Lecture IV Matrix Algebra

Lei Du
dulei@dlut.edu.cn

School of Mathematical Sciences
Dalian University of Technology

September, 2022

Contents

Chapter summary

The objective of this chapter is to present an algebraic framework for the study of matrices and to introduce bilinear/sesquilinear maps and bilinear/sesquilinear forms with their associated matrices.

- main notations and a few special matrices
- transposition and vectorization operations
- inner product, norm, and orthogonality
- matrix multiplication, periodic, nilpotent, and idempotent matrices
- Matrix trace and Frobenius norm
- the fundamental subspaces associated with a matrix
- determinant, inverse, auto-inverse, and generalized inverse
- properties of the Moore-Penrose pseudo-inverse
- structures of multiplicative groups of matrices
- matrix representations of a linear map and a bilinear/sesquilinear form, Quadratic forms and Hermitian forms
- eigenvalue and eigenvector, and generalized eigenvalue problem

Notations and definitions

- Scalars, column vectors, matrices, and hypermatrices/tensors of order higher than two will be denoted by lowercase letters $(a, b, \cdots)$, bold lowercase letters $(\mathbf{a}, \mathbf{b}, \cdots)$, bold uppercase letters $(\mathbf{A}, \mathbf{B}, \cdots)$, and calligraphic letters $(\mathcal{A}, \mathcal{B}, \cdots)$, respectively.
- $\mathbf{1}_I$ is the column vector of dimension I, all the elements are equal to 1. Matrices $\mathbf{0}_{I\times J}$ and $\mathbf{1}_{I\times J}$ of dimensions $I \times J$ have all their elements equal to 0 and 1 , respectively. $\mathbf{I}_N = [\delta_{ij}]$, designates the Nth-order identity matrix, δ_{ij} being the Kronecker delta.
- A matrix $\mathbf{A}$ of dimensions $I \times J$, denoted by $\mathbf{A}(I, J)$, is an array of IJ elements stored in I rows and J columns. Its ith row and jth column, denoted by $\mathbf{A}_{i.}$ and $\mathbf{A}_{.j}$, respectively. The element located at the intersection of $\mathbf{A}_{i}$. and $\mathbf{A}_{.j}$ is designated by a_{ij} or $a_{i,j}$ or $(\mathbf{A})_{ij}$ or still $(\mathbf{A})_{i,j}$. We will use the notation $\mathbf{A} = [a_{ij}]$.

Notations and definitions

- A matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$. If $I = J$, it is said that $\mathbf{A}$ is square of order I, whereas if $I \neq J$, $\mathbf{A}$ is said to be rectangular. Coefficients $a_{ii}, i \in \langle I \rangle$, of a square matrix of order I, are called the diagonal elements; they form the main diagonal. The elements $a_{i,I-i+1}, i \in \langle I \rangle$, constitute the secondary diagonal also known as the antidiagonal.
- The special cases $I = 1$ and $J = 1$ correspond respectively to row vectors of dimension J and to column vectors of dimension I.

The matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ can be defined as a map f from the Cartesian product $\langle I \rangle \times \langle J \rangle$ to $\mathbb{K}$ such that $(i, j) \mapsto f(i, j) = a_{ij} \in \mathbb{K}$, $f(i, j)$ representing the value of f at position (i, j) of $\mathbf{A}$.

Partitioned matrices

Another notation consists in partitioning a matrix into blocks which can be matrices or vectors themselves. Such a matrix is written as:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \dots & \mathbf{A}_{1,S} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \dots & \mathbf{A}_{2,S} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{R,1} & \mathbf{A}_{R,2} & \dots & \mathbf{A}_{R,S} \end{bmatrix} \in \mathbb{K}^{(I_1+\cdots+I_R)\times(J_1+\cdots+J_S)},$$

where $\mathbf{A}_{r,s} \in \mathbf{K}^{I_r \times J_x}$, with $r \in \langle R \rangle$ and $s \in \langle S \rangle$.

Matrix vector spaces

In the following, we present addition and scalar multiplication operations that allow to equip the set of matrices $\mathbb{K}^{I \times J}$ with a v.s. structure:

- Addition of two matrices $\mathbf{A}$ and $\mathbf{B}$ of $\mathbb{K}^{I \times J}$ such that:

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}] \in \mathbb{K}^{I \times J}.$$

- Multiplication of a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ by a scalar $\alpha \in \mathbb{K}$ such that:

$$\alpha \mathbf{A} = [\alpha a_{ij}] \in \mathbb{K}^{I \times J}.$$

  Equipped with these two operations, the set $\mathbb{R}^{I \times J}$ is a $\mathbb{R}$-v.s., while $\mathbb{C}^{I \times J}$ is a C-v.s.

Every real or complex matrix $\mathbf{A} = [a_{ij}] \in \mathbb{K}^{I \times J}$ can be written in the canonical basis as:

$$\mathbf{A} = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} \mathbf{E}_{ij}^{(I \times J)}$$

The elements a_{ij} represent the coordinates of $\mathbf{A}$ in the canonical basis $\left\{ \mathbf{E}_{ij}^{(I \times J)} \right\}$, $\mathbf{E}_{ij}^{(I \times J)} \in \mathbb{R}^{I \times J}$ containing 1 at position (i, j) and 0s elsewhere.

Some special matrices

Properties		Conditions
$\mathbf{A}$ positive (denoted $\mathbf{A} > 0$)	If	$a_{ij} > 0 \ \forall i \in \langle I \rangle \ , \ \forall j \in \langle J \rangle$
$\mathbf{A}$ non-negative (denoted $\mathbf{A} \geq 0$)	If	$a_{ij} \geq 0 \ \forall i \in \langle I \rangle \ , \ \forall j \in \langle J \rangle$
$\mathbf{A}$ (square) diagonal	If	$a_{ij} = 0 \quad \forall i \neq j$
$\mathbf{A}$ diagonally dominant	If	$\mid a_{ii} \mid \geq \sum_{j \neq i} \mid a_{ij} \mid \quad \forall i \in \langle I \rangle$
$\mathbf{A}$ strongly diagonally dominant	If	the above inequality is strict for a value of i at least
$\mathbf{A}$ strictly diagonally dominant	If	the inequality is strict for all i
$\mathbf{A}$ (square) stochastic	If	$\mathbf{A} \geq 0$ and $\sum_{j=1}^{J} a_{ij} = 1, \quad \forall i \in \langle I \rangle$
$\mathbf{A}$ (square) doubly stochastic	If	$\mathbf{A} \geq 0$ and $\sum_{j=1}^{I} a_{ij} = \sum_{j=1}^{I} a_{ji} = 1, \quad \forall i \in \langle I \rangle$

图: Special matrices

Structures		Conditions
Upper triangular	If	$a_{ij} = 0$ for $i > j$
Unit upper triangular	If	$a_{ij} = 0$ for $i > j$ and $a_{ii} = 1$
Strictly upper triangular	If	$a_{ij} = 0$ for $i \geq j$
Upper Hessenberg	If	$a_{ij} = 0$ for $i > j + 1$

图: Upper triangular matrices

Transposition and conjugate transposition

The transpose of $\mathbf{A} \in \mathbb{K}^{I \times J}$ is the matrix denoted by $\mathbf{A}^T$, of dimensions $J \times I$, such that $\mathbf{A}^T = [a_{ji}]$, with $i \in \langle I \rangle$ and $j \in \langle J \rangle$.

In the case of a complex matrix, the conjugate transpose, also known as Hermitian transpose and denoted by $\mathbf{A}^H$, is defined as:

$$\mathbf{A}^H = (\mathbf{A}^*)^T = \left(\mathbf{A}^T\right)^* = \left[a_{ji}^*\right],$$

where $\mathbf{A}^* = \left[a_{ij}^*\right]$ is the conjugate of $\mathbf{A}$. By decomposing $\mathbf{A}$ using its real and imaginary parts, we have:

$$\mathbf{A} = \mathrm{Re}(\mathbf{A}) + j\mathrm{Im}(\mathbf{A}) \Rightarrow \left\{ \begin{array}{l} \mathbf{A}^T = [\mathrm{Re}(\mathbf{A})]^T + j[\mathrm{Im}(\mathbf{A})]^T \\ \mathbf{A}^H = [\mathrm{Re}(\mathbf{A})]^T - j[\mathrm{Im}(\mathbf{A})]^T \end{array} \right\}.$$

Transposition and conjugate transposition

The operations of transposition and conjugate transposition satisfy:

$$
\begin{aligned}
&\left(\mathbf{A}^T\right)^T = \mathbf{A}, \quad \left(\mathbf{A}^H\right)^H = \mathbf{A}, \\
&(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T, \quad (\mathbf{A}+\mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H, \\
&(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T, \quad (\alpha\mathbf{A})^H = \alpha^*\mathbf{A}^H,
\end{aligned}
$$

for any matrix $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{I \times J}$ and any scalar $\alpha \in \mathbb{C}$.

Vectorization

A very widely used operation in matrix computation is vectorization which consists of stacking the columns of a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ on top of each other to form a column vector of dimension JI:

$$\mathbf{A} = \left[\begin{array}{ccc} \mathbf{A}_{.1} & \cdots & \mathbf{A}_{.J} \end{array} \right] \in \mathbb{K}^{I \times J} \Rightarrow \mathrm{vec}(\mathbf{A}) = \left[\begin{array}{c} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_{.J} \end{array} \right] \in \mathbb{K}^{JI}.$$

Since the operator vec satisfies $\mathrm{vec}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\mathrm{vec}(\mathbf{A}) + \beta\mathrm{vec}(\mathbf{B})$ for all $\alpha, \beta \in \mathbb{K}$, it is linear.

Inner product

We recall the definition of the inner product (also called dot or scalar product) of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{K}^I$, denoted by $\langle \mathbf{a}, \mathbf{b} \rangle$ or $\mathbf{b}^T\mathbf{a}$ if $\mathbb{K} = \mathbb{R}$, and $\mathbf{b}^H\mathbf{a}$ if $\mathbb{K} = \mathbb{C}$. This binary operation satisfies the properties of an Euclidean inner product. In $\mathbb{R}^I$, it is defined as:

$$\langle \cdot, \cdot \rangle : \mathbb{R}^I \times \mathbb{R}^I \to \mathbf{R}, \quad (\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^T\mathbf{a} = \sum_{i=1}^{I} a_i b_i.$$

It is easy to verify that this operation is a bilinear form. In $\mathbb{C}^I$, the definition of the inner product is given by:

$$\langle \cdot, \cdot \rangle : \mathbb{C}^I \times \mathbb{C}^I \to \mathbf{C}, \quad (\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{b}^H\mathbf{a} = \sum_{i=1}^{I} a_i b_i^*.$$

Inner product

In $\mathbb{C}^I$, the inner product is a sesquilinear form, which means:

- linearity with respect to the first argument:

$$\langle \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2, \mathbf{b} \rangle = \alpha_1 \langle \mathbf{a}_1, \mathbf{b} \rangle + \alpha_2 \langle \mathbf{a}_2, \mathbf{b} \rangle \qquad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C}.$$

- Hermitian symmetry:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle^*.$$

The inner product in $\mathbb{C}^I$ is said to be semilinear with respect to the second argument, namely:

$$\langle \mathbf{a}, \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 \rangle = \alpha_1^* \langle \mathbf{a}, \mathbf{b}_1 \rangle + \alpha_2^* \langle \mathbf{a}, \mathbf{b}_2 \rangle \qquad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C}.$$

By equipping $\mathbb{K}^I$ with an inner product, we obtain a Euclidean R.-v.s. if $\mathbb{K} = \mathbb{R}$ or Hermitian $\mathbb{C}$-v.s. if $\mathbb{K} = \mathbb{C}$.

Euclidean/Hermitian norm

The Euclidean (Hermitian) norm of a vector a, denoted $\|a\|$, associates to $\mathbf{a} \in \mathbb{R}^{\boldsymbol{I}}\left(\mathbf{a} \in \mathbb{C}^{\boldsymbol{I}}\right)$ a non-negative real number according to the following definition:

$$\|\cdot\|_2 : \mathbb{K}^{\boldsymbol{I}} \to \mathbb{R}^+$$

$$\mathbf{a} \mapsto \|\mathbf{a}\|_2 = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}.$$

The properties of the inner product guarantee that $\|\mathbf{a}\|_2 = 0$ if and only if $\mathbf{a} = 0$.

Orthogonality

Two vectors $\mathbf{a}$ and $\mathbf{b}$ of $\mathbb{K}^I$ are said to be orthogonal if and only if $\langle \mathbf{a}, \mathbf{b} \rangle = 0$. This terminology comes from the fact that the inner product in the Euclidean space R^I satisfies:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \|\mathbf{a}\|_2 \|\mathbf{b}\|_2 \cos\theta,$$

where θ is the angle between vectors $\mathbf{a}$ and $\mathbf{b}$.

Definition and properties

Given matrices $\mathbf{B} \in \mathbb{K}^{I \times J}$ and $\mathbf{A} \in \mathbb{K}^{J \times K}$, the product of $\mathbf{B}$ by $\mathbf{A}$ gives a matrix $\mathbf{C} = \mathbf{BA} \in \mathbb{K}^{I \times K}$ such that $c_{ik} = \sum_{j=1}^{J} b_{ij} a_{jk}$, for $i \in \langle I \rangle; k \in \langle K \rangle$. It is said that $\mathbf{A}$ is pre-multiplied by $\mathbf{B}$ and $\mathbf{B}$ is post-multiplied by $\mathbf{A}$.

This product can be written in terms of the outer products of column vectors of $\mathbf{B}$ with transposed row vectors of $\mathbf{A}$:

$$\mathbf{C} = \sum_{j=1}^{J} \mathbf{B}_{.j} \circ \mathbf{A}_{j.}^{T} = \sum_{j=1}^{J} \mathbf{B}_{.j} \mathbf{A}_{j.} \tag{1}$$

From equation (1), it can be deduced that row i and column j of $\mathbf{C}$ are given by $\mathbf{C}_{i.} = [\mathbf{BA}]_{i.} = \sum_{j=1}^{J} b_{ij} \mathbf{A}_{j.}$ and $\mathbf{C}_{.k} = [\mathbf{BA}]_{.k} = \sum_{j=1}^{J} a_{jk} \mathbf{B}_{.j}$ which shows that the rows of $\mathbf{BA}$ are linear combinations of the rows of $\mathbf{A}$, while the columns of $\mathbf{BA}$ are linear combinations of the columns of B. We also have:

$$\mathbf{C}_{i.} = \mathbf{B}_{i.} \mathbf{A}, \quad \mathbf{C}_{.k} = \mathbf{B} \mathbf{A}_{.k}.$$

Definition and properties

The matrix product satisfies the following properties:

- The product is not commutative, since in general $\mathbf{BA} \neq \mathbf{AB}$.
- The product is associative:

$$(\mathbf{BA})\mathbf{C} = \mathbf{B}(\mathbf{AC}) = \mathbf{BAC}.$$

- The product is distributive over the addition:

$$\mathbf{B}(\mathbf{A}+\mathbf{C}) = \mathbf{BA}+\mathbf{BC},\ (\mathbf{B}+\mathbf{D})\mathbf{C} = \mathbf{BC}+\mathbf{DC}.$$

- The product is associative over scalar multiplication:

$$\mathbf{B}(\lambda\mathbf{A}) = (\lambda\mathbf{B})\mathbf{A} = \lambda(\mathbf{BA}), \quad \forall\lambda \in \mathbb{K}.$$

- The transpose and conjugate transpose of a product of matrices are such that:

$$(\mathbf{BA})^T = \mathbf{A}^T\mathbf{B}^T, \quad (\mathbf{BA})^H = \mathbf{A}^H\mathbf{B}^H.$$

Powers of a matrix

The nth power of a square matrix is defined for every natural integer n as follows:

$$\mathbf{A}^n = \begin{cases} \mathbf{I}, & n = 0, \\ \mathbf{A}, & n = 1, \\ \mathbf{A}\mathbf{A}^{n-1}, & n > 1, \end{cases}$$

thus satisfying $\mathbf{A}^n\mathbf{A}^p = \mathbf{A}^{n+p}$ and $(\mathbf{A}^n)^p = \mathbf{A}^{np}$ for all $n, p \in \mathbb{N}$. It is easy to also show that the nth power of a transposed matrix is equal to the transpose of the nth power of the matrix: $\left(\mathbf{A}^T\right)^n = (\mathbf{A}^n)^T$.

定理 (Binomial theorem)

For $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times I}$ are commutative, we have:

$$(\mathbf{A} + \mathbf{B})^n = \sum_{q=0}^{n} C_n^q \mathbf{A}^{n-q}\mathbf{B}^q, \ \textit{where } C_n^q = \frac{n!}{q!(n-q)!}.$$

Powers of a matrix

Matrices	Conditions	
$\mathbf{A}$ periodic of period $n \geq 1$	If	$\mathbf{A}^{n+1} = \mathbf{A}$
$\mathbf{A}$ nilpotent of degree (or index) $n \geq 1$	If	$\mathbf{A}^{n-1} \neq \mathbf{0}$ and $\mathbf{A}^{n} = \mathbf{0}$
$\mathbf{A}$ idempotent	If	$\mathbf{A}^2 = \mathbf{A}$

图: Periodic/Nilpotent/Idempotent matrices

定理

If $\mathbf{A}$ is nilpotent, then the trace of $\mathbf{A}$ is zero, that is, tr($\mathbf{A}$) = 0.

Examples

- The matrix $\mathbf{x}\mathbf{y}^T$, with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^I$, is nilpotent of degree 2 if and only if vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.
- Any strictly upper (or lower) triangular matrix of order n is nilpotent of degree $k \leq n$. Thus, lower $(\mathbf{D}_n)$ shift matrix of order n :

$$\mathbf{D}_n = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

is nilpotent of degree n, because $\mathbf{D}_n^k \neq \mathbf{0}$, $\forall k \in \langle n-1 \rangle$, and $\mathbf{D}_n^n = \mathbf{0}$.

Examples

- Computation of powers of a unit triangular matrix $\mathbf{A} = \mathbf{I}_n + \mathbf{B}$, where $\mathbf{B}$ is a strictly upper (or lower) triangular square matrix of order n.
- Application to the decomposition of a lower triangular Toeplitz matrix in the form of a matrix polynomial:

$$\mathbf{A} = \begin{bmatrix} t_0 & 0 & \cdots & \cdots & 0 \\ t_1 & t_0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ t_{n-1} & \cdots & \cdots & t_1 & t_0 \end{bmatrix}.$$

定理

If $\mathbf{A}$ and $\mathbf{B}$ are lower triangular Toeplitz matrices of order n, their product $\mathbf{AB}$ is commutative and gives a lower triangular Toeplitz matrix.

Definition and properties of the trace

The trace of a square matrix $\mathbf{A}$ of order I is defined as the sum of its diagonal elements: $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{I} a_{ii}$.

