ALGEBRA
5. ORTHOGONALITY

Grau de Matemàtiques. FME
2018-2019

Marta Casanellas
Departament de Matemàtiques

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

OUTLINE

Outline

The standard scalar product

Definition

The dot product (or scalar product) $< u, v >$ of two vectors

$$u = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, v = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n \text{ is}$$

$$< u, v > := u^t v = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

Properties:

1. $< u, u > \geq 0 \ \forall u$ and $< u, u > = 0 \Leftrightarrow u = 0$ (*positive definite*)
2. $< u, v > = < v, u >$ (*symmetric*).
3. *bilineal*:
   - $< a_1 u_1 + a_2 u_2, v > = a_1 < u_1, v > + a_2 < u_2, v >$;
   - $< u, a_1 v_1 + a_2 v_2 > = a_1 < u, v_1 > + a_2 < u, v_2 >$.

Any function $\mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ that satisfies these properties is called a *scalar product*:

Let E be an $\mathbb{R}$-e.v. A bilinear form on E is a map $\varphi E \times E \longrightarrow \mathbb{R}$ such that, $\forall\ u, v, w \in E$ and $\lambda \in \mathbb{R}$:

(A) $\varphi(u + v, w) = \varphi(u, w) + \varphi(v, w)$
$\varphi(\lambda u, w) = \lambda\varphi(u, w)$,

(B) $\varphi(w, u + v) = \varphi(w, u) + \varphi(w, v)$
$\varphi(w, \lambda u) = \lambda\varphi(w, u)$.

If $\mathbf{u} = \{u_1, \dots, u_n\}$ is a basis of E, then the matrix of φ in the basis $\mathbf{u}$ is defined as

$$M_{\mathbf{u}}(\varphi) = \begin{pmatrix} \varphi(u_1, u_1) & \cdots & \varphi(u_1, u_n) \\ \vdots & & \vdots \\ \varphi(u_n, u_1) & \cdots & \varphi(u_n, u_n) \end{pmatrix}.$$

Properties:

1. If $v_{\mathbf{u}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, w_{\mathbf{u}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \Rightarrow \varphi(v, w) = (x_1 \dots x_n) M_{\mathbf{u}}(\varphi) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$
2. If $\mathbf{v}$ is another basis, then

$$M_{\mathbf{v}}(\varphi) = A^t_{\mathbf{v}\to\mathbf{u}} M_{\mathbf{u}}(\varphi) A_{\mathbf{v}\to\mathbf{u}}$$

A bilinear form φ is symmetric if $\varphi(u, v) = \varphi(v, u)$ for all u, v.
A bilinear form is symmetric $\Leftrightarrow$ $M_{\mathbf{u}}(\varphi)$ is a symmetric matrix for any basis $\mathbf{u}$.

Let E be an $\mathbb{R}$-e.v. and φ a bilinear form on E. One says that φ is positive definite if $\varphi(u, u) \geq 0$ with equality only when $u = 0$.

Definition

A scalar product on E is a symmetric, positive definite bilinear form $<,>: E \times E \longrightarrow \mathbb{R}$. An $\mathbb{R}$-e.v together with a scalar product is called a euclidean vector space.

Examples:

- The usual standard product
- $E = \mathcal{F}([a, b], \mathbb{R}) = \{$ continuous real functions defined on $[a, b]\}$, then the following defines a scalar product on E:

$$< f, g >:= \int_a^b f(x)g(x)dx.$$

Let E be an $\mathbb{R}$-e.v. with scalar product $<,>$. The norm of $u \in \mathbb{E}$ is $\|u\| = \sqrt{< u, u >}$.

If $<,>$ is the standard product, the norm is called the *usual, euclidean or 2-norm* and is also denoted as $||u||_2$.

