Mathématiques pour l'intelligence artificielle (MIA)

M2 IAAA, AMU, 2021-2022

Thomas Schatz

Plan du cours

1. Introduction générale

2. Preuves (revue)

3. Algèbre linéaire (revue)

4. Optimisation (revue)

5. Optimisation sous contraintes

6. Probabilités (revue)

7. Statistique

8. Théorie de l’apprentissage

9. Optimisation pour l’apprentissage

Partie 3 — Algèbre linéaire (revue)

Orthogonalité

Intérêt ?

Soit E un espace vectoriel de dimension finie n muni d'un produit scalaire $(\, . \mid . \,)$ et $V = (v_1, \ldots, v_n)$ une base orthonormale de E. Alors, pour tout $v \in E$,

$$v = \sum_{i=1}^{n} (v \mid v_i) v_i.$$

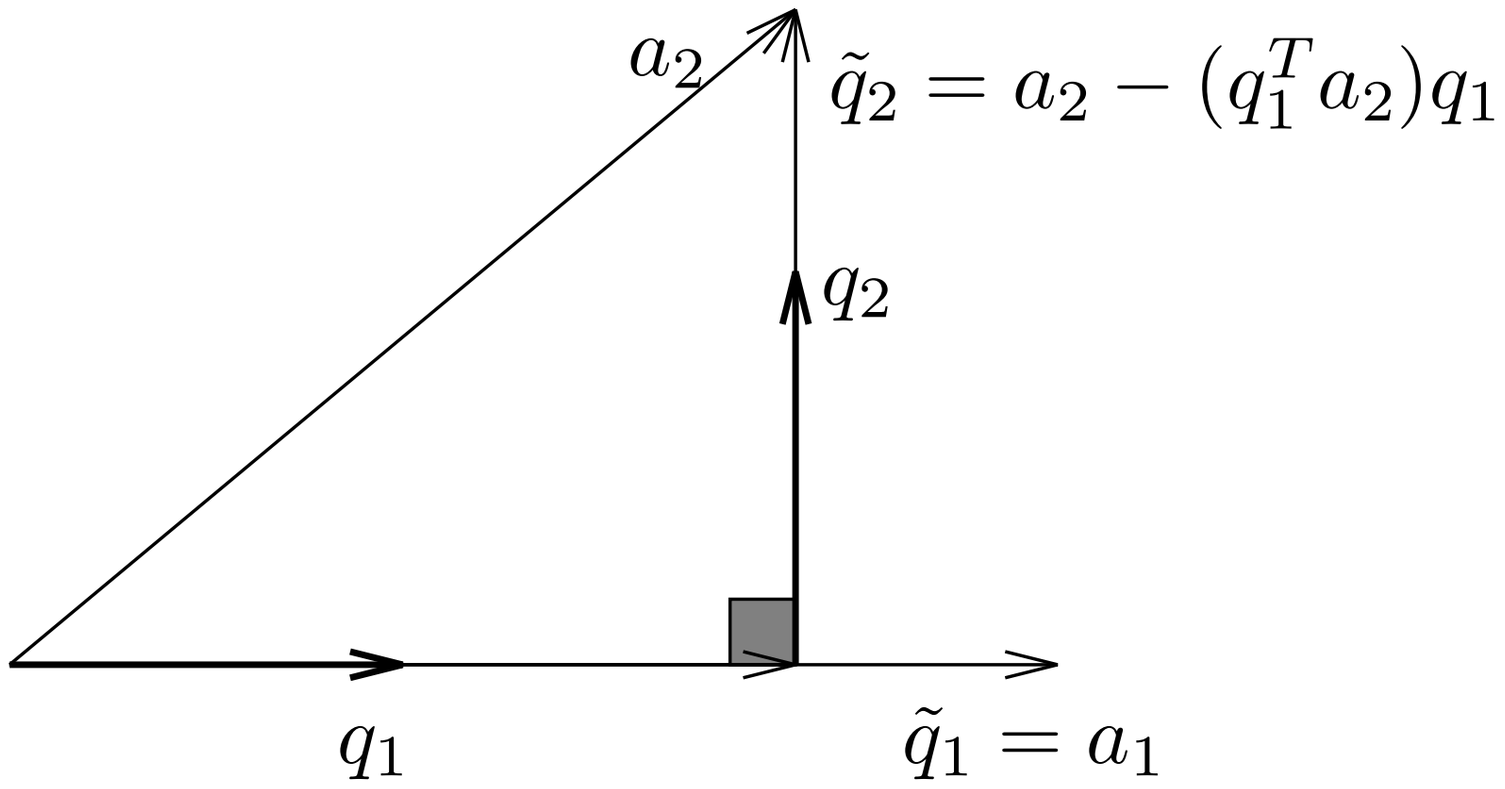
Orthogonalité

Comment construire des familles orthonormales?

Gram-Schmidt procedure : étant donné k vecteurs indépendants $(a_1, a_2, \ldots, a_k) \in E^k$, trouver des vecteurs orthonormaux $(q_1, q_2, \ldots, a_k)$, tels que pour tout $r \leq k$, $(q_1, q_2, \ldots, q_r)$ soit une base orthonormale de $\text{Vect}(a_1, a_2, \ldots, a_r)$

- step 1a. $\tilde{q}_1 := a_1$
- step 1b. $q_1 := \tilde{q}_1 / \|\tilde{q}_1\|$ (normalize)
- step 2a. $\tilde{q}_2 := a_2 - (q_1^T a_2) q_1$ (remove q_1 component from a_2)
- step 2b. $q_2 := \tilde{q}_2 / \|\tilde{q}_2\|$ (normalize)
- step 3a. $\tilde{q}_3 := a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2$ (remove q_1, q_2 components)
- step 3b. $q_3 := \tilde{q}_3 / \|\tilde{q}_3\|$ (normalize)
- etc.

Orthogonalité



Gram-Schmidt et factorisation QR

$$\underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{kk} \end{bmatrix}}_{R}$$

k : rang de la matrice A

$r_{ij} = q_i^T a_j$ Calculé durant Gram-Schmidt