The trace satisfies the following properties:

- $\mathrm{tr}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha\mathrm{tr}(\mathbf{A}) + \beta\mathrm{tr}(\mathbf{B})$,
- $\mathrm{tr}\left(\mathbf{A}^T\right) = \mathrm{tr}(\mathbf{A})$,
- $\mathrm{tr}\left(\mathbf{A}^*\right) = \mathrm{tr}\left(\mathbf{A}^H\right) = [\mathrm{tr}(\mathbf{A})]^*$,
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}b_{ji}$,
- $\mathrm{tr}\left(\mathbf{AC}^T\right) = \mathrm{tr}\left(\mathbf{C}^T\mathbf{A}\right) = \mathrm{tr}\left(\mathbf{A}^T\mathbf{C}\right) = \mathrm{tr}\left(\mathbf{CA}^T\right) = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij}c_{ij}$,
- $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB}) = \mathrm{tr}(\mathbf{BCA})$. This property is called the cyclic invariance of the trace, and it can be generalized to the product of any number of matrices. In general, $\mathrm{tr}(\mathbf{ABC}) \neq \mathrm{tr}(\mathbf{ACB})$ and $\mathrm{tr}(\mathbf{ABC}) \neq \mathrm{tr}(\mathbf{BAC})$.

Examples

例

Expression of a bilinear form and a quadratic form in terms of a matrix trace: from the relation $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ and from the fact that the trace of a scalar is the scalar itself, it can be deduced that:

$$\mathbf{y}^H\mathbf{x} = \sum_{i=1}^{I} y_i^* x_i = \mathrm{tr}\left(\mathbf{y}^H\mathbf{x}\right) = \mathrm{tr}\left(\mathbf{x}\mathbf{y}^H\right),$$

$$\mathbf{y}^H\mathbf{A}\mathbf{x} = \sum_{i=1}^{I}\sum_{j=1}^{J} a_{ij} y_i^* x_j = \mathrm{tr}\left(\mathbf{y}^H\mathbf{A}\mathbf{x}\right) = \mathrm{tr}\left(\mathbf{x}\mathbf{y}^H\mathbf{A}\right) = \mathrm{tr}\left(\mathbf{A}\mathbf{x}\mathbf{y}^H\right).$$

Matrix inner product

Similarly to the v.s. $\mathbb{K}^I$, the v.s. $\mathbb{K}^{I\times J}$ can be equipped with an inner product in order to give it a Euclidean space structure if $\mathbb{K} = \mathbb{R}$. This inner product is defined by:

$$\begin{gathered}\mathbb{R}^{I\times J} \times \mathbb{R}^{I\times J} \to \mathbb{R} \\ (\mathbf{A}, \mathbf{B}) \mapsto \langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr}\left(\mathbf{B}^T \mathbf{A}\right) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} b_{ij}\end{gathered}$$

in the real case. The following identities can be easily verified:

$$\operatorname{tr}\left(\mathbf{B}^T \mathbf{A}\right) = \operatorname{tr}\left(\mathbf{A}\mathbf{B}^T\right) = \langle \operatorname{vec}(\mathbf{A}), \operatorname{vec}(\mathbf{B}) \rangle = \operatorname{vec}^T(\mathbf{B}) \operatorname{vec}(\mathbf{A})$$

in $\mathbb{R}$.

Frobenius norm

The previous definition of the matrix inner product induces the Euclidean/Hermitian matrix norm as:

$$\begin{aligned} \|\cdot\|_F : & \mathbb{K}^{I\times J} \to \mathbb{R}^+ \\ \mathbf{A} \mapsto \|\mathbf{A}\|_F = \sqrt{\langle \mathbf{A}, \mathbf{A}\rangle} & = \sqrt{\sum_{i=1}^{I}\sum_{j=1}^{J} |a_{ij}|^2}. \end{aligned}$$

This norm is called the Frobenius norm. It is also called the Hilbert–Schmidt norm or the Schur norm. A natural link can be found between this norm and the Euclidean/Hermitian norm of a vector: $\mathrm{tr}\left(\mathbf{A}\mathbf{A}^H\right) = \mathrm{tr}\left(\mathbf{A}^H\mathbf{A}\right) = \sum_{i,j=1}^{I} |a_{ij}|^2 = \sum_{i=1}^{I} \|\mathbf{A}_{i.}\|_2^2 = \sum_{j=1}^{J} \|\mathbf{A}_{.j}\|_2^2 = \|\mathrm{vec}(\mathbf{A})\|_2^2$.

Subspaces associated with a matrix

Given the matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$, its column space, denoted $C(\mathbf{A})$, is the subspace of $\mathbb{K}^{I}$ spanned by the column vectors of $\mathbf{A}$:

$$C(\mathbf{A}) = \mathrm{lc}\left(\mathbf{A}_{.1}, \cdots, \mathbf{A}_{,J}\right).$$

Subspaces associated with a matrix

Given the matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$, its column space, denoted $C(\mathbf{A})$, is the subspace of $\mathbb{K}^{I}$ spanned by the column vectors of $\mathbf{A}$:

$$C(\mathbf{A}) = \mathrm{lc}\left(\mathbf{A}_{.1}, \cdots, \mathbf{A}_{,J}\right).$$

Similarly, the column space of $\mathbf{A}^{T}$, denoted $C\left(\mathbf{A}^{T}\right)$, is the subspace of $\mathbb{K}^{J}$ spanned by the column vectors of $\mathbf{A}^{T}$, in other words, the row vectors of $\mathbf{A}$. This space is called the row space of $\mathbf{A}$ and is denoted by $L(\mathbf{A})$. It is such that:

$$\begin{aligned} L(\mathbf{A}) &= \mathrm{lc}\left(\mathbf{A}_{1.}, \cdots, \mathbf{A}_{I.}\right) \\ &= C\left(\mathbf{A}^{T}\right). \end{aligned}$$

Subspaces associated with a matrix

$\mathbf{A}$ can be seen as the matrix of the linear map $L : \mathbb{K}^J \to \mathbb{K}^I$ such that $\mathrm{x} \mapsto \mathbf{A}\mathbf{x}$. The image space of the map, denoted by $\mathrm{Im}(\mathbf{A})$, can then be defined as:

$$\mathrm{Im}(\mathbf{A}) = \left\{ \mathbf{y} \in \mathbb{K}^I : \exists \mathbf{x} \in \mathbb{K}^J \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{x} \right\}$$

This subspace is also called the range of $\mathbf{A}$. We thus have:

$$C(\mathbf{A}) = \mathrm{Im}(\mathbf{A}) \subseteq \mathbb{K}^I.$$

Subspaces associated with a matrix

$\mathbf{A}$ can be seen as the matrix of the linear map $L : \mathbb{K}^J \to \mathbb{K}^I$ such that $\mathrm{x} \mapsto \mathbf{A}\mathbf{x}$. The image space of the map, denoted by $\mathrm{Im}(\mathbf{A})$, can then be defined as:

$$\mathrm{Im}(\mathbf{A}) = \left\{\mathbf{y} \in \mathbb{K}^I : \exists \mathbf{x} \in \mathbb{K}^J \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{x}\right\}$$

This subspace is also called the range of $\mathbf{A}$. We thus have:

$$C(\mathbf{A}) = \mathrm{Im}(\mathbf{A}) \subseteq \mathbb{K}^I.$$

Similarly, we have:

$$C\left(\mathbf{A}^T\right) = \mathrm{Im}\left(\mathbf{A}^T\right) = \left\{\mathbf{x} \in \mathbb{K}^J : \exists \mathbf{y} \in \mathbb{K}^I \text{ such that } \mathbf{x} = \mathbf{A}^T\mathbf{y}\right\} \subseteq \mathbb{K}^J.$$

Subspaces associated with a matrix

The kernel of $\mathbf{A}$, also called the null space of $\mathbf{A}$ and denoted by $\mathcal{N}(\mathbf{A})$, is defined as:

$$\mathcal{N}(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{K}^J : \mathbf{A}\mathbf{x} = \mathbf{0}_I \right\} \subseteq \mathbb{K}^J$$

The kernel $\mathcal{N}(\mathbf{A})$ can be viewed as the set of all solutions of the homogeneous system of equations $\mathbf{A}\mathbf{x} = \mathbf{0}_I$.

Similarly, the kernel of $\mathbf{A}^T$ can be defined as:

$$\mathcal{N}\left(\mathbf{A}^T\right) = \left\{ \mathbf{y} \in \mathbb{K}^I : \mathbf{A}^T\mathbf{y} = \mathbf{0}_J \right\} \subseteq \mathbb{K}^I$$

$\mathcal{N}\left(\mathbf{A}^T\right)$ is also called the left-hand kernel (or left nullspace) of $\mathbf{A}$ because it is the set of all solutions to $\mathbf{y}^T\mathbf{A} = \mathbf{0}_J^T$. It is worth noting that, in the case of complex matrices $(\mathbb{K} = \mathbb{C})$, $\mathbf{A}^T$ should be replaced by $\mathbf{A}^H$ in the previous definitions.

Subspaces associated with a matrix

In summary, with any matrix $\mathbf{A}$ we can associate the four fundamental subspaces given in the following table:

Subspaces	Notations
Column space	$C(\mathbf{A}) = \mathrm{Im}(\mathbf{A}) \subseteq \mathbf{K}^I$
Row space	$L(\mathbf{A}) = C\left(\mathbf{A}^T\right) \subseteq \mathbb{K}^I$
Kernel	$\mathcal{N}(\mathbf{A}) \subseteq \mathbf{K}^J$
Left-hand kernel	$\mathcal{N}\left(\mathbf{A}^T\right) \subseteq \mathbf{K}^I$

Table: Subspaces associated with a matrix

Subspaces associated with a matrix

定理 (Fundamental theorem of linear algebra)

Any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ *satisfies:*

$$\dim[\mathcal{C}(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A})] = J,$$

or equivalently:

$$\dim[\mathcal{L}(\mathbf{A})] + \dim\left[\mathcal{N}\left(\mathbf{A}^T\right)\right] = I.$$

Subspaces associated with a matrix

$C(\mathbf{A}) = [\mathcal{N}(\mathbf{A}^T)]^\perp = \mathrm{Im}(\mathbf{A})$	$C(\mathbf{A}^T) = [\mathcal{N}(\mathbf{A})]^\perp = L(\mathbf{A})$
$C(\mathbf{A}) \cap \mathcal{N}(\mathbf{A}^T) = \{\mathbf{0}_I\}$	$\mathcal{N}(\mathbf{A}) \cap C(\mathbf{A}^T) = \{\mathbf{0}_J\}$
$\dim[C(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A})] = J$	
$\dim[L(\mathbf{A})] + \dim[\mathcal{N}(\mathbf{A}^T)] = I$	

Table: Relations linking subspaces $C(\mathbf{A}), L(\mathbf{A})$, and $\mathcal{N}(\mathbf{A})$.

The equalities $C(\mathbf{A}) \cap \mathcal{N}\left(\mathbf{A}^T\right) = \{\mathbf{0}_I\}$ and $\mathcal{N}(\mathbf{A}) \cap C\left(\mathbf{A}^T\right) = \{\mathbf{0}_j\}$ arise from the fact that the intersection of two orthogonal subspaces is made up of the null vector. From this table, it can be deduced that any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ provides an orthogonal decomposition of the spaces $\mathbb{K}^I$ and $\mathbb{K}^J$ such as:

$$\mathbb{K}^I = C(\mathbf{A}) \oplus [C(\mathbf{A})]^\perp = C(\mathbf{A}) \oplus \mathcal{N}\left(\mathbf{A}^T\right),$$
$$\mathbb{K}^J = \mathcal{N}(\mathbf{A}) \oplus [\mathcal{N}(\mathbf{A})]^\perp = \mathcal{N}(\mathbf{A}) \oplus C\left(\mathbf{A}^T\right).$$

Definition and properties

A basic result of linear algebra states that the column rank and row rank of a matrix are equal, which leads to the following proposition.

定理

Any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ *is such that:*

$$\dim[\mathbf{C}(\mathbf{A})] = \dim[\mathbf{L}(\mathbf{A})].$$

From this result, the rank of a matrix can be defined as:

$$\mathrm{r}(\mathbf{A}) = \dim[\boldsymbol{C}(\mathbf{A})] = \dim[L(\mathbf{A})],$$

implying that the rank is equal to the maximal number of linearly independent columns or rows, such that for any matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$, we have:

$$\mathrm{r}(\mathbf{A}) \leq \min(I, J).$$

定理 (Rank theorem)

$$\mathrm{r}(\mathbf{A}) = J - \dim[\mathcal{N}(\mathbf{A})],$$
$$\mathrm{r}(\mathbf{A}) = I - \dim[\mathcal{N}(\mathbf{A}^T)].$$

Definition and properties

定理 (Rank theorem)

$$\mathrm{r}(\mathbf{A}) = J - \dim[\mathcal{N}(\mathbf{A})],$$
$$\mathrm{r}(\mathbf{A}) = I - \dim[\mathcal{N}(\mathbf{A}^T)].$$

- $\mathbf{A}$ is full column rank if and only if $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}_J\}$ and full row rank if $\mathcal{N}\left(\mathbf{A}^T\right) = \{\mathbf{0}_I\}$.
- $\mathrm{r}\left(\mathbf{A}^T\right) = \mathrm{r}\left(\mathbf{A}^*\right) = \mathrm{r}\left(\mathbf{A}^H\right) = \mathrm{r}(\mathbf{A})$.

Sum and difference rank

Operations	Ranks
Sum	$\mathrm{r}(\mathbf{A}+\mathbf{B}) \leq \mathrm{r}(\mathbf{A})+\mathrm{r}(\mathbf{B})$
Difference	$\mathrm{r}(\mathbf{A}-\mathbf{B}) \leq \lvert \mathrm{r}(\mathbf{A})-\mathrm{r}(\mathbf{B}) \rvert$

Table: Rank of a sum and difference of matrices

Subspaces associated with a matrix product

Matrices	Relations
$\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$	$C(\mathbf{A}_1\mathbf{A}_2) \subseteq C(\mathbf{A}_1)$
$\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$	$L(\mathbf{A}_1\mathbf{A}_2) \subseteq L(\mathbf{A}_2)$
$\mathbf{A}_1 \in \mathbb{K}^{I\times J}, \mathbf{A}_2 \in \mathbb{K}^{J\times K}$	$\mathcal{N}(\mathbf{A}_2) \subseteq \mathcal{N}(\mathbf{A}_1\mathbf{A}_2)$
$\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$	$C(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) \subseteq C(\mathbf{A}_1)$
$\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$	$L(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) \subseteq L(\mathbf{A}_N)$
$\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$	$\mathcal{N}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N) = \mathrm{Im}(\mathbf{A}_2\mathbf{A}_3\cdots\mathbf{A}_N) \cap \mathcal{N}(\mathbf{A}_1)$
$\mathbf{A}_1 \in \mathbb{K}^{I_1\times I_2}, \cdots, \mathbf{A}_N \in \mathbb{K}^{I_N\times I_{N+1}}$	$\mathcal{N}(\mathbf{A}_N) \subseteq \mathcal{N}(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_N)$
$\mathbf{A} \in \mathbb{K}^{I\times I}$	$C\left(\mathbf{A}^k\right) \subseteq C(\mathbf{A})$
$\mathbf{A} \in \mathbb{K}^{I\times I}$	$\mathcal{N}(\mathbf{A}) \subseteq \mathcal{C}\left(\mathbf{A}^k\right)$
$\mathbf{A} \in \mathbb{K}^{I\times J}$	$C\left(\mathbf{A}^H\mathbf{A}\right) = C\left(\mathbf{A}^H\right), \mathcal{N}\left(\mathbf{A}^H\mathbf{A}\right) = \mathcal{N}(\mathbf{A})$
$\mathbf{A} \in \mathbb{K}^{I\times J}$	$C\left(\mathbf{A}\mathbf{A}^H\right) = C(\mathbf{A}), \mathcal{N}\left(\mathbf{A}\mathbf{A}^H\right) = \mathcal{N}\left(\mathbf{A}^H\right)$

Table: Subspaces associated with a matrix product

From this table, it can be concluded that the Gram matrix $\mathbf{A}^H\mathbf{A}$ has the same column space as $\mathbf{A}^H$ and the same kernel as $\mathbf{A}$, while the Gram matrix $\mathbf{A}\mathbf{A}^H$ has the same column space as $\mathbf{A}$ and the same kernel as $\mathbf{A}^H$. Consequently, the Gram matrices $\mathbf{A}^H\mathbf{A}$ and $\mathbf{A}\mathbf{A}^H$ have the same rank as $\mathbf{A}$.

Product rank

定理

The rank of the product of two matrices is less than or equal to the minimum between the ranks of the two matrices:

$$\mathrm{r}(\mathbf{AB}) \leq \min(\mathrm{r}(\mathbf{A}), \mathrm{r}(\mathbf{B})).$$

From this result, one can deduce that the rank of the outer product of two non-zero vectors is equal to 1:

$$\mathrm{r}(\mathbf{a} \circ \mathbf{b}) = \mathrm{r}\left(\mathbf{a}\mathbf{b}^{T}\right) = 1.$$

Product rank

Properties	Ranks
$\mathbf{B}$ and $\mathbf{C}$ non-singular	$\mathrm{r}(\mathbf{CA}) = \mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{CAB}) = \mathrm{r}(\mathbf{A})$
	$\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{B}) - \dim\left(\mathrm{Im}(\mathbf{B}) \bigcap \mathcal{N}(\mathbf{A})\right)$
$\mathbf{A} \in \mathbb{K}^{I\times J}, \mathbf{B} \in \mathbb{K}^{J\times K}$	$\mathrm{r}(\mathbf{A}) + \mathrm{r}(\mathbf{B}) - J \leq \mathrm{r}(\mathbf{AB}) \leq \min\{\mathrm{r}(\mathbf{A}), \mathrm{r}(\mathbf{B})\}$
	$\mathrm{r}(\mathbf{ABC}) \geq \mathrm{r}(\mathbf{AB}) + \mathrm{r}(\mathbf{BC}) - \mathrm{r}(\mathbf{B})$
Gram matrices	$\mathrm{r}(\mathbf{A}^H\mathbf{A}) = \mathrm{r}(\mathbf{A}\mathbf{A}^H) = \mathrm{r}(\mathbf{A})$.

Table: Rank of a matrix product.

Determinant

The study of determinants has played an important role in the development of matrix algebra. The determinant of a square matrix:

- provides a criterion for invertibility of a matrix;
- can be employed for the calculation of the inverse and thereby for solving linear systems of equations;
- enables the measurement of areas and volumes;
- is directly related to the notion of characteristic polynomial and thus of eigenvalue.

Determinant

The determinant of the square matrix $\mathbf{A}$ of order I is the scalar given by:

$$\det(\mathbf{A}) = \sum_{p} \sigma(p) a_{1,p_1} a_{2,p_2} \cdots a_{I,p_I} \tag{2}$$

(132)

(1

where the summation involves the $I!$ permutations $p = (p_1, p_2, \cdots, p_I)$ of the set $\{1, 2, \cdots, I\}$, and $\sigma(p)$ refers to the signature of permutation p defined as:

$$\sigma(p) = \begin{cases} +1, & \text{if the initial order } 1, \cdots, I \text{ can be found from } p \\ & \text{using an even number of elementary permutations;} \\ -1, & \text{if the initial order } 1, \cdots, I \text{ can be found from } p \\ & \text{using an odd number of elementary permutations,} \end{cases}$$

an elementary permutation corresponding to the inversion of two elements of the I-tuple $(p_1, p_2, \cdots, p_I)$. The expression (2) of the determinant is called the Leibniz formula.

