Skew–Symmetric Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{R})$ is **normal** if

$$A^\intercal A = AA^\intercal.$$

Definition

Matrix $Q \in M(n \times n; \mathbb{R})$ is **orthogonal** if

$$Q^\intercal Q = QQ^\intercal = I_n.$$

Normal, Orthogonal, Symmetric and Skew–Symmetric Matrix

Definition

Matrix $A \in M(n \times n; \mathbb{R})$ is **normal** if

$$A^\intercal A = AA^\intercal.$$

Definition

Matrix $Q \in M(n \times n; \mathbb{R})$ is **orthogonal** if

$$Q^\intercal Q = QQ^\intercal = I_n.$$

Definition

Matrix $H \in M(n \times n; \mathbb{R})$ is **symmetric** if

$$H = H^\intercal.$$

Matrix $H \in M(n \times n; \mathbb{R})$ is **skew–symmetric** if

$$H = -H^\intercal.$$

Normal, Orthogonal, Symmetric and Skew–Symmetric Matrix (continued)

Proposition

Orthogonal, symmetric and skew-symmetric real matrices are normal.

Characterization of Real Normal Matrices

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Then

i) λ *is an eigenvalue of* $A \iff \overline{\lambda}$ *is an eigenvalue of* A,

ii) $v = \operatorname{Re} v + i \operatorname{Im} v$ *is an eigenvector for the eigenvalue* λ *of* $A \iff \overline{v} = \operatorname{Re} v - i \operatorname{Im} v$ *is an eigenvector for the eigenvalue* $\overline{\lambda}$ *of* A.

Characterization of Real Normal Matrices (continued)

Proof.

The characteristic polynomial of A has real coefficients hence its strictly complex roots form pairs $\lambda, \overline{\lambda}$. Let $\lambda = a + bi$ where $a, b \in \mathbb{R}$.

$$Av = \lambda v \Leftrightarrow A(\operatorname{Re} v + i \operatorname{Im} v) = (a + bi)(\operatorname{Re} v + i \operatorname{Im} v) \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} A \operatorname{Re} v = a \operatorname{Re} v - b \operatorname{Im} v \\ A \operatorname{Im} v = b \operatorname{Re} v + b \operatorname{Im} v \end{cases},$$

where the right–hand side remains invariant under changing the sign of b and $\operatorname{Im} v$. $\square$

Characterization of Real Normal Matrices (continued)

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Let $v, w \in \mathbb{C}^n$ be two complex eigenvectors corresponding, respectively, to eigenvalues λ and μ of A. Assume $\overline{\lambda} \neq \mu$. Then

$$(\mathrm{Re}\, v)\cdot(\mathrm{Re}\, w) = (\mathrm{Im}\, v)\cdot(\mathrm{Im}\, w) = (\mathrm{Re}\, v)\cdot(\mathrm{Im}\, w) = (\mathrm{Im}\, v)\cdot(\mathrm{Re}\, w) = 0.$$

Characterization of Real Normal Matrices (continued)

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Let $v, w \in \mathbb{C}^n$ be two complex eigenvectors corresponding, respectively, to eigenvalues λ and μ of A. Assume $\overline{\lambda} \neq \mu$. Then

$$(\operatorname{Re} v)\cdot(\operatorname{Re} w) = (\operatorname{Im} v)\cdot(\operatorname{Im} w) = (\operatorname{Re} v)\cdot(\operatorname{Im} w) = (\operatorname{Im} v)\cdot(\operatorname{Re} w) = 0.$$

Proof.

Assume $w \notin \mathbb{R}$. Then $w, \overline{w}$ are eigenvectors of A, both unitary orthogonal to v.

$$\begin{cases} \langle \operatorname{Re} v + i \operatorname{Im} v, \operatorname{Re} w + i \operatorname{Im} w \rangle = 0 \\ \langle \operatorname{Re} v + i \operatorname{Im} v, \operatorname{Re} w - i \operatorname{Im} w \rangle = 0 \end{cases} \Leftrightarrow$$

$$\Leftrightarrow \begin{cases} (\langle \operatorname{Re} v, \operatorname{Re} w \rangle + \langle \operatorname{Im} v, \operatorname{Im} w \rangle) + i(\langle \operatorname{Im} v, \operatorname{Re} w \rangle - \langle \operatorname{Re} v, \operatorname{Im} w \rangle) = 0 \\ (\langle \operatorname{Re} v, \operatorname{Re} w \rangle - \langle \operatorname{Im} v, \operatorname{Im} w \rangle) + i(\langle \operatorname{Im} v, \operatorname{Re} w \rangle + \langle \operatorname{Re} v, \operatorname{Im} w \rangle) = 0 \end{cases}$$

Characterization of Real Normal Matrices (continued)

Proof.

If $\lambda, \mu \in \mathbb{R}$ then they are different, and $(\operatorname{Re} v) \cdot (\operatorname{Re} w) = 0$ since $v = \operatorname{Re} v, w = \operatorname{Re} w$ are real and unitary orthogonal.

Characterization of Real Normal Matrices (continued)

Proof.

If $\lambda, \mu \in \mathbb{R}$ then they are different, and $(\operatorname{Re} v) \cdot (\operatorname{Re} w) = 0$ since $v = \operatorname{Re} v, w = \operatorname{Re} w$ are real and unitary orthogonal. If $\lambda \in \mathbb{R}$ and $\mu \notin \mathbb{R}$ then the above proof works as well. $\square$

Characterization of Real Normal Matrices (continued)

Proof.

If $\lambda, \mu \in \mathbb{R}$ then they are different, and $(\mathrm{Re}\, v) \cdot (\mathrm{Re}\, w) = 0$ since $v = \mathrm{Re}\, v, w = \mathrm{Re}\, w$ are real and unitary orthogonal. If $\lambda \in \mathbb{R}$ and $\mu \notin \mathbb{R}$ then the above proof works as well. $\square$

Corollary

If $v, w \in \mathbb{C}$ are complex eigenvectors for the strictly complex eigenvalue λ, and $\langle v, w \rangle = 0$ (i.e. unitary orthogonal) then

$$(\mathrm{Re}\, v) \cdot (\mathrm{Re}\, w) = (\mathrm{Im}\, v) \cdot (\mathrm{Im}\, w) = (\mathrm{Re}\, v) \cdot (\mathrm{Im}\, w) = (\mathrm{Im}\, v) \cdot (\mathrm{Re}\, w) = 0.$$

Characterization of Real Normal Matrices (continued)

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Let $v \in \mathbb{C}^n$ be a **unit** *complex eigenvector corresponding to a strictly complex eigenvalue $\lambda \notin \mathbb{R}$. Then*

$$(\operatorname{Re} v) \cdot (\operatorname{Im} v) = 0,$$

and

$$\|\operatorname{Re} v\| = \|\operatorname{Im} v\| = \frac{1}{\sqrt{2}}.$$

Characterization of Real Normal Matrices (continued)

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Let $v \in \mathbb{C}^n$ be a **unit** *complex eigenvector corresponding to a strictly complex eigenvalue $\lambda \notin \mathbb{R}$. Then*

$$(\mathrm{Re}\, v) \cdot (\mathrm{Im}\, v) = 0,$$

and

$$\|\mathrm{Re}\, v\| = \|\mathrm{Im}\, v\| = \frac{1}{\sqrt{2}}.$$

Proof.

Then $\overline{v}$ is a unit eigenvector, unitary orthogonal to v

$$0 = \langle \mathrm{Re}\, v + i\, \mathrm{Im}\, v, \mathrm{Re}\, v - i\, \mathrm{Im}\, v \rangle =$$

$$= (\langle \mathrm{Re}\, v, \mathrm{Re}\, w \rangle - \langle \mathrm{Im}\, v, \mathrm{Im}\, w \rangle) + 2i \langle \mathrm{Re}\, v, \mathrm{Im}\, w \rangle,$$

moreover

$$1 = \|v\|^2 = \|\mathrm{Re}\, v\|^2 + \|\mathrm{Im}\, v\|^2.$$

$\square$

Characterization of Real Normal Matrices (continued)

Corollary

Let $A \in M(n \times n; \mathbb{R})$ be a normal matrix. Let $\mu_1, \dots, \mu_m \in \mathbb{R}$ be (possibly repeating) real eigenvalues of A. Let $\lambda_1, \overline{\lambda_1}, \lambda_2, \overline{\lambda_2}, \dots, \lambda_k, \overline{\lambda_k} \in \mathbb{C}$ be (possibly repeating) strictly complex eigenvalues of A, where $\alpha_j = a_j + ib_j$ for $j = 1, \dots, k$. Let $u_1, \dots, u_m, v_1, \overline{v_1}, v_2, \overline{v_2}, \dots, v_k, \overline{v_k} \in \mathbb{C}^n$ be the corresponding unitary orthonormal basis of $\mathbb{C}^n$, consisting of the corresponding eigenvectors, such that $u_j = \operatorname{Re} u_j$ for $j = 1, \dots, m$. Then

$$\mathcal{A} = \Big(u_1, \dots, u_k,$$

$$\sqrt{2}\operatorname{Re} v_1, \sqrt{2}\operatorname{Im} v_1, \sqrt{2}\operatorname{Re} v_2, \sqrt{2}\operatorname{Im} v_2, \dots, \sqrt{2}\operatorname{Re} v_k, \sqrt{2}\operatorname{Im} v_k\Big),$$

is an real orthogonal basis of $\mathbb{R}^n$.

