

1. VECTOR SPACES

1.1. LINEAR VECTOR SPACES

1.2. SUBSPACES. EXAMPLES.

1.3. LINEAR DEPENDENCE AND INDEPENDENCE. BASIS AND DIMENSION.

1.4. GEOMETRICAL CONSIDERATIONS: VECTOR NORM, INNER PRODUCT AND ORTHOGONALITY

The set of matrices with real or complex components and with M rows and N columns can be seen as a particular *linear vector space*. In this chapter we gain more insight in the geometric nature of vector spaces. Vector space theory will provide a useful geometric understanding of matrices introducing concepts such as length, angle, orthogonality and orthogonal projections.

1.1 LINEAR VECTOR SPACE

A linear vector space V over a set of scalars is a collection of objects known as vectors, together with an additive operation $+$ and a scalar multiplication operation \cdot , that satisfies the conditions below. A *real vector space* is a vector space in which the scalars comprise the set R of real numbers, and a *complex vector space* is a vector space in which the scalars comprise the set C of complex numbers.

- V forms a group under addition. That is, the following properties are satisfied.

a) For any v_1 and $v_2 \in V$, $v_1 + v_2 \in V$.

b) There is an identity element in S , which we will denote as 0 , such that for any $v_1 \in V$,

$$v_1 + 0 = 0 + v_1 = v_1$$

c) For every element $v_1 \in V$ there is another element $v_2 \in V$ such that

$$v_1 + v_2 = 0$$

The element v_2 is the additive inverse of v_1 .

d) The addition operation is associative; for any v_1 , v_2 and $v_3 \in V$,

$$(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$$

Vector Spaces

1.3

- For any $\alpha, \beta \in \mathbb{R}$ and any v_1 and $v_2 \in V$,

entonces

$$\alpha \cdot v_1 \in V,$$

$$\alpha \cdot (\beta \cdot v_1) = (\alpha\beta) \cdot v_1 = \beta \cdot (\alpha \cdot v_1)$$

$$(\alpha + \beta) \cdot v_1 = \alpha \cdot v_1 + \beta \cdot v_1$$

$$\alpha \cdot (v_1 + v_2) = \alpha \cdot v_1 + \alpha \cdot v_2$$

$$1 \cdot v_1 = v_1$$

$$0 \cdot v_1 = 0$$

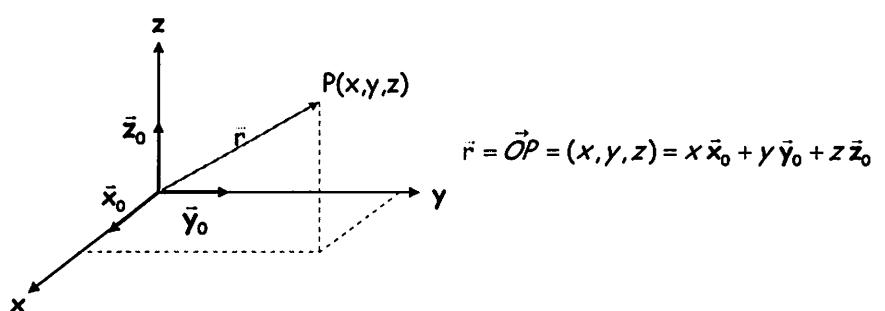
la propietat del producte per t.z. V sigui e.vectorial

Vector Spaces

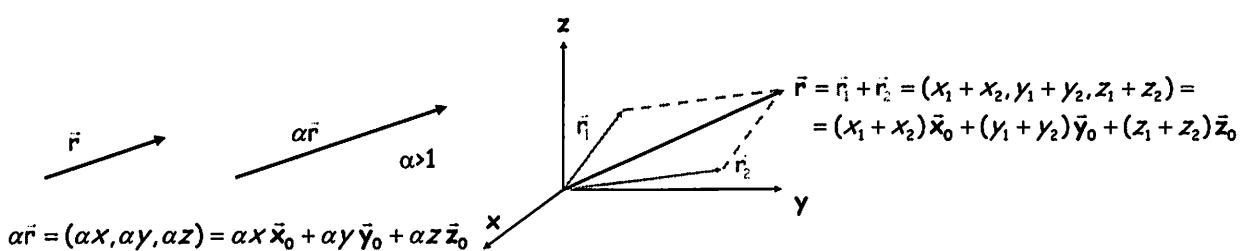
1.4

Examples (exemples d'espais vectorials)

- \mathbb{R}^3 Geometrical Vectors. The real vector space of all geometrical vectors in three-dimensional physical space



with vector addition and multiplication by a scalar.



2. \mathbb{R}^3 Free Geometrical Vectors with the same rules for addition and multiplication by a scalar.

3. \mathbb{R}^M (\mathbb{C}^M), the real (complex) vector space of all real (complex) M ordered numbers with the following rules:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \quad \alpha\mathbf{x} + \beta\mathbf{y} = \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \\ \vdots \\ \alpha x_M + \beta y_M \end{bmatrix}$$

4. The vector space of all real infinite sequences

$$\{x_i\} = x_1, x_2, \dots,$$

with $\{x_i\} + \{y_i\}$ defined as $\{x_i + y_i\}$ and similarly for $\alpha\{x_i\}$. Restrictions are usually placed on the behaviour of x_i as i tends to infinity.

5. The vector space of all real-valued functions $f(x)$ that are continuous, together with their first k derivatives, on the closed interval $[a, b]$ and with the usual rules for the addition and multiplication.

6. The vector space of all polynomials of arbitrary degree and having real coefficients. (polinomis d'urunt grām)

7. The vector space of all the functions $y(x)$ that solve a linear homogeneous differential equation

$$a_0(x) \frac{d^n y}{dx^n} + a_1(x) \frac{d^{n-1} y}{dx^{n-1}} + \dots + a_n(x) y = 0.$$

8. The complex vector space of all complex-valued continuous functions that solve the integral equation

$$\int_0^x K(x, t) y(t) dt + \lambda y(x) = 0, \quad 0 \leq x \leq a.$$

9. Matrices of dimension $M \times N$, with the usual rules of vector addition and multiplication by a scalar.

1.2 VECTOR SUBSPACE

Let $W \subset V$ be a subset of the vector space V and let define the vector addition and multiplication by a scalars for elements of W exactly as in V . Then, W is a vector subspace of V if and only if the following three conditions hold:

- W is nonempty.
- $\alpha \cdot w \in W$ for all $w \in W$ and all scalars α .
- $w + w' \in W$ for all pair $\{w, w'\} \in W$.

Example: Linear combination of vectors

Suppose that $L = \{v_1, v_2, \dots, v_k\}$ is some nonempty set of vector from V , with V the C^M space. Then, the set W of all linear combinations

$$w = \alpha_1 \cdot v_1 + \alpha_2 \cdot v_2 + \dots + \alpha_k \cdot v_k, \quad k \leq M$$

is a vector subspace of V . We say that $\{v_1, v_2, \dots, v_k\}$ spans (or generates) the W subspace or, equivalently, that $\{v_1, v_2, \dots, v_k\}$ is a spanning set of W . □

1.3 LINEAR DEPENDENCE AND LINEAR INDEPENDENCEDefinition

Let $L = \{v_1, v_2, \dots, v_k\}$ be a nonempty set of vectors.

1. The set $L = \{v_1, v_2, \dots, v_k\}$ is said to be linearly independent if

$$\sum_{j=1}^k \alpha_j \cdot v_j = 0 \quad \text{implies that} \quad \alpha_1 = \alpha_2 = \dots = \alpha_k = 0.$$

2. A set that is not linearly independent is said to be linearly dependent. Equivalently, there are scalars $\alpha_1, \dots, \alpha_k$ not all zero with $\sum_{j=1}^k \alpha_j \cdot v_j = 0$. In that case, the vector v_i with $\alpha_i \neq 0$ can be written as some linear combination of the set $\{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k\}$ as follows

$$v_i = \frac{\alpha_1}{\alpha_i} v_1 + \dots + \frac{\alpha_{i-1}}{\alpha_i} v_{i-1} + \frac{\alpha_{i+1}}{\alpha_i} v_{i+1} + \dots + \frac{\alpha_k}{\alpha_i} v_k$$

Theorems

1. Every subset of a linearly independent set is linearly independent.
2. Suppose that $L = \{v_1, v_2, \dots, v_k\}$ is a finite set of vectors and that some subset $L' \subset L$ is linearly dependent. Then, L is linearly dependent.

3. The set $L = \{v_1, v_2, \dots, v_k\}$ is a linearly dependent set if and only if one vector v_i is linearly dependent of the rest.

4. The set W of all linear combinations of $\{v_1, v_2, \dots, v_k\}$ with $v_i \in C^M$ for $i=1, \dots, k$. If $\{v_1, v_2, \dots, v_k\}$ spans a given vector subspace W , we may select a subset of linearly independent vectors among them that generates the same subspace.

BASIS AND DIMENSION

Definition: Basis for a vector space

A basis for a vector space V is any linearly independent set of vectors that spans V .

Theorems

1. Let $B = \{v_1, v_2, \dots, v_p\}$ be a basis. Then the representation of each v with respect to B is unique. That is, if

$$v = \sum_{j=1}^p \alpha_j \cdot v_j \quad \text{and also} \quad v = \sum_{j=1}^p \alpha'_j \cdot v_j$$

then $\alpha_j = \alpha'_j$ for $1 \leq j \leq p$.

dem: $\underline{v} - \underline{v} = 0$
 $\sum_{j=1}^p \alpha_j v_j - \sum_{j=1}^p \alpha'_j v_j = \sum_j (\alpha_j - \alpha'_j) v_j = 0 \Leftrightarrow \alpha_j - \alpha'_j = 0 \quad \text{pq. } v_1 \dots v_p$
són l.i.

2. If $\{u_1, u_2, \dots, u_L\}$ is some linearly independent subset of V , then $L \leq p$.

Definition: Dimension of a vector space

The number of vectors in a basis for a vector space is known as the dimension of the space. If the dimension of V is p , we also say that V is p -dimensional.

Theorem

1. The real vector space R^M and the complex vector space C^M are M -dimensional.

Base de C^M : $\{e_1, e_2, \dots, e_m\}$ on $e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i \quad \begin{cases} \text{base} \\ \text{euclídea} \end{cases}$

$$\text{si } \underline{v} \in C^n, \quad \underline{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^n v_i e_i \quad \therefore v_i \in C$$

1.4 GEOMETRICAL CONSIDERATIONS OF VECTORS

When dealing with vector spaces it is common to talk about the geometric concepts of the length and the direction of a vector. The mathematical concept associated with the length of a vector is the norm whereas the inner product provides an interpretation of angle between vectors, and hence direction.

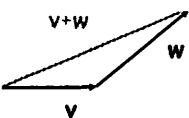
VECTOR NORM

Definition

Let V be a vector space with elements v . A real-valued function $\|v\|$ is said to be a norm if $\|v\|$ satisfies the following properties

- $\|v\| \geq 0$ for any $v \in V$.
- $\|v\| = 0$ if and only if $v = 0$.
- $\|\alpha v\| = |\alpha| \|v\|$, where α is an arbitrary scalar.
- $\|v + w\| \leq \|v\| + \|w\|$ (triangle inequality).

* In the various definitions
de norma, para
sempre ha de cumplir
las propiedades



p-Norm or Holder's Norm

For vectors $v = [v_1 \ v_2 \ \dots \ v_M]^T$ in R^M or C^M , the p-norm is defined as follows

$$\|v\|_p = \left(|v_1|^p + \dots + |v_M|^p \right)^{1/p}$$

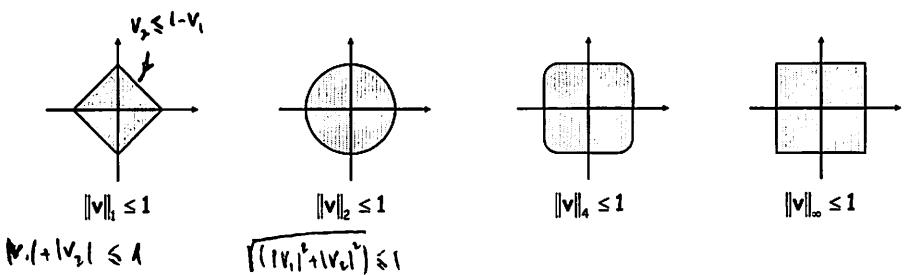
Some useful particular cases are:

- 1-norm: $\|v\|_1 = |v_1| + \dots + |v_M|$.
- 2-norm: $\|v\|_2 = \left(|v_1|^2 + \dots + |v_M|^2 \right)^{1/2}$.
- ∞ -norm: $\|v\|_\infty = \max_i |v_i|$.

Each of these norms introduces its own geometry. For example, let us define the unit "sphere" as follows

$$S_p = \{v \in R^2 : \|v\|_p \leq 1\}$$

Next figure shows the shape of such spheres for various values of p .



Example. Norm of functions over the interval $[a, b]$

1. The L_1 1-norm: $\|x(t)\|_1 = \int_a^b |x(t)| dt$.
2. The L_p p-norm: $\|x(t)\|_p = \left(\int_a^b |x(t)|^p dt \right)^{1/p}$ for $1 \leq p < \infty$.
3. The L_∞ ∞ -norm: $\|x(t)\|_\infty = \sup_{t \in [a, b]} |x(t)|$.

Hölder Inequality

$$|\mathbf{v}^H \mathbf{w}| \leq \|\mathbf{v}\|_p \|\mathbf{w}\|_q \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} = 1$$

with $\mathbf{v}^H \mathbf{w} = \sum_{i=1}^n v_i w_i$, that is, the "conventional" inner product.

$\{$ (vectors) \rightarrow scalar

INNER PRODUCT

An inner product is an operation on two vectors that returns a scalar value. Inner products can be used to provide geometrical interpretation of the "direction" of a vector in an arbitrary vector space.

Definition

Let V be a complex vector space. An inner product on V is a function that assigns to each ordered pair of vectors \mathbf{u} and \mathbf{v} in V a possibly complex number, denoted by $\langle \mathbf{u}, \mathbf{v} \rangle$, satisfying:

1. Hermitian symmetry:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle^* \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ in } V.$$

2. Hermitian linearity:

- $\langle \alpha \mathbf{u} + \beta \mathbf{w}, \mathbf{v} \rangle = \alpha^* \langle \mathbf{u}, \mathbf{v} \rangle + \beta^* \langle \mathbf{w}, \mathbf{v} \rangle$
- $\langle \mathbf{v}, \alpha \mathbf{u} + \beta \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle + \beta \langle \mathbf{v}, \mathbf{w} \rangle$

for all \mathbf{u}, \mathbf{v} and \mathbf{w} in V and all complex numbers α, β .

3. Positive definite:

$$\langle \mathbf{u}, \mathbf{u} \rangle > 0 \text{ if } \mathbf{u} \neq \mathbf{0}, \text{ and } \langle \mathbf{u}, \mathbf{u} \rangle = 0 \text{ if and only if } \mathbf{u} = \mathbf{0}.$$

It can be shown that $\|\mathbf{u}\| = \langle \mathbf{u}, \mathbf{u} \rangle^{1/2}$ is a norm on V .

Examples

1. Inner product for the space of functions defined in $[a, b]$.

$$\langle f(t), g(t) \rangle = \int_a^b \int_a^b f^*(x) w(x, y) g(y) dx dy$$

$$\langle f(t), g(t) \rangle = \int_a^b f^*(t) g(t) dt$$

Both are inner products

2. Expectation as an inner product[^]. Let X and Y be two r.v. with joint density $f_{XY}(x, y)$. We define an inner product between them as

$$\langle X, Y \rangle = \int xy f_{X,Y}(x, y) dx dy$$

which is the an expectation $\langle X, Y \rangle = E[XY]$. If X is a zero-mean r.v., then $\langle X, X \rangle = \text{var}(X)$ is a norm.

3. Euclidean Inner product. For vectors in C^M , the conventional inner product between the vectors

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_M \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_M \end{bmatrix}$$

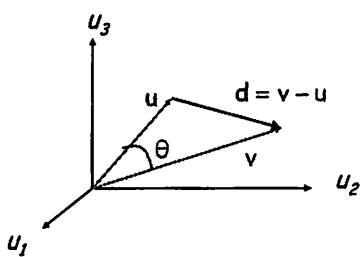
is

$$\left[\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^M u_i v_i \right] \leftarrow \begin{array}{l} \text{product} \\ \text{scalar} \\ \text{conventional} \end{array}$$

The 2-norm can be defined by this scalar product as follows

$$\|\mathbf{v}\|_2 = \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} = (\mathbf{v}^H \cdot \mathbf{v})^{1/2} = \left(\sum_{i=1}^M |v_i|^2 \right)^{1/2}$$

The angle between two vectors can be defined by this scalar product.



The 2-norm of d is equal to

$$\|d\|_2^2 = \|v - u\|_2^2 = \|\mathbf{v}\|_2^2 + \|\mathbf{u}\|_2^2 - 2\mathbf{u}^H \mathbf{v}$$

The law of cosines:

$$\|d\|_2^2 = \|\mathbf{v}\|_2^2 + \|\mathbf{u}\|_2^2 - 2\|\mathbf{v}\|_2 \|\mathbf{u}\|_2 \cos \theta.$$



$$\cos \theta = \frac{\mathbf{u}^H \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}$$

The same approach can be applied to define the angle between two vectors in any space V with inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ as follows

$$\cos \theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle^{1/2} \langle \mathbf{v}, \mathbf{v} \rangle^{1/2}}$$

$$\begin{aligned} \|\mathbf{v} - \mathbf{u}\|_2^2 &= \langle \mathbf{v} - \mathbf{u}, \mathbf{v} - \mathbf{u} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{u} \rangle - 2\langle \mathbf{v}, \mathbf{u} \rangle \\ R &= \|\mathbf{v}\|_2^2 + \|\mathbf{u}\|_2^2 - \underbrace{\mathbf{u}^H \mathbf{v} - \mathbf{v}^H \mathbf{u}}_{-2\mathbf{u}^H \mathbf{v}} \\ &= \|\mathbf{v}\|_2^2 + \|\mathbf{u}\|_2^2 - \underbrace{\mathbf{u}^H \mathbf{v} - \mathbf{v}^H \mathbf{u}}_{-2\operatorname{Re}\{\mathbf{u}^H \mathbf{v}\}} \end{aligned}$$

In elementary geometry, an specially important role is played by perpendicular or orthogonal vectors.

Definition

- Two vectors u and v in V are said to be orthogonal if and only if $\langle u, v \rangle = 0$.
- A set of vectors $\{u_1, u_2, \dots, u_L\}$ is said to be orthogonal if and only if every two vectors from the set are orthogonal. That is, $\langle u_i, u_j \rangle = 0 \quad i \neq j$.
- Additionally, a set of vectors is said to be orthonormal if and only if the set is orthogonal and $\|u_i\| = \langle u_i, u_i \rangle^{1/2} = 1$ for all u in the set.

Normalization of a vector: $u^N = \frac{u}{\|u\|}$ ← el factor de orthonormalidad → - norma unitaria
- l'angle queda igual

Theorem

Let $B = \{v_1, v_2, \dots, v_p\}$ be an orthogonal (or orthonormal) basis. Then the representation of any vector v with respect to B is

$$v = \sum_{j=1}^p \alpha_j \cdot v_j$$

with $\alpha_k = \frac{\langle v_k, v \rangle}{\langle v_k, v_k \rangle}$.

$$\text{dem: } v_k \in B, \quad \langle v_k, \sum_{j=1}^p \alpha_j \cdot v_j \rangle = \sum_j \alpha_j \langle v_k, v_j \rangle = \alpha_k \langle v_k, v_k \rangle \rightarrow \alpha_k = \frac{\langle v_k, v \rangle}{\langle v_k, v_k \rangle}$$

CREATING ORTHOGONAL AND ORTHONORMAL BASIS: GRAM-SCHMIDT PROCEDURE

Let us assume $\{u_1, u_2, \dots, u_L\}$ is a set of linearly independent vectors that span a given subspace W . How we can find an orthogonal or orthonormal basis, let say $\{v_1, v_2, \dots, v_L\}$, for that subspace?

Traditional Gram-Schmidt Procedure

1. Define $v_1 = u_1$.
2. For $2 \leq i \leq L$, define

$$v_i = u_i - \sum_{k=1}^{i-1} \alpha_{k,i} v_k$$

where $\alpha_{k,i} = \frac{\langle v_k, u_i \rangle}{\langle v_k, v_k \rangle}$ if $v_k \neq 0$ and $\alpha_{k,i} = 0$ if $v_k = 0$.

At each step, v_i is a vector that belongs to the subspace spanned by the set $\{v_1, \dots, v_{i-1}, u_i\}$ but orthogonal to vectors $\{v_1, \dots, v_{i-1}\}$.

2. BASIC MATRIX ALGEBRA

- 2.1. TERMINOLOGY AND BASIC OPERATIONS OF MATRICES.
- 2.2. SUBMATRICES AND PARTITIONED MATRICES.
- 2.3. SPECIAL MATRICES.
- 2.4. LINEAR SPACES: RANGE AND NULL SUBSPACES, RANK OF MATRICES AND NORM OF MATRICES.
- 2.5. INVERSE MATRIX, DETERMINANT AND TRACE.
- 2.6. KRONECKER PRODUCT AND VEC OPERATOR.

TERMINOLOGY AND BASIC OPERATIONS OF MATRICES

Definition

A $M \times N$ matrix is a rectangular array of MN scalars defined in C and enclosed in square brackets.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & a_{ij} & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MN} \end{bmatrix}$$

N

M

a_{ij} : is the entry of the i^{th} row and j^{th} column. In general, $a_{ij} \in C$ and it is said that $A \in C^{M \times N}$.

A matrix can also be represented as a set of row or column vectors:

$$A = [a_1 \ a_2 \ \cdots \ a_N] \quad \text{where } a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{Mj} \end{bmatrix}, \quad A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_M^T \end{bmatrix} \quad \text{where } a_i^T = [a_{i1} \ a_{i2} \ \cdots \ a_{iN}],$$

* els vectors son , per defecte, columnas .

Equality, addition and multiplication by an scalar

The matrix addition and the multiplication by an scalar define the set of matrices $C^{M \times N}$ as a linear space vector

1. Equality. $B = A$ if and only if $A, B \in C^{M \times N}$ and $b_{ij} = a_{ij} \quad \forall i, j$.

2. Addition: $C = A + B$ then $c_{ij} = a_{ij} + b_{ij} \quad \forall i, j$.

Theorems: Addition Laws

- Commutative law: $A + B = B + A$.
- Associative law: $A + (B + C) = (A + B) + C$.
- Null or Zero matrix: $A + 0 = A$.
- Negation matrix: $A + B = 0 \Rightarrow B = -A$.

3. Multiplication by a scalar: $B = \alpha A$ then $b_{ij} = \alpha a_{ij} \quad \forall i, j$.

Theorems: Laws of Multiplication by a Scalar

- | | |
|--|---|
| <ul style="list-style-type: none"> • $(\alpha + \beta)A = \alpha A + \beta A$. • $\alpha(\beta A) = (\alpha\beta)A$ • $\alpha(A + B) = \alpha A + \alpha B$. | <ul style="list-style-type: none"> • $(-1)A = -A$. • $0A = 0$. • $\alpha 0 = 0$ |
|--|---|

Matrix Multiplication

Let $A \in C^{M \times N}$ and $B \in C^{N \times Q}$. Then the product $C = AB$ is $C \in C^{M \times Q}$ whose (i,j) -entry is the product of the i^{th} row of A and the j^{th} column of B .

$$\boxed{C = A \cdot B \quad \text{where} \quad c_{ij} = \sum_{k=1}^N a_{ik} b_{kj} \quad \text{for } i = 1, 2, \dots, M \text{ and } j = 1, 2, \dots, Q}$$

$$\begin{array}{c}
 \begin{array}{ccccc}
 & A & \cdot & B & = & C \\
 \begin{matrix} 3^{\text{rd}} \text{ row} \rightarrow \\ \hline \end{matrix} &
 \left[\begin{array}{ccccc} x & x & x & x & x \\ x & x & x & x & x \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{array} \right] & \cdot &
 \left[\begin{array}{ccccc} x & \cdot & x & x \\ x & \cdot & x & x \end{array} \right] & = &
 \left[\begin{array}{ccccc} x & x & x & x \\ x & x & x & x \\ x & \cdot & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{array} \right] \leftarrow 3^{\text{rd}} \text{ row}
 \end{array} \\
 \uparrow 2^{\text{nd}} \text{ column} \qquad \uparrow 2^{\text{nd}} \text{ column}
 \end{array}$$

Theorems: Matrix Multiplication Laws

- Associative Law: $A(BC) = (AB)C$.
- Matrix Multiplication is not commutative: $AB \neq BA$.
- Distributive Law: $A(B \pm C) = AB \pm AC$.
- Multiplicative cancellation is not in general valid. That is,

if $AB = AC \not\Rightarrow B = C$ and

if $AB = 0 \not\Rightarrow A = 0 \text{ or } B = 0$.



A pot ser singular

Let $A \in \mathbb{C}^{MxN}$.

- The transpose of A is denoted by $A^T \in \mathbb{C}^{NxM}$ obtained by interchanging the rows and columns of A . That is, if $B = A^T$ then $b_{ij} = a_{ji} \forall i, j$.
- The conjugate of A is denoted by $A^* \in \mathbb{C}^{MxN}$ obtained by taking the complex conjugates of the entries in A . That is, if $B = A^*$ then $b_{ij} = \bar{a}_{ij} \forall i, j$.
- The hermitian transpose of A is denoted by $A^H \in \mathbb{C}^{NxM}$ obtained by taking the complex conjugates of A^T . That is, if $B = A^H$ then $b_{ij} = \bar{a}_{ji} \forall i, j$.

Theorems: Laws for Transposes

- $(A^H)^H = A$.
- $(A \pm B)^H = A^H \pm B^H$.
- $(\alpha A)^H = \bar{\alpha} A^H$.
- $(AB)^H = B^H A^H$ (note reversed order).

$$\begin{array}{c} \cancel{\text{if } AB = AC} \\ \cancel{\text{if } AB = 0} \end{array} \quad \begin{array}{c} \cancel{\text{if } B = C} \\ \cancel{\text{if } A = 0 \text{ or } B = 0} \end{array}$$

The same properties hold for $(\cdot)^T$ instead of $(\cdot)^H$.

SUBMATRICES AND PARTITIONED MATRICES.

It is useful to think of a matrix as being partitioned into *submatrices*.

$$A = \left[\begin{array}{cc|cc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{array} \right] = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}$$

Addition of partitioned matrices

Let $A, B \in \mathbb{C}^{MxN}$ be partitioned in the same way by

$$A = [A_1 \ A_2] \text{ and } B = [B_1 \ B_2] \implies A + B = [A_1 + B_1 \ A_2 + B_2]$$

where $A_1, B_1 \in \mathbb{C}^{MxQ}$ and $A_2, B_2 \in \mathbb{C}^{Mx(N-Q)}$.

Product of partitioned matrices

Let $A \in \mathbb{C}^{MxQ}$ and $B \in \mathbb{C}^{QxN}$ both partitioned as follows

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \implies AB = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix}$$

as long as the products and sums of the submatrices make sense.

Example

Suppose $A \in \mathbb{C}^{MxQ}$ is partitioned into its columns and $B \in \mathbb{C}^{QxN}$ into its rows

Basic Matrix Algebra

2.7

$$A = [a_1 \ a_2 \ \dots \ a_Q] \text{ and } B = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_Q^T \end{bmatrix}, \Rightarrow AB = a_1 b_1^T + a_2 b_2^T + \dots + a_Q b_Q^T. \quad M \times N$$

$M \times Q$ 1×1
 $Q \times N$ $Q \times N$
 $1 \times N$

Transposition of partitioned matrices

Let $A \in C^{M \times N}$ be partitioned by

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix}, \text{ then } A^H = \begin{bmatrix} A_{11}^H & A_{21}^H \\ A_{12}^H & A_{22}^H \\ A_{13}^H & A_{23}^H \end{bmatrix}.$$

* és una d'una matrius particionada, les propietats de matrius element a element també són aplicables.

Basic Matrix Algebra

2.8

SPECIAL MATRICES

Square Matrices

A matrix having the same number of rows and columns is called a square matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix}$$

Properties:

- Hermitian matrix: $A^H = A$.
- Symmetric matrix: $A^T = A$.
- Skew-symmetric: $A^T = -A$.

• Identity matrix: $I_N = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$

$N \times N$

Diagonal matrices

A diagonal matrix is a matrix whose off-diagonal elements are all equal to 0.

$$a_{ij} = 0 \quad i \neq j \Rightarrow A = \begin{bmatrix} x & 0 & 0 & 0 & 0 \\ 0 & x & 0 & 0 & 0 \\ 0 & 0 & x & 0 & 0 \\ 0 & 0 & 0 & x & 0 \end{bmatrix}$$

els elements fora de la diagonal són 0

Tridiagonal matrices

$$a_{ij} = 0 \quad |i - j| > 1 \Rightarrow A = \begin{bmatrix} x & x & 0 & 0 & 0 \\ x & x & x & 0 & 0 \\ 0 & x & x & x & 0 \\ 0 & 0 & x & x & x \end{bmatrix}$$

Upper triangular matrices

$$a_{ij} = 0 \quad i > j \Rightarrow A = \begin{bmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \end{bmatrix}$$

per sota de la diagonal
els elements són 0

Superior bidiagonal matrices

$$a_{ij} = 0 \quad \begin{cases} i > j \\ j > i+1 \end{cases} \Rightarrow A = \begin{bmatrix} x & x & 0 & 0 & 0 \\ 0 & x & x & 0 & 0 \\ 0 & 0 & x & x & 0 \\ 0 & 0 & 0 & x & x \end{bmatrix}$$

Lower triangular matrices

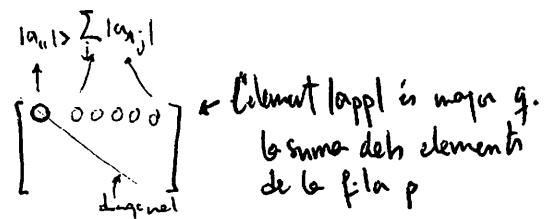
$$a_{ij} = 0 \quad i < j \Rightarrow A = \begin{bmatrix} x & 0 & 0 & 0 & 0 \\ x & x & 0 & 0 & 0 \\ x & x & x & 0 & 0 \\ x & x & x & x & 0 \end{bmatrix}$$

Upper Hessenberg matrices

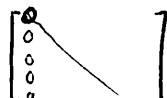
$$a_{ij} = 0 \quad i < j+1 \Rightarrow A = \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \end{bmatrix}$$

Dominant Diagonal Matrices

- If $|a_{pp}| > \sum_{j=1}^N |a_{pj}| \quad \forall p$ then A is called row dominant diagonal.



- If $|a_{pp}| > \sum_{i=1}^N |a_{ip}| \quad \forall p$ then A is called column dominant diagonal.



- If A is either column or row dominant diagonal, it is defined to be a dominant diagonal matrix.

• Util per trobar matrins singulars (+ endavant)

Basic Matrix Algebra
Permutation matrices

2.11

The column permutation matrix $P \in \mathbb{C}^{N \times N}$ results from permuting the columns of an $N \times N$ identity matrix.
The matrix that results from permuting the rows of an identity matrix is a row permutation matrix.
The product PA (AP) permutes the rows (columns) of A :

$$PA = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} a_{41} & a_{42} & a_{43} & a_{44} \\ a_{11} & a_{12} & a_{13} & a_{14} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \begin{matrix} 4^a \\ 1^a \\ 3^a \\ 2^a \end{matrix} \quad \text{permute files}$$

$$AP = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} a_{12} & a_{14} & a_{13} & a_{11} \\ a_{22} & a_{24} & a_{23} & a_{21} \\ a_{32} & a_{34} & a_{33} & a_{31} \\ a_{42} & a_{44} & a_{43} & a_{41} \end{bmatrix} \begin{matrix} 2^a \\ 4^a \\ 3^a \\ 1^a \end{matrix} \quad \text{permute columns}$$

Unitary (orthogonal) matrices

Let $A \in \mathbb{C}^{N \times N}$ be a matrix for which $A^{-1} = A^H$, so that $A^H A^{-1} = A^H A^H = I$, then A is said to be unitary.
An orthogonal matrix is a real unitary matrix A , so that $A^{-1} = A^T$ and $A^T A^{-1} = A^T A^T = I$.

The columns (rows) of a unitary matrix form an orthonormal set.

$$A^H A = \begin{bmatrix} a_1^H \\ a_2^H \\ \vdots \\ a_N^H \end{bmatrix} \begin{bmatrix} a_1, \dots, a_N \end{bmatrix} = \begin{bmatrix} a_1^H a_1 & a_1^H a_2 & \dots & a_1^H a_N \\ a_2^H a_1 & a_2^H a_2 & \dots & a_2^H a_N \\ \vdots & \vdots & \ddots & \vdots \\ a_N^H a_1 & a_N^H a_2 & \dots & a_N^H a_N \end{bmatrix} = I \quad \text{pq. } i, j \Rightarrow a_i^H a_j = \delta_{i,j} \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

Basic Matrix Algebra
Vandermonde matrices

An $M \times M$ Vandermonde matrix $V \in \mathbb{C}^{M \times M}$ has the form

$$V = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{M-1} & z_1^{M-1} & \dots & z_{M-1}^{M-1} \end{bmatrix}.$$

The determinant of the Vandermonde matrix is

$$\det(V) = \prod_{\substack{i,j=0 \\ i \neq j}}^{M-1} (z_i - z_j).$$

$$|\mathcal{V}| = \begin{vmatrix} 1 & 0 & \dots & 0 \\ z_0 & z_1 - z_0 & \dots & z_{M-1} - z_0 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{M-1} & z_1^{M-1} - z_0^{M-1} & \dots & z_{M-1}^{M-1} - z_0^{M-1} \end{vmatrix} \quad \left\{ \begin{array}{l} \text{observe q. } z_j^k - z_0^k = (z_j - z_0)(z_j^{k-1} + z_j^{k-2} z_0 + \dots + z_j z_0^{k-2} + z_0^{k-1}) \\ \text{desarrollar per la 1^a fila i traer factor común } (z_j - z_0) \end{array} \right.$$

$$\begin{aligned} & \text{restar la 1^a columna a las otras} \\ & = \prod_{j=0}^{M-1} (z_j - z_0) \begin{vmatrix} 1 & & \dots & 1 \\ z_1 + z_0 & & \dots & z_{M-1} + z_0 \\ \vdots & & & \vdots \\ z_1^{n-2} + z_0^{n-2} & \dots & z_{M-1}^{n-2} + z_0^{n-2} \end{vmatrix} = \prod_{\substack{i,j=0 \\ i \neq j}}^{M-1} (z_j - z_i) \end{aligned}$$

order-1

Fourier matrices

Define the symbol w by $w = \cos \frac{2\pi}{k} - i \sin \frac{2\pi}{k}$, where $i = \sqrt{-1}$. Then $\frac{1}{\sqrt{k}}$ times the $k \times k$ Vandermonde matrix F with elements $1, w, w^2, \dots, w^{k-1}$ is defined to be a Fourier matrix.

$$F = \frac{1}{\sqrt{k}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{k-1} \\ 1 & w^2 & w^4 & \cdots & w^{2k-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & w^{k-1} & w^{2k-2} & \cdots & w^{(k-1)(k-1)} \end{bmatrix}$$

$$\downarrow$$

$$z_i = e^{-j \frac{2\pi}{k} i}$$

Note that the pq -th element of F is $\left(\frac{1}{\sqrt{k}}\right)w^{(p-1)(q-1)}$.

Toeplitz Matrices

An $N \times N$ matrix is said to be a Toeplitz matrix if the entries are constant along each diagonal.

$$a_{ij} = a_{i-j} \quad \forall i, j \Rightarrow A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \dots \\ a_{+1} & a_0 & a_{-1} & a_{-2} \\ a_{+2} & a_{+1} & a_0 & a_{-1} \\ \dots & a_{+2} & a_{+1} & a_0 \end{bmatrix}$$

els elements de les
diagonals són iguals.

Block Toeplitz matrices

Let have the set of matrices $\{R_{-p+1}, R_{-p+2}, \dots, R_{p-1}\}$. The matrix A is called a block Toeplitz matrix if

$$A_{ij} = R_{i-j} \quad \forall i, j = 1, \dots, p \Rightarrow A = \begin{bmatrix} R_0 & R_1 & \cdots & R_{p-1} \\ R_{-1} & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_{-p+1} & \cdots & R_{-1} & R_0 \end{bmatrix}$$

In the practical problems¹ in which the block Toeplitz matrix arises, the matrices $\{R_{-p+1}, R_{-p+2}, \dots, R_{p-1}\}$ are usually Toeplitz.

Sylvester Matrices

The $(n+m) \times (n+m)$ matrix S is called Sylvester if S is in the following form, where the left submatrix of S is made of m columns and the right submatrix is made of n columns. It is important to observe that both submatrices are Toeplitz.

$$S = \begin{bmatrix} a_n & 0 & b_m & 0 \\ \vdots & \ddots & \vdots & \ddots \\ a_0 & a_n & b_0 & b_m \\ 0 & \ddots & \vdots & \ddots \\ & & 0 & b_0 \end{bmatrix}$$

¹ [Kay88] S. M. Kay, "Modern Spectral Estimation. Theory and Applications", Prentice Hall, Eng. Cliff, NJ, 1988

Hankel Matrices

The matrix $H \in \mathbb{R}^{n \times n}$ is called Hankel if H is in the form

$$H = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_n \\ h_1 & h_2 & h_3 & \cdots & h_{n-1} \\ h_2 & h_3 & h_4 & \cdots & h_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n-1} & h_{n-2} & \cdots & h_0 \end{bmatrix}$$

elements equals on b.
anti-diagonals

The general term of Hankel matrix H is $h_{ij} = h_{j+i-2}$ for some given sequence $h_0, h_1, \dots, h_{n-1}, h_n$. The entries of H are constant along the anti-diagonals.

Companion matrices

A matrix A_c of the form

$$A_c = \begin{bmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_n \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

Basic Matrix Algebra

2.16

is said to be a *companion matrix*, or a matrix companion to the polynomial $s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n$. There can be four forms of companion matrices depending on whether the $\{a_i\}$ occur in the first or last row or first or last column. We refer to them as top-, bottom-, left- or right-companion matrices.

Circulant matrices

A circulant matrix C is of the form

$$C = \begin{bmatrix} c_1 & c_2 & \cdots & c_m \\ c_m & c_1 & \cdots & c_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & \cdots & c_1 \end{bmatrix}$$

where each row is obtained by cyclically shifting to the right the previous row.

Lemma

1. Circulant matrices commute: if C_1 and C_2 are circulant matrices of the same size, then $C_1 C_2 = C_2 C_1$.
2. Circulant matrices are normal matrices. (A normal matrix C is a matrix such that $CC^H = C^H C$).

↓
les circulants
sont en podien
diagonalisation

def. $C = R^H T R$ & matrice C "est unitaire" $R^H R = I$
 T triangulaire supérieure

$C C^H = C^H C$

$R^H T R^H T^H R = R^H T^H R R^H T R$

$R^H T^H R = R^H T^H T R$

$T^H T = T^H T$

T est diagonal !

LINEAR SPACES: RANGE AND NULL SPACES

A matrix $A \in C^{M \times N}$ can be viewed as both a collection of N column vectors or as a collection of M row vectors. In fact, the linear transformation

$$\begin{array}{c} y = Ax \quad \forall x \\ \text{is the linear subspace spanned by the column vectors of } A. \end{array} \quad \left(\begin{array}{l} y = Ax = \sum_{j=1}^N x_j a_j \quad \forall x. \\ A x = [a_1 \cdots a_N] \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \sum_{j=1}^N x_j a_j \end{array} \right)$$

We define the column (row) space of A as the vector space spanned by the N column (M row) vectors of matrix A .