Une matrice est dite semi-orthogonale si ces colonnes sont orthonormales.

Si Q est une matrice semi-orthogonale de taille $m \times k$, alors $Q^T Q = I_k$

Une matrice semi-orthogonale carrée est dite orthogonale.

Transposée

Transposée de A : obtenue en échangeant les lignes et les colonnes de A

$$(A+B)^\mathsf{T} = A^\mathsf{T} + B^\mathsf{T}, \qquad (\alpha A)^\mathsf{T} = \alpha A^\mathsf{T}$$

$$(AB)^\mathsf{T} = B^\mathsf{T}\, A^\mathsf{T}$$

Matrice par blocs

La matrice

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 2 & 2 & 2 \\ 1 & 1 & 2 & 2 & 2 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 4 & 4 & 4 \end{bmatrix}$$

peut être partitionnée en quatre blocs

$$\mathbf{P}_{11} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \mathbf{P}_{12} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \mathbf{P}_{21} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \\ 3 & 3 \end{bmatrix}, \mathbf{P}_{22} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}.$$

On peut alors écrire la matrice par bloc comme :

$$\mathbf{P}_{\text{partitionnee}} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{bmatrix}.$$

Interpretation des opérations matricielles

write $A \in \mathbf{R}^{m \times n}$ in terms of its columns:

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

then $y = Ax$ can be written as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

write A in terms of its rows:

$$A = \begin{bmatrix} \tilde{a}_1^T \\ \tilde{a}_2^T \\ \vdots \\ \tilde{a}_n^T \end{bmatrix}$$

then $y = Ax$ can be written as

$$y = \begin{bmatrix} \tilde{a}_1^T x \\ \tilde{a}_2^T x \\ \vdots \\ \tilde{a}_m^T x \end{bmatrix}$$