Characterization of Real Normal Matrices (continued)

Corollary

Moreover, if $Q = M(\mathrm{id})^{st}_{\mathcal{A}}$ *then* $Q \in M(n \times n; \mathbb{R})$ *is an (real) orthogonal matrix (i.e.* $Q^\intercal Q = QQ^\intercal = I$*) and*

$$Q^\intercal A Q = \left[\begin{array}{c|c|c|c|cc|cc|c|cc}
\mu_1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & \mu_2 & \cdots & 0 & 0 & 0 & 0 & 0 & & 0 & 0 \\ \hline
\vdots & & \ddots & & & & & & & \vdots & \vdots \\ \hline
0 & & & \mu_m & & & & & & 0 & 0 \\ \hline
0 & & & & a_1 & b_1 & 0 & 0 & \cdots & 0 & 0 \\
0 & & & & -b_1 & a_1 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & & & & 0 & 0 & a_2 & b_2 & & 0 & 0 \\
0 & & & & 0 & 0 & -b_2 & a_2 & & 0 & 0 \\ \hline
\vdots & & & & \vdots & \vdots & & & \ddots & \vdots & \vdots \\ \hline
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & a_k & b_k \\
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & -b_k & a_k
\end{array}\right].$$

Characterization of Real Orthogonal Matrices (continued)

Corollary

Matrix $A \in M(n \times n; \mathbb{R})$ is orthogonal if and only if there exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ and numbers $\varphi_1, \ldots, \varphi_k \in \mathbb{R}$ such that

$$
Q^\intercal A Q = \left[\begin{array}{c|c|c|c|cc|cc|c|cc}
\pm 1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & \pm 1 & & & 0 & 0 & 0 & 0 & & 0 & 0 \\ \hline
\vdots & & \ddots & & & & & & & \vdots & \vdots \\ \hline
0 & & & \pm 1 & & & & & & 0 & 0 \\ \hline
0 & & & & \cos\varphi_1 & \sin\varphi_1 & 0 & 0 & \cdots & 0 & 0 \\
0 & & & & -\sin\varphi_1 & \cos\varphi_1 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & & & & 0 & 0 & \cos\varphi_2 & \sin\varphi_2 & & 0 & 0 \\
0 & & & & 0 & 0 & -\sin\varphi_2 & \cos\varphi_2 & & 0 & 0 \\ \hline
\vdots & & & & \vdots & \vdots & & & \ddots & \vdots & \vdots \\ \hline
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & \cos\varphi_k & \sin\varphi_k \\
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & -\sin\varphi_k & \cos\varphi_k
\end{array}\right].
$$

Characterization of Real Symmetric Matrices (continued)

Corollary

Matrix $A \in M(n \times n; \mathbb{R})$ is symmetric if and only if there exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ and numbers $\mu_1, \dots, \mu_n \in \mathbb{R}$ such that

$$Q^\intercal A Q = \begin{bmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{bmatrix}.$$

Characterization of Real Skew–Symetric Normal Matrices

Corollary

Matrix $A \in M(n \times n; \mathbb{R})$ is skew–symmetric if and only if there exists an orthogonal matrix $Q \in M(n \times n; \mathbb{R})$ and numbers $b_1, \ldots, b_k \in \mathbb{R}$ such that

$$Q^\intercal A Q = \left[\begin{array}{c|c|c|c|cc|cc|c|cc}
0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & 0 & & & 0 & 0 & 0 & 0 & & 0 & 0 \\ \hline
\vdots & & \ddots & & & & & & & \vdots & \vdots \\ \hline
0 & & & 0 & & & & & & 0 & 0 \\ \hline
0 & & & & 0 & b_1 & 0 & 0 & \cdots & 0 & 0 \\
0 & & & & -b_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline
0 & & & & 0 & 0 & 0 & b_2 & & 0 & 0 \\
0 & & & & 0 & 0 & -b_2 & 0 & & 0 & 0 \\ \hline
\vdots & & & & \vdots & \vdots & & & \ddots & \vdots & \vdots \\ \hline
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & 0 & b_k \\
0 & 0 & 0 & & 0 & 0 & 0 & 0 & \cdots & -b_k & 0
\end{array}\right].$$

Example

Let

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

Then

$$A^{\intercal}A = AA^{\intercal} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

Moreover

$$w_A(x) = \det(A - xI) = -x^3 + 3x^2 - 3x + 2 = -(x-2)\left(x^2 - x + 1\right),$$

therefore

$$\mu = 2, \quad \lambda = \frac{1 + i\sqrt{3}}{2} = e^{\frac{i\pi}{3}}, \quad \overline{\lambda} = \frac{1 - i\sqrt{3}}{2} = e^{-\frac{i\pi}{3}}.$$

It can be checked that

$$V_{(\mu)} = \mathrm{lin}((1,1,1)), \quad V_{(\lambda)} = \mathrm{lin}((1,\lambda^2,-\lambda)), \quad V_{(\overline{\lambda})} = \mathrm{lin}((1,-\lambda,\lambda^2)),$$

(note that $\lambda^3 + 1 = 0, \lambda^2 = \lambda - 1, \overline{\lambda} = \frac{1}{\lambda} = -\lambda^2$).

Example (continued)

Since

$$|(1,1,1)| = \left|(1,\lambda^2,-\lambda)\right| = \left|(1,-\lambda,\lambda^2)\right| = \sqrt{3},$$

we have

$$u_1 = \frac{1}{\sqrt{3}}(1,1,1),$$

$$v_1 = \frac{1}{\sqrt{3}}(1,\lambda^2,-\lambda),$$

$$\overline{v}_1 = \frac{1}{\sqrt{3}}(1,-\lambda,\lambda^2).$$

If

$$U = \frac{1}{\sqrt{3}}\left[\begin{array}{ccc} 1 & 1 & 1 \\ 1 & \lambda^2 & -\lambda \\ 1 & -\lambda & \lambda^2 \end{array}\right],$$

Then

$$U^* U = UU^* = I.$$

Example (continued)

If

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \overline{\lambda} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -\lambda^2 \end{bmatrix},$$

then

$$UDU^* = A,$$

i.e.

$$\frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \lambda^2 & -\lambda \\ 1 & -\lambda & \lambda^2 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & -\lambda^2 \end{bmatrix} \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -\lambda & \lambda^2 \\ 1 & \lambda^2 & -\lambda^2 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Example (continued)

Let

$$u_1 = v_1 = \frac{1}{\sqrt{3}}(1,1,1),$$

$$u_2 = \sqrt{2}\,\mathrm{Re}\, v_1 = \frac{\sqrt{2}}{\sqrt{3}}\left(1, -\frac{1}{2}, -\frac{1}{2}\right),$$

$$u_3 = \sqrt{2}\,\mathrm{Im}\, v_1 = \frac{\sqrt{2}}{\sqrt{3}}\left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}\right).$$

If

$$Q = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & \sqrt{2} & 0 \\ 1 & -\frac{1}{\sqrt{2}} & \frac{\sqrt{6}}{2} \\ 1 & -\frac{1}{\sqrt{2}} & -\frac{\sqrt{6}}{2} \end{bmatrix},$$

then

$$Q^{\mathsf{T}} Q = Q Q^{\mathsf{T}} = I.$$

Example (continued)

Moreover, let

$$B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Then

$$QBQ^\mathsf{T} = A,$$

i.e.

$$\frac{1}{\sqrt{3}} \begin{bmatrix} 1 & \sqrt{2} & 0 \\ 1 & -\frac{1}{\sqrt{2}} & \frac{\sqrt{6}}{2} \\ 1 & -\frac{1}{\sqrt{2}} & -\frac{\sqrt{6}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 & 0 \\ \sqrt{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & -\frac{\sqrt{6}}{2} \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Rayleigh Quotient

Definition

For any matrix $M \in M(n \times n; \mathbb{C})$ and any vector $x \in \mathbb{C}^n, x \neq \mathbf{0}$, the **Rayleigh quotient** $R(M, x)$ is equal to

$$R(M, x) = \frac{x^* M x}{x^* x}.$$

Rayleigh Quotient

Definition

For any matrix $M \in M(n \times n; \mathbb{C})$ and any vector $x \in \mathbb{C}^n, x \neq \mathbf{0}$, the **Rayleigh quotient** $R(M, x)$ is equal to

$$R(M, x) = \frac{x^* M x}{x^* x}.$$

Proposition

For any complex number $\alpha \in \mathbb{C}$

$$R(M, \alpha x) = R(M, x).$$

Rayleigh Quotient (continued)

Proposition

For any Hermitian matrix $M \in M(n \times n; \mathbb{C})$ (i.e., $M^* = M$)

$$R(M, x) \in \mathbb{R},$$

$$\lambda_{min} \leqslant R(M, x) \leqslant \lambda_{max},$$

where $\lambda_{min}, \lambda_{max} \in \mathbb{R}$ are the smallest and the greatest (real) eigenvalues of matrix M. Moreover, those bounds are attained by $R(M, x)$ by the corresponding eigenvectors $x \in \mathbb{C}^n$.

Rayleigh Quotient (continued)

Proposition

For any Hermitian matrix $M \in M(n \times n; \mathbb{C})$ *(i.e.,* $M^* = M$*)*

$$R(M, x) \in \mathbb{R},$$

$$\lambda_{min} \leqslant R(M, x) \leqslant \lambda_{max},$$

where $\lambda_{min}, \lambda_{max} \in \mathbb{R}$ *are the smallest and the greatest (real) eigenvalues of matrix* M*. Moreover, those bounds are attained by* $R(M, x)$ *by the corresponding eigenvectors* $x \in \mathbb{C}^n$*.*

Proof.

As $R(M, x)^* = R(M, x)$, it follows that $R(M, x) \in \mathbb{R}$. Let $v_1, \ldots, v_n \in \mathbb{C}^n$ be a unitary orthonormal basis of $\mathbb{C}^n$, in which matrix of M is diagonal (i.e., it consist of eigenvectors v_i of matrix M such that $Mv_i = \lambda_i v_i$ and $v_j^\intercal M v_i = 0$ for $i \neq j$). Let

$$x = \alpha_1 v_1 + \ldots + \alpha_n v_n.$$

Rayleigh Quotient (continued)

Proof.

Then

$$R(M,x) = \frac{\sum_{i=1}^{n} \lambda_i |\alpha_i|^2}{\sum_{i=1}^{n} |\alpha_i|^2}.$$

Since $\lambda_{min} \leqslant \lambda_i \leqslant \lambda_{max}$, it follows that

$$\lambda_{min} \leqslant \frac{\sum_{i=1}^{n} \lambda_i |\alpha_i|^2}{\sum_{i=1}^{n} |\alpha_i|^2} \leqslant \lambda_{max}.$$

The bounds are attained for $x = v_i$ where $Mv_i = \lambda_{min} v_i$ and for $x = v_j$ where $Mv_j = \lambda_{max} v_j$. □

Rayleigh Quotient (continued)

Proposition

For any matrix $M \in M(n \times n; \mathbb{C})$ and any vector $x \in \mathbb{C}^n$ the Rayleigh quotient

$$\lambda = R(M, x) = \frac{x^* M x}{x^* x},$$

is the least square solution of the (possibly inconsistent) equation

$$Mx = \lambda x.$$

Rayleigh Quotient (continued)

Proposition

For any matrix $M \in M(n \times n; \mathbb{C})$ and any vector $x \in \mathbb{C}^n$ the Rayleigh quotient

$$\lambda = R(M, x) = \frac{x^* M x}{x^* x},$$

is the least square solution of the (possibly inconsistent) equation

$$Mx = \lambda x.$$

Proof.

