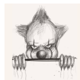
Some recalls

Lectures for PHD course on
Numerical Optimization

Enrico Bertolazzi

DII – Università di Trento



Notes

Outline

Notes

Outline

1 Determinant
- Some property of determinant
- Esistence and uniqueness
- Matrix product and determinant

Notes

- We always work with finite dimensional Eucledian vector spaces $\mathbb{R}^n$, the natural number n denote the dimension of the space.
- Elements $\boldsymbol{v} \in \mathbb{R}^n$ will be referred to as vectors, and we think them as composed of n real numbers stacked on top of each other, i.e.,

$$\boldsymbol{v} = \big(v_1, v_2, \ldots, v_n\big)^T = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

v_k being real numbers, and T denotes the transpose operator.

Notes

Basic operation

The basic operations defined for two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, and an arbitrary scalar $\alpha \in \mathbb{R}$

$$\boldsymbol{a} = \big(a_1, a_2, \ldots, a_n\big)^T \qquad \boldsymbol{b} = \big(b_1, b_2, \ldots, b_n\big)^T$$

are as follows:

1. addition: $\boldsymbol{a} + \boldsymbol{b} = \big(a_1 + b_1, \ldots, a_n + b_n\big)^T \in \mathbb{R}^n$;
2. multiplication by a scalar: $\alpha\boldsymbol{a} = \big(\alpha a_1, \ldots, \alpha a_n\big)^T \mathbb{R}^n$;
3. A linear subspace $L \subset \mathbb{R}^n$ is a set enjoying the following two properties:
    1. for every $\boldsymbol{a}, \boldsymbol{b} \in L$ it holds that $\boldsymbol{a} + \boldsymbol{b} \in L$;
    2. and for every $\alpha \in \mathbb{R}$, $\boldsymbol{a} \in L$ it holds that $\alpha\boldsymbol{a} \in L$.
4. An affine subspace $A \subset \mathbb{R}^n$ is any set that can be represented as $\boldsymbol{v} + L := \{\boldsymbol{v} + \boldsymbol{x} | \boldsymbol{x} \in L\}$ for some vector $\boldsymbol{v} \in \mathbb{R}^n$ and some linear subspace $L \subset \mathbb{R}^n$.

Notes

Scalar Product (real case)

A scalar product is a map $\mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}$ with the following properties:

1. Linearity

$$(\boldsymbol{a}, \alpha\boldsymbol{b} + \beta\boldsymbol{c}) = \alpha(\boldsymbol{a}, \boldsymbol{b}) + \beta(\boldsymbol{a}, \boldsymbol{c}) \qquad (\alpha\boldsymbol{b} + \beta\boldsymbol{c}, \boldsymbol{a}) = \alpha(\boldsymbol{b}, \boldsymbol{a}) + \beta(\boldsymbol{c}, \boldsymbol{a})$$

2. Simmetry

$$(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{b}, \boldsymbol{a})$$

3. Positivity

$$(\boldsymbol{a}, \boldsymbol{a}) \geq 0 \qquad (\boldsymbol{a}, \boldsymbol{a}) = 0 \quad \text{iif} \quad \boldsymbol{a} = \boldsymbol{0}$$

for example, the following product is a scalar product:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i \in \mathbb{R}.$$

Notes

Scalar Product (complex case)

A scalar product is a map $\mathbb{C}^n \times \mathbb{C}^n \mapsto \mathbb{C}$ with the following properties:

1. Linearity

$$(\boldsymbol{a}, \alpha\boldsymbol{b} + \beta\boldsymbol{c}) = \alpha(\boldsymbol{a}, \boldsymbol{b}) + \beta(\boldsymbol{a}, \boldsymbol{c}) \qquad (\alpha\boldsymbol{b} + \beta\boldsymbol{c}, \boldsymbol{a}) = \overline{\alpha}(\boldsymbol{b}, \boldsymbol{a}) + \overline{\beta}(\boldsymbol{c}, \boldsymbol{a})$$

2. (Conjugate) Simmetry

$$(\boldsymbol{a}, \boldsymbol{b}) = \overline{(\boldsymbol{b}, \boldsymbol{a})}$$

3. Positivity

$$(\boldsymbol{a}, \boldsymbol{a}) \geq 0 \qquad (\boldsymbol{a}, \boldsymbol{a}) = 0 \quad \text{iif} \quad \boldsymbol{a} = \boldsymbol{0}$$

for example, the following product is a scalar product:

$$\boldsymbol{a} \cdot \boldsymbol{b} = \overline{\boldsymbol{a}}^T \boldsymbol{b} = \sum_{i=1}^{n} \overline{a_i} b_i \in \mathbb{R}.$$

Notes

Norm

A norm is a map $\mathbb{R}^n \mapsto \mathbb{R}^+$ with the following properties:

1. Positivity

$$\|\boldsymbol{a}\| \geq 0 \qquad \|\boldsymbol{a}\| = 0 \quad \text{iif} \quad \boldsymbol{a} = \mathbf{0}$$

2. Homogeneity

$$\|\lambda\boldsymbol{a}\| = |\lambda|\,\|\boldsymbol{a}\|$$

3. Triangle inequality

$$\|\boldsymbol{a}+\boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\|$$

Notes

Most used norm in $\mathbb{R}^n$

1. Euclidean norm or 2-norm

$$\|\boldsymbol{a}\|_2 = \sqrt{\sum_{i=1}^{n} a_i^2}$$

2. 1-norm

$$\|\boldsymbol{a}\|_1 = \sum_{i=1}^{n} |a_i|$$

3. ∞-norm

$$\|\boldsymbol{a}\|_\infty = \max_{i=1}^{n} |a_i|$$

Notes

Cauchy–Bunyakowski–Schwarz inequality

Lemma

The Cauchy–Bunyakowski–Schwarz inequality says that

$$\left|(\boldsymbol{a},\boldsymbol{b})\right|^2 \leq (\boldsymbol{a},\boldsymbol{a})(\boldsymbol{b},\boldsymbol{b})$$

with equality iif $\boldsymbol{a} = \alpha\boldsymbol{b}$, i.e. $\boldsymbol{a}$ and $\boldsymbol{b}$ are parallel.