Range space of a matrix

The range space of $A \in C^{M \times N}$, denoted by $\mathfrak{R}(A)$, is the vector subspace spanned by the N columns of A .

$$\mathfrak{R}(A) = \left\{ y : y = Ax = \sum_{j=1}^N x_j a_j; \quad x \in C^{N \times 1} \right\} \quad \begin{array}{l} \text{Range: e.vect. generat per tots} \\ \text{els vectors column de } A \end{array}$$

Null space of a matrix

The null space of $A \in C^{M \times N}$ is denoted by $N(A)$ and is a vector subspace defined by the set of vectors that

$$N(A) = \left\{ x : Ax = 0; \quad x \in C^{N \times 1} \right\} \quad \begin{array}{l} \text{Null: e.vect. generat per } x \text{ t.q. } Ax = 0 \end{array}$$

$\text{Rank} = \text{range} \neq \text{Range}$

Matrix Rank

The dimension of the column space of A is called the rank of A and is equal to the number of linearly independent columns of A .

$$\text{rank}(A) = \dim[\mathfrak{R}(A)]$$

Rank: dimensió del range

Theorems

1. The number of linearly independent column vectors is equal to the number of linearly independent row vectors. That is

$$\text{qualsevol} \quad \text{rank}(A) = \text{rank}(A^T) \quad \text{rank per columnes} = \text{rank per filers}$$

2. $\text{rank}(A) + \dim[N(A)] = N$. (min(M, N) \leftarrow rectangle)
3. $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$ unless B is singular and then $\text{rank}(AB) = \text{rank}(A)$.
4. Sylvester Inequality. Let $A \in C^{M \times Q}$ and $B \in C^{Q \times N}$, then

$$\text{rank}(A) + \text{rank}(B) - Q \leq \text{rank}(AB) \leq \min[\text{rank}(A), \text{rank}(B)]$$

5. In general,

$$\text{rank}(A+B) \leq \text{rank}(A) + \text{rank}(B)$$

6. Let $A \in C^{N \times N}$ be a square matrix, then the following three statements are equivalent:

- A is non-singular
- $N(A) = \{0\}$.
- $\text{rank}(A) = N$.

NORM OF A MATRIX

The norm of a matrix A must satisfy the properties of any vector norm (see section 1.4):

- $\|A\| \geq 0$ for any $A \in V$.
- $\|A\| = 0$ if and only if $A = 0$.
- $\|\alpha A\| = |\alpha| \|A\|$, where α is an arbitrary scalar.
- $\|A + B\| \leq \|A\| + \|B\|$ (triangle inequality).

Usual Norm of a matrix

If we think in a matrix $A \in C^{M \times N}$ as an element of a linear space V of $M \times N$ matrices, then we may define the norm of a matrix from the inner product.

Let $A, B \in C^{M \times N}$ be defined in a linear space V of $M \times N$ matrices. The inner product $\langle A, B \rangle$ is

$$\langle A, B \rangle = \text{tr}(A^H B),$$

and the so-called usual norm of $A \in C^{M \times N}$ is

$$\|A\|_F = \langle A, A \rangle^{1/2} = [\text{tr}(A^H A)]^{1/2}.$$

Matrix Norm as a Linear Operator Norm

A matrix $A \in C^{M \times N}$, such that

\downarrow
Norma de
Frobenius

(es la norma asociada
al producto escalar
de las matrices (e.v.)

$$A^H = \begin{bmatrix} a_1^H \\ a_2^H \\ \vdots \\ a_N^H \end{bmatrix} \quad B = [b_1 \quad \cdots \quad b_N]$$

$$\text{tr}(A^H B) = \sum_{i=1}^N a_i^H b_i$$

$$y = Ax,$$

can be seen as a linear operator that transforms a vector $x \in C^N$ defined in the vector space V , named domain space, to another vector $y \in C^M$ defined in the vector space W , named image space.

An operator norm can be defined so that it provides an indication of the maximal amount of change on the magnitude of the inputs.

Definition

Let $A \in C^{M \times N}$ be a linear transformation from V to W having norms $\|\cdot\|_V$ and $\|\cdot\|_W$. The norm of the matrix is the supremum of the norm increase.

$$\|A\|_{V,W} = \sup_{x \neq 0} \frac{\|Ax\|_W}{\|x\|_V}$$

norma de y
 supremo \rightarrow mínimo de los superiores
 norma de x

□

- If $\|A\|_{V,W}$ is small, then all vectors are reduced in length.
- If $\|A\|_{V,W}$ is large, then some vectors are increased in length.

$$\|A\|_{V,W} \geq \frac{\|Ax\|}{\|x\|}, \forall x$$

done informació de com afecta a la norma la transformació
 amb A ($x \rightarrow y$)

P-norm of a matrix

The p operator norm of A is

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

Special cases:

- 1-norm: $\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$ (maximum absolute column sum).
- 2-norm: $\|A\|_2 = (\text{maximum eigenvalue of } A^H A)^{1/2}$
- ∞ -norm: $\|A\|_\infty = \max_i \sum_{j=1}^N |a_{ij}|$ (maximum absolute row sum).

dem. 1-norm: $y = Ax \quad \|\underline{A}\underline{x}\|_1 = \sum_{i=1}^n |y_i| = \sum_{i=1}^n \left| \sum_{j=1}^N a_{ij} x_j \right| \leq \sum_{i=1}^n \sum_{j=1}^N |a_{ij}| |x_j| \leq \sum_{i=1}^n \max_j |a_{ij}| \sum_{j=1}^N |x_j| = \sum_{i=1}^n \max_j |a_{ij}| \|\underline{x}\|_1$

per tant, $\|\underline{A}\|_1 = \frac{\|\underline{A}\underline{x}\|_1}{\|\underline{x}\|_1} = \sum_{j=1}^N \max_i |a_{ij}|$

INVERSE MATRICESSome basic definitions

Let $A \in C^{M \times N}$, then

- the right-inverse of A is an $N \times M$ matrix R such that

$$AR = I_M.$$

- the left-inverse of A is an $N \times M$ matrix L such that

$$LA = I_N.$$

A matrix may or may not have a right or left inverse. If they exist, they may be not unique.

Theorem

- An $M \times N$ matrix A has a right-inverse if and only if $\text{rank}(A) = M$ (i.e., if and only if A has full ^{row} rank) and has a left-inverse if and only if $\text{rank}(A) = N$ (i.e., if and only if A has full column rank).

A rectangular matrix $A \in C^{M \times N}$ cannot possibly have a two-sided inverse. If $M < N$ ($N < M$), then A cannot possibly have a left-inverse (right-inverse), but may possibly have a right-inverse (left-inverse).

$M \neq N \Rightarrow$ Nunca podrá tener ma
de los dos inversos
(\log . tiene
full rank)

$A = M \times N \left\{ \begin{array}{ll} \text{rank}(A) = M & \text{full row rank} \\ \text{rank}(A) = N & \text{full column rank} \end{array} \right.$

Basic Matrix Algebra

2.23

- If both a left-inverse L and a right-inverse R exist for A , then they are equal and it is a two-sided inverse, denoted by $A^{-1} = L = R$.
den. $L = L \cdot I = L(A \cdot R) = (L \cdot A) \cdot R = I \cdot R = R$

A square matrix $A \in C^{N \times N}$, either has a unique two-sided inverse A^{-1} , or A has neither a left- nor a right inverse.

$\boxed{\text{if } A \text{ is quadrade} \rightarrow \text{it has inverse}}$

Definition of singular and non-singular matrices

- A non-singular matrix is a necessarily square matrix A that has an inverse denoted by A^{-1} .
- A singular matrix is a square matrix that does not possess an inverse.

Properties of Inverse Matrices

Let $A \in C^{N \times N}$ be a non-singular matrix, then

- $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$.
- $(A^{-1})^{-1} = A$
- $(A^T)^{-1} = (A^{-1})^T$ and $(A^H)^{-1} = (A^{-1})^H$.

Let $A, B \in C^{N \times N}$ be non-singular matrices, then

- $(AB)^{-1} = B^{-1}A^{-1}$.

Basic Matrix Algebra

2.24

DETERMINANTS

The determinant only exists for square matrices. It is a scalar quantity that can provide some useful information about a matrix, for instance, about its invertability.

Definition 1

It is a recursive definition.

- If $A \in C^{1 \times 1}$, then $\det(A) = |A| = a_{11}$.
- If $A \in C^{N \times N}$, then

$$\det(A) = |A| = \sum_{n=1}^N a_{np} A_{np} = \sum_{n=1}^N a_{pn} A_{pn} \quad \forall p.$$

$$\begin{vmatrix} + & - & + & \dots \\ - & + & - & \dots \\ + & - & + & \dots \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

The scalar A_{ij} is the cofactor of a_{ij} defined as

$$A_{ij} = (-1)^{i+j} \det(M_{ij}); \quad \det(M_{ij}) \triangleq \text{minor of } a_{ij}$$

where M_{ij} is the submatrix formed by deleting the i^{th} row and the j^{th} column of A .

Definition 2

Another useful definition of the determinant of a matrix is the following:

$$\det(A) = |A| = \sum (-1)^{\phi_N(j_1, \dots, j_N)} a_{1j_1} \cdot a_{2j_2} \cdots a_{Nj_N}$$

where j_1, \dots, j_N is a permutation of the first N positive integers and the summation is over all such permutations. The function $\phi_N(j_1, \dots, j_N) = p_1 + \dots + p_N$ where p_k represents the number of integers in the subsequence j_{k+1}, \dots, j_N that are smaller than j_k . For instance, we may have

$$\phi_5(3, 5, 2, 1, 4) = 2 + 0 + 1 + 0 + 0 = 3.$$

Example

Let $A \in C^{3 \times 3}$ with a determinant

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}. \quad \left(\begin{array}{ccc} \bullet & \times & \times \\ \times & \bullet & \times \\ \times & \times & \bullet \end{array} \right) - \left(\begin{array}{ccc} \bullet & \times & \times \\ \times & \bullet & \times \\ \times & \times & \bullet \end{array} \right)$$

The determinant of A has 6 terms, equal to $3!$, which is the number of permutations of $(1, 2, 3)$.

$$\text{Permutations} = \{123, 132, 213, 231, 312, 321\}.$$

$$\begin{aligned} \det(A) &= (-1)^0 a_{11}a_{22}a_{33} + (-1)^1 a_{11}a_{23}a_{32} + (-1)^1 a_{12}a_{21}a_{33} + \\ &\quad + (-1)^2 a_{12}a_{23}a_{31} + (-1)^3 a_{13}a_{22}a_{31} + (-1)^2 a_{13}a_{21}a_{32} = \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} \end{aligned}$$

Properties of the determinant

- Product. For any pair of matrices $A, B \in C^{N \times N}$, then

$$\det(AB) = \det(A)\det(B).$$

- Transposes. For any $A \in C^{N \times N}$,

$$\det(A^T) = \det(A) \text{ but}$$

$$\det(A^H) = \det(A^*) = (\det(A))^*.$$

Multiplication by scalars.

- Let $B \in C^{N \times N}$ be a matrix formed from another matrix $A \in C^{N \times N}$ by multiplying all of the elements of one row or one column of A by the same scalar α (and leaving the other elements unchanged), then

$$\det(B) = \alpha \det(A).$$

- For any $A \in C^{N \times N}$ and any scalar $\alpha \in C$, $\det(\alpha A) = \alpha^N \det(A)$.
- If one or more rows (or columns) of a matrix $A \in C^{N \times N}$ are null, then $\det(A) = 0$.

Elementary row or column manipulations.

- If a matrix $B \in C^{N \times N}$ is formed from a matrix $A \in C^{N \times N}$ by interchanging two rows of two columns of A , then $\det(B) = -\det(A)$.
- If two rows (columns) are the same, the determinant is 0.
- If a row (column) is a linear combination of other rows (columns), the determinant is 0.
- If a multiple of one row (column) is subtracted from another row (column), the determinant remains the same.

Basic Matrix Algebra

Inverse.

- Let $A \in C^{N \times N}$ be a nonsingular matrix. Then, the inverse of A may be computed as

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)} \quad \text{with } [\text{adj}(A)]_{ij} = A_{ji}^T$$

that is, the cofactor of the transpose.

$$\begin{matrix} A = & \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \rightarrow & A^{-1} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \end{matrix} \quad 2.27$$

$$A^{-1} = \frac{1}{|A|} \begin{pmatrix} + (a_{22}a_{33}) \\ - (a_{23}a_{32}) \\ + (a_{32}a_{23}) \end{pmatrix}$$

Partitioned matrices.

- Let $A_{11} \in C^{M \times M}$ and $A_{22} \in C^{N \times N}$, then,

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{vmatrix} = |A_{11}| |A_{22}|.$$

- The determinant of an arbitrary square upper block-triangular matrix with square diagonal blocks is equal to

$$\begin{vmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ 0 & A_{22} & \cdots & A_{2N} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{NN} \end{vmatrix} = |A_{11}| |A_{22}| \cdots |A_{NN}|.$$

The same property holds for a lower block-triangular matrix.

Basic Matrix Algebra

2.28

TRACE OF A MATRIX

Definition

For a square matrix $A \in C^{M \times M}$, the trace of A is the sum of the entries of the main diagonal of A .

$$\text{tr}(A) = \sum_{i=1}^n a_{ii}$$

Properties of the trace

- $\text{tr}(AA^H) = \|A\|_F^2$
- $\text{tr}(AB) = \text{tr}(BA)$
- $a^H b = \text{tr}(ba^H) = (\text{tr}(ab^H))^*$
- $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$
- $\sum_{i=1}^L w_i^H A w_i = \text{tr}(AW) \quad \text{con } W = \sum_{i=1}^L w_i w_i^H$
- $e^{\text{tr}(A)} = \sum_{k=0}^{\infty} \frac{\text{tr}(A^{[k]})}{k!} \quad \text{where } A^{[k+1]} = A \otimes A^{[k+1]}.$
- $\text{tr}(A) = \sum_{i=1}^m \alpha_i \quad \text{where } \{\alpha_i; i=1, \dots, m\} \text{ are the eigenvalues of } A.$

KRONECKER PRODUCT

The Kronecker product has recently been used in the expression and development of some fast signal processing algorithm.

Definition

Let $A \in C^{MxP}$ and $B \in C^{NxQ}$. Then, the matrix $A \otimes B \in C^{MNxPQ}$ is defined as follows

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1P}B \\ a_{21}B & a_{22}B & \cdots & a_{2P}B \\ \vdots & \vdots & & \vdots \\ a_{N1}B & a_{N2}B & \cdots & a_{NP}B \end{bmatrix}$$

and is called the Kronecker product of A and B.

Basic properties of the Kronecker product.

- $A \otimes B \neq B \otimes A$, in general.
- For a scalar α

$$(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B).$$

- Distributive properties

$$(A + B) \otimes C = (A \otimes C) + (B \otimes C)$$

$$A \otimes (B + C) = (A \otimes B) + (A \otimes C)$$

- Associative property

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

- Transposes

$$(A \otimes B)^T = A^T \otimes B^T$$

$$(A \otimes B)^H = A^H \otimes B^H$$

} different del producto de matrices
(no suman)

- Trace (for square A and B). $MN=PQ \leftarrow$ igualadas!

$$\text{tr}(A \otimes B) = \text{tr}(A) \otimes \text{tr}(B)$$

- Determinant, for $A \in C^{MxM}$ and $B \in C^{NxN}$

$$\det(A \otimes B) = \det(A)^N \det(B)^M.$$

- The Kronecker product theorem, with $A \in C^{MxN}$, $B \in C^{PxQ}$, $C \in C^{NxR}$ and $D \in C^{QxS}$.

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

- Inverses: If A and B are nonsingular, then $A \otimes B$ is non-singular and

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

i.e. la inversa existeix

$$\text{dem. } (A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I_M \otimes I_N = I_{MN}$$

$$\underline{\underline{A}}^{(k+l)} = \underline{\underline{A}}^k \otimes \underline{\underline{A}}^l$$

Theorem

Let be $A \in \mathbb{C}^{M \times M}$ with eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$ and corresponding eigenvectors $\{x_1, x_2, \dots, x_M\}$ and let $B \in \mathbb{C}^{N \times N}$ with eigenvalues $\{\mu_1, \mu_2, \dots, \mu_N\}$ and corresponding eigenvectors $\{y_1, y_2, \dots, y_N\}$. Then the MN eigenvalues of $A \otimes B$ are $\lambda_i \mu_j$ for $i=1, \dots, M, j=1, \dots, N$ and the corresponding eigenvectors are $x_i \otimes y_j$.

[Bre78] J. W. Brewer, "Kronecker Products and Matrix Calculus in System Theory", IEEE Trans. On Circuits and Systems, Vol. CAS-25, No. 9, September 1978.

dem. pg 234, Bellman 30

$$M=N=2 \rightarrow A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\begin{array}{l} \lambda_1 x = Ax \\ \mu_1 y = By \end{array} \quad \left\{ \begin{array}{l} \lambda_1 \mu_1 x^T y^T = A x^T y^T B^T \\ \dots \end{array} \right.$$

$$\left[\begin{array}{l} \lambda_1 \mu_1 z = Cz \\ \dots \end{array} \right] \quad \text{on} \quad z = \begin{pmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \end{pmatrix} \quad \text{if } C = A \otimes B$$

VEC(-) OPERATOR

The vec operator is useful in restructuring matrix equations by turning matrices into vectors.

Definition

For an $A \in \mathbb{C}^{M \times N}$ matrix with $A = [a_1 \ a_2 \ \dots \ a_N]$, the vec operator convert the matrix to a column vector by stacking the columns of A to obtain a vector of MN elements.

$$\text{vec}(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}.$$

Relations of vec(-) operator

- Let $A \in \mathbb{C}^{R \times M}$, $Y \in \mathbb{C}^{M \times N}$ and $B \in \mathbb{C}^{N \times Q}$, then $\text{vec}(AYB) = (B^T \otimes A) \text{vec}(Y)$
- Let $A, B \in \mathbb{C}^{M \times M}$, then $\text{vec}(AB) = (\mathbf{I}_M \otimes A) \text{vec}(B) = (B^T \otimes \mathbf{I}_M) \text{vec}(A) = (B^T \otimes A) \text{vec}(\mathbf{I}_M)$

Relations with the trace

- $\text{tr}(AB) = (\text{vec}(A^T))^T \text{vec}(B)$.
- $\text{tr}(AYB) = (\text{vec}(A^T))^T (\mathbf{I}_M \otimes Y) \text{vec}(B)$
- $\text{tr}(BX^T CXD) = (\text{vec}(X))^T (B^T D^T \otimes C) \text{vec}(X) = (\text{vec}(X))^T (DB \otimes C^T) \text{vec}(X)$

A.I APPLICATIONS VECTOR SPACES AND BASIC LINEAR ALGEBRAA.I.1. PROJECTION MATRIXA.I.2. QR FACTORIZATION

\

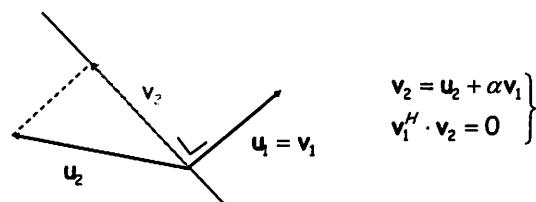
PROJECTION MATRIX

We are interested in the projection of $u \in C^M$ onto a given subspace W spanned by the set of independent vectors $\{u_1, u_2, \dots, u_L\}$. The Gram-Schmidt orthogonalization procedure provides a graphic interpretation of the projection matrices.

Conventional Gram -Schmidt Procedure

The conventional Gram-Schmidt procedure obtains an orthogonal basis, denoted by $\{v_1, v_2, \dots, v_L\}$, of a given subspace W given a non-orthogonal basis of the same subspace, denoted by $\{u_1, u_2, \dots, u_L\}$.

1. Define $v_1 = u_1$.
2. Find $v_2 \in C^M$ as the projection of $u_2 \in C^M$ onto the orthogonal subspace spanned by $\{v_1\} \in C^M$.



en quede amb la component
de u_2 q. in \perp a v_1
(i.e. v_2)

Substituting the first expression into the second, we have

$$\mathbf{v}_1^H \mathbf{u}_2 + \alpha \mathbf{v}_1^H \mathbf{v}_1 = 0 \Rightarrow \alpha = -\frac{\mathbf{v}_1^H \mathbf{u}_2}{\mathbf{v}_1^H \mathbf{v}_1}$$

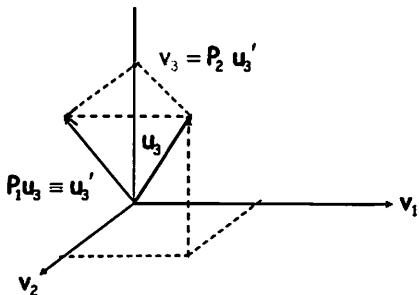
and substituting α back in the first expression we get \mathbf{v}_2 .

$$\mathbf{v}_2 = \mathbf{u}_2 - \frac{\mathbf{v}_1^H \mathbf{u}_2}{\mathbf{v}_1^H \mathbf{v}_1} \mathbf{v}_1 = \mathbf{u}_2 - \mathbf{v}_1 \frac{\mathbf{v}_1^H \mathbf{u}_2}{\mathbf{v}_1^H \mathbf{v}_1} = \underbrace{\left(\mathbf{I} - \frac{\mathbf{v}_1 \mathbf{v}_1^H}{\mathbf{v}_1^H \mathbf{v}_1} \right)}_{P_1} \mathbf{u}_2$$

Projection Matrix

- \mathbf{v}_2 is la componente de \mathbf{u}_2
- q. es orthogonal a \mathbf{v}_1
- P_1 is the matrix q. projecta \mathbf{u}_2 a l'espai ortogonal

3. Similarly, the projection of vector $\mathbf{u}_3 \in \mathbb{C}^M$ onto the orthogonal subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2\}$ is denoted by \mathbf{v}_3 and is equal to



$$\mathbf{v}_3 = P \mathbf{u}_3 = P_2 P_1 \mathbf{u}_3$$

$$P_2 = \mathbf{I} - \frac{\mathbf{v}_2 \mathbf{v}_2^H}{\mathbf{v}_2^H \mathbf{v}_2}$$

- primer projectem \mathbf{u}_3 a l'espai ortogonal de \mathbf{v}_1 (P_1) i després sobre el subespai ortogonal generat per \mathbf{v}_2 (P_2)

4. Iterating the procedure we may obtain \mathbf{v}_L as the projection of a vector $\mathbf{u}_L \in \mathbb{C}^M$ onto the orthogonal subspace spanned by the orthogonal set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{L-1}\}$ as follows

$$\boxed{\mathbf{v}_L = P \mathbf{u}_L}$$

Matrix P is defined as

$$\boxed{P = P_{L-1} P_{L-2} \cdots P_1 = \prod_{l=1}^{L-1} \left(\mathbf{I} - \frac{\mathbf{v}_l \mathbf{v}_l^H}{\mathbf{v}_l^H \mathbf{v}_l} \right) = \mathbf{I} - \sum_{l=1}^{L-1} \frac{\mathbf{v}_l \mathbf{v}_l^H}{\mathbf{v}_l^H \mathbf{v}_l}}$$

since $\mathbf{v}_l^H \mathbf{v}_m = 0 \quad \forall l \neq m$.

matrix de projecció sobre l'e.v. generat per $\mathbf{v}_1 \cdots \mathbf{v}_{L-1}$

Projection Matrices

Generalizing the Gram-Schmidt procedure, we obtain the projection of any vector $\mathbf{u} \in \mathbb{C}^M$ onto the orthogonal subspace of the subspace W spanned by the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_L\}$ as follows

$$\mathbf{u}_\perp = P \mathbf{u} = \left(\mathbf{I} - \sum_{l=1}^L \frac{\mathbf{v}_l \mathbf{v}_l^H}{\mathbf{v}_l^H \mathbf{v}_l} \right) \mathbf{u} \quad \begin{array}{l} \text{• Projecció de } \mathbf{u} \text{ sobre el} \\ \text{subespai generat per } \mathbf{v}_1 \cdots \mathbf{v}_L \end{array}$$

However, we are also interested in the projection matrix onto the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_L\}$. To find that matrix we have the following.

$$P: \text{Matrícula de projecció al subespai ortogonal a } \{\mathbf{v}_1, \dots, \mathbf{v}_L\} = \mathbf{I} - \sum_{i=1}^L \frac{\mathbf{v}_i \mathbf{v}_i^H}{\mathbf{v}_i^H \mathbf{v}_i}$$

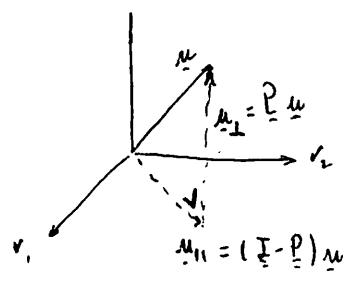
Theorem

Any vector $u \in C^M$ with $M > L$ is the sum of two components

$$u = u_{\perp} + u_{\parallel}$$

with $u_{\parallel} \in W$ and u_{\perp} a component orthogonal to W . These components are equal to

$$\begin{cases} u_{\perp} = P u = \left(I - \sum_{i=1}^L \frac{v_i v_i^H}{v_i^H v_i} \right) u \\ u_{\parallel} = u - u_{\perp} = (I - P) u = \sum_{i=1}^L \frac{v_i v_i^H}{v_i^H v_i} u \end{cases}$$



with $\{v_1, v_2, \dots, v_L\}$ an orthogonal basis of the W subspace and P the projection matrix onto the W subspace.

Both P and \tilde{P} depend on the orthogonal basis $\{v_1, v_2, \dots, v_L\}$. It would be more useful an expression of the projection matrices directly based on the non-orthogonal set $\{u_1, u_2, \dots, u_L\}$.

Since $\{v_1, v_2, \dots, v_L\}$ is an orthogonal set, matrix \tilde{P} can be expressed as

vectors columnas \rightarrow

$$\tilde{P} = \sum_{i=1}^L \frac{v_i v_i^H}{v_i^H v_i} =$$

$$= [v_1 \ v_2 \ \dots \ v_L] \begin{bmatrix} (v_1^H v_1)^{-1} & 0 & \dots & 0 \\ 0 & (v_2^H v_2)^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & (v_L^H v_L)^{-1} \end{bmatrix} \begin{bmatrix} v_1^H \\ v_2^H \\ \vdots \\ v_L^H \end{bmatrix} = V(V^H V)^{-1} V^H$$

Matriz de Proyección

As $\{v_1, v_2, \dots, v_L\}$ is a basis of W , we have

$$u_k = \sum_{i=1}^L \alpha_{ik} v_i \quad \forall k = 1, \dots, L$$

where $\{\alpha_{ik}\}$ are unique and $\alpha_{ik} = \frac{v_i^H u_k}{v_i^H v_i}$. According to that,

escalar

$$U = [u_1 \ u_2 \ \dots \ u_L] = \begin{bmatrix} u_{11} & u_{21} & \dots & u_{L1} \\ u_{12} & u_{22} & \dots & u_{L2} \\ \vdots & \vdots & & \vdots \\ u_{1L} & u_{2L} & \dots & u_{LL} \end{bmatrix} =$$

in general, no es cuadrada

$$= \begin{bmatrix} v_{11} & v_{21} & \dots & v_{L1} \\ v_{12} & v_{22} & \dots & v_{L2} \\ \vdots & \vdots & & \vdots \\ v_{1L} & v_{2L} & \dots & v_{LL} \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{L1} \\ \alpha_{12} & \alpha_{22} & \dots & \alpha_{L2} \\ \vdots & \vdots & & \vdots \\ \alpha_{1L} & \alpha_{2L} & \dots & \alpha_{LL} \end{bmatrix} = V \cdot A.$$

escalar

A is no singular \rightarrow invertible

Since $\mathbf{U} = \mathbf{V} \cdot \mathbf{A}$ with \mathbf{V} and \mathbf{A} non-singular matrices, $\mathbf{V} = \mathbf{U} \cdot \mathbf{A}^{-1}$. Substituting this into $\tilde{\mathbf{P}}$, we obtain

$$\begin{aligned}\tilde{\mathbf{P}} &= \mathbf{U}\mathbf{A}^{-1}(\mathbf{A}^{-1H}\mathbf{U}^H\mathbf{U}\mathbf{A}^{-1})^{-1}\mathbf{A}^{-1H}\mathbf{U}^H = \mathbf{U}\mathbf{A}^{-1}\mathbf{A}(\mathbf{U}^H\mathbf{U})^{-1}\mathbf{A}^H\mathbf{A}^{-1H}\mathbf{U}^H \\ \boxed{\tilde{\mathbf{P}} = \mathbf{U}(\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H} &\quad \leftarrow \text{Matriz de proyección}\end{aligned}$$

The expressions of the projection matrices remain the same even with an non-orthogonal set of independent vectors $\{u_1, u_2, \dots, u_L\}$!!!

Orthogonal Decomposition

Let's assume the vector space of all the vectors $\mathbf{u} \in \mathbb{C}^M$. Let us also assume $\{u_1, u_2, \dots, u_L\}$ is a basis of the vector subspace W with dimension L and $\{u_{L+1}, u_{L+2}, \dots, u_M\}$ is a basis of the complementary vector subspace \hat{W} of dimension $M-L$.

If we define matrices \mathbf{U} and $\hat{\mathbf{U}}$ as follows

$$\mathbf{U} = [u_1 \ u_2 \ \dots \ u_L] \text{ and } \hat{\mathbf{U}} = [u_{L+1} \ u_{L+2} \ \dots \ u_M]$$

such that

$$\mathbf{U}^H\hat{\mathbf{U}} = \mathbf{0}, \quad \text{nón ortogonals}$$

then any vector $\mathbf{u} \in \mathbb{C}^M$ can be decomposed in two components,

$$\mathbf{u} = \tilde{\mathbf{P}}_W \mathbf{u} + \tilde{\mathbf{P}}_{\hat{W}} \mathbf{u}.$$

Matrices $\tilde{\mathbf{P}}_W$ and $\tilde{\mathbf{P}}_{\hat{W}}$ are projection matrices onto the subspaces W and \hat{W} , respectively.

$$\tilde{\mathbf{P}}_W = \mathbf{U}(\mathbf{U}^H\mathbf{U})^{-1}\mathbf{U}^H \text{ and } \tilde{\mathbf{P}}_{\hat{W}} = \hat{\mathbf{U}}(\hat{\mathbf{U}}^H\hat{\mathbf{U}})^{-1}\hat{\mathbf{U}}^H$$

They have the following properties:

1. $\tilde{\mathbf{P}}_W$ and $\tilde{\mathbf{P}}_{\hat{W}}$ are hermitian. (i.e. $\tilde{\mathbf{P}}^H = \tilde{\mathbf{P}}$)
2. $\tilde{\mathbf{P}}_W$ and $\tilde{\mathbf{P}}_{\hat{W}}$ are idempotent. (i.e. $\tilde{\mathbf{P}} \cdot \tilde{\mathbf{P}} = \tilde{\mathbf{P}}$)
3. $\tilde{\mathbf{P}}_W$ and $\tilde{\mathbf{P}}_{\hat{W}}$ are orthogonal. (i.e. $\tilde{\mathbf{P}}_W^H \cdot \tilde{\mathbf{P}}_{\hat{W}} = \mathbf{0}$)

QR DECOMPOSITIONNormalized Gram-Schmidt procedure

For convenience, we now formulate the problem of the Gram-Schmidt procedure in a more general way. Let's assume a set of vectors $\{p_1, p_2, \dots, p_N\} \in C^M$ (not necessarily independent!); we want to find a set of vectors $\{q_1, q_2, \dots, q_L\}$ with $L \leq N$ so that

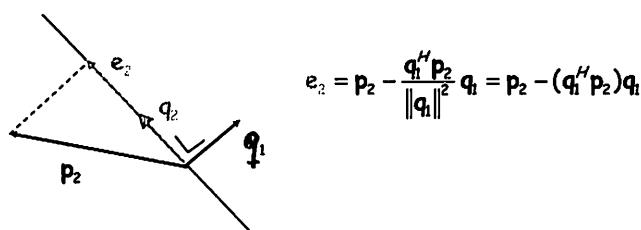
$$\text{span}\{p_1, p_2, \dots, p_N\} = \text{span}\{q_1, q_2, \dots, q_L\}$$

and

$$q_i^H q_j = \delta_{i,j} \rightarrow \text{independents.}$$

To solve this problem we propose the normalized Gram-Schmidt procedure, which is very similar to the conventional one except that it normalizes the vector q_i at each step. The process will be developed stepwise and the norm of a vector u is defined as $\|u\| = \langle u, u \rangle^{1/2}$.

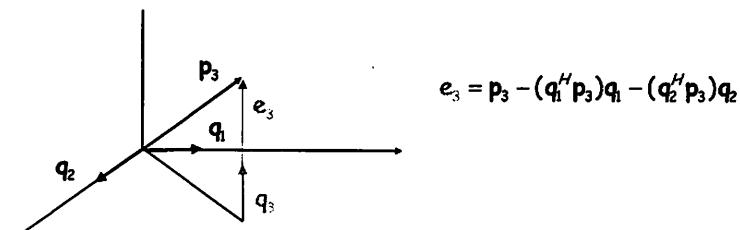
1. Normalize the first vector $q_1 = \frac{p_1}{\|p_1\|}$.
2. Compute the difference between p_2 and the projection of p_2 onto q_1 . The resulting vector e_2 is orthogonal to p_1 .



If $e_2 = 0$, then $q_2 \in \text{span}\{q_1\}$ and can be discarded; we will assume that such discards are done as necessary in what follows. If $e_2 \neq 0$, then normalize

$$q_2 = \frac{e_2}{\|e_2\|}.$$

3. At the next stage, a vector orthogonal to q_1 and q_2 is obtained from the error between p_3 and its projection onto $\text{span}\{q_1, q_2\}$.



Vector e_3 is normalized to produce q_3

$$q_3 = \frac{e_3}{\|e_3\|}.$$

4. Now proceed inductively. To form the next orthogonal vector using p_k , determine the component orthogonal to all found vectors

$$\left[e_k = p_k - \sum_{i=1}^{k-1} (q_i^H p_k) q_i \right] \quad (1)$$

and normalize

$$\left[q_k = \frac{e_k}{\|e_k\|} \right] \quad (2)$$

This Gram-Schmidt process can be represented in a matrix form. Let $A = [p_1 \ p_2 \ \dots \ p_N] \in C^{M \times N}$. The orthogonal vectors obtained by the Gram-Schmidt process are stacked in a matrix $Q = [q_1 \ q_2 \ \dots \ q_L] \in C^{M \times L}$, with $L \leq N$ because e_k could be equal to 0 for some values of k . Therefore in that case, and without loss of generality, we may consider that the first L vectors $\{p_1, p_2, \dots, p_L\}$ span the same subspace than the whole set $\{p_1, p_2, \dots, p_N\}$. We define the upper triangular matrix R with the inner products and norms of (1) and (2) as follows

$$\begin{array}{c} R \\ \boxed{\begin{matrix} \|p_1\| & (q_1^H p_2) & (q_1^H p_3) & \cdots & \cdots & (q_1^H p_N) \\ \|e_2\| & (q_2^H p_3) & \cdots & \cdots & (q_2^H p_N) \\ \vdots & & & & & \\ \|e_L\| & (q_L^H p_3) & \cdots & (q_L^H p_N) \\ 0 & \cdots & 0 \\ & \ddots & \vdots \\ & & 0 \end{matrix}} \end{array}$$

Then, the matrix formulation of the normalized Gram-Schmidt process is

$$A = [p_1 \ p_2 \ \dots \ p_N] = \left[q_1 \ q_2 \ \dots \ q_L \underbrace{0 \ \dots \ 0}_{M-L} \right] R = QR$$

$\uparrow \quad \nwarrow \quad M \times L$

which is known as QR factorization of matrix A .

$$[p_1 \ p_2 \ \dots \ p_N] = [q_1 \ \|q_1\| \quad q_2 \ \|q_2\| + q_1 (q_1^H p_2) \ \dots]$$

The QR factorization

The QR factorization arises commonly in matrix analysis and signal processing routines. It becomes useful

- to solve least squares problems.
- to compute eigenvectors and eigenvalues
- to obtain other matrix factorizations.

QR Decomposition

A general matrix A can be factored as

$$A = QR$$

where Q is a unitary matrix, $QQ^H = I$, and R is upper triangular.

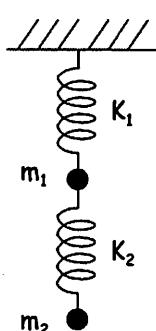
So far we have presented the normalized Gram-Schmidt algorithm as a way of computing the QR factorization. Other methods are based on the so-called Householder transformation and Givens rotations. The Gram-Schmidt method is the least complex computationally, but is also the most poorly conditioned.

3. EIGENVALUES AND EIGENVECTORS

- 3.1. DEFINITION, AND BASIC PROPERTIES.
- 3.2. DIAGONALIZATION. SIMILARITY TRANSFORMATIONS. LEFT EIGENVECTORS
- 3.3. DEFECTIVE MATRICES. JORDAN FORM.
- 3.4. SCHUR DECOMPOSITION.
- 3.5. NORMAL MATRICES: DIAGONALIZATION. EIGENVECTORS AND EIGENVALUES OF THE CORRELATION MATRIX.

The main reason for studying the so-called *eigenvalues* and *eigenvectors* of matrices is that these basic concepts arise when models are studied from a wide variety of viewpoints. These are some examples of applied problems where the eigenvalues and eigenvectors become fundamental.

- Singularity of $(A - \lambda I)$ where λ is a parameter.



This problem arises in modelling oscillatory phenomena such as the motion of airplane wings under various aerodynamic forces or the swings in the economy under various market forces.

Example: Modelling the vertical oscillations of two masses suspended and coupled by springs.

This system oscillates when the matrix $(K - \omega^2 M)$ is singular, where K and M are known matrices and ω is unknown.

More generally, if M is non-singular, the problem is equivalent to finding a parameter ω for which $(A - \lambda I)$ is singular, where $A = M^{-1}K$ and $\lambda = \omega^2$.

- Invariant Subspaces.

Mathematical models of applied problems often lead to the study of linear transformations defined by matrices, i.e. $T(x) = Ax$. In real situations, it is very often the case that the dimension of A is extremely large. In such cases, it is useful to break the large system into a collection of much smaller subsystems that can be dealt with more easily. Algebraically, this corresponds to the problem of finding low-dimensional subspaces W so that Ax is also in W . As we will see, the eigenvalues and eigenvectors of those matrices become involved in this problem.

Eigenvalues and Eigenvectors

3.3

• Simple Representations of Transformations.

Another approach to simplifying the study of linear transformations is to represent A by a simpler matrix that is easier to study. For instance,

$$A' = Q^{-1}AQ$$

with A' a diagonal or triangular matrix.

• Decompositions of matrices.

Matrix A can be decomposed into the product

$$A = QA'Q^{-1}$$

with A' having a special form. This decomposition can be extremely useful, for instance, in analysing the behaviour of large powers of A' .

Eigenvalues and Eigenvectors

3.4

3.1. DEFINITIONS AND BASIC PROPERTIES

Definition of eigenvector and eigenvalue

The *eigenvector* of $A \in C^{M \times M}$ are the nonzero vectors, that is, $x \neq 0$, for which there is a number λ with

$$Ax = \lambda x. \quad (3.1)$$

The scalar $\lambda \in C$ is the *eigenvalue* associated with the eigenvector x , and vice versa.

Equation (3.1) can be written as

$$(A - \lambda I)x = 0.$$

This problem has a nonzero solution $x \neq 0$ if and only if $(A - \lambda I)$ is singular, which is equivalent to

$$(A - \lambda I) \text{ is singular} \Leftrightarrow \det(A - \lambda I) = 0.$$

Definition of characteristic polynomial

The $\det(A - \lambda I)$ is a polynomial of degree M in the variable λ , denoted by $P_M(\lambda)$ and called *characteristic polynomial* of A .

$$\det(A - \lambda I) = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2M} \\ \vdots & \ddots & \ddots & \vdots \\ a_{M1} & a_{M2} & \cdots & a_{MM} - \lambda \end{bmatrix} = P_M(\lambda)$$

Eigenvalues and Eigenvectors

Therefore, the eigenvalues of matrix $A \in \mathbb{C}^{M \times M}$ are the roots of the polynomial $P_M(\lambda)$. From the theory of polynomials we know that such a polynomial has M real or complex roots $\{\lambda_1, \dots, \lambda_M\}$. Then, we can factor the polynomial as

$$\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_M - \lambda).$$

Note that λ_m need not be distinct!

Basic properties of the eigenvalues

- The coefficient λ^M in $P_M(\lambda)$ equals $(-1)^M$.
- The constant term of $P_M(\lambda)$ equals $\det(A)$.

$$P_M(\lambda)|_{\lambda=0} = \det(A) = \lambda_1 \lambda_2 \cdots \lambda_M.$$

- The coefficient λ^{M-1} in $P_M(\lambda)$ equals

$$(-1)^M \sum_{m=1}^M \lambda_m = (-1)^M \sum_{m=1}^M a_{mm}.$$

Therefore,

$$\sum_{m=1}^M \lambda_m = \text{tr}(A).$$

*Eigenvalues and Eigenvectors**Example.*

Consider the 5×5 matrix

$$A = \begin{bmatrix} 7 & 0 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

The characteristic polynomial of A is $(7-\lambda)^2(4-\lambda)^3$, so that the eigenvalues of A are $\lambda_1=7$ with algebraic multiplicity 2 and $\lambda_2=4$ of algebraic multiplicity 3. We consider the eigenvectors associated with these eigenvalues.