Properties: for any $u, v \in E$ and $c \in \mathbb{R}$

1. $\|u\| \geq 0 \ \forall u$ and $\|u\| = 0 \Leftrightarrow u = 0$;
2. $\|cu\| = |c|\|u\| \ c \in \mathbb{R}$;
3. $| < u, v > | \leq \|u\|\|v\|$ (Cauchy-Schwarz inequality)
4. $\|u + v\| \leq \|u\| + \|v\|$ (triangular inequality);

Any function $f : E \longrightarrow \mathbb{R}$ that satisfies properties 1,2,4 is called a *norm* (and is not necessary defined through a scalar product).

Other norms

If $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, one defines:

1. The 1-norm (also called taxicab or Manhattan norm):

$$||x||_1 = |x_1| + \cdots + |\mathbf{x}_n|.$$

2. The maximum norm (also called infinite norm):

$$||x||_\infty = \max(|x_1|, \dots, |\mathbf{x}_n|).$$

Distances and Angles

Let E be an $\mathbb{R}$-e.v. with scalar product $<,>$.

- A vector u is called a unit vector if $\|u\| = 1$. Given a vector $v \neq 0$, we can always find a unit vector in its direction: $v/\|v\|$ (we say that we have normalized v).
- The distance between two vectors $u, v \in E$, is $d(u, v) = \|u - v\|$.
- The (unoriented) angle between two nonzero vectors $u, v \in E$ is the unique $\alpha \in [0, \pi]$ such that $\cos(\widehat{uv}) = \frac{<u,v>}{||u||\cdot||v||}$ (the sign of $\widehat{uv}$ depends on the orientation we choose).
- Two vectors u, v are orthogonal (also denoted $u \perp v$) if $< u, v >= 0$.
- Two orthogonal vectors have $\widehat{uv} = \pm\frac{\pi}{2}$.
- If $u \perp v$ and $u, v \neq 0 \Rightarrow u, v$ are l.i.

Definition

An orthonormal basis (b.o.n) of a subspace $F \subset E$ is a basis $\{v_1, \ldots, v_d\}$ of F such that its vectors are

- pairwise orthogonal: $< v_i, v_j >= 0$ if $i \neq j$
- and normalized: $\|v_i\| = 1$ for $i = 1, 2, \ldots, d$.

- If they are pairwise orthogonal but not normalized, then the basis is called orthogonal.
- Example: the standard basis is an orthonormal basis of $\mathbb{R}^n$ with the standard product.
- If $\mathbf{v} = \{v_1, \ldots, v_n\}$ is a b.o.n. of E, then the coordinates of $v \in E$ in basis $\mathbf{v}$ are

$$< v, v_1 >, \ldots, < v, v_n > .$$

- If $\mathbf{v}$ is b.o.n. and $\mathbf{u}$ is another basis of E then,

$$\mathbf{u} \text{ is b.o.n} \quad \text{if and ony if } A^t_{\mathbf{v}\to\mathbf{u}} A_{\mathbf{v}\to\mathbf{u}} = Id.$$

An $n \times n$ matrix that satisfies $A^t A = Id$ is called an orthogonal matrix.

- If we call the columns $v_1, \dots, v_n$, $A = (v_1 \dots v_n)$, then,

$$A^t A = Id \quad \text{if and ony if } \{v_1, \dots, v_n\} \text{ is an orthonormal basis.}$$

- A is orthogonal if and only if $A^{-1} = A^t$.
- A is orthogonal $\Rightarrow \det A = \pm 1$.
- If A is orthogonal, then the corresponding endomorphism **preserves the usual scalar product**:

$$< Au, Av >=< u, v > \text{ for all } u, v$$

- In particular, A preserves norms and angles (and hence preserves orthogonality) and so it is a transformation that does not deform objects.