Determinant

The determinant can also be computed using the following formula called the Laplace expansion with respect to row i :

$$\det(\mathbf{A}) = \sum_{j=1}^{I} (-1)^{i+j} a_{ij} \det\left(\mathbf{A}_i^j\right), \quad i \in \langle I \rangle$$

where $\mathbf{A}_i^j$ is a square submatrix of $\mathbf{A}$, of order $I-1$, obtained by removing the i th row and the j th column in $\mathbf{A}$. The determinant $\det(\mathbf{A}_i^j)$ is called the minor associated with coefficient a_{ij}, and the minor preceded by the signature depending on the element's position (i,j), namely, $(-1)^{i+j}\det(\mathbf{A}_i^j)$, is called the cofactor of a_{ij}. A similar formula based on the expansion of the determinant with respect to column j can also be used.

Determinant

KA

In general, the definition based on the notion of permutation is easier to apply than the Laplace formula for the calculation of the determinant of a high-order square matrix. However, for certain matrices, such as sparse matrices, the Laplace expansion may lead to faster computation.

The determinant satisfies the following properties:

- The determinant changes the sign if two rows (or columns) of $\mathbf{A}$ are swapped.
- If a row (or column) is multiplied by a scalar k, then the determinant is multiplied by k.
- The determinant remains unmodified if k times row (or column) i is added to row (or column) j.
- In addition, for a square matrix of order I, we have:

$$\det\left(\mathbf{A}^T\right) = \det(\mathbf{A}); \quad \det\left(\mathbf{A}^H\right) = \det\left(\mathbf{A}^*\right) = [\det(\mathbf{A})]^*,$$

$$\det(\alpha\mathbf{A}) = \alpha^I \det(\mathbf{A}), \quad \forall \alpha \in \mathbb{K},$$

$$\det(-\mathbf{A}) = (-1)^I \det(\mathbf{A}).$$

Determinant

- The determinant is a multiplicative mapping, that is, it satisfies the following properties.

定理

For all square matrices $\mathbf{A}, \mathbf{B}$, *and* $\mathbf{C}$, *we have:*

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$
$$\det(\mathbf{ABC}) = \det(\mathbf{A})\det(\mathbf{B})\det(\mathbf{C}).$$

Matrix inversion

$\boxed{Ax = 0}$

The inverse of a square matrix $\mathbf{A}$ of order I is a matrix $\mathbf{X}$ such that $\mathbf{AX} = \mathbf{XA} = \mathbf{I}_I$. When such a matrix exists, it is said that $\mathbf{A}$ is invertible or non-singular, or still regular, and it is denoted by $\mathbf{A}^{-1}$. Otherwise, it is said that $\mathbf{A}$ is singular.

The inverse $\mathbf{A}^{-1}$ is unique and given by:

$\det(A) =$ A $\boxed{A_C^T}$

$$\boxed{\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \mathbf{A}_C^T,}$$

where the square matrix $\mathbf{A}_C$, of order I, is the matrix formed by the cofactors of $\mathbf{A}$.

Matrix inversion: an example

Inversion of a unit triangular matrix $\mathbf{A} = \mathbf{I}_n + \mathbf{B}$, where $\mathbf{B}$ is nilpotent of degree n, that is, $\mathbf{B}^n = \mathbf{0}_n$

$\rho(B) < \|B\| < 1$

$$(\mathbf{I}_n + \mathbf{B})^{-1} = \mathbf{I}_n - \mathbf{B} + \mathbf{B}^2 + \cdots + (-1)^{n-1}\mathbf{B}^{n-1}.$$

Indeed, taking the nilpotence property of $\mathbf{B}$ into account, we verify that:

$$(\mathbf{I}_n + \mathbf{B})\left(\mathbf{I}_n - \mathbf{B} + \mathbf{B}^2 + \cdots + (-1)^{n-1}\mathbf{B}^{n-1}\right) = \mathbf{I}_n + (-1)^{n-1}\mathbf{B}^n = \mathbf{I}_n.$$

For instance, for $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$, we obtain:

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}.$$

Matrix inversion

For a square matrix $\mathbf{A}$ of order I, the properties listed in the following table are equivalent in terms of invertibility. One should bear in mind that $\dim[\mathcal{N}(\mathbf{A})] = 0$ is the result of the rank theorem applied to a square matrix $\mathbf{A}$ of full rank, that is, when $\mathbf{r}(\mathbf{A}) = I$.

Column vectors are linearly independent
Row vectors are linearly independent
$\mathrm{r}(\mathbf{A}) = I$
$\det(\mathbf{A}) \neq 0$
$\dim[\mathcal{N}(\mathbf{A})] = 0$

满秩

full rank

Table: Several manners of characterizing the invertibility of a square matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$

Matrix inversion

$2x_1 + x_2 = 3$
$x_1 + 2x_2 = 3$

$(A\ B)$

$n \times n$ $Ax = b$ $n!$ n^3

$\frac{2}{3}n^3$ LU QR $A = LU$

Let $\mathbf{A}, \mathbf{B} \in \mathbb{K}^{I \times I}$ denote two invertible matrices. We have:

- $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$
- $\left(\mathbf{A}^{*}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{*}$
- $\left(\mathbf{A}^{T}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{T}, \quad \left(\mathbf{A}^{H}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{H}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $\left(\mathbf{A}^{k}\right)^{-1} = \left(\mathbf{A}^{-1}\right)^{k}$

$\frac{1}{10}$ 0.1

$L \cdot \underbrace{U \cdot x}_{y} = b$

n^2 $Ly = b$

SVD

$Ux = y$

$\det(\lambda I - A) = 0$

$\lambda^n + \cdots + a_0 = 0$

$A = U \Sigma V^T$

$\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$

Solution of a homogeneous system of linear equations

As indicated in previous table, for a non-singular matrix $\mathbf{A}$, we have $\dim[\mathcal{N}(\mathbf{A})] = 0$, in other words, $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$. Consequently, the homogeneous system $\mathbf{Ax} = \mathbf{0}$ then only admits the trivial solution $\mathbf{x} = \mathbf{0}$.

More generally, if $\mathbf{A}$ and $\mathbf{B}$ are square matrices and $\mathbf{AB} = \mathbf{0}$, then either $\mathbf{A}$ is singular, or $\mathbf{B} = \mathbf{0}$. By symmetry, the same reasoning applies to $\mathbf{B}$: indeed, $\mathbf{AB} = \mathbf{0}$ implies $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T = \mathbf{0}$, and so either $\mathbf{B}^T$ is singular (which amounts to saying that $\mathbf{B}$ is singular), or $\mathbf{A}^T = \mathbf{0}$, and thus $\mathbf{A} = \mathbf{0}$. It can be, therefore, concluded that $\mathbf{AB} = \mathbf{0}$ when (at least) one of the following conditions is met: (i) $\mathbf{A} = \mathbf{0}$; (ii) $\mathbf{B} = \mathbf{0}$; (iii) $\mathbf{A}$ and $\mathbf{B}$ are both singular.

$A = U_1 V_1^T$

$B = U_2 V_2^T$

Complex matrix inverse

$a = x_1 + i y_1$

$b = x_2 + i y_2$

$(X + iY)(C + iD) = I$

$XC - YD = I$ $\quad$ $XD + YC = 0$

Let $\mathbf{A} \in \mathbb{C}^{I \times I}$ be a complex matrix decomposed using its real and imaginary parts as $\mathbf{A} = \mathbf{X} + j\mathbf{Y}$, with $\mathbf{X} \in \mathbb{R}^{I \times I}$ and $\mathbf{Y} \in \mathbb{R}^{I \times I}$.

- If $\mathbf{X}$ is invertible, we have:

$$\mathbf{A}^{-1} = \left(\mathbf{I}_I - j\mathbf{X}^{-1}\mathbf{Y}\right)\left(\mathbf{X} + \mathbf{Y}\mathbf{X}^{-1}\mathbf{Y}\right)^{-1}.$$

- Similarly, if $\mathbf{Y}$ is invertible, we can write:

$$\mathbf{A}^{-1} = \left(\mathbf{Y}^{-1}\mathbf{X} - j\mathbf{I}_I\right)\left(\mathbf{Y} + \mathbf{X}\mathbf{Y}^{-1}\mathbf{X}\right)^{-1}.$$

$$a = \begin{pmatrix} x_1 & y_1 \\ -y_1 & x_1 \end{pmatrix} \qquad b = \begin{pmatrix} x_2 & y_2 \\ -y_2 & x_2 \end{pmatrix}$$

Orthogonal and unitary matrices

$A^T A x = A^T b$

$QR x = Q^T b$? $\boxed{QR}$ SVD

A square matrix $\mathbf{A} \in \mathbb{R}^{I \times I}$ is said to be orthogonal if it satisfies $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$. A matrix is orthogonal if and only if one of the following conditions is satisfied:

- its rows are orthonormal, that is, $\langle \mathbf{A}_{i.}^T, \mathbf{A}_{j.}^T \rangle = \delta_{i,j}$ for all $i, j \in \langle I \rangle$.
- its columns are orthonormal, that is, $\langle \mathbf{A}_{.i}, \mathbf{A}_{.j} \rangle = \delta_{i,j}$ for all $i, j \in \langle I \rangle$.

Another way of characterizing orthogonal matrices consists of saying that A is orthogonal if and only if its inverse is equal to its transpose $\left(\mathbf{A}^{-1} = \mathbf{A}^T\right)$, which implies that the inverse of an orthogonal matrix is itself orthogonal.

Orthogonal and unitary matrices

For an orthogonal matrix $(\mathbf{A}\mathbf{A}^T = \mathbf{I}\ )$ and of a unitary matrix $(\mathbf{A}\mathbf{A}^H = \mathbf{I})$, the following results can be deduced:

- The determinant of an orthogonal matrix is equal to ± 1.
- The determinant of a unitary matrix is of modulus 1.

Another important property of an orthogonal/unitary matrix $\mathbf{A}$ is the invariance of the norm of a vector when it is multiplied by $\mathbf{A}$:

- $\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \in \mathbb{R}^I$,
- $\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{x}^H\mathbf{A}^H\mathbf{A}\mathbf{x} = \|\mathbf{x}\|_2^2 \quad \forall \mathbf{x} \in \mathbb{C}^I$.

Involutory matrices and anti-involutory matrices

$A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$ $A^T = A$ $\boxed{A^T = -A}$ bi

A square matrix $\mathbf{A}$ of order I is said to be involutory (对合) if it is equal to its inverse

$QΛQ^T$

$$\mathbf{A}^{-1} = \mathbf{A} \Rightarrow \mathbf{A}^2 = \mathbf{I}_I.$$

$H = I - 2ww^T$ $\|w\|_2 = 1$

When $\mathbf{A}$ is equal to its inverse with the opposite sign, that is, $\mathbf{A}^{-1} = -\mathbf{A}$ (which implies $\mathbf{A}^2 = -\mathbf{I}_I$), it is said to be anti-involutory.

QR方法

A real (complex) symmetric (Hermitian) matrix is involutory if and only if it is orthogonal (unitary). Indeed, when $\mathbf{A} \in \mathbb{R}^{I \times I}$ and $\mathbf{A}^T = \mathbf{A}$, we have:

$$\mathbf{A}^{-1} = \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_I.$$

$B = \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & \\ & & 1 & -a_{n-1} \end{pmatrix}$

Similarly, if $\mathbf{A} \in \mathbb{C}^{I \times I}$ and $\mathbf{A}^H = \mathbf{A}$, we get:

$$\mathbf{A}^{-1} = \mathbf{A} \quad \Leftrightarrow \quad \mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A} = \mathbf{I}_I.$$

$\det(\lambda I - B)$

roots

$\det(\lambda I - A) = \boxed{\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0}$

Left and right inverses of a rectangular matrix

A matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$ is said to be left (or right) invertible if there exists a matrix $\mathbf{A}_G^{-1}$ (or $\mathbf{A}_D^{-1}$) $\in \mathbb{K}^{J\times I}$ such that: $\mathbf{A}_G^{-1}\mathbf{A} = \mathbf{I}_J$ (or $\mathbf{A}\mathbf{A}_D^{-1} = \mathbf{I}_I$). Matrices $\mathbf{A}_G^{-1}$ and $\mathbf{A}_D^{-1}$ are, respectively, called left-inverse and right-inverse of $\mathbf{A}$.

B

$J\times 1 \Leftarrow BA = I_J$

$I\times J$

When $\mathbf{A}$ is square and non-singular, the two inverses are equal: $\mathbf{A}_G^{-1} = \mathbf{A}_D^{-1} = \mathbf{A}^{-1}$. For $I \neq J$, these inverses are not unique, and they can exist on one side only: on the left if $I > J$ or on the right if $I < J$.

定理 (Existence theorem of left and right inverses (Perlis 1958))

Given a matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$, we have:

- *A has a left-inverse if and only if $I \geq J$ and $\mathrm{r}(\mathbf{A}) = J$, that is, if $\mathbf{A}$ has full column rank.*
- *A has a right-inverse if and only if $J \geq I$ and $\mathrm{r}(\mathbf{A}) = I$, that is, if $\mathbf{A}$ has full row rank.*

Formulae for left and right inverses

Given a matrix $\mathbf{A} \in \mathbb{K}^{I \times J}$ of full rank, we have:

- If $I \geq J$, all left inverses are given by the partitioned formula:

$$\mathbf{A}_G^{-1} = \left[\mathbf{A}_1^{-1} - \mathbf{Y}\mathbf{A}_2\mathbf{A}_1^{-1} \quad \mathbf{Y}\right]\mathbf{P},$$

  where $\mathbf{P} \in \mathbb{K}^{I \times I}$ is a permutation matrix such that:

$$\boxed{A_1^{-1} \mid 0} \quad BP^{-1} \qquad \mathbf{P}\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \qquad = I_J$$

  with $\mathbf{A}_1 \in \mathbb{K}^{J \times J}$ non-singular, $\mathbf{A}_2 \in \mathbb{K}^{(I-J) \times J}$, and $\mathbf{Y} \in \mathbb{K}^{J \times (I-J)}$ arbitrary.

- If $J \geq I$, all right-inverses are given by:

$$\mathbf{A}_D^{-1} = \mathbf{P}\begin{bmatrix} \mathbf{A}_1^{-1} - \mathbf{A}_1^{-1}\mathbf{A}_2\mathbf{Y} \\ \mathbf{Y} \end{bmatrix},$$

  where $\mathbf{P} \in \mathbb{K}^{J \times J}$ is a permutation matrix such that $\mathbf{A}\mathbf{P} = [\mathbf{A}_1 \mathbf{A}_2]$, with $\mathbf{A}_1 \in \mathbb{K}^{I \times I}$ non-singular, $\mathbf{A}_2 \in \mathbb{K}^{I \times (J-I)}$, and $\mathbf{Y} \in \mathbb{K}^{(J-I) \times I}$ arbitrary.

Generalized inverses

The left and right inverses of a rectangular matrix exist if and only if the matrix has full rank. A generalized inverse of $\mathbf{A} \in \mathbb{K}^{I \times J}$ is defined as a matrix $\mathbf{A}^{\#} \in \mathbb{K}^{J \times I}$ that satisfies the following equations:

$$\mathbf{A}\mathbf{A}^{\#}\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^{\#}\mathbf{A}\mathbf{A}^{\#} = \mathbf{A}^{\#}.$$

It can be shown that any matrix has a generalized inverse. Nonetheless, this generalized inverse is not unique in general. In fact, it is unique if and only if $\mathbf{A} = \mathbf{0}$ or if $\mathbf{A}$ is a regular square matrix.

Any generalized inverse of a singular matrix $\mathbf{A}$ can be obtained using a bordering technique by building a non-singular bordered matrix $\mathbf{M}$:

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{D} & \mathbf{B} \end{bmatrix}$$

The submatrix corresponding to $\mathbf{A}$ in $\mathbf{M}^{-1}$ is a generalized inverse of $\mathbf{A}$.

$Ax=b$ $\quad x=A^{-1}b$ $\quad \min \|b-Ax\|_2$ $\quad \boxed{x=(A^TA)^{-1}A^Tb}$

If $\mathbf{A}$ is of full column rank and $I > J$, then $\mathbf{A}^T\mathbf{A}$ is invertible, and the generalized inverse of $\mathbf{A}$ is given by:

$Ax = A(A^TA^{-1})A^Tb$

$[A \; b]$

$$\mathbf{A}^{\#} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T.$$

$R(A)$ $\quad P^T = P$

It is indeed possible to find a matrix $\mathbf{C} \in \mathbb{K}^{I\times(I-J)}$ of full column rank $\left(\det\left(\mathbf{C}^T\mathbf{C}\right) \neq 0\right)$ such that $\mathbf{M} = [\mathbf{A}\,\mathbf{C}]$ is regular, with $\mathbf{C}^T\mathbf{A} = \mathbf{0}$. Equivalently, we have $\mathbf{A}^T\mathbf{C} = \mathbf{0}$. Since $\mathbf{M}^{-1} = \left(\mathbf{M}^T\mathbf{M}\right)^{-1}\mathbf{M}^T$ for any regular matrix $\mathbf{M}$, we have:

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^T\mathbf{C} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{C}^T \end{bmatrix} = \begin{bmatrix} (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \\ (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{\#} \\ \mathbf{C}^{\#} \end{bmatrix}.$$

$\begin{bmatrix} A^T \\ C^T \end{bmatrix}$ $\quad [A \; C]$

Generalized inverses: full row rank

If $\mathbf{A}$ is of full row rank and $J > I$, then similarly $\mathbf{A}\mathbf{A}^T$ is regular and the generalized inverse of $\mathbf{A}$ is given by:

$$\mathbf{A}^{\#} = \mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1}.$$

Indeed, we can find a matrix $\mathbf{D} \in \mathbb{K}^{(J-I)\times J}$ of full row rank $\left(\det\left(\mathbf{D}\mathbf{D}^T\right) \neq 0\ \right)$ such that

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} \\ \mathbf{D} \end{bmatrix}$$

is regular, with $\mathbf{A}\mathbf{D}^T = \mathbf{0}$. By writing that $\mathbf{M}^{-1} = \mathbf{M}^T \left(\mathbf{M}\mathbf{M}^T\right)^{-1}$, a similar reasoning to the previous case leads to the following formula:

$$\mathbf{M}^{-1} = \left[\mathbf{A}^T \left(\mathbf{A}\mathbf{A}^T\right)^{-1} \ \mathbf{D}^T \left(\mathbf{D}\mathbf{D}^T\right)^{-1}\right] = \left[\mathbf{A}^{\#} \ \mathbf{D}^{\#}\right].$$

Moore-Penrose pseudo-inverse

伪

$\boxed{\text{latex}}$ 中 dagger

The uniqueness of the generalized inverse can be obtained by imposing additional constraints. This is the case of the Moore-Penrose pseudo-inverse, denoted by $\mathbf{A}^\dagger$, which is defined using the four following relations:

$$\boxed{\begin{array}{l} \bullet\ \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, \\ \bullet\ \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger, \\ \bullet\ (\mathbf{A}^\dagger\mathbf{A})^T = \mathbf{A}^\dagger\mathbf{A}, \checkmark \\ \bullet\ (\mathbf{A}\mathbf{A}^\dagger)^T = \mathbf{A}\mathbf{A}^\dagger. \checkmark \end{array}}$$

When $\mathbb{K} = \mathbb{C}$, that is, in the case of a complex matrix, transposition must be replaced by conjugate transposition. Relations (iii) and (iv) denote that $\mathrm{A}^\dagger\mathrm{A}$ and $\mathbf{A}\mathbf{A}^\dagger$ are symmetric if $\mathbb{K} = \mathbb{R}$ and Hermitian if $\mathbb{K} = \mathbb{C}$, respectively.