- $\lambda_1=7$. Consider the equation $(A - 7I)x = 0$ that we must solve to find the eigenvector x associated with $\lambda_1=7$. This system is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

with solution

$$x = [\alpha \ 0 \ 0 \ \beta \ 0]^T = \alpha e_1^T + \beta e_4^T \quad \forall \alpha, \beta.$$

Eigenvalues and Eigenvectors

3.7

Therefore, e_1 and e_4 are eigenvectors associated with $\lambda_1 = 7$, and all other eigenvectors are linearly dependent on e_1 and e_4 . We have found two linearly independent eigenvectors associated with this eigenvalue of algebraic multiplicity two.

- $\lambda_2 = 4$. Similarly, we have to solve the system $(A - 4I)x = 0$ equal to

$$\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The solution of this system is

$$x = [0 \ \alpha \ 0 \ 0 \ \beta]^T = \alpha e_2^T + \beta e_5^T \quad \forall \alpha, \beta$$

as the set of all the eigenvectors associated with $\lambda_2 = 4$. Therefore, e_2 and e_5 are eigenvectors associated with $\lambda_2 = 4$, and all other eigenvectors are linearly dependent on e_2 and e_5 . We have found only two linearly independent eigenvectors associated with this eigenvalue of algebraic multiplicity three.

We cannot find a third eigenvector associated with $\lambda_2 = 4$ that is linearly independent of the two already found. Note, however, that we can find a third eigenvector, let say e_3 , so that the set $\{e_3, e_2, e_5\}$ form a basis for an invariant subspace of A of dimension 3.

$$A(\alpha e_3 + \beta e_2 + \gamma e_5) = 4\alpha e_3 + (4\beta + \alpha)e_2 + 4\gamma e_5.$$

Eigenvalues and Eigenvectors

3.8

The situation illustrated in this example is typical: an eigenvalue λ_i of algebraic multiplicity m_i may or may not have a linearly independent set of m_i associated eigenvectors, although it will always have an invariant subspace of dimension m_i .

□

Theorems

1. There exists at least one eigenvector x_i associated with each distinct eigenvalue λ_i .
2. If $\{\lambda_1, \dots, \lambda_s\}$ is a collection of distinct eigenvalues, and if x_m is the eigenvector associated with λ_m for each m , then $\{x_1, \dots, x_s\}$ is linearly independent.
3. The set of all the eigenvectors associated with the same eigenvalue forms an invariant subspace of A . Let assume λ_m is an eigenvalue with associated eigenvectors $\{x_1, \dots, x_s\}$. Then, for any $y \in \text{span}(x_1, \dots, x_s)$ we have

$$Ay = A \sum_{i=1}^s \alpha_i x_i = \sum_{i=1}^s \alpha_i Ax_i = \lambda_m \sum_{i=1}^s \alpha_i x_i = \lambda_m y.$$

4. If a matrix $A \in \mathbb{C}^{M \times M}$ has M distinct eigenvalues, then any set of M eigenvectors - one of which associated with each eigenvalue - is linearly independent and is a basis of vector space \mathbb{C}^M .

3.2. DIAGONALIZATION AND SIMILARITY TRANSFORMATIONSExample

Consider the matrix

$$A = \begin{bmatrix} 0.8 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.1 & 0.6 \end{bmatrix}$$

whose eigenvalues are $\lambda_1 = 0.5$, $\lambda_2 = 0.6$ and $\lambda_3 = 1$, and the respective associated eigenvectors are

$$x_1 = [1 \ -2 \ 1]^T, x_2 = [1 \ -1 \ 0]^T \text{ and } x_3 = [9 \ 7 \ 4]^T.$$

These eigenvectors are linearly independent. We can write the equations $Ax_i = \lambda_i x_i$ in the block form

$$A[x_1 \ x_2 \ x_3] = [\lambda_1 x_1 \ \lambda_2 x_2 \ \lambda_3 x_3],$$

or equivalently,

$$AQ = Q\Lambda \Rightarrow A = Q\Lambda Q^{-1}$$

with matrix Q and matrix Λ defined as follows

$$Q = [x_1 \ x_2 \ x_3] \text{ and } \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3).$$

□

Theorem

A matrix $A \in C^{M \times M}$ has linearly independent set of M eigenvectors if and only if there exists a non-singular matrix $Q \in C^{M \times M}$ and a diagonal matrix $\Lambda \in C^{M \times M}$ for which

$$A = Q\Lambda Q^{-1}. \quad (3.2)$$

The columns of Q are a set of linearly independent vectors of A and the elements of the diagonal of Λ are the eigenvalues.

$$\underbrace{A[x_1 \ \dots \ x_M]}_Q = [x_1 \ \dots \ x_M] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M \end{bmatrix} \Downarrow \begin{array}{l} \text{de } (3) \text{ definic} \\ \text{d'autovectores} \end{array}$$

$$[Ax_1 \ \dots \ Ax_M] = [\lambda_1 x_1 \ \dots \ \lambda_M x_M]$$

Moreover, if $\{x_1, \dots, x_M\}$ are orthonormal, then matrix Q is unitary, that is, $Q^{-1} = Q^H$, and we have

$$\|A\|_2 = \|Q^{-1}A\|_2 = \|AQ\|_2 = \|Q^{-1}AQ\|_2.$$

$$\text{ex: } \tilde{A}^{-1} = (\tilde{Q}\Lambda\tilde{Q}^{-1})^{-1} = (\tilde{Q}^{-1})^{-1}(\tilde{Q}\Lambda)^{-1} = \tilde{Q}^{-1}\tilde{\Lambda}^{-1}\tilde{Q} \rightarrow \tilde{A}^{-1} \text{ es la } \tilde{\Lambda}^{-1}$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

Definition of Similarity Transformations

Matrix B is similar to matrix A if there exists a non-singular matrix P such that $P^{-1}AP = B$.

Eigenvalues and Eigenvectors of Similarity Transformations

Similar matrices have the same eigenvalues, and the eigenvectors are related as follows:

$$Ax = \lambda x \Rightarrow APP^{-1}x = \lambda x \Rightarrow \underbrace{P^{-1}AP}_{B} \underbrace{P^{-1}x}_{y} = \lambda \underbrace{P^{-1}x}_{y}$$

$$y = P^{-1}x.$$

LEFT-EIGENVECTORS

Since matrix Q of equation (3.2) is non-singular, we also have

$$Q^{-1}A = \Lambda Q^{-1}.$$

Let y_m^H denote the m^{th} row of Q^{-1} , so that

$$Q^{-1} = \begin{bmatrix} y_1^H \\ \vdots \\ y_m^H \\ \vdots \\ y_n^H \end{bmatrix} \Rightarrow y_m^H A = \lambda_m y_m^H.$$

Definition

A nonzero vector y for which $y^H A = \lambda y^H$ is called a left-eigenvector associated with the eigenvalue λ of A . (Note that the eigenvalues associated with the left and "right" eigenvalues remain the same).

Properties

- The number of left-eigenvectors and right-eigenvectors are the same.
- The eigenvalues of A and A^T are the same. d.m.: $\det(A - \lambda I) = \det(A^T - \lambda I^T) = 0$
- The left-eigenvectors of A are equal to the right-eigenvectors of A^H .
- The eigenvalues of A^H are the complex conjugate of the eigenvalues of A .
- The left-eigenvectors of A are the complex conjugate of the right-eigenvectors of A^T .

$$\text{d.m.: } (y^H A)^T = (\lambda y^H)^T \rightarrow A^T y^* = \lambda y^*$$

$$(y^H A)^H = (\lambda y^H)^H \rightarrow A^H y = \lambda^* y$$

\downarrow
 pu vu
 weiter vobr
 de λ

Theorem

Let x_i denote the right-eigenvector associated with the eigenvalue λ_i and let y_j the left-eigenvector associated with the same eigenvalue, then

$$y_j^H x_i = \delta_{ij}.$$

Note: These properties of the left and right eigenvalues still hold for matrices $A \in C^{M \times M}$ with less than M linearly independent eigenvectors.

i.e. se x_i e y_j estan asociados a un vector λ $\rightarrow y_j^H x_i = \delta_{ij}$

$$\text{dem: } y_j^H x_i = \frac{1}{\lambda_i} y_j^H \underbrace{\lambda_i x_i}_{= x_i} = \frac{1}{\lambda_i} y_j^H \underbrace{A x_i}_{= x_i} *$$

$$y_j^H x_i = \frac{\lambda_i y_j^H x_i}{\lambda_i} \Leftrightarrow \lambda_j = \lambda_i$$

3.3. DEFECTIVE MATRICES.

So far, we have seen that a matrix $A \in C^{M \times M}$ is similar to a diagonal matrix Λ , that is, A can be decomposed as $A = Q \Lambda Q^{-1}$, if and only if A has a linearly independent set of M eigenvectors. What remains to study is the more complicated case of matrices with at least one eigenvalue λ_i with algebraic multiplicity m_i and less than m_i linearly independent associated eigenvectors. These matrices are named defective matrices.

Definition

A matrix that fails to have a linearly independent set of M eigenvectors is said to be defective. Otherwise, it is nondefective.

Note that the matrices $M \times M$ with M distinct eigenvalues are nondefective!

i.e. que un matriz tiene repetidos

$$\text{se } A \text{ es non-defectivo} \rightarrow A = Q \Lambda Q^{-1}$$

↑
Diagonal

Example

Consider the matrix

$$A = \begin{bmatrix} 2 & \alpha & 0 & 0 \\ 0 & 2 & \beta & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

with $\lambda_1=2$ of algebraic multiplicity 3 and $\lambda_2=3$ with associated eigenvector e_4 . Regarding the eigenvectors associated with $\lambda_1=2$, we have different cases depending on the values of $\{\alpha, \beta\}$.

- $\alpha = \beta = 0$. Three linearly independent vectors, i.e. $\{e_1, e_2, e_3\}$, associated with $\lambda_1=2$.

In this case, $\mathfrak{N}(A)$ is decomposed in 4 invariant subspaces, i.e., $\{\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}\}$.

to operations to partayer de subesp.

- $\alpha = 0; \beta \neq 0$. Two eigenvectors, i.e. e_1 and e_2 , associated with $\lambda_1=2$. The Euclidean vector e_3 is not an eigenvector. However, since

$$Ae_3 = 2e_3 + \beta e_2 \Rightarrow A(\alpha e_2 + \beta e_3) = (2\alpha + \beta\beta)e_2 + 2\beta e_3$$

the set $\{e_2, e_3\}$ span an invariant subspace of dimension 2. In this case, $\mathfrak{N}(A)$ is decomposed in three invariant subspaces, i.e., $\{\{e_1\}, \{e_2, e_3\}, \{e_4\}\}$.

- $\alpha \neq 0; \beta \neq 0$. Only one eigenvector associated with $\lambda_1=2$. Similarly, we obtain the invariant subspaces $\{\{e_1, e_2, e_3\}, \{e_4\}\}$. \square

JORDAN FORM

We develop a special form to which each matrix can be reduced by a similarity transformation that exploit the structure of the eigensystems of defective and nondefective matrices.

Theorem

A matrix $A \in \mathbb{C}^{M \times M}$ is similar to a matrix J in Jordan form: $J = Q^{-1}AQ$ and $A = QJQ^{-1}$ with

$$J = Q^{-1}AQ = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & J_L \end{bmatrix}$$

where each J_i is a $M_i \times M_i$ Jordan block

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ 0 & \vdots & \ddots & \lambda_i & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}$$

and $M_1 + M_2 + \cdots + M_L = M$.

L subespacios invariantes de A

ex:
$$\begin{bmatrix} J_1 & J_2 \\ J_2 & 0 \end{bmatrix} \quad \begin{bmatrix} J_2 & 0 \\ 0 & J_3 \end{bmatrix}$$

- Each block has one eigenvalue.
- The dimension of each Jordan block is unique although the Jordan blocks could be ordered differently.
- The same distinct eigenvalue may occur in different Jordan blocks J_i , but each Jordan block represents a different eigenvector and the total number of blocks with the same eigenvalue equals the number of linearly independent eigenvectors associated to that eigenvalue.

Computation of the Jordan Form

The previous theorem is an existence theorem, although the special form of J enables us to compute it. Since $Q^{-1}AQ = J$, we have $AQ = QJ$. Writing Q in terms of its columns as

$$Q = [q_1 \ q_2 \ \dots \ q_M]$$

shows that $AQ = QJ$ is equivalent to

$$Aq_i = \lambda q_i + v_i q_{i-1} \quad \text{for } i = 1, 2, \dots, M \quad (3.3)$$

where λ is the eigenvalue in the Jordan block affecting q_i and v_i equals 1 or 0. Thus we may obtain matrix Q from the resulting M equations, and knowing the structure of the Jordan form J .

ex:

$$\left(\begin{array}{cccccc} \lambda_1 & 1 & 0 & & & \\ 0 & \lambda_1 & & & & \\ 0 & 0 & \lambda_1 & & & \\ & & & \ddots & & \\ & & & & \lambda_2 & 1 \\ & & & & 0 & \lambda_2 \end{array} \right) \quad \begin{aligned} (1) \quad Aq_i &= \lambda q_i \\ (2) \quad Aq_i &= \lambda q_i + 1 q_{i-1} \end{aligned}$$

$\uparrow \uparrow$
 $i \quad i-1$

Example

Consider a matrix $A \in \mathbb{C}^{6 \times 6}$ with the following Jordan form

$$J = \begin{bmatrix} J_1 & 0 & 0 \\ 0 & J_2 & 0 \\ 0 & 0 & J_3 \end{bmatrix}$$

$$QJ = [q_1 \ \dots \ q_6] \begin{bmatrix} \alpha & 1 & 0 & & & \\ 0 & \alpha & 1 & & & \\ 0 & 0 & \alpha & & & \\ & & & \alpha & & \\ & & & 0 & \beta & 1 \\ & & & & 0 & \beta \end{bmatrix}$$

with

$$J_1 = \begin{bmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix}, \quad J_2 = [\alpha], \quad J_3 = \begin{bmatrix} \beta & 1 \\ 0 & \beta \end{bmatrix}$$

$$[\alpha q_1, \alpha q_2 + q_1, \alpha q_3 + q_2, q_4, \beta q_5, \beta q_6 + q_5]$$

The set of equations following (3.3) are

$$\begin{aligned} Aq_1 &= \alpha q_1 \rightarrow (A - \alpha I)q_1 = 0 \quad i=1 \\ Aq_2 &= \alpha q_2 + q_1 \rightarrow (A - \alpha I)q_2 = q_1 \quad i=2 \\ &\quad (A - \alpha I)q_3 = q_2 \quad i=3 \\ (A - \alpha I)q_4 &= 0 \quad i=4 \\ (A - \beta I)q_5 &= 0 \quad i=5 \\ (A - \beta I)q_6 &= q_5 \quad i=6 \end{aligned}$$

where $\{q_1, q_4, q_5\}$ are eigenvectors of A and $\{q_2, q_3, q_6\}$ are the so-called generalized eigenvectors.

□

3.4. SCHUR DECOMPOSITION

In this section we aim at studying the specific structure of the eigenvalues and eigenvectors of "symmetric" matrices that often arise in mathematical models of real world. First, we present a simple form to which any square matrix can be transformed.

Theorem

Any matrix $A \in C^{M \times M}$ is unitarily similar to an upper-triangular matrix R as follows

$$Q^H A Q = R \quad R = \begin{bmatrix} \text{Upper Triangular} \\ 0 \end{bmatrix}$$

with Q unitary and with the eigenvalues of A (repeated according to their algebraic multiplicities) on the main diagonal of R . If A and its eigenvalues are real, then R may be taken real and hence orthogonal.

Proof

Both parts of the theorem are true if $M=1$. We proceed by induction, assuming the theorem is true for $M=k$ and seeking to prove it for $M=k+1$.

So suppose that $A \in C^{(k+1) \times (k+1)}$. Let λ_1 be an eigenvalue of A with associated eigenvector x_1 normalized so

Eigenvalues and Eigenvectors

3.20

that $\|x_1\|_2 = 1$. There is a set of vectors $\{w_1, \dots, w_k\}$ such that $\{x_1, w_1, \dots, w_k\}$ are orthonormal. Thus the matrix

$$Q = \begin{bmatrix} x_1 & \underbrace{w_1 \cdots w_k}_W \end{bmatrix} = [x_1 \ W]$$

is unitary. We now compute

$$\begin{aligned} Q^H A Q &= [x_1 \ W]^H A [x_1 \ W] = [x_1 \ W]^H [Ax_1 \ AW] \\ &= [x_1 \ W]^H [\lambda_1 x_1 \ AW] \\ &= \begin{bmatrix} \lambda_1 & x_1^H AW \\ 0 & W^H AW \end{bmatrix} = \begin{bmatrix} \lambda_1 & b^H \\ 0 & C \end{bmatrix} \end{aligned}$$

since $\|x_1\|_2 = 1$ and $x_1^H W = 0$ because Q is unitary. Since matrix C is $k \times k$, we can find a unitary matrix V so that $V^H C V$ is upper triangular. If we now define the unitary matrix U as

$$U = \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix},$$

then

$$R = U^H (Q^H A Q) U = \begin{bmatrix} 1 & 0 \\ 0 & V^H \end{bmatrix} \begin{bmatrix} \lambda_1 & b^H \\ 0 & C \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & V \end{bmatrix} = \begin{bmatrix} \lambda_1 & b^H V \\ 0 & V^H C V \end{bmatrix}.$$

Matrix R is upper triangular with the eigenvalues in the main diagonal (note that R and A are similar matrices because the product $U^H Q^H$ is unitary).

□

Note: This is an existence theorem, but does not say that R is easily computed; in practice, one needs to know the eigenvalues to obtain R . The Schur decomposition is not unique!

3.5. NORMAL MATRICES: DIAGONALIZATION

Definition

A normal matrix is a $M \times M$ matrix A satisfying $A^H A = A A^H$.

Examples of normal matrices: symmetric matrices, hermitian matrices, unitary matrices, skew-symmetric matrices $A^H = -A$.

□

Theorem: Schur decomposition of normal matrices

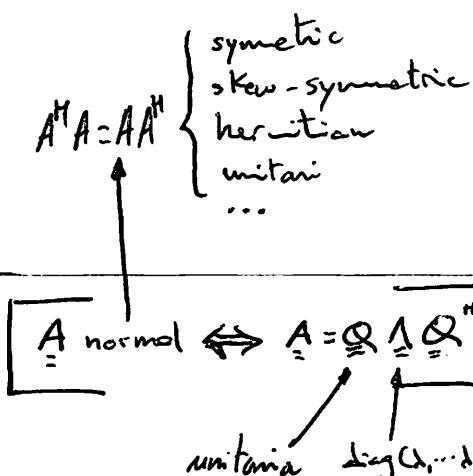
A matrix $A \in \mathbb{C}^{M \times M}$ is normal if and only if A is unitarily similar to a diagonal (Schur form) matrix $\Lambda = Q^H A Q$, where Q is unitary and Λ is diagonal with the eigenvalues of A on the main diagonal (repeated according to their algebraic multiplicities).

Note: Only the normal matrices are unitarily similar to a diagonal matrix!

Proof

(diagonal form \rightarrow normal). If there exists a unitary Q with $\Lambda = Q^H A Q$ diagonal, then A is certainly normal.

$$\begin{aligned} A^H A &= (Q \Lambda Q^H)^H (Q \Lambda Q^H) = Q \Lambda^H Q^H Q \Lambda Q^H = Q \Lambda^H \Lambda Q^H = Q \Lambda \Lambda^H Q^H \\ &= Q \Lambda Q^H Q \Lambda^H Q^H = (Q \Lambda Q^H)(Q \Lambda Q^H)^H = A A^H \end{aligned}$$



Eigenvalues and Eigenvectors

3.23

(normal \rightarrow diagonal form). Suppose that A is normal, then R the Schur form of A is also normal

$$\begin{aligned} R^H R &= (Q^H A Q)^H (Q^H A Q) = Q^H A^H A Q \stackrel{A \text{ is normal}}{=} Q^H A A^H Q = \\ &= Q^H A Q Q^H A^H Q = R R^H \end{aligned}$$

Since R is both upper-triangular and normal, then it is also diagonal.

The decomposition is unique!

□

Corollaries

- The set of eigenvectors of a normal matrix is an orthogonal basis of the C^M vector space.
- All the normal matrices (and only them) are unitarily similar to a diagonal matrix.
- The right-eigenvectors and ^{right}_{left}-eigenvectors of a normal matrix are equal

Theorem

The eigenvalues of a hermitian matrix $A \in C^{M \times M}$ are real.

Proof

Since $A \in C^{M \times M}$ is hermitian, it is normal and we have

Eigenvalues and Eigenvectors

3.24

$$\Lambda = Q^H A Q \stackrel{A \text{ hermitian}}{=} Q A^H Q^H = \Lambda^H.$$

Thus Λ is real.

□

Theorem

If $A \in C^{M \times M}$ is normal and λ is an eigenvalue associated with the eigenvector x of A , then λ^* is the eigenvalue associated with the eigenvector x of A^H .

Proof

If $A \in C^{M \times M}$ is normal, we have

$$\Lambda = Q^H A Q \Rightarrow \Lambda^H = Q A^H Q^H.$$

and since Λ is diagonal, $\Lambda^H = \Lambda^*$.

□

SPECTRAL DECOMPOSITION

If $A \in \mathbb{C}^{M \times M}$ is normal, we have

$$\begin{aligned} A &= Q \Lambda Q^H \\ A &= [q_1 \ \dots \ q_M] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M \end{bmatrix} \begin{bmatrix} q_1^H \\ \vdots \\ q_M^H \end{bmatrix} = \\ &= [\lambda_1 q_1 \ \dots \ \lambda_M q_M] \begin{bmatrix} q_1^H \\ \vdots \\ q_M^H \end{bmatrix} = \sum_{i=1}^M \lambda_i q_i q_i^H \end{aligned}$$

where $q_i q_i^H = P_i$ is the projection matrix onto the subspace spanned by $\{q_i\}$ with $\text{rank}(P_i)=1$.

$$P_i = q_i q_i^H = \frac{q_i q_i^H}{q_i^H q_i}$$

Then, we define the Spectral Decomposition of matrix A as follows

$$A = \sum_{i=1}^M \lambda_i P_i \quad \text{where} \quad P_i = q_i q_i^H$$

APPLICATION: ENGINVECTORS AND EIGENVALUES OF THE CORRELATION MATRIX

Let $\{u(n)\}$ be the samples of a discrete process. If we store them in a vector with length M :

$$u^T(n) = [u(n) \ u(n-1) \ \dots \ u(n-M+1)]$$

the correlation matrix of the process can be written as follows:

$$R = E\{u(n)u^H(n)\} = \begin{bmatrix} r(0) & r(1) & \dots & r(M-1) \\ r(-1) & r(0) & \dots & r(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-M+1) & r(-M+2) & \dots & r(0) \end{bmatrix}.$$

Properties of R

1. If the process is stationary, R will be hermitic and Toeplitz.

2. R is definite positive:

$$x^H R x \geq 0 \ \forall x.$$

Proof:

Suppose

$$y = x^H u(n) \Rightarrow y^* = u^H(n)x$$

then,

$$E\{|y|^2\} = E\{yy^*\} = E\{x^H u(n) u^H(n) x\} = x^H R x \geq 0.$$

□

3. The eigenvalues of R are real and non-negative.

Proof:

$$Rq_i = \lambda_i q_i \quad i = 1, 2, \dots, M$$

$$q_i^H R q_i = \lambda_i q_i^H q_i$$

then

$$\lambda_i = \frac{q_i^H R q_i}{q_i^H q_i} \geq 0$$

due to the 2nd property. □

4. If $\lambda_1, \lambda_2, \dots, \lambda_M$ are eigenvalues of $R \in \mathbb{C}^{M \times M}$, the sum of the eigenvalues equals to the trace of R .

Proof:

Decomposition of R : $Q^H R Q = \Lambda$

$$\text{tr}[Q^H R Q] = \text{tr}[\Lambda] = \sum_{i=1}^M \lambda_i$$

Since $\text{tr}(AB) = \text{tr}(BA)$, then

$$\text{tr}[Q^H R Q] = \text{tr}[Q^H Q R] = \text{tr}[R].$$

□

5. The eigenvalues of the correlation matrix of a discrete process are limited by the maximum and the minimum values of the power spectrum of the process.

Proof:

$$\lambda_i = \frac{q_i^H R q_i}{q_i^H q_i} \quad i = 1, 2, \dots, M$$

$$\text{with } q_i = \begin{bmatrix} q_{i1} \\ q_{i2} \\ \vdots \\ q_{iM} \end{bmatrix}$$

Since the correlation is the inverse Fourier Transform of the power spectrum, that is

$$r(l-k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(w) e^{jw(l-k)} dw,$$

we have

$$\begin{aligned} q_i^H R q_i &= \frac{1}{2\pi} \sum_{k=1}^M \sum_{l=1}^M q_{ik}^* q_{il} \int_{-\pi}^{\pi} S(w) e^{jw(k-l)} dw \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} S(w) \sum_{k=1}^M q_{ik}^* e^{jwk} \sum_{l=1}^M q_{il} e^{-jwl} dw \end{aligned}$$

If we denote the Fourier Transform of vector \mathbf{q}_i by $Q_i(e^{jw}) = \sum_{k=1}^M q_{ik} e^{-jkw}$, then

$$\mathbf{q}_i^H \mathbf{R} \mathbf{q}_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 S(w) dw$$

and

$$\mathbf{q}_i^H \mathbf{q}_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 dw.$$

From the Rayleigh quotient, we have that

$$\lambda_i = \frac{\int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 S(w) dw}{\int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 dw}.$$

Therefore, we can bound the integrals in the following way:

$$\int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 S(w) dw \geq S_{MIN} \int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 dw \quad (\text{lower bound})$$

$$\int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 S(w) dw \leq S_{MAX} \int_{-\pi}^{\pi} |Q_i(e^{jw})|^2 dw \quad (\text{upper bound})$$

So, we can guarantee that:

$$S_{MIN} \leq \lambda_i \leq S_{MAX}$$

□

4. SVD DECOMPOSITION AND ADVANCED TOPICS ON EIGENSYSTEMS

4.1. SVD DECOMPOSITION.

4.2. GENERALIZED EIGENVALUE PROBLEM.

4.3. PERTURBATION OF EIGENSYSTEMS.

4.1. SINGULAR VALUE DECOMPOSITION (SVD)

So far we have presented two useful decompositions that use unitary matrices:

- The QR-decomposition $A = QR$ for $A \in \mathbb{C}^{M \times N}$ (non square!).
- The Schur decomposition $A = QRQ^H$ for $A \in \mathbb{C}^{M \times M}$ (square!). $\leftarrow A \text{ normal } // \& \text{ square}$

The SVD decomposition provides a similar factorization for all matrices, even matrices that are not square or have repeated eigenvalues, becoming one of the most important tools of matrix applications.

Theorem

Any matrix $A \in \mathbb{C}^{M \times N}$ has the so-called *singular value decomposition* defined as

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

where $U \in \mathbb{C}^{M \times M}$ is a unitary matrix (orthogonal if A is real) and $V \in \mathbb{C}^{N \times N}$ is a unitary matrix (orthogonal if A is real).

$$U^H U = I_{(M \times M)} \text{ and } V^H V = I_{(N \times N)}. \leftarrow \text{Unitaries}$$

Matrix $\Sigma \in \mathbb{R}^{M \times N}$ is a diagonal matrix of the form

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$$

where $p = \min(M, N)$ and σ_i are the *singular values*, usually ordered so that

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0.$$

The matrix Σ is written as a diagonal matrix, even though is not square, that is, if $M \neq N$.

$$\Sigma = \begin{cases} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} & \text{If } M > N \\ \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & & \\ 0 & 0 & \cdots & \sigma_M & \cdots & 0 \end{bmatrix} & \text{If } N > M \end{cases}$$

Properties of the SVD decomposition

- Note that $\Sigma = U^H A V \Rightarrow A V = U \Sigma$ and this means

$$A v_i = \sigma_i u_i \text{ for } i = 1, \dots, \min(M, N).$$

Therefore, we refer to v_i as the *right singular vectors* of A .

Similarly, $\Sigma = U^H A V \Rightarrow U^H A = \Sigma V^H$ and this is equivalent to

$$u_i^H A = \sigma_i v_i^H \text{ for } i = 1, \dots, \min(M, N).$$

Therefore, we refer to u_i as the *left singular vectors* of A .

- Note that we have

$$A^H A = V \Sigma^H U^H U \Sigma V^H = V (\Sigma^H \Sigma) V^H$$

where $\Sigma^H \Sigma \in R^{N \times N}$ is diagonal with real nonnegative entries on its main diagonal. Matrix V is unitary and its columns are the eigenvectors of $A^H A$ associated with the real nonnegative eigenvalues of $A^H A$ that are on the diagonal of $\Sigma^H \Sigma$. This decomposition always exists since $A^H A$ is normal!

- Similarly, we have

$$A A^H = U \Sigma V^H V \Sigma^H U^H = U (\Sigma \Sigma^H) U^H$$

where $\Sigma \Sigma^H \in R^{M \times M}$ is diagonal with real nonnegative entries on its main diagonal. Matrix U is unitary and its columns are the eigenvectors of $A A^H$ associated with the real nonnegative eigenvalues of $A A^H$ that are on the diagonal of $\Sigma \Sigma^H$.

- The diagonal entries of $\Sigma^H \Sigma \in R^{N \times N}$ and $\Sigma \Sigma^H \in R^{M \times M}$ are just the square singular values σ_i^2 (perhaps with some zeros added).

- Also note that $\Sigma = U^H A V$ is equivalent to $A = U \Sigma V^H$ and this leads to

$$\xrightarrow{M>N} A = [u_1 \dots u_M] \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1^H \\ v_2^H \\ \vdots \\ v_N^H \end{bmatrix} = \sum_{i=1}^{\min(M,N)} \sigma_i u_i v_i^H.$$

This expression holds for both $M>N$ and $N>M$.

Theorems

- $A^H A$ and AA^H have the same nonzero eigenvalues.

$$A^H A v = \lambda v \Rightarrow AA^H \frac{Av}{\|v\|} = \lambda \frac{Av}{\|v\|}$$

$$AA^H u = \lambda u$$

- U and V are not unique

Both matrices U and V arise in the computation of the eigenvectors of AA^H and $A^H A$, and the eigenvectors are not uniquely defined.

U i V es troben de la EVD de $A^H A$ i $A^H A$ respectivament

COMPUTATION OF THE SVD DECOMPOSITION

- We diagonalize the Gramm matrix

$$A^H A = V \Sigma^H U^H U \Sigma V^H = V (\Sigma^H \Sigma)_{N \times N} V^H$$

- We diagonalize the Gramm matrix

$$A A^H = U \Sigma V^H V \Sigma^H U^H = U (\Sigma \Sigma^H)_{M \times M} U^H$$

- The solution must follow: $AV = U\Sigma$

$$A v_i = U \Sigma V^H v_i = U \Sigma \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = U \begin{bmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{bmatrix} = \sigma_i u_i$$

The sign of σ_i has to be verified for each singular vector.

ExampleSVD decomposition of $A \in \mathbb{R}^{2 \times 3}$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -3 & 1 & -2 \end{bmatrix}_{2 \times 3} \quad A = \underline{U} \sum \underline{V}^H$$

$$A^H A = V (\Sigma \Sigma^H) V^H \rightarrow \underline{V} = \begin{bmatrix} \cdot & \cdot & \cdot \end{bmatrix}_{3 \times 3}$$

$$A A^H = U (\Sigma \Sigma^H) V^H \rightarrow \underline{U} = \begin{bmatrix} \cdot & \cdot & \cdot \end{bmatrix}_{2 \times 2}$$

$$A \underline{V}_i = \sigma_i \underline{U}_i \rightarrow \underbrace{\begin{matrix} A & \underline{V}_1 & ? \\ 2 \times 3 & 3 \times 1 & 2 \times 1 \\ (\text{column}) & & (\text{column}) \end{matrix}}_{\text{identificare i valori}} = \sigma_1 \underline{U}_1 \quad \dots \quad \underbrace{\begin{matrix} A & \underline{V}_2 & ? \\ 2 \times 3 & 3 \times 1 & 2 \times 1 \\ (\text{column}) & & (\text{column}) \end{matrix}}_{\text{trovare le parelle}} = \sigma_2 \underline{U}_2 \leftrightarrow \begin{bmatrix} \underline{V}_1 & \underline{V}_2 & \underline{V}_3 \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

$$\begin{matrix} A & \underline{V} & = & \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \\ 2 \times 3 & 3 \times 3 & & \text{identificare i valori} \\ \underline{U} & \underline{\Sigma} & = & \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \\ 2 \times 2 & 2 \times 3 & & \text{trovare le parelle} \\ & & & \text{VAP - VGP} \\ & & & (\text{amb el segno aderente}) \end{matrix}$$

$$= [\underline{U}, \underline{U}_2] \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix}$$

[V, D1]=eig(A'*A)

V =

$$\begin{matrix} 0.5345 & -0.5774 & 0.6172 \\ -0.8018 & -0.5774 & 0.1543 \\ -0.2673 & 0.5774 & 0.7715 \end{matrix}$$

D1 =

$$\begin{matrix} 7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 21 \end{matrix}$$

A'*V(:,3)=

$$\begin{matrix} 3.2404 \\ -3.2404 \end{matrix}$$

>>sqrt(D(1,1))*U(:,1) =

$$\begin{matrix} -3.2404 \\ 3.2404 \end{matrix}$$

Reordering:

V=[V(:,3) V(:,1) V(:,2)]

V =

$$\begin{matrix} 0.6172 & 0.5345 & -0.5774 \\ 0.1543 & -0.8018 & -0.5774 \\ 0.7715 & -0.2673 & 0.5774 \end{matrix}$$

S =

$$\begin{matrix} -4.5826 & 0 & 0 \\ 0 & 2.6458 & 0 \end{matrix}$$

U =

$$\begin{matrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{matrix}$$

A'*V(:,1) =

$$\begin{matrix} -1.8708 \\ -1.8708 \end{matrix}$$

>>sqrt(D(2,2))*U(:,2) =

$$\begin{matrix} -1.8708 \\ -1.8708 \end{matrix}$$

A=U*S*V' =

$$\begin{matrix} 1.0000 & 2.0000 & 3.0000 \\ -3.0000 & 1.0000 & -2.0000 \end{matrix}$$

SINGULAR VALUES AND RANK

The SVD decomposition contains a great deal of information about the matrix A . Before going into that, let us assume that the lowest ($p-r$) singular values are zero such as $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ and $\sigma_{r+1} = \dots = \sigma_p = 0$. Then we can break the Σ matrix into two parts, corresponding to the nonzero singular values and the zero singular values, as follows

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}$$

where $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r) \in R^{rr}$ and $\Sigma_2 = \text{diag}(\sigma_{r+1}, \dots, \sigma_p) = 0 \in R^{(M-r) \times (N-r)}$.

Then the SVD can be written as

$$A = [U_1 \ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix} = U_1 \Sigma_1 V_1^H$$

where U_1 is $M \times r$, U_2 is $M \times (M-r)$, V_1 is $N \times r$ and V_2 is $N \times (N-r)$.

Theorems

1. The rank r of A equals the number r of nonzero singular values of A .

$$\text{rank}(A) = \text{rank}(A^H A)$$

2. The first r left singular vectors $\{u_1, \dots, u_r\}$ form an orthonormal basis for the column space of A .
The range (column space) of a matrix is

$$\mathcal{R}(A) = \{b : b = Ax; x \in C^{Nr}\}.$$

Substituting in the SVD, we obtain

$$\begin{aligned} \mathcal{R}(A) &= \{b : b = U\Sigma V^H x\} \\ &= \left\{ b : b = U\Sigma \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \right\} \\ &= \{b : b = U_1 y_1\} = \text{span}\{U_1\} \end{aligned}$$

3. The last $N-r$ right singular vectors $\{v_{k+1}, \dots, v_N\}$ form an orthonormal basis for the null space of A .
The null of a matrix is

$$\mathcal{N}(A) = \{x : Ax = 0; x \in C^{Nr}\}$$

Substituting in the SVD, we obtain

$$\begin{aligned} \mathcal{N}(A) &= \{x : U\Sigma V^H x = 0\} \\ &= \{x : \Sigma V^H x = 0\} \\ &= \left\{ x : \begin{bmatrix} \Sigma_1 V_1^H \\ \Sigma_2 V_2^H \\ 0 \end{bmatrix} x = 0 \right\} \\ &\quad \downarrow \begin{matrix} V_1^H V_2 = 0 \\ V_2^H V_2 = 0 \end{matrix} \\ &= \{x : \Sigma_1 V_1^H x = 0\} = \mathcal{N}(V_1^H) = \mathcal{R}(V_2) \end{aligned}$$

4. Similarly, $\mathcal{R}(A^H) = \text{span}(V_1)$ and $\mathcal{N}(A^H) = \text{span}(U_2)$.

4.2. GENERALIZED EIGENVALUE PROBLEM

Let A and B be two matrices that belong to $C^{M \times M}$. The so-called generalized eigenvalues and generalized eigenvectors are the solution of the problem

$$Ax = \lambda Bx \quad x \neq 0.$$

The set of all matrices of the form $A - \lambda B$ with $\lambda \in C$ is said to be a *pencil matrix*.

Depending on the rank of B we switch between two different cases:

- If $\text{rank}(B) = M$, then the problem becomes a simpler one as follows

$$B^{-1}Ax = \lambda x.$$

In this case, there are M eigenvalues.

- If $\text{rank}(B) < M$ (B is rank deficient), the number of generalized eigenvalues may be finite, empty or infinite.

Example

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \lambda(A, B) = \{1\}$$

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \lambda(A, B) = \{1\} \text{ There are no eigenvalues}$$

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \Rightarrow \lambda(A, B) = C \text{ } \infty \text{ eigenvalues}$$

Methods for solving the $A - \lambda B$ problem

1. Method for $\text{rank}(B) = M$.

Note that $\lambda(A, B) = \lambda(B^{-1}A, I) = \lambda(B^{-1}A)$

- Solve $BC = A$ for C .
- Use the QR algorithm to compute the eigenvalues of C . \rightarrow tröben eine base orthonormal de \subseteq

2. Method for $\text{rank}(B) < M$.

To solve this problem we need the Generalized Schur Decomposition.

Theorem

If A and B are in $C^{M \times M}$, then there exists unitary Q and Z such that $Q^H AZ = T$ and $Q^H BZ = S$ are upper triangular. If for some i , T_{ii} and S_{ii} are both zero, then $\lambda(A, B) = C$. Otherwise

$$\lambda(A, B) = \left\{ \frac{T_{ii}}{S_{ii}} \mid S_{ii} \neq 0 \right\}$$

Assuming there exists unitary Q and Z such that $Q^H AZ = T$ and $Q^H BZ = S$ are upper triangular,

$$\begin{aligned} \det(A - \lambda B) &= \det(QQ^H AZZ^H - \lambda QQ^H BZZ^H) = \\ &= \det[Q(Q^H AZ - \lambda Q^H BZ)Z^H] = \\ &= \det(QZ^H) \det(T - \lambda S) = \\ &= \det(QZ^H) \prod_{i=1}^M (T_{ii} - \lambda S_{ii}). \end{aligned}$$

4.3. PERTURBATION THEORY

CONDITION OF LINEAR SYSTEMS

We consider here the condition of the solution \mathbf{x} to the system of equations $A\mathbf{x} = \mathbf{b}$ in terms of the data \mathbf{b} and A . We want to know by how much the solution \mathbf{x} changes - say from \mathbf{x} to $\mathbf{x} + \delta\mathbf{x}$ - when the data \mathbf{b} and A change - say $\mathbf{b} + \delta\mathbf{b}$ and $A + \delta A$.

If small changes in the data \mathbf{b} and A always lead to reasonably small changes in the answer to the problem, the problem is said to be well-conditioned. If small changes in the data \mathbf{b} and A can sometimes lead to unacceptably large changes in the answer to the problem, the problem is said to be ill-conditioned.

The condition concept is important in applied problems because data is almost always inaccurate because of measuring and modeling errors. Thus it is crucial for us to know what effect such inaccuracies in the data have on the answer to the problem.