The orthogonal projection of Mx onto $V = \mathrm{lin}(x)$ is equal to

$$P_V(Mx) = \frac{x^*(Mx)}{x^* x} x.$$

Rayleigh Quotient (continued)

Proposition

For any fixed symmetric matrix $M = M^{\intercal} \in M(n \times n; \mathbb{R})$ *the eigenvectors of* M *are stationary points of the Rayleigh quotient, that is if* $Mx = \lambda x$ *for some* $x \in \mathbb{R}^n, x \neq \mathbf{0}$ *then*

$$\boldsymbol{\nabla}_x R(M, x) = \mathbf{0}.$$

Rayleigh Quotient (continued)

Proposition

For any fixed symmetric matrix $M = M^\intercal \in M(n \times n; \mathbb{R})$ *the eigenvectors of* M *are stationary points of the Rayleigh quotient, that is if* $Mx = \lambda x$ *for some* $x \in \mathbb{R}^n, x \neq \mathbf{0}$ *then*

$$\boldsymbol{\nabla}_x R(M, x) = \mathbf{0}.$$

Proof.

$$\frac{\partial R}{\partial x_j}(M, x) = \frac{\frac{\partial}{\partial x_j}(x^\intercal M x)(x^\intercal x) - (x^\intercal M x)\frac{\partial}{\partial x_j}(x^\intercal x)}{(x^\intercal x)^2} =$$

$$= \frac{2(Mx)_j(x^\intercal x) - (x^\intercal M x)2x_j}{(x^\intercal x)^2} =$$

$$= \frac{2}{x^\intercal x}\left(Mx - R(M, x)x\,\right)_j,$$

where $(Mx)_j$ denotes the j-th entry of the vector Mx. $\square$

Eigenvalue Decomposition

Proposition

Let $M \in M(n \times n; \mathbb{C})$ *be a matrix such that there exists basis* $\mathcal{A} = (v_1, \ldots, v_n)$ *of* $\mathbb{C}^n$ *and numbers* $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ *such that*

$$M = CDC^*,$$

where $C = M(\mathrm{id})^{st}_{\mathcal{A}}$ *and* $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. *Then*

$$M = \sum_{i=1}^{n} \lambda_i v_i v_i^*,$$

where $v_i v_i^*$ *are rank* 1 *matrices.*

Eigenvalue Decomposition (continued)

Proof.

$$M = \sum_{i=1}^{n} CD_i C^* = \left(\sum_{i=1}^{n} \lambda_i C_i \right) C^* = \sum_{i=1}^{n} \lambda_i C_i C^*,$$

where $D_i = \mathrm{diag}(0, \ldots, 0, \lambda_i, 0, \ldots, 0)$ and $C_i \in M(n \times n; \mathbb{C})$ is a zero matrix with $i-$th column replaced with eigenvector v_i. Then

$$C_i C^* = C_i C_i^* = v_i v_i^*.$$

$\square$

Eigenvalue Decomposition (continued)

Corollary

Let $M \in M(n \times n; \mathbb{C})$ *be a Hermitian matrix (i.e.* $M^* = M$*). Let* $v_1, \ldots, v_n \in \mathbb{C}^n$ *be a unitary orthonormal basis consisting of eigenvectors of* M *corresponding to eigenvalues* $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$*. Then*

$$M = \sum_{i=1}^{n} \lambda_i v_i v_i^*.$$

Sherman–Morrison Formula

The following formula expresses the inverse of rank 1 update of matrix A.

Proposition

For any matrix invertible $A \in M(n \times n; \mathbb{C})$ *and vectors* $v, w \in \mathbb{C}$ *such that* $1 + w^* A v \neq 0$ *the matrix* $A + vw^*$ *is invertible and*

$$(A + vw^*)^{-1} = A^{-1} - \frac{A^{-1} v w^* A^{-1}}{1 + w^* A^{-1} v}.$$

Sherman–Morrison Formula

The following formula expresses the inverse of rank 1 update of matrix A.

Proposition

For any matrix invertible $A \in M(n \times n; \mathbb{C})$ and vectors $v, w \in \mathbb{C}$ such that $1 + w^*Av \neq 0$ the matrix $A + vw^*$ is invertible and

$$(A + vw^*)^{-1} = A^{-1} - \frac{A^{-1}vw^*A^{-1}}{1 + w^*A^{-1}v}.$$

Proof.

First we show that

$$(I + uw^*)^{-1} = I - \frac{uw^*}{1 + w^*u}.$$

$$(I + uw^*)\left(I - \frac{uw^*}{1 + w^*u}\right) =$$

$$= I - \frac{uw^*}{1 + w^*u} + uw^* - w^*u\frac{uw^*}{1 + w^*u} = I.$$

Sherman–Morrison Formula (continued)

Proof.

Since A is invertible there exists $u \in \mathbb{C}^n$ such that $v = Au$, i.e. $u = A^{-1}v$. Then

$$A + vw^* = A(I + uw^*),$$

and the matrix $A + vw^*$ is invertible if and only if the matrix $I + uw^*$ is invertible. Moreover

$$(A + vw^*)^{-1} = (I + uw^*)^{-1}A^{-1} = \left(I - \frac{uw^*}{1 + w^*u}\right)A^{-1} =$$

$$= A^{-1} - \frac{A^{-1}vw^*A^{-1}}{1 + w^*A^{-1}v}.$$

$\square$

Singular Value Decomposition – SVD

Theorem

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exist unitary matrices* $U \in M(m \times m; \mathbb{C}), V \in M(n \times n; \mathbb{C})$ *and a unique (real) generalized diagonal matrix* $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R})$ *such that*

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0,$$

where $r = r(A)$ *oraz*

$$A = U \Sigma V^*.$$

Singular Value Decomposition – SVD

Theorem

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exist unitary matrices* $U \in M(m \times m; \mathbb{C}), V \in M(n \times n; \mathbb{C})$ *and a unique (real) generalized diagonal matrix* $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R})$ *such that*

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0,$$

where $r = r(A)$ *oraz*

$$A = U \Sigma V^*.$$

Remark

Matrices U, V *are not uniquely determined (unlike the matrix* Σ*).*

Singular Value Decomposition – SVD (continued)

Proof.

Let $\sigma_1 = \|A\|_2$. By the definition of $\|\cdot\|_2$ and the compactness of a ball in $\mathbb{C}^m$ there exist vectors $v_1 \in \mathbb{C}^m$ and vectors $u_1 \in \mathbb{C}^n$ such that $\|v_1\|_2 = \|u_1\|_2 = 1$, and

$$Av_1 = \sigma_1 u_1.$$

Let $V_1 \in M(n \times n; \mathbb{C})$ be a unitary matrix with the first column equal to vector v_1, and let $U_1 \in M(m \times m; \mathbb{C})$ be a unitary matrix with first column equal to u_1. Then

$$U_1^* A V_1 = \begin{bmatrix} \sigma_1 & w^* \\ \mathbf{0} & B \end{bmatrix},$$

where $w \in \mathbb{C}^{n-1}$ and $B \in M((m-1) \times (n-1); \mathbb{C})$. □

[0] see L. N. Trefethen, D. Bau, III, *Numerical Linear Algebra*, SIAM

Singular Value Decomposition – SVD (continued)

Proof.

Then

$$\left\| \begin{bmatrix} \sigma_1 & w^* \\ \mathbf{0} & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 \geqslant \sigma_1^2 + w^* w = \sqrt{\sigma_1^2 + w^* w} \left\| \begin{bmatrix} \sigma_1 \\ w \end{bmatrix} \right\|_2 .$$

It follows that $w = \mathbf{0}$, otherwise σ_1 is not maximal. By the inductive assumption there exists unitary matrices $V_2 \in M((n-1) \times (n-1); \mathbb{C})$ and $U_2 \in M((m-1) \times (m-1); \mathbb{C})$ such that

$$A = U_1 \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0} \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V_2 \end{bmatrix}^* V_1^* .$$

□

Singular Value Decomposition – SVD (continued)

Proof.

To prove uniqueness of Σ, assume there exists a vector w corresponding to the singular value σ_1, such that v_1, w are linearly independent (i.e., $\|Aw\|_2 = \sigma_1$) such that $\|w\|_2 = 1$ (otherwise the subspace $\mathrm{lin}(v_1)^\perp$ is uniquely determined). Then the vector

$$v_2 = \frac{w - (v_1^* w)v_1}{\|w - (v_1^* w)v_1\|_2},$$

equal to the unit vector of the projection of vector w onto the subspace $\mathrm{lin}(v_1)^\perp \subset \mathbb{C}^m$, satisfies the condition

$$w = \alpha v_1 + \beta v_2,$$

where $|\alpha|^2 + |\beta|^2 = 1$ (vector w is a unit vector and vectors v_1, v_2 are orthogonal). $\square$

Singular Value Decomposition – SVD (continued)

Proof.

Then $\|Av\|_2 \leqslant \sigma_1$, and if $\|Av\|_2 < \sigma_1$, then

$$\|Aw\|_2^2 = |\alpha|^2 \|Av_1\|_2^2 + |\beta|^2 \|Av_2\|_2^2 < \sigma_1,$$

which leads to contradiction. Therefore, vector w is a vector corresponding to the singular value σ_1 of matrix B. The claim follow by induction. $\square$

Real Singular Value Decomposition

Theorem

For any matrix $A \in M(m \times n; \mathbb{R})$ *there exists orthogonal matrices* $U \in M(m \times m; \mathbb{R}), V \in M(n \times n; \mathbb{C})$ *and a uniquely determined generalized diagonal matrix* $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R})$ *such that*

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0,$$

where $r = r(A)$ *and*

$$A = U \Sigma V^\intercal.$$

Real Singular Value Decomposition

Theorem

For any matrix $A \in M(m \times n; \mathbb{R})$ there exists orthogonal matrices $U \in M(m \times m; \mathbb{R})$, $V \in M(n \times n; \mathbb{C})$ and a uniquely determined generalized diagonal matrix $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R})$ such that

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0,$$

where $r = r(A)$ and

$$A = U \Sigma V^\intercal.$$

Remark

As before, the orthogonal matrices U, V are not uniquely determined.