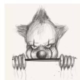
Proof: consider the vector $\boldsymbol{a} - \beta\boldsymbol{b}$:

$$0 \leq (\boldsymbol{a} - \beta\boldsymbol{b}, \boldsymbol{a} - \beta\boldsymbol{b}) = (\boldsymbol{a},\boldsymbol{a}) + \beta\overline{\beta}(\boldsymbol{b},\boldsymbol{b}) - \beta(\boldsymbol{a},\boldsymbol{b}) - \overline{\beta}(\boldsymbol{b},\boldsymbol{a})$$

choosing $\beta = \overline{(\boldsymbol{a},\boldsymbol{b})}/(\boldsymbol{b},\boldsymbol{b})$

$$0 \leq (\boldsymbol{a},\boldsymbol{a}) - \left|(\boldsymbol{a},\boldsymbol{b})\right|^2/(\boldsymbol{b},\boldsymbol{b})$$

if $\boldsymbol{a} = \alpha\boldsymbol{b}$ than $\beta = \alpha$ and inequality becomes equality. $\square$



Notes

Induced norm

A scalar product $(\cdot,\cdot)$ induce a norm $\|\cdot\|$ as follows:

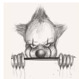
$$\|\boldsymbol{v}\| = \sqrt{(\boldsymbol{v},\boldsymbol{v})}$$

1. Positivity $\|\boldsymbol{a}\| = \sqrt{(\boldsymbol{a},\boldsymbol{a})}$ follows from property 3 of scalar product.
2. Homogeneity, from properties 1 and 2 of scalar product

$$\|\lambda\boldsymbol{a}\| = \sqrt{(\lambda\boldsymbol{a},\lambda\boldsymbol{a})} = \sqrt{\lambda^2(\boldsymbol{a},\boldsymbol{a})} = |\lambda|\,\sqrt{(\boldsymbol{a},\boldsymbol{a})}$$

3. Triangle inequality (by using Cauchy inequality for real case)

$$\begin{aligned}
\|\boldsymbol{a}+\boldsymbol{b}\|^2 &= (\boldsymbol{a}+\boldsymbol{b},\boldsymbol{a}+\boldsymbol{b}) = (\boldsymbol{a},\boldsymbol{a}) + (\boldsymbol{b},\boldsymbol{b}) + 2(\boldsymbol{a},\boldsymbol{b}) \\
&\leq (\boldsymbol{a},\boldsymbol{a}) + (\boldsymbol{b},\boldsymbol{b}) + 2\sqrt{(\boldsymbol{a},\boldsymbol{a})(\boldsymbol{b},\boldsymbol{b})} \\
&= \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 + 2\,\|\boldsymbol{a}\|\,\|\boldsymbol{b}\| = (\|\boldsymbol{a}\| + \|\boldsymbol{b}\|)^2
\end{aligned}$$



Notes

Orthogonality

- By the Cauchy inequality the number $\dfrac{(\boldsymbol{a},\boldsymbol{b})}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|}$ is in the interval $[-1,1]$
- The angle θ between two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ is defined as

$$\theta = \arccos \frac{(\boldsymbol{a},\boldsymbol{b})}{\|\boldsymbol{a}\|\,\|\boldsymbol{b}\|}.$$

- We say that $\boldsymbol{a}$ is orthogonal to $\boldsymbol{b}$ if and only if $(\boldsymbol{a},\boldsymbol{b}) = 0$.
- The only vector orthogonal to itself is $\boldsymbol{0} = (0,\ldots,0)^T$; moreover, this is the only vector with zero norm.

Notes

Linear and affine dependence

- A collection of vectors $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is said to be linearly independent if and only if

$$\sum_{i=1}^{k} \alpha_i \boldsymbol{v}_i = \boldsymbol{0} \qquad \Rightarrow \qquad \alpha_1 = \cdots = \alpha_k = 0.$$

- Similarly, a collection of vectors $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ is said to be affinely independent if and only if the collection

$$(\boldsymbol{v}_2 - \boldsymbol{v}_1, \boldsymbol{v}_3 - \boldsymbol{v}_1, \dots, \boldsymbol{v}_k - \boldsymbol{v}_1)$$

is linearly independent.

Notes

Basis

- The largest number of linearly independent vectors in $\mathbb{R}^n$ is n;
- n linearly independent vectors from $\mathbb{R}^n$ is referred to as basis.
- The basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is said to be orthogonal if $(\boldsymbol{v}_i, \boldsymbol{v}_j) = 0$ for all $i \neq j$. If, in addition $\|\boldsymbol{v}_i\| = 1$ for $i = 1, \ldots, n$, the basis is called orthonormal.
- Given the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ every vector $\boldsymbol{v}$ can be written in a unique way as $\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i$, and the n-tuple $(\alpha_1, \ldots, \alpha_n)$ will be referred to as coordinates of $\boldsymbol{v}$ in this basis.
- If the basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is orthonormal, the coordinates α_i are computed as $\alpha_i = (\boldsymbol{v}, \boldsymbol{v}_i)$.
- The space $\mathbb{R}^n$ will be typically equipped with the standard basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$.
- For every vector $\boldsymbol{v} = (v_1, \ldots, v_n)^T$ we have $(\boldsymbol{v}, \boldsymbol{e}_i) = v_i$ which allows us to identify vectors and their coordinates.