Theorem of Bauer-Fike I

$A \in \mathbb{R}^{n \times n}$

Let A be nonsingular and let $\|\cdot\|$ denote any vector norm and its corresponding matrix norm. Suppose that \mathbf{x} solves $A\mathbf{x} = \mathbf{b}$. While

$$\mathbf{x} + \delta\mathbf{x} \text{ solves } (A + \delta A)(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$$

for some perturbations δA and $\delta\mathbf{b}$ in the data. Suppose that the perturbation δA is small enough that $\|A^{-1}(\delta A)\|_p < 1$ or $\|(\delta A)A^{-1}\|_p < 1$ then the change $\delta\mathbf{x}$ in the solution satisfies

$$\frac{\|\delta\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \leq M \cdot c(A) \cdot \left(\frac{\|\delta\mathbf{b}\|_p}{\|\mathbf{b}\|_p} + \frac{\|\delta A\|_p}{\|A\|_p} \right),$$

where $M = \left(1 - \|(\delta A)A^{-1}\|_p\right)^{-1}$ and $c(A) = \|A\|_p \|A^{-1}\|_p$ is the so-called condition number of A .

To see what this theorem says about condition, suppose that we decide to measure the size of the changes in, let's say, \mathbf{y} by examining $\|\delta\mathbf{y}\|_p / \|\mathbf{y}\|_p$. Then, the theorem compare the changes in the solution \mathbf{x} to those in the data \mathbf{b} and A .

- If the change δA in A is small enough, the constant M in the theorem is near 1. In this case, the bound on the change $\|\delta\mathbf{x}\|_p / \|\mathbf{x}\|_p$ will not be much larger than the changes in the data if the condition number $c(A)$ is not too large. This means that a moderate condition number $c(A)$ guarantees that the equations are well-conditioned: small changes in the data produce reasonable small changes in the solution.
- However, if $c(A)$ is large, the changes in \mathbf{x} caused by changes in the data may be much larger than the changes in the data (whether this actually occurs depends on the specific \mathbf{b} , but it is usually wise to view such a system as ill-conditioned).
- The meaning of the condition number $c(A)$ can be examined from another viewpoint. It can be seen that

$$A + \delta A \text{ is non singular as long as } \frac{\|\delta A\|_p}{\|A\|_p} < \frac{1}{c(A)}.$$

Put differently,

$$c^{-1}(A) \geq \frac{\|\delta A\|_p}{\|A\|_p} \text{ whenever } A + \delta A \text{ is singular.}$$

4.15

ens dóna una idea de com
de l'imatge pot ser un singular
o net en A

In fact, $c^{-1}(A)$ equals the supremum of $\|\delta A\|_p / \|A\|_p$, taken over all δA for which $A + \delta A$ is singular. Thus the condition number measures how far A is from the nearest singular matrix: if $c^{-1}(A)$ is small, or equivalently $c(A)$ is large, then a small perturbation of A may lead to a singular matrix.

- The condition number for the 2-norm is

$$c(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\text{maximum singular value of } A}{\text{minimum singular value of } A}.$$

dispersió dels VAPS
 $\lambda_{\max} \gg \lambda_{\min} \rightarrow$ matríc mal condicionada

4.16

Example

Unfortunately, in practice it is not always easy to recognize an ill-conditioned matrix A just by looking at A . For instance, the following matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} & \frac{1}{11} \\ \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} & \frac{1}{11} & \frac{1}{12} \\ \frac{1}{7} & \frac{1}{8} & \frac{1}{9} & \frac{1}{10} & \frac{1}{11} & \frac{1}{12} & \frac{1}{13} \end{bmatrix}$$

has an inverse with an infinite norm of the order of 10^8 . Therefore, the condition number of A is about 10^8 , so there is a singular matrix within about 10^8 of A - not something we would easily guess from looking at A .

PERTURBATION OF EIGENSYSTEMS

com d'iferents sen els VAPs
si hi ha perturbacions.

In this section we study the condition of eigensystems.

Theorem (Gershgorin circles)

Each eigenvalue λ of a matrix $A \in \mathbb{C}^{N \times N}$ satisfies at least one of the inequalities

$$|\lambda - a_{ii}| \leq r_i \text{ where } r_i = \sum_{j \neq i}^N |a_{ij}| \text{ for } i = 1, \dots, N.$$

↓ Diagonals ↑ to file i

That is, each eigenvalue lies in at least one of the discs with centre a_{ii} and radius r_i in the complex plane.

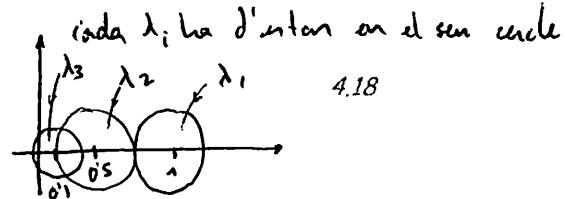
Example

For the matrix

$$A = \begin{bmatrix} 1 & -10^{-5} & 2 \cdot 10^{-5} \\ 4 \cdot 10^{-5} & 0.5 & -3 \cdot 10^{-5} \\ -10^{-5} & 3 \cdot 10^{-5} & 0.21 \end{bmatrix}, \quad \begin{array}{c} \xrightarrow{+} 3 \cdot 10^{-5} \\ \xrightarrow{+} 7 \cdot 10^{-5} \\ \xrightarrow{+} 4 \cdot 10^{-5} \end{array}$$

the eigenvalues satisfy

$$\begin{aligned} |\lambda_1 - 1| &\leq 3 \cdot 10^{-5} \\ |0.5 - \lambda_2| &\leq 7 \cdot 10^{-5} \\ |0.1 - \lambda_3| &\leq 4 \cdot 10^{-5} \end{aligned}$$



Since this theorem provides a bound on the error of the eigenvalues, we may compute the eigenvalues of a matrix with the iterative QR algorithm with a finite number of steps. \square

Now we consider the problem of how the eigenvalues and eigenvectors of a matrix A change in terms of changes in the data A .

Theorem of Bauer-Fike II

tants VEPs indep. com VAPs

Assume μ is an eigenvalue of $A + \delta A$ and also that A is a non-defective matrix such that $Q^{-1}AQ = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$, then

$$c(Q) \downarrow \Rightarrow \mu \approx \lambda$$

$$\min_{\lambda \in \sigma(A)} |\lambda - \mu| \leq c(Q) \|\delta A\|_p.$$

That is, a lower value of the condition number of Q will not arise to large perturbations of the eigenvalues of $A + \delta A$. Conversely, if $c(Q)$ is large, we can not guarantee anything.

Moreover, note that if Q is unitary we have

$$\min_{\lambda \in \sigma(A)} |\lambda - \mu| \leq \|\delta A\|_p.$$

Q unitari $\Rightarrow \mu \approx \lambda$
 $\delta A \downarrow$

That is, the eigenvalues of normal matrices are always well-conditioned.



The non-normal matrices can have a mixture of ill-conditioned eigenvalues and well-conditioned eigenvalues.
 Assume a given eigenvalue λ of matrix A ; then we have

$$Ax = \lambda x \text{ and } y^H A = \lambda y^H \text{ with } \|x\|_2 = \|y\|_2 = 1.$$

x : VEP per l'eigenvalue
 y : " " la dreta

If we consider a perturbation matrix δA , then

$$(A + \varepsilon \delta A) \cdot x(\varepsilon) = \lambda(\varepsilon)x(\varepsilon) \text{ where } \|\delta A\|_2 = \|x(\varepsilon)\|_2 = 1.$$

Deriving with respect to ε and forcing $\varepsilon = 0$, we obtain

$$A \cdot \dot{x}(0) + \delta A \cdot x(0) = \dot{\lambda}(0)x(0) + \lambda(0)\dot{x}(0).$$

We may multiply both sides by y^H and afterwards divide by $y^H x$ to have

$$\dot{\lambda}(0) = \left| \frac{y^H \delta A x}{y^H x} \right| \leq \frac{1}{\|y^H x\|},$$

$$y^H A \dot{x} + y^H \delta A x = \dot{\lambda} x + y^H \lambda \dot{x}$$

$$|\dot{\lambda}(0)| = \frac{|\Delta \lambda|}{\varepsilon} \leq \frac{1}{\|y^H x\|}$$

with equality if $\delta A = y^H x$. Therefore, the condition number of that eigenvalue is defined as follows

$$s(\lambda) = |y^H x| \text{ with } \|x\|_2 = \|y\|_2 = 1.$$

* el condicionament d'un VAP depen de l'angle entre VEPs per la dreta; per l'esquerra.

VAPs + robustes a interaccions amb perturbacions
 q. en VEPs

com varien els VEPs
 amb una perturbació.

The sensibility of an eigenvector x_k depends upon eigenvalue sensitivity and the separation of λ_k from the other eigenvalues.

$$x_k(\varepsilon) = x_k(0) + \varepsilon \sum_{i=1, i \neq k}^N \left\{ \frac{y_i^H \delta A x_k}{|\lambda_k - \lambda_i| y_i^H x_i} \right\} x_i + O(\varepsilon^2)$$

dopen de la distància
 entre VEPs.
 $\Rightarrow x_k \approx \lambda_i \Rightarrow$ la perturbació
 afecta molt.

If λ is a non-defective multiple eigenvalue, then there are an infinite number of possible eigenvector bases for the associated invariant subspace. The preceding result indicates that this indeterminacy begins to be felt as the eigenvalues collapse.

Example

For matrix A we have

$$A = \begin{bmatrix} 1.01 & 0.01 \\ 0.00 & 0.99 \end{bmatrix} \Rightarrow \begin{aligned} \lambda_1 &= 1.01 & \frac{1}{s(\lambda_2)} &= 1.118 \\ \lambda_2 &= 0.99 & x_2 &= [0.4472 \quad -0.8944] \end{aligned}$$

VEPs

But, for matrix $A + \delta A$ we obtain

$$A + \delta A = \tilde{A} = \begin{bmatrix} 1.01 & 0.01 \\ 0.00 & 1.00 \end{bmatrix} \Rightarrow \begin{aligned} \tilde{\lambda}_1 &= 1.01 \\ \tilde{\lambda}_2 &= 1.00 \end{aligned}$$

$$\tilde{x}_2 = [0.7071 \quad -0.7071]$$

A.II APPLICATIONS OF EIGENSYSTEMS OF GENERAL MATRICES

A.II.1. POPULATION GROWTH: POWERS OF A MATRIX

A.II.2. QUADRATIC FORMS. GEOMETRIC INTERPRETATION

A.II.3. SIGNAL SUBSPACE TECHNIQUES FOR FREQUENCY ESTIMATION

POPULATION GROWTH: POWERS OF A MATRIX

In this section, we aim at presenting the tools provided by the eigensystem theory in the problem of powers of a matrix. For this purpose, we study a simple model of growth of populations over time.

Foxes and Chickens

Specifically, we consider the growth of two populations, let's say *Foxes* and *Chickens*, that compete with one another. Imagine that we count the number of these populations every year and we denote them by F_i and C_i , respectively. Assume that the chickens, without any foxes to harass them, would have a birth rate exceeding the death rate, such as $C_{i+1} = 1.2 \cdot C_i$. Conversely, without chickens to feed upon, we would expect the foxes to die out, let's say $F_{i+1} = 0.6 \cdot F_i$.

We want to model what happens when the foxes succeed in eating some number of chickens every year, and we assume this would allow an increase in the fox population proportional to the number of chickens killed. Similarly, the chicken population would decrease because of the raiding foxes. To be specific we assume

$$\begin{aligned} F_{i+1} &= 0.6 \cdot F_i + 0.5 \cdot C_i \\ C_{i+1} &= -k \cdot F_i + 1.2 \cdot C_i \end{aligned} \tag{1}$$

for $i = 1, 2, \dots$, and with $F_1 = 100$ and $C_1 = 1000$. Therefore, variable k reflects the kill rate of chickens by foxes.

We use matrices to analyze the behaviour of these populations as time passes. If we let

$$A = \begin{bmatrix} 0.6 & 0.5 \\ -k & 1.2 \end{bmatrix}, \quad x_i = \begin{bmatrix} F_i \\ C_i \end{bmatrix}, \quad \text{so that } x_1 = \begin{bmatrix} 100 \\ 1000 \end{bmatrix},$$

our model (1) becomes

$$x_{i+1} = Ax_i, \quad \text{for } i=1,2,\dots \text{ with } x_1 = [100 \ 1000].$$

Equivalently,

$$x_{i+1} = A^i x_1, \quad \text{for } i=1,2,\dots$$

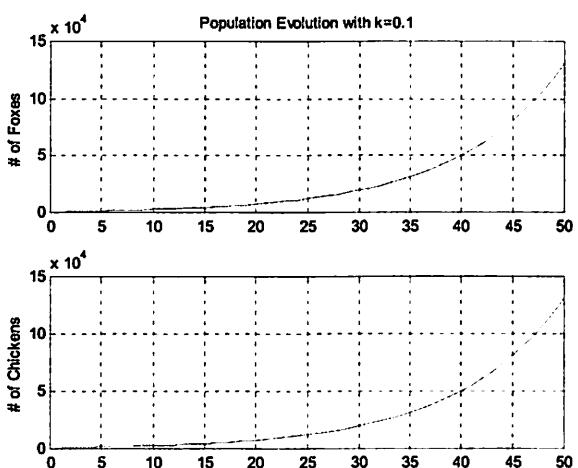
which means that studying the behaviour of x_i as i increases is essentially equivalent to studying the behaviour of the powers A^i of A . Matrix A^i behaves quite differently for different values of k .

Example with $k=0.1$

With $k=0.1$, the transition matrix is equal to

$$A = \begin{bmatrix} 0.6 & 0.5 \\ -0.1 & 1.2 \end{bmatrix}.$$

The figure below shows the number of foxes and chickens at every year for several values of i . The model verifies that, for a low kill rate k , the chicken population grows without bound and this allows the fox population also to grow without bound. The two populations become equal.



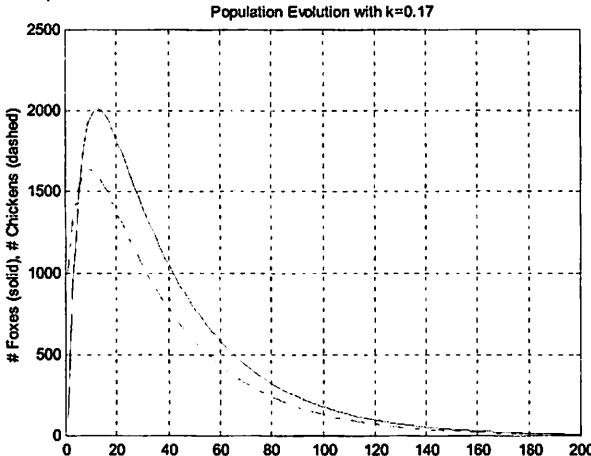
□

Example with $k = 0.17$

With $k = 0.17$, the transition matrix is equal to

$$A = \begin{bmatrix} 0.6 & 0.5 \\ -0.17 & 1.2 \end{bmatrix}.$$

The figure below shows the number of foxes and chickens at every year for several values of i . The chicken population is killed off, which in turn leads to the death of the foxes. That is, both populations die.



□

These two examples show that our model can produce different types of behaviour for different values of k . We would like to know, for instance, how x_i behaves as i grows for any specific value of k . Or equivalently, how the powers A^i of A behaves for different values of k .

Matrix Powers

The mathematical model of the previous example, that is,

$$x_i = A^i x_0, \quad \text{for } i \geq 0,$$

arises in a broad class of applications (biology, economy, mechanics, control...). We can be interested in knowing what happens with the state variables, x_i , as time passes: if it tends to an equilibrium point or, otherwise, if it oscillates or becomes arbitrarily large.

The Jordan form of A , $A = QJQ^{-1}$, is extremely useful to study the evolution of such systems. The Jordan form of A^i is

$$A^i = QJ^iQ^{-1}$$

with J^i equal to

i.e. A^i similar to J^i

$$\mathbf{J}^i = \begin{bmatrix} \mathbf{J}_1^i & 0 & \cdots & 0 \\ 0 & \mathbf{J}_2^i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{J}_L^i \end{bmatrix}.$$

Therefore, we only need to study \mathbf{J}^i . We assume \mathbf{J} is strictly diagonal (for defective matrices, see [Noble, pp. 372-374]), then

$$\mathbf{J}^i = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_L \end{bmatrix} = \begin{bmatrix} \lambda_1^i & 0 & \cdots & 0 \\ 0 & \lambda_2^i & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_L^i \end{bmatrix}.$$

For any number λ , it is clear how λ^i behaves as i tends to infinity

$$\lim_{i \rightarrow \infty} \lambda^i = \begin{cases} 0 & |\lambda| < 1 \\ 1 & |\lambda| = 1 \\ \infty & |\lambda| > 1 \end{cases}$$

Using the spectral decomposition of a matrix, \mathbf{A}^i can be decomposed as follows

$$\mathbf{A}^i = \mathbf{Q} \mathbf{J}^i \mathbf{Q}^{-1} = \sum_{l=1}^L \lambda_l^i q_l (q_l^i)^T$$

$$\text{dem: } \mathbf{A} = \mathbf{Q} \Lambda \mathbf{Q}^{-1} = [\mathbf{q}_1 \cdots \mathbf{q}_L] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_L \end{bmatrix} [\mathbf{q}_1 \cdots \mathbf{q}_L]^H = \sum_{l=1}^L \lambda_l q_l (q_l^i)^H$$

Applications of Eigensystems of General Matrices

$\rightarrow \left\{ \mathbf{A} = \mathbf{Q} \Lambda \mathbf{Q}^{-1} \rightarrow \mathbf{Q} \text{ non sh. VEPS: for non unitary base} \rightarrow \begin{array}{l} \text{to put orthonormalization} \\ \text{for unitary} \end{array} \right. \downarrow$

and if \mathbf{Q} is unitary we have

$$\mathbf{A}^i = \mathbf{Q} \mathbf{J}^i \mathbf{Q}^H = \sum_{l=1}^L \lambda_l^i q_l q_l^H.$$

$$\mathbf{A} = \mathbf{Q} \Lambda \mathbf{Q}^H$$

As $i \rightarrow \infty$, we have

$$\lim_{i \rightarrow \infty} \mathbf{A}^i = \lim_{i \rightarrow \infty} \lambda_{\max}^i q(q^i)^T = \lim_{i \rightarrow \infty} \lambda_{\max}^i q q^H,$$

or, equivalently, for negative powers

$$\lim_{i \rightarrow \infty} \mathbf{A}^{-i} = \lim_{i \rightarrow \infty} \lambda_{\min}^{-i} q(q^i)^T = \lim_{i \rightarrow \infty} \lambda_{\min}^{-i} q q^H.$$

The matrix \mathbf{A} for the competing-populations model has eigenvalues $\lambda = \{0.7, 1.1\}$ for $k = 0.1$. Looking at the maximum eigenvalue, larger than 1, it can be understood the unstable behaviour of the chicken and fox populations. For $k = 0.17$, the largest eigenvalue is equal to $\lambda = 0.97$ and, therefore, the populations go to 0.

λ_{\max} is e.g. zero to tendance per $i \rightarrow \infty$

$$\left\{ \begin{array}{l} k = 0.1 \Rightarrow \lambda_{\max} > 1 \Rightarrow \mathbf{A}^i \rightarrow \infty \\ k = 0.17 \Rightarrow \lambda_{\max} < 1 \Rightarrow \mathbf{A}^i \rightarrow 0 \end{array} \right.$$

QUADRATIC FORMS.

The quadratic functions arise in diverse applications areas (dynamics, signal processing, statistics), and mostly of those applications involve the optimization of some goal. Matrix methods allow a unified study of their properties.

Definition.

A quadratic form of a hermitian matrix $A \in \mathbb{C}^{N \times N}$ is a scalar of the form

$$Q(x) = \langle x^H, Ax \rangle = x^H Ax.$$

The vector is, in general, $x \in \mathbb{C}^N$ and $\langle \cdot, \cdot \rangle$ denotes the conventional inner product of \mathbb{C}^N . Notice that even when A and x are complex, $Q(x)$ will necessarily be real since A is hermitian.

Example.

In the statistical analysis of data described by random variables $\{x_1, \dots, x_N\}$, we have

- $E[x_i]$, the expected value of x_i .
- $E[x_i^2]$, the variance of x_i (we have assumed that $E[x_i] = 0$ for all i).

- S , the covariance matrix with components $s_{ij} = E[x_i x_j]$ $i, j = 1, \dots, N$.
- P , the total variance defined as $P = E[x_1^2 + \dots + x_N^2] = \text{tr}(S)$.

It is often important in the statistical analysis of data to seek some small set of new variables that explain the experimental results. A common approach is to seek a new variable that depends linearly on the set $\{x_1, \dots, x_N\}$ as follows,

$$y = a_1 x_1 + \dots + a_N x_N = [a_1 \ \dots \ a_N] \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = a^T x$$

whose variations in some sense reflect as much as possible of the variations in the x_i . This requires that a be chosen so as to maximize the variance $E[y^2]$ subject to the constraint that $\|a\|_2 = 1$. Notice that

$$E[y^2] = E[a^T x x^T a] \stackrel{E \text{ is linear}}{=} a^T E[x x^T] a = a^T S a.$$

Then, the problem reduces to an optimization problem of a quadratic form

$$\arg \max_{a \in \mathbb{C}^N} \langle a, S a \rangle \quad \text{subject to } \langle a, a \rangle = 1.$$

□

Geometric interpretation

By diagonalizing the matrix A in $A = Q\Lambda Q^T$, the quadratic form $Q(y)$ is transformed to a new coordinate system in which the geometry becomes more apparent. For convenience we will assume that real vectors are used. Then, we have

$$Q(y) = y^T A y = y^T Q \Lambda Q^T y = z^T \Lambda z = \sum_{i=1}^N \lambda_i z_i^2 \quad (2)$$

where $z = Q^T y$ and $Q = [q_1 \ \cdots \ q_N]$.

The variable z can be interpreted geometrically in two dimensions by observing that when

- $z = [1 \ 0]^T$, then $y = q_1$, the first eigenvalue of A
- $z = [0 \ 1]^T$, then $y = q_2$, the second eigenvalue of A .

The orthogonal eigenvectors of A provide the orthogonal bases of a new coordinate system. Next figure illustrates this concept.

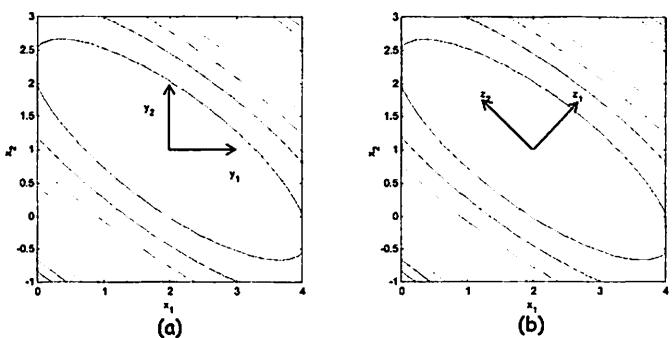


Figure (a) shows the level curves, $Q(x) = ct$, of the following quadratic form

$$(x - x_0)^T A (x - x_0) \text{ where } A = \begin{bmatrix} 3.88 & 3.84 \\ 3.84 & 6.12 \end{bmatrix} \text{ and } x_0 = [2 \ 1].$$

The new coordinates y_1 and y_2 are obtained by the translation $y = x - x_0$. Matrix A has the eigendecomposition

$$\lambda_1 = 9 \quad x_1 = \frac{1}{5} [3 \ 4]^T$$

$$\lambda_2 = 1 \quad x_2 = \frac{1}{5} [-4 \ 3]^T$$

Using (2), the level curves can be obtained by solving the equation

$$9z_1^2 + z_2^2 = ct \Rightarrow \frac{z_1^2}{1} + \frac{z_2^2}{9} = ct' \Leftarrow \text{Ellipse!}$$

The ratio λ_1 / λ_2 controls the shape of the ellipse.

In the general two-dimensional case, the level curves of the quadratic form $Q(x) = ct$ are of the form

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 = ct \Rightarrow \begin{cases} \lambda_1 \lambda_2 > 0 & \text{Ellipse} \\ \lambda_1 \lambda_2 < 0 & \text{Hyperbola} \\ \lambda_1 \lambda_2 = 0 & \text{Parallel lines} \end{cases}$$

Definite Quadratic Forms

In general, a quadratic form $Q(x)$ will assume positive, negative and zero values for various values of x . However, quadratic forms may represent physical phenomena for which a particular sign is meaningless.

1. A quadratic form $Q(x)$ is said to be positive definite if $Q(x) > 0$ for all x .
2. A quadratic form $Q(x)$ is said to be positive semidefinite if $Q(x) \geq 0$ for all x .
3. A quadratic form $Q(x)$ is said to be negative definite or negative semidefinite if and only if $-Q(x)$ is positive definite or positive semidefinite, respectively.

4. A matrix A is said to be positive definite (et cetera) when the quadratic form $Q(x) = \langle x^H, Ax \rangle$ is positive definite (et cetera).

$$\boxed{A \text{ positive def.} \Leftrightarrow x^H A x \text{ positive def.}}$$

Theorem

Suppose that A is hermitian. Then

1. A is positive (negative) definite if and only if all its eigenvalues are strictly positive (negative).
2. A is positive (negative) semidefinite if and only if all its eigenvalues are nonnegative (nonpositive).
3. A is indefinite if and only if A has both strictly positive and strictly negative eigenvalues.

The Rayleigh Quotient

The Rayleigh quotient of a hermitian matrix A is the function defined as

$$\rho(A) = \frac{x^H A x}{x^H x} \quad \text{for } x \neq 0.$$

The Rayleigh quotient is upper and lower bounded by the minimum and maximum eigenvalues of matrix A . That is,

$$\lambda_{\min} \leq \frac{x^H A x}{x^H x} \leq \lambda_{\max} \quad \text{for } x \neq 0$$

Proof

Method of constrained optimization using Lagrange multipliers. Our problem is equivalent to maximize (minimize) $\mathbf{x}^H \mathbf{A} \mathbf{x}$ subject to the constraint $\|\mathbf{x}\|_2 = 1$. Therefore, we form the function

$$J(\mathbf{x}) = \mathbf{x}^H \mathbf{A} \mathbf{x} - \mu \mathbf{x}^H \mathbf{x}$$

where μ is a Lagrange multiplier. Taking the gradient and equating to zero, we obtain

$$\frac{\partial}{\partial \mathbf{x}} J(\mathbf{x}) = \mathbf{A} \mathbf{x} - \mu \mathbf{x} = \mathbf{0}.$$

The maximizing (minimizing) solution \mathbf{x}_{opt} is an eigenvector of \mathbf{A} ,

$$\mathbf{A} \mathbf{x}_{opt} = \mu \mathbf{x}_{opt}, \quad \begin{cases} \max : \mu = \lambda_{\max} \\ \min : \mu = \lambda_{\min} \end{cases}$$

and μ is the associated eigenvalue. Then we have,

$$Q(\mathbf{x}_{opt}) = \mathbf{x}_{opt}^H \mathbf{A} \mathbf{x}_{opt} = \mathbf{x}_{opt}^H \mu \mathbf{x}_{opt} \stackrel{\|\mathbf{x}\|=1}{=} \mu$$

Maximization (minimization) of this, subject to the constraint $\|\mathbf{x}\|_2 = 1$, requires that we choose μ to be the largest (lowest) eigenvalue, and \mathbf{x}_{opt} the associated eigenvector.

□

SIGNAL SUBSPACE TECHNIQUES FOR FREQUENCY ESTIMATION

In this section we will review some methods which can provide good spectral resolution considering the presence of noise. We will break the signal out in terms of *signal subspace* component and a *noise subspace* component in order to approach methods for sinusoidal frequency estimation.

The signal model

Suppose a signal that can be expressed as the sum of exponentials plus a noise component:

$$y(n) = x(n) + w(n) = \sum_{i=1}^P a_i e^{j(2\pi f_i n + \phi_i)} + w(n)$$

We assume that the phase angles ϕ_i are pairwise independent and uniformly distributed over $[-\pi, \pi]$ and $w(n)$ is assumed to be a zero-mean white Gaussian noise with variance σ_w^2 and uncorrelated with the signal.

Let us define the vectors

$$\mathbf{y}[n] = \begin{bmatrix} y(n) \\ y(n+1) \\ \vdots \\ y(n+M-1) \end{bmatrix} \text{ and } \mathbf{w}[n] = \begin{bmatrix} w(n) \\ w(n+1) \\ \vdots \\ w(n+M-1) \end{bmatrix}.$$

Then we have

$$\mathbf{y}[n] = \sum_{k=1}^P a_k e^{j\omega_k n} \cdot \mathbf{s}_k + \mathbf{w}[n],$$

where

$$\mathbf{s}_k = [1 \ e^{-j2\pi f_k} \ \dots \ e^{-j2\pi(M-1)f_k}]^H \quad \text{for } k=1, \dots, P.$$

The autocorrelation matrix is equal to

$$\begin{aligned} \mathbf{R}_{yy} &= E[\mathbf{y}[n]\mathbf{y}^H[n]] = E\left[\left(\sum_{k=1}^P a_k e^{j\omega_k n} \cdot \mathbf{s}_k + \mathbf{w}[n]\right)\left(\sum_{k=1}^P a_k^* e^{-j\omega_k n} \cdot \mathbf{s}_k^H + \mathbf{w}^H[n]\right)\right] = \\ &= E\left[\sum_{k=1}^P \sum_{j=1}^P a_k a_j^* e^{j(\omega_k - \omega_j)n} \cdot \mathbf{s}_k \mathbf{s}_j^H\right] + E[\mathbf{w}[n] \mathbf{w}^H[n]] = \\ &= \sum_{k=1}^P |a_k|^2 \mathbf{s}_k \mathbf{s}_k^H + E[\mathbf{w}[n] \mathbf{w}^H[n]] = \\ &= \underbrace{\mathbf{S} \mathbf{A} \mathbf{S}^H}_{\mathbf{R}_{xx}} + \underbrace{\sigma_w^2 \mathbf{I}}_{\mathbf{R}_{ww}} \end{aligned}$$

where $\mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2 \ \dots \ \mathbf{s}_P]$ and $\mathbf{A} = \text{diag}(|a_1|^2, \dots, |a_P|^2)$.

Signal and Noise Subspaces

It can be shown that if the frequencies f_i are all distinct modulo 2π and if the size of the autocorrelation matrix M is chosen such that $M \geq P$, then \mathbf{R}_{xx} has rank P . Let $\lambda(\mathbf{R}_{xx})$ denote the non-zero eigenvalues of \mathbf{R}_{xx} ordered so that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$, and let $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M\}$ be the associated eigenvectors. As \mathbf{R}_{xx} is a normal matrix, it can be factorized as

$$\mathbf{R}_{xx} = \mathbf{U} \Lambda \mathbf{U}^H$$

where $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_M)$ and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_M]$ is a unitary matrix, that is, $\mathbf{U}\mathbf{U}^H = \mathbf{I}$. Then, the spectral decomposition of \mathbf{R}_{xx} with rank P is equal to

$$\mathbf{R}_{xx} = \sum_{k=1}^P \lambda_k \mathbf{u}_k \mathbf{u}_k^H,$$

since only the first P eigenvalues are distinct of zero, that is $\lambda_{P+1} = \lambda_{P+2} = \dots = \lambda_M = 0$. The set of orthonormal eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P\}$ are called the *principal eigenvectors* of \mathbf{R}_{xx} and they span the signal subspace. Notice that the set of vectors $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_P\}$ span the same subspace, since $\mathbf{R}_{xx} = \mathbf{S} \mathbf{A} \mathbf{S}^H$ can be seen as a linear combination of them.

The autocorrelation matrix $\mathbf{R}_{yy} = \mathbf{R}_{xx} + \sigma_w^2 \mathbf{I}$ is full rank because $\sigma_w^2 \mathbf{I}$ is full rank. Considering $\mathbf{U}\mathbf{U}^H = \mathbf{I}$, we have

$$\mathbf{R}_{yy} = \mathbf{U} \Sigma \mathbf{U}^H$$

$$\begin{aligned} R_{yy} &= R_{xx} + \sigma_w^2 I = \sum_{k=1}^P \lambda_k u_k u_k^H + \sigma_w^2 U U^H = \\ &= \underbrace{\sum_{k=1}^P (\lambda_k + \sigma_w^2) u_k u_k^H}_{\text{Signal Subspace}} + \underbrace{\sum_{k=P+1}^M \sigma_w^2 u_k u_k^H}_{\text{Noise Subspace}} \end{aligned}$$

The signal autocorrelation is separated in two orthogonal subspaces, the signal subspace and the noise subspace. Therefore, any vector from the signal subspace, i.e. s_k , is orthogonal to the noise subspace.

$$\left. \begin{array}{c} \text{SIGNAL} \\ \text{SUBSPACE} \end{array} \right\} \Rightarrow \text{span}\{u_1, u_2, \dots, u_p\} = \text{span}\{s_1, s_2, \dots, s_p\}$$

$$\left. \begin{array}{c} \text{NOISE} \\ \text{SUBSPACE} \end{array} \right\} \Rightarrow \text{span}\{u_{p+1}, \dots, u_M\}$$

Pisarenko harmonic decomposition

Based on the observation that the signal subspace is orthogonal to the noise subspace, there are various means that can be employed to estimate the signal components in the presence of noise. In the Pisarenko harmonic decomposition, the orthogonality is exploited directly. Suppose that the number of modes P is

known. Then, setting $M = P + 1$, the noise subspace is spanned by the single vector u_M , which must be orthogonal to all of the signal subspace vectors. Therefore, we have

(Again el domínio
VEP del NOISEsubspace)

$$s_k^H u_M = \sum_{m=1}^M u_M(m) e^{-j2\pi f_k(m-1)} = 0 \quad \text{for } k = 1, \dots, P$$

where $u_M(m)$ is the m component of vector u_M .

This method uses an autocorrelation matrix of size $M = P + 1$ and estimates the frequencies as the peaks of the following frequency estimation function corresponding to the Pisarenko harmonic decomposition.

$$P_{PHD}(f) = \frac{1}{\left| \sum_{m=1}^M u_M(m) z^{(m-1)} \right|^2} \quad \text{where } z = e^{-j2\pi f}.$$

The P roots of the denominator correspond to the frequencies of the sinusoidal signal.

Once the frequencies are obtained from the roots of the polynomial, the squared amplitudes can be obtained by solving:

$$\begin{bmatrix} e^{j2\pi f_1} & e^{j2\pi f_2} & \dots & e^{j2\pi f_P} \\ e^{j2\pi f_1} & e^{j2\pi f_2} & \dots & e^{j2\pi f_P} \\ \vdots & & & \vdots \\ e^{j2\pi f_1} & e^{j2\pi f_2} & \dots & e^{j2\pi f_P} \end{bmatrix} \begin{bmatrix} a_1^2 \\ a_2^2 \\ \vdots \\ a_P^2 \end{bmatrix} = \begin{bmatrix} r_{yy}(1) \\ r_{yy}(2) \\ \vdots \\ r_{yy}(P) \end{bmatrix}. \quad (A3.3)$$

The noise strength is obtained from the M^{th} eigenvalue of \mathbf{R}_{yy} . Of course, in practice the correlation matrix \mathbf{R}_{yy} must be estimated based on received signals.

MUSIC method

MUSIC stands for Multiple Signal Classification. It relies on the fact that the signal subspace is orthogonal to the noise subspace. The main idea is really just using averaging to improve the performance of the estimator of Pisarenko. Now we choose $M \geq P+1$. Let $s^H(f)$ be a row vector that depends on f as follows

$$s^H(f) = [1 \ e^{-j2\pi f} \ e^{-j2\pi 2f} \ \dots \ e^{-j2\pi(M-1)f}]$$

When $f = f_k$ (one of the input frequencies), then for any vector x in the noise subspace we have

$$\underset{x \in \text{noise}}{x} \underset{\substack{\text{freq.} \\ \text{at } x_k}}{f=f_k} \Rightarrow s^H(f)x = 0$$

Then, the function

$$P_{\text{MUSIC}}(f) = \frac{1}{\sum_{k=P+1}^M |s^H(f)u_k|^2}$$

(from noise tots els
 VEPS del Noise
 Subspace)

is equal infinite at the frequencies $f = f_k$ since the denominator is zero. Thus, a plot of $P_{\text{MUSIC}}(f)$ should have a tall peak at $f = f_i$, one for each of the input frequencies. By locating the peaks, the frequencies can be identified as for the Pisarenko methods, see (A3.3).

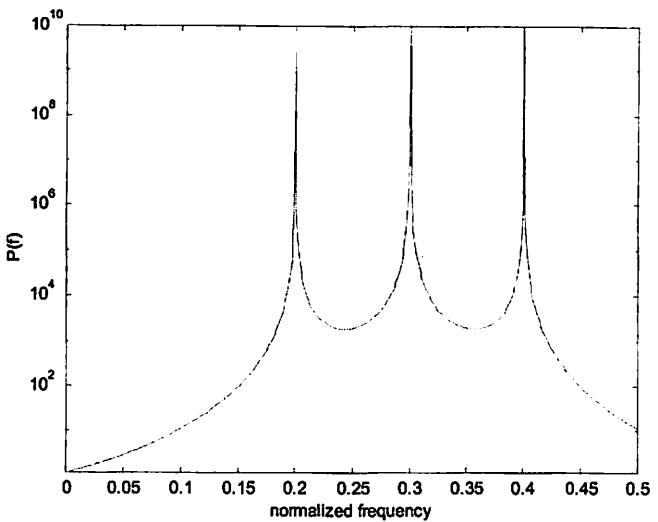
Example

A source $x(n)$ is known to produce $P=3$ sinusoids. The correlation matrix \mathbf{R}_{yy} is estimated to be

$$\mathbf{R}_{yy} = \begin{bmatrix} 6.4000 & -2.7361 + 4.6165j & -1.5000 - 3.4410j & 1.7361 + 1.0898j \\ 2.7361 - 4.6165j & 6.4000 & 2.7361 + 4.6165j & -1.5000 - 3.4410j \\ -1.5000 + 3.4410j & 2.7361 - 4.6165j & 6.4000 & 2.7361 + 4.6165j \\ 1.7361 - 1.0898j & -1.5000 + 3.4410j & 2.7361 - 4.6165j & 6.4000 \end{bmatrix}$$

The Pisarenko method results in $\sigma^2 = 0.4$ and frequencies $\{0.2, 0.3, 0.4\}$. The amplitudes are $\{1.0003, 1.9999, 2.9998\}$.

The MUSIC spectrum results as follow:



The peaks are clearly at 0.2, 0.3, and 0.4.

ESPRIT method

One application of generalized eigenvalue decompositions is the sinusoidal estimation technique ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques). The method assumes that there are p sinusoids in white noise, and deals with an eigendecomposition. Let $\mathbf{y}[n] = \mathbf{x}[n] + \mathbf{w}[n]$ be the vector result of summing the vector of signal samples and the vector of noise samples. We also introduce a new vector of delayed samples as follows

$$\mathbf{z}[n] = \mathbf{y}[n+1] = \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(n+M) \end{bmatrix}.$$

As before, we can write

$$\mathbf{R}_{yy} = E[\mathbf{y}[n]\mathbf{y}^H[n]] = \underbrace{\mathbf{S}\mathbf{A}\mathbf{S}^H}_{\mathbf{R}_{xx}} + \frac{\sigma_w^2}{\mathbf{R}_{ww}} \mathbf{I}$$

Also,

$$\mathbf{R}_{yz} = E[\mathbf{y}[n]\mathbf{z}^H[n]] = E[\mathbf{y}[n]\mathbf{y}^H[n+1]] = \mathbf{S}\mathbf{A}\psi^H\mathbf{S}^H + \mathbf{R}_w$$

where

$$\mathbf{R}_w = E[\mathbf{w}[n]\mathbf{w}^H[n+1]] = \sigma_w^{-2} \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

and $\psi \in \mathbb{C}^{p \times p}$ represents the phase shift between successive samples and appears as a result of the delayed vector,

$$\psi = \text{diag}(e^{j2\pi f_1}, e^{j2\pi f_2}, \dots, e^{j2\pi f_p}).$$

If $M \geq p$, the matrix

$$\mathbf{R}_{xx} = \mathbf{R}_{yy} - \sigma_w^{-2} \mathbf{I} = \mathbf{S} \mathbf{A} \mathbf{S}^H$$

has rank p . Let

$$\mathbf{C}_{yz} = \mathbf{R}_{yz} - \mathbf{R}_w = \mathbf{S} \mathbf{A} \psi^H \mathbf{S}^H.$$

Now consider the generalized eigenvalue problem

$$\mathbf{R}_{xx} \mathbf{u} = \lambda \mathbf{C}_{yz} \mathbf{u};$$

that is,

$$(\mathbf{R}_{xx} - \lambda \mathbf{C}_{yz}) \mathbf{u} = 0.$$

This can be written as

$$\mathbf{S} \mathbf{A} (\mathbf{I} - \lambda \psi^H) \mathbf{S}^H \mathbf{u} = 0.$$

Since ψ is diagonal, it is clear that the value of λ for what $\mathbf{I} - \lambda \psi^H = 0$ is $\lambda = e^{j2\pi f_i}$ which is an eigenvalue. From the p generalized eigenvalues that lie on the unity circle can be obtained the frequencies f_i , $i = 1, 2, \dots, p$.