Real Singular Value Decomposition (continued)

The following proof, using the spectral theorem, after a slight modification works in the complex case too.

Proof.

Matrix $A^{\intercal}A \in M(n \times n; \mathbb{R})$ is symmetric and positive semidefinite hence there exists orthonormal basis (not uniquely determined) $v_1, \ldots, v_n \in \mathbb{R}^n$ of $\mathbb{R}^n$ consisting of eigenvectors of $A^{\intercal}A$ such that

$$v_i^{\intercal} A^{\intercal} A v_j = \begin{cases} 0 & i \neq j, \\ \lambda_i & i = j, \end{cases}, \quad \text{for} \quad i, j = 1, \ldots, r,$$

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_r > 0,$$

$$\lambda_{r+1} = \ldots \lambda_n = 0,$$

where $\lambda_i \geqslant 0$ is an eigenvalue of $A^{\intercal}A$ corresponding to eigenvector $v_i \in \mathbb{R}^n$ and $r \in \mathbb{N}$ is some natural number such that $1 \leqslant r \leqslant n$.

Real Singular Value Decomposition (continued)

Proof.

Let

$$\sigma_i = \sqrt{\lambda_i}, \quad \text{for} \quad i = 1, \dots, n,$$

and

$$u_i = \frac{1}{\sigma_i} A v_i \in \mathbb{R}^m, \quad \text{for} \quad i = 1, \dots, r.$$

Then

$$u_i^\mathsf{T} u_j = \frac{1}{\sigma_i \sigma_j} v_i^\mathsf{T} A^\mathsf{T} A v_j = \begin{cases} 0 & i \neq j, \\ 1_i & i = j, \end{cases}, \quad \text{for} \quad i, j = 1, \dots, r.$$

Moreover

$$A v_i = \mathbf{0}, \quad \text{for} \quad i = r+1, \dots, n,$$

as $\|A v_i\|^2 = v_i^\mathsf{T} A^\mathsf{T} A v_i = 0$.

Real Singular Value Decomposition (continued)

Proof.

Let $u_1, \dots, u_r, u_{r+1}, \dots, u_m \in \mathbb{R}^m$ be an extension of some orthonormal basis of $imA \subset \mathbb{R}^m$ to some othonormal basis $\mathbb{R}^m$ (both not uniquely determined). Let $U \in M(m \times m; \mathbb{R})$ be an orthogonal matrix which columns are equal to $u_1, \dots, u_m \in \mathbb{R}^m$, respectively and let $V \in M(n \times n; \mathbb{R})$ be an orthogonal matrix which columns are equal to $v_1, \dots, v_n \in \mathbb{R}^n$, respectively. Let

$$\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in M(m \times n; \mathbb{R}).$$

Real Singular Value Decomposition (continued)

Proof.

Then

$$U\Sigma V^{\mathsf{T}} v_i = U\Sigma_i e_i = \sigma_i u_i = \sigma_i \left(\frac{1}{\sigma_i} A v_i \right) = A v_i,$$

for $i = 1, \ldots, r$, and

$$U\Sigma V^{\mathsf{T}} v_i = U\Sigma_i e_i = 0 u_i = \mathbf{0},$$

for $i = r + 1, \ldots, n$. Therefore

$$A = U\Sigma V^{\mathsf{T}},$$

and $r(A) = r$ as $r(\Sigma) = r$ and matrices U, V are non–singular. For the uniqueness of matrix Σ proceed like in the complex case. $\square$

Real Singular Value Decomposition (continued)

Remark

The proof implies that

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^\mathsf{T}.$$

Real Singular Value Decomposition (continued)

Remark

The proof implies that

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^\mathsf{T}.$$

Remark

The preceding proof works after small modification in the complex case.

Pseudoinverse

Definition

With the same notation
$\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in M(m \times n; \mathbb{R})$ set

$$\Sigma^+ = \mathrm{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}, 0, \dots, 0) \in M(n \times m; \mathbb{R}),$$

$$A^+ = V\Sigma^+ U^*.$$

Matrix A^+ is called **pseudoinverse** or **Moore–Penrose pseudoinverse** of A (note that matrix Σ^+ is of the same size as $\Sigma^\intercal$).

Pseudoinverse (continued)

Proposition

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exists at most one matrix* $A^+ \in M(n \times m; \mathbb{C})$ *such that*

i) $AA^+A = A$,

Pseudoinverse (continued)

Proposition

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exists at most one matrix* $A^+ \in M(n \times m; \mathbb{C})$ *such that*

i) $AA^+A = A$,

ii) $A^+AA^+ = A^+$,

Pseudoinverse (continued)

Proposition

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exists at most one matrix* $A^+ \in M(n \times m; \mathbb{C})$ *such that*

i) $AA^+A = A$,

ii) $A^+AA^+ = A^+$,

iii) $(AA^+)^* = AA^+$,

Pseudoinverse (continued)

Proposition

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exists at most one matrix* $A^+ \in M(n \times m; \mathbb{C})$ *such that*

i) $AA^+A = A$,

ii) $A^+AA^+ = A^+$,

iii) $(AA^+)^* = AA^+$,

iv) $(A^+A)^* = A^+A$,

Pseudoinverse (continued)

Proposition

For any matrix $A \in M(m \times n; \mathbb{C})$ *there exists at most one matrix* $A^+ \in M(n \times m; \mathbb{C})$ *such that*

i) $AA^+A = A$,

ii) $A^+AA^+ = A^+$,

iii) $(AA^+)^* = AA^+$,

iv) $(A^+A)^* = A^+A$,

(in particular matrices AA^+, A^+A *are Hermitian). Moreover, matrix*

$$A^+ = V\Sigma^+U^*,$$

satisfies the above conditions.

Pseudoinverse (continued)

Proof.

Let

$$A = U\Sigma V^*,$$

$$A^+ = V\Sigma^+ U^*,$$

be the singular value decomposition of A, where

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R}),$$

$$\Sigma^+ = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0) \in M(n \times m; \mathbb{R}).$$

Then

$$\Sigma\Sigma^+ = \left[\begin{array}{c|c} I_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \in M(m \times m; \mathbb{R}), \quad \Sigma^+\Sigma = \left[\begin{array}{c|c} I_r & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \in M(n \times n; \mathbb{R}).$$

In particular

$$\Sigma\Sigma^+\Sigma = \Sigma, \quad \Sigma^+\Sigma\Sigma^+ = \Sigma^+.$$

Pseudoinverse (continued)

Proof.

Then

i)

$$AA^+A = (U\Sigma V^*)V\Sigma^+U^*(U\Sigma V^*) = U(\Sigma\Sigma^+\Sigma)V^* = A,$$

Pseudoinverse (continued)

Proof.

Then

i)

$$AA^+A = (U\Sigma V^*)V\Sigma^+U^*(U\Sigma V^*) = U(\Sigma\Sigma^+\Sigma)V^* = A,$$

ii)

$$A^+AA^+ = (V\Sigma^+U^*)U\Sigma V^*(V\Sigma^+U^*) = V(\Sigma^+\Sigma\Sigma^+)U^* = A^+,$$

Pseudoinverse (continued)

Proof.

Then

i)

$$AA^+A = (U\Sigma V^*)V\Sigma^+U^*(U\Sigma V^*) = U(\Sigma\Sigma^+\Sigma)V^* = A,$$

ii)

$$A^+AA^+ = (V\Sigma^+U^*)U\Sigma V^*(V\Sigma^+U^*) = V(\Sigma^+\Sigma\Sigma^+)U^* = A^+,$$

iii)

$$\left(A^+A\right)^* = A^*(A^+)^* = (U\Sigma V^*)^*(V\Sigma^+U^*)^* =$$

$$= (V\Sigma^*U^*)(U(\Sigma^+)^*V^*) = V(\Sigma^+\Sigma)^*V^* = V(\Sigma^+\Sigma)V^* =$$

$$= (V\Sigma^+U^*)(U\Sigma V^*) = V(\Sigma^+\Sigma)V^* = A^+A.$$

Pseudoinverse (continued)

Proof.

Then

i)

$$AA^+A = (U\Sigma V^*)V\Sigma^+U^*(U\Sigma V^*) = U(\Sigma\Sigma^+\Sigma)V^* = A,$$

ii)

$$A^+AA^+ = (V\Sigma^+U^*)U\Sigma V^*(V\Sigma^+U^*) = V(\Sigma^+\Sigma\Sigma^+)U^* = A^+,$$

iii)

$$\left(A^+A\right)^* = A^*(A^+)^* = (U\Sigma V^*)^*(V\Sigma^+U^*)^* =$$

$$= (V\Sigma^*U^*)(U(\Sigma^+)^*V^*) = V(\Sigma^+\Sigma)^*V^* = V(\Sigma^+\Sigma)V^* =$$

$$= (V\Sigma^+U^*)(U\Sigma V^*) = V(\Sigma^+\Sigma)V^* = A^+A.$$

iv) j.w.

Pseudoinverse (continued)

Proof.