The following maps $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ are linear and preserve norms:

- f= symmetry with respect to a line l passing through the origin, $l = [v]$. E.g. $f(x, y) = (x, -y)$.
- f=rotation counterclockwise of angle α with respect to the origin; then
$$M_e(f) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$$

Examples of 3×3 orthogonal matrices

Similarly,

- The symmetry in $\mathbb{R}^3$ with respect to a plane through the origin preserves norms and hence has an orthogonal matrix.
- The rotation in $\mathbb{R}^3$ of angle α with respect to an axis through the origin preserves norms and hence has an orthogonal matrix.
- The composition of these types of maps also preserves norms and hence has orthogonal matrix (if A and B are orthogonal matrices, so is AB.)

Given a subspace F of a euclidean space E, the following algorithm produces a b.o.n. of F:

1. Take any basis of F: $u_1, \ldots, u_d$ and define:
2. $v_1 := u_1$
3. $v_2 := u_2 - \frac{<v_1,u_2>}{<v_1,v_1>} v_1 \quad (\Rightarrow [v_1, v_2] = [u_1, u_2])$.
4. $v_3 := u_3 - \frac{<v_1,u_3>}{<v_1,v_1>} v_1 - \frac{<v_2,u_3>}{<v_2,v_2>} v_2 \quad (\Rightarrow [v_1, v_2, v_3] = [u_1, u_2, u_3])$.

$$\vdots$$

5. $v_d := u_d - \frac{<v_1,u_d>}{<v_1,v_1>} v_1 - \cdots - \frac{<v_{d-1},u_d>}{<v_{d-1},v_{d-1}>} v_{d-1}$.
6. Then $v_1, \ldots, v_d$ are orthogonal vectors and $[v_1, \ldots, v_d] = [u_1, \ldots, u_d]$.
7. We normalize to obtain the desired b.o.n. $w_1, \ldots, w_d$:

$$w_1 = v_1/\|v_1\|,\ w_2 = v_2/\|v_2\|,\ \ldots,\ w_d = v_d/\|v_d\|.$$

Warning: Roundoff errors make vectors not orthogonal! $\rightarrow$ **modified Gram-Schmidt** algorithm solves this issue ($\rightsquigarrow$ Algebra lineal numèrica)

Theorem (Spectral Theorem)

*Let A be a **symmetric** $n \times n$ matrix. Then A has **real eigenvalues**, **diagonalizes**, and there exists an **orthonormal basis** of eigenvectors.*

The orthonormal basis of eigenvectors is not difficult to find:

- If u, v are eigenvectors of A of eigenvalues $\lambda \neq \mu$, then $u \perp v$.
- If the eigenvalues are all distinct, then normalizing the eigenvectors we obtain an orthonormal basis of eigenvectors.
- If the eigenvalues are not all distinct, we have to use Gram-Schmidt algorithm on each subspace of eigenvectors.

OUTLINE

1. SCALAR PRODUCT
2. CROSS-PRODUCT
3. ORTHOGONAL COMPLEMENT AND PROJECTION
4. SINGULAR VALUE DECOMPOSITION
5. APPLICATIONS
   - Rank approximation
   - Linear least squares
   - Pseudoinverse
   - Principal component analysis
6. BIBLIOGRAPHY

The cross-product between two vectors $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$ of $\mathbb{R}^3$ is the following vector (in standard basis)

$$\begin{aligned}(u_1, u_2, u_3) \times (v_1, v_2, v_3) &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \\ &= (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).\end{aligned}$$

Main properties:

- $v \times u = -u \times v$ (anti-commutative)
- $u \times v$ is orthogonal to both u and v
- If u, v are orthogonal and normalized, then $u, v, u \times v$ is an orthonormal basis.

Outline

Orthogonal complement

The orthogonal complement to a given subspace F of a euclidean space E is the subspace

$$F^{\perp} = \{u \in E \mid u \perp v \text{ for all } v \in F\}.$$

Properties:

- If $F = [v_1, \ldots, v_d]$, then
$$F^{\perp} = \left\{ u \in \mathbb{E} \;\middle|\; \begin{array}{c} < u, v_1 >= 0 \\ \vdots \\ < u, v_d >= 0 \end{array} \right\}$$
- $F \cap F^{\perp} = 0$.