Notes

Matrices

- All linear functions from $\mathbb{R}^n$ to $\mathbb{R}^k$ can be represented by using a linear space of real matrices $\mathbb{R}^{k\times n}$ (i.e., with k row and n columns).
- Given a matrix $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ it will often be convenient to view it as a row of its columns, which are thus vectors in $\mathbb{R}^k$.
- Let $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ have elements A_{ij} we write $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$, where $\boldsymbol{a}_i = (A_{1i}, \ldots, A_{ki})^T \in \mathbb{R}^k$.
- The addition of two matrices and scalar-matrix multiplication are defined in a straightforward way. For $\boldsymbol{v} = (v_1, \ldots, v_n)^T \in \mathbb{R}^n$ we define

$$\boldsymbol{A}\boldsymbol{v} = \sum_{i=1}^{n} v_i \boldsymbol{a}_i \in \mathbb{R}^k$$

Notes

Matrix norm

Let be $\boldsymbol{A}$ an $n \times m$ matrix. If we have two vector norms $\|\cdot\|_a$ and $\|\cdot\|_b$ defined in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, we can define a matrix norm as follows:

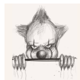
$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{v}\|_a=1} \|\boldsymbol{A}\boldsymbol{v}\|_b \qquad (*)$$

This is a norm and has the property

$$\|\boldsymbol{A}\boldsymbol{v}\|_b \leq \|\boldsymbol{A}\| \, \|\boldsymbol{v}\|_a$$

We say that matrix norm $\|\cdot\|$ is compatible with the vector norms $\|\cdot\|_a$ and $\|\cdot\|_b$. A compatible matrix norm not necessarily must be defined by a relation like (**??**), for example Frobenius norm

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} A_{ij}^2}$$



is compatible with the norm $\|\cdot\|_2$.

Notes

Most used matrix norm

1. 1-norm

$$\|\boldsymbol{A}\|_1 = \max_{\|\boldsymbol{v}\|_1=1} \|\boldsymbol{A}\boldsymbol{v}\|_1 = \max_{j=1}^{m} \sum_{i=1}^{n} |A_{ij}|$$

2. ∞-norm

$$\|\boldsymbol{A}\|_\infty = \max_{\|\boldsymbol{v}\|_\infty=1} \|\boldsymbol{A}\boldsymbol{v}\|_\infty = \max_{i=1}^{n} \sum_{j=1}^{m} |A_{ij}|$$

3. 2-norm

$$\|\boldsymbol{A}\|_2 = \max_{\|\boldsymbol{v}\|_2=1} \|\boldsymbol{A}\boldsymbol{v}\|_2 = \sqrt{\varrho(\boldsymbol{A}^T\boldsymbol{A})}$$

$\varrho(\boldsymbol{B})$ is the spectral ratio of matrix $\boldsymbol{B}$ defined forward.

Notes

Matrix norm and transpose

Definition

For a given matrix $\boldsymbol{A} \in \mathbb{R}^{k \times n}$ we define $\boldsymbol{A}^T \in \mathbb{R}^{n \times k}$ with elements $(\boldsymbol{A}^T)_{ij} = A_{ji}$ as matrix transpose

Definition

A more elegant definition: $\boldsymbol{A}^T$ is the unique matrix, satisfying the equality $(\boldsymbol{A}\boldsymbol{v}) \cdot \boldsymbol{u} = \boldsymbol{v} \cdot (\boldsymbol{A}^T\boldsymbol{u})$ for all $\boldsymbol{v} \in \mathbb{R}^n$ and $\boldsymbol{u} \in \mathbb{R}^k$.

Remark

Using different scalar products in the previous definition produces different transpose matrices.

Remark

From this definition it follows $(\boldsymbol{A}^T)^T = \boldsymbol{A}$

Notes

Matrix product

- Given two matrices $\boldsymbol{A} \in \mathbb{R}^{k\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, we define the product matrix product $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{k\times m}$ elementwise by

$$C_{ij} = \sum_{\ell=1}^{n} A_{i\ell} B_{\ell j}, \qquad i = 1,\ldots,k \quad j = 1,\ldots,m.$$

- In other words, $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ iff for all $\boldsymbol{v} \in \mathbb{R}^n$, $\boldsymbol{C}\boldsymbol{v} = \boldsymbol{A}(\boldsymbol{B}\boldsymbol{v})$.
- The matrix product is:
  - associative i.e., $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{C}) = (\boldsymbol{A}\boldsymbol{B})\boldsymbol{C}$;
  - not commutative i.e., $\boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A}$ in general;

  for matrices of compatible sizes.