Assume that matrices A^+, A'^+ satisfy conditions i) – iv). Then

$$A^+ = A^+AA^+ = A^+(A)A^+ = A^+(AA'^+A)A^+ = A^+((A)A'^+(A))A^+ =$$

$$= A^+((AA'^+A)A'^+(AA'^+A))A^+ = (A^+A)^*(A'^+A)^*A'^+(AA'^+)^*(AA^+)^* =$$

$$= (A^*(A^+)^*)(A^*(A'^+)^*)A'^+((A'^+)^*A^*)((A^+)^*A^*) =$$

$$= (A^*(A^+)^*A^*)(A'^+)^*A'^+(A'^+)^*(A^*(A^+)^*A^*) =$$

$$= (A(A^+)A)^*(A'^+)^*A'^+(A'^+)^*(A(A^+)A)^* =$$

$$= A^*(A'^+)^*A'^+(A'^+)^*A^* =$$

Pseudoinverse (continued)

Proof.

$$
\begin{aligned}
&= A^*(A'^+)^*A'^+(A'^+)^*A^* = \\
&= (A'^+A)^*A'^+(AA'^+)^* = (A'^+A)A'^+(AA'^+) = \\
&= A'^+(AA'^+A)A'^+ = A'^+AA'^+ = A'^+.
\end{aligned}
$$

Singular Value Decomposition – Remarks

Remarks

i) *if matrix A is real then there exists real orthogonal matrices U, V such that $A = U\Sigma V^{\intercal}$,*

ii) *when $\sigma_1 > \sigma_2 > \ldots > \sigma_r > 0$, that is the singular values are pairwise different then the columns $1, 2, \ldots, r$ of U i V are uniquely determined up to a constant $\alpha_i \in \mathbb{C}$ (respectively $\alpha_i \in \mathbb{R}$, when A is real) such that $|\alpha_i| = 1$,*

iii) *when $A \in M(n \times n; \mathbb{C})$ and $\det A \neq 0$ then $A^+ = A^{-1}$,*

iv) *the following matrix norms of A are determined by the singular values of A, i.e.,*

$$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2},$$

$$\|A\|_2 = \sigma_1,$$

Singular Value Decomposition – Remarks (continued)

Remarks

i) *let* $A = U\Sigma V^*$, *that is*

$$AV = U\Sigma.$$

Denote by $u_1, \ldots, u_m$ *the columns of matrix* $U \in M(m \times m; \mathbb{R})$ *and by* $v_1, \ldots, v_n$ *the columns of matrix* $V \in M(n \times n; \mathbb{R})$. *Then for* $i = 1, \ldots, \max m, n$

$$Av_i = \sigma_i u_i.$$

Moreover

$$\ker A = \operatorname{lin}(v_{r+1}, \ldots, v_n)$$

$$imA = \operatorname{lin}(u_1, \ldots, u_r).$$

Singular Value Decomposition – Remarks (continued)

Remarks

vi) *for any* $k \leqslant r$ *let* $\Sigma_k = \mathrm{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) \in M(m \times n; \mathbb{R})$. *Then the matrix*

$$A_k = U\Sigma_k V^*,$$

satisfies the condition: for any matrix $B \in M(m \times n; \mathbb{C})$ *of rank* k

$$\|A - B\|_2 \geqslant \|A - A_k\|_2 = \sigma_{k+1},$$

$$\|A - B\|_F \geqslant \|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \ldots + \sigma_r^2},$$

where, assuming $A = [a_{ij}] \in M(m \times n; \mathbb{C})$ *the norms are defined as follows*

$$\|A\|_F = \sqrt{\mathrm{Tr}(A^*A)} = \sqrt{\mathrm{Tr}(AA^*)} = \sqrt{\sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} |a_{ij}|^2},$$

$$\|A\|_2 = \sup\{\|Ax\|_2 \in \mathbb{R} \mid x \in \mathbb{R}^n, \ \|x\|_2 = 1\} = \sqrt{\lambda_{max}(A^*A)},$$

$$\|x\|_2 = \sqrt{x^*x}.$$

The Best Low Rank Approximation

Proposition

Let $A \in M(m \times n; \mathbb{C})$ *be any matrix and let* $A = U\Sigma V^*$ *be its singular value decomposition, where*

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0) \in M(m \times n; \mathbb{R}),$$

and $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0$*, i.e.,* $r(A) = r$*. Then, for any* k *such that* $0 \leqslant k < r$ *and for any matrix* $B \in M(m \times n; \mathbb{C})$ *such that* $r(B) = k$ *it holds*

$$\|A - A_k\| \leqslant \|A - B\|,$$

where

$$A_k = U\Sigma_k V^*,$$

$$\Sigma_k = \operatorname{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) \in M(m \times n; \mathbb{R}),$$

that is, matrix A_k *of rank* k *is the best approximation of matrix* A *among matrices of rank* k *of the same size as matrix* A *(in the norm* $\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2$*).*

The Best Low Rank Approximation (continued)

Proof.

Obviously $r(A_k) = k$. Moreover

$$\|A - A_k\| = \|U \operatorname{diag}(0, \ldots, \sigma_{k+1}, \ldots, \sigma_r, 0 \ldots, 0) V^*\| =$$

$$= \|\operatorname{diag}(0, \ldots, \sigma_{k+1}, \ldots, \sigma_r, 0 \ldots, 0)\| = \sigma_{k+1}.$$

Let $B \in M(m \times n; \mathbb{C})$ be any matrix such that $r(B) = k$. Let

$$W = \{w \in \mathbb{R}^m \mid Bw = \mathbf{0}\}.$$

Let $w_1, \ldots, w_{n-k} \in \mathbb{C}^n$ be an unitary orthonormal basis of subspace $W \subset \mathbb{C}^n$. Let $v_1, \ldots, v_n \in \mathbb{C}^n$ denote columns of matrix V. Let

$$v \in \operatorname{lin}(v_1, \ldots, v_{k+1}) \cap W \neq \{\mathbf{0}\},$$

be any (non–zero) vector such that

$$\|v\| = 1.$$

The Best Low Rank Approximation (continued)

Proof.

Then

$$\|A - B\| \geqslant \|(A - B)v\| = \|Av\| =$$

$$= \left\| \sum_{i=1}^{r} (u_i \sigma_i v_i^*) \, v \right\| = \left\| \sum_{i=k+1}^{r} ((v_i^* v) u_i \sigma_i) \right\| =$$

$$= \sum_{i=k+1}^{r} \sigma_i^2 (v_i^* v)^2 \geqslant \sigma_{k+1} \sum_{i=k+1}^{r} (v_i^* v)^2 \geqslant \sigma_{k+1},$$

since

$$\|Vv\|^2 = \sum_{i=1}^{n} (v_i^* v)^2 = 1 \geqslant \sum_{i=k+1}^{r} (v_i^* v)^2.$$

$\square$

Singular Value Decomposition – Example

Let

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix}.$$

Then, assuming $A = U\Sigma V^*$, we have

$$A^*A = (V\Sigma^* U^*)(U\Sigma V^*) = V\Sigma^*\Sigma V^* = \begin{bmatrix} 26 & 18 \\ 18 & 74 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 80 & 0 \\ 0 & 20 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix},$$

$$AA^* = (U\Sigma V^*)(V\Sigma^* U^*) = U\Sigma\Sigma^* U^* = \begin{bmatrix} 50 & 30 \\ 30 & 50 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 80 & 0 \\ 0 & 20 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Singular Value Decomposition – Example (continued)

Hence

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix},$$

that is

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix},$$

$$A = U\Sigma V^*,$$

therefore the best rank 1 approximation of matrix A in the norm $\|\cdot\|_2$ oraz $\|\cdot\|_F$ is

$$A_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 2 & 6 \end{bmatrix}.$$

Singular Value Decomposition – Example (continued)

$$A = \begin{bmatrix} 5 & 5 \\ -1 & 7 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{5} & 0 \\ 0 & 2\sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{10}} \end{bmatrix},$$

$$A_1 = \begin{bmatrix} 2 & 6 \\ 2 & 6 \end{bmatrix},$$

$$B = A - A_1 = \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix},$$

and

$$\|B\|_F = \sqrt{3^2 + (-1)^2 + 3^2 + (-1)^2} = 2\sqrt{5} = \sigma_2(A),$$

$$\det(B^* B - \lambda I) = \det \begin{bmatrix} 18 - \lambda & -6 \\ -6 & 2 - \lambda \end{bmatrix} = \lambda(\lambda^2 - 20),$$

hence

$$\|B\|_2 = \sqrt{\lambda_{max}(B^* B)} = \sqrt{20} = 4\sqrt{5}.$$

Optimal Solution of a System of Linear Equations

Definition

For any system of linear equations $Ax = b$ where
$A \in M(m \times n; \mathbb{C}), b \in M(m \times 1; \mathbb{C})$ the vector $x \in \mathbb{C}^n$ is called the optimal solution if

$$\|Ax - b\|_2 \leqslant \|Ay - b\|_2 \text{ for any } y \in \mathbb{C}^n,$$

and if $\|Ax - b\|_2 = \|Ax' - b\|_2$ then $\|x\|_2 \leqslant \|x'\|_2$,

Proposition

For any matrices $A \in M(m \times n; \mathbb{C}), b \in M(m \times 1; \mathbb{C})$ the vector

$$x = A^+ b$$

is the optimal solution of the system $Ax = b$.

Optimal Solution of a System of Linear Equations

Proof.

Let $P = AA^+$ be the matrix of orthogonal projection onto $\operatorname{im}(A)$. Then for any x

$$\|Ax - b\|_2 = \|Ax - Pb + (P - I)b\|_2 =$$

$$= \|Ax - Pb\|_2 + \|(P - I)b\|_2 \geqslant \|(P - I)b\|_2.$$

The lower bound (which does not depend on x) is attained when $x = A^+b$. Assume that $Ax = Ax'$ where $x = A^+b \in \operatorname{im}(A^*)$. Therefore there exists $n \in \ker A$ such that $x' = x + n$ where x and n are perpendicular. Therefore

$$\|x'\|_2 = \|x\|_2 + \|n\|_2 \geqslant \|x\|_2.$$

$\square$

Example

For

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}.$$

$$A^+ = \begin{bmatrix} \frac{3}{2} & -1 & \frac{3}{2} \\ -1 & 1 & -1 \end{bmatrix}$$

It follows

$$A^+A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A^+B = \begin{bmatrix} 3 \\ -2 \end{bmatrix},$$

which is the optimal solution of $AX = B$.

Hadamard's inequality

Proposition

For any matrix $A \in M(n \times n; \mathbb{R})$

$$|\det A| \leqslant \|c_1\| \cdot \ldots \cdot \|c_n\|,$$

where c_i *is the* $i-$*th column of matrix* A *and*

$$\|c_i\| = \sqrt{c_i^{\mathsf{T}} c_i},$$

is the (Euclidean) length of the $i-$*th column, for* $i = 1, \ldots, n$*.*
Moreover, the equality holds if and only if

$$c_i \perp c_j, \quad \text{for} \quad i \neq j.$$

Hadamard's inequality (continued)

Proof.