In $\mathbb{R}^n$ with the standard scalar product,

- If F is defined by generators $\Rightarrow$ the equations of $F^\perp$ are easy to get: their coefficients are the generators coordinates.
- If F is given by equations $\Rightarrow$ the generators of $F^\perp$ are easy to get: their coordinates are the coefficients of the equations.

F	$F^\perp$
$[(1,3,2),(-2,1,8)]$	$\begin{cases} x+3y+2z=0 \\ -2x+y+8z=0 \end{cases}$
$3x-5y+\frac{11}{2}z=0$	$[(3,-5,\frac{11}{2})]$

Let E be a euclidean space of dimension n.

Theorem (Orthogonal Decomposition)

$E = F \oplus F^{\perp}$ for any subspace F. This is, any $v \in \mathbb{E}$ can be ***written in a unique way*** *as $v = w + w'$ where $w \in F$ and $w' \in F^{\perp}$.*

- w is called the *orthogonal projection* of v on F and is denoted as $proj_F(v)$,
- w' is called the *orthogonal projection* of v on $F^{\perp}$ and is denoted as $proj_{F^{\perp}}(v)$.
- Thus, $v = proj_F(v) + proj_{F^{\perp}}(v)$ and $proj_F(v)$ is the unique vector of F such that $v - proj_F(v)$ belongs to $F^{\perp}$.
- If $F \subseteq E$ has dimension $d \Rightarrow F^{\perp}$ has dimension $n - d$.

Proposition

The orthogonal projection of v on F is the vector of F that is closest to v; this is,

$$\|v - proj_F(v)\| = \min_{w \in F}\{\|v - w\|\}$$

(and this equals $\|proj_{F^\perp}(v)\|$). The orthogonal projection $proj_F(v)$ is the best approximation to v in F.

Computation of the orthogonal projection

Proposition

$proj_F(v)$ is the unique vector w that satisfies $w \in F$ and $v - w \in F^\perp$. If F has basis $u_1, \dots, u_d$, then $proj_F(v)$ is the unique vector w such that

$$w = c_1 u_1 + \dots c_d u_d \in F \quad \textit{and} \quad \begin{cases} < u_1, w > = < u_1, v > \\ \vdots \\ < u_d, w > = < u_d, v > \end{cases}$$

Thus, $proj_F(v)$ is the vector $c_1 u_1 + \cdots + c_d u_d$ such that $c_1, \dots, c_d$ are solution to the system

$$\begin{pmatrix} < u_1, u_1 > & \dots & < u_1, u_d > \\ \vdots & \vdots & \vdots \\ < u_d, u_1 > & \dots & < u_d, u_d > \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_d \end{pmatrix} = \begin{pmatrix} < v, u_1 > \\ \vdots \\ < v, u_d > \end{pmatrix}$$

Orthogonal projection with orthogonal basis

Corollary

If $\dim F = 1$, $F = [u]$, *then* $proj_F(v) = \frac{<v,u>}{<u,u>}\, u$.

Proposition

If $u_1, \ldots, u_d$ *is an orthogonal basis of* F *and* $v \in \mathbb{R}^n$, *then*

$$proj_F(v) = \frac{< v, u_1 >}{< u_1, u_1 >} u_1 + \cdots + \frac{< v, u_d >}{< u_d, u_d >} u_d.$$

Proposition

If $u_1, \ldots, u_d$ *is an b.o.n. of* F *and* $v \in \mathbb{R}^n$, *then*

$$proj_F(v) = < v, u_1 > u_1 + \cdots + < v, u_d > u_d.$$

That is, the coordinates of $proj_F(v)$ *in the basis* $u_1, \ldots, u_d$ *are* $< v, u_1 >, \ldots, < v, u_d >$.