Moore-Penrose pseudo-inverse: uniqueness

定理

The Moore-Penrose pseudo-inverse of a matrix A is unique.

证明.

Let $\mathbf{X}$ and $\mathbf{Y}$ denote two pseudo-inverses of $\mathbf{A}$. The use of the relations of definition of Moore-Penrose pseudo-inverse for $\mathbf{X}$ and $\mathbf{Y}$, as indicated in the following equalities:

$$\mathbf{X} \overset{\text{(ii)}}{=} \mathbf{XAX} \overset{\text{(iv)}}{=} \mathbf{XX}^H\mathbf{A}^H \overset{\text{(i)}}{=} \mathbf{XX}^H(\mathbf{AYA})^H = \mathbf{X}(\mathbf{AX})^H(\mathbf{AY})^H$$
$$\overset{\text{(iv)}}{=} (\mathbf{XAX})(\mathbf{AY}) \overset{\text{(i)}}{=} \mathbf{XAY} \overset{\text{(iii)}}{=} \mathbf{A}^H\mathbf{X}^H\mathbf{Y} \overset{\text{(i)}}{=} (\mathbf{AYA})^H\mathbf{X}^H\mathbf{Y}$$
$$= (\mathbf{YA})^H(\mathbf{XA})^H\mathbf{Y} \overset{\text{(iii)}}{=} \mathbf{Y}(\mathbf{AXA})\mathbf{Y} \overset{\text{(i)}}{=} \mathbf{YAY} \overset{\text{(ii)}}{=} \mathbf{Y}$$

□

Moore-Penrose pseudo-inverse: properties

	Properties	Conditions
(i)	$\mathbf{A}^\dagger = \mathbf{A}^{-1}$	$\mathbf{A}$ non-singular
(ii)	$(\mathbf{A}^\dagger)^\dagger = \mathbf{A}$	
(iii)	$(\mathbf{A}^H)^\dagger = (\mathbf{A}^\dagger)^H$	
(iv)	$(\alpha\mathbf{A})^\dagger = \alpha^{-1}\mathbf{A}^\dagger, \forall\alpha \in \mathbb{K}, \alpha \neq 0$	
(v)	$(\mathbf{A}\mathbf{A}^\dagger)^\dagger = \mathbf{A}\mathbf{A}^\dagger, (\mathbf{A}^\dagger\mathbf{A})^\dagger = \mathbf{A}^\dagger\mathbf{A}$	
(vi)	$(\mathbf{A}\mathbf{A}^H)^\dagger = (\mathbf{A}^\dagger)^H\mathbf{A}^\dagger, (\mathbf{A}^H\mathbf{A})^\dagger = \mathbf{A}^\dagger\left(\mathbf{A}^\dagger\right)^H$	
(vii)	$\mathbf{A}^H\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^H = \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^H$	
(viii)	$\mathbf{A}^H(\mathbf{A}^\dagger)^H\mathbf{A}^\dagger = \mathbf{A}^\dagger = \mathbf{A}^\dagger(\mathbf{A}^\dagger)^H\mathbf{A}^H$	
(ix)	$\mathbf{A}(\mathbf{A}^H\mathbf{A})^\dagger\mathbf{A}^H\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^\dagger\mathbf{A}$	
(x)	$\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^\dagger\mathbf{A}^H = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^\dagger$	
(xi)	$\mathbf{A}^\dagger = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$	$\mathbf{A}$ of full column rank
(xii)	$\mathbf{A}^\dagger = \mathbf{A}^H\left(\mathbf{A}\mathbf{A}^H\right)^{-1}$	$\mathbf{A}$ of full row rank
(xiii)	$(\mathbf{ACB})^\dagger = \mathbf{B}^\dagger\mathbf{C}^{-1}\mathbf{A}^\dagger$	$\mathbf{A} \in \mathbb{K}^{I\times K}, \mathbf{B} \in \mathbb{K}^{K\times J}$, $\mathbf{C} \in \mathbb{K}^{K\times K}$ of full rank
(xiv)	$(\mathbf{ACB})^\dagger = \mathbf{B}^H\mathbf{C}^\dagger\mathbf{A}^H$	$\mathbf{A}$ and $\mathbf{B}$ unitary, $\mathbf{C} \in \mathbb{K}^{I\times J}$
(xv)	$\mathrm{r}(\mathbf{A}^\dagger) = \mathrm{r}(\mathbf{A})$	
(xvi)	$(\mathbf{0}_{I\times J})^\dagger = \mathbf{0}_{J\times I}$	

Table: Key properties of the Moore-Penrose pseudoinverse of a matrix $\mathbf{A} \in \mathbb{K}^{I\times J}$

Multiplicative groups of matrices

Groups	Sets of matrices	Structural constraints
$GL_n \subseteq \mathbb{K}^{n\times n}$	Invertible	$\det(\mathbf{A}) \neq 0$
$SL_n \subseteq \mathbb{K}^{n\times n}$	Invertible with determinant 1	$\det(\mathbf{A}) = 1$
$O_n \subseteq \mathbb{R}^{n\times n}$	Orthogonal	$\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n$
$U_n \subseteq \mathbb{C}^{n\times n}$	Unitary	$\mathbf{A}^H\mathbf{A} = \mathbf{A}\mathbf{A}^H = \mathbf{I}_n$
$SO_n \subseteq \mathbb{R}^{n\times n}$	Orthogonal with determinant 1	$\mathbf{A}^T\mathbf{A} = \mathbf{I}_n$ and $\det(\mathbf{A}) = 1$
$SU_n \subseteq \mathbb{C}^{n\times n}$	Unitary of determinant 1	$\mathbf{A}^H\mathbf{A} = \mathbf{I}_n$ and $\det(\mathbf{A}) = 1$
$SP_n \subseteq \mathbb{R}^{2n\times 2n}$	Symplectic	$\mathbf{A}^T\widetilde{\mathbf{J}}\mathbf{A} = \widetilde{\mathbf{J}}$

Table: Multiplicative groups of matrices

Matrix representation of a linear map

Any linear map $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ from a $\mathbb{K}$-v.s. $\mathcal{U}$, of dimension J with the ordered basis $\{\mathbf{u}\} = \{\mathbf{u}_1, \cdots, \mathbf{u}_J\}$, to a $\mathbb{K}$-v.s. $\mathcal{V}$, of dimension I with the basis $\{\mathbf{v}\} = \{\mathbf{v}_1, \cdots, \mathbf{v}_I\}$, is defined using the images $\mathcal{L}(\mathbf{u}_j), j \in \langle J \rangle$, expressed in the basis $\{\mathbf{v}\}$, namely:

$$\mathcal{L}(\mathbf{u}_j) = \sum_{i=1}^{I} a_{ij}\mathbf{v}_i, \quad j \in \langle J \rangle.$$

The coefficients a_{ij} are the coordinates of $\mathcal{L}(\mathbf{u}_j)$ in the basis $\{\mathbf{v}\}$. The matrix $\mathbf{A} = [a_{ij}]$, of dimensions $I \times J$, is called the matrix associated with the linear map $\mathcal{L}$, relative to the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$. This dependency of $\mathbf{A}$ with respect to bases will be highlighted by denoting $\mathbf{A}_{\mathrm{vu}}$ if necessary.

Matrix representation of a linear map

By defining the matrix whose columns are the images $\mathcal{L}\left(\mathbf{u}_j\right)$, the following equation is obtained:

$$
\begin{aligned}
[\mathcal{L}(\mathbf{u}_1) \cdots \mathcal{L}(\mathbf{u}_j) \cdots \mathcal{L}(\mathbf{u}_J)] &= [\mathbf{v}_1 \cdots \mathbf{v}_I] \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1,J} \\ \vdots & & \vdots & & \vdots \\ a_{I1} & \cdots & a_{Ij} & \cdots & a_{IJ} \end{bmatrix} \\
&= [\mathbf{v}_1 \cdots \mathbf{v}_I]\, \mathbf{A}.
\end{aligned}
$$

Matrix representation of a linear map

定理

The matrix $\mathbf{A}$ *completely specifies the map* $\mathcal{L}$ *with respect to the bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$*. In other words,* $\forall \mathbf{x} \in \mathcal{U}$ *and its image* $\mathbf{y} = \mathcal{L}(\mathbf{x}) \in \mathcal{V}$*, vectors* $\mathbf{x}_\mathbf{u}$ *and* $\mathbf{y}_\mathbf{v}$ *of the coordinates of* $\mathbf{x}$ *and* $\mathbf{y}$ *in the respective bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$ *are related by the following relationship:*

$$\mathbf{y}_\mathbf{v} = \mathbf{A}\mathbf{x}_\mathbf{u} = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_\mathbf{u}.$$

Matrix representation of a linear map

定理

The matrix $\mathbf{A}$ *completely specifies the map* $\mathcal{L}$ *with respect to the bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$*. In other words,* $\forall \mathbf{x} \in \mathcal{U}$ *and its image* $\mathbf{y} = \mathcal{L}(\mathbf{x}) \in \mathcal{V}$*, vectors* $\mathbf{x}_\mathbf{u}$ *and* $\mathbf{y}_\mathbf{v}$ *of the coordinates of* $\mathbf{x}$ *and* $\mathbf{y}$ *in the respective bases* $\{\mathbf{u}\}$ *and* $\{\mathbf{v}\}$ *are related by the following relationship:*

$$\mathbf{y}_\mathbf{v} = \mathbf{A}\mathbf{x}_\mathbf{u} = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_\mathbf{u}.$$

证明.

Let $\mathbf{x} = \sum_{j=1}^{J} x_j \mathbf{u}_j$ and $\mathbf{y} = \sum_{i=1}^{I} y_i \mathbf{v}_i$. Thus

$$\mathbf{y} = \mathcal{L}(\mathbf{x}) = \sum_{j=1}^{J} x_j \mathcal{L}\left(\mathbf{u}_j\right) = \sum_{j=1}^{J} x_j \left(\sum_{i=1}^{I} a_{ij}\mathbf{v}_i\right) = \sum_{i=1}^{I} \left(\sum_{j=1}^{J} a_{ij} x_j\right) \mathbf{v}_i$$

from which it can be deduced that: $y_i = \sum_{j=1}^{J} a_{ij} x_j$. By defining the vectors of coordinates: $\mathbf{x}_\mathbf{u} = [x_1 \cdots x_J]^T$ and $\mathbf{y}_\mathbf{v} = [y_1 \cdots y_I]^T$, we get the relation: $\mathbf{y}_\mathbf{v} = \mathbf{A}\mathbf{x}_\mathbf{u} = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_\mathbf{u}$. □

Matrix representation of a linear map

Let us define the mappings $\mathcal{E}_{\mathbf{u}} : \mathcal{U} \rightarrow \mathbb{K}^J$ and $\mathcal{F}_{\mathbf{v}} : \mathcal{V} \rightarrow \mathbb{K}^I$ that transform an element $\mathbf{x}$ of $\mathcal{U}$ and an element $\mathbf{y}$ of $\mathcal{V}$ into their vectors of coordinates: $\mathcal{U} \ni \mathbf{x} \mapsto \mathbf{x}_{\mathbf{u}} \in \mathbb{K}^J$ and $\mathcal{V} \ni \mathbf{y} \mapsto \mathbf{y}_{\mathbf{v}} \in \mathbb{K}^I$, respectively. The linear map $\mathcal{L}$ can also be viewed as the composition $\mathcal{F}_{\mathbf{v}}^{-1} \circ L_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}$, where $\mathcal{L}_{\mathbf{vu}} : \mathbb{K}^J \rightarrow \mathbb{K}^I$ is a linear map such that $\mathbf{x}_{\mathbf{u}} \mapsto \mathbf{y}_{\mathbf{v}} = \mathcal{L}_{\mathbf{vu}}\left(\mathbf{x}_{\mathbf{u}}\right) = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_{\mathbf{u}}$

Indeed, for all $\mathbf{x} \in \mathcal{U}$, we have $\mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}(\mathbf{x}) = \mathcal{F}_{\mathbf{v}}^{-1} \circ L_{\mathbf{vu}}\left(\mathbf{x}_{\mathbf{u}}\right) = \mathcal{F}_{\mathbf{v}}^{-1}\left(\mathbf{y}_{\mathbf{v}}\right) = \mathbf{y} = \mathcal{L}(\mathbf{x})$, and therefore:

$$\mathcal{L} = \mathcal{F}_{\mathbf{v}}^{-1} \circ \mathcal{L}_{\mathbf{vu}} \circ \mathcal{E}_{\mathbf{u}}.$$

Matrix representation of a linear map

Let $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ be the v.s. of the linear maps from $\mathcal{U}$ to $\mathcal{V}$. The following properties can be demonstrated:

- If $\mathcal{U} = \mathbb{K}^J$ and $\mathcal{V} = \mathbb{K}^I$, and $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ are the canonical bases, i.e., $\{\mathbf{u}_j = \mathbf{e}_j^{(J)}\}$ and $\{\mathbf{v}_i = \mathbf{e}_i^{(I)}\}$, we then have $\mathbf{x} = \mathbf{x}_{\mathbf{u}}$ and $\mathbf{y} = \mathbf{y}_{\mathbf{v}}$, and thus $\mathbf{y} = \mathbf{A}\mathbf{x}$.
- Given the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$ of $\mathcal{U}$ and $\mathcal{V}$, the mapping:

$$\varphi : \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V}) \to \mathbb{K}^{I \times J}$$
$$\mathcal{L} \mapsto \varphi(\mathcal{L}) = \mathbf{A}_{\mathbf{vu}},$$

  which makes the correspondence between an element $\mathcal{L} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ and its associated matrix $\mathbf{A}_{\mathbf{vu}}$, is bijective. In addition, $\forall \mathcal{L}, \mathcal{K} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$ such that $\varphi(\mathcal{L}) = \mathbf{A}_{\mathbf{vu}}$ and $\varphi(\mathcal{K}) = \mathbf{B}_{\mathbf{vu}}$, and $\forall \alpha \in \mathbb{K}$, we have:

$$\varphi(\alpha\mathcal{L}) = \alpha\mathbf{A}_{\mathbf{vu}} = \alpha\varphi(\mathcal{L}), \varphi(L + \mathcal{K}) = \mathbf{A}_{\mathbf{vu}} + \mathbf{B}_{\mathbf{vu}} = \varphi(\mathcal{L}) + \varphi(\mathcal{K}),$$

  that is, the mapping φ preserves the operations of addition and multiplication by a scalar in the vector space $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$. Therefore, φ is an isomorphism between the vector spaces $\mathbb{L}_{K}(\mathcal{U}, \mathcal{V})$ and $\mathbb{K}^{I \times J}$.

Matrix representation of a linear map

- $\mathcal{U}$ and $\mathcal{V}$ being two v.s. of the same dimension, with bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_I\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_I\}$, the linear map $\mathcal{L} \in \mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{V})$, with the associated matrix $\mathbf{A}_{\mathbf{vu}}$, is bijective if and only if $\mathbf{A}_{\mathbf{vu}}$ is invertible, and we have $\varphi\left(\mathcal{L}^{-1}\right) = [\varphi(\mathcal{L})]^{-1} = \mathbf{A}_{\mathbf{vu}}^{-1}$.

定理

Let $\mathcal{U}, \mathcal{V}$, and $\mathcal{W}$ be three $\mathbb{K}$-v.s., of respective dimensions J, I, and K. Consider two linear maps $\mathcal{L} : \mathcal{U} \to \mathcal{V}$ and $\mathcal{K} : \mathcal{V} \to \mathcal{W}$ with the associated matrices $\mathbf{A}_{\mathbf{vu}} \in \mathbb{K}^{I \times J}$ and $\mathbf{B}_{\mathbf{wv}} \in \mathbb{K}^{K \times I}$, then the composite map $\mathcal{K} \circ \mathcal{L}$ is itself a linear map of $\mathbb{L}_{\mathbb{K}}(\mathcal{U}, \mathcal{W})$ with the associated matrix $\mathbf{C}_{\mathbf{wu}} = \mathbf{B}_{\mathbf{wv}} \mathbf{A}_{\mathbf{vu}} \in \mathbb{K}^{K \times J}$.

Matrix representation of a linear map

证明.

Consider the bases $\{\mathbf{u}_1, \ldots, \mathbf{u}_J\}$, $\{\mathbf{v}_1, \ldots, \mathbf{v}_I\}$, and $\{\mathbf{w}_1, \ldots, \mathbf{w}_K\}$ of $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$, respectively, and define $\varphi(\mathcal{K}) = \mathbf{B}_{\mathbf{wv}} = [b_{ki}]$ and $\varphi(\mathcal{L}) = \mathbf{A}_{\mathbf{vu}} = [a_{ij}]$ associated with the linear maps $\mathcal{K}$ and $\mathcal{L}$. Then,

$$\begin{aligned}
\mathcal{K} \circ \mathcal{L}(\mathbf{u}_j) &= \mathcal{K}\left(\sum_{i=1}^{I} a_{ij}\mathbf{v}_i\right) = \sum_{i=1}^{I} a_{ij}\mathcal{K}(\mathbf{v}_i) \\
&= \sum_{i=1}^{I} a_{ij}\left(\sum_{k=1}^{K} b_{ki}\mathbf{w}_k\right) = \sum_{k=1}^{K}\left(\sum_{i=1}^{I} b_{ki}a_{ij}\right)\mathbf{w}_k.
\end{aligned}$$

Accordingly, the matrix associated with $\mathcal{K} \circ \mathcal{L}$ is such that $c_{kj} = \sum_{i=1}^{I} b_{ki}a_{ij}$, from which it can be concluded that $\mathbf{C}_{\mathbf{wu}} = \mathbf{B}_{\mathbf{wv}}\mathbf{A}_{\mathbf{vu}}$. □

Change of basis

Let $\{\mathbf{u}\}$ and $\{\underline{\mathbf{u}}\}$ be two bases of the vector space $\mathcal{U}$, linked by the following formula:

$$\underline{\mathbf{u}}_j = \sum_{i=1}^{J} p_{ij}\mathbf{u}_i, \quad j \in \langle J \rangle.$$

Define the matrix $\mathbf{P} = [p_{ij}]$, of dimensions $J \times J$, called the change of basis matrix from $\{\mathbf{u}\}$ to $\{\underline{\mathbf{u}}\}$. The components of its j th column are the coordinates of $\underline{\mathbf{u}}_j$ in the basis $\{\mathbf{u}\}$, and the equation for the change of basis is written as:

$$\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1 \cdots \underline{\mathbf{u}}_J] = [\mathbf{u}_1 \cdots \mathbf{u}_J]\,\mathbf{P} = \mathbf{U}\mathbf{P}.$$

$\mathbf{P}$ is invertible, and its inverse $\mathbf{P}^{-1}$ is the transformation matrix from $\{\underline{\mathbf{u}}\}$ to $\{\mathbf{u}\}$ such that:

$$\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_J] = [\underline{\mathbf{u}}_1 \cdots \underline{\mathbf{u}}_J]\,\mathbf{P}^{-1} = \underline{\mathbf{U}}\mathbf{P}^{-1}.$$

Its columns contain the coordinates of the vectors of the basis $\{\mathbf{u}\}$ in the basis $\{\underline{\mathbf{u}}\}$.