Notes

Matrix norm and product (1/2)

Consider the vector spaces $\mathbb{R}^n$, $\mathbb{R}^k$ and $\mathbb{R}^m$ with norms $\|\cdot\|_a$, $\|\cdot\|_b$ and $\|\cdot\|_c$ respectively. We can define the matrix norms

$$
\begin{aligned}
\|\boldsymbol{A}\|_{ab} &= \max_{\|\boldsymbol{v}\|_a=1} \|\boldsymbol{A}\boldsymbol{v}\|_b \\
\|\boldsymbol{A}\|_{bc} &= \max_{\|\boldsymbol{v}\|_b=1} \|\boldsymbol{A}\boldsymbol{v}\|_c \\
\|\boldsymbol{A}\|_{ac} &= \max_{\|\boldsymbol{v}\|_a=1} \|\boldsymbol{A}\boldsymbol{v}\|_c
\end{aligned}
$$

If $\boldsymbol{A} \in \mathbb{R}^{n\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{k\times m}$ it is easy (and instructive) to check that

$$
\|\boldsymbol{A}\boldsymbol{B}\|_{ac} \leq \|\boldsymbol{A}\|_{ab} \|\boldsymbol{B}\|_{bc}
$$

Notes

Matrix norm and product

- Vectors $\boldsymbol{v} \in \mathbb{R}^n$ can be (and sometimes will be) viewed as matrices $\boldsymbol{v} \in \mathbb{R}^{n\times 1}$.
- Check that this embedding is norm-preserving, i.e., the norm of $\boldsymbol{v}$ viewed as a vector equals the norm of $\boldsymbol{v}$ viewed as a matrix with one column.
  In fact consider the definition of the matrix norm $\|\cdot\|'$ starting with the vector norm $\|\cdot\|$

$$\|\boldsymbol{v}\|' = \max_{|\alpha|=1} \|\alpha\boldsymbol{v}\| = \max_{|\alpha|=1} |\alpha|\,\|\boldsymbol{v}\| = \|\boldsymbol{v}\|$$

Notes

Matrix inverse

For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ we can discuss the existence of the unique matrix $\boldsymbol{A}^{-1}$, called the inverse of $\boldsymbol{A}$, verifying $\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{R}^n$. Or equivalently $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$ the identity matrix. If the inverse of a given matrix exist, we call the latter nonsingular.

Theorem

The inverse matrix exists iff

- *the columns of $\boldsymbol{A}$ are linearly independent;*
- *the columns of $\boldsymbol{A}^T$ are linearly independent;*
- *the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{v}$ has a unique solution for every $\boldsymbol{v} \in \mathbb{R}^n$;*
- *the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has $\boldsymbol{x} = \boldsymbol{0}$ as its unique solution.*

Notes

Matrix inverse

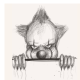
Lemma

From this definition it follows that $\boldsymbol{A}$ is nonsingular iff $\boldsymbol{A}^T$ is nonsingular, and, furthermore, $(\boldsymbol{A}^{-1})^T = (\boldsymbol{A}^T)^{-1}$ and therefore will be denoted simply as $\boldsymbol{A}^{-T}$.

At last, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are two nonsingular matrices of the same size, then $\boldsymbol{A}\boldsymbol{B}$ is nonsingular and $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

Notes

Outline



Notes

Determinant

Definition

Is a function on matrix

$$\left|\cdot\right| : \mathbb{K}^{n\times n} \mapsto \mathbb{K},$$

a law that for each (square) matrix $\boldsymbol{A} \in \mathbb{K}^{n}$ that return a scalar. The field $\mathbb{K}$ can be $\mathbb{R}$ or $\mathbb{C}$. Some properties must be verified.

Notes

To simplify notation split matrices by columns. Let $\boldsymbol{A}_{\bullet j}$ the j-th column of matrix $\boldsymbol{A}$,

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_{n-1n} \\ A_{n1} & \cdots & A_{nn-1} & A_{nn} \end{bmatrix}, \qquad \boldsymbol{A}_{\bullet j} = \begin{bmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{nj} \end{bmatrix},$$

in such a way matrix can be thought as column partitioned.

$$\boldsymbol{A} = (\boldsymbol{A}_{\bullet 1}, \ldots, \boldsymbol{A}_{\bullet n}),$$

so that

$$|\boldsymbol{A}| := |\boldsymbol{A}_{\bullet 1}, \ldots, \boldsymbol{A}_{\bullet n}|.$$

Notes

Determinant: axiomatic definition

Definition (Determinant properties)

'

1. *Is a multi-linear function of the columns*

$$\big|\ldots,\lambda\boldsymbol{a},\ldots\big| = \lambda\big|\ldots,\boldsymbol{b},\ldots\big|,$$

$$\big|\ldots,\boldsymbol{a}+\boldsymbol{b},\ldots\big| = \big|\ldots,\boldsymbol{a},\ldots\big| + \big|\ldots,\boldsymbol{b},\ldots\big|.$$

2. *Is null if two consecutive columns are equal*

$$\big|\ldots,\boldsymbol{a},\boldsymbol{a},\ldots\big| = 0.$$

3. *The determinant of identity matrix is* 1*:*

$$\big|\boldsymbol{I}\big| = \big|\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\big| = 1,$$

Notes

Determinant: particular cases

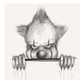
Observation

1. $n = 1, 2$ $\quad \left|A_{11}\right| = A_{11}, \qquad \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = A_{11}\,A_{22} - A_{21}\,A_{12};$
2. $n = 3$

$$\begin{vmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{vmatrix} = \left\{ \begin{array}{l} A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{21}A_{32}A_{13} \\ -A_{13}A_{22}A_{31} - A_{12}A_{21}A_{33} - A_{11}A_{23}A_{32}. \end{array} \right.$$

3. *Upper triangular*

$$\begin{vmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ 0 & A_{22} & \dots & A_{2n} \\ \vdots & & & \\ 0 & 0 & \dots & A_{nn} \end{vmatrix} = A_{11}A_{22}\dots A_{nn}.$$