If $c_i = \mathbf{0}$ or $\det A = 0$ then there is nothing to prove. Dividing each column of matrix A by its length the problem reduces to the following one

$$|\det A| \leqslant 1,$$

where $\|c_i\| = 1$ for $i = 1, \ldots, n$. Let

$$M = A^{\mathsf{T}} A.$$

Then matrix M is a positive definite symmetric matrix. Moreover,

$$\operatorname{Tr} M = \sum_{i=1}^{n} m_{ii} = n,$$

where $M = [m_{ij}]$ as columns of matrix A are of length 1.

Hadamard's inequality (continued)

Proof.

By spectral theorem matrix M is diagonalizable and therefore

$$\det M = \lambda_1 \cdot \ldots \cdot \lambda_n.$$

Moreover

$$\det M = \det(A^\intercal A) = (\det A)^2 = \lambda_1 \cdot \ldots \cdot \lambda_n \leqslant \left(\frac{\lambda_1 + \ldots + \lambda_n}{n}\right)^n = 1,$$

by the Arithmetic-Geometric Mean Inequality. The upper bound is achieved when

$$\lambda_1 = \lambda_2 = \ldots = \lambda_n = 1,$$

i.e., when $M = A^\intercal A = I$ that is when columns of A or pairwise perpendicular.

□

Cauchy–Schwarz Inequality

Proposition

Let $A \in M(n \times n; \mathbb{R})$ be a positive semidefinite symmetric matrix. Then for any $x, y \in \mathbb{R}^n$

$$|x^\intercal A y| \leqslant (x^\intercal A x)^{\frac{1}{2}} \, (y^\intercal A y)^{\frac{1}{2}} \, .$$

Proof.

For any $t \in \mathbb{R}$,

$$0 \leqslant (x - ty)^\intercal A (x - ty) = (y^\intercal A y) \, t^2 - 2 \, (x^\intercal A y) \, t + (x^\intercal A x) \, .$$

Hence, the discriminant

$$\Delta = 4 \, (x^\intercal A y)^2 - 4 \, (x^\intercal A x) \, (y^\intercal A y) \leqslant 0.$$

$\square$

Cauchy–Schwarz Inequality (continued)

Definition

Vector $x \in \mathbb{R}^n$ is **isotropic** (with respect to a symmetric matrix A) if $x^\mathsf{T} A x = 0$.

Corollary

Let $A \in M(n \times n; \mathbb{R})$ be a symmetric positive semidefinite matrix. Then $x \in \mathbb{R}^n$ is isotropic if and only if $Ax = \mathbf{0}$.

Proof.

Assume $y \in \mathbb{R}^n$ is istotropic in the proof of Cauchy–Schwarz inequality. Then the linear function

$$-2\,(x^\mathsf{T} A y)\,t + (x^\mathsf{T} A x) \geqslant 0,$$

is non–negative for any $x \in \mathbb{R}^n$. This implies $x^\mathsf{T} A y = 0$ for any $x \in \mathbb{R}^n$, i.e. $Ay = \mathbf{0}$. □

Convex Cone

Definition

A subset $C \subset \mathbb{R}^n$ is a **cone**, if

i) for any $v, w \in C$
$$v + w \in C,$$

ii) for any $v \in C$ and any $\alpha \in \mathbb{R}$ such that $\alpha \geqslant 0$,
$$\alpha \in C.$$

The cone C is **pointed** if it does not contain a one–dimensional subspace of $\mathbb{R}^n$ (i.e, a line). The cone C is **(closed) polyhedral** if it equal to the intersection of finite (closed) half–spaces in $\mathbb{R}^n$.

Dual Cone

Definition

Let $A \subset \mathbb{R}^n$ be any subset. Let $v \cdot w$ be a scalar product in bR^n. Then the set

$$A^\vee = \{v \in \mathbb{R}^n \mid v \cdot w \geqslant 0 \text{ for any } w \in A\},$$

is called the dual cone of the set A.

Proposition

For any subset $A \subset \mathbb{R}^n$ the set $A^\vee$ is a closed convex cone.

Proof.

Exercise. $\square$

Cone Spanned by Set

Definition

A cone $C \subset \mathbb{R}^n$ is spanned by set $A \subset \mathbb{R}^n$ if

$$C = \{\alpha_1 v_1 + \dots \alpha_k v_k \in \mathbb{R}^n \mid v_1, \dots, v_k \in A,\ \alpha_1, \dots, \alpha_k \geqslant 0,\ k \geqslant 1\}.$$

We write

$$C = \mathrm{cone}(A),$$

and if $A = \{v_1, \dots, v_k\}$

$$C = \mathrm{cone}(v_1, \dots, v_k).$$

Extremal Rays of a Cone

Definition

Let $C \subset \mathbb{R}^n$ be a (convex) cone. Vector (or a half–line spanned by it) $v \in C, v \neq \mathbf{0}$ is an **extremal ray** of cone C, if for any $v_1, v_2 \in V$, if $v = v_1 + v_2$ then $v_1 = tv$ or $v_2 = tv$ for some $t \geqslant 0$.

The Positive Semidefinite Cone

Definition

Let

$$\mathbb{S}^n = \{A \in M(n \times n; \mathbb{R}^n) \mid A^\intercal = A\} \subset M(n \times n; \mathbb{R}),$$

be the $\binom{n+1}{2}$ subspace of symmetric matrices with the (standard) scalar product given by

$$A \cdot B = \mathrm{Tr}(AB),$$

for any $A, B \in \mathbb{S}^n$.

Definition

Let

$$C_{\geqslant 0} = \{A \in \mathbb{S}^n \mid A \text{ is postive semidefinite}\},$$

$$C_{>0} = \{A \in \mathbb{S}^n \mid A \text{ is postive definite}\},$$

denote the **positive semidefinite** and **positive definite** cones, respectively.

The Positive Semidefinite Cone (continued)

Proposition

i) *the positive semidefinite cone* $C_{\geqslant 0}$ *is a closed convex pointed cone,*

ii) *the positive semidefinite cone* $C_{\geqslant 0}$ *is self–dual, i.e.*

$$C_{\geqslant 0}^{\vee} = C_{\geqslant 0},$$

with respect to the scalar product given by the trace,

iii) *the positive semidefinite cone* $C_{\geqslant 0}$ *is spanned by rank 1 matrices* vv^{T}*, i.e.,*

$$C_{\geqslant 0} = \operatorname{cone}\left(\{vv^{\mathsf{T}} \in \mathbb{S}^n \mid v \in \mathbb{R}^n\}\right),$$

iv) *the matrices* vv^{T} *are exactly the extremal rays of the cone* $C_{\geqslant 0}$*,*

v)

$$\operatorname{int} C_{\geqslant 0} = C_{>0}.$$

The Positive Semidefinite Cone (continued)

Proposition

i) *the positive semidefinite cone $C_{\geqslant 0}$ is a closed convex pointed cone,*

ii) *the positive semidefinite cone $C_{\geqslant 0}$ is self–dual, i.e.*

$$C_{\geqslant 0}^{\vee} = C_{\geqslant 0},$$

with respect to the scalar product given by the trace,

iii) *the positive semidefinite cone $C_{\geqslant 0}$ is spanned by rank 1 matrices vv^{T}, i.e.,*

$$C_{\geqslant 0} = \mathrm{cone}\left(\{vv^{\mathsf{T}} \in \mathbb{S}^n \mid v \in \mathbb{R}^n\}\right),$$

iv) *the matrices vv^{T} are exactly the extremal rays of the cone $C_{\geqslant 0}$,*

v)

$$\mathrm{int}\, C_{\geqslant 0} = C_{>0}.$$

Proof.

Omitted. Involves mostly eigenvalue decomposition. $\square$

The Positive Semidefinite Cone (continued)

Remark

The positive semidefinite cone is described by polynomial inequalities given by the all principal minors (Sylvester's criterion). For example matrix

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

is positive semidefinite if and only if

$$\begin{cases} a \geqslant 0, \\ c \geqslant 0, \\ ac - b^2 \geqslant 0. \end{cases}$$

The extremal rays of the positive semidefinite cone are exactly of the form

$$\begin{bmatrix} s \\ t \end{bmatrix} \begin{bmatrix} s & t \end{bmatrix} = \begin{bmatrix} s^2 & st \\ st & t^2 \end{bmatrix},$$

for any $s, t \in \mathbb{R}$.

The Positive Semidefinite Cone (continued)

Remark

When $\|v\| = 1$ *the matrix* $vv^\intercal$ *is the matrix of the orthogonal (linear) projections onto* $\mathrm{lin}(v)$, *i.e.*

$$M(P_{\mathrm{lin}(v)})_{st}^{st} = vv^\intercal.$$

In general, for any $v \neq \mathbf{0}$

$$M(P_{\mathrm{lin}(v)})_{st}^{st} = \frac{vv^\intercal}{v^\intercal v}.$$

Non–negative Polynomials

Definitions

Let $d \geqslant 1$. A polynomial $p(x)$ of degree $2d$ is **non–negative** if for any $x \in \mathbb{R}$

$$p(x) \geqslant 0.$$

Proposition

A polynomial $p(x)$ of degree $2d$ is non–negative if and only if all its real roots are of even multiplicity and if $a_{2d} > 0$ where $p(x) = a_{2d}x^{2n} + \dots$ (that is the leading coefficient is positive).

Proof.

Exercise. □

Non–negative Polynomials (continued)

Proposition

A polynomial $p(x) = \sum_{i=0}^{2d} a_i x^i$ of degree $2d$ is non–negative if and only if there exists a symmetric positive semidefinite matrix $M = [m_{ij}] \in M((d+1) \times (d+1); \mathbb{R})$ such that

$$a_k = \sum_{i+j=k} m_{ij},$$

for any $k = 0, \ldots, 2d$ where rows and columns of matrix M are numbered from 0 to d. Moreover the correspondence is one–to–one.

Non–negative Polynomials (continued)

Proof.