Example

For the data-correlation matrix of Pisarenko and MUSIC example, the cross correlation matrix \mathbf{R}_{yz} is determined to be

$$\mathbf{R}_{yz} = \begin{bmatrix} -2.7361 - 4.6165j & 6.0000 & -2.7361 + 4.6165j & -1.5000 - 3.4410j \\ -1.1000 + 3.4410j & -2.7361 - 4.6165j & 6.0000 & -2.7361 + 4.6165j \\ 1.7361 - 1.0898j & -1.1000 + 3.4410j & -2.7361 - 4.6165j & 6.0000 \\ -1.5000 + 0.8123j & 1.7361 - 1.0898j & -1.1000 + 3.4410j & -2.7361 - 4.6165j \end{bmatrix}$$

5. SOLUTIONS OF LINEAR SYSTEMS. THEORY

5.1. EXISTENCE AND NUMBER OF SOLUTIONS TO $Ax = b$

5.2. SOLUTION OF TRIANGULAR SYSTEMS. GAUSS ELIMINATION

5.3. LU DECOMPOSITION. CHOLESKY DECOMPOSITION

Many physical problems may be formulated in terms of a set of linear equations. In fact, there is perhaps no field of mathematical inquiry in which systems of linear equations do not play an important role. We will denote a system of M equations and N unknowns as

$$A_{(M \times N)} x_{(N \times 1)} = b_{(M \times 1)} \quad (5.1)$$

The important problems for our concern are:

1. For a given matrix $A \in \mathbb{R}^{M \times N}$ and a given vector $b \in \mathbb{R}^{M \times 1}$, does there exist a vector $x \in \mathbb{R}^{M \times 1}$ that satisfies equation (5.1)?
2. If the answer to (1) is "yes", the next question is, "how many solution vectors x are there"? When can we guarantee the uniqueness of the solution? Otherwise, what can we do when the linear system has infinite solutions?
3. If the answer to (1) is "no", the next question is, "does there exist a vector x so that equation (5.1) is approximately satisfied for a suitable definition of approximate"?

If the answer to (5.1) is "yes", the system is said to be consistent; otherwise the system is said to inconsistent.

5.1. EXISTENCE AND NUMBER OF SOLUTIONS TO $Ax=b$ Theorem

A solution to the system $Ax=b$ exists if and only if the rank of the coefficient matrix A is equal to the rank of the augmented matrix $[A \ b]$.

\longleftrightarrow és a dir b és una combinació lineal de x ($Ax=b$)

Corollary

The system $Ax=b$ has a solution if and only if the vector b is in the column space of A .

Depending upon the relation between the $\text{rank}(A)$ and the $\text{rank}([A \ b])$, three possible cases seem to occur

1. Inconsistent system (no solution exists).

$$\text{rank}([A \ b]) > \text{rank}(A) \quad b \text{ no c. l. de } x$$

2. Consistent with a unique solution

$$\text{rank}([A \ b]) = \text{rank}(A) = N \quad b \text{ c. l. de } x$$

3. Consistent with ∞ solutions

$$\text{rank}([A \ b]) = \text{rank}(A) < N \quad \begin{matrix} \text{les columnas de } A \text{ són linealment} \\ \text{dependent de les altres} \end{matrix}$$

Example

Consider the system of equations $Ax=b$ below with α an arbitrary parameter.

$$x - 3y = -2$$

$$2x + y = 3$$

$$3x - 2y = \alpha$$

Therefore,

$$A = \begin{bmatrix} 1 & -3 \\ 2 & 1 \\ 3 & -2 \end{bmatrix} \quad b = \begin{bmatrix} -2 \\ 3 \\ \alpha \end{bmatrix} \quad [A \ b] = \begin{bmatrix} 1 & -3 & -2 \\ 2 & 1 & 3 \\ 3 & -2 & \alpha \end{bmatrix}$$

The determinant of the augmented matrix is $\det([A \ b]) = 7\alpha - 7$. Then,

- If $\alpha = 1$, then the ranks of both A and $[A \ b]$ are equal to $N=2$. In this case we have a unique solution
- If $\alpha \neq 1$, then the $\text{rank}([A \ b]) = 3$ and $\text{rank}(A) = 2$. The system is inconsistent and there are no solutions.

5.2. SOLUTION OF TRIANGULAR SYSTEMS. GAUSS ELIMINATION

If the coefficient matrix $A \in \mathbb{R}^{N \times N}$ is square and either upper-triangular or lower-triangular, the solution can be computed with the back-substitution algorithm or the forward elimination algorithm, respectively.

- **Upper-triangular case**

Consider the linear system

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ 0 & u_{22} & \cdots & u_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Back-substitution algorithm:

$$\begin{aligned} x_N &= \frac{y_N}{u_{NN}} \\ &\vdots \quad \vdots \\ x_n &= \frac{(y_n - \sum_{i=n+1}^N u_{ni} \cdot x_i)}{u_{nn}} \quad \text{for } n=N-1, \dots, 1 \end{aligned}$$

It needs $O\left(\frac{N^2}{2}\right)$ multiplications $N + [1+2+\dots+(N-1)] \otimes = \frac{N(N+1)}{2}$

- **Lower-triangular case**

Consider the linear system

$$\begin{bmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{N1} & l_{N2} & \cdots & l_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

Forward elimination algorithm:

$$\begin{aligned} x_1 &= \frac{y_1}{l_{11}} \\ &\vdots \quad \vdots \\ x_n &= \frac{(y_n - \sum_{i=1}^{n-1} l_{ni} \cdot x_i)}{l_{nn}} \quad \text{for } n=2, \dots, N \end{aligned}$$

It also needs $O\left(\frac{N^2}{2}\right)$ multiplications.

GAUSS ELIMINATION → si \underline{A} no es triangular ←

In general, the matrix A is not triangular and we need to develop other techniques to solve it. The method of Gauss elimination is a straightforward but powerful procedure for reducing systems of linear equations to a upper-triangular form easily solved by back-substitution.

Example:

$$\begin{array}{l} x_1 - 1x_2 + 2x_3 + 3x_4 = 2 \\ -3x_1 + 6x_2 + 3x_3 - 15x_4 = -3 \\ 5x_1 - 8x_2 - x_3 + 17x_4 = 17 \\ x_1 + x_2 + 11x_3 + 7x_4 = 7 \end{array} \left. \begin{array}{l} a \\ b \\ c \\ d \end{array} \right\} \Rightarrow \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ -3 & 6 & 3 & -15 & -3 \\ 5 & -8 & -1 & 17 & 9 \\ 1 & 1 & 11 & 7 & 7 \end{array} \right] \begin{array}{l} 3a+d \\ -5a+c \\ -a+d \end{array}$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 3 & 9 & -6 & 3 \\ 0 & -3 & -11 & 2 & -1 \\ 0 & 2 & 9 & 4 & 5 \end{array} \right] \xrightarrow{-3} \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 1 & 3 & -2 & 1 \\ 0 & -3 & -11 & 2 & -1 \\ 0 & 2 & 9 & 4 & 5 \end{array} \right] \begin{array}{l} 3b+c \\ -2b+d \end{array}$$

$$\left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 1 & 3 & -2 & 1 \\ 0 & 0 & -2 & -4 & 2 \\ 0 & 0 & 3 & 8 & 3 \end{array} \right] \xrightarrow{-2} \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 1 & 3 & -2 & 1 \\ 0 & 0 & 1 & 2 & -1 \\ 0 & 0 & 3 & 8 & 3 \end{array} \right] \xrightarrow{-3c+d} \left[\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 2 \\ 0 & 1 & 3 & -2 & 1 \\ 0 & 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 & 3 \end{array} \right]$$

This completes the Gauss elimination and we now may use the back-substitution algorithm to produce the solution. \square

The Gauss elimination procedure discovers all (if any) the solutions to any concrete system of equations.

Example

$$\begin{array}{l} x_1 + 2x_2 - 5x_3 = 2 \\ 2x_1 - 3x_2 + 4x_3 = 4 \\ 4x_1 + x_2 - 6x_3 = 8 \end{array} \Rightarrow \left[\begin{array}{ccc|c} 1 & 2 & -5 & 2 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

In this case, the system is consistent with infinite solutions of the type

$$x_1 = 2 + \alpha ; x_2 = 2\alpha ; x_3 = \alpha$$

with α an arbitrary parameter. \square

Example

$$\begin{array}{l} x_1 + 2x_2 - x_3 + 2x_4 = 4 \\ 2x_1 + 7x_2 + x_3 + x_4 = 14 \\ 3x_1 + 8x_2 - x_3 + 4x_4 = 17 \end{array} \Rightarrow \left[\begin{array}{cccc|c} 1 & 2 & -1 & 2 & 4 \\ 0 & 1 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right]$$

In this case the system consists of 3 equations and 4 unknowns. It also has infinite solutions. \square

The Gauss elimination process is equivalent to writing A as a product LU of a lower-triangular L and an upper-triangular U .

5.3. LU DECOMPOSITION

We have seen that if the coefficient matrix A can be reduced to a triangular system, the solution is trivial. In this section we present the LU decomposition of a matrix A into the product LU of a lower-triangular L and an upper-triangular U . Then,

$$\boxed{Ax = b \Rightarrow LUx = b}$$

The LU decomposition is useful to reduce a linear system into two triangular systems:

$$\boxed{Ly = b \text{ and } Ux = y.}$$

LU DECOMPOSITION WITHOUT INTERCHANGES

The procedure applies $(N-1)$ transformations

$$\boxed{M_{N-1} \cdots M_2 M_1 \cdot A = U} \Rightarrow \boxed{L^{-1} A = U} \Rightarrow \boxed{A = LU}$$

where the Gaussian transformation M_k is equal to

$$\boxed{M_k = I - v^{(k)} \cdot e_k^T}$$

with

$$\boxed{\begin{aligned} v^{(k)} &= [0 \ \dots \ 0 \ \ell_{k+1,k} \ \dots \ \ell_{kk}]^T \quad \ell_{i,k} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \text{ for } i=k+1, \dots, N \\ e_k &= [0 \ \dots \ 1 \ \dots \ 0]^T \end{aligned}}$$

The diagonal element of A , $a_{kk}^{(k-1)}$, is the pivot. We'll see later how the procedure changes if the pivot is zero.

In the first step, we have

$$\begin{aligned} v^{(1)} &= [0 \ \ell_{21} \ \dots \ \ell_{2N}]^T \quad \ell_{i,1} = \frac{a_{i1}}{a_{11}} \text{ for } i=2, \dots, N \\ M_1 &= \begin{bmatrix} 1 & 0 & \dots & 0 \\ -\ell_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\ell_{2N} & 0 & \dots & 1 \end{bmatrix} \Rightarrow M_1 A = \begin{bmatrix} a_{11}^{(1)} & | & A_{12}^{(1)} \\ 0 & | & \vdots \\ \vdots & | & A_{22}^{(1)} \\ 0 & | & \vdots \end{bmatrix} \end{aligned}$$

where the vector $A_{12}^{(1)}$ has been modified. Since the matrices M_k are lower-triangular, their product is lower-triangular as well, and then

$$L^{-1} A = U.$$

In general, in the k^{th} step we have

$$M_k A^{(k-1)} = M_k \cdots M_1 A = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix}_{(k) \times (N-k)}$$

where $A_{11}^{(k)}$ is upper-triangular and

$$M_k = \begin{bmatrix} I_k & & 0 \\ & 1 & 0 & \cdots & 0 \\ 0 & -l_{k+1,k} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & -l_{N,k} & 0 & \cdots & 1 \end{bmatrix} \quad \left. \begin{array}{l} l_{i,k} = \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} \\ M_{N-1} \cdots M_2 M_1 = L^{-1} \\ A = LU \end{array} \right\} \text{LU decomposition}$$

□

The LU decomposition could become unstable (moreover, could not exist) if the pivot $a_{kk}^{(k-1)}$ is near zero. Next theorem tells us when the LU decomposition exists.

Theorem

The LU decomposition of a matrix $A \in R^{N \times N}$ exists if and only if the principal submatrices are non singular. That is, if $\det(A[1:k, 1:k]) \neq 0$ for $k=1, \dots, N-1$.

(For rectangular matrices, we have that the LU decomposition of a matrix $A \in R^{M \times N}$ exists if and only if the principal submatrices are non singular. That is, if $\det(A[1:k, 1:k]) \neq 0$ for $k=1, \dots, \min(M, N)$. (For details, see [Golub, p.102]).

This condition guarantees all the pivots will be zero and, therefore, the LU decomposition may be computed as explained before. Note that, in general, the LU decomposition is not unique.

If any pivot is zero, then the LU decomposition may be obtained if we allow row interchanges in order to avoid null pivots.

LU DECOMPOSITION WITH INTERCHANGES

If we allow row permutations before applying the Gaussian transformation M_k , the LU decomposition becomes

$$\boxed{M_{N-1} P_{N-1} \cdots M_2 P_2 M_1 P_1 \cdot A = U} \quad (5.2)$$

where P_k is a row permutation matrix. In the procedure, the permutations should be kept to solve the linear system of equations properly.

To obtain the lower-triangular matrix L from equation (5.2) we proceed as follows. First, considering that $P_i^{-1} = P_i$, we have

$$P_{N-2} M_{N-3} \cdots M_2 P_2 M_1 P_1 \cdot A = M_{N-2}^{-1} P_{N-1} M_{N-1}^{-1} U. \quad (5.3)$$

Matrix M_{N-1}^{-1} is lower-triangular, but $M_{N-2}^{-1} P_{N-1}$ is not. However, if we multiply (5.3) on the right-hand side by P_{N-1} , we obtain

$$P_{N-1} P_{N-2} M_{N-3} \cdots M_2 P_2 M_1 P_1 \cdot A = (P_{N-1} M_{N-2}^{-1} P_{N-1}) (M_{N-1}^{-1}) U$$

with both $P_{N-1} M_{N-2}^{-1} P_{N-1}$ and the product $(P_{N-1} M_{N-2}^{-1} P_{N-1}) (M_{N-1}^{-1})$ lower-triangular.

A new iteration leads us to

$$P_{N-1} P_{N-2} P_{N-3} M_{N-4} \cdots M_2 P_2 M_1 P_1 \cdot A = \underbrace{(P_{N-1} P_{N-2} M_{N-3}^{-1} P_{N-2} P_{N-1})}_{L_{N-3}} \underbrace{(P_{N-1} M_{N-2}^{-1} P_{N-1})}_{L_{N-2}} \underbrace{(M_{N-1}^{-1})}_{L_{N-1}} U$$

At the last iteration, we obtain the permutation matrix P equal to

$$P = P_{N-1} P_{N-2} \cdots P_1,$$

the lower-triangular matrix L of the LU decomposition equal to

$$L = L_1 L_2 \cdots L_{N-1}$$

where L_i defined as follows

$$L_i = (P_{N-1} P_{N-2} \cdots P_{i+1} \cdot M^{-1} \cdot P_{i+1} \cdots P_{N-2} P_{N-1}).$$

From the computational point of view, two aspects are relevant:

- The permutations do not involve computational load.
- Since $M_k = I - v^{(k)} \cdot e_k^T$, then $M_k^{-1} = I + v^{(k)} \cdot e_k^T$.

Example

$$A = \begin{bmatrix} 2 & 8 & -4 \\ -4 & -12 & 11 \\ 3 & 14 & -16 \end{bmatrix}$$

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$X = P_1 * A = \begin{bmatrix} -4 & -12 & 11 \\ 2 & 8 & -4 \\ 3 & 14 & -16 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 1.0000 & 0 & 0 \\ 0.5000 & 1.0000 & 0 \\ 0.7500 & 0 & 1.0000 \end{bmatrix}$$

$$\begin{aligned} X = M_1 * X = \\ -4.0000 & -12.0000 & 11.0000 \\ 0 & 2.0000 & 1.5000 \\ 0 & 5.0000 & -7.7500 \end{aligned}$$

$$\begin{aligned} P_2 = \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{aligned}$$

$$\begin{aligned} X = P_2 * X = \\ -4.0000 & -12.0000 & 11.0000 \\ 0 & 5.0000 & -7.7500 \\ 0 & 2.0000 & 1.5000 \end{aligned}$$

$$\begin{aligned} M_2 = \\ 1.0000 & 0 & 0 \\ 0 & 1.0000 & 0 \\ 0 & -0.4000 & 1.0000 \end{aligned}$$

$$\begin{aligned} U = M_2 * X = M_2 * P_2 * M_1 * P_1 * A = \\ -4.0000 & -12.0000 & 11.0000 \\ 0 & 5.0000 & -7.7500 \\ 0 & 0 & 4.6000 \end{aligned}$$

$$\begin{aligned} L = P_2 * \text{inv}(M_1) * P_2 * \text{inv}(M_2) = \\ 1.0000 & 0 & 0 \\ -0.7500 & 1.0000 & 0 \\ -0.5000 & 0.4000 & 1.0000 \end{aligned}$$

$$\begin{aligned} P = P_2 * P_1 = \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{aligned}$$

$$\begin{aligned} P * A = \\ -4 & -12 & 11 \\ 3 & 14 & -16 \\ 2 & 8 & -4 \end{aligned}$$

$$\begin{aligned} L * U = \\ -4 & -12 & 11 \\ 3 & 14 & -16 \\ 2 & 8 & -4 \end{aligned}$$

SOLVING A GENERAL LINEAR SYSTEMS WITH THE LU DECOMPOSITION:

The procedure based on the LU decomposition to solve the linear system of equations given by

$$Ax = b,$$

is as follows:

1. Compute the LU decomposition with interchanges.

$$PA = LU.$$

2. New linear system of equations

$$PAx = L \underbrace{Ux}_{y} = Pb.$$

3. Solve the lower-triangular system (forward elimination):

$$Ly = Pb.$$

4. Solve the upper-triangular system (back-substitution):

$$Ux = y.$$

to obtain \underline{x}

Computational aspects.

It can be seen that the computational burden is equal to

- Cost of $PA = LU$

$$N^3/3 - N/3 \text{ multiplications and divisions}$$

$$N^3/3 - N^2/2 + N/6 \text{ additions and subtractions}$$

- Cost of $Ly = Pb$ and $Ux = y$

$$N^2 \text{ multiplications and divisions}$$

$$N^2 - N \text{ additions and subtractions}$$

Therefore, the computational burden of solving a linear system of N equations with N unknowns is $O(N^3/3)$ operations.

In fact, it is known that no method using only row and column operations can take less work on general matrices than the LU decomposition or the Gauss elimination.

CHOLESKY DECOMPOSITION

For hermitian matrices, the LU decomposition becomes

$$A = LDL^H$$

where D is a diagonal matrix. Moreover, if A is hermitian and positive definite, then all the eigenvalues are nonnegative and we get the Cholesky decomposition

$$A = LDL^H = GG^H$$

with G lower-triangular.

$$\text{Cholesky : } \begin{aligned} & * \quad A^H = A \quad \rightarrow \quad A = \underbrace{L}_{\substack{\text{diagonal}}} \underbrace{D}_{\substack{\uparrow \\ \text{lower triangular}}} \underbrace{L^H}_{\substack{\leftarrow \\ \text{lower triangular}}} \\ & * \quad \begin{aligned} & A^H = A \\ & x^H A x \geq 0 \end{aligned} \quad \left. \begin{aligned} & \} \rightarrow A = \underbrace{G}_{\substack{\text{lower triangular}}} \underbrace{G^H}_{\substack{\leftarrow \\ \text{lower triangular}}} \\ & (\text{positive def.}) \end{aligned} \right. \end{aligned}$$

{ des del punt de vista
 de resolució de sist. d'eq.
 No com a estimador

6. LEAST SQUARES PROBLEM

6.1. INTRODUCTION

6.2. THE LEAST SQUARES PROBLEM

6.3. THE MOORE-PENROSE GENERALIZED INVERSE

6.4. MINIMUM NORM SOLUTION TO THE UNDERDETERMINED LEAST SQUARES

6.5. EFFICIENT LEAST SQUARES COMPUTATION

6.6. WEIGHTED LEAST SQUARES PROBLEM

6.7. REDUCED-RANK LEAST SQUARES PROBLEM

6.8. TOTAL LEAST SQUARES PROBLEM

6.1 INTRODUCTION

Consider the set of linear equations:

$$A_{(M \times N)} x_{(N \times 1)} = b_{(M \times 1)}$$

If the matrix A is squared ($M=N$) and is full rank, then the unique solution for the unknowns x can be found as

$$x = A^{-1}b$$

However, in all other cases the solution is not so straightforward:

- 1. If the matrix is "tall" ($M>N$), has full column rank ($\text{rank}(A)=N$) and the system is consistent, it has a unique solution, but A^{-1} is not defined for non-squared matrices...
- 2. If the matrix does not have full column rank ($\text{rank}(A)<N$) and the system is consistent, then infinite solutions exist, so the general expression for all of them must be found and a criterion must be defined to select one of them.
- 3. If the matrix has full column rank ($\text{rank}(A)=N$) and the system is inconsistent, it has no solution. Hence, a criterion must be defined to choose what is the vector x that provides the best fit to the set of equations in an approximate sense.
- 4. If the matrix does not have full column rank ($\text{rank}(A)<N$) and the system is inconsistent, it has no solution. Hence, a criterion must be defined to choose what is the vector x that provides the best fit to the set of equations in an approximate sense. However, it will be shown that many (infinite) vectors may provide the same "degree of fit", so the general expression for all of them must be found and a criterion must be defined to select one of them.

In all these cases, the tools provided in this chapter will be useful: Least Squares solution (3,4), Moore-Penrose generalized inverse (1,2,3,4), minimum-norm solution (2,4).

↗
 solucions
 ← tables

quan \underline{b} no pertany al subespai generat per les columnes de \underline{A}

volem trobar una solució a $\underline{A}\underline{x} = \underline{b}$ el q. es fa
en perturbar \underline{b} per t.g. $(\underline{b} + \Delta\underline{b})$ pertangi al subespai
generat per \underline{A}

Consider the following optimization problem:

$$\min_{\hat{x}} \|A\hat{x} - b\|_p$$

or equivalently to the problem

$$\min_{\hat{b}: \hat{b}=A\hat{x}} \|\hat{b} - b\|_p$$

Different norms render different optimum solutions. E.g:

Example:

$$\begin{aligned} A &= [1 \ 1 \ 1]^T & \Rightarrow & p=1: \quad x_{\text{opt}} = b_2 \\ b &= [b_1, b_2, b_3]^T, \text{ with } b_1 \geq b_2 \geq b_3 \geq 0 & p=2: \quad x_{\text{opt}} = (b_1 + b_2 + b_3)/3 \\ & & p=\infty: \quad x_{\text{opt}} = (b_1 + b_3)/3 \end{aligned}$$

However, optimization in the 1-norm and the ∞ -norm is complicated because the function $\|A\hat{x} - b\|_p$ is not differentiable for these values of p .

In contrast, the 2-norm problem is more tractable because:

- $\|\hat{x}\|^2$ is a differentiable function \Rightarrow gradient algorithms can be applied for minimization
- The 2-norm is preserved under unitary transformation, so we can formulate the problem as

$$\min_{\hat{x}} \|Q\hat{x} - Qb\|_2$$

and we can select a unitary transformation Q such that the new problem is easy to solve.

So hereafter we will only deal with the 2-norm case.

2-norm: funció derivable amb un únic mínim.

The Least Squares problem can be formulated as

$$\min_{\hat{x}} \|A\hat{x} - b\|_2^2 = \min_{\hat{x}} \|e(x)\|_2^2 = \min_{\hat{x}} \sum_{i=1}^M |e_i|^2 \quad e(x) = A\hat{x} - b \quad (1)$$

or as

$$\min_{\hat{b}: \hat{b}=A\hat{x}} \|\hat{b} - b\|_2^2 \quad (2)$$

Equation (1) originates the name Least squares and can be regarded as an inversion or a parameter estimation problem. Equation (2) can be regarded as a filtering or a signal extraction problem, the solutions \hat{b} and \hat{x} to both problems are related as $\hat{b} = A\hat{x}$. Note that the degree of fit in both cases (i.e. $x - \hat{x}$ vs $Ax - \hat{b}$) may not be the same.

It will be shown in this chapter that

... the solution to $\min_{\hat{x}} \|A\hat{x} - b\|_2$ is provided by the Moore-Penrose generalized inverse

... the solution to $\min_{\hat{b}: \hat{b}=A\hat{x}} \|\hat{b} - b\|_2$ is provided by the projection matrix P_A

- Usos:
- ajustar unes mesures a una certa corba (recta, paràbola, ...) minimitzant el MSE de les mesures.
 - estimar els paràmetres d'un model linear amb soroll blanc gaussià

$$\underline{x} \rightarrow \boxed{A} \xrightarrow{\oplus} \underline{y} = A\underline{x} + \underline{e} \Rightarrow \text{ML: } \max_{\underline{x}} f_{\underline{y}|\underline{x}}(\underline{y}|\underline{x})$$

↑

gaussian noise: $f_{\underline{e}}(\underline{e}) = \frac{1}{\pi^n |C_e|} e^{-\underline{e}^T C_e^{-1} \underline{e}}$

$$\text{ML: } \max_{\underline{x}} f_{\underline{e}}(\underline{y} - A\underline{x}) = K e^{-(\underline{y} - A\underline{x})^T C_e^{-1} (\underline{y} - A\underline{x})}$$

($\min_{\underline{x}} (\underline{y} - A\underline{x})^T C_e^{-1} (\underline{y} - A\underline{x})$)

soroll blanco $\underline{e} = \sigma^2 \underline{I}$ $\min_{\underline{x}} (\underline{y} - A\underline{x})^T (\underline{y} - A\underline{x}) = \min_{\underline{x}} \|\underline{y} - A\underline{x}\|_2^2$

unitaria $\sigma^2 = 1$

6.2. THE LEAST SQUARES PROBLEM

Let the linear model be

$$\mathbf{b} = \mathbf{A} \mathbf{x} + \mathbf{e} = \mathbf{y} + \mathbf{e}$$

Where $\mathbf{e} = \mathbf{b} - \mathbf{A} \mathbf{x}$ is the measurement error we do not know. We try to fit \mathbf{x} to the observation \mathbf{b}

The Least Squares formulation looks for the vector \mathbf{x} that minimizes the norm of this error vector:

$$\|\mathbf{e}\|_2^2 = \|\mathbf{Ax} - \mathbf{b}\|_2^2 = (\mathbf{Ax} - \mathbf{b})^H (\mathbf{Ax} - \mathbf{b}) = \mathbf{x}^H \mathbf{A}^H \mathbf{Ax} + \mathbf{b}^H \mathbf{b} - \mathbf{x}^H \mathbf{A}^H \mathbf{b} - \mathbf{b}^H \mathbf{Ax}$$

The optimum vector can be found through a gradient algorithm:

$$\nabla_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2 = 0 \Rightarrow \mathbf{A}^H \mathbf{Ax}_{opt} - \mathbf{A}^H \mathbf{b} = 0 \Rightarrow \boxed{\mathbf{A}^H \mathbf{Ax}_{opt} = \mathbf{A}^H \mathbf{b}}$$

These are the so-called normal equations.

If $\mathbf{A}^H \mathbf{A}$ is invertible (i.e. if \mathbf{A} has full column rank, $\text{rank}(\mathbf{A}) = N \leq M$), then

$$\mathbf{x}_{opt} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$$

$$\mathbf{b}_{opt} = \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b} = \mathbf{P}_A \mathbf{b}$$

$$\mathbf{e} = \mathbf{b} - \mathbf{b}_{opt} = \mathbf{b} - \mathbf{A} (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b} = (\mathbf{I} - \mathbf{P}_A) \mathbf{b} = \mathbf{P}_A^\perp \mathbf{b}$$

Recordatori

$$\frac{\partial \mathbf{x}^H \mathbf{C} \mathbf{x}}{\partial \mathbf{x}^H} = \mathbf{C} \mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{x}^H} \mathbf{x}^H \mathbf{a} = \mathbf{a}$$

$$\frac{\partial}{\partial \mathbf{x}^H} \mathbf{a}^H \mathbf{x} = 0$$

where \mathbf{P}_A , \mathbf{P}_A^\perp are the matrix that project onto the subspace spanned by the columns of \mathbf{A} and its orthogonal complement.

\mathbf{P}_A : matríc de projecció al subespai generat per les columnes de \mathbf{A}

Graphical interpretation: Orthogonality

$$\mathbf{b} = \mathbf{A} \mathbf{x} + \mathbf{e}$$

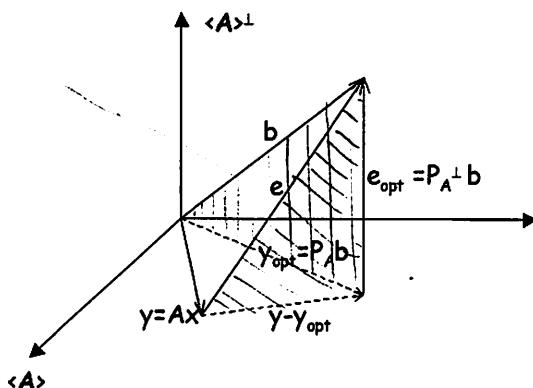
$3 \times 1 \quad 3 \times 2 \times 1 \quad 3 \times 1$

$$\mathbf{x}_{opt} = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}$$

$\mathbf{y}_{opt} = \mathbf{P}_A \mathbf{b}$ ← projecto \mathbf{b} a $\langle \mathbf{A} \rangle$ per obtenir \mathbf{b}_{opt}

$\mathbf{e}_{opt} = \mathbf{b} - \mathbf{b}_{opt} = \mathbf{P}_A^\perp \mathbf{b}$ ← l'error mínim més \perp a $\langle \mathbf{A} \rangle$

\mathbf{b}_{opt} i \mathbf{y}_{opt} pertanyen al subespai generat per \mathbf{A}



$$\mathbf{b} = \mathbf{P}_A^\perp \mathbf{b} + \mathbf{P}_A \mathbf{b} = \mathbf{e}_{opt} + \mathbf{y}_{opt}$$

$$\mathbf{e} = \mathbf{P}_A^\perp \mathbf{e} + \mathbf{P}_A \mathbf{e} = \mathbf{e}_{opt} + (\mathbf{y} - \mathbf{y}_{opt}) \Rightarrow \|\mathbf{e}_{opt}\| \leq \|\mathbf{e}\|$$

↑
proof. Pitagoras.

! La solució ha de pertànyer al subespai generat per les columnes de \mathbf{A} . LS troba el vector q. hi pertany i té distància mínima a \mathbf{b} , \mathbf{y}_{opt} .

$$\begin{aligned} c^2 &= a^2 + b^2 \\ c^2 &= (\mathbf{y} - \mathbf{y}_{opt})^2 + \mathbf{e}_{opt}^2 \Rightarrow \mathbf{e}_{opt}^2 \geq \mathbf{e}^2 \end{aligned}$$

6.3. THE MOORE-PENROSE GENERALISED INVERSE

Let $A_{(M \times N)}$ be a generic matrix. The Moore-Penrose generalised inverse, also called pseudoinverse and denoted $A^\#$ is defined as the unique matrix that satisfies the Moore-Penrose conditions:

$$\underline{A}^t = \underline{A}^\#$$

- 1) $A A^\# A = A$
- 2) $A^\# A A^\# = A^\#$
- 3) $A^\# A$ is Hermitian
- 4) $A A^\#$ is Hermitian

Alternatively, this matrix can also be defined as the linear operator $A^\#$ that maps vector b to the minimum norm vector x such that minimizes $\|Ax - b\|_2$.

If matrix A has full column rank, the pseudoinverse can be expressed as:

$$(M > N) \quad A^\# = (A^H A)^{-1} A^H$$

Otherwise, it can be defined in terms of the Singular Value Decomposition of matrix A : Let $\text{rank}(A)=r$ and define $\Sigma_1 = \text{diag}(\sigma_1, \dots, \sigma_r) \in R^{r \times r}$, $\Sigma_2 = \text{diag}(0, \dots, 0) \in R^{(M-r) \times (N-r)}$, then

$$A = U \Sigma V^H = [U_A \ U_\perp] \begin{bmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_0 \end{bmatrix} [V_A \ V_\perp]^H = U_A \Sigma_+ V_A^H$$

$$\Downarrow$$

$$A^\# = V \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & \Sigma_0^T \end{bmatrix} U^H = V_A \Sigma_+^{-1} U_A^H$$

Σ_+ - VAPS no nuls
- sempre quadrada

Σ_0 - matriz de 0's
- no quadrada en general

The Moore-Penrose pseudoinverse satisfies the following properties (among others):

- 1) A square ($M=N$), full rank $\Rightarrow A^\# = A^{-1}$
- 2) A fat ($M \leq N$), full row rank $\Rightarrow A^\# = A^H (A A^H)^{-1}$
- 3) A tall ($M \geq N$), full column rank $\Rightarrow A^\# = (A^H A)^{-1} A^H$
- 4) A is Hermitian $\Rightarrow A^\#$ is Hermitian
- 5) $(A^\#)^\# = A$

una matriz de proyección tiene r
VAPS = 1

- 6) $\text{rank}(A^\#) = \text{rank}(A)$
- 7) $A A^\# = P_A$ and $A^\# A = P_{A^H}$

$$A A^\# = U \begin{bmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_0 \end{bmatrix} V^H V \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & \Sigma_0^T \end{bmatrix} U^H = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^H = U_A U_A^H$$

$$A^\# A = V \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & \Sigma_0^T \end{bmatrix} U^H U \begin{bmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_0 \end{bmatrix} V^H = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^H = V_A V_A^H$$

Applying the pseudoinverse definition to the Least Squares solution we get that $x \in \text{span}(V_A)$, $y_{\text{opt}} \in \text{span}(U_A)$, $e_{\text{opt}} \in \text{span}(V_\perp)$,

$$x_{\text{opt}} = A^\# b = V_A \Sigma_+^{-1} U_A^H b = \sum_{i=1}^r v_i \frac{1}{\sigma_i} u_i^H b$$

$$y_{\text{opt}} = P_A b = U_A U_A^H b = \sum_{i=1}^r u_i u_i^H b \quad e_{\text{opt}} = P_A^\perp b = U_\perp U_\perp^H b = \sum_{i=r+1}^{\min(M,N)} u_i u_i^H b$$

6.4. MINIMUM NORM SOLUTION TO THE UNDETERMINED LEAST SQUARES

Particular formulation: Consider the consistent set of linear equations:

pero no col-rank

$$Ax=b$$

- el sistema tiene ∞ soluciones.
- A no es de rango completo

Where A is a full row-rank matrix. The vector x that satisfies this set of equations and has minimum norm can be found as the solution to the optimisation problem:

$$\min_x \|x^H x\| \text{ subject to } Ax = b$$

Using Lagrange multipliers and equating the gradient to zero we can solve for x :

$$\frac{\partial}{\partial x^H} (x^H x - (Ax - b)^H \lambda) = 0$$

$$x - A^H \lambda = 0 \Rightarrow x = A^H \lambda$$

Imposing the constraint:

$$Ax - b = AA^H \lambda - b = 0 \Rightarrow \lambda = (AA^H)^{-1} x$$

$$x = A^H \lambda = A^H (AA^H)^{-1} b = A^\# b$$

la solución es la pseudoinversa

The other solutions \tilde{x} to this set of equations are generated adding arbitrary components in $\text{span}(V_\perp)$ to the minimum norm solution:

$$\begin{cases} \tilde{x} = x + w \\ A\tilde{x} = b \end{cases} \Rightarrow Aw = 0 \Rightarrow w \in \text{span}(V_\perp)$$

General formulation: Consider the set of linear equations:

$$Ax=b$$

- A es full-rank (m rows n cols.)
- ∞ soluciones
- el caso general

Where A is an arbitrary matrix with $\text{rank}(A)=r$. In order to derive the minimum norm vector x that minimizes the error we express the problem in terms of the SVD of matrix A :

$$\|Ax - b\|^2 = \|\Sigma V^H x - b\|^2 = \|\Sigma V^H x - U^H b\|^2 \quad \text{de multiplicar por } U^H \text{ (pq. es unitaria)}$$

$$\Downarrow z = V^H x, c = U^H b$$

$$\|Ax - b\|^2 = \| \Sigma z - c \|^2 = \sum_{j=1}^r |\sigma_j z_j - c_j|^2 + \sum_{j=r+1}^{\min(M,N)} |c_j|^2$$

where the second term is zero when the set of equations is consistent and, therefore, $b \in \text{span}(A)$. The error is minimized choosing

$$z_j = \frac{c_j}{\sigma_j} \quad j = 1, \dots, r$$

and letting z_j take arbitrary values for $j=r+1, \dots, \min(M,N)$. Since $\|x\| = \|z\|$ is minimized over all such solutions by setting $z_j=0$ for $j=r+1, \dots, \min(M,N)$ we get that

$$z = \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} c \Rightarrow V^H x = \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H b \Rightarrow x = A^\# b \quad \text{la solución torna a ser la pseudoinversa.}$$

The other vectors \tilde{x} that also minimize the 2-norm are also generated adding arbitrary components in $\text{span}(V_\perp)$ to the minimum norm solution.

6.5. EFFICIENT LEAST SQUARES COMPUTATION

- * La solució pseudoinversa de negades NO és fàcil d'implementar \hat{A}^{-1} -inversa $[(A^H A)^{-1} A^H]$
- * Mètodes pràctics:

Descomposició QR: $A = QR$ on $x = R^{-1}Q^H b$

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \quad Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}$$

$$\|Ax - b\|^2 = \|QRx - b\|^2 = \|R_x - Q^H b\|^2 = \|R_x - \underbrace{Q^H b}_{\in \mathbb{C}^m}\|^2 = \|R_x - c_1\|^2 + \|c_2\|^2$$

↓
és molt + fàcil invertir una triangular

6.6. WEIGHTED LEAST SQUARES PROBLEM

- tenim un conjunt de dades, però n hi ha de + fiables q. d'altres.

The weighted least squares problem can be stated as follows:

$$\min_x e^H W e = (Ax - b)^H W (Ax - b)$$

where W is a square (positive) definite matrix. This problem reduces to the least squares formulation for $W=I$.

This formulation can be useful in ML estimation dealing with coloured Gaussian noise or in problems where the data in b have different reliability. In the later case the weighting matrix will be diagonal, so

$$e^H W e = (Ax - b)^H W (Ax - b) = \sum_{j=1}^m w_j |e_j|^2$$

- W en general NO diagonal
(i.e. sense colores)

Minimizing the error norm leads to the weighted normal equations:

$$A^H W A x = A^H W b$$

If matrix $A^H W A$ is invertible (A has full column rank and W is full rank) then the weighted least squares solution is

$$x = (A^H W A)^{-1} A^H W b$$

This formulation can also be used to find iteratively the vector x minimizing $\|A^* \hat{x} - b\|_p$ for other values of p -norm.

6.7. REDUCED-RANK LEAST SQUARES PROBLEM

Motivation:

Consider the following signal model:

$$\mathbf{y} = \mathbf{x} + \mathbf{n} = \mathbf{A}\boldsymbol{\theta} + \mathbf{n} \quad \mathbf{n} : N(\mathbf{0}, \sigma^2 \mathbf{I})$$

notas:
 - cuando generación
 - en linear
 - scroll incorrecto → único parámetro
 probabilístico

where \mathbf{x} is an unknown deterministic vector. Then, assuming \mathbf{A} has full column rank, the LS estimate of \mathbf{x} and $\boldsymbol{\theta}$ are distributed as:

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^\# \mathbf{y} \Rightarrow \hat{\boldsymbol{\theta}} : N(\mathbf{A}^\# \mathbf{A}\boldsymbol{\theta}, \sigma^2 \mathbf{A}^\# \mathbf{A}^\# H) \quad \text{verse biasa } (\mathbf{A}^\# \mathbf{A} \neq \mathbf{I})$$

$$\hat{\mathbf{x}} = \mathbf{P}_A \mathbf{y} \Rightarrow \hat{\mathbf{x}} : N(\mathbf{A}\mathbf{A}^\# \mathbf{x}, \sigma^2 \mathbf{P}_A)$$

and the total variance of each one is

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\} = \text{tr}[\sigma^2 \mathbf{V}_A \Sigma_+^{-2} \mathbf{V}_A^H] = \sigma^2 \text{tr}[\Sigma_+^{-2}] = \sigma^2 \sum_{j=1}^r \frac{1}{\sigma_j^2}$$

$$E\{(\hat{\mathbf{x}} - \mathbf{A}\mathbf{x})^H (\hat{\mathbf{x}} - \mathbf{A}\mathbf{x})\} = \text{tr}[\sigma^2 \mathbf{P}_A] = r\sigma^2$$

Hence, the variance of the estimate of $\boldsymbol{\theta}$ can be very large if there are singular values close to zero.