Notes

Lemma (Multiply by a scalar)

$$\left|\lambda \boldsymbol{A}\right| = \lambda^n \left|\boldsymbol{A}\right|$$

Proof:

$$\begin{aligned}
\left|\lambda \boldsymbol{A}_{\bullet 1}, \lambda \boldsymbol{A}_{\bullet 2}, \lambda \boldsymbol{A}_{\bullet 3}, \ldots, \lambda \boldsymbol{A}_{\bullet n}\right| &= \lambda \left|\boldsymbol{A}_{\bullet 1}, \lambda \boldsymbol{A}_{\bullet 2}, \lambda \boldsymbol{A}_{\bullet 3}, \ldots, \lambda \boldsymbol{A}_{\bullet n}\right| \\
&= \lambda^2 \left|\boldsymbol{A}_{\bullet 1}, \boldsymbol{A}_{\bullet 2}, \lambda \boldsymbol{A}_{\bullet 3}, \ldots, \lambda \boldsymbol{A}_{\bullet n}\right| \\
&= \lambda^3 \left|\boldsymbol{A}_{\bullet 1}, \boldsymbol{A}_{\bullet 2}, \boldsymbol{A}_{\bullet 3}, \ldots, \lambda \boldsymbol{A}_{\bullet n}\right| \\
&= \ldots \\
&= \lambda^n \left|\boldsymbol{A}_{\bullet 1}, \boldsymbol{A}_{\bullet 2}, \ldots, \boldsymbol{A}_{\bullet n}\right|
\end{aligned}$$

□

Notes

Observation (Somma di matrici)

Notice that

$$|\boldsymbol{A}+\boldsymbol{B}| \neq |\boldsymbol{A}| + |\boldsymbol{B}|$$

for example

$$\begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 \\ 3 & 4 \end{vmatrix} \neq \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$$

in fact

$$\begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} = 1\cdot 0 - 2\cdot 0 = 0, \qquad \begin{vmatrix} 0 & 0 \\ 3 & 4 \end{vmatrix} = 0\cdot 4 - 0\cdot 3 = 0,$$

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 1\cdot 4 - 2\cdot 3 = 4 - 6 = -2$$

Notes

Lemma

If a column is 0 the determinant is 0.

Proof:

$$
\begin{aligned}
\big|\ldots,\mathbf{0},\ldots\big| &= \big|\ldots,0\cdot\mathbf{0},\ldots\big|,\\
&= 0\cdot\big|\ldots,\mathbf{0},\ldots\big|,\\
&= 0.
\end{aligned}
$$

$\square$

Notes

Lemma

If two consecutive columns are exchanged determinant sign change

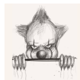
Proof: from property 2 it follows $\big|\ldots,\boldsymbol{w}+\boldsymbol{z},\boldsymbol{w}+\boldsymbol{z},\ldots\big| = 0$ and using multi-linearity

$$
\begin{aligned}
0 &= \big|\ldots,\boldsymbol{w}+\boldsymbol{z},\boldsymbol{w}+\boldsymbol{z},\ldots\big|,\\
&= \big|\ldots,\boldsymbol{w},\boldsymbol{w}+\boldsymbol{z},\ldots\big| + \big|\ldots,\boldsymbol{z},\boldsymbol{w}+\boldsymbol{z},\ldots\big|,\\
&= \big|\ldots,\boldsymbol{w},\boldsymbol{w},\ldots\big| + \big|\ldots,\boldsymbol{w},\boldsymbol{z},\ldots\big| + \big|\ldots,\boldsymbol{z},\boldsymbol{w},\ldots\big| + \big|\ldots,\boldsymbol{z},\boldsymbol{z},\ldots\big|,
\end{aligned}
$$

from property 2 $\big|\ldots,\boldsymbol{w},\boldsymbol{w},\ldots\big| = \big|\ldots,\boldsymbol{z},\boldsymbol{z},\ldots\big| = 0$ so that

$$
0 = \big|\ldots,\boldsymbol{w},\boldsymbol{z},\ldots\big| + \big|\ldots,\boldsymbol{z},\boldsymbol{w},\ldots\big|,
$$

$\square$



Notes

Lemma

If two columns equal determinant is 0.

Proof: Let $\boldsymbol{v}_i = \boldsymbol{v}_j$ for two column such that $i < j$. By exchanging consecutive columns it is possible to move $\boldsymbol{v}_i$ close to $\boldsymbol{v}_j$.

$$
\begin{aligned}
\left|\ldots,\boldsymbol{v}_i,\boldsymbol{v}_{i+1},\boldsymbol{v}_{i+2},\ldots,\boldsymbol{v}_j,\ldots\right| &= (-1)\left|\ldots,\boldsymbol{v}_{i+1},\boldsymbol{v}_i,\boldsymbol{v}_{i+2},\ldots,\boldsymbol{v}_j,\ldots\right|,\\
&= (-1)^2\left|\ldots,\boldsymbol{v}_{i+1},\boldsymbol{v}_{i+2},\boldsymbol{v}_i,\ldots,\boldsymbol{v}_j,\ldots\right|,\\
&= \ldots\\
&= \sigma\left|\ldots,\boldsymbol{v}_{i+1},\boldsymbol{v}_{i+2}\ldots,\boldsymbol{v}_i,\boldsymbol{v}_j,\ldots\right|
\end{aligned}
$$

where $\sigma = (-1)^{j-i} = \pm 1$. Let $\boldsymbol{v}_i = v^j = \boldsymbol{a}$, then

$$
\left|\ldots,\boldsymbol{a},\boldsymbol{a},\ldots\right| = 0.
$$

$\square$

Notes

Lemma

If two columns are exchanged (e.g. i-th and j-th with $i \neq j$) determinant change sign.