($\Leftarrow$) Let $\mathbf{x} = (1, x, x^2, \ldots, x^d)$. Then

$$p(x) = \mathbf{x}^\mathsf{T} M \mathbf{x} \geqslant 0.$$

($\Rightarrow$)

$$p(z) = a_{2d} \prod_{i=1}^{d} (z - z_i)(z - \overline{z_i}),$$

where $z_i, \overline{z_i} \in \mathbb{C}$ are complex roots of $p(x)$. Let

$$q(x) = \sqrt{a_{2d}} \prod_{i=1}^{d} (x - z_i) = \sum_{i=0}^{d} c_i x^i.$$

Let

$$q_1(x) = \operatorname{Re} q(x) = \sum_{i=0}^{d} (\operatorname{Re} c_i)\, x^i,$$

$$q_2(x) = \operatorname{Im} q(x) = \sum_{i=0}^{d} (\operatorname{Im} c_i)\, x^i,$$

Non–negative Polynomials (continued)

Proof.

i.e.

$$q(x) = q_1(x) + \sqrt{-1}q_2.$$

Then for any $x \in \mathbb{R}$

$$p(x) = q(x)\overline{q(x)} = |q(x)|^2 = q_1^2(x) + q_2^2(x) =$$

$$= (v^\mathsf{T}\mathbf{x})^2 + (w^\mathsf{T}\mathbf{x})^2 = \mathbf{x}^\mathsf{T}(vv^\mathsf{T} + ww^\mathsf{T})\mathbf{x},$$

where

$$v = (\operatorname{Re} c_0, \operatorname{Re} c_1, \ldots, \operatorname{Re} c_d) \in \mathbb{R}^{d+1},$$

$$w = (\operatorname{Im} c_0, \operatorname{Im} c_1, \ldots, \operatorname{Im} c_d) \in \mathbb{R}^{d+1}.$$

$\square$

Example

Let

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

Then

$$M = A^{\intercal} A = \begin{bmatrix} 2 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 2 \end{bmatrix},$$

is positive definite. Therefore, the polynomial

$$p(x) = \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 2 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} = 2x^4 - 4x^3 + 7x^2 - 4x + 2,$$

is non–negative. In fact,

$$f(x) \geqslant f(0.3768669139161389\ldots) \approx 1.312973699214175\ldots > 0.$$

Quiz

Is it possible to find $n \geqslant 1$ and $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ such that

$$\|\mathbf{x}_1 - \mathbf{x}_2\| = \|\mathbf{x}_2 - \mathbf{x}_3\| = 1, \quad \|\mathbf{x}_1 - \mathbf{x}_3\| = 3?$$

Quiz

Is it possible to find $n \geqslant 1$ and $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ such that

$$\|\mathbf{x}_1 - \mathbf{x}_2\| = \|\mathbf{x}_2 - \mathbf{x}_3\| = 1, \quad \|\mathbf{x}_1 - \mathbf{x}_3\| = 3?$$

No, it is not possible as

$$\|\mathbf{x}_1 - \mathbf{x}_2\| \leqslant \|\mathbf{x}_1 - \mathbf{x}_2 + \mathbf{x}_2 - \mathbf{x}_3\| \leqslant$$

$$\leqslant \|\mathbf{x}_1 - \mathbf{x}_2\| + \|\mathbf{x}_2 - \mathbf{x}_3\|,$$

(triangle inequality) but it is not true that $3 \leqslant 1 + 1 = 2$.

Properties of Pseudoinverses

Proposition

Let $A \in M(m \times n; \mathbb{R})$ be any matrix. Then $P = AA^+$ is a matrix of the orthogonal projection onto imA and $Q = A^+A$ is a matrix of the orthogonal projection onto $imA^\intercal$.

Proof.

Let $A = U\Sigma V^\intercal$ be an SVD decomposition of A. Then

$$P = U\Sigma V^\intercal V\Sigma^+ U^\intercal = U_{:,1:r} U^\intercal_{:,1:r},$$

is symmetric where $r = r(A)$ and $U_{:,1:r}$ denotes first r columns of matrix A (orthonormal basis or imA). Moreover

$$P^2 = AA^+AA^+ = AA^+ = P.$$

Similarly for Q. $\square$

Properties of Pseudoinverses (continued)

Proposition

Let $A \in M(m \times n; \mathbb{R})$ be any matrix.

$$A^+ = (A^\intercal A)^+ A^\intercal, \quad A^+ = A^\intercal (A A^\intercal)^+.$$

Proof.

Let $A = U\Sigma V^\intercal$ be an SVD decomposition of A. Then

$$A^\intercal A = V\Sigma^2 V^\intercal,$$

$$(A^\intercal A)^+ = V\left(\Sigma^2\right)^+ V^\intercal,$$

$$(A^\intercal A)^+ A = V\left(\Sigma^2\right)^+ V^\intercal V \Sigma U^\intercal = A^+.$$

The second part is similar. $\square$

Properties of Pseudoinverses (continued)

Proposition

Let $A \in M(m \times n; \mathbb{R})$ be a matrix. If $r(A) = m$ (full row rank) then

$$A^+ = A^\intercal (AA^\intercal)^{-1}.$$

If $r(A) = n$ (full column rank) then

$$A^+ = (A^\intercal A)^{-1} A^\intercal.$$

Proof.

Follows from the above proposition (matrices $AA^\intercal$ and $A^\intercal A$ are invertible).

$\square$

Properties of Pseudoinverses (continued)

The following lemma will be subsequently used in the proof of Greville's conditions.

Proposition

Let $A, B \in M(m \times n; \mathbb{R})$ be any matrices. Then

$$A^\intercal = A^+ A A^\intercal,$$

$$B^\intercal = B^\intercal B B^+.$$

Proof.

Since A^+A is a matrix of (orthogonal) projection onto $\mathrm{im}(A^\intercal)$ and BB^+ is a matrix of (orthogonal) projection onto $\mathrm{im}(B)$

$$A^\intercal = A^+ A A^\intercal,$$

$$B = BB^+ B.$$

Conjugating the last equation finishes the proof. $\square$

Inverse Law

Theorem (Greville)

Let $A \in M(m \times n; \mathbb{R})$, $B \in M(n \times k; \mathbb{R})$. *If* $(AB)^+ = B^+A^+$ *then* $\operatorname{im}(A^\intercal AB) \subset \operatorname{im}(B)$ *and* $\operatorname{im}(BB^\intercal A^\intercal) \subset \operatorname{im}(A^\intercal)$.

Proof.

By the above lemma applied to AB (the second case) using the main assumption

$$B^\intercal A^\intercal = B^\intercal A^\intercal ABB^+A^+,$$

Multiplying on the right by $AA^\intercal AB$ gives

$$B^\intercal A^\intercal AA^\intercal AB = B^\intercal A^\intercal ABB^+A^+AA^\intercal AB.$$

By the above lemma

$$B^\intercal A^\intercal AA^\intercal AB = B^\intercal A^\intercal ABB^+(A^+AA^\intercal)AB = B^\intercal A^\intercal ABB^+A^\intercal AB,$$

i.e.,

$$B^\intercal A^\intercal A(I - BB^+)A^\intercal AB = 0.$$

□

Inverse Law(continued)

Proof.

$$B^\intercal A^\intercal A(I - BB^+)A^\intercal AB = 0.$$

The middle matrix is idempotent and symmetric hence

$$\|(I - BB^+)A^\intercal AB\|_2^2 = 0,$$

which is equivalent to

$$\mathrm{im}(A^\intercal AB) \subset \mathrm{im}(B).$$

The rest is similar to the previous argument. $\square$

Inverse Law(continued)

In fact, the converse holds.

Theorem (Greville)

Let $A \in M(m \times n; \mathbb{R}),\ B \in M(n \times k; \mathbb{R})$*. If* $\mathrm{im}(BB^{\intercal}A^{\intercal}) \subset \mathrm{im}(A^{\intercal})$ *and* $\mathrm{im}(A^{\intercal}AB) \subset \mathrm{im}(B)$ *then* $(AB)^{+} = B^{+}A^{+}$*.*

Proof.

The assumptions imply that

$$A^{+}ABB^{\intercal}A^{\intercal} = BB^{\intercal}A^{\intercal},$$

$$BB^{+}A^{\intercal}AB = A^{\intercal}AB,$$

Multiplying the first equation on the right by $((AB)^{\intercal})^{+}$ and on the left by B^{+} gives

$$B^{+}A^{+}ABB^{\intercal}A^{\intercal}((AB)^{\intercal})^{+} = B^{+}BB^{\intercal}A^{\intercal}((AB)^{\intercal})^{+}.$$

Inverse Law(continued)

Proof.

$$B^+A^+ABB^\intercal A^\intercal((AB)^\intercal)^+ = B^+BB^\intercal A^\intercal((AB)^\intercal)^+.$$

$$B^+A^+AB(AB)^\intercal((AB)^\intercal)^+ = \left(B^+BB^\intercal\right)A^\intercal((AB)^\intercal)^+.$$

By the previous lemma this is equivalent to

$$B^+A^+AB = (AB)^\intercal((AB)^\intercal)^+,$$

therefore the matrix B^+A^+AB is symmetric.

Inverse Law(continued)

Proof.

Similarly, by multiplying

$$BB^+A^\intercal AB = A^\intercal AB,$$

on the left by $\left((A)^+\right)^\intercal$

$$\left((A)^+\right)^\intercal BB^+A^\intercal AB = \left(\left((A)^+\right)^\intercal A^\intercal A\right)B,$$

$$\left((A)^+\right)^\intercal BB^+A^\intercal AB = AB.$$

Multiplying the above on the right by $(AB)^+$ and using on the left hand side $B^+ = (B^\intercal B)^+B^\intercal$ gives

$$\left((A)^+\right)^\intercal B(B^\intercal B)^+B^\intercal A^\intercal(AB)(AB)^+ = (AB)(AB)^+,$$

$$\left((A)^+\right)^\intercal \left((B)^+\right)^\intercal (AB)^\intercal = (AB)(AB)^+,$$

which, after conjugating side-wise implies that ABB^+A^+ is symmetric.

Inverse Law(continued)

Proof.

The first Penrose condition is easily verified.

$$ABB^{+}A^{+}AB = AB\big(B^{+}A^{+}AB\big) =$$

$$= AB(AB)^{\intercal}((AB)^{\intercal})^{+} = AB.$$

Note that

$$\mathrm{im}(BB^{*}A^{*}) \subset \mathrm{im}(A^{*}) \Longrightarrow \mathrm{im}(BB^{+}A^{+}) \subset \mathrm{im}(A^{+}).$$

($\mathrm{im}(A^{+}) = \mathrm{im}(A^{*})$ and any eigenvector of BB^{*} is an eigenvector of BB^{+}, moreover any linear combination of eigenvectors of BB^{*} corresponding to non–zero eigenvalues is an eigenvalue of BB^{+}.)

Inverse Law(continued)

Proof.