Change of basis

定理

Given a vector $\mathbf{x} \in \mathcal{U}$, *the coordinate vectors* $\mathbf{x_u}$ *and* $\mathbf{x_{\underline{u}}}$ *in the bases* $\{\mathbf{u}\}$ *and* $\{\underline{\mathbf{u}}\}$ *are related by the following relationship:*

$$\mathbf{x_u} = \mathbf{P}\mathbf{x_{\underline{u}}}.$$

定理

Consider the linear map $L : \mathcal{U} \to \mathcal{V}$ *and the changes of basis in* $\mathcal{U}$ *and* $\mathcal{V}$ *characterized by matrices* $\mathbf{P}$ *and* $\mathbf{Q}$, *respectively, and designate by* $\mathbf{A_{vu}}$ *and* $\mathbf{A_{\underline{vu}}}$ *the matrices associated with* L *in the bases* $(\{\mathbf{u}\}, \{\mathbf{v}\})$ *on the one hand, and* $(\{\underline{\mathbf{u}}\}, \{\underline{\mathbf{v}}\})$ *on the other. We have the following relationship:*

$$\mathbf{A_{\underline{vu}}} = \mathbf{Q}^{-1}\mathbf{A_{vu}}\mathbf{P}, \tag{3}$$

and it is said that the matrices $\mathbf{A_{vu}}$ *and* $\mathbf{A_{\underline{vu}}}$ *are equivalent.*

Change of basis

$$
\begin{tikzcd}[column sep=huge, row sep=large]
\mathbf{x} = \sum_{j=1}^{J} x_j \mathbf{u}_j \in \mathcal{U} \arrow[r, "\mathcal{L}"] \arrow[d, "\mathcal{E}_{\mathbf{u}}"'] & \mathbf{y} = \sum_{i=1}^{I} y_i \mathbf{v}_i \in \mathcal{V} \arrow[d, "\mathcal{F}_{\mathbf{v}}"] \\
\mathbf{x}_{\mathbf{u}} \arrow[r, "\mathbf{y}_{\mathbf{v}} = \mathbf{A}_{\mathbf{vu}} \mathbf{x}_{\mathbf{u}} = \mathcal{L}_{\mathbf{vu}}(\mathbf{x}_{\mathbf{u}})"] \arrow[d, "\mathbf{P}^{-1}"'] & \mathbf{y}_{\mathbf{v}} \arrow[d, "\mathbf{Q}^{-1}"] \\
\mathbf{x}_{\underline{\mathbf{u}}} \arrow[r, "\mathbf{y}_{\underline{\mathbf{v}}} = \mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} = \mathbf{Q}^{-1} \mathbf{A}_{\mathbf{vu}} \mathbf{P} \mathbf{x}_{\underline{\mathbf{u}}}"] & \mathbf{y}_{\underline{\mathbf{v}}}
\end{tikzcd}
$$

Figure: Effect of changes of bases with a linear map $\mathcal{L}$

Endomorphisms

An endomorphism $f \in \mathcal{L}(\mathcal{U})$ is a linear map from $\mathcal{U}$ into itself. By choosing the basis $\{v\}$ identical to $\{\mathbf{u}\}$, the matrix of an endomorphism in the basis $\{\mathbf{u}\}$ is a square matrix, denoted by $\mathbf{A}_{\mathbf{u}}$. Relation (3) then becomes:

$$\mathbf{A}_{\underline{\mathbf{u}}} = \mathbf{P}^{-1}\mathbf{A}_{\mathbf{u}}\mathbf{P},$$

where $\mathbf{P}$ is the change of basis matrix in U.
An important problem in matrix calculus consists in finding the transformation matrix $\mathbf{P}$ allowing to reduce the matrix of a linear map to its simplest form possible, as for example the diagonal form. When there exists $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}_{\mathbf{u}}\mathbf{P}$ is diagonal, it is said that $\mathbf{A}_{\mathbf{u}}$ is diagonalizable. The matrix $\mathbf{P}$ is then obtained from the eigendecomposition of $\mathbf{A}_{\mathrm{u}}$, its columns being formed of eigenvectors, while the diagonal elements of the diagonal matrix $\mathbf{A}_{\underline{\mathbf{u}}}$ are the eigenvalues.

Nilpotent endomorphisms

It is said that $f \in \mathcal{L}(\mathcal{U})$ is a nilpotent endomorphism of index n if there exists a smallest integer $n > 0$ such that $f^n = 0$, which amounts to saying that the composition of f by itself n times gives the zero morphism.

定理

The endomorphism $f \in \mathcal{L}(\mathcal{U})$ is nilpotent of index n if and only if its associated matrix $\mathbf{A_u}$ is strictly upper triangular.

This results implies that the matrix associated with f^n is equal to $(\mathbf{A_u})^n$. As a consequence, the matrix associated with a nilpotent endomorphism of index n is a nilpotent matrix of index n defined such that $(\mathbf{A_u})^n = \mathbf{0}$.

Equivalent, similar and congruent matrices

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times J}$ are said to be equivalent when there exists two non-singular matrices $\mathbf{P} \in \mathbb{K}^{J \times J}$ and $\mathbf{Q} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{QAP}. \tag{4}$$

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times I}$ are called similar if there exists a non-singular matrix $\mathbf{P} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP} \text{ or } \mathbf{PB} = \mathbf{AP}.$$

Equivalent, similar and congruent matrices

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times J}$ are said to be equivalent when there exists two non-singular matrices $\mathbf{P} \in \mathbb{K}^{J \times J}$ and $\mathbf{Q} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{QAP}. \tag{4}$$

Two matrices $\mathbf{A}$ and $\mathbf{B} \in \mathbb{K}^{I \times I}$ are called similar if there exists a non-singular matrix $\mathbf{P} \in \mathbb{K}^{I \times I}$ such that:

$$\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP} \text{ or } \mathbf{PB} = \mathbf{AP}.$$

It is easy to deduce the following properties:

- If $\mathbf{A}$ and $\mathbf{B}$ are equivalent, then $r(\mathbf{A}) = r(\mathbf{B})$.
- If $\mathbf{A}$ and $\mathbf{B}$ are similar, then $r(\mathbf{A}) = r(\mathbf{B})$, and $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{B})$, because $\mathrm{tr}(\mathbf{B}) = \mathrm{tr}\left(\mathbf{P}^{-1}\mathbf{AP}\right) = \mathrm{tr}\left(\mathbf{PP}^{-1}\mathbf{A}\right) = \mathrm{tr}(\mathbf{A})$.

Equivalent, similar and congruent matrices

$\mathbf{A}$ and $\mathbf{B}$ are said to be orthogonally similar if there exists an orthogonal matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{A}\mathbf{P}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^T.$$

Similarly, matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be unitarily similar if there exists a unitary matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^H\mathbf{A}\mathbf{P}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^H.$$

$\mathbf{A}$ and $\mathbf{B}$ are said to be congruent if there exists a non-singular matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{A}\mathbf{P}.$$

Equivalent, similar and congruent matrices

$\mathbf{A}$ and $\mathbf{B}$ are said to be orthogonally similar if there exists an orthogonal matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{A}\mathbf{P}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^T.$$

Similarly, matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be unitarily similar if there exists a unitary matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^H\mathbf{A}\mathbf{P}, \quad \text{with} \quad \mathbf{P}^{-1} = \mathbf{P}^H.$$

$\mathbf{A}$ and $\mathbf{B}$ are said to be congruent if there exists a non-singular matrix $\mathbf{P}$ such that:

$$\mathbf{B} = \mathbf{P}^T\mathbf{A}\mathbf{P}.$$

Equivalent, similar, and congruent matrices are such that:

- Two matrices are equivalent (similar) if they can represent the same linear map (the same endomorphism) in different bases.
- Two matrices are congruent if they can represent the same symmetric bilinear form in two different bases.

Definition of a bilinear/sesquilinear map

Let $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ be three $\mathbb{R}$-v.s., of respective dimensions J, I, and K. A bilinear map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$, such that $\mathcal{U} \times \mathcal{V} \ni (x, y) \mapsto f(x, y) \in \mathcal{W}$, is linear with respect to each of the variables x and y, when the other variable (y and x, respectively) is fixed.

Similarly, when $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ are three $\mathbb{C}$-v.s., the map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$ is said to be sesquilinear if it is linear with respect to the first variable $x \in \mathcal{U}$, and semilinear with respect to the second variable $y \in \mathcal{V}$.

Definition of a bilinear/sesquilinear map

Let $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ be three $\mathbb{R}$-v.s., of respective dimensions J, I, and K. A bilinear map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$, such that $\mathcal{U} \times \mathcal{V} \ni (x, y) \mapsto f(x, y) \in \mathcal{W}$, is linear with respect to each of the variables x and y, when the other variable (y and x, respectively) is fixed.

Similarly, when $\mathcal{U}, \mathcal{V}$ and $\mathcal{W}$ are three $\mathbb{C}$-v.s., the map $f : \mathcal{U} \times \mathcal{V} \to \mathcal{W}$ is said to be sesquilinear if it is linear with respect to the first variable $x \in \mathcal{U}$, and semilinear with respect to the second variable $y \in \mathcal{V}$.

- linearity with respect to the first argument:

$$f(\alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2, \mathbf{b}) = \alpha_1 f(\mathbf{a}_1, \mathbf{b}) + \alpha_2 f(\mathbf{a}_2, \mathbf{b}) \quad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C}.$$

- semilinear with respect to the second argument:

$$f(\mathbf{a}, \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2) = \alpha_1^* f(\mathbf{a}, \mathbf{b}_1) + \alpha_2^* f(\mathbf{a}, \mathbf{b}_2) \quad \text{for all } \alpha_1, \alpha_2 \in \mathbb{C}.$$

Definition of a bilinear/sesquilinear map

We thus have the following properties:

- f is additive with respect to the two variables, that is, for all $x, x_1, x_2 \in \mathcal{U}$ and $y, y_1, y_2 \in \mathcal{V}$, we have:

$$f(x_1 + x_2 + y) = f(x_1, y) + f(x_2, y)\,,$$
$$f(x, y_1 + y_2) = f(x, y_1) + f(x, y_2)\,.$$

- f is homogeneous in the first variable

$$f(\alpha x, y) = \alpha f(x, y), \quad \alpha \in \mathbb{K}.$$

- In the case of a bilinear map $(\mathbb{K} = \mathbb{R})$, f is homogeneous in the second variable

$$f(x, \alpha y) = \alpha f(x, y), \quad \alpha \in \mathbb{R}.$$

- In the case of a sesquilinear map $(\mathbb{K} = \mathbb{C})$, f is conjugate-homogeneous in the second variable

$$f(x, \alpha y) = \alpha^* f(x, y), \quad \alpha \in \mathbb{C}.$$

Definition of a bilinear/sesquilinear map

定理

The set of bilinear maps, denoted by $\mathcal{BL}(\mathcal{U}, \mathcal{V}; W)$, is a $\mathbb{K}$-v.s. of dimension: $\dim[\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathcal{W})] = \dim[\mathcal{U}]\dim[\mathcal{V}]\dim[\mathcal{W}]$.

Definition of a bilinear/sesquilinear map

定理

The set of bilinear maps, denoted by $\mathcal{BL}(\mathcal{U}, \mathcal{V}; W)$*, is a* $\mathbb{K}$*-v.s. of dimension:* $\dim[\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathcal{W})] = \dim[\mathcal{U}]\dim[\mathcal{V}]\dim[\mathcal{W}]$.

In the following, we consider the case $\mathcal{W} = \mathbb{K}$, corresponding to a bilinear form if $\mathbb{K} = \mathbb{R}$, and a sesquilinear form if $\mathbb{K} = \mathbb{C}$. Vectors x and y of v.s. $\mathcal{U}$ and $\mathcal{V}$ will be denoted in bold to distinguish them from their scalar components in a given basis. The notation (x, y) will be used instead of $(\mathbf{x}, \mathbf{y})$ when there is no ambiguity.

定理

The set of bilinear forms, denoted by $\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathbb{K})$*, is a* $\mathbb{K}$*-v.s. of dimension:* $\dim[\mathcal{BL}(\mathcal{U}, \mathcal{V}; \mathbb{K})] = \dim[\mathcal{U}]\dim[\mathcal{V}]$.

Definition of a bilinear/sesquilinear map

例

Given a matrix $\mathbf{B} \in \mathbb{R}^{I \times I}$, the following expression:

$$\mathbf{y}^T \mathbf{B} \mathbf{x} = \sum_{i,j=1}^{I} b_{ij} y_i x_j \tag{5}$$

is a bilinear form on the vector space $\mathbb{R}^I$, that is, $f : \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}$, such that $\mathbb{R}^I \times \mathbb{R}^I \ni (\mathbf{x}, \mathbf{y}) \mapsto f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T \mathbf{B} \mathbf{x} \in \mathbb{R}$.

In the next section, we shall see that all bilinear forms can be represented by means of an equation like Equ. (5).

Matrix associated to a bilinear/sesquilinear form

Consider the $\mathbb{R}$-v.s. $\mathcal{U}$ and $\mathcal{V}$, of dimensions J and I, with the respective ordered bases $\{\mathbf{u}\} = \{\mathbf{u}_1, \cdots, \mathbf{u}_J\}$ and $\{\mathbf{v}\} = \{\mathbf{v}_1, \cdots, \mathbf{v}_I\}$. The vectors $\mathbf{x} \in \mathcal{U}$ and $\mathbf{y} \in \mathcal{V}$ can then be written as:

$$\mathbf{x} = \sum_{j=1}^{J} x_j \mathbf{u}_j, \quad \mathbf{y} = \sum_{i=1}^{I} y_i \mathbf{v}_i.$$

Using the bilinearity property of the bilinear form $f : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$ gives:

$$f(\mathbf{x}, \mathbf{y}) = f\left(\sum_{j=1}^{J} x_j \mathbf{u}_j, \sum_{i=1}^{I} y_i \mathbf{v}_i\right) = \sum_{i=1}^{I} \sum_{j=1}^{J} x_j y_i f(\mathbf{u}_j, \mathbf{v}_i) . \tag{6}$$

Matrix associated to a bilinear/sesquilinear form

By defining the matrix $\mathbf{B}_{\mathbf{vu}} \in \mathbb{R}^{I \times J}$ such that $(\mathbf{B}_{\mathbf{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i) = [b_{ij}]$, $f(\mathbf{x}, \mathbf{y})$ can be rewritten as:

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{I} \sum_{j=1}^{J} b_{ij} x_j y_i = \mathbf{y}_{\mathbf{v}}^T \mathbf{B}_{\mathbf{vu}} \mathbf{x}_{\mathbf{u}} = \mathbf{x}_{\mathbf{u}}^T \mathbf{B}_{\mathbf{vu}}^T \mathbf{y}_{\mathbf{v}}, \tag{7}$$

where $\mathbf{x}_{\mathbf{u}} = [x_1, \cdots, x_J]^T$ and $\mathbf{y}_{\mathbf{v}} = [y_1, \cdots, y_I]^T$ are the vectors of coordinates of $\mathbf{x}$ and $\mathbf{y}$ in the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$, respectively. The matrix $\mathbf{B}_{\mathbf{vu}}$ is called the matrix associated with the bilinear form f, relative to the bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$.

Changes of bases with a bilinear form

Let the changes of bases be defined as:

$$\underline{\mathbf{U}} = [\underline{\mathbf{u}}_1, \cdots, \underline{\mathbf{u}}_J] = [\mathbf{u}_1, \cdots, \mathbf{u}_J]\,\mathbf{P} = \mathbf{U}\mathbf{P}, \tag{8}$$

$$\underline{\mathbf{V}} = [\underline{\mathbf{v}}_1, \cdots, \underline{\mathbf{v}}_I] = [\mathbf{v}_1, \cdots, \mathbf{v}_I]\,\mathbf{Q} = \mathbf{V}\mathbf{Q}. \tag{9}$$

We have:

$$\mathbf{x} = \mathbf{U}\mathbf{x}_{\mathbf{u}} = \underline{\mathbf{U}}\mathbf{x}_{\mathbf{u}} = \mathbf{U}\mathbf{P}\underline{\mathbf{x}}_{\mathbf{u}},$$
$$\mathbf{y} = \mathbf{V}\mathbf{y}_{\mathbf{v}} = \underline{\mathbf{V}}\mathbf{y}_{\underline{\mathbf{v}}} = \mathbf{V}\mathbf{Q}\mathbf{y}_{\underline{\mathbf{v}}},$$

from which are inferred the relations: $\mathbf{x}_{\mathbf{u}} = \mathbf{P}\mathbf{x}_{\underline{\mathbf{u}}}, \mathbf{y}_{\mathbf{v}} = \mathbf{Q}\mathbf{y}_{\underline{\mathbf{v}}}$, where $\mathbf{x}_{\underline{\mathbf{u}}}$ and $\mathbf{y}_{\underline{\mathbf{v}}}$ are the vectors of coordinates of $\mathbf{x}$ and $\mathbf{y}$ in the bases $\underline{\mathbf{u}}$ and $\underline{\mathbf{v}}$, and $\mathbf{P}$ and $\mathbf{Q}$ are the change of basis matrices from $\mathbf{u}$ to $\underline{\mathbf{u}}$ and from $\mathbf{v}$ to $\underline{\mathbf{v}}$, respectively.