Proof:

$$
\begin{aligned}
0 &= \left|\ldots, \boldsymbol{w}+\boldsymbol{z}, \ldots, \boldsymbol{w}+\boldsymbol{z}, \ldots\right|, \\
&= \left|\ldots, \boldsymbol{w}, \ldots, \boldsymbol{w}+\boldsymbol{z}, \ldots\right| + \left|\ldots, \boldsymbol{z}, \ldots, \boldsymbol{w}+\boldsymbol{z}, \ldots\right|, \\
&= \left|\ldots, \boldsymbol{w}, \ldots, \boldsymbol{w}, \ldots\right| + \left|\ldots, \boldsymbol{w}, \ldots, \boldsymbol{z}, \ldots\right| + \left|\ldots, \boldsymbol{z}, \ldots, \boldsymbol{w}, \ldots\right| + \\
&\quad \left|\ldots, \boldsymbol{z}, \ldots, \boldsymbol{z}, \ldots\right|.
\end{aligned}
$$

so that

$$
0 = \left|\ldots, \boldsymbol{w}, \ldots, \boldsymbol{z}, \ldots\right| + \left|\ldots, \boldsymbol{z}, \ldots, \boldsymbol{w}, \ldots\right|,
$$

$\square$

Notes

Lemma

If to a column of the determinant we add a linear combination of the others the value of the determinat do not change.

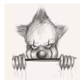
Proof: Let $\boldsymbol{b} = \sum\limits_{\substack{j=1\\j\neq i}}^{n} \beta_j \boldsymbol{v}_j$, with $\beta_1, \ldots, \beta_n$ scalars, then

$$\left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_i + \boldsymbol{b}, \boldsymbol{v}_{i+1}, \ldots\right| = \left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_i + \sum_{\substack{j=1\\j\neq i}}^{n} \beta_j \boldsymbol{v}_j, \boldsymbol{v}_{i+1}, \ldots\right|,$$

$$= \left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_i, \boldsymbol{v}_{i+1}, \ldots\right| + \sum_{\substack{j=1\\j\neq i}}^{n} \beta_j \left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_j, \boldsymbol{v}_{i+1}, \ldots\right|.$$

but $\left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_j, \boldsymbol{v}_{i+1}, \ldots\right| = 0$ for $j \neq i$ and

$$\left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_i + \boldsymbol{b}, \boldsymbol{v}_{i+1}, \ldots\right| = \left|\ldots, \boldsymbol{v}_{i-1}, \boldsymbol{v}_i, \boldsymbol{v}_{i+1}, \ldots\right|.$$



Notes

Theorem

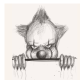
There exists a unique function that satisfy properties $1, 2, 3$ of the determinant.

Let

$$
\boldsymbol{A}_{\bullet j} = \sum_{k=1}^{n} A_{kj}\boldsymbol{e}_k,
$$

and from multi-linearity

$$
\begin{aligned}
\left|\boldsymbol{A}_{\bullet 1}, \ldots, \boldsymbol{A}_{\bullet n}\right| &= \left|\textstyle\sum_{i_1=1}^{n} A_{i_1 1}\boldsymbol{e}_{i_1}, \sum_{i_2=1}^{n} A_{i_2 2}\boldsymbol{e}_{i_2}, \ldots, \sum_{i_n=1}^{n} A_{i_n n}\boldsymbol{e}_{i_n}\right|, \\
&= \sum_{i_1=1}^{n} A_{i_1 1}\left|\boldsymbol{e}_{i_1}, \textstyle\sum_{i_2=1}^{n} A_{i_2 2}\boldsymbol{e}_{i_2}, \ldots, \sum_{i_n=1}^{n} A_{i_n n}\boldsymbol{e}_{i_n}\right|, \\
&= \sum_{i_1=1}^{n} A_{i_1 1} \sum_{i_2=1}^{n} A_{i_2 2} \cdots \sum_{i_n=1}^{n} A_{i_n n}\left|\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \ldots, \boldsymbol{e}_{i_n}\right|.
\end{aligned}
$$



Notes

The summation containts n^n termini, but obly $n!$ are not null. The term of the form

$$\left|\ldots, \boldsymbol{e}_{i_s}, \ldots, \boldsymbol{e}_{i_t}, \ldots\right|$$

with $i_s = i_t$ are 0. The only non zero term are the one with $i_1, \ldots i_n$ all differents, i.e. are permutation of $1, 2, \ldots, n$. Each permutation can be obtained by column exchange it follows

$$\left|\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \ldots, \boldsymbol{e}_{i_n}\right| = \sigma(i_1, i_2, \ldots, i_n)\left|\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\right|,$$

where $\sigma(i_1, i_2, \ldots, i_n) = \pm 1$ is called sign of the permutation. From $\left|\boldsymbol{I}\right| = 1$ it follows

$$\left|\boldsymbol{A}\right| = \sum_{\sigma \in \Pi(n)} \sigma(i_1, i_2, \ldots, i_n) A_{i_1 1} A_{i_2 2} \cdots A_{i_n n}$$

Notes

Corollary

Let $\mathcal{D}(\boldsymbol{A})$ a function that satisfy property 1 and 2 (not 3) then

$$\mathcal{D}(\boldsymbol{A}) = \left|\boldsymbol{A}\right|\mathcal{D}(\boldsymbol{I}).$$