Outline

THEOREM (SINGULAR VALUE DECOMPOSITION)

Let A be a real $m \times n$ matrix. There there exists a decomposition $A = U \cdot D \cdot V^t$, where U is $m \times m$, V is $n \times n$, U, V are ***orthogonal*** *and D is the following $m \times n$ matrix*

$$D = \begin{pmatrix} \sigma_1 & & 0 & \cdots & 0 \\ & \ddots & & & \vdots \\ 0 & & \sigma_r & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix}$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ and $r = \operatorname{rank} A$.
$\sigma_1, \ldots, \sigma_r$ are called singular values of A.

How to get the SVD?

The singular values are determined by the matrix A:

$$A = UDV^t \Rightarrow A^tA = VD^tU^tUDV^t = VD^tDV^t$$

but U and V are not (almost determined in most cases). How do we compute the SVD?

(1) Diagonalize the symmetric matrix $S = A^t \cdot A$

(2) If $\lambda_1 \geq \cdots \geq \lambda_r$ are the non-zero eigenvalues of $S \Rightarrow$ the singular values are $\sigma_1 = \sqrt{\lambda_1}, \ldots \sigma_r = \sqrt{\lambda_r}$ (fact: A^tA always has non-negative eigenvalues).

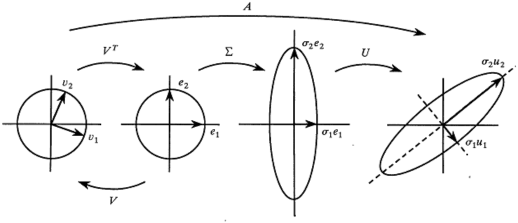
(3) The columns of V are an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of S.

(4) $u_1 = \frac{1}{\sigma_1}Av_1, \ldots, u_r = \frac{1}{\sigma_r}Av_r$ are orthonormal vectors in $\mathbb{R}^m$ (which can be completed to an orthonormal basis of $\mathbb{R}^m$ if necessary) and they form the columns of U.

If A is the standard matrix of a linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, and we call $u_1, \ldots, u_m$, $v_1, \ldots, v_n$, the columns of U and V respectively, then D the matrix associated to f in orthonormal basis $v_1, \ldots, v_n$ and $u_1, \ldots, u_m$:

$$A = M_e(f) = \underbrace{U}_{A_{u\to e}} * \underbrace{D}_{M_{v,u}(f)} * \underbrace{V^t}_{A_{e\to v}}$$

(note that $V^t = V^{-1} = A_{e\to v}$).



[T. Hern, C. Long]

Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a linear map and let A be its standard matrix. Then $\mathbb{R}^n = \text{Null}(A) \oplus \text{Im}(A^t)$ ($\text{Im}(A^t)$=row space of A), $\mathbb{R}^m = \text{Im}(A) \oplus \text{Null}(A^t)$, these decompositions give orthogonal complements and there exist b.o.n.'s $v_1, \dots, v_n$ (of $\mathbb{R}^n$) and $u_1, \dots, u_m$ (of $\mathbb{R}^m$) such that

1. $\text{Im}(A) = [u_1, \dots, u_r]$
2. $\text{Null}(A) = [v_{r+1}, \dots, v_n]$
3. $\text{Im}(A^t) = [v_1, \dots, v_r]$
4. $\text{Null}(A^t) = [u_{r+1}, \dots, u_m]$

Moreover, the restriction of the map f to the row space $\text{Im}(A^t)$ and onto $\text{Im}(A)$ in the bases $v_1, \dots, v_r$, $u_1, \dots, u_r$ (left and right, respectively) is the diagonal matrix

$$D = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{pmatrix}$$

Outline

2-NORM OF A MATRIX

To "measure" a linear map we measure how big the image of the unit sphere is under this map:

DEFINITION

The 2-norm of an $m \times n$ matrix A is

$$\|A\|_2 = \max_{||x||=1} ||Ax||.$$

- $\|A\|_2 = \max_{||x||\neq 0} \frac{||Ax||}{||x||}$
- $\|Av\| \leq \|A\|_2 \|v\| \ \forall v.$
- $\|AX\|_2 = \|A\|_2$ if X is an orthogonal matrix.
- $\|YA\|_2 = \|A\|_2$ if Y is an orthogonal matrix.
- $\|AB\|_2 \leq \|A\|_2 \|B\|_2.$

Geometric consequence of the *SVD*:

PROPOSITION

- $\|A\|_2 = \sigma_1$
- *The maximum is attained at* $\pm v_1$*:* $\max_{||x||=1} \|Ax\| = \|Av_1\|$.
- $\min_{||x||=1} \|Ax\| = \begin{cases} \sigma_n & \textit{if A has rank n, and is attained at } \pm v_n \\ 0 & \textit{if A has rank } < n \end{cases}$
- *If A is invertible,* $\|A^{-1}\|_2 = \frac{1}{\sigma_r}$.

SVD AND RANK APPROXIMATION

THEOREM (ECKHART-YOUNG)

Let A be any matrix. If $A = UDV^t$ and the singular values of A are $\sigma_1, \ldots, \sigma_r$ then for any $k \leq r$,

$$M = U \begin{pmatrix} \sigma_1 & & 0 & \cdots & 0 \\ & \ddots & & & \vdots \\ 0 & & \sigma_k & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} V^t$$

is the matrix of rank k closest to A (in the sense that $||A - M||_2$ is minimal among matrices M of rank k). Note that $||A - M||_2 = \sigma_{k+1}$.

This is used in image compression, for example. Note that $A = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.

OUTLINE

Problem: $Ax = b$ might be incompatible due to measure errors in b, but we would still like to have an approximated solution:

$$\begin{array}{c}\text{Incompatible}\\ Ax = b\\ \text{system}\end{array} \quad \Leftrightarrow \quad b \notin \mathrm{Im}(A)$$

Want: $\tilde{x}$ such that $A\tilde{x}$ is as close to b as possible.

DEFINITION

A least squares solution of $Ax = b$ is a vector $\tilde{x}$ that minimizes $\|Ax - b\|$, that is

$$\|A\tilde{x} - b\| \leq \|Ax - b\| \text{ for all } x$$

Solution given by Gauss (1801)

- Change b by the vector of $\mathrm{Im}(A)$ that is closest to b: the *orthogonal projection* of b in $\mathrm{Im}(A)$, $proj_{\mathrm{Im}(A)}(b)$.
- Find a solution $\tilde{x}$ to the compatible system $Ax = proj_{\mathrm{Im}(A)}(b)$
- Then $\tilde{x}$ is a least square solution of $Ax = b$.
- $\tilde{x}$ does not satisfy $Ax - b = \vec{0}$, but minimizes the norm $\|Ax - b\|$
- The residual measures how far $\tilde{x}$ is from a solution to the system:

  residual = $A\tilde{x} - b$ (which is $= proj_{\mathrm{Im}(A)}(b) - b$).

  norm of the residual: $\|A\tilde{x} - b\|$
- Important point: we do not need to compute $proj_{\mathrm{Im}(A)}(b)$ (see next slide).

Theorem

- $\tilde{x}$ *is a least squares solution of* $Ax = b$ *if and only if it is a solution of the* *normal equations:*

$$A^t Ax = A^t b.$$

- *If the rank of* A *equals the number of columns, then* $A^t A$ *is invertible and the least squares solution is unique and given by*

$$\tilde{x} = (A^t A)^{-1} A^t b$$

  (although computing the inverse is not efficient)
- *If the original system is compatible,* $\tilde{x}$ *is a solution to the original system as well.*

Outline

Definition

If A is an $m \times n$ matrix that has singular value decomposition $A = UDV^t$, then the pseudo-inverse of A is the $n \times m$ matrix

$$A^+ = VD^+U^t$$

where D^+ is the $n \times m$ matrix with diagonal entries $\frac{1}{\sigma_i}$ and 0's. Let A be an $m \times n$ matrix with linearly independent columns. The pseudo-inverse of A is:

$$A^+ = (A^tA)^{-1}A.$$

Properties:

- $AA^+A = A$
- $A^+AA^+ = A^+$
- AA^+ and A^+A are symmetric.