Changes of bases with a bilinear form

After the changes of bases, expression (7) of the bilinear form becomes:

$$
\begin{aligned}
f(\mathbf{x}, \mathbf{y}) &= \mathbf{y}_{\underline{\mathbf{v}}}^{T} \mathbf{Q}^{T} \mathbf{B}_{\mathbf{vu}} \mathbf{P} \mathbf{x}_{\underline{\mathbf{u}}} \\
&= \mathbf{y}_{\underline{\mathbf{y}}}^{T} \mathbf{B}_{\underline{\mathbf{v}}\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}}
\end{aligned}
\tag{10}
$$

where

$$
\mathbf{B}_{\underline{\mathbf{vu}}} = \mathbf{Q}^{T} \mathbf{B}_{\mathbf{vu}} \mathbf{P} \tag{11}
$$

is the matrix associated to f relatively to the bases $\{\underline{\mathbf{u}}\}$ and $\{\underline{\mathbf{v}}\}$. Matrices $\mathbf{B}_{\mathbf{vu}}$ and $\mathbf{B}_{\underline{v}}\underline{u}$, which represent the same bilinear form f in two different sets of bases, are equivalent according to the definition (4), and therefore have the same rank. This rank is called the rank of f.

Changes of bases with a sesquilinear form

In the case where $\mathcal{U}$ and $\mathcal{V}$ are two $\mathbb{C}$-v.s., with the respective bases $\{\mathbf{u}\}$ and $\{\mathbf{v}\}$, equations (6) and (7) for a sesquilinear form become:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \sum_{i=1}^{I} \sum_{j=1}^{J} x_j y_i^* f(\mathbf{u}_j, \mathbf{v}_i) \\ &= \mathbf{y}_{\mathbf{v}}^H \mathbf{B}_{\mathbf{vu}} \mathbf{x}_{\mathbf{u}}, \end{aligned}$$

and after the changes of bases (8) and (9), the relation (11) becomes:

$$\mathbf{B}_{\underline{\mathrm{vu}}} = \mathbf{Q}^H \mathbf{B}_{\mathrm{vu}} \mathbf{P}.$$

Matrices associated with linear, bilinear, and sesquilinear forms

The main results established for matrices associated with a linear map, a bilinear form, and a sesquilinear form are given in the following table.

Linear map	Bilinear form	Sesquilinear form
$\mathcal{L} : \mathcal{U} \to \mathcal{V}$	$f : \mathcal{U} \times \mathcal{V} \to \mathbb{R}$	$f : \mathcal{U} \times \mathcal{V} \to \mathbb{C}$
Matrix associated to $\mathcal{L}$	**Matrix associated to f**	**Matrix associated to f**
$\mathbf{A}_{.j} = \mathcal{L}(\mathbf{u}_j)$ $\mathbf{y}_{\mathbf{v}} = \sum_{j=1}^{J} x_j \mathcal{L}(\mathbf{u}_j) = \mathbf{A}_{\mathbf{vu}}\mathbf{x}_{\mathbf{u}}$	$(\mathbf{B}_{\mathbf{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i)$ $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_{\mathbf{v}}^T \mathbf{B}_{\mathbf{vu}}\mathbf{x}_{\mathbf{u}}$	$(\mathbf{B}_{\mathbf{vu}})_{ij} = f(\mathbf{u}_j, \mathbf{v}_i)$ $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_{\mathbf{v}}^H \mathbf{B}_{\mathbf{vu}}\mathbf{x}_{\mathbf{u}}$
	Changes of bases	
	$\begin{cases} \underline{\mathbf{U}} = \mathbf{UP} \\ \underline{\mathbf{V}} = \mathbf{VQ} \end{cases}$	
$\mathbf{A}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{-1}\mathbf{A}_{\mathbf{vu}}\mathbf{P}$	$\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{T}\mathbf{B}_{\mathbf{vu}}\mathbf{P}$	$\mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^{H}\mathbf{B}_{\mathbf{vu}}\mathbf{P}$

Table: Matrices associated with linear, bilinear, and sesquilinear forms.

Symmetric bilinear/sesquilinear forms

From now on, we consider that both $\mathbb{R}$-v.s. $\mathcal{U}$ and $\mathcal{V}$ are identical and $\{\mathbf{v}\} = \{\mathbf{u}\}$. The matrix associated with the bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$, denoted by $\mathbf{B}_\mathbf{u}$, is then square, and equation (7) becomes:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}. \tag{12}$$

After the basis change with matrix $\mathbf{P}$, equations (10) and (11) can be written as:

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_{\underline{u}}^T \mathbf{B}_{\underline{u}} \mathbf{x}_{\underline{u}}, \tag{13}$$

$$\mathbf{B}_{\underline{u}} = \mathbf{P}^T \mathbf{B}_\mathbf{u} \mathbf{P}. \tag{14}$$

Matrices $\mathbf{B}_\mathbf{u}$ and $\mathbf{B}_{\underline{u}}$ which represent f in two different bases are said to be congruent.

Symmetric bilinear/sesquilinear forms

For a sesquilinear form $f: \mathcal{U}^2 \to \mathbb{C}$, where $\mathcal{U}$ is a $\mathbb{C}$-v.s., equations (12)–(14) become:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \mathbf{y}_{\mathbf{u}}^H \mathbf{B}_{\mathbf{u}} \mathbf{x}_{\mathbf{u}} \\ &= \mathbf{y}_{\underline{u}}^H \mathbf{B}_{\underline{u}} \mathbf{x}_{\underline{u}} \\ \mathbf{B}_{\underline{u}} &= \mathbf{P}^H \mathbf{B}_{\mathbf{u}} \mathbf{P}. \end{aligned}$$

Matrices $\mathbf{B}_{\mathbf{u}}$ and $\mathbf{B}_{\underline{u}}$ are again congruent.

Symmetric/antisymmetric bilinear forms

It is said that the bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$ is:

- symmetric, if: $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = f(\mathbf{x}, \mathbf{y})$;
- antisymmetric, if: $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = -f(\mathbf{x}, \mathbf{y})$.

If $\mathbf{B_u}$ is the matrix associated with f in the basis $\{\mathbf{u}\}$, it is then easy to verify from (12) that $\mathbf{B_u}$ is such that:

$$\begin{aligned} f \text{ symmetric } &\iff \mathbf{B_u}^T = \mathbf{B_u} \\ f \text{ antisymmetric } &\iff \mathbf{B_u}^T = -\mathbf{B_u} \end{aligned}$$

in other words, $\mathbf{B_u}$ is respectively symmetric and antisymmetric.

Symmetric/antisymmetric bilinear forms

定理

Any alternating bilinear form is antisymmetric.

证明.

Indeed, if f is an alternating bilinear form, according to definition of an alternating multilinear form, the bilinear form f is alternating if: $\forall \mathbf{x} \in \mathcal{U}$, we have $f(\mathbf{x}, \mathbf{x}) = 0$. Thus,

$$
\begin{aligned}
f(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) &= f(\mathbf{x}, \mathbf{x}) + f(\mathbf{y}, \mathbf{y}) + f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{x}) \\
&= f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{x}) = 0 \\
&\downarrow \\
f(\mathbf{y}, \mathbf{x}) &= -f(\mathbf{x}, \mathbf{y})
\end{aligned}
$$

which proves that f is antisymmetric. □

Conjugate symmetric/antisymmetric sesquilinear forms

In the case of a sesquilinear form $f : \mathcal{U}^2 \to \mathbb{C}$, it is said that f has:

- a Hermitian symmetry if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{y})$;
- a skew-Hermitian (also called antihermitian) symmetry if: $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = -f^*(\mathbf{x}, \mathbf{y})$.

Conjugate symmetric/antisymmetric sesquilinear forms

In the case of a sesquilinear form $f : \mathcal{U}^2 \to \mathbb{C}$, it is said that f has:

- a Hermitian symmetry if: $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = f^*(\mathbf{x}, \mathbf{y})$;
- a skew-Hermitian (also called antihermitian) symmetry if: $\forall (\mathbf{x}, \mathbf{y}) \in \mathcal{U}^2$, we have $f(\mathbf{y}, \mathbf{x}) = -f^*(\mathbf{x}, \mathbf{y})$.

定理

Based on equation $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}_{\mathbf{u}}^H \mathbf{B}_{\mathbf{u}} \mathbf{x}_{\mathbf{u}}$*, we can deduce that* $\mathbf{B}_{\mathbf{u}}$ *is such that:*

$$f \text{ of Hermitian symmetry} \iff \mathbf{B}_{\mathbf{u}}^H = \mathbf{B}_{\mathbf{u}}$$

$$f \text{ of skew-Hermitian symmetry} \iff \mathbf{B}_{\mathbf{u}}^H = -\mathbf{B}_{\mathbf{u}}.$$

Conjugate symmetric/antisymmetric sesquilinear forms

定理

Any bilinear form f with the associated matrix $\mathbf{B_u}$ can be decomposed into the sum of a symmetric bilinear form with an antisymmetric bilinear form:

$$f(x,y) = \frac{1}{2}[f(x,y) + f(y,x)] + \frac{1}{2}[f(x,y) - f(y,x)]$$

where the first term in square brackets is a symmetric bilinear form, the second being an antisymmetric bilinear form. From this decomposition, it results that any real square matrix $\mathbf{B_u}$ can be decomposed into the sum of a symmetric matrix and an antisymmetric matrix such as:

$$\mathbf{B_u} = \frac{1}{2}\left[\mathbf{B_u} + \mathbf{B_u}^T\right] + \frac{1}{2}\left[\mathbf{B_u} - \mathbf{B_u}^T\right].$$

Similarly, any complex square matrix can be decomposed into the sum of a Hermitian matrix and an antihermitian matrix.

Quadratic forms

Given a symmetric bilinear form f over a $\mathbb{R}$-v.s. U, the quadratic form associated with f designates the mapping $q : \tau \to \mathbb{R}$ defined by:

$$\begin{aligned} q(\mathbf{x}) &= f(\mathbf{x}, \mathbf{x}) \\ &= x_{\mathbf{u}}^T \mathbf{B}_{\mathbf{u}} \mathbf{x}_{\mathbf{u}}, \quad \text{according to } (12) \end{aligned}$$

the symmetry property of f implying that: $\mathbf{B}_{\mathbf{u}}^T = \mathbf{B}_{\mathbf{u}}$. By defining the coordinate vector $\mathbf{x}_{\mathbf{u}}^T = [x_1, \cdots, x_J]$, with $\mathbf{B}_{\mathbf{u}} = [b_{ij}]\,, i, j \in \langle J \rangle$, the quadratic form can be developed as:

$$\begin{aligned} q(\mathbf{x}) &= \sum_{i=1}^{J} \sum_{j=1}^{J} b_{ij} x_i x_j, \quad \text{with} \quad b_{ji} = b_{ij} \\ &= \sum_{i=1}^{J} b_{ii} x_i^2 + 2 \sum_{1 \le i < j \le J} b_{ij} x_i x_j. \end{aligned} \tag{15}$$

$q(\mathbf{x})$ is thus a homogeneous polynomial of second degree in the coordinates $\{x_j, \, , \, j \in \langle J \rangle\}$ of vector $\mathbf{x}$ in the basis $\{\mathbf{u}\}$.

Quadratic forms

As it can be easily verified, shifting from one quadratic form $q(\mathbf{x})$ to form $f(\mathbf{x}, \mathbf{y})$ can be performed using the following rule:

- the terms x_i^2 are replaced by $x_i y_i$;
- the terms $x_i x_j$ are replaced by $\frac{1}{2}\left[x_i y_j + x_j y_i\right]$,

which gives:

$$f(x, y) = \sum_{i=1}^{J} b_{ii} x_i y_i + \sum_{1 \leq i < j \leq J} b_{ij}\left(x_i y_j + x_j y_i\right).$$

By exploiting the bilinearity and symmetry properties of f, it can be deduced that:

$$\begin{aligned} q(\lambda\mathbf{x}) &= f(\lambda\mathbf{x}, \lambda\mathbf{x}) = \lambda^2 q(\mathbf{x}) \\ q(\mathbf{x}+\mathbf{y}) &= f(\mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}) = q(\mathbf{x}) + 2f(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}) \\ q(\mathbf{x}-\mathbf{y}) &= f(\mathbf{x}-\mathbf{y}, \mathbf{x}-\mathbf{y}) = q(\mathbf{x}) - 2f(\mathbf{x}, \mathbf{y}) + q(\mathbf{y}). \end{aligned}$$

Quadratic forms

定理

The symmetric bilinear form f associated with the quadratic form q can be determined using the following formulae:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= \frac{1}{4}[q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}-\mathbf{y})] \\ &= \frac{1}{2}[q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y})] \\ &= \frac{1}{2}[q(\mathbf{x}) + q(\mathbf{y}) - q(\mathbf{x}-\mathbf{y})]. \end{aligned}$$

Hermitian forms

In the case of a sesquilinear form with Hermitian symmetry, we call Hermitian form associated to f the map $q : \mathcal{U} \to \mathbb{R}$ defined by:

$$\begin{aligned} q(\mathbf{x}) &= f(\mathbf{x}, \mathbf{x}) \\ &= \mathbf{x}_{\mathbf{u}}^{H} \mathbf{B}_{\mathbf{u}} \mathbf{x}_{\mathbf{u}}, \end{aligned}$$

Hermitian symmetry induces that: $\mathbf{B}_{\mathbf{u}}^{H} = \mathbf{B}_{\mathbf{u}}$, which implies that the diagonal coefficients of $\mathbf{B}_{\mathbf{u}}$ are real. In addition, we have:

$$\forall \mathbf{x} \in \mathcal{U}, f(\mathbf{x}, \mathbf{x}) = f^{*}(\mathbf{x}, \mathbf{x}) \Longrightarrow q(\mathbf{x}) = f(\mathbf{x}, \mathbf{x}) \in \mathbb{R}.$$

The relation (15) then becomes:

$$q(\mathbf{x}) = \sum_{i=1}^{J} b_{ii} \left|x_i\right|^2 + 2 \sum_{1 \le i < j \le I} \operatorname{Re}\left(b_{ij} x_i x_j^{*}\right)$$

Hermitian forms

Using the symmetry property $f(\mathbf{y};\mathbf{x}) = f^*(\mathbf{x},\mathbf{y})$, we have

$$\begin{aligned}
q(\lambda\mathbf{x}) &= f(\lambda\mathbf{x},\lambda\mathbf{x}) = |\lambda|^2 q(\mathbf{x}) \\
q(\mathbf{x}+\mathbf{y}) &= f(\mathbf{x}+\mathbf{y},\mathbf{x}+\mathbf{y}) = q(\mathbf{x}) + f(\mathbf{x},\mathbf{y}) + f^*(\mathbf{x},\mathbf{y}) + q(\mathbf{y}) \\
q(\mathbf{x}-\mathbf{y}) &= f(\mathbf{x}-\mathbf{y},\mathbf{x}-\mathbf{y}) = q(\mathbf{x}) - f(\mathbf{x},\mathbf{y}) - f^*(\mathbf{x},\mathbf{y}) + q(\mathbf{y}) \\
q(\mathbf{x}+j\mathbf{y}) &= f(\mathbf{x}+j\mathbf{y},\mathbf{x}+j\mathbf{y}) = q(\mathbf{x}) - jf(\mathbf{x},\mathbf{y}) + jf^*(\mathbf{x},\mathbf{y}) + q(\mathbf{y}) \\
q(\mathbf{x}-j\mathbf{y}) &= f(\mathbf{x}-j\mathbf{y},\mathbf{x}-j\mathbf{y}) = q(\mathbf{x}) + jf(\mathbf{x},\mathbf{y}) - jf^*(\mathbf{x},\mathbf{y}) + q(\mathbf{y}).
\end{aligned}$$

Hermitian forms

From above equations, it is easy to deduce the following identities:

$$
\begin{aligned}
q(\mathbf{x}+\mathbf{y}) &= q(\mathbf{x}) + 2\,\mathrm{Re}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}-\mathbf{y}) &= q(\mathbf{x}) - 2\,\mathrm{Re}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}+j\mathbf{y}) &= q(\mathbf{x}) + 2\,\mathrm{Im}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}-j\mathbf{y}) &= q(\mathbf{x}) - 2\,\mathrm{Im}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R}.
\end{aligned}
$$

Hermitian forms

From above equations, it is easy to deduce the following identities:

$$
\begin{aligned}
q(\mathbf{x}+\mathbf{y}) &= q(\mathbf{x}) + 2\,\mathrm{Re}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}-\mathbf{y}) &= q(\mathbf{x}) - 2\,\mathrm{Re}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}+j\mathbf{y}) &= q(\mathbf{x}) + 2\,\mathrm{Im}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R} \\
q(\mathbf{x}-j\mathbf{y}) &= q(\mathbf{x}) - 2\,\mathrm{Im}[f(\mathbf{x},\mathbf{y})] + q(\mathbf{y}) \in \mathbb{R}.
\end{aligned}
$$

From these identities, it is easy to deduce the following theorem.

定理

The sesquilinear form f associated with the Hermitian form q can be determined using the following formula:

$$
f(\mathbf{x},\mathbf{y}) = \frac{1}{4}[q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}-\mathbf{y})] + \frac{j}{4}[q(\mathbf{x}+j\mathbf{y}) - q(\mathbf{x}-j\mathbf{y})]
$$

Hermitian forms

Form	Bilinear/quadratic	Sesquilinear/Hermitian
non-sym. bil/sesq	$\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{v}}}^T \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^T \mathbf{B}_{\mathbf{vu}} \mathbf{P} \end{cases}$	$\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{v}}}^H \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{v}}\,\underline{\mathbf{u}}} = \mathbf{Q}^H \mathbf{B}_{\mathbf{vu}} \mathbf{P} \end{cases}$
sym. bil/sesq	$\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{u}}}^T \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^T \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$	$\begin{cases} f(\mathbf{x},\mathbf{y}) = \mathbf{y}_{\underline{\mathbf{u}}}^H \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^H \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$
quadrat/Hermit	$\begin{cases} q(\mathbf{x}) = \mathbf{x}_{\underline{\mathbf{u}}}^T \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^T \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$	$\begin{cases} q(\mathbf{x}) = \mathbf{x}_{\underline{\mathbf{u}}}^H \mathbf{B}_{\underline{\mathbf{u}}} \mathbf{x}_{\underline{\mathbf{u}}} \\ \mathbf{B}_{\underline{\mathbf{u}}} = \mathbf{P}^H \mathbf{B}_{\mathbf{u}} \mathbf{P} \end{cases}$

Table: Change of basis formulae for billnear/sesquillnear forms and quadratic/Hermitian forms.

A quadratic form q is said to be positive if $q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{U}$. It is said to be positive definite if $q(\mathbf{x}) = f(\mathbf{x}, \mathbf{x}) > 0$ for all $\mathbf{x} \neq 0$. The associated matrix $\mathbf{B_u}$ then satisfies the condition $\langle \mathbf{B}_{\mathrm{u}}\mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T\mathbf{B}_{\mathrm{u}}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and it is said that $\mathbf{B}_{\mathrm{u}}$ is positive definite, which is written as $\mathbf{B_u} > 0$.

In the same way, it is said that a quadratic form q is positive semi-definite if $\mathbf{x}^T\mathbf{B}_{\mathrm{u}}\mathbf{x} \geq 0$, negative definite if $\mathbf{x}^T\mathbf{B}_{\mathrm{u}}\mathbf{x} < 0$, for all $\mathbf{x} \neq 0$ and negative semidefinite if $\mathbf{x}^T\mathbf{B_u}\mathbf{x} \leq 0$. It is written $\mathbf{B_u} \geq 0, \mathbf{B_u} < 0$, and $\mathbf{B_u} \leq 0$, respectively.