The reduced rank approach provides a trade-off between bias and variance in order to reduce the mean squared error of the estimator

- el Reduced-Rank LS prescindeix dels VAPs petits.

Define

$\Gamma = \text{diag}(\gamma_i)$ Weighting matrix

$\Sigma_r = \Sigma_+ \Gamma^{-1}$ Diagonal matrix with weighted singular values

$\mathbf{A}_r^\# = \mathbf{V} \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & \Sigma_0^T \end{bmatrix} \mathbf{U}^H = \mathbf{V} \Sigma_r^{-1} \mathbf{U}_A^H$ Reduced rank pseudo inverse of \mathbf{A}

Then, if we define the reduced-rank estimate of the parameter vector as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_r &= \mathbf{A}_r^\# \mathbf{y} \\ \hat{\mathbf{x}} &= \mathbf{A} \hat{\boldsymbol{\theta}}_r \end{aligned}$$

it can easily be shown that this estimate is Gaussian distributed according to

$$\hat{\boldsymbol{\theta}}_r = \mathbf{A}_r^\# \mathbf{y} \Rightarrow \hat{\boldsymbol{\theta}}_r : N(\mathbf{A}_r^\# \mathbf{A}\boldsymbol{\theta}, \sigma^2 \mathbf{A}_r^\# \mathbf{A}_r^\# H)$$

$$\hat{\mathbf{x}} = \mathbf{A} \hat{\boldsymbol{\theta}}_r \Rightarrow \hat{\mathbf{x}} : N(\mathbf{A}\mathbf{A}_r^\# \mathbf{x}, \sigma^2 \mathbf{A}\mathbf{A}_r^\# \mathbf{A}_r^\# H \mathbf{A}^H)$$

so the mean squared error in the estimate of $\boldsymbol{\theta}$ is

$$\xi_\theta^2 = E\{(\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta})^H (\hat{\boldsymbol{\theta}}_r - \boldsymbol{\theta})\} = \mathbf{c}_r^H \mathbf{c}_r + \sigma^2 \text{tr}[\mathbf{A}_r^\# \mathbf{A}_r^\# H] = \mathbf{c}_r^H \mathbf{c}_r + \sigma^2 \text{tr}[\Gamma^2 \Sigma_+^{-2}] = \mathbf{c}_r^H \mathbf{c}_r + \sigma^2 \sum_{j=1}^r \frac{\gamma_j^2}{\sigma_j^2}$$

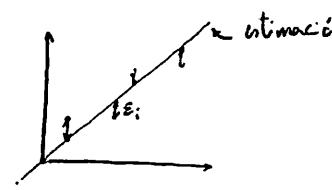
where the unknown bias term is $\mathbf{c}_r = (\mathbf{A}^\# \mathbf{A} - \mathbf{A}_r^\# \mathbf{A}) \boldsymbol{\theta}$. Hence, selecting γ_j we can trade bias and variance in the estimate.

Quina W hauria de triar pel WLS per t. j. faci de Reduced-Rank LS?

6.8. TOTAL LEAST SQUARES PROBLEM

The theory of least squares tries to find a solution to the set of equations $Ax \approx b$ so

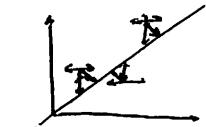
$$\min_{x, \Delta b} \|\Delta b\| \quad \text{subject to } Ax = b + \Delta b$$



If the set of equations is not consistent, it is assumed that the errors appear in b , so we try to modify b as little as possible in order to guarantee that $b \in \text{span}(A)$.

However, if both A, b may have errors, it seems more appropriate to state the problem as

$$\min_{x, E, r} \|E r\|_F \quad \text{subject to } (A + E)x = b + r$$



The total least squares theory provides the framework for the resolution of this problem. We will derive the solution to this problem from two different approaches:

First Approach:

Theorem: Let A be a matrix with $\text{rank}(A)=r$ and define its reduced rank approximation A_k as:

$$A_k = \sum_{j=1}^k u_j \sigma_j v_j^H$$

Then, $\|A - A_k\|_2 = \sigma_{k+1}$ and A_k is the nearest matrix of rank k to A (in the 2-norm): $\min_B \|A - B\|_2 = \|A - A_k\|_2$
 B is de rank $k = K$

The Least Squares problem

6.16

We can re-state our problem as that one of finding the matrix $[E r]$ such that satisfies

$$[A + E \ b + r] \begin{bmatrix} x \\ -1 \end{bmatrix} = 0 \Rightarrow ([A \ b] + [E \ r]) \begin{bmatrix} x \\ -1 \end{bmatrix} = 0 \iff ([A+b] + [E+r]) \text{ is de rank deficient}$$

and has minimum norm.

According to the previous theorem the solution is

$$[E \ r] = -\sigma_{\min} u_{\min} v_{\min}^H$$

$$\begin{bmatrix} x \\ -1 \end{bmatrix} = \alpha v_{\min} \quad \text{similar a un problema d'estimació espectral de mínima variança (AR)}$$

from where x can be obtained.

Second Approach:

We now regard the LS estimate of Ax as a projection P_s that is not perfectly matched to A :

$$\begin{aligned} \hat{y} &= P_s b = P_s y + P_s e = P_s A x + P_s e \\ b = Ax + e &= y + e \Rightarrow \hat{e} = P_s^\perp b = P_s^\perp y + P_s^\perp e \end{aligned}$$

The projection P_s is selected to minimize the sum of the squares of \hat{e} and $P_s^\perp A$, that is:

$$\xi^2 = \text{tr} \left[\begin{bmatrix} A^H \\ b^H \end{bmatrix} P_s^\perp [A \ b] \right]$$

$$\xi^2 = \|P_s^\perp [A \ b]\|_F^2 \quad \text{on } P_s^\perp P_s^\perp = P_s^\perp$$

$$b = \hat{y} + \hat{e} \dots b \begin{cases} P_s \\ P_s^\perp \end{cases} \begin{cases} \hat{y} \\ \hat{e} \end{cases}$$

$P_s^\perp A = 0 \Rightarrow$ mas en el nodd
 ers memoria el grau d'rror del model

This Frobenius norm is minimized selecting P_s^\perp to project into the subspace spanned by the singular vector associated to the smallest singular value of $[A \ b]$:

$$[A \ b] = U \Sigma V^H$$

$$P_s = I - u_{\min} u_{\min}^H$$

$$\hat{\theta} = [(P_s A)^H (P_s A)]^{-1} (P_s A)^H \hat{x} = (A^H P_s A)^{-1} A^H P_s x$$

So we get an estimate that has the same structure as the weighted least squares.

• en general, minimization $\|B^H P_s^\perp B\|$ equivalent a aigafar P_s^\perp t.g. projecti en el subespai generat pel VAP i més de B

* TLS 1st approach. is similar a un problema d'estimacio espectral basat en model AR

$$a_0 x_n + a_1 x_{n-1} + a_2 x_{n-2} + \dots \stackrel{F=0}{=} \begin{bmatrix} x \\ x_{n-1} \\ \vdots \end{bmatrix}$$

7. DERIVATIVES AND GRADIENTS. REAL AND COMPLEX VARIABLE

7.1. DERIVATIVES AND GRADIENTS. REAL VARIABLE.

7.2. DERIVATIVE OF COMPLEX FUNCTIONS WITH RESPECT TO COMPLEX VARIABLES.

7.3. DERIVATIVE OF REAL FUNCTIONS WITH RESPECT TO COMPLEX VARIABLES

7.4. MODIFICATIONS FOR DERIVATIVES OF COMPLEX VECTORS AND MATRICES.

In most signal processing applications, we have to derivate a function of cost, not negative, with respect to a parameter vector or a parameter matrix.

In this situation, the gradient of vectors and matrices play an important role in signal matters. In this chapter, we develop some fundamental theoretical concepts and practical tools

7.1 DERIVATIVES AND GRADIENTS. REAL VARIABLE

DERIVATIVES OF VECTORS AND SCALARS WITH RESPECT TO A REAL VECTOR

Let $a(x)$ be a scalar function dependent on vector x with dimension $N \times 1$, that is, $x \in \mathbb{R}^{N \times 1}$.

The gradient of a with respect to x is the vector:

$$\frac{\partial a(x)}{\partial x} = \nabla_x(a(x)) = \begin{bmatrix} \frac{\partial a(x)}{\partial x_1} \\ \vdots \\ \frac{\partial a(x)}{\partial x_N} \end{bmatrix}$$

This definition could be generalized to the gradient of a vector $a^T(x) = [a_1(x) \dots a_M(x)] \in \mathbb{R}^{M \times 1}$ with respect to x as follows

$$\frac{\partial a(x)}{\partial x} = \nabla_x(a(x)) = \begin{bmatrix} \frac{\partial a_1(x)}{\partial x} & \dots & \frac{\partial a_M(x)}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial a_1(x)}{\partial x_1} & \dots & \frac{\partial a_M(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_1(x)}{\partial x_N} & \dots & \frac{\partial a_M(x)}{\partial x_N} \end{bmatrix}$$

Each row corresponds to a gradient of the scalar $a_m(x)$ with respect to the vector x .

Particular cases

$$\underline{x} \in \mathbb{R}^{N \times 1}$$

1. Identity matrix:

$$\nabla_{\underline{x}}(\underline{x}) = \mathbf{I}_{N \times N}$$

2. Linear operation (\mathbf{b} does not depend on \underline{x}):

$$\nabla_{\underline{x}}(\underline{x}^T \mathbf{b}) = \nabla_{\underline{x}}(\mathbf{b}^T \underline{x}) = \mathbf{b}$$

3. Scalar product of vectors that depend on \underline{x} :

$$\nabla_{\underline{x}}(\underline{a}^T(\underline{x}) \underline{b}(\underline{x})) = \nabla_{\underline{x}}(\underline{a}(\underline{x})) \underline{b}(\underline{x}) + \nabla_{\underline{x}}(\underline{b}(\underline{x})) \underline{a}(\underline{x})$$

scalar = red. fcto • vect. val.

4. Quadratic form:

Symmetric matrix, $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{N \times N}$: $\underline{a}^T : 3. s b \in \text{den}$

$$\nabla_{\underline{x}}(\underline{x}^T \mathbf{Q} \underline{x}) = 2 \mathbf{Q} \underline{x}$$

Non-symmetric matrix, $\mathbf{Q} \in \mathbb{R}^{N \times N}$

$$\nabla_{\underline{x}}(\underline{x}^T \mathbf{Q} \underline{x}) = \mathbf{Q} \underline{x} + \mathbf{Q}^T \underline{x}$$

5. Parametric quadratic form ($\underline{x} \in \mathbb{R}^{N \times 1}$ and $\underline{a}(\underline{x}) \in \mathbb{R}^{M \times 1}$):Symmetric matrix, $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{N \times N}$:

$$\nabla_{\underline{x}}(\underline{a}^T(\underline{x}) \mathbf{Q} \underline{a}(\underline{x})) = 2 \nabla_{\underline{x}}(\underline{a}(\underline{x})) \mathbf{Q} \underline{a}(\underline{x})$$

$$3b. \quad \nabla_{\underline{x}}(\underline{A} \cdot \underline{x}) = \underline{A}^T$$

$$\begin{aligned}
 \textcircled{3} \text{ den: } \underline{a}^T(\underline{x}) \underline{b}(\underline{x}) &= \sum_{i=1}^M a_i(\underline{x}) b_i(\underline{x}) \\
 \frac{\partial(\underline{a}^T \underline{b})}{\partial \underline{x}_e} &= \sum_i \left(\frac{\partial a_i}{\partial \underline{x}_e} \underline{b}_i + a_i \frac{\partial \underline{b}_i}{\partial \underline{x}_e} \right) = \\
 &= \begin{bmatrix} \frac{\partial a_1}{\partial \underline{x}_e} & \dots & \frac{\partial a_M}{\partial \underline{x}_e} \end{bmatrix} \begin{bmatrix} \underline{b}_1 \\ \vdots \\ \underline{b}_M \end{bmatrix} + \begin{bmatrix} a_1 & \dots & a_M \end{bmatrix} \begin{bmatrix} \frac{\partial \underline{b}_1}{\partial \underline{x}_e} \\ \vdots \\ \frac{\partial \underline{b}_M}{\partial \underline{x}_e} \end{bmatrix} \\
 \frac{\partial(\underline{a}^T \underline{b})}{\partial \underline{x}} &= \begin{bmatrix} \frac{\partial a_1}{\partial \underline{x}_1} & \dots & \frac{\partial a_M}{\partial \underline{x}_N} \end{bmatrix} \begin{bmatrix} \underline{b}_1 \\ \vdots \\ \underline{b}_M \end{bmatrix} + \begin{bmatrix} a_1 & \dots & a_M \end{bmatrix} \begin{bmatrix} \frac{\partial \underline{b}_1}{\partial \underline{x}_1} \\ \vdots \\ \frac{\partial \underline{b}_M}{\partial \underline{x}_N} \end{bmatrix} \\
 &= \frac{\partial \underline{a}}{\partial \underline{x}} \cdot \underline{b} + \frac{\partial \underline{b}}{\partial \underline{x}} \underline{a} = (\underline{a}^T \frac{\partial \underline{b}}{\partial \underline{x}})^T \\
 &= \nabla_{\underline{x}}(\underline{a}) \cdot \underline{b} + \nabla_{\underline{x}}(\underline{b}) \underline{a}
 \end{aligned}$$

Non-symmetric matrix, $\mathbf{Q} \in \mathbb{R}^{N \times N}$

$$\nabla_{\underline{x}}(\underline{a}^T(\underline{x}) \mathbf{Q} \underline{a}(\underline{x})) = \nabla_{\underline{x}}(\underline{a}(\underline{x})) \mathbf{Q} \underline{a}(\underline{x}) + \nabla_{\underline{x}}(\underline{a}(\underline{x})) \mathbf{Q}^T \underline{a}(\underline{x})$$

6. Chain Rule:

$$\nabla_{\underline{x}}(f(\underline{a}(\underline{x}))) = \nabla_{\underline{x}}(\underline{a}(\underline{x})) \nabla_{\underline{a}}(f(\underline{a}))$$

where $\nabla_{\underline{x}}(f(\underline{a}(\underline{x}))) \in \mathbb{R}^{N \times 1}$, $\nabla_{\underline{x}}(\underline{a}(\underline{x})) \in \mathbb{R}^{N \times M}$ and $\nabla_{\underline{a}}(f(\underline{a})) \in \mathbb{R}^{M \times 1}$.

$$\textcircled{4} \text{ den: } \nabla(\underline{A}^T \underline{Q} \underline{x}) = \frac{\partial(\underline{Q}^T \underline{x})}{\partial \underline{x}} \underline{x} + \frac{\partial \underline{x}}{\partial \underline{x}} \underline{Q}^T \underline{x} = \underline{Q} \underline{x} + \underline{Q}^T \underline{x} = 2 \underline{Q} \underline{x}$$

$\textcircled{3}$ $\textcircled{5}$ $\textcircled{6}$

sym.

$$\textcircled{5} \text{ den: } \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}_1} \\ \vdots \\ \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \underline{a}} & \frac{\partial a(\underline{x})}{\partial \underline{x}_1} \\ \vdots & \vdots \\ \frac{\partial f}{\partial \underline{a}} & \frac{\partial a(\underline{x})}{\partial \underline{x}_N} \end{bmatrix} = \frac{\partial f}{\partial \underline{a}} \cdot \frac{\partial a(\underline{x})}{\partial \underline{x}}$$

Regla de la cadena

$$\textcircled{6} \text{ den: } \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}_1} \\ \vdots \\ \frac{\partial f(\underline{a}(\underline{x}))}{\partial \underline{x}_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial \underline{a}_1} \frac{\partial a(\underline{x})}{\partial \underline{x}_1} + \dots + \frac{\partial f}{\partial \underline{a}_M} \frac{\partial a(\underline{x})}{\partial \underline{x}_1} \\ \vdots \\ \frac{\partial f}{\partial \underline{a}_1} \frac{\partial a(\underline{x})}{\partial \underline{x}_N} + \dots + \frac{\partial f}{\partial \underline{a}_M} \frac{\partial a(\underline{x})}{\partial \underline{x}_N} \end{bmatrix} = \frac{\partial a(\underline{x})}{\partial \underline{x}} \cdot \frac{\partial f(\underline{a})}{\partial \underline{a}} = \nabla_{\underline{x}}(a(\underline{x})) \nabla_{\underline{a}}(f(\underline{a}))$$

DERIVATIVES OF REAL-VALUED FUNCTIONS OF REAL MATRICES:

Let $a(\mathbf{R})$ be a scalar function that depends on a matrix $\mathbf{R} \in \mathbb{R}^{M \times N}$. The gradient function of a with respect to the matrix is defined as follows

$$\nabla_{\mathbf{R}}(a) = \begin{bmatrix} \frac{\partial a(\mathbf{R})}{r_{11}} & \frac{\partial a(\mathbf{R})}{r_{12}} & \dots & \frac{\partial a(\mathbf{R})}{r_{1N}} \\ \frac{\partial a(\mathbf{R})}{r_{21}} & \frac{\partial a(\mathbf{R})}{r_{22}} & \dots & \frac{\partial a(\mathbf{R})}{r_{2N}} \\ \dots & \dots & \dots & \dots \\ \frac{\partial a(\mathbf{R})}{r_{M1}} & \frac{\partial a(\mathbf{R})}{r_{M2}} & \dots & \frac{\partial a(\mathbf{R})}{r_{MN}} \end{bmatrix}.$$

Each row vector corresponds to a vector gradient of the scalar a with respect to a row of \mathbf{R} .

Gradient and derivative of the determinant:

Let $\mathbf{R} \in \mathbb{R}^{N \times N}$, then the determinant is defined as

$$\det(\mathbf{R}) = |\mathbf{R}| = \det(\mathbf{R}) = \sum_{j=1}^N r_{ij} \cdot A_{ij}$$

with the scalar A_{ij} the cofactor.

7.6 $\text{① dem. } \frac{\partial \det(\mathbf{R})}{\partial \mathbf{R}} = \begin{bmatrix} \frac{\partial \det(\mathbf{R})}{\partial r_{11}} & \dots & \frac{\partial \det(\mathbf{R})}{\partial r_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \det(\mathbf{R})}{\partial r_{M1}} & \dots & \frac{\partial \det(\mathbf{R})}{\partial r_{MN}} \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \frac{|\mathbf{R}|}{|\mathbf{R}|} =$

$$= \frac{1}{|\mathbf{R}|} \frac{|\mathbf{R}|}{|\mathbf{R}|} = \frac{1}{|\mathbf{R}|} \cdot |\mathbf{R}| = \frac{1}{|\mathbf{R}|} \cdot \frac{1}{|\mathbf{R}|} = \frac{1}{|\mathbf{R}|} \cdot \frac{1}{|\mathbf{R}|} = \frac{1}{|\mathbf{R}|} \cdot \frac{1}{|\mathbf{R}|} = \frac{1}{|\mathbf{R}|} \cdot \frac{1}{|\mathbf{R}|} =$$

$$\tilde{\mathbf{R}} = \frac{1}{|\mathbf{R}|} \mathbf{A}^T$$

1. Gradient of the determinant

- If the matrix is non-symmetric:

$$\nabla_{\mathbf{R}}(\det(\mathbf{R})) = \det(\mathbf{R}) \mathbf{R}^{-T}$$

- If the matrix is symmetric ($\mathbf{R} = \mathbf{R}^T$):

$$\nabla_{\mathbf{R}}(\det(\mathbf{R})) = \det(\mathbf{R}) (2\mathbf{R}^{-1} - \text{diag}(\mathbf{R}^{-1}))$$

2. Derivative of the determinant of a parameterized matrix:

$$(\mathbf{R})_{ij} = f_{ij}(c); \quad \mathbf{B} = \frac{\partial \mathbf{R}}{\partial c}; \quad \frac{\partial \det(\mathbf{R})}{\partial c} = \text{trace}(\mathbf{A}_R \mathbf{B}^T)$$

$$3. \quad \nabla_{\mathbf{R}}(\ln(\det(\mathbf{R}))) = \frac{1}{\det(\mathbf{R})} \nabla_{\mathbf{R}}(\det(\mathbf{R}))$$

$$4. \quad \nabla_{\mathbf{R}}((\det(\mathbf{R}))^n) = n(\det(\mathbf{R}))^{n-1} \nabla_{\mathbf{R}}(\det(\mathbf{R}))$$

DERIVATIVE OF MATRICES WITH RESPECT TO SCALARS AND VICE VERSA

$$1. \frac{\partial R}{\partial r_{ij}} = E_{ij} \quad E_{ij} = \begin{cases} 0 & e_{mn} \neq e_{ij} \\ 1 & e_{ij} \end{cases}$$

$$2. \frac{\partial ARB}{\partial r_{ij}} = AE_{ij}B$$

$$3. \nabla_R (ARB)_{ij} = A^T E_{ij} B^T$$

$$4. \frac{\partial AR^{-1}B}{\partial r_{ij}} = -AR^{-1}E_{ij}R^{-1}B$$

$$5. \nabla_R (AR^{-1}B)_{ij} = -R^{-T}A^T E_{ij} B^T R^{-T}$$

$$6. \frac{\partial AB}{\partial r} = \frac{\partial A}{\partial r}B + A\frac{\partial B}{\partial r}$$

$$7. \frac{\partial R^T AR}{\partial r_{ij}} = E_{ij}^T AR + R^T AE_{ij}$$

$$8. \nabla_R (R^T AR)_{ij} = ARE_{ij}^T + A^T RE_{ij}$$

$$9. \frac{\partial R^n}{\partial r_{ij}} = \sum_{k=0}^{n-1} R^k E_{ij} R^{n-k-1}$$

$$10. \nabla_R (R^n)_{ij} = \sum_{k=0}^{n-1} (R^T)^k E_{ij} (R^T)^{n-k-1}$$

$$\frac{\partial(AA^{-1})}{\partial a_{ij}} \triangleq 0 \quad \text{dem.} \quad \frac{\partial(AA^{-1})}{\partial a_{ij}} = \frac{\partial A}{\partial a_{ij}} A^{-1} + A \frac{\partial A^{-1}}{\partial a_{ij}} \rightarrow \frac{\partial A^{-1}}{\partial a_{ij}} = A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1}$$

Derivatives involving the trace

The trace of a matrix $A \in \mathbb{C}^{M \times M}$ is $\text{tr}(A) = \sum_{i=1}^n a_{ii}$.

1. $\nabla_R (\text{trace}(R)) = I_{N \times N}$
2. $\text{trace}\left(\frac{\partial R}{\partial x}\right) = \frac{\partial \text{trace}(R)}{\partial x}$
3. Non-symmetric matrix: $\nabla_R (\text{trace}(AR)) = A^T$
Symmetric matrix: $\nabla_R (\text{trace}(AR)) = A^T + A - \text{diag}A$
4. $\nabla_R (\text{trace}(ARB)) = A^T B^T$
5. $\nabla_R (\text{trace}(RR^T)) = 2R; \quad \nabla_R (\text{trace}(RR)) = 2R^T$
6. $\nabla_R (\text{trace}(R^n)) = 2(R^{n-1})^T$
7. $\nabla_R (\text{trace}(AR^{-1}B)) = -(R^{-1}BAR^{-1})^T$
8. $\nabla_R (\text{trace}(\exp(R))) = \exp(R^T)$

7.2. DERIVATIVE OF COMPLEX FUNCTIONS WITH RESPECT TO COMPLEX VARIABLES

In this section we summarize the complex variable basics needed to generalize the results of derivative rules of vectors and matrices presented so far to the complex case.

When we work with complex functions of complex variable, we have two different approaches:

Approach 1:

Let z be a complex scalar: $z = x + jy$ and $f(x, y)$ a generic function:

$$f: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C} \quad f(x, y) = u(x, y) + jv(x, y)$$

with $u, v: f: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$

f : func δ s g. naja \mathbb{R}^2 als \mathbb{C}

Approach 2:

Let z be a complex scalar: $z = x + jy$ and $z^* = x - jy$ its complex conjugate, and $g(z, z^*)$ a generic function:

$$g: \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C} \quad g(z, z^*)$$

f : func δ s g. naja \mathbb{C}^2 a \mathbb{C}

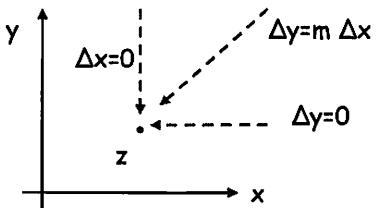
Note that $f(x, y)$ and $g(z, z^*)$ are the same function, although formally they are very different.

Derivative of $f(x,y)$ eC.

The derivative of $f(x,y)$ with respect to the complex variable z is defined by:

$$\frac{\partial f}{\partial z} = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \quad (7.1)$$

This definition is formally identical to a real derivative. However, since $\Delta z = \Delta x + j\Delta y$ is a complex variable, the limit $\Delta z \rightarrow 0$ presents additional aspects to be considered.

Example

Consider the function $f(z) = z^* = x - jy$. Then, using (7.1) we obtain

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} = \frac{(x + \Delta x) - j(y + \Delta y) - (x - jy)}{\Delta x + j\Delta y} = \frac{\Delta x - j\Delta y}{\Delta x + j\Delta y}.$$

Then,

- $\Delta z \in R$

$$\lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\Delta x \rightarrow 0} \frac{\Delta x}{\Delta x} = 1 \quad \Delta y = 0$$

- $\Delta z \in I$

$$\lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\Delta y \rightarrow 0} \frac{-j\Delta y}{j\Delta y} = -1 \quad \Delta x = 0$$

- $\Delta y = m\Delta x$

$$\lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\Delta x \rightarrow 0} \frac{\Delta x - jm\Delta x}{\Delta x + jm\Delta x} = \frac{1 - jm}{1 + jm} \quad \Delta y = m \Delta x$$

In this case, we say that the derivative of this function does not exist. Moreover, we say it is a non-analytic function.

Conditions per f.g. b derivada existen:



□

Cauchy-Riemann Equations. Analytic functions

Assume u and v are uniform real functions of x and y . Assume also that these functions and their four partial derivatives are continuous in a region R . Then, the Cauchy-Riemann equations

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} : \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}} \quad (7.2) \quad f(x,y) = u(x,y) + j v(x,y)$$

• No complex $\Rightarrow f(x,y)$ es analítica

are necessary and sufficient conditions to $f(x,y) = u(x,y) + jv(x,y)$ be analytic in z . That is, if Cauchy-Riemann equations hold for $z = z_0$, then the derivative of $f(x,y)$ in $z = z_0$ exists (in this case, we say that z_0 is a regular point of $f(x,y) = f(z)$).

The above condition implies that for any analytic function $g(z, z^*) = g(z)$, that is, it depends on z but not on z^* .

Theorem

If we replace the variables $x = (z + z^*)/2$ and $y = (z - z^*)/2j$ in any analytic function $g(z, z^*) = f(x, y) = u(x, y) + jv(x, y)$, then $g(z, z^*) = g(z)$ only depends on z .

Proof:

Since $x = (z + z^*)/2$ and $y = (z - z^*)/2j$, we can consider that g is a function of two new independent variables z and z^* . To prove that g only depends on z , we need to check if $\partial g / \partial z^* = 0$. Using the Chain Rule we have

$$\frac{\partial g}{\partial z^*} = \frac{\partial u}{\partial z^*} + j \frac{\partial v}{\partial z^*} = \left(\frac{\partial u}{\partial x} \frac{\partial x}{\partial z^*} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial z^*} \right) + j \left(\frac{\partial v}{\partial x} \frac{\partial x}{\partial z^*} + \frac{\partial v}{\partial y} \frac{\partial y}{\partial z^*} \right).$$

Considering that $\frac{\partial x}{\partial z^*} = \frac{1}{2}$ and $\frac{\partial y}{\partial z^*} = \frac{j}{2}$, we can certainly write

$$\frac{\partial g}{\partial z^*} = \left(\frac{1}{2} \frac{\partial u}{\partial x} + \frac{j}{2} \frac{\partial u}{\partial y} \right) + j \left(\frac{1}{2} \frac{\partial v}{\partial x} + \frac{j}{2} \frac{\partial v}{\partial y} \right) = \frac{1}{2} \left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) + \frac{j}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)$$

The proof is completed if we replace the Cauchy-Riemann equations (7.2). □

Example

Assume that $f(x, y) = x - jy = z$. Since

$$\frac{\partial u}{\partial x} = 1; \frac{\partial u}{\partial y} = 0; \frac{\partial v}{\partial x} = 0; \frac{\partial v}{\partial y} = -1$$

$$u = x \\ v = -y$$

$f(x, y)$ is not an analytic function in z . In fact, $f(x, y)$ is not an explicit function of z . □

Example

For $f(x, y) = zz^* = x^2 + y^2$, we have

$$\frac{\partial u}{\partial x} = 2x; \frac{\partial u}{\partial y} = 2y; \frac{\partial v}{\partial x} = 0; \frac{\partial v}{\partial y} = 0.$$

Then, $f(x, y)$ is an analytic function at $z = 0$. □

Example

Assume the analytic function $f(x, y) = z^2$, then $\frac{\partial f}{\partial z} = 2z$.

si $f(x, y)$ es analítica, entonces
el derivado respecto a

However, lots of signal processing problems deals with the optimization of real functions that depend on both z and z^* . Therefore, an alternative approach is developed for those cases.

Derivative of $g(z, z^*)$ c.c. Non-analytic functionsTheorem

Suppose that $\begin{array}{l} g: C \times C \rightarrow C \\ g: z, z^* \rightarrow g(z, z^*) \end{array}$ where g is analytic (or derivable) with respect to z and z^* independently.

Suppose that $\begin{array}{l} f: R \times R \rightarrow C \\ f: x, y \rightarrow f(x, y) \end{array}$ and $f(x, y) = g(z, z^*)$ with $z = x + jy$, then

$$\frac{\partial g}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right) \text{ and } \frac{\partial g}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right). \quad (7.3)$$



$$\frac{\partial g}{\partial z} \Big|_{z^*=\text{ctt}}$$



$$\frac{\partial g}{\partial z^*} \Big|_{z=\text{ctt}}$$

$g(z, z^*)$ derivable respecto a
a t.o. respecto z^* , quan
 z^* ; z son c.t.o. respecte.

Proof:

Since $x = (z + z^*)/2$ and $y = (z - z^*)/2j$, we can consider that g is a function of two independent variables, z and z^* . Using the chain rule, the derivative with respect to x and y is equal to:

$$\begin{cases} \frac{\partial f}{\partial x} = \frac{\partial g}{\partial z} \frac{\partial z}{\partial x} + \frac{\partial g}{\partial z^*} \frac{\partial z^*}{\partial x} \\ \frac{\partial f}{\partial y} = \frac{\partial g}{\partial z} \frac{\partial z}{\partial y} + \frac{\partial g}{\partial z^*} \frac{\partial z^*}{\partial y} \end{cases}$$

Since

$$\frac{\partial z}{\partial x} = \frac{\partial z^*}{\partial x} = 1 \text{ and } \frac{\partial z}{\partial y} = -\frac{\partial z^*}{\partial y} = j,$$

we have

$$\begin{cases} \frac{\partial f}{\partial x} = \frac{\partial g}{\partial z} + \frac{\partial g}{\partial z^*} \\ \frac{\partial f}{\partial y} = j \left(\frac{\partial g}{\partial z} - \frac{\partial g}{\partial z^*} \right) \end{cases}$$

Reorganising the equations and isolating the terms in which we are interested, we proof the theorem:

$$\begin{cases} \frac{\partial g}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + \frac{1}{j} \frac{\partial f}{\partial y} \right) & \leftarrow \text{per } z^* \text{ ctt} \\ \frac{\partial g}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - \frac{1}{j} \frac{\partial f}{\partial y} \right) & \leftarrow \text{per } z \text{ ctt} \end{cases}$$

These expressions are valid for analytic functions as well.

Particular cases:

$$1. \frac{\partial z}{\partial z^*} = \frac{\partial z^*}{\partial z} = 0 \text{ but } \frac{\partial z}{\partial z} = \frac{\partial z^*}{\partial z^*} = 1$$

$$2. \frac{\partial(zz^*)}{\partial z} = z^* \text{ but } \frac{\partial(zz^*)}{\partial z^*} = z$$

$$3. g(z, z^*) = z^2 z^* = \underbrace{(x^3 + xy^2)}_{u(x,y)} + j \underbrace{(x^2y + y^3)}_{v(x,y)}. \text{ Then,}$$

$$\frac{\partial g}{\partial z} = \frac{1}{2} \left(\frac{\partial f}{\partial x} + \frac{1}{j} \frac{\partial f}{\partial y} \right) = 2(x^2 + y^2) = 2zz^*$$

$$\frac{\partial g}{\partial z^*} = \frac{1}{2} \left(\frac{\partial f}{\partial x} - \frac{1}{j} \frac{\partial f}{\partial y} \right) = x^2 + 2jxy + y^2 = z^2$$

□

It can be seen that, in general, the derivative of non-analytic functions $g(z, z^*)$ with respect to z is different from the derivative with respect to z^* .

7.3. DERIVATIVE OF REAL FUNCTIONS WITH RESPECT TO COMPLEX VARIABLES

In this section we study the particular case of real functions that depend on a given complex variables. We assume this function is independently analytic in z and z^* . Next theorem shows that, in this case, we can find the stationary points of the function either with the derivative with respect to z or with respect to z^* .

Derivative of $g(z, z^*) \in \mathbb{R}$

g real i no-analitica

Theorem 3

Let $f: C \times C \rightarrow \mathbb{R}$; $f(z) = g(z, z^*)$ be a real function that depends on a complex variable. A necessary and sufficient condition to guarantee the existence of a stationary point of $f(z)$ in $z = z_0$ is that

$$\left. \frac{\partial g}{\partial z} \right|_{z=z_0} = 0.$$

A necessary and sufficient condition to guarantee the existence of a stationary point of $f(z)$ in $z^* = z_0^*$ is:

$$\left. \frac{\partial g}{\partial z^*} \right|_{z^*=z_0^*} = 0.$$

Un 2 sim
vinden
tenen
igual
resultat

Proof:

The stationary condition is given by $\frac{\partial u}{\partial x} = \frac{\partial u}{\partial y} = 0$, since $v(x, y) = 0$. The theorem is proved substituting in the derivatives (7.3). □

$f(z)$
vector
func scalar

7.4. MODIFICATIONS FOR DERIVATIVES OF COMPLEX VECTORS AND MATRICES.

Assume $f(z)$ with $z \in C^N$. If $f: C^N \rightarrow R$; $f(z) = g(z, z^*)$ is independently analytic in z and z^* , we may define the gradient as follows. If

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_N \end{bmatrix} = \begin{bmatrix} x_1 + jy_1 \\ x_2 + jy_2 \\ \dots \\ x_N + jy_N \end{bmatrix},$$

then the gradient

$$\nabla_z(g) = \begin{bmatrix} \frac{\partial g}{\partial z_1} \\ \frac{\partial g}{\partial z_2} \\ \dots \\ \frac{\partial g}{\partial z_N} \end{bmatrix} \text{ and } \nabla_{z^*}(g) = \begin{bmatrix} \frac{\partial g}{\partial z_1^*} \\ \frac{\partial g}{\partial z_2^*} \\ \dots \\ \frac{\partial g}{\partial z_N^*} \end{bmatrix},$$

with

$$\frac{\partial g}{\partial z_k} = \frac{1}{2} \left(\frac{\partial g}{\partial x_k} - j \frac{\partial g}{\partial y_k} \right) \text{ and } \frac{\partial g}{\partial z_k^*} = \frac{1}{2} \left(\frac{\partial g}{\partial x_k} + j \frac{\partial g}{\partial y_k} \right).$$

Gradient of a linear function of a complex vector ∇ : vector column

1. $\nabla_z (\mathbf{a}^H \mathbf{z}) = \mathbf{a}^*$ and $\nabla_{z^*} (\mathbf{a}^H \mathbf{z}) = 0.$

2. $\nabla_z (\mathbf{z}^H \mathbf{a}) = \mathbf{0}$ and $\nabla_{z^*} (\mathbf{z}^H \mathbf{a}) = \mathbf{a}.$

Gradient of a quadratic function of a complex vector

1. $\nabla_z (\mathbf{z}^H \mathbf{R} \mathbf{z}) = \mathbf{R}^T \mathbf{z}^*$ and $\nabla_{z^*} (\mathbf{z}^H \mathbf{R} \mathbf{z}) = \mathbf{R} \mathbf{z}.$

2. Gradient of Vectors:

• $\nabla_z (\mathbf{z}) = \mathbf{I}$ and $\nabla_{z^*} (\mathbf{z}) = \nabla_z (\mathbf{z}^*) = \mathbf{0}$

Derivative involving the trace. $\mathbf{R} \in \mathbb{C}^{N \times N}$

1. $\nabla_{\mathbf{R}} (\text{trace}(\mathbf{R})) = \mathbf{I}_{N \times N}$

2. $\nabla_{\mathbf{R}} (\text{trace}(\mathbf{R}^H \mathbf{R})) = \mathbf{R}^*$

3. $\nabla_{\mathbf{R}^*} (\text{trace}(\mathbf{R}^H \mathbf{R})) = \mathbf{R}$

4. $\nabla_{\mathbf{R}^H} (\text{trace}(\mathbf{R}^H \mathbf{R})) = \mathbf{R}^T$

5. $\nabla_{\mathbf{R}} (\text{trace}(\mathbf{A} \mathbf{R} \mathbf{B})) = \mathbf{A}^T \mathbf{B}^T$

6. $\nabla_{\mathbf{R}^*} (\text{trace}(\mathbf{A} \mathbf{R}^* \mathbf{B})) = \mathbf{A}^T \mathbf{B}^T$

7. $\nabla_{\mathbf{R}^*} (\text{trace}(\mathbf{A} \mathbf{R} \mathbf{B})) = \mathbf{0}$

8. $\nabla_{\mathbf{R}^*} (\text{trace}(\mathbf{A} \mathbf{R}^H \mathbf{B})) = \mathbf{B} \mathbf{A}$

Stationary pointsTheorem¹

Assume $f: C^N \rightarrow R$
 $f(z) = g(z, z^*)$ is independently analytic in z and z^* . Then, $\nabla_z(g) = 0$ or $\nabla_{z^*}(g) = 0$ are both necessary and sufficient conditions to find the stationary points of g .

qualssevol de les
 2 condicions no bé
 i dona resultats iguals

Theorem²

The gradient defines the direction of maximum slope of f with respect to z , and z^* .

pendent

Important per LMS (p.ex.)

^{1,2} Proofs of these theorems are in [Brandwood, IEE Proc, Feb 83]

8. LINEAR PROGRAMMING

8.1. INTRODUCTION

8.2. BASIC PROPERTIES OF LINEAR PROGRAMS

8.3. THE SIMPLEX METHOD

8.1 INTRODUCTION

A linear programming problem is a mathematical program in which the objective function is linear in the unknowns and the constraints are linear equalities and linear inequalities. In fact, any linear program can be transformed into the following *standard form*:

$$\begin{aligned} & \text{maximize} && c_1x_1 + \cdots + c_nx_n \\ & \text{subject to} && \begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases} \\ & \text{and} && x_1 \geq 0, \dots, x_n \geq 0 \end{aligned} \quad (8.1)$$

With a vector notation, this standard problem becomes

$$\begin{aligned} & \text{maximize} && c^T x \\ & \text{subject to} && Ax = b \text{ and } x \geq 0' \end{aligned}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ have constant real components and $x \in \mathbb{R}^n$. We also assume $b_i \geq 0$ for $i = 1, 2, \dots, m$ (if not, the i^{th} equation should be multiplied by -1).

• es un maximizar una función lineal con restricciones lineales

Converting other forms of linear programs into the standard form1. The minimizing problem

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax = b \text{ and } x \geq 0 \end{aligned}$$

is equivalent to

$$\begin{aligned} & \text{maximize } -c^T x \\ & \text{subject to } Ax = b \text{ and } x \geq 0 \end{aligned}$$

2. Consider the problem in which the constraint set are linear inequalities

$$\begin{aligned} & \text{maximize } c_1 x_1 + \dots + c_n x_n \\ & \text{subject to } \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n \leq b_1 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \leq b_m \end{cases} \\ & \text{and } x_1 \geq 0, \dots, x_n \geq 0. \end{aligned}$$

The problem may be alternatively expressed as

$$\begin{aligned} & \text{maximize } c_1 x_1 + \dots + c_n x_n \\ & \text{subject to } \begin{cases} a_{11}x_1 + \dots + a_{1n}x_n + y_1 = b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n + y_2 = b_2 \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n + y_m = b_m \end{cases} \\ & \text{and } x_1 \geq 0, \dots, x_n \geq 0 \\ & \text{and } y_1 \geq 0, y_2 \geq 0, \dots, y_m \geq 0. \end{aligned}$$

→ el problema tiene: $(n+m)$ variables

The new positive variables $\{y_i\}$ are called *slack variables*.