Proof: In the theorem without using property 3 it follows

$$\begin{aligned}
\mathcal{D}(\boldsymbol{A}) &= \sum_{\sigma\in\Pi(n)} \sigma(i_1,i_2,\ldots,i_n)A_{i_1 1}A_{i_2 2}\cdots A_{i_n n}\mathcal{D}(\boldsymbol{e}_{i_1},\boldsymbol{e}_{i_2},\ldots,\boldsymbol{e}_{i_n}) \\
&= \sum_{\sigma\in\Pi(n)} \sigma(i_1,i_2,\ldots,i_n)A_{i_1 1}A_{i_2 2}\cdots A_{i_n n}\mathcal{D}(\boldsymbol{I}) \\
&= \left|\boldsymbol{A}\right|\mathcal{D}(\boldsymbol{I})
\end{aligned}$$

$\square$

Notes

Theorem (of Jacques Philippe Marie Binet 1786–1856)

Siano $\boldsymbol{A}$ e $\boldsymbol{B}$ due matrici quadrate dello stesso ordine allora

$$
\left|\boldsymbol{A}\boldsymbol{B}\right| = \left|\boldsymbol{A}\right|\left|\boldsymbol{B}\right|.
$$

Proof: Let $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \boldsymbol{A}\left[\boldsymbol{B}_{\bullet 1}, \ldots \boldsymbol{B}_{\bullet n}\right] = \left[\boldsymbol{A}\boldsymbol{B}_{\bullet 1}, \ldots \boldsymbol{A}\boldsymbol{B}_{\bullet n}\right],$. The determinant of the product is

$$
\left|\boldsymbol{C}\right| = \left|\boldsymbol{A}\boldsymbol{B}\right| = \left|\boldsymbol{A}\boldsymbol{B}_{\bullet 1}, \boldsymbol{A}\boldsymbol{B}_{\bullet 2}, \ldots, \boldsymbol{A}\boldsymbol{B}_{\bullet n}\right|.
$$

The function

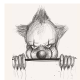
$$
\mathcal{D}_{\boldsymbol{A}}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \left|\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_n\right|,
$$

satisfy property 1 and 2 of the determinant an thust

$$
\mathcal{D}_{\boldsymbol{A}}(\boldsymbol{B}) = \left|\boldsymbol{A}\right|\mathcal{D}_{\boldsymbol{A}}(\boldsymbol{I}),
$$

and finally

$$
\mathcal{D}_{\boldsymbol{A}}(\boldsymbol{I}) = \left|\boldsymbol{A}_{\bullet 1}, \ldots, \boldsymbol{A}_{\bullet n}\right| = \left|\boldsymbol{A}\right|
$$



Notes

Outline

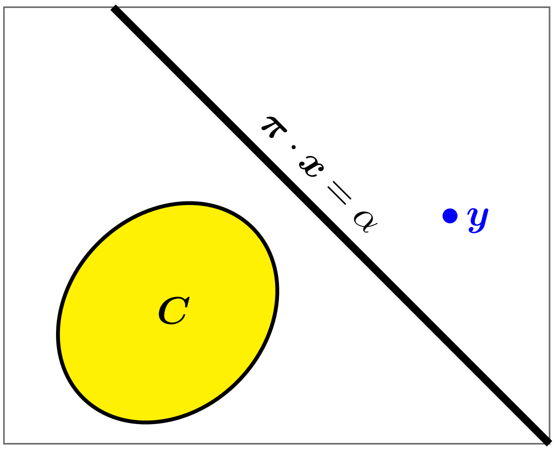
Notes

The Separation Theorem

Theorem (Separation Theorem)

Let be $C \subseteq \mathbb{R}^n$ closed and convex, and $\boldsymbol{y} \notin C$.
Then there exist a real α and a vector $\boldsymbol{\pi} \neq \mathbf{0}$ such that:

1. $\boldsymbol{\pi} \cdot \boldsymbol{y} > \alpha$;
2. $\boldsymbol{\pi} \cdot \boldsymbol{x} \leq \alpha$ for all $\boldsymbol{x} \in C$.



Notes

Proof: Define the function $\mathsf{f} : \mathbb{R}^n \mapsto \mathbb{R}$ by $\mathsf{f}(\boldsymbol{x}) = \frac{1}{2}\left\|\boldsymbol{x}-\boldsymbol{y}\right\|^2$. Now by the Weierstrass Theorem there exists $\boldsymbol{z} \in C$ such that:

$$\mathsf{f}(\boldsymbol{z}) \leq \mathsf{f}(\boldsymbol{x}), \qquad \forall \boldsymbol{x} \in C$$

due to the convexity of C we have $\boldsymbol{z} + t(\boldsymbol{x}-\boldsymbol{z}) \in C$ for all $t \in [0,1]$ and then

$$0 \leq \frac{\mathsf{f}(\boldsymbol{z}+t(\boldsymbol{x}-\boldsymbol{z})) - \mathsf{f}(\boldsymbol{z})}{t},$$

taking the limit $t \to 0$ and noticing that $\nabla \mathsf{f}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{y}$ we have

$$0 \leq \nabla \mathsf{f}(\boldsymbol{z})(\boldsymbol{x}-\boldsymbol{z}) = (\boldsymbol{z}-\boldsymbol{y}) \cdot (\boldsymbol{x}-\boldsymbol{z})$$

Now setting $\boldsymbol{\pi} = \boldsymbol{y} - \boldsymbol{z}$ and $\alpha = \boldsymbol{\pi} \cdot \boldsymbol{z}$ gives the result. $\square$

Notes

The Farkas's lemma

Lemma (Farkas's lemma)

Let $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, $\boldsymbol{b} \in \mathbb{R}^{n}$ and consider the following two problems

- (A) *Find $\boldsymbol{x} \in \mathbb{R}^{m}$ such that:* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *and* $\boldsymbol{x} \geq 0$;
- (B) *Find $\boldsymbol{\pi} \in \mathbb{R}^{n}$ such that:* $\boldsymbol{A}^T\boldsymbol{\pi} \leq \boldsymbol{0}$ *and* $\boldsymbol{\pi}\cdot\boldsymbol{b} > 0$;

then exactly only one of them has a solution.