If A is invertible, then $A^+ = A^{-1}$.

- If A has linearly independent columns ($\Rightarrow r = n \leq m$), then $A^+ = (A^tA)^{-1}A^t$:

$$A^tA = Vdiag(\sigma_i^2)V^t \Rightarrow (A^tA)^{-1}A^t = Vdiag(\sigma_i^{-2})V^tVD^tU^t = A^+.$$

- The least squares solution to an incompatible system $Ax = b$ (if A has linearly independent columns) is

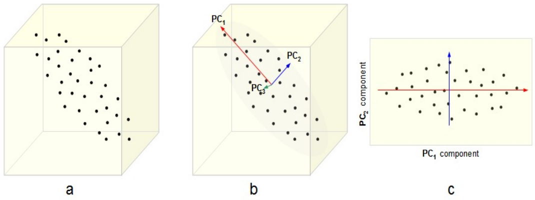
$$\tilde{x} = A^+b.$$

- The standard matrix of the orthogonal projection p onto the columns of A is

$$M_e(p) = AA^+.$$

Outline

Goal: Given N data points in $\mathbb{R}^3$, $p_i = (x_i, y_i, z_i)$, $i = 1, \ldots, N$ highly correlated, one wants to find a linear transformation $ax + by + cz$ such that the set $\{t_i = ax_i + by_i + cz_i\}_i$ has maximum variance:



- $v_1 = (a, b, c)$ is called the first principal component.
- Then one can look for $v_2 \in [v_1]^\perp$ (2nd principal component) maximizing variance of $proj_{[v_1]^\perp}(p_i)$.
- Keep going or project down to the first components in order to reduce the dimension of the problem.

Procedure

Assume that set $\{p_i\}$ is centered at the origin. Let

$$M = \begin{pmatrix} x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & z_N \end{pmatrix} \text{ so that } \sum_i x_i = \sum_i y_i = \sum_i z_i = 0.$$

-
- Want $v_1 = (a, b, c)$ of norm 1 such that $\sum_i t_i^2 = \sum_i (ax_i + by_i + cz_i)^2 = ||Mv_1||$ is maximum.
- v_1 is the first column vector of V in the SVD: $M = UDV^t$.
- Then the matrix $M_2 = M - Mv_1 v_1^t$ has $proj_{[v_1]^\perp}(p_i)$ in its rows.
- $M_2 = \sigma_2 u_2 v_2^t + \ldots + \sigma_r u_r v_r^t$.
- The direction which maximizes the variance is v_2 (2nd vector in V).
- Keep going.

Remarks:

- If the set $\{p_i\}$ is not centered at the origin we center it: let $(\bar{x}, \bar{y}, \bar{z}) = \sum_i (x_i, y_i, z_i)/N$, and consider

$$M = \begin{pmatrix} x_1 - \bar{x} & y_1 - \bar{y} & z_1 - \bar{z} \\ \vdots & \vdots & \vdots \\ x_N - \bar{x} & y_N - \bar{y} & z_N - \bar{z} \end{pmatrix}.$$

  Proceed as before with this M and then sum $(\bar{x}, \bar{y}, \bar{z})$ to the final result.
- The same can be done for clouds of points in $\mathbb{R}^n$.

Outline

BIBLIOGRAPHY

Basic:

- D. Poole, Linear Algebra, A modern introduction, 3rd ed., Brooks/Cole, 2011.

Additional

- Friedberg, S., Insel, A., Spence, L. Linear Algebra (2nd edition). Prentice Hall, 1989.