3. If the linear inequalities of the previous example are reversed so that

$$a_{11}x_1 + \dots + a_{1n}x_n \geq b_1.$$

It is clear that this is equivalent to

$$a_{11}x_1 + \dots + a_{1n}x_n - y_1 = b_1 \text{ and } y_1 \geq 0.$$

4. If a linear program is given in standard form except that one or more of the unknown variables is not required to be nonnegative, we can proceed as follows. Let us assume x_1 is unconstrained in sign. We may take any of m constraint equations of (8.1) to eliminate x_1 together with one of the constraint equations. For instance, we consider the problem

$$\begin{array}{ll} \text{maximize} & x_1 + 3x_2 + 4x_3 \\ \text{subject to} & \begin{cases} x_1 + 2x_2 + x_3 = 5 \\ 2x_1 + 3x_2 + x_3 = 6 \end{cases} \\ \text{and} & x_2 \geq 0, x_3 \geq 0. \end{array}$$

Then, since x_1 is free we solve from the first constraint

$$x_1 = 5 - 2x_2 - x_3.$$

Substituting this into the objective function and the second constraint, we obtain the equivalent problem in standard form.

$$\begin{array}{ll} \text{maximize} & x_2 + 3x_3 \\ \text{subject to} & 2x_1 + 3x_2 + x_3 = 6 \\ \text{and} & x_2 \geq 0, x_3 \geq 0. \end{array}$$

per substitució treiem una variable, reduim el problema.

8.2. BASIC PROPERTIES OF LINEAR PROGRAMS:

Consider the system of equalities

$$Ax = b \quad (8.2)$$

where $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. In practical problems $n > m$, so we can assume, without loss of generality, that matrix A is full rows rank, that is, $\text{rank}(A) = m$ or equivalently that the m rows of columns A are linearly independent.

For notational simplicity assume that the m first columns of A are linearly independent. Then, we can always find a solution of the system (8.2) setting $n-m$ components of x to 0. That is, there always exists a vector x_B such that

$$Ax = \begin{bmatrix} A_1 & A_2 \\ m \times m & m \times (n-m) \end{bmatrix} \begin{bmatrix} x_B \\ 0 \end{bmatrix}_{n-m} = A_1 x_B = b$$

since A_1 is nonsingular and may uniquely solve the equation $A_1 x_B = b$.

Definition

pg. $A_1^{m \times m}$ té cols. l.i.

$$A_{m \times n} = \begin{bmatrix} \nearrow n \\ \uparrow m \\ \searrow m \\ \downarrow n \end{bmatrix} \quad \text{l.i.}$$

Consider the system of m linear equations and n unknowns, $Ax = b$, and let $A_1 \in \mathbb{R}^{m \times m}$ be any nonsingular submatrix made up of columns of A . Then, if all $n-m$ components of x not associated with columns of A_1 are set equal to 0, the solution to the resulting set of equations is said to be a

basic solution to (8.2). The components of x associated with the columns of A_1 are called basic variables.

□

However, in the discussion of basic solutions we have made no reference to positivity constraints on the variables. Thus, considering the system of constraints

$$Ax = b \text{ and } x \geq 0 \quad (8.3)$$

we define a feasible solution as a vector x that satisfies (8.3). If it is a basic solution, then it is said to be a basic feasible solution.

FUNDAMENTAL THEOREM OF LINEAR PROGRAMMING

The fundamental theorem of linear programming establishes the importance of basic feasible solutions in solving linear programming problems. Assuming a linear program in standard form, i.e.,

$$\begin{aligned} & \text{maximize } c^T x \\ & \text{subject to } Ax = b \text{ and } x \geq 0 \end{aligned} \quad (8.4)$$

A feasible solution to the constraints that achieves the minimum value of the objective function subject to those constraints is said to be an optimal feasible solution. If the solution is basic, it is an optimal basic feasible solution.

Theorem: Fundamental theorem on linear programming.

Given a linear program in standard form (8.4) where $A \in \mathbb{R}^{m \times n}$ of rank m ,

1. If there is a feasible solution, there is a basic feasible solution.
2. If there is an optimal feasible solution, there is an optimal basic feasible solution.

□

Proof

Proof of (1). Denote the columns of A by a_1, a_2, \dots, a_n . Suppose $x = (x_1, x_2, \dots, x_n)$ is a feasible solution. Then, in terms of the columns of A , this solution satisfies:

$$x_1 a_1 + x_2 a_2 + \dots + x_n a_n = b$$

Assume that exactly p of the variables x_i are greater than zero, and for convenience, that they are the first p variables. Thus

$$x_1 a_1 + x_2 a_2 + \dots + x_p a_p = b$$

There are now two cases, corresponding as to whether the set a_1, a_2, \dots, a_p is linearly independent or linearly dependent.

CASE 1: Assume a_1, a_2, \dots, a_p are linearly independent. Then clearly, $p \leq m$. If $p = m$, the solution is basic and the proof is complete. If $p < m$, then, since A has rank m , $m-p$ vectors can be found from the remaining $n-p$ vectors so that the resulting set of m vectors is linearly independent.

Assigning the value zero to the corresponding $m - p$ variables yields a (degenerate) basic feasible solution.

CASE 2: Assume a_1, a_2, \dots, a_p are linearly dependent. Then there is a nontrivial linear combination of these vectors that is zero. Thus there are constants y_1, y_2, \dots, y_p , at least one of which can be assumed to be positive, such that

$$y_1 a_1 + y_2 a_2 + \dots + y_p a_p = 0.$$

Multiplying this equation by a scalar ε and subtracting it from $x_1 a_1 + x_2 a_2 + \dots + x_p a_p = b$, we obtain

$$(x_1 - \varepsilon y_1) a_1 + (x_2 - \varepsilon y_2) a_2 + \dots + (x_p - \varepsilon y_p) a_p = b.$$

This question equation holds for every ε , and for each ε the components $x_i - \varepsilon y_i$ correspond to a solution of the linear equalities - although they may violate $x_i - \varepsilon y_i \geq 0$. Denoting $y = (y_1, y_2, \dots, y_p, 0, 0, \dots, 0)$, we see that for any ε

$$x - \varepsilon y$$

is a solution to the equalities. For $\varepsilon = 0$, this reduces to the original feasible solution. As ε is increased from zero, the various components increase, decrease, or remain constant, depending upon whether the corresponding y_i is negative, positive, or zero. Since we assume at least one y_i is

positive, at least one component will decrease as ε is increased. Increase ε to the first point where one or more components become zero. Specifically, set

$$\varepsilon = \min\{x_i / y_i : y_i \geq 0\}$$

For this value of ε the solution given by $x - \varepsilon y$ is feasible and has at most $p - 1$ positive variables. Repeating this process if necessary, we can eliminate positive variables until we have a feasible solution with corresponding columns that are linearly independent. At that point Case 1 applies. \square

Proof of (2). Let $x = (x_1, \dots, x_n)$ be an optimal feasible solution and, as in proof of (1) above, suppose there are exactly p positive variables (x_1, \dots, x_p) . Again there are two cases; and Case 1, corresponding to linear independence is exactly the same as before.

Case 2 also goes exactly the same as before, but it must be shown that for any ε the solution $x - \varepsilon y$ is optimal. To show this, note that the value of this solution is

$$c^T x - \varepsilon c^T y.$$

For ε sufficiently small in magnitude, $x - \varepsilon y$ is a feasible solution for positive or negative values of ε . Thus we conclude that $c^T y = 0$. For, if $c^T y \neq 0$, an ε of small magnitude and proper sign could be determined so as to render $c^T x - \varepsilon c^T y$ smaller than $c^T x$ while maintaining feasibility. This would violate the assumption of optimality hence we must have $c^T y = 0$.

Having established that the new feasible solution with fewer positive components is also optimal, the remainder of the proof may be completed exactly as in part (1).

□

This theorem reduces the task of solving a linear programming problem to that of searching over basic feasible solutions. Thus for a problem with n variables and m constraints there are at most

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

basic solutions. The fundamental theorem provides an obvious but terribly inefficient search technique.

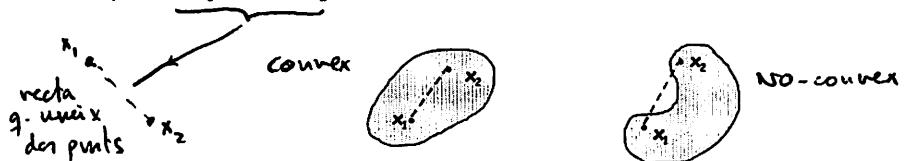
RELATIONS TO CONVEXITY

The fundamental theorem has an interesting geometric interpretation in terms of the theory of convex sets. The main link between the algebraic and geometric theories is the formal relation between basic feasible solutions of linear inequalities in standard form and extreme points of polytopes. Let first introduce some basic concepts related to convexity before establishing the correspondence between basic feasible solutions and extreme points

Basics of Convex sets

Concepts related to convex sets dominate the theory of optimization and it is essential to have knowledge of their most fundamental properties.

1. A set D in R^n is said to be *convex* if for every $x_1, x_2 \in D$ and every real number α , $0 < \alpha < 1$, the point $\alpha x_1 + (1 - \alpha) x_2 \in D$.



The most important type of convex sets is the hyperplane which generalizes the concept of a two-dimensional plane in the n -dimensional space.

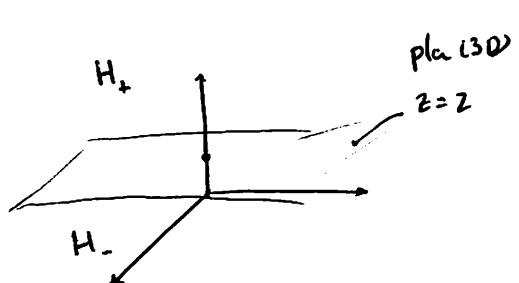
2. Let a be a nonzero n -dimensional column vector, and let c be a real number. The set

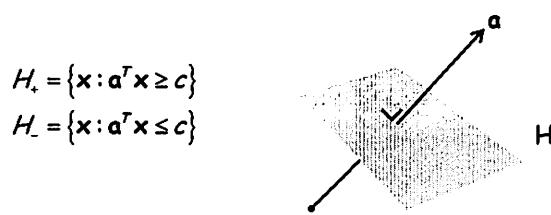
$$H = \{x \in R^n : a^T x = c\}$$

is a *hyperplane* in R^n .

Corresponding to this hyperplane are the *positive and negative closed half spaces*.

↑
per soche
del pla





3. A set which can be expressed as the intersection of a finite number of closed half spaces is said to be a *convex polytope*.

We see that convex polytopes are the sets obtained as the family of solutions to a set of linear inequalities of the form

$$\begin{aligned} a_1^T x \leq b_1 \\ a_2^T x \leq b_2 \\ \vdots \\ a_m^T x \leq b_m \end{aligned} \quad \text{intersecció de closed half spaces.}$$



la intersecció de H_+ i H_- formen un convex polytope (l'hipòpla)

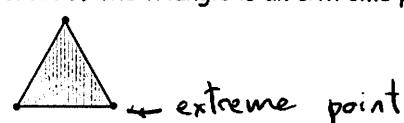
As an example, let $A \in \mathbb{R}^{m \times n}$ of rank m and $b \in \mathbb{R}^m$. All the n -vectors x satisfying

$$Ax = b \text{ and } x \geq 0$$

- form a convex polytope.
4. A point x in a convex set D is said to be an *extreme point* of D if there are no two distinct points x_1 and x_2 in D such that $x = \alpha x_1 + (1 - \alpha)x_2$ for some α , $0 < \alpha < 1$.
- An extreme point is thus a point that does not lie strictly within the line segment connecting two other points of the set. E.g. each vertex of the triangle is an extreme point.



els vèrtexos d'un
convex polígon són
extrem points



Equivalence of extreme points and basic solutions

Theorem

Let A be an $m \times n$ matrix of rank m and b and m -vector. Let K be the convex polytope consisting of all n -vectors x satisfying

$$\begin{aligned} Ax = b \\ x \geq 0 \end{aligned} \quad (8.5)$$

A vector \mathbf{x} is an extreme point of K if and only if \mathbf{x} is a Basic feasible solution to (8.5).

Proof

(i) \leftarrow Suppose $\mathbf{x} = [x_1 \dots x_m \ 0 \ \dots \ 0]^T$ is a feasible solution. Then

$$x_1 \mathbf{a}_1 + \dots + x_m \mathbf{a}_m = \mathbf{b},$$

where $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$, the first columns of A , are linearly independent. Suppose that \mathbf{x} could be expressed as a convex combination of two other points in K , that is,

$$\mathbf{x} = \alpha \mathbf{y} + (1 - \alpha) \mathbf{z}$$

with $0 < \alpha < 1$ and $\mathbf{y} \neq \mathbf{z}$. Since $\mathbf{x}, \mathbf{y}, \mathbf{z} \geq \mathbf{0}$, we have that the last $n-m$ components of \mathbf{y} and \mathbf{z} are zero. Thus, in particular, we have

$$y_1 \mathbf{a}_1 + \dots + y_m \mathbf{a}_m = \mathbf{b} \text{ and } z_1 \mathbf{a}_1 + \dots + z_m \mathbf{a}_m = \mathbf{b}.$$

Since the vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}$ are linearly independent, it follows that $\mathbf{x} = \mathbf{y} = \mathbf{z}$ and hence \mathbf{x} is an extreme point

(ii) \rightarrow Assume \mathbf{x} is an extreme point of K . Let us assume that the nonzero components of \mathbf{x} are the first k components, so

$$x_1 \mathbf{a}_1 + \dots + x_k \mathbf{a}_k = \mathbf{b}$$

with $k \leq n$ and $x_i > 0$ for $i = 1, 2, \dots, k$. To show that \mathbf{x} is a basic feasible solution it must be shown that $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ are linearly independent. We do this by contradiction

Suppose $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ are linearly dependent, then

$$y_1 \mathbf{a}_1 + \dots + y_k \mathbf{a}_k = \mathbf{0} \text{ with } y_i \neq 0 \text{ for } i = 1, 2, \dots, k.$$

Let us define vector $\mathbf{y} = [y_1 \ \dots \ y_k \ 0 \ \dots \ 0]^T$. Since $x_i > 0$ for $i = 1, 2, \dots, k$, it is always possible to find ε such that

$$\mathbf{x} + \varepsilon \mathbf{y} \geq \mathbf{0} \text{ and } \mathbf{x} - \varepsilon \mathbf{y} \geq \mathbf{0}.$$

We then have $\mathbf{x} = \frac{1}{2}(\mathbf{x} + \varepsilon \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \varepsilon \mathbf{y})$ which expresses \mathbf{x} as a convex combination of two distinct vectors in K . This cannot occur since \mathbf{x} is an extreme point. Thus $\{\mathbf{a}_1, \dots, \mathbf{a}_k\}$ are linearly independent and \mathbf{x} is a basic feasible solution.

□

Example: A production planning problem

Suppose an industrial plant has three types (M_1, M_2 and M_3) of machines, each of which must be used in manufacturing the plant's products, of which there are two products (P_1 and P_2). The problem is to decide how many of each product to produce each week so as to maximize weekly profits. To formulate the problem we assume several data, summarized in the following table

Machine Type	Hours Needed by One Unit of P ₁	Hours Needed by One Unit of P ₂	Total Hours Available
M ₁	2	1	70
M ₂	1	1	40
M ₃	1	3	90
Profit per unit of P ₁ = 40		Profit per unit of P ₂ = 60	

Let denote x_1 the number of units of P₁ to be produced each week while x_2 denotes the number of units of P₂. Then, the mathematical version of our planning program is

$$\begin{aligned} & \text{maximize } M = 40x_1 + 60x_2 \\ & \text{subject to } \begin{cases} 2x_1 + x_2 \leq 70 \\ x_1 + x_2 \leq 40 \\ x_1 + 3x_2 \leq 90 \\ \text{and } x_1 \geq 0, x_2 \geq 0 \end{cases}, \\ & \quad (8.6) \end{aligned}$$

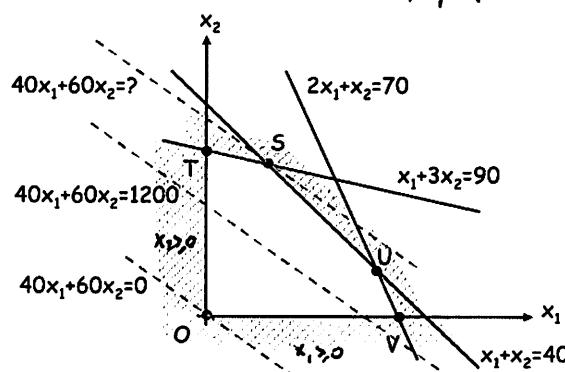
which in matrix notation is equivalent to

$$\begin{aligned} & \text{maximize } M = c^T x \\ & \text{subject to } Ax \leq b \\ & \quad x \geq 0 \end{aligned}$$

where

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 3 \end{bmatrix}; \quad b = \begin{bmatrix} 70 \\ 40 \\ 90 \end{bmatrix}; \quad c = \begin{bmatrix} 40 \\ 60 \end{bmatrix}; \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

The constraint set is illustrated in next Figure (convex polytope)



It is geometrically obvious that solutions will always occur at vertices of the constraint set, no matter what c might equal: a last point at which the graph of $M = c^T x$ intersects the constraint set will always be a vertex. A possible approach to solving linear programs is simply evaluate $c^T x$ at each vertex and keep the maximum, but this technique would result too expensive computationally since a general problem with n variables and m constraints has at most

$\{O, T, S, U, V\}$ non extreme points: possibles punts q. maximizar la funció.

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}$$

extreme points solutions.

The simplex method uses the basic idea of examining the vertices but avoiding examining vertices with smaller values of $c^T x$ than have already been found. Thus, while the first technique would test all five vertices {O,S,T,U,V}, the simplex method starting at O might examine {O,T,S}. But you need to identify all vertices anyway! How do you know which ones are the largest and the smallest? This is done in the simplex.

A more recent approach is the Karmarkar's method and other interior methods claimed to be several times faster than the simplex methods on special classes of linear programming problems.

8.2. THE SIMPLEX METHOD

Let us solve the problem of (8.6) by the simplex method.

$$\begin{aligned} & \text{maximize } M = 40x_1 + 60x_2 \\ \text{subject to } & \begin{cases} 2x_1 + x_2 \leq 70 \\ x_1 + x_2 \leq 40 \\ x_1 + 3x_2 \leq 90 \end{cases} \quad \text{and } x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

The simplex method moves from extreme point to extreme point and thus must start at an extreme point, that is, a basic feasible solution. The slack variables provide a basic feasible solution.

$$\begin{aligned} & \text{maximize } M = 40x_1 + 60x_2 \\ \text{subject to } & \begin{cases} 2x_1 + x_2 + x_3 = 70 \\ x_1 + x_2 + x_4 = 40 \\ x_1 + 3x_2 + x_5 = 90 \end{cases} \\ \text{and } & x_i \geq 0 \quad i=1, \dots, 5 \end{aligned} \tag{8.7}$$

We have a constraint system of 3 equations and 5 unknowns, so a basic solution $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]^T$ should have 2 components equal to 0. Notice that (8.7) provide a starting feasible solution:

$$x_1 = x_2 = 0; x_3 = 70; x_4 = 40; x_5 = 90$$

Now we include the objective function in the system of equations, leading to

$$\begin{array}{ccccccc} M & x_1 & x_2 & x_3 & x_4 & x_5 & \\ \left[\begin{array}{cccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 70 \\ 0 & 1 & 1 & 0 & 1 & 0 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 90 \\ 1 & -40 & -60 & 0 & 0 & 0 & 0 \end{array} \right] \end{array}$$

Since $x_1 = x_2 = 0$ at this point $M = 0$. To increase $M = 40x_1 + 60x_2$, we can make either x_1 or x_2 positive. We decide to increase x_2 since it is multiplied by 60. We keep $x_1 = 0$ but change x_2 in order to be a basic variable (that is, put all zeros in the column associated with x_2 except one component equal to 1). This implies changes in the rest of variables (x_3, x_4, x_5). The pivot is selected in such a way that M is increased as much as x_2 may increase it, but without leading to a non-positive solution of \mathbf{x} .

$$\begin{array}{cccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 70 & 70/1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 40 & 40/1 \\ 0 & 1 & 3 & 0 & 0 & 1 & 90 & 90/3 \text{ the lowest} \\ 1 & -40 & -60 & 0 & 0 & 0 & 0 & \text{positive!} \end{array}$$

$$\begin{array}{ccccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 70 \\ 0 & 1 & 1 & 0 & 1 & 0 & 40 \\ 0 & 1/3 & 1 & 0 & 0 & 1/3 & 30 \\ 1 & -40 & -60 & 0 & 0 & 0 & 0 \end{array} \Rightarrow \begin{array}{ccccccc|c} 0 & 5/3 & 0 & 1 & 0 & -1/3 & 40 \\ 0 & 2/3 & 0 & 0 & 1 & -1/3 & 10 \\ 0 & 1/3 & 1 & 0 & 0 & 1/3 & 30 \\ 1 & -20 & 0 & 0 & 0 & 20 & 1800 \end{array}$$

Now, we have the basic feasible solution:

$$x_1 = 0; x_2 = 30; x_3 = 40; x_4 = 10; x_5 = 0 \text{ and } M = 1800.$$

We can proceed with x_1 in the same way.

$$\begin{array}{ccccccc|c} 0 & \frac{5}{3} & 0 & 1 & 0 & -\frac{1}{3} & 40 \\ 0 & \frac{2}{3} & 0 & 0 & 1 & -\frac{1}{3} & 10 \\ 0 & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & 30 \\ 1 & -20 & 0 & 0 & 0 & 20 & 1800 \end{array} \Rightarrow \begin{array}{ccccccc|c} 0 & \frac{5}{3} & 0 & 1 & 0 & -\frac{1}{3} & 40 \\ 0 & 1 & 0 & 0 & \frac{3}{2} & -\frac{1}{2} & 15 \\ 0 & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & 30 \\ 1 & -20 & 0 & 0 & 0 & 20 & 1800 \end{array}$$

$$\begin{array}{ccccccc|c} 0 & 0 & 0 & 1 & -\frac{5}{2} & \frac{1}{2} & 15 \\ 0 & 1 & 0 & 0 & \frac{3}{2} & -\frac{1}{2} & 15 \\ 0 & 0 & 1 & 0 & -\frac{1}{2} & \frac{1}{2} & 25 \\ 1 & 0 & 0 & 0 & 30 & 10 & 2100 \end{array}.$$

The final solution is

$$x_1 = 15; x_2 = 25; x_3 = 15; x_4 = 0; x_5 = 0 \text{ and } M = 2100$$

ARTIFICIAL VARIABLES

An initial basic feasible solution is not always apparent for all the types of linear programs. The so-called artificial variables provide the way to initiate the simplex method. Any linear programming problem can be always expressed in the form

$$Ax = b \text{ and } x \geq 0 \quad (8.8)$$

with $b \geq 0$. In order to find an initial basic feasible solution to (8.8), we consider the artificial minimization problem

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^m y_i \\ & \text{subject to} && \begin{cases} Ax + y = b \\ x \geq 0 \\ y \geq 0 \end{cases} \end{aligned} \quad (8.9)$$

where $y = [y_1 \dots y_m]^T$ is a vector of artificial variables. If there is a feasible solution to (8.8), then it is clear that (8.9) has a minimum of zero with $y = 0$. If (8.8) has no feasible solution, then the minimum value of (8.9) is greater than zero. Therefore the search of an initial basic feasible solution of (8.8) becomes itself a linear programming problem as established in (8.9).

Example

Consider the previous program with a modification

$$\begin{aligned} & \text{maximize} && M = 40x_1 + 60x_2 \\ & \text{subject to} && \begin{cases} 2x_1 + x_2 \leq 70 \\ -x_1 - x_2 \leq -40 \\ x_1 + 3x_2 \leq 90 \end{cases} \quad \text{and } x_1 \geq 0, x_2 \geq 0 \end{aligned} \quad (8.10)$$

If we introduce the slack variables in the constraint system as follows

$$\begin{cases} 2x_1 + x_2 + x_3 = 70 \\ -x_1 - x_2 + x_4 = -40 \\ x_1 + 3x_2 + x_5 = 90 \end{cases} \quad x_i \geq 0 \quad i = 1, \dots, 5$$

we do not obtain the expected feasible solution, since with $x_1 = x_2 = 0$ the slack variable $x_4 = -40$ and the positive constraint is not fulfilled. Then we introduce an artificial variable

$$\begin{cases} 2x_1 + x_2 + x_3 = 70 \\ -x_1 - x_2 + x_4 - x_6 = -40 \\ x_1 + 3x_2 + x_5 = 90 \end{cases} \quad x_i \geq 0 \quad i = 1, \dots, 6$$

and define the maximization function $M' = -x_6$. The solution of the linear programming problem

$$\begin{aligned} & \text{maximization } M' = -x_6 \\ & \text{subject to } \begin{cases} 2x_1 + x_2 + x_3 = 70 \\ -x_1 - x_2 + x_4 - x_6 = -40 \\ x_1 + 3x_2 + x_5 = 90 \end{cases} \\ & \text{and } x_i \geq 0 \quad i=1, \dots, 6 \end{aligned}$$

leads to a basic feasible solution, if it exists. We proceed as follows

$$\left[\begin{array}{cccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 0 & 70 \\ 0 & 1 & 1 & 0 & -1 & 0 & 1 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 0 & 90 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \Rightarrow \left[\begin{array}{cccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 0 & 70 \\ 0 & 1 & 1 & 0 & -1 & 0 & 1 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 0 & 90 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & -40 \end{array} \right]$$

Since $M' = -40 + x_1 + x_2$, we try to increase for instance x_1 .

$$\left[\begin{array}{cccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 35 \\ 0 & 1 & 1 & 0 & -1 & 0 & 1 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 0 & 90 \\ 1 & -1 & -1 & 0 & 1 & 0 & 0 & -40 \end{array} \right] \Rightarrow \left[\begin{array}{cccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 35 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & -1 & 0 & 1 & 5 \\ 0 & 0 & \frac{3}{2} & \frac{1}{2} & 0 & 1 & 0 & 55 \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 & -5 \end{array} \right]$$

Now we are at the solution $x_1 = 35; x_2 = x_3 = x_4 = 0; x_5 = 55; x_6 = 5$ and $M' = -5 + \frac{1}{2}x_2 - \frac{1}{2}x_3 - x_4$. We may further increase M' increasing x_2 .

$$\left[\begin{array}{cccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 35 \\ 0 & 0 & 1 & -1 & -2 & 0 & 2 & 10 \\ 0 & 0 & \frac{5}{2} & \frac{1}{2} & 0 & 1 & 0 & 55 \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 & -5 \end{array} \right] \Rightarrow \left[\begin{array}{cccccc|c} 0 & 1 & 0 & 1 & 1 & 0 & -1 & 30 \\ 0 & 0 & 1 & -1 & -2 & 0 & 2 & 10 \\ 0 & 0 & 0 & 2 & 5 & 1 & -5 & 30 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{array} \right]$$

We obtain a basic feasible solution $x_1 = 30; x_2 = 10; x_3 = x_4 = 0; x_5 = 30; x_6 = 0$ of the initial problem of (8.10) formulated with slack variables as follows

$$\text{maximization } M = 40x_1 + 60x_2$$

$$\begin{aligned} & \text{subject to } \begin{cases} 2x_1 + x_2 + x_3 = 70 \\ +x_1 + x_2 - x_4 = 40 \\ x_1 + 3x_2 + x_5 = 90 \end{cases} \\ & \text{and } x_i \geq 0 \quad i=1, \dots, 5 \end{aligned}$$

Then, we have

$$\left[\begin{array}{cccccc|c} 0 & 2 & 1 & 1 & 0 & 0 & 70 \\ 0 & 1 & 1 & 0 & -1 & 0 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 90 \\ 1 & -40 & -60 & 0 & 0 & 0 & 0 \end{array} \right]$$

We know there is a basic feasible solution with basic variables x_1, x_2, x_5 . So we put all zeroes in the columns associated with these variables except a single entry equal to 1.

$$\left[\begin{array}{ccccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 35 \\ 0 & 1 & 1 & 0 & -1 & 0 & 40 \\ 0 & 1 & 3 & 0 & 0 & 1 & 90 \\ 1 & -40 & -60 & 0 & 0 & 0 & 0 \end{array} \right] \Rightarrow \left[\begin{array}{ccccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 35 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & -1 & 0 & 5 \\ 0 & 0 & \frac{3}{2} & \frac{1}{2} & 0 & 1 & 55 \\ 1 & 0 & -40 & 20 & 0 & 0 & 1400 \end{array} \right]$$

$$\left[\begin{array}{ccccccc|c} 0 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 35 \\ 0 & 0 & 1 & -1 & -2 & 0 & 10 \\ 0 & 0 & \frac{3}{2} & \frac{1}{2} & 0 & 1 & 55 \\ 1 & 0 & -40 & 20 & 0 & 0 & 1400 \end{array} \right] \Rightarrow \left[\begin{array}{ccccccc|c} 0 & 1 & 0 & 1 & 1 & 0 & 30 \\ 0 & 0 & 1 & -1 & -2 & 0 & 10 \\ 0 & 0 & 0 & 2 & 5 & 1 & 30 \\ 1 & 0 & 0 & -20 & -80 & 0 & 1800 \end{array} \right]$$

We increase x_3

$$\left[\begin{array}{ccccccc|c} 0 & 1 & 0 & 1 & 1 & 0 & 30 \\ 0 & 0 & 1 & -1 & -2 & 0 & 10 \\ 0 & 0 & 0 & 1 & \frac{5}{2} & \frac{1}{2} & 15 \\ 1 & 0 & 0 & -20 & -80 & 0 & 1800 \end{array} \right] \Rightarrow \left[\begin{array}{ccccccc|c} 0 & 1 & 0 & 0 & \frac{-3}{2} & \frac{-1}{2} & 15 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & 25 \\ 0 & 0 & 0 & 1 & \frac{5}{2} & \frac{1}{2} & 15 \\ 1 & 0 & 0 & 0 & -30 & 10 & 2100 \end{array} \right]$$

Finally, we increase x_4

$$\left[\begin{array}{ccccccc|c} 0 & 1 & 0 & 0 & \frac{-3}{2} & \frac{-1}{2} & 15 \\ 0 & 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & 25 \\ 0 & 0 & 0 & \frac{3}{5} & 1 & \frac{1}{5} & 6 \\ 1 & 0 & 0 & 0 & -30 & 10 & 2100 \end{array} \right] \Rightarrow \left[\begin{array}{ccccccc|c} 0 & 1 & 0 & 0 & 0 & \frac{-1}{5} & 24 \\ 0 & 0 & 1 & 0 & \frac{-1}{5} & 0 & 22 \\ 0 & 0 & 0 & \frac{3}{5} & 1 & \frac{1}{5} & 6 \\ 1 & 0 & 0 & 12 & 0 & 16 & 2280 \end{array} \right]$$

The final solution is

$$x_1 = 24; x_2 = 22; x_4 = 6; x_3 = x_5 = 0 \text{ and } M = 2280.$$

SECTION I

Exercise 1:

Theorem. Let $B = \{v_1, \dots, v_p\}$ and $B' = \{v'_1, \dots, v'_p\}$ be two ordered bases for the same p -dimensional vector space V . For any vector v in V , the B -coordinates v_B of v are related by

$$v_B = Mv_{B'} \text{ and } v_{B'} = M^{-1}v_B,$$

where M is the $p \times p$ nonsingular matrix with $(M)_{ij} = m_{ij}$ where $v'_j = m_{ij}v_i + \dots + m_{pj}v_p$.

Exercise. Suppose that

$$B = \{1; r, r^2\} \text{ and } B' = \{1; 1+r, 1+r^2\}$$

are two ordered bases for P^3 .

- (a) Find the Matrix M that translates between coordinates with respect to these two bases.

- (b) Find the coordinates of $7 - 3t + 4t^2$ with respect to B and using the change of basis find them with respect to B' .

Exercise 2

If, in partitioned form,

$$A = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$$

where A and P are nonsingular, prove that

$$A^{-1} = \begin{bmatrix} X & -P^{-1}QW \\ -WRP^{-1} & W \end{bmatrix}$$

where

$$W = (S - RP^{-1}Q)^{-1} \quad X = P^{-1} + P^{-1}QWRP^{-1}$$

Similarly, if A and S are nonsingular, prove that

$$A^{-1} = \begin{bmatrix} X & -XQS^{-1} \\ -S^{-1}RX & W \end{bmatrix}$$

where

$$X = (P - QS^{-1}R)^{-1} \quad W = S^{-1} + S^{-1}RXQS^{-1}.$$

If P and S are both nonsingular, prove directly that the forms are equivalent. Show that this proves that

$$(S - RP^{-1}Q)^{-1} = W = S^{-1} + S^{-1}RXQS^{-1},$$

a result, called the “matrix inversion lemma”, that is useful in applications.

Exercise 3:

If the entries $\langle A \rangle_{ij}$ of a matrix A are functions of a variable t , then we define the derivative $\frac{dA}{dt}$ as the obvious matrix of derivatives: $\langle \frac{dA}{dt} \rangle_{ij} = \frac{d\langle A \rangle_{ij}}{dt}$

- (a) If AB is defined, show that $\frac{d(AB)}{dt} = A(\frac{dB}{dt}) + (\frac{dA}{dt})B$.
 (b) If A is nonsingular, differentiate $AA^{-1} = I$ so as to obtain

$$\frac{d(A^{-1})}{dt} = -A^{-1}(\frac{dA}{dt})A^{-1}.$$

Exercise 4:

For a square matrix A , define the *trace* of A , $\text{tr}(A)$, as the sum of the entries on the main diagonal of A .

- (a) Prove that $\text{tr}(AB) = \text{tr}(BA)$.

(b) Prove that $\text{tr}(\mathbf{S}^{-1}\mathbf{A}\mathbf{S}) = \text{tr}(\mathbf{A})$ if \mathbf{S} is nonsingular.

Exercise 5:

Definition. The $p \times p$ elementary rotation matrix $\mathbf{R} = \mathbf{R}_{xz}(\alpha)$ is defined by:

- $\langle \mathbf{R} \rangle_{kk} = \langle \mathbf{R} \rangle_{nn} = \cos(\alpha)$
- $\langle \mathbf{R} \rangle_{kn} = -\sin(\alpha)$
- $\langle \mathbf{R} \rangle_{nk} = \sin(\alpha)$
- $\langle \mathbf{R} \rangle_{ij} = 1$ for $i \neq k, n$
- $\langle \mathbf{R} \rangle_{ij} = 0$, for all entries other than those defined above.

Theorem (rotation matrices). Let $\mathbf{R} = \mathbf{R}_{xz}(\alpha)$ be a $p \times p$ elementary rotation matrix. Then:

- \mathbf{R} is orthogonal
- For each \mathbf{x} , \mathbf{Rx} equals the vector that results from rotating \mathbf{x} through an angle α in the x_1 - x_n -plane in the direction from the x_1 -axis towards the x_n -axis.
- $\det \mathbf{R} = 1$.

Exercise. Let \mathbf{v} be the vector $\mathbf{v} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$. Use MATLAB or similar software to

find the result of rotating \mathbf{v} through an angle of 15° in all possible planes xy , xz , yz .

Exercise 6:

A Householder matrix is a $p \times p$ matrix \mathbf{H} of the form $\mathbf{H} = \mathbf{I}_p - 2\mathbf{w}\mathbf{w}^H$ for a real nonzero $p \times 1$ vector \mathbf{w} .

- (a) Show that every Householder matrix is symmetric.

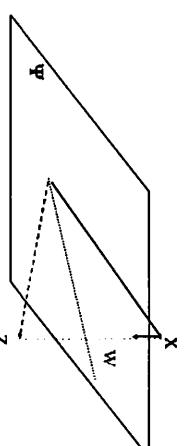
- (b) Show that every Householder matrix is non-singular and that $\mathbf{H}^{-1} = \mathbf{H}$.

- (c) Use (b) to show that the columns of a Householder matrix form an orthonormal set.

Exercise 7:

Geometrical Interpretation

The reflection of \mathbf{x} across the plane Ψ is the vector \mathbf{z} that lies on the other side of Ψ from \mathbf{x} , exactly the same distance from Ψ as is \mathbf{x} , and having the same projection into Ψ as \mathbf{x} . That is, the difference between \mathbf{x} and its reflection is exactly twice the projection of \mathbf{x} along the unitary vector \mathbf{w} normal to Ψ .



Graphically, we can see that \mathbf{z} is a linear combination of \mathbf{x} and \mathbf{w} .

This observation yields the equation:

$$\mathbf{z} = \mathbf{x} + \alpha \mathbf{w},$$

where

$$\alpha = -2\mathbf{w}^H \mathbf{x}$$

so,

$$\mathbf{z} = \mathbf{x} - 2\mathbf{w}^H \mathbf{x} \mathbf{w} = \mathbf{x} - 2\mathbf{w}\mathbf{w}^H \mathbf{x} = \mathbf{H}\mathbf{x}.$$

\mathbf{H} is the Householder matrix.

- (d) Compute the reflection of the vector $\mathbf{v} = (-1; 3; -4)$ across the plane $2x - y + 7z = 0$.

The computations of the Gram-Schmidt algorithm can be reorganized to be more stable numerically. In these modified computations, a column of \mathbf{Q} and a row of \mathbf{R} are produced at each iteration. (The regular Gram-Schmidt process produces

a column of \mathbf{Q} and a column of \mathbf{R} at each iteration.) Let the k th column of \mathbf{Q} be denoted as \mathbf{q}_k and let the k th row of \mathbf{R} be denoted as \mathbf{r}_k^T .

(a) Show that for an $m \times n$ matrix \mathbf{A} ,

$$\mathbf{A} - \sum_{i=1}^{k-1} \mathbf{q}_i \mathbf{r}_i^T = \sum_{i=k}^n \mathbf{q}_i \mathbf{r}_i^T = [\mathbf{0} \quad \mathbf{A}^{(k)}],$$

where $\mathbf{A}^{(k)}$ is $m \times (n-k+1)$.

(b) Let $\mathbf{A}^{(k)} = [\mathbf{I}_k \quad \mathbf{B}]$, where \mathbf{B} is $m \times n-k$, and explain why the k th column of \mathbf{Q} and the k th row of \mathbf{R} are given by

$$\mathbf{r}_k = \|\mathbf{e}_k\| \quad \mathbf{q}_k = \frac{\mathbf{e}_k}{\|\mathbf{e}_k\|} \quad r_{1,1}, \dots, r_{1,n}$$

(c) Then show that the next iteration can be started by computing

$$\mathbf{A} - \sum_{i=1}^k \mathbf{q}_i \mathbf{r}_i^T = [\mathbf{0} \quad \mathbf{A}^{(k+1)}],$$

where $\mathbf{A}^{(k+1)} = \mathbf{B} - \mathbf{q}_k (r_{1,k+1}, \dots, r_{1,n})$.

(d) Code the modified Gram-Schmidt algorithm in MATLAB.

* Ex M. Programa funció per calcular determinants

X Exercise 8:

Definition. The projection matrix \mathbf{P}_0 satisfies:

1. \mathbf{P}_0 is symmetric (or hermitian).
2. $\mathbf{P}_0^2 = \mathbf{P}_0$.
3. $\mathbf{P}_0 (\mathbf{I}_p - \mathbf{P}_0) = (\mathbf{I}_p - \mathbf{P}_0) \mathbf{P}_0 = 0$.
4. $(\mathbf{I}_p - \mathbf{P}_0) \mathbf{Q} = 0$.

Exercise. Prove all these properties.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad \left(\begin{array}{cccc} + & + & + & \\ + & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{array} \right)$$

$$|A| = \sum_{j=1}^n a_{ij} A_{ij} = \sum_{j=1}^n (-1)^{i+j} |M_{ij}| \quad \text{on } A_{ij} = (-1)^{i+j} |M_{ij}|$$

per files

ordenant l'element de a_{ij} següent la Laplace.

X Exercise 9:

Definition. \mathbf{A} is said to be idempotent if $\mathbf{A}^2 = \mathbf{A}$.

Exercise. Show that if an $n \times n$ matrix \mathbf{A} is idempotent, then

- (a) for any $n \times n$ non-singular matrix \mathbf{B} , $\mathbf{B}^{-1} \mathbf{A} \mathbf{B}$ is idempotent;
- (b) for any integer k greater than or equal to 2, $\mathbf{A}^k = \mathbf{A}$.