In the case of a Hermitian form, the previous definitions remain valid by replacing $\mathbf{x}^T$ by $\mathbf{x}^H$.

Special cases of bilinear/sesquilinear forms and quadratic/Hermitian forms

Types	Definitions	Properties
	Bilinear form: $f : \mathcal{U}^2 \to \mathbb{R}$	
Symmetric	$f(\mathbf{y},\mathbf{x}) = f(\mathbf{x},\mathbf{y}) = \mathbf{y}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$	$\mathbf{B}_\mathbf{u}^T = \mathbf{B}_\mathbf{u}$
Antisymmetric	$f(\mathbf{y},\mathbf{x}) = -f(\mathbf{x},\mathbf{y})$	$\mathbf{B}_\mathbf{u}^T = -\mathbf{B}_\mathbf{u}.$
	Sesquilinear form: $f : \mathcal{U}^2 \to \mathbb{C}$	
Hermitian symmetry	$f(\mathbf{y},\mathbf{x}) = f^*(\mathbf{x},\mathbf{y}) = \mathbf{y}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$	$\mathbf{B}_\mathbf{u}^H = \mathbf{B}_\mathbf{u}$
Antihermitian symmetry	$f(\mathbf{y},\mathbf{x}) = -f^*(\mathbf{x},\mathbf{y})$	$\mathbf{B}_\mathbf{u}^H = -\mathbf{B}_\mathbf{u}$

	Quadratic form : $q : \mathcal{U} \to \mathbb{R}$	
Symmetric	$q(\mathbf{x}) = f(\mathbf{x},\mathbf{x}) = \mathbf{x}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$	$\mathbf{B}_\mathbf{u}^T = \mathbf{B}_\mathbf{u}$
Positive definite	$q(\mathbf{x}) = \mathbf{x}_\mathbf{u}^T \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u} > 0\ ,\ \forall \mathbf{x}_\mathbf{u} \in \mathcal{U}\ ,\ \mathbf{x}_\mathbf{u} \neq \mathbf{0}$	$\mathbf{B}_\mathbf{u} > 0$
	Hermitian form : $q : \mathcal{U} \to \mathbb{R}$	
Hermitian symmetry	$q(\mathbf{x}) = f(\mathbf{x},\mathbf{x}) = \mathbf{x}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u}$	$\mathbf{B}_\mathbf{u}^H = \mathbf{B}_\mathbf{u}$
Positive definite	$q(\mathbf{x}) = \mathbf{x}_\mathbf{u}^H \mathbf{B}_\mathbf{u} \mathbf{x}_\mathbf{u} > 0\ ,\ \forall \mathbf{x}_\mathbf{u} \in \mathcal{U}\ ,\ \mathbf{x}_\mathbf{u} \neq \mathbf{0}$	$\mathbf{B}_\mathbf{u} > 0$

Examples of positive definite quadratic forms

For instance:

- In $\mathbb{R}^I$:
  $$f(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{I} x_i y_i \Rightarrow q(\mathbf{x}) = \|\mathbf{x}\|_2^2 = \sum_{i=1}^{I} x_i^2.$$
- In $\mathcal{C}^0([a, b], \mathbb{R})$, with $[a, b] \subset \mathbb{R}$:
  $f(x, y) = \langle x, y \rangle = \int_a^b x(t)y(t)dt \Rightarrow q(x) = \|x\|_2^2 = \int_a^b x^2(t)dt.$
- In $\mathbb{R}^{I \times I}$:
  $$f(\mathbf{A}, \mathbf{B}) = \mathrm{tr}\left(\mathbf{B}^T \mathbf{A}\right) = \sum_{i,j=1}^{I} a_{ij} b_{ij} \Rightarrow q(\mathbf{A}) = \|\mathbf{A}\|_F^2 = \mathrm{tr}\left(\mathbf{A}^T \mathbf{A}\right) = \sum_{i,j=1}^{I} a_{ij}^2.$$

定理 (Cauchy-Schwarz inequality)

Given a $\mathbb{R}$-v.s. $\mathcal{U}$ and a positive definite quadratic form q over $\mathcal{U}$, we have:

$$\forall (x, y) \in u^2, [f(x, y)]^2 \leq q(x) + q(y),$$

with the equality if and only if x and y are linearly dependent in the v.s. $\mathcal{U}$.

In the case where the quadratic form q is positive definite, it can be interpreted as a norm ($\|x\| = \sqrt{q(x)}$), and the associated form f as an inner product.

Cauchy-Schwarz and Minkowski inequalities

定理 (Cauchy-Schwarz inequality)

Given a $\mathbb{R}$-v.s. $\mathcal{U}$ and a positive definite quadratic form q over $\mathcal{U}$, we have:

$$\forall (x, y) \in u^2, [f(x, y)]^2 \leq q(x) + q(y),$$

with the equality if and only if x and y are linearly dependent in the v.s. $\mathcal{U}$.

In the case where the quadratic form q is positive definite, it can be interpreted as a norm ($\|x\| = \sqrt{q(x)}$), and the associated form f as an inner product.

定理 (Minkowski inequality)

If the quadratic form q is positive definite, we have:

$$\forall (x, y) \in \mathcal{U}^2, \sqrt{q(x+y)} \leq \sqrt{q(x)} + \sqrt{q(y)}$$

with the equality if and only if there exists $\lambda \geq 0$ such that $y = \lambda x$, or if $x = 0$.

Consider a $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension J, with the basis $\{u\}$, a symmetric bilinear form $f : \mathcal{U}^2 \to \mathbb{R}$, and its associated quadratic form q.

- Two vectors $x, y \in \mathcal{U}$ are called orthogonal relatively to f if $f(x, y) = 0$. It is also said that the vectors are f-orthogonal, and it is written $x \perp_f y$.
- When $q(x) = f(x, x) = 0$, that is, when x is orthogonal to itself, it is then said that x is isotropic, and the set of isotropic vectors is called the isotropic cone of f, denoted by C_f and such that $C_f = \{x \in \mathcal{U} : q(x) = 0\}$.
- If x is f-orthogonal to a set of vectors $\{y_1, \cdots, y_p\}$, then it is f-orthogonal to any linear combination of these vectors.
- Two subspaces $\mathcal{V}, \mathcal{W} \subset \mathcal{U}$ are said to be f-orthogonal, and denoted $\mathcal{V} \perp_f \mathcal{W}$, if:

$$\forall x \in \mathcal{V} \text{ and } \forall y \in \mathcal{W}, f(x, y) = 0.$$

- Let $\mathcal{W} \subset U$ be a subset of U. The f-orthogonal complement of $\mathcal{W}$, denoted by $\mathcal{W}^{\perp_f}$, is defined as the set:

$$\mathcal{W}^{\perp_f} = \{y \in \mathcal{U} : \forall x \in \mathcal{W}, x \perp_f y\}$$

This set $\mathcal{W}^{\perp_f}$ is a subspace of u. Indeed, given $u, v \in \mathcal{W}^{\perp_f}$, and $\lambda \in \mathbb{R}$, then for any $x \in \mathcal{W}$, we have $f(x, u + \lambda v) = f(x, u) + \lambda f(x, v) = 0$, and therefore $u + \lambda v \in \mathcal{W}^{\perp_f}$, implying that $\mathcal{W}^{\perp_f}$ is a subspace of $\mathcal{U}$.

- A basis $\{u\} = \{u_1, \cdots, u_J\}$ of $\mathcal{U}$ is said to be f-orthogonal if:

$$f(u_i, u_j) = 0, \quad \forall i, j \in \langle J \rangle, i \neq j.$$

We then have $q(x) = q\left(\sum_{j=1}^{J} x_j u_j\right) = \sum_{j=1}^{J} x_j^2 q(u_j)$. The matrix associated with f, in the basis $\{u\}$, is diagonal, having as diagonal elements the coefficients $q(u_j), j \in \langle J \rangle$. Conversely, if $\mathbf{B}_u$ is diagonal, then $f(u_i, u_j) = 0$ for $i \neq j$, which implies the f-orthogonality of the basis.

- The rank of q is equal to the number of vectors of the basis such that $q\left(u_j\right) \neq 0$. It is also the rank of the associated matrix $\mathbf{B}_{\mathcal{U}}$.
- The kernel of f is the subspace of $\mathcal{U}$, denoted by $\mathrm{Ker}(f)$ and defined as:

$$\mathrm{Ker}(f) = \{x \in \mathcal{U} : \forall y \in \mathcal{U}, f(x, y) = 0\}$$

- The bilinear form f is said to be non-degenerate if $\mathrm{Ker}(f) = \{0\}$. Otherwise, f is said to be degenerate.
- According to the rank theorem applied to f, the rank of f, denoted by $\mathrm{r}(f)$, is the integer defined by:

$$\mathrm{r}(f) = \dim(\mathcal{U}) - \dim[\mathrm{Ker}(f)],$$

  and if f is non-degenerate:

$$\mathrm{r}(f) = \dim(\mathcal{U}) = \mathrm{r}\left(\mathbf{B}_{\mathcal{U}}\right).$$

  Therefore, f is non-degenerate if and only if its matrix $\mathbf{B}_{\mathcal{U}}$, in any basis of $\mathcal{U}$, is regular.

Gauss reduction method

In the following, we assume that $\{u_1, \cdots u_J\}$ is an f-orthogonal basis of the $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension J.

定理 (Gauss reduction)

Any quadratic form can be written as a linear combination of squares of independent linear forms, the number of squares being equal to the rank of the quadratic form.

Indeed, considering the dual basis $\{u_1^*, \cdots u_J^*\}$ associated with the basis $\{u_1, \cdots u_J\}$, such that $u_i^*(x) = x_i$, we have:

$$\forall x \in \mathcal{U}, \quad q(x) = \sum_{j=1}^{J} q(u_j)\, x_j^2 = \sum_{j=1}^{J} \lambda_j u_j^*(x)^2, \quad \lambda_j = q(u_j) \in \mathbb{R}. \tag{16}$$

Since vectors $\{u_1^*, \cdots u_J^*\}$ form a basis of the dual space u^*, the linear forms $u_j^*(x)$ are linearly independent in $\mathcal{U}^*$. The rank of q is given by the number of non-zero coefficients λ_j. The Gauss method proceeds iteratively on j to diagonalize the matrix of q.

Sylvester's inertia law

定理 (Sylvester's law of inertia)

For any f-orthogonal basis of a $\mathbb{R}$-v.s. $\mathcal{U}$ of dimension J, the quadratic form associated with f can be written as a linear combination of squares of independent linear forms such as Equ. (16), with a sum of m squares (for $\lambda_j > 0$), a difference of n squares (for $\lambda_j < 0$), and $J - m - n$ zero terms corresponding to zero coefficients λ_j. Then, there exists a basis of $\mathcal{U}$ in which the matrix of the quadratic form q is block-diagonal of the form:

$$\begin{bmatrix} \mathbf{I}_m & & \mathbf{0} \\ & -\mathbf{I}_n & \\ \mathbf{0} & & \mathbf{0}_{k\times k} \end{bmatrix},$$

where $k = J - m - n$. This matrix is called the canonical form of q over $\mathbb{R}$.

The pair (m, n) is called the signature of the (real) quadratic form q.

Sylvester's inertia law

Sylvester's inertia law

From the signature (m, n), the quadratic form can be classified as follows:

- q is of rank $m+n$, and thereby nondegenerate if $m+n=J$;
- q is degenerate if $m+n<J$;
- q is positive definite if $(m, n)=(J, 0)$; the canonical form (4.83) is then the identity matrix of order J;
- q is negative definite if $(m, n)=(0, J)$.

An illustrated example

Consider the quadratic form $q : \mathbb{R}^2 \ni (x_1, x_2) \mapsto x_1 x_2 \in \mathbb{R}$. It is obvious that this quadratic form can be rewritten as:

$$(x_1, x_2) \mapsto \left(\frac{x_1 + x_2}{2}\right)^2 - \left(\frac{x_1 - x_2}{2}\right)^2,$$

namely, a linear combination, in the form of a difference of squares of the two following linear forms:

$$(x_1, x_2) \mapsto \frac{x_1 + x_2}{2} \text{ and } (x_1, x_2) \mapsto \frac{x_1 - x_2}{2}.$$

This result can be recovered from a diagonalization of the matrix of the original quadratic form which can be written as:

$$(x_1, x_2) \mapsto [x_1 \;\; x_2] \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{x}^T \mathbf{B} \mathbf{x}.$$

The characteristic polynomial of the square matrix $\mathbf{A}$, of order n, is the polynomial defined as:

$$p_{\mathbf{A}}(\lambda) = \det\left(\lambda \mathbf{I}_n - \mathbf{A}\right) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

The equation $p_A(\lambda) = 0$ is called the characteristic equation of A, and its roots are the eigenvalues of A. These roots may be all distinct, or some of them may be repeated.

The characteristic polynomial of the square matrix $\mathbf{A}$, of order n, is the polynomial defined as:

$$p_{\mathbf{A}}(\lambda) = \det\left(\lambda \mathbf{I}_n - \mathbf{A}\right) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

The equation $p_A(\lambda) = 0$ is called the characteristic equation of A, and its roots are the eigenvalues of A. These roots may be all distinct, or some of them may be repeated.

The following relations can be shown:

$$\det(\mathbf{A}) = (-1)^n a_0, \quad \operatorname{tr}(\mathbf{A}) = -a_{n-1}.$$

Cayley-Hamilton theorem

定理 (Cayley-Hamilton theorem)

Any matrix A *is a zero of its characteristic polynomial, in other words the matrix* $p(\mathbf{A}) = \mathbf{A}^n + a_{k-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I}_n$ *is zero. The characteristic polynomial is therefore an annihilating polynomial for* $\mathbf{A}$*.*

A polynomial p such that $p(\mathbf{A}) = 0$ is called an annihilating polynomial of $\mathbf{A}$, and the annihilating polynomial of lowest degree is called the minimal polynomial of $\mathbf{A}$, denoted by $p_{\min}$. Every annihilating polynomial is a multiple of $p_{\min}$. We see that the Cayley–Hamilton theorem provides an annihilating polynomial, called the characteristic polynomial.

Cayley-Hamilton theorem

定理 (Cayley-Hamilton theorem)

Any matrix A *is a zero of its characteristic polynomial, in other words the matrix* $p(\mathbf{A}) = \mathbf{A}^n + a_{k-1}\mathbf{A}^{n-1} + \cdots + a_1\mathbf{A} + a_0\mathbf{I}_n$ *is zero. The characteristic polynomial is therefore an annihilating polynomial for* $\mathbf{A}$*.*

A polynomial p such that $p(\mathbf{A}) = 0$ is called an annihilating polynomial of $\mathbf{A}$, and the annihilating polynomial of lowest degree is called the minimal polynomial of $\mathbf{A}$, denoted by $p_{\min}$. Every annihilating polynomial is a multiple of $p_{\min}$. We see that the Cayley–Hamilton theorem provides an annihilating polynomial, called the characteristic polynomial.

From the Cayley-Hamilton theorem, it is easy to show that: For all $p \geq n$, $\mathbf{A}^p$ can be expressed as a linear combination of $\mathbf{I}_n, \mathbf{A}, \cdots, \mathbf{A}^{n-1}$.

Cayley-Hamilton theorem

It should be noted that this theorem must not be stated using the trivial formula $\det(\mathbf{A}\mathbf{I}_n - \mathbf{A}) = 0$, but as $\det(\mathbf{M}) = 0$, with:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} - a_{11}\mathbf{I}_m & -a_{12}\mathbf{I}_n & \dots & -a_{1n}\mathbf{I}_n \\ -a_{21}\mathbf{I}_n & \mathbf{A} - a_{22}\mathbf{I}_n & \dots & -a_{2n}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1}\mathbf{I}_n & -a_{n2}\mathbf{I}_n & \dots & \mathbf{A} - a_{nn}\mathbf{I}_n \end{bmatrix} = \mathbf{I}_n \otimes \mathbf{A} - \mathbf{A} \otimes \mathbf{I}_n,$$

where the symbol $\otimes$ denotes the Kronecker product. The elements of $\mathbf{M}$ are matrices of $\mathbb{K}^{n\times n}$, and the determinant of $\mathbf{M}$ is itself a matrix of $\mathbb{K}^{n\times n}$ and not a scalar.

Right eigenvectors

定义

Given a square matrix $\mathbf{A} \in \mathbb{K}^{I \times I}$, a right eigenvector (or simply eigenvector) is a non-zero vector $\mathbf{v}_k \in \mathbb{K}^I$ satisfying:

$$\mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k \tag{17}$$

where the scalar λ_k is called the eigenvalue of $\mathbf{A}$ associated with the eigenvector $\mathbf{v}_k$.

The definition (17) can also be written as $(\mathbf{A} - \lambda_k \mathbf{I}_I)\,\mathbf{v}_k = 0$, which implies that the matrix $\mathbf{A} - \lambda_k \mathbf{I}_I$ is singular, and therefore its determinant is zero. The eigenvector $\mathbf{v}_k$ is in the nullspace of $\mathbf{A} - \lambda_k \mathbf{I}_I$, and the eigenvalues of $\mathbf{A}$ are the I roots of the characteristic polynomial, $p_{\mathbf{A}}(\lambda) = \det(\lambda \mathbf{I}_I - \mathbf{A}) = 0$. They will be denoted by $\lambda_i(\mathbf{A})$. If λ_i is a zero of order n_i of the characteristic equation, it is said that λ_i is a multiple eigenvalue, with an (algebraic) multiplicity n_i. Otherwise, it is said that λ_i is a simple eigenvalue.

Right eigenvectors

One defines the geometric multiplicity m_i of λ_i as the dimension of the eigensubspace associated with λ_i, namely, the maximal number of linearly independent eigenvectors associated with λ_i.

It can be shown that the geometric multiplicity of λ_i corresponds to the dimension of the kernel of $\mathbf{A} - \lambda_i \mathbf{I}$. Subsequently, using the rank theorem, it can be deduced that $m_i = \dim (\mathbf{A} - \lambda_i \mathbf{I}) = I - \mathrm{r}\,(\mathbf{A} - \lambda_i \mathbf{I})$.

A few properties of eigenvalues

- If $\mathrm{r}(\mathbf{A}) = k$, then $\lambda = 0$ is an eigenvalue with a geometric multiplicity $I - \mathrm{r}(\mathbf{A}) = I - k$.
- If $\mathbf{A}$ has k (distinct) non-zero eigenvalues, then $r(\mathbf{A}) \geq k$. Therefore, $\lambda = 0$ is an eigenvalue with an algebraic multiplicity $I - k$, and a geometric multiplicity $I - \mathrm{r}(\mathbf{A}) \leq I - k$.
- The algebraic multiplicity of an eigenvalue is therefore at least equal to its geometric multiplicity, that is, $n_i \geq m_i$.
- In the case where $\mathbf{A} \in \mathbb{K}^{I \times I}$ has $J \leq I$ distinct eigenvalues $\lambda_1, \cdots, \lambda_J$, of respective algebraic multiplicities $n_1, \cdots, n_J$, with $\sum_{j=1}^{J} n_j = I$, the characteristic polynomial can be written as $p(\lambda) = \prod_{j=1}^{J} (\lambda - \lambda_j)^{n_j}$.