The second Penrose condition follows from $\mathrm{im}(BB^+A^+) \subset \mathrm{im}(A^+)$. Fix any vector u and let

$$v = B^+A^+ABB^+A^+u = B^+A^+A\big(BB^+A^+\big)u.$$

There exists vector w such that $(BB^+A^+)u = A^+w$, i.e.,

$$v = B^+A^+AA^+w = B^+A^+w = B^+BB^+A^+u = B^+A^+u.$$

Since vector u was arbitrary

$$B^+A^+ABB^+A^+ = B^+A^+.$$

□

Inverse Law(continued)

Remark

This also shows that condition $\operatorname{im}(A^\intercal AB) \subset \operatorname{im}(B)$ *implies conditions* $i), ii)$ *and* $iii)$ *for* B^+A^+.

Positive Semidefinite Block Matrix

Proposition

For any matrices
$A \in M(m \times m; \mathbb{R}), B \in M(n \times m; \mathbb{R}), C \in M(n \times n; \mathbb{R})$ *where* A *and* C *are symmetric, let*

$$M = \begin{bmatrix} A & B^\intercal \\ B & C \end{bmatrix},$$

be a symmetric positive semidefinite matrix. Then

$$B^\intercal = AA^+B^\intercal, \quad B = (CC^+)^\intercal B.$$

Positive Semidefinite Block Matrix (continued)

Proof.

By spectral decomposition there exist $N \in M((m+n) \times (m+n); \mathbb{R})$ such that $M = N^\intercal N$. Assume that $N = \begin{bmatrix} N_1 & N_2 \end{bmatrix}$, where $N_1 \in M((m+n) \times m; \mathbb{R}), N_2 \in M((m+n) \times n; \mathbb{R})$. Then

$$A = N_1^\intercal N_1, \quad B^\intercal = N_1^\intercal N_2, \quad C = N_2^\intercal N_2.$$

Moreover,

$$AA^+ B^\intercal = (N_1^\intercal N_1)(N_1^\intercal N_1)^+ N_1^\intercal N_2 = N_1^\intercal N_2 = B^\intercal,$$

as $(N_1^\intercal N_1)(N_1^\intercal N_1)^+$ is an orthogonal projection onto $\mathrm{im}(N_1^\intercal N_1) = \mathrm{im}(N_1^\intercal)$. Similarly,

$$\left(CC^+\right)^\intercal B = (N_2^\intercal N_2)^+ (N_2^\intercal N_2) N_2^\intercal N_1 = N_2^\intercal N_1.$$

$\square$

Schur Complement

Definition

For any matrices $A \in M(m \times m; \mathbb{R}), B \in M(m \times n; \mathbb{R}), C \in M(n \times m; \mathbb{R}), D \in M(n \times n; \mathbb{R})$ and the matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix},$$

the **Schur complement** of matrix A with respect to M is

$$M|A = D - CA^{+}B.$$

Schur Complement (continued)

Proposition

A positive symmetric semidefinite matrix

$$M = \begin{bmatrix} A & B^\intercal \\ B & C \end{bmatrix},$$

is conjugate to the matrix $\mathrm{diag}(A, M|A)$, *where* $M|A = C - BA^+ B^\intercal$.

Proof.

$$\begin{bmatrix} I & 0 \\ BA^+ & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & M|A \end{bmatrix} \begin{bmatrix} I & A^+ B^\intercal \\ 0 & I \end{bmatrix} =$$

$$= \begin{bmatrix} A & 0 \\ BA^+ A & M|A \end{bmatrix} \begin{bmatrix} I & A^+ B^\intercal \\ 0 & I \end{bmatrix} =$$

$$= \begin{bmatrix} A & AA^+ B^\intercal \\ BA^+ A & BA^+ AA^+ B^\intercal + M|A \end{bmatrix} = \begin{bmatrix} A & AA^+ B^\intercal \\ BA^+ A & C \end{bmatrix} = M.$$

Schur Complement (continued)

Corollary

If a symmetric matrix

$$M = \begin{bmatrix} A & B^\intercal \\ B & C \end{bmatrix},$$

is positive semidefinite then matrix A is positive semidefinite and the Schur complement $M|A$ is postive semidefinite. If matrix A is positive semidefinite and the Schur complement $M|A$ is postive semidefinite for symmetric matrix M and $BA^+A = B$ (for example when A is invertible) then M is positive semidefinite. Similar theorem is true for positive definite matrices.

Quiz (continued)

Is it possible to find $n \geqslant 1$ and $\mathbf{x}_0, \mathbf{x}_2, \ldots, \mathbf{x}_5 \in \mathbb{R}^n$ such that (addition modulo 6)

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm 1}\| = 1,$$

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm 2}\| = \sqrt{3},$$

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm 3}\| = 2?$$

Quiz (continued)

Is it possible to find $n \geqslant 1$ and $\mathbf{x}_0, \mathbf{x}_2, \ldots, \mathbf{x}_5 \in \mathbb{R}^n$ such that (addition modulo 6)

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm1}\| = 1,$$

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm2}\| = \sqrt{3},$$

$$\|\mathbf{x}_i - \mathbf{x}_{i\pm3}\| = 2?$$

Yes, it is. Those are vertices of a regular hexagon with sides of length 1 and $n = 2$.

Multidimensional Scaling

Definition

A symmetric non–negative matrix $D = [d_{ij}] \in M(n \times n; \mathbb{R}_{\geqslant 0})$ is called **Euclidean distance matrix** if there exist $m \geqslant 1$ and

$$\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m,$$

such that

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|.$$

Definition

Let

$$H = I - n\mathbb{1}\mathbb{1}^\mathsf{T} \in M(n \times n; \mathbb{R}),$$

be the **centering matrix**.

[0]based on K. V. Mardia, J. T. Kent, J. M. Bibby *Mulitvariate Analysis*

Multidimensional Scaling

Proposition

Let $D = [a_{ij}] \in M(n \times n; \mathbb{R}_{\geqslant 0})$ *be a non–negative symmetric matrix. Let* $A = [a_{ij}] \in M(n \times n; \mathbb{R})$ *be a matrix given by the condition*

$$a_{ij} = -\frac{1}{2} d_{ij}^2.$$

Let

$$B = HAH.$$

Multidimensional Scaling (continued)

Proposition

Then D is an Euclidean distance matrix if and only if matrix B is postive semidefinite. Moreover, in this case, let

$$\lambda_1 \geqslant \ldots \geqslant \lambda_m > 0,$$

denote (all) positive eigenvalues of B (i.e., eigenvalue of multiplicity k appear exactly k times) with corresponding pairwise orthogonal eigenvectors $w_1, \ldots, w_m$ such that for $i = 1, \ldots, m$

$$w_i \cdot w_i = \lambda_i.$$

Then $\mathbf{x}_i \in \mathbb{R}^m$ and $\mathbf{x}_i$ lie in the rows of the matrix $\begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix}$. Moreover the barycenter of $v_1, \ldots, v_m$ is $\mathbf{0}$ and B is the Gram matrix of vectors $v_1, \ldots, v_m$, i.e. $b_{ij} = v_i \cdot v_j$.

Example 1

Let

$$D = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ 3 & 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & -\frac{1}{2} & -\frac{9}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{9}{2} & -\frac{1}{2} & 0 \end{bmatrix}.$$

Then

$$B = HAH = \frac{1}{18} \begin{bmatrix} 38 & 5 & -43 \\ 5 & -10 & 5 \\ -43 & 5 & 38 \end{bmatrix},$$

which has eigenvalues $\lambda = -\frac{5}{6}$ or $\lambda = 0$ or $\lambda = \frac{9}{2}$, i.e. it is not positive semidefinite.

Example 2

Let

$$D = \begin{bmatrix} 0 & 1 & \sqrt{3} & 2 & \sqrt{3} & 1 \\ 1 & 0 & 1 & \sqrt{3} & 2 & \sqrt{3} \\ \sqrt{3} & 1 & 0 & 1 & \sqrt{3} & 2 \\ 2 & \sqrt{3} & 1 & 0 & 1 & \sqrt{3} \\ \sqrt{3} & 2 & \sqrt{3} & 1 & 0 & 1 \\ 1 & \sqrt{3} & 2 & \sqrt{3} & 1 & 0 \end{bmatrix},$$

$$A = \begin{bmatrix} 0 & -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{3}{2} \\ -\frac{3}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} & -\frac{3}{2} & -2 \\ -2 & -\frac{3}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} & -\frac{3}{2} \\ -\frac{3}{2} & -2 & -\frac{3}{2} & -\frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{3}{2} & -2 & -\frac{3}{2} & -\frac{1}{2} & 0 \end{bmatrix}.$$

Example 2 (continued)

Then

$$B = HAH = \frac{1}{2}\begin{bmatrix} 2 & 1 & -1 & -2 & -1 & 1 \\ 1 & 2 & 1 & -1 & -2 & -1 \\ -1 & 1 & 2 & 1 & -1 & -2 \\ -2 & -1 & 1 & 2 & 1 & -1 \\ -1 & -2 & -1 & 1 & 2 & 1 \\ 1 & -1 & -2 & -1 & 1 & 2 \end{bmatrix},$$

which has eigenvalues $\lambda = 0$ (of multiplicity 4) and $\lambda = 3$ (of multiplicity 2), i.e. it is positive semidefinite.

Example 2 (continued)

Moreover

$$V_{(3)} = \text{lin}((1,0,-1,-1,0,1),(0,1,1,0,-1,-1)),$$

which, after Gram-Schmidt process gives orthogonal basis

$$V_{(3)} = \text{lin}((1,0,-1,-1,0,1),(1,2,1,-1,-2,-1)).$$

Let

$$w_1 = \frac{\sqrt{3}}{2}(1,0,-1,-1,0,1),$$

$$w_2 = \frac{1}{2}(1,2,1,-1,-2,-1).$$

Then $w_1 \cdot w_2 = 0$ and $w_1 \cdot w_1 = w_2 \cdot w_2 = 3$.

Example 2 (continued)

Vectors $\mathbf{x}_0, \ldots, \mathbf{x}_5 \in \mathbb{R}^2$ can be read from the rows of the matrix

$$\begin{bmatrix} w_1 & w_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ 1 & 0 \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -1 & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}.$$

Those are exactly the (complex) sixth roots of unity (clockwise).