Exercise 10:

Theorem. Let \mathbf{T} represent an $m \times m$ matrix, \mathbf{U} an $m \times m$ matrix, and \mathbf{W} an $n \times n$ matrix. Then,

$$\begin{vmatrix} T & 0 \\ V & W \end{vmatrix} = \begin{vmatrix} W & V \\ 0 & T \end{vmatrix} = |T||W|$$

Exercise. Let \mathbf{T} represent an $m \times m$ matrix, \mathbf{U} an $m \times n$ matrix, \mathbf{V} an $n \times m$ matrix, and \mathbf{W} an $n \times n$ matrix. If \mathbf{T} is nonsingular, prove that

$$\begin{vmatrix} T & U \\ V & W \end{vmatrix} = \begin{vmatrix} W & V \\ U & T \end{vmatrix} = |T||W - VT^{-1}U|$$

SECTION II

Exercise 5:

Let

$$F = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{\frac{-j2\pi}{N}} & e^{\frac{-j2\pi(N-1)}{N}} & & \\ e^{\frac{-j4\pi}{N}} & \dots & e^{\frac{-j2\pi(2N-1)}{N}} & \\ \vdots & & \vdots & \\ e^{\frac{-j2\pi(N-1)}{N}} & \dots & e^{\frac{-j2\pi(N-1)^2}{N}} & \end{bmatrix}$$

If $f(x)$ is a polynomial in x and A is a square matrix, then $f(A)$ denotes the matrix obtained by replacing x^i by A^i (and x^0 by I) in the formula for $f(x)$. Prove that $f(\lambda)$ is an eigenvalue of $f(A)$ associated with the eigenvector x if λ is an eigenvalue of A associated with x .

Exercise 2:

Let Λ be a diagonal matrix with diagonal entries $\{\lambda_1, \dots, \lambda_n\}$, $\Lambda = \text{diag}\{\lambda_i, i=1, \dots, n\}$. Prove that its characteristic polynomial $\alpha(\Lambda) = 0$, and show that this result also holds for matrices of the form $A = P\Lambda P^{-1}$.

The (i,j) th element of this is $f_{ij} = e^{\frac{-j2\pi(i-j(N-1))}{N}}$. For a vector $x = [x_0, \dots, x_{n-1}]^T$, the product $X = Fx$ is the DFT of x .

Prove that the matrix $\frac{F}{\sqrt{N}}$ is unitary.

Hint: use the following.

$$\sum_{n=0}^{N-1} e^{\frac{j2\pi kn}{N}} = \begin{cases} N & k = 0 \bmod N \\ 0 & k \neq 0 \bmod N \end{cases}$$

Exercise 6:

A circulant matrix C has the form

$$C = \begin{bmatrix} c_1 & c_2 & \dots & c_n \\ c_n & c_1 & \dots & c_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & \dots & c_1 \end{bmatrix}.$$

*a circulant matrix is normal
 $C^H C = C C^H$
 2. n Poden
 diagonalisieren*

Exercise 4:

Let A be an $n \times n$ hermitian matrix and $\lambda_1 \leq \dots \leq \lambda_n$ its eigenvalues. Prove that

$$\lambda_1 \leq \frac{x^H A x}{x^H x} \leq \lambda_n$$

for any vector x .

Hint: the function $\left(\frac{x^H A x}{x^H x} \right)$ is known as the Rayleigh quotient.

$$C = \frac{1}{m} F A F^H, \quad (1)$$

where F is the so-called Fourier transformed matrix defined in exercise 5 and A is a diagonal matrix with the values

$$\Lambda = \text{diag}\{p_r(1), p_r(a), \dots, p_r(a^{n-1})\}.$$

The function $p_r(z)$ is the polynomial

$$P_c(z) = c_1 + c_2 z + \dots + c_m z^{m-1}.$$

Hint: since $\mathbf{F}\mathbf{F}^H = m\mathbf{I}$, note that (1) is equivalent to

$$\frac{1}{m} \mathbf{F}^H \mathbf{C} \mathbf{F} = \mathbf{A}.$$

Exercise 7:

We aim at analyzing the sensitivity of the eigenvalues and eigenvectors of a correlation matrix obtained from measures. Consider the next signal

$$x(n) = s(n)\cos(2\pi fn) + w(n),$$

with $s(n) = \cos(2\pi fn)$ and $w(n)$ a zero-mean white Gaussian noise with variance equal to one (the MATLAB instruction to generate it is `randn(m)`).

We propose the following three correlation matrix estimators:

$$1. \quad \mathbf{R} = \frac{1}{N} \sum_{i=1}^N \mathbf{S}(i) \mathbf{S}(i)^T + \text{eye}(4) \text{ with } N = 1000;$$

$$2. \quad \mathbf{R} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}(i) \mathbf{X}(i)^T \text{ with } N = 10;$$

$$3. \quad \mathbf{R} = \frac{1}{N} \sum_{i=1}^N \mathbf{X}(i) \mathbf{X}(i)^T \text{ with } N = 100;$$

with

$$\mathbf{S}(i)^T = [s(i) \ s(i+1) \ s(i+2) \ s(i+3)]$$

$$\mathbf{X}(i)^T = [x(i) \ x(i+1) \ x(i+2) \ x(i+3)].$$

Design a MATLAB routine that computes the unitary eigenvectors and eigenvalues of the three correlation matrices for $\omega = 0.17$.

Hint: use the MATLAB function `eig`.

Exercise 8:

Design a routine in MATLAB that computes the k largest eigenvalues and associated unitary eigenvectors of a $n \times n$ hermitian matrix using the Frost algorithm.

Exercise 9:

Let the column vectors of $\mathbf{A}_{M,N}$ be the signal subspace generators inside the space \mathbb{R}^M and let column vectors of $\mathbf{H}_{M \times (m-rank(\mathbf{A}))}$ be the subspace generators orthogonal to the signal subspace.

(a) Show, based on the SVD decomposition of \mathbf{A} , that the projection matrix \mathbf{P}_A accomplishes:

$$(\mathbf{I} + \alpha \mathbf{P}_A)^{-1} = \mathbf{I} - \frac{\alpha}{1+\alpha} \mathbf{P}_A$$

(b) Calculate the pseudoinverse matrix of $\mathbf{I} + \alpha \mathbf{P}_A$. Particularize the two above expressions for $\alpha = -1$.

(c) Using the SVD decomposition of $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^H$, matrices \mathbf{U} and \mathbf{V} could be decomposed in $\mathbf{U} = [\mathbf{U}_A \quad \mathbf{U}_W]$ and $\mathbf{V} = [\mathbf{V}_A \quad \mathbf{V}_W]$ respectively. \mathbf{U}_A and \mathbf{V}_A contain the vectors associated to non-zero singular values of \mathbf{A} . Obtain the expression of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and \mathbf{D} in

$$\begin{bmatrix} \mathbf{U}_A \\ \mathbf{U}_W \end{bmatrix} = \mathbf{P}_A \begin{bmatrix} \mathbf{U}_A & \mathbf{U}_W \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

(d) Let $\mathbf{A}_{M,N}$ be the matrix with $N = 5$ partitions of $M = 20$ points each of a realization of the process $x(n) = \cos(2\pi fn) + w(n)$, where $w(n)$ is a white Gaussian noise with variance σ_n^2 and $n = 1:100$. Show all the above relations using MATLAB identifying the signal subspace from the single values of the matrix \mathbf{A} in each relation for next values: $f = 0.17$ and $f = 0.25$ and for $SNR = 0.10dB$ and $SNR = 100dB$

SECTION III

$(a_{11} \dots a_{NN})$ with the highest absolute value. Use the routine designed in exercise 2.1 in order to permute the corresponding rows. Minimise the number of flops.

Exercise 1:

The population of a country has grown in last decades as follows:

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990
Pop. Mill.	76.0	92.0	105.7	123.2	131.7	150.7	179.3	203.2	226.5

We aim at predicting the population in the year 2000. For this, we propose to model it by square curve fitting:

t = time in years

y = millions of inhabitants

$$y(t) = a + bt + ct^2$$

Working with MATLAB:

- (a) By using the singular value decomposition (SVD), find the least square solution (LSQ) of the prediction model, i.e., $[a \ b \ c]^T$. Predict the population in the year 2000.
- (b) By using the SVD decomposition, find the total least square solution (TLS) of the prediction model, i.e., $[a \ b \ c]^T$. Predict the population in the year 2000.

Exercise 2:

Y 2.1. Design a routine in MATLAB which permutes the rows of any input matrix.

Use the following variables:

Input variables:

- * $N \times N$ real square matrix A .
- * Permutation vector of dimension N .

Output variables: ($B = PA$)

- * Permutation matrix P .
- * Permutated matrix B .

2.2. Design a routine in MATLAB which computes the LU decomposition of any input matrix. Use a_i as a pivot where a_i is the element of the k^{th} -column

Input variables:

* real square matrix A .

Output variables: ($PA = LU$)

* Lower triangular matrix L .

* Upper triangular matrix U .

* Permutation Matrix P .

* Number of flops required by the routine.

Exercise 3:

Let $\{u_0, u_1, \dots, u_T\}$ be the sequence transmitted through a communication system using pulses $\{h_t\}$ for $t \in [0, \infty)$. The transmitted sequence which corresponds to the symbol u_n is $\{u_n h_{t-nT}\}$ where T is the so-called baud interval. The output of the composed source is written as:

$$s_t = \sum_{n=0}^T u_n h_{t-nT} \quad t = 0, 1, \dots$$

The signal type models a channel with ISI.

The noisy measure corresponds to p symbols transmitted:

$$y_t = y_t' + \sum_{m=0}^p u_m h_{t-mT} + n_t.$$

Theoretical development
 (a) If the output samples are represented in the vector $\mathbf{y} = \mathbf{H}\mathbf{u} + \mathbf{n}$, determine the samples of \mathbf{H} , \mathbf{u} and \mathbf{n} .

- (b) Find the LSQ estimation of \mathbf{u} .
- (c) Justify the suitability of using the QR decomposition in (b).
- (d) Justify the suitability of using the SVD decomposition in (b).

Practical development
 (e) Design the above estimation in MATLAB. Use a FIR filter \mathbf{h} , and an appropriate input vector \mathbf{u} . Obtain the LSQ and TLS solutions and compare them.

- (f) Make another estimation of the input vector by applying the QR decomposition.
- (g) Compare the obtained solutions and the computational efficiency in number of flops of the three solutions.

SECTION IV

Exercise 1:

In terms of estimation of parameters, θ , it is often interesting to calculate the Fisher Information Matrix $J(\theta)$. This matrix measures the quantity of information that data introduce in the estimation and that is why it is useful in some iterative optimization algorithms, such as the Newton's one, and to measure the lower bound that unbiased algorithms have (CRLB or Cramer-Rao Lower Bound).

Let $\mathbf{x} = (x_1, x_2, \dots, x_M)$ be the data vector and let x_i be the random variable where $x_i \sim N(m(\theta), R_i(\theta))$. Let $p(\theta; \mathbf{x})$ be the probability density function defined as:

$$p(\theta; \mathbf{x}) = (2\pi)^{-MK/2} |\mathbf{R}|^{-M/2} \exp\left\{-M/2 \text{tr}(\mathbf{R}^{-1}\mathbf{S})\right\}$$

$$\mathbf{S} = \frac{1}{M} \sum_{i=1}^M (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T$$

Suppose that we know \mathbf{R} (meaning independency from θ),

a) Show that the gradient vector is

$$\mathbf{S}(\theta, \mathbf{x}) = \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \frac{\partial \mathbf{m}}{\partial \theta} \mathbf{R}^{-1} \sum_{i=1}^M (\mathbf{x}_i - \mathbf{m})$$

b) Find the Fisher Matrix, whose components are defined as:

$$[\mathbf{J}(\theta)]_{ij} = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta_i \partial \theta_j} \right]$$

Exercise 2:

- a) Use artificial variables to find an initial basic feasible point for the linear program below.

b) Solve it.

$$\begin{aligned} & \text{Maximize } x_1 + x_2 \\ & \text{Subject to the constraints} \\ & \quad -x_1 + x_2 \leq 10 \\ & \quad x_1 + x_2 \geq 5 \\ & \quad 2x_1 + x_2 \leq 40 \\ & \quad x_1 \geq 0, \quad x_2 \geq 0. \end{aligned}$$

Exercise 3:

a) Express the following function in the quadratic form:

$$f(x, y, z) = 2x^2 + xy + y^2 + yz + z^2 - 6x - 7y - 8z + 9$$

Hint: the expression should follow this pattern

$$f(x, y, z) = \mathbf{x}^H \mathbf{A} \mathbf{x} + \mathbf{b}^H \mathbf{x} + c \text{ where } \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

- b) Using the first-order necessary conditions, find a minimum point of the function.
- c) Verify that the point is a relative minimum point by verifying that the second-order sufficiency conditions hold.
- d) Prove that the point is a global minimum point.

Exercise 4:

Design a routine in MATLAB which develops the Frost algorithm for the m largest eigenvalues.

Exercise 5:

Show that the quadratic function $\mathbf{x}^H \mathbf{Q} \mathbf{x} + \mathbf{b}^H \mathbf{x} + \mathbf{x}^H \mathbf{b} + c$, with $\mathbf{Q} = \mathbf{Q}^H$ hermitian and positive definite and $c \in \mathbb{R}$, is strictly convex.

Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(I.1)

$$B = \{1, t, t^2\}$$

$$B' = \{1, 1+t, 1+t+t^2\}$$

bases de P^3

a)

$$v_B = \begin{matrix} M \\ 3 \times 3 \end{matrix} v_{B'} \quad ; \quad v_{B'} = \begin{matrix} M^{-1} \\ 3 \times 3 \end{matrix} v$$

$$v = a + bt + ct^2$$

$$v_B = \begin{bmatrix} a \\ bt \\ ct^2 \end{bmatrix}$$

$$v = \sum_{j=1}^3 \alpha_j v_{B'_j} = \alpha_1 \cdot 1 + \alpha_2 t + \alpha_3 t^2$$

$$\begin{aligned} v &= \sum_{j=1}^3 \beta_j v_{B'_j} = \beta_1 \cdot 1 + \beta_2 (1+t) + \beta_3 (1+t+t^2) = \\ &= (\underbrace{\beta_1 + \beta_2 + \beta_3}_a) \cdot 1 + (\underbrace{\beta_2 + \beta_3}_b) t + \underbrace{\beta_3 t^2}_c \end{aligned}$$

$$\begin{cases} b'_1 = m_{11} b_1 + m_{21} b_2 + m_{31} b_3 = 1 \cdot b_1 \\ b'_2 = m_{12} b_1 + m_{22} b_2 + m_{32} b_3 = 1 \cdot b_1 + 1 \cdot b_2 \\ b'_3 = m_{13} b_1 + m_{23} b_2 + m_{33} b_3 = 1 \cdot b_1 + 1 \cdot b_2 + 1 \cdot b_3 \end{cases}$$

$$\begin{matrix} \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} \\ 3 \times 1 \end{matrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ t \\ t^2 \end{pmatrix} \quad \rightarrow \quad \begin{bmatrix} M \\ 3 \times 3 \end{bmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$b) \quad v = 7 - 3t + 4t^2$$

$$\begin{cases} v_B = (7, -3, 4)^T \\ v_{B'} = \begin{matrix} M^{-1} \\ 3 \times 3 \end{matrix} v_B = \begin{bmatrix} 7 \\ -10 \\ 7 \end{bmatrix} \end{cases}$$

I.2 $\underline{A} = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}$

a) $\underline{A} : P$ No singulars.

b) $\underline{A} : S$ " "

$$\underline{A}' \underline{A} = \underline{I}$$

a) $\begin{bmatrix} X & -P'QW \\ -WRP' & W \end{bmatrix} \begin{bmatrix} P & Q \\ R & S \end{bmatrix} = \begin{bmatrix} XP - P'QWR & XQ - P'QWS \\ -WRP'P + WRPR & -WRPQ + WRS \end{bmatrix} = \underline{I}$

$$XW - P'QWS = 0$$

I.3 $\underline{A} = \underline{A}(t) \quad \frac{d\underline{A}}{dt} \Rightarrow \frac{d\underline{A}}{dt} \Big|_{ij} = \frac{dA_{ij}}{dt}$

a) $\frac{dAB}{dt}$

$$\underline{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mp} \end{bmatrix}_{n \times p} \quad \underline{B} = \begin{bmatrix} b_{11} & \dots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{p1} & \dots & b_{pm} \end{bmatrix}_{p \times m} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}^T$$

$$\underline{AB} = \begin{bmatrix} a_1 b_1 & \dots & a_1 b_m \\ \vdots & \ddots & \vdots \\ a_n b_1 & \dots & a_n b_m \end{bmatrix}_{n \times m}$$

$$\frac{dAB}{dt} = \begin{bmatrix} a_1' b_1 + a_2' b_1 & \dots & a_1' b_m + a_2' b_m \\ \vdots & \ddots & \vdots \\ a_n' b_1 + a_1' b_1 & \dots & a_n' b_m + a_1' b_m \end{bmatrix} = \underline{A}' \underline{B} + \underline{A} \underline{B}'$$

b) $\frac{d(AA')} {dt} = \frac{dA}{dt} \underline{A}' + A \frac{dA'}{dt} \stackrel{\frac{dA}{dt} = 0}{=} 0 \rightarrow \frac{dA'}{dt} = -\underline{A}' \frac{dA}{dt} \underline{A}'$

I.4 \underline{A} square $\text{tr}(\underline{A}) = \sum \text{diag}(A) \quad c_{ii} = a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31}$

a) $\text{tr}\{AB\} = \text{tr}\{BA\}$ $\cancel{AB = BA} \quad (AB)_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \cancel{(AB)_{ij} = \sum_k b_{ik} a_{kj}}$

$$\underline{AB} = \begin{bmatrix} a_1 b_1 & & \\ & a_2 b_2 & \\ & & \ddots & \\ & & & a_n b_n \end{bmatrix}_{n \times n} \quad \underline{BA} = \begin{bmatrix} b_1 a_1 & & \\ & \ddots & \\ & & b_n a_n \end{bmatrix}$$

b) $\text{tr}(S^{-1}AS) = \text{tr}(A)$ $\Leftrightarrow S$ NO singular

$$\text{tr}(S^{-1}AS) = \text{tr}(SS^{-1}A) = \text{tr}(A)$$

Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(I.5)

elementary rotation matrix $\underline{\underline{R}} = \underline{\underline{R}}_{kn}(\alpha)$

(I.6) Householder matrix $p \times p$ $\underline{\underline{H}} = \underline{\underline{I}}_p - 2\underline{w}\underline{w}^H$ $\underline{w} \in \mathbb{R}^{p \times 1}$

a) Symmetric: $\underline{\underline{H}} = \underline{\underline{H}}^T$

$$(\underline{\underline{AB}})^T = \underline{\underline{B}}^T \underline{\underline{A}}^T$$

$$\underline{\underline{H}}^T = (\underline{\underline{I}}_p - 2\underline{w}\underline{w}^H)^T = (\underline{\underline{I}}^T - 2(\underline{w}\underline{w}^T)^T) = \underline{\underline{I}}_p - 2\underline{w}\underline{w}^T = \underline{\underline{H}}$$

b) $\underline{\underline{H}}^{-1} = \underline{\underline{H}}$

$$\underline{\underline{H}}\underline{\underline{H}}^{-1} = \underline{\underline{I}} \quad \underline{\underline{H}}^{-1} = (\underline{\underline{I}} - 2\underline{w}\underline{w}^H)^{-1} = \underline{\underline{I}} - \frac{1}{2}(\underline{w}\underline{w}^H)^{-1}$$

$$\begin{aligned} \underline{\underline{H}}\underline{\underline{H}}^{-1} &= (\underline{\underline{I}} - 2\underline{w}\underline{w}^H)(\underline{\underline{I}} - \frac{1}{2}(\underline{w}\underline{w}^H)^{-1}) = \underline{\underline{I}}\underline{\underline{I}} - \frac{1}{2}\underline{\underline{I}}(\underline{w}\underline{w}^H)^{-1} - 2\underline{w}\underline{w}^H\underline{\underline{I}} + \frac{1}{2}(\underline{w}\underline{w}^H)(\underline{w}\underline{w}^H)^{-1} \\ &= \underline{\underline{I}} - \end{aligned}$$

$$|\underline{\underline{H}}| = |\underline{\underline{I}}_p - 2\underline{w}\underline{w}^T| = |\underline{\underline{I}}_p| - 2|\underline{w}\underline{w}^T| = 1 - 2|R_{ww}| = 1$$

$$\underline{h}_i^H \underline{h}_i = \underline{\underline{I}}_p - 2$$

"0" hermitica

c) Orthonormal Set: $\underline{\underline{H}}\underline{\underline{H}}^H = (\underline{\underline{I}}_p - 2\underline{w}\underline{w}^H)(\underline{\underline{I}}^H - 2\underline{w}\underline{w}^H) = \underline{\underline{I}} - 2\underline{I}\underline{w}\underline{w}^H - 2\underline{w}\underline{w}^H\underline{\underline{I}} + 4\underline{w}\underline{w}^H\underline{w}\underline{w}^H$

I.7

Normalized Gram-Schmidt : $\underline{\underline{A}} = \underline{\underline{Q}} \underline{\underline{R}}$

$M \times N$ $M \times L$ $L \times N$

upper triangular.

$$\underline{\underline{A}} = [\underline{\underline{P}}_1, \underline{\underline{P}}_2, \dots, \underline{\underline{P}}_N] \xrightarrow{\text{orthogonal}} \underline{\underline{Q}} = [\underline{\underline{q}}_1, \underline{\underline{q}}_2, \dots, \underline{\underline{q}}_L]$$

$$\underline{\underline{q}}_k = \frac{\underline{\underline{e}}_k}{\|\underline{\underline{e}}_k\|} \quad \text{on} \quad \underline{\underline{e}}_k = \underline{\underline{P}}_k - \sum_{i=1}^{k-1} (\underline{\underline{q}}_i^H \underline{\underline{P}}_k) \underline{\underline{q}}_i$$

$$\underline{\underline{R}} = \begin{bmatrix} \|\underline{\underline{P}}_1\| & (\underline{\underline{q}}_1^H \underline{\underline{P}}_2) & (\underline{\underline{q}}_1^H \underline{\underline{P}}_3) & \dots & (\underline{\underline{q}}_1^H \underline{\underline{P}}_N) \\ 0 & \|\underline{\underline{e}}_2\| & (\underline{\underline{q}}_2^H \underline{\underline{P}}_3) & \dots & (\underline{\underline{q}}_2^H \underline{\underline{P}}_N) \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & (\underline{\underline{q}}_L^H \underline{\underline{P}}_N) \end{bmatrix}$$

$$\underline{\underline{A}} = \underline{\underline{A}} - \sum_{i=1}^{k-1} \underline{\underline{q}}_i \underline{\underline{c}}_i^T = \sum_{i=k}^N \underline{\underline{q}}_i \underline{\underline{c}}_i^T = \begin{bmatrix} 0 & \underline{\underline{A}}^{[k:N]} \end{bmatrix}$$

$\underline{\underline{A}}^{[k:N]}$ $M \times (N-k+1)$

Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(I.8)

P_o projection matrix

$$P = U(V^H V)^{-1} V^H$$

a) $P^H = P$ heritice

$$P^H = (U(V^H V)^{-1} V^H)^H = V^{H^H} (V^H V)^{-1} V^H = U(V^H V)^{-1} V^H = P$$

b) $PP = P$ idempotent

$$PP = U(V^H V)^{-1} V^H U(V^H V)^{-1} V^H$$

$$\mu_{ii} = \rho_{ii} : \mu_{ii} = \mu - \mu_{\perp} = (I - P)_{ii}$$

generen espais ortogonals

c) $P(I - P) = P(V^H V)^{-1} V^H + (I - P)P = 0 \Leftrightarrow P \perp (I - P)$

d) $(I - P)Q = 0$

(I.9)

$$n \in \mathbb{A}^2 = \mathbb{A} \Rightarrow \mathbb{A} \text{ is idempotent}$$

i) $\mathbb{A}_{n \times n} = \mathbb{A}$

ii) $\mathbb{B}_{n \times n}$ non-singular $\rightarrow \mathbb{B}^{-1} \mathbb{A} \mathbb{B} = \mathbb{A}$ idempotent

$$(\mathbb{B}^{-1} \mathbb{A} \mathbb{B})^2 = (\mathbb{B}^{-1} \mathbb{A} \mathbb{B})(\mathbb{B}^{-1} \mathbb{A} \mathbb{B}) = \mathbb{B}^{-1} \mathbb{A} \mathbb{B} \mathbb{B}^{-1} \mathbb{A} \mathbb{B} = \mathbb{B}^{-1} \mathbb{A}^2 \mathbb{B} = \mathbb{B}^{-1} \mathbb{A} \mathbb{B}$$

iii) $k > 2 \quad \mathbb{A}^k = \mathbb{A}$

$$\mathbb{A}^2 = \mathbb{A}$$

$$\mathbb{A}^3 = \mathbb{A}^2 \mathbb{A} = \mathbb{A} \mathbb{A} = \mathbb{A}^2 = \mathbb{A}$$

$$\mathbb{A}^k = \mathbb{A}^{k-1} \mathbb{A} = \mathbb{A}^{k-2} \underbrace{\mathbb{A} \mathbb{A}}_A = \mathbb{A}^{k-3} \underbrace{\mathbb{A} \mathbb{A}}_A = \dots = \mathbb{A}$$

$$\mathbb{A} = 0$$

(I.10)

$$\begin{vmatrix} T & V \\ V & W \end{vmatrix} = |T||W| - |V||U| \stackrel{U=0}{=} |T||W| = |T||W| \stackrel{|U=(-1)^3|W|}{=} |T(-1)^2|W| + (-1)^3|W|$$



II.1 $f(x) \text{ is } A \text{ square} \quad f(A) \text{ t.g. } x^i = A^i$

prove, $\begin{cases} f(\lambda) \text{ VAP} \\ \times \text{ VEP} \end{cases} \text{ de } f(A) \text{ is } \begin{cases} \text{VAP} \\ \times \text{ VEP} \end{cases} \text{ de } A$

- * $Ax = \lambda x \rightarrow (A - \lambda I)x = 0 \rightarrow \det(A - \lambda I) = 0$
- * $f(A)x = \lambda' x \rightarrow \det(f(A) - \lambda' I) = 0$
 $\stackrel{x?}{\lambda'}$ $\stackrel{x?}{f(A)}$

$$f(x) = a_n x^n + \dots + a_1 x + a_0 \quad f(A) = a_n A^n + \dots + a_1 A + a_0 I$$

$$f(A)x = a_n A^n x + \dots + a_1 A x + a_0 I x \quad (A - \lambda I)x = 0$$

$$f(A)x = a_n \lambda^n x + \dots + a_1 \lambda x + a_0 x$$

$$\left. \begin{array}{l} A\mathcal{Q} = \mathcal{Q}\Lambda \rightarrow A = \mathcal{Q}\Lambda\mathcal{Q}^{-1} \\ f(A) = f(\mathcal{Q}\Lambda\mathcal{Q}^{-1}) = \mathcal{Q} f(\Lambda) \mathcal{Q}^{-1} \end{array} \right\}$$

$$Ax = \lambda x \rightarrow f(A) = X f(\lambda) X^{-1}$$

II.2 $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \quad P_n(\lambda) = 0 \quad \text{prove it!}$

$$P_n(\lambda) = \det(\Lambda - \lambda I) = \begin{vmatrix} \lambda_1 - \lambda & & & & 0 \\ & \lambda_2 - \lambda & & & \\ & & \ddots & & \\ 0 & & & & \lambda_n - \lambda \end{vmatrix} = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda)$$

$\lambda_1 \dots \lambda_n \text{ VAPS}$

$P_n(\lambda_i) = 0$

$$\det(\Lambda - \lambda I) = 0 \Leftrightarrow \Lambda - \lambda I \text{ is singular} \Leftrightarrow (\Lambda - \lambda I)x = 0 \Leftrightarrow \Lambda x = \lambda x$$

$$A = P \Lambda P^{-1}$$

$$\begin{matrix} P & | \\ \text{VEPs} & \text{VAPS} \end{matrix} \Rightarrow P_n(\lambda)|_A = P_n(\lambda)_A = 0$$



Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(II.3)

 $\underline{A} \underset{n \times n}{=}$ $\lambda_1, \dots, \lambda_n$ VAPS de \underline{A} $\underline{B} \underset{p \times p}{=}$ μ_1, \dots, μ_p VAPS de \underline{B}
 $\left\{ \begin{array}{l} \text{prote VAPS de } \underline{A} \otimes \underline{B} \\ \lambda_i \mu_j \end{array} \right\}_{i=1 \dots n}^{j=1 \dots p}$

 $\underbrace{}_{np \times np}$

$$\underline{A} \otimes \underline{B} = \begin{bmatrix} a_{11} \underline{B} & \cdots & a_{1n} \underline{B} \\ \vdots & \ddots & \vdots \\ a_{n1} \underline{B} & \cdots & a_{nn} \underline{B} \end{bmatrix} = \begin{bmatrix} a_{11} b_{11} & \cdots & a_{11} b_{1p} & a_{1n} \underline{B} \\ \vdots & \ddots & \vdots & \vdots \\ a_{11} b_{p1} & \cdots & a_{11} b_{pp} & a_{1n} \underline{B} \\ a_{1n} \underline{B} & & a_{22} b_{11} & \vdots \\ & & \vdots & \vdots \\ & & a_{22} b_{pp} & a_{2n} \underline{B} \\ & & \vdots & \vdots \\ & & a_{nn} b_{11} & \vdots \\ & & \vdots & \vdots \\ & & a_{nn} b_{pp} & a_{nn} \underline{B} \end{bmatrix}$$

$$|\underline{A} \otimes \underline{B} - \lambda \underline{I}| = \begin{vmatrix} a_{11} b_{11} - \lambda & \cdots & a_{11} b_{1p} - \lambda & a_{1n} \underline{B} \\ \vdots & \ddots & \vdots & \vdots \\ a_{11} b_{p1} - \lambda & \cdots & a_{11} b_{pp} - \lambda & a_{1n} \underline{B} \\ a_{1n} \underline{B} & & a_{22} b_{11} - \lambda & \vdots \\ & & \vdots & \vdots \\ & & a_{22} b_{pp} - \lambda & a_{2n} \underline{B} \\ & & \vdots & \vdots \\ & & a_{nn} b_{11} - \lambda & \vdots \\ & & \vdots & \vdots \\ & & a_{nn} b_{pp} - \lambda & a_{nn} \underline{B} \end{vmatrix} = 0$$

$$\underline{A} \otimes \underline{B} = (\underline{Q} \Lambda \underline{Q}) \otimes (\underline{P} \Lambda' \underline{P})$$

$$n=p=2 \rightarrow \underline{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \underline{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \quad \underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \underline{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\lambda \underline{x} = \underline{A} \underline{x} \rightarrow \lambda_1 x_1 = a_{11} x_1 + a_{12} x_2$$

$$\lambda_1 x_2 = a_{21} x_1 + a_{22} x_2$$

$$\mu \underline{y} = \underline{B} \underline{y} \rightarrow \mu_1 y_1 = b_{11} y_1 + b_{12} y_2$$

$$\mu_2 y_2 = b_{21} y_1 + b_{22} y_2$$

productes:

$\lambda_1 \mu_1 x_1 y_1 = a_{11} b_{11} x_1 y_1 + a_{11} b_{12} x_1 y_2 + a_{12} b_{11} x_2 y_1 + a_{12} b_{12} x_2 y_2$	$\lambda_1 \mu_1 x_1 y_2 =$
$\lambda_1 \mu_1 x_2 y_1 =$	\vdots
$\lambda_1 \mu_1 x_2 y_2 =$	\vdots

$$\left[\lambda_1 \mu_1 \underline{z} = \underline{C} \underline{z} \right] \text{ on } \underline{z} = \begin{pmatrix} x_1 y_1 \\ x_1 y_2 \\ x_2 y_1 \\ x_2 y_2 \end{pmatrix} \text{ if } \underline{C} = \underline{A} \otimes \underline{B}$$

II.4

$A_{n \times n} \quad \lambda_1, \dots, \lambda_n$ VAPS t.g. $A = A^H$ hermitica

$\max x^H A x$ subject to $x^H x = 1 = \|x\|$

$$\mathcal{L} = x^H A x - \lambda x^H x \rightarrow \nabla_x \mathcal{L} = Ax - \lambda x = 0 \rightarrow \begin{matrix} Ax \\ x \end{matrix} = \lambda x$$

\uparrow VGP \uparrow VAP

$x^H A x$ $\left\{ \begin{array}{l} \max \\ \min \end{array} \right\}$ in λ VAP $\left\{ \begin{array}{l} \max \\ \min \end{array} \right\}$

$$Q(x) = x^H A x = x^H \lambda x = x^H x \lambda = \|x\| \lambda = \lambda$$

$$\lambda_{\min} \leq \frac{x^H A x}{x^H x} \leq \lambda_{\max}$$

II.5

$F^{1 \times N} = A$ mitaria $\rightarrow A^{-1} = A^H \Leftrightarrow A^H A^H = A^H A = I$

$$\frac{1}{N} F F^H = I$$

$$F = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-i\frac{2\pi}{N}} & \cdots & e^{-i\frac{2\pi}{N}(N-1)} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & e^{-i\frac{2\pi}{N}(N-1)} & \cdots & e^{-i\frac{2\pi}{N}(N-1)} \end{pmatrix}$$

$$f_{ij} = e^{-i\frac{2\pi}{N}(i-1)(j-1)}$$

$$f_{ij}^H = f_{ji}^* = e^{i\frac{2\pi}{N}(j-1)(i-1)}$$

~~$$f_{ij} f_{ij}^* = f_{ij} f_{ji}^* = e^{i\frac{2\pi}{N}(i-1)(j-1)} e^{i\frac{2\pi}{N}(j-1)(i-1)}$$~~

column j
 $j, i = 1, \dots, N$

$c_{ij} = \sum_{k=1}^N f_{ik} f_{kj}^H$	$= \sum_{k=1}^N f_{ik} f_{jk}^* = \sum_{k=1}^N e^{-i\frac{2\pi}{N}(i-1)(k-1)} e^{i\frac{2\pi}{N}(j-1)(k-1)} = \sum_{k=1}^N e^{i\frac{2\pi}{N}(j-1-(i-1))}$
---	--

$c_{ij} = \begin{cases} j=i & \rightarrow N \\ j \neq i & \rightarrow 0 \end{cases}$

$\Rightarrow \frac{1}{N} F F^H = \frac{1}{N} \begin{pmatrix} N & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & N \end{pmatrix} = I$

Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(II.6)

$$\underline{C} = \begin{bmatrix} c_1 & c_2 & \dots & c_m \\ c_m & c_1 & \dots & c_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_2 & c_3 & \dots & c_1 \end{bmatrix}$$

Matrīn Circulant

$$\underline{C} = \frac{1}{m} \underline{F} \underline{\Lambda} \underline{F}^H$$

Fourier transf. (ex 5)

$$P_c(z) = c_1 + c_2 z + \dots + c_m z^{m-1}$$

$$\text{NOTA : } \underline{F} \underline{F}^H = m \underline{I} \rightarrow \frac{1}{m} \underline{F}^H \underline{C} \underline{F} = \underline{\Lambda}$$

$$\underline{F}^H \underline{C} \underline{F} = \underbrace{\begin{bmatrix} f_{11} & \dots & f_{1N} \\ \vdots & \ddots & \vdots \\ f_{N1} & \dots & f_{NN} \end{bmatrix}}_{(1)} \underbrace{\begin{bmatrix} c_1 & \dots & c_N \\ c_N & \dots & c_1 \\ \vdots & \ddots & \vdots \\ c_2 & \dots & c_1 \end{bmatrix}}_{(2)} \underbrace{\begin{bmatrix} f_{11} & \dots & f_{1N} \\ \vdots & \ddots & \vdots \\ f_{N1} & \dots & f_{NN} \end{bmatrix}}_{(3)} = (\lambda_j)$$

j=1, ..., N

$$\text{(1)} \rightarrow \sum_{k=1}^N f_{ik}^H c_{kj} = \sum_{k=1}^N f_{ki}^* c_{kj} \quad // \quad \boxed{(\lambda_j)_{ij} \rightarrow \sum_{l=1}^N \left(\sum_{k=1}^N f_{ki}^* c_{kj} \right) f_{lj}}$$

f_{ki}
columns

$$\underline{F}^H \underline{C} \underline{F} = \sum_{l=1}^N \sum_{k=1}^N f_{ki}^* c_{kj} f_{lj} \quad i', j', j'' = 1, \dots, N$$

$$\underline{F}^H \underline{C} \underline{F} \Big|_{ii'} = \sum_{k=1}^N \sum_{l=1}^N f_{ki}^* c_{kj} f_{lj}$$

$$\begin{aligned} \underline{F}^H \underline{C} \underline{F} \Big|_{ii'} &= \sum_{k=1}^N f_{ik}^H c_{kj} \cdot f_{ji} + \sum_{k=1}^N f_{ik}^H c_{kj} \cdot f_{2j} + \dots + \sum_{k=1}^N f_{ik}^H c_{kj} f_{Nj} = \\ &= \sum_{l=1}^N \sum_{k=1}^N f_{ik}^H c_{kj} \cdot f_{lj} - \sum_{l=1}^N \sum_{k=1}^N c_{kj} f_{lk}^H f_{lj} = \begin{cases} k=l \rightarrow N^2 \\ k \neq l \rightarrow 0 \end{cases} \end{aligned}$$

II.7

$$x(n) = s(n) \cos(2\pi w_n n) + w(n) \quad , \quad s(n) = \cos(2\pi w_n)$$

$$\hat{R}_{xx} = U \Sigma V^H$$

SVD de R_{xx}

$$RR^H = U(\Sigma \Sigma^H)V^H \longrightarrow U$$

$$R^H R = V(\Sigma^H \Sigma)V^H \longrightarrow V$$

$$R = V \xrightarrow{\text{per observació}} U \Sigma$$

identifiquem

les parelles

VAP - VEP

amb el signe adjacent i formen $R = U \Sigma V^H$

parametres
//

II.9

$$P_A = I - A(A^H A)^{-1} A^H = I - \sum_{i=1} \frac{\alpha_i \alpha_i^H}{\alpha_i^H \alpha_i}$$

$$A_{mn} = [a_1 \dots a_n] \text{ span signal subspace } R^m$$

$$\Rightarrow (I + \alpha P_A)^{-1} = (I + \alpha(I - A(A^H A)^{-1} A^H))^{-1} = (I + \alpha I - \alpha A(A^H A)^{-1} A^H)^{-1}$$

$$A = U \Sigma V^H \quad A(A^H A)^{-1} A^H = U \Sigma V^H (\Sigma \Sigma^H)^{-1} V^H U^H \xrightarrow{(AB)^{-1} = B^{-1} A^{-1}}$$

$$= U \Sigma V^H (\Sigma \Sigma^H)^{-1} (\Sigma \Sigma^H)^{-1} V^H U^H = U \Sigma (\Sigma^H)^{-1} V^H U^H$$

$$\Sigma' = \Sigma \Sigma^H$$

$$= U \Sigma V^H (\Sigma \Sigma^H)^{-1} V^H \Sigma^H U^H = U \Sigma (\Sigma^H)^{-1} \Sigma^{-1} \Sigma^H V^H$$

$$\bar{\Sigma}^{-1} = \Sigma^H \Sigma^{-1}$$



--

Cognoms

Nom

Centre

Assignatura / especialitat

DNI

Núm. matrícula

Curs

Grup

Data

(III.1)

predicting population with $y(t) = a + bt + ct^2$

$$\text{III.2.1} \quad B = P A \quad : \quad A \text{ permuted in rows}$$

$$[P, B] = \text{permute}(A, v)$$

↑ ↑
 perm. matrix $N \times N$
 perm. vector $(1 \times N)$

$$v = [v_1, \dots, v_N]$$

$$P = \text{zeros}(N) \longrightarrow \left. \begin{array}{l} P(1, v_1) = 1 \\ P(2, v_2) = 2 \\ \dots \\ P(N, v_N) = 1 \end{array} \right\} \quad \begin{array}{l} \text{for } k=1:N \\ P(k, v(k)) = 1; \\ \text{end} \end{array}$$

aixi
funktion

$$[P, B] = \text{permute}(A, v)$$

$$P = \text{zeros}(\text{length}(A));$$

$$\text{for } K=1: \text{length}(A)$$

$$P(K, v(k)) = 1;$$

$$\text{end}$$

$$B = P \cdot A;$$

$$\text{III.2.2}$$

the first time in history that the world's population has reached 6 billion. This is a momentous occasion, and it is a reminder of the responsibility we have to care for the planet and its people. As we look towards the future, let us remember that our actions today will have a lasting impact on the world we leave behind.