Spectrum and regularity/singularity conditions

The spectrum of $\mathbf{A}$ is the set of eigenvalues of $\mathbf{A}$. It is denoted as $\mathrm{sp}(\mathbf{A}) = \{\lambda_i(\mathbf{A}), i \in \langle I \rangle\}$. The spectral radius of $\mathbf{A}$ is the maximal value of the moduli of the eigenvalues:

$$\rho(\mathbf{A}) = \max_{i \in \langle I \rangle} |\lambda_i(\mathbf{A})| .$$

Spectrum and regularity/singularity conditions

The spectrum of $\mathbf{A}$ is the set of eigenvalues of $\mathbf{A}$. It is denoted as $\mathrm{sp}(\mathbf{A}) = \{\lambda_i(\mathbf{A}), i \in \langle I \rangle\}$. The spectral radius of $\mathbf{A}$ is the maximal value of the moduli of the eigenvalues:

$$\rho(\mathbf{A}) = \max_{i \in \langle I \rangle} |\lambda_i(\mathbf{A})| \, .$$

- The spectral radius is such that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for any matrix norm.
- The matrix $\mathbf{A}$ is singular if and only if $0 \in \mathrm{sp}(\mathbf{A})$, that is, $\lambda = 0$ is an eigenvalue of $\mathbf{A}$.
- $\mathbf{A}$ is regular if and only if $0 \notin \mathrm{sp}(\mathbf{A})$.

Left eigenvectors

A left eigenvector of $\mathbf{A}$ associated with the eigenvalue μ_i, is a vector $\mathbf{u}_i \neq \mathbf{0}_I$ satisfying:

$$\mathbf{u}_i^H \mathbf{A} = \mu_i \mathbf{u}_i^H$$

or equivalently:

$$\mathbf{u}_i^H (\mu_i \mathbf{I}_I - \mathbf{A}) = 0,$$

which implies $\det(\mu_i \mathbf{I}_I - \mathbf{A}) = 0$, therefore the characteristic equation, demonstrating that left and right eigenvectors are associated with the same eigenvalues. However, it should be noted that right and left eigenvectors associated with the same eigenvalue are different in general.

Properties of eigenvectors

The eigenvectors of a matrix $\mathbf{A}$ satisfy the following properties.

- If $\mathbf{v}_k$ is an eigenvector, then any multiple $\alpha \mathbf{v}_k$ with $\alpha \neq 0$ is also an eigenvector. The set of all eigenvectors $\{\mathbf{v}_k\}$ associated with λ_k is a subspace of K^I, of dimension at least equal to 1 , called eigensubspace of $\mathbf{A}$ associated with λ_k.
- The p eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ associated with p distinct eigenvalues $\lambda_1, \cdots, \lambda_p$ are linearly independent.
- Given a right eigenvector $\mathbf{v}_k$ and a left eigenvector $\mathbf{u}_i$ associated with two distinct eigenvalues λ_k and λ_i, respectively, we have the following relation of orthogonality according to the Hermitian inner product:

$$\langle \mathbf{v}_k, \mathbf{u}_i \rangle = \mathbf{u}_i^H \mathbf{v}_k = 0 \tag{18}$$

  This relation means that any left eigenvector is orthogonal to any right eigenvector when the two vectors are associated with distinct eigenvalues.

Properties of eigenvectors

定理

Let $\mathbf{A} \in \mathbb{K}^{I \times I}$ *be such that all its eigenvalues are distinct. Let us define matrices* $\mathbf{P}, \mathbf{Q} \in \mathbb{K}^{I \times I}$ *formed by the (normalized) right and left eigenvectors of* $\mathbf{A}$*, respectively:*

$$\mathbf{P} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_I \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_I \end{bmatrix}$$

Then, from the orthogonality property (18), it can be inferred that $\mathbf{P}$ *and* $\mathbf{Q}$ *satisfy the following orthogonality relation:*

$$\mathbf{Q}^H \mathbf{P} = \mathbf{I}_I$$

that is, $\mathbf{P}$ *and* $\mathbf{Q}$ *are bi-unitary, and their inverses are such that* $\mathbf{P}^{-1} = \mathbf{Q}^H$ *and* $\mathbf{Q}^{-1} = \mathbf{P}^H$

Properties of eigenvectors

引理

For $\mathbf{A} \in \mathbb{K}^{I \times J}$ *and* $\mathbf{B} \in \mathbb{K}^{J \times I}$*, we have the following property:*

$$\lambda^J \det\left(\lambda \mathbf{I}_I - \mathbf{AB}\right) = \lambda^I \det\left(\lambda \mathbf{I}_J - \mathbf{BA}\right). \tag{19}$$

Properties of eigenvectors

引理

For $\mathbf{A} \in \mathbb{K}^{I \times J}$ *and* $\mathbf{B} \in \mathbb{K}^{J \times I}$, *we have the following property:*

$$\lambda^{J} \det\left(\lambda \mathbf{I}_{I} - \mathbf{AB}\right) = \lambda^{I} \det\left(\lambda \mathbf{I}_{J} - \mathbf{BA}\right). \tag{19}$$

定理

Given $\mathbf{A} \in \mathbb{K}^{I \times J}$ *and* $\mathbf{B} \in \mathbb{K}^{J \times I}$, *with* $J \geq I$, *the spectrum of* $\mathbf{BA}$ *consists of eigenvalues of* $\mathbf{AB}$ *and* $J - I$ *zeros.*

Properties of eigenvectors

引理

For $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times I}$, we have the following property:

$$\lambda^J \det(\lambda \mathbf{I}_I - \mathbf{A}\mathbf{B}) = \lambda^I \det(\lambda \mathbf{I}_J - \mathbf{B}\mathbf{A}). \tag{19}$$

定理

Given $\mathbf{A} \in \mathbb{K}^{I \times J}$ and $\mathbf{B} \in \mathbb{K}^{J \times I}$, with $J \geq I$, the spectrum of $\mathbf{B}\mathbf{A}$ consists of eigenvalues of $\mathbf{A}\mathbf{B}$ and $J - I$ zeros.

证明.

According to (19), we have $\det(\lambda \mathbf{I}_J - \mathbf{B}\mathbf{A}) = \lambda^{J-I} \det(\lambda \mathbf{I}_I - \mathbf{A}\mathbf{B})$, which implies:

$$\det(\lambda \mathbf{I}_J - \mathbf{B}\mathbf{A}) = 0 \Leftrightarrow \begin{cases} \det(\lambda \mathbf{I}_I - \mathbf{A}\mathbf{B}) = 0, \\ \lambda = 0 (\text{ with multiplicity } J - I), \end{cases}$$

that is, the non-zero eigenvalues of $\mathbf{A}\mathbf{B}$ and $\mathbf{B}\mathbf{A}$ are identical, and $\mathbf{B}\mathbf{A}$ has in addition $J - I$ zero eigenvalues. $\square$

Eigenvalues and eigenvectors of a regularized matrix

In many applications, we have to invert an ill-conditioned matrix $\mathbf{A}$, namely, such that the ratio $|\lambda_{\max}(\mathbf{A})| / |\lambda_{\min}(\mathbf{A})|$ is large, where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$, respectively, denote the largest and smallest eigenvalue (in modulus). The result of the numerical computation of $\mathbf{A}^{-1}$ is then very inaccurate. To improve this computation, a so-called regularization method consists in adding a constant $\alpha \neq 0$ to each diagonal term of $\mathbf{A}$. As a result, a so-called regularized matrix is obtained of the form $\mathbf{A} + \alpha \mathbf{I}_I$, which satisfies the following property.

Eigenvalues and eigenvectors of a regularized matrix

定理

Given the matrix A having eigenpairs $(\lambda_i, \mathbf{v}_i)$, $i \in \langle l \rangle$, *then the matrix* $\mathbf{A} + \alpha \mathbf{I}_l$ *has the eigenpairs* $(\lambda_i + \alpha, \mathbf{v}_i)$, $i \in \langle l \rangle$, *which means that the regularization operation has the effect of leaving eigenvectors unchanged, whereas eigenvalues are modified by adding the constant* α.

Other properties of eigenvalues

For all $\mathbf{A} \in \mathbb{K}^{I \times I}$, we have the following relations:

- $\lambda_i\left(\mathbf{A}^T\right) = \lambda_i(\mathbf{A}), \quad \lambda_i\left(\mathbf{A}^*\right) = \lambda_i^*(\mathbf{A}) \quad \Rightarrow \quad \lambda_i\left(\mathbf{A}^H\right) = \lambda_i^*(\mathbf{A})$
- $\lambda_i(-\mathbf{A}) = -\lambda_i(\mathbf{A}), \quad \lambda_i(k\mathbf{A}) = k\lambda_i(\mathbf{A}), \quad k \in \mathbb{K}$
- $\lambda_i\left(\mathbf{A}^{-1}\right) = \left[\lambda_i(\mathbf{A})\right]^{-1}$
- $\lambda_i\left(\mathbf{A}^k\right) = \left[\lambda_i(\mathbf{A})\right]^k, k \geq 2$
- $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{I} \lambda_i(\mathbf{A})$
- $\mathrm{tr}\left(\mathbf{A}^k\right) = \sum_{i=1}^{I} \lambda_i^k(\mathbf{A}),$ with $k \geq 2$
- $\det(\mathbf{A}) = \prod_{i=1}^{I} \lambda_i(\mathbf{A})$
- If $\mathbf{A}$ is a lower or upper triangular matrix, its eigenvalues are equal to its diagonal elements.
- Two similar matrices have the same characteristic polynomial, and the same eigenvalues with the same multiplicities.
- A matrix $\mathbf{A} \in \mathbb{K}^{I \times I}$ is diagonalizable if there exists a regular matrix $\mathbf{P} \in \mathbb{K}^{I \times I}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix.
- Any square matrix of order I that has I distinct eigenvalues is diagonalizable.

Symmetric/Hermitian matrices

定理

For a Hermitian matrix $\mathbf{A} \in \mathbb{C}^{I \times I}$, the eigenvalues are real-valued, and any right eigenvector of $\mathbf{A}$ is also a left eigenvector, associated with the same eigenvalue.

定理

The orthogonality relation (18) then implies that two eigenvectors of a Hermitian matrix associated with two distinct eigenvalues are orthogonal according to the Hermitian inner product. In the case where the I eigenvalues are distinct, it is thus possible to construct an orthonormal basis of eigenvectors such that $\mathbf{u}_i^H \mathbf{u}_j = \delta_{ij}$.

Symmetric/Hermitian matrices

定理

The eigenvalues of an anti-hermitian matrix are either zero, or pure imaginary numbers, and left and right eigenvectors associated with a same eigenvalue are identical.

Properties of eigenvalues of some special matrices

Classes of matrices	Properties of eigenvalues
$\mathbf{A} \in \mathbb{C}^{I \times I}$ Hermitian	$\lambda_i \in \mathbb{R}$
$\mathbf{A} \in \mathbb{C}^{I \times I}$ antihermitian	$\lambda_i^* = -\lambda_i$
$\mathbf{A} \in \mathbb{R}^{I \times I}$ symmetric	$\lambda_i \in \mathbb{R}$
$\mathbf{A} \in \mathbb{R}^{I \times I}$ antisymmetric	$\lambda_i^* = -\lambda_i$
$\mathbf{A} \in \mathbb{R}^{I \times I}$ positive definite	$\lambda_i > 0$
$\mathbf{A} \in \mathbb{R}^{I \times I}$ positive semi-definite	$\lambda_i \geq 0$
$\mathbf{A} \in \mathbb{R}^{I \times I}$ orthogonal	$\lvert\lambda_i\rvert = 1\,(\lambda_i = \pm 1 \text{ if } \lambda_i \in \mathbb{R})$
$\mathbf{A} \in \mathbb{C}^{I \times I}$ unitary	$\lvert\lambda_i\rvert = 1$

Table: Properties of eigenvalues of some special matrices

Orthogonal/unitary matrices

- The eigenvalues of a real orthogonal matrix are of modulus 1; The eigenvalues of a real symmetric and orthogonal matrix are equal to 1 or -1.
- The eigenvalues of a unitary matrix are all of unit modulus; The eigenvalues of a Hermitian and unitary matrix are equal to 1 or -1.

Eigenvalues and extrema of the Rayleigh quotient

定理

Let $\mathbf{A} \in \mathbb{R}^{I \times I}$ *be a real positive definite (or semi-definite) symmetric matrix. The extrema of the Rayleigh quotient* λ *of vector* x *with respect to* $\mathbf{A}$*, defined as:*

$$\lambda = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \qquad \textit{with } \mathbf{x} \neq \mathbf{0} \tag{20}$$

are eigenvalues of $\mathbf{A}$*. The extrema of this ratio are therefore given by the largest and smallest eigenvalues of* $\mathbf{A}$*.*

Eigenvalues and extrema of the Rayleigh quotient

定理

Let $\mathbf{A} \in \mathbb{R}^{I \times I}$ *be a real positive definite (or semi-definite) symmetric matrix. The extrema of the Rayleigh quotient* λ *of vector* x *with respect to* $\mathbf{A}$, *defined as:*

$$\lambda = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \qquad \textit{with } \mathbf{x} \neq \mathbf{0} \tag{20}$$

are eigenvalues of $\mathbf{A}$. *The extrema of this ratio are therefore given by the largest and smallest eigenvalues of* $\mathbf{A}$.

证明.

Re-expressing the quotient as $\lambda = \left(\mathbf{x}^T \mathbf{A} \mathbf{x}\right) \left(\mathbf{x}^T \mathbf{x}\right)^{-1}$, the gradient of λ with respect to $\mathbf{x}$ is:

$$\frac{\partial \lambda}{\partial \mathbf{x}} = 2 \frac{\mathbf{A}\mathbf{x}}{\left(\mathbf{x}^T \mathbf{x}\right)} - 2 \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\left(\mathbf{x}^T \mathbf{x}\right)^2} \mathbf{x} = \frac{2}{\mathbf{x}^T \mathbf{x}} \left(\mathbf{A} - \lambda \mathbf{I}_I\right) \mathbf{x}.$$

Eigenvalues and extrema of the Rayleigh quotient

证明.

The necessary condition to have an extremum of λ is obtained by canceling the gradient, which gives $(\mathbf{A} - \lambda \mathbf{I}_l)\,\mathbf{x} = \mathbf{0}_l$. Therefore, any extremum (global minimum or global maximum) of (20) is a root of the characteristic equation of $\mathbf{A}$, in other words, an eigenvalue of A. Concerning the second part of the proposition, it should be observed that a global maximum and a global minimum of the Rayleigh ratio are obtained by replacing in (20) the vector $\mathbf{x}$ by eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_l$ associated with the largest and smallest eigenvalues, $\lambda_1 \geq 0$ and $\lambda_l \geq 0$, respectively, that is:

$$0 \leq \lambda_l \leq \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_1, \quad \text{with} \quad \lambda_1 = \frac{\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1}{\mathbf{x}_1^T \mathbf{x}_1}, \quad \lambda_l = \frac{\mathbf{x}_l^T \mathbf{A} \mathbf{x}_l}{\mathbf{x}_l^T \mathbf{x}_l}.$$

□

Eigenvalues and extrema of the Rayleigh quotient

The eigenvalues can also be interpreted in terms of extrema of the quadratic form $q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$, under the constraint $\|\mathbf{x}\|_2^2 = 1$. Indeed, this constrained optimization problem of the equality type can be solved by using the Lagrangian $L(\mathbf{x}, \lambda) = q(\mathbf{x}) - \lambda\left(\mathbf{x}^T\mathbf{x} - 1\right)$, where λ is the Lagrange multiplier. The Karush-Kuhn-Tucker optimality conditions give:

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = 0, \frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} = 1 - \mathbf{x}^T\mathbf{x} = 0$$

from which we deduce that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, with $\mathbf{x}^T\mathbf{x} = 1$. By carrying over these expressions into $q(\mathbf{x})$, we obtain the following value of the optimized cost function: $q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} = \lambda$

Eigenvalues and extrema of the Rayleigh quotient

The eigenvalues can also be interpreted in terms of extrema of the quadratic form $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A}\mathbf{x}$, under the constraint $\|\mathbf{x}\|_2^2 = 1$. Indeed, this constrained optimization problem of the equality type can be solved by using the Lagrangian $L(\mathbf{x}, \lambda) = q(\mathbf{x}) - \lambda \left(\mathbf{x}^T\mathbf{x} - 1\right)$, where λ is the Lagrange multiplier. The Karush-Kuhn-Tucker optimality conditions give:

$$\frac{\partial L(\mathbf{x}, \lambda)}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x} = 0, \frac{\partial L(\mathbf{x}, \lambda)}{\partial \lambda} = 1 - \mathbf{x}^T\mathbf{x} = 0$$

from which we deduce that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, with $\mathbf{x}^T\mathbf{x} = 1$. By carrying over these expressions into $q(\mathbf{x})$, we obtain the following value of the optimized cost function: $q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} = \lambda$

Therefore, optimizing the quadratic form $q(\mathbf{x})$, under the constraint $\|\mathbf{x}\|_2^2 = 1$, is equivalent to calculating the eigenvalues of $\mathbf{A}$, and the value of the optimized criterion is equal to the largest eigenvalue (λ_1) or to the smallest eigenvalue (λ_I), depending on whether the optimization corresponds to a maximization or minimization of $q(\mathbf{x})$.

Generalized eigenvalues

Given $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I \times I}$, the notion of generalized eigenvalue is defined by means of the following equation:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \tag{21}$$

The generalized eigenvalues λ are determined by solving the equation:

$$\det(\lambda \mathbf{B} - \mathbf{A}) = 0. \tag{22}$$

Generalized eigenvalues

Given $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I \times I}$, the notion of generalized eigenvalue is defined by means of the following equation:

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{B}\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}. \tag{21}$$

The generalized eigenvalues λ are determined by solving the equation:

$$\det(\lambda \mathbf{B} - \mathbf{A}) = 0. \tag{22}$$

Similarly, it is easy to show that the calculation of the generalized eigenvalues defined by equation (21) is equivalent to searching for the extrema of the following ratio of quadratic forms:

$$\lambda = \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}}$$

where $\mathbf{A}$ and $\mathbf{B}$ are, respectively, a positive definite or semi-definite symmetric matrix, and a positive definite symmetric matrix. The absolute extrema of the ratio (22) are solutions to the equation:

$$(\mathbf{A} - \lambda \mathbf{B})\mathbf{x} = \mathbf{0} \text{ or equivalently } \left(\lambda \mathbf{I}_n - \mathbf{B}^{-1}\mathbf{A}\right)\mathbf{x} = \mathbf{0}.$$

that is, λ and x are an eigenvalue and an eigenvector of $\mathrm{B}^{-1}\mathrm{A}$.