Remark

$\boldsymbol{x} \geq 0$ *is intended component-wise, i.e., $x_k \geq 0$ for all k.*

Proof: $\Rightarrow$ If (A) IS feasible the (B) IS NOT feasible:
Let (A) has a feasible solution, say $\boldsymbol{x} \geq \boldsymbol{0}$, then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ so if there is a solution to (B), say $\boldsymbol{\pi}$, then $\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{\pi} = \boldsymbol{\pi}\cdot\boldsymbol{b} > 0$. But then $\boldsymbol{A}^T\boldsymbol{\pi} > \boldsymbol{0}$ (since $\boldsymbol{x} \geq \boldsymbol{0}$), a contradiction. Hence (B) is infeasible. $\square$

Notes

Proof: $\Rightarrow$ If (A) IS NOT feasible then (B) IS feasible:
Let $C = \{\boldsymbol{z} \in \mathbb{R}^m \mid \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$. If (B) is infeasible then $\boldsymbol{b} \notin C$.
The set C is convex and closed (see next slide) so by the Separation Theorem there exists a real α and a vector $\boldsymbol{\pi}$ such that $\boldsymbol{\pi} \cdot \boldsymbol{b} > \alpha$ and $\boldsymbol{\pi} \cdot \boldsymbol{z} \leq \alpha$ for all $\boldsymbol{z} \in C$, that is,

$$\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\pi} \leq \alpha, \qquad \forall \boldsymbol{x} \geq \boldsymbol{0}$$

Since $\boldsymbol{0} \in C$ it follows that $\alpha \geq 0$, so $\boldsymbol{\pi} \cdot \boldsymbol{b} > 0$. If there exists an $\boldsymbol{z} \geq \boldsymbol{0}$ such that $\boldsymbol{z}^T \boldsymbol{A}^T \boldsymbol{\pi} > 0$ then

$$\lim_{\lambda \to \infty} (\lambda \boldsymbol{z}^T) \boldsymbol{A}^T \boldsymbol{\pi} = \infty$$

Therefore we must have $\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\pi} \leq 0$ for all $\boldsymbol{x} \geq \boldsymbol{0}$, and this holds if and only if $\boldsymbol{A}^T \boldsymbol{\pi} \leq \boldsymbol{0}$, which means that (B) is feasible. $\square$

Notes

Proof: The set C is convex:
Let $C = \{\boldsymbol{z} \in \mathbb{R}^m \mid \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \geq \boldsymbol{0}\}$. Let $\boldsymbol{z}_1$ and $\boldsymbol{z}_2 \in C$ then there exists $\boldsymbol{x}_1 \geq \boldsymbol{0}$ and $\boldsymbol{x}_2 \geq \boldsymbol{0}$ such that $\boldsymbol{z}_1 = \boldsymbol{A}\boldsymbol{x}_1$ and $\boldsymbol{z}_2 = \boldsymbol{A}\boldsymbol{x}_2$.
Moreover

$$\begin{aligned} \alpha\boldsymbol{z}_1 + (1-\alpha)\boldsymbol{z}_2 &= \boldsymbol{A}\big(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2\big), \\ \alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2 &\geq \boldsymbol{0}, \qquad \forall\alpha \in [0,1]. \end{aligned}$$

so that C is convex. $\square$

Notes

Proof: The set C is closed:
To see that the set C is closed we prove that the complementary set is open. Let be $\boldsymbol{z} \in \mathbb{R}^m \setminus C$ than we have:

$$\inf_{\boldsymbol{x}\geq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{z}\| = \epsilon > 0$$

If $\epsilon > 0$ consider a $\boldsymbol{w}$ such that $\|\boldsymbol{w}-\boldsymbol{z}\| < \epsilon/2$ than we have

$$\begin{aligned}
\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{w}\| &= \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{w}+\boldsymbol{z}-\boldsymbol{z}\| \\
&\geq \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{z}\| - \|\boldsymbol{w}-\boldsymbol{z}\| = \epsilon - \epsilon/2 = \epsilon/2
\end{aligned}$$

so that

$$\inf_{\boldsymbol{x}\geq\boldsymbol{0}} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{w}\| \geq \epsilon/2 > 0$$

for all $\boldsymbol{w}$ such that $\|\boldsymbol{w}-\boldsymbol{z}\| < \epsilon/2$. $\square$

Notes

Proof: If $\epsilon = 0$ ...
TO BE COMPLETED
...so that $z \in \mathbb{R}^m \setminus C$ is open and thus C closed. $\square$

Notes

References

- R. Tyrrell Rockafellar
  Convex Analysis
  Princeton University Press, 1996.
- J. Farkas
  Theorie der einfachen Ungleichungen
  Journal für die reine und angewandte Mathematik, pp.1–27, **124**, 1902.
- `http://en.wikipedia.org/wiki/Multi-index_notation`

Notes