Interpretation des opérations matricielles

$$c_{ij} = \tilde{a}_i^T b_j = \langle \tilde{a}_i, b_j \rangle$$

$$C = \left[\begin{array}{c} c_1 \cdots c_p \end{array} \right] = AB = \left[\begin{array}{c} Ab_1 \cdots Ab_p \end{array} \right]$$

$$C = \left[\begin{array}{c} \tilde{c}_1^T \\ \vdots \\ \tilde{c}_m^T \end{array} \right] = AB = \left[\begin{array}{c} \tilde{a}_1^T B \\ \vdots \\ \tilde{a}_m^T B \end{array} \right]$$

$$C = \sum_i a_i \tilde{b}_i^T$$

Décomposition en valeurs singulières

Etant donné une matrice rectangulaire quelconque

(à coefficients réels)

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

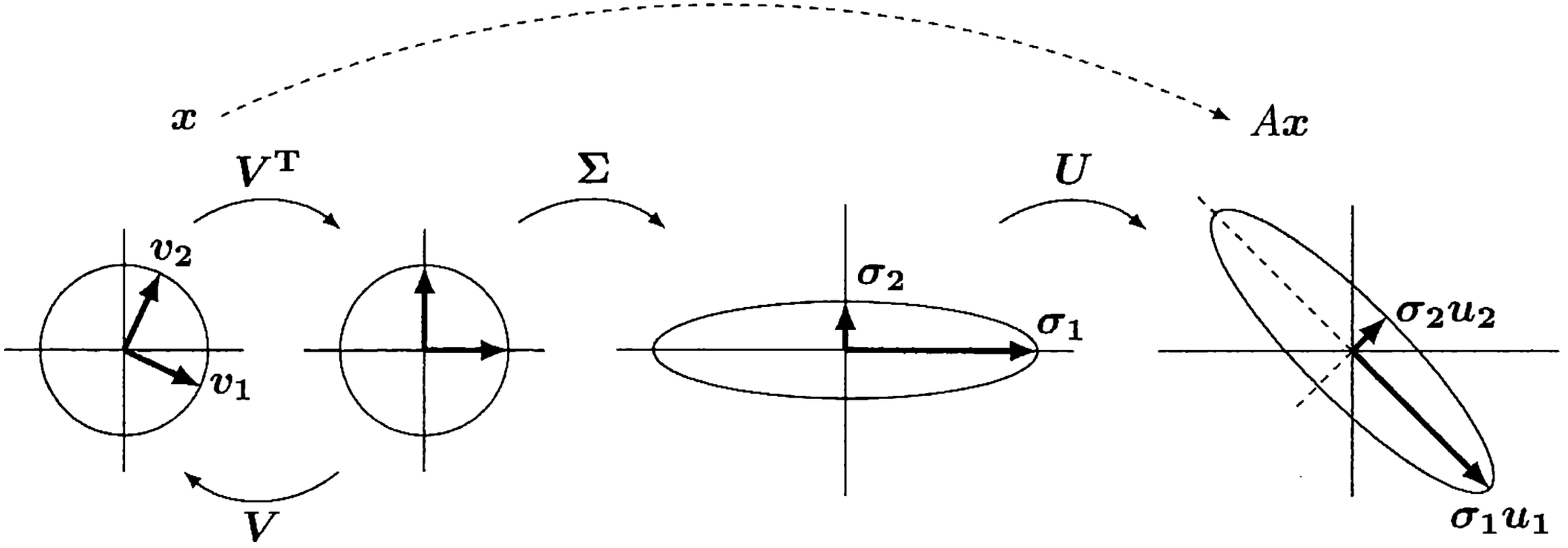
Quelle transformation linéaire représente-t-elle ?

Décomposition en valeurs singulières

$$AV = U\Sigma \qquad A \begin{bmatrix} v_1 \,..\, v_r \,..\, v_n \end{bmatrix} = \begin{bmatrix} u_1 \,..\, u_r \,..\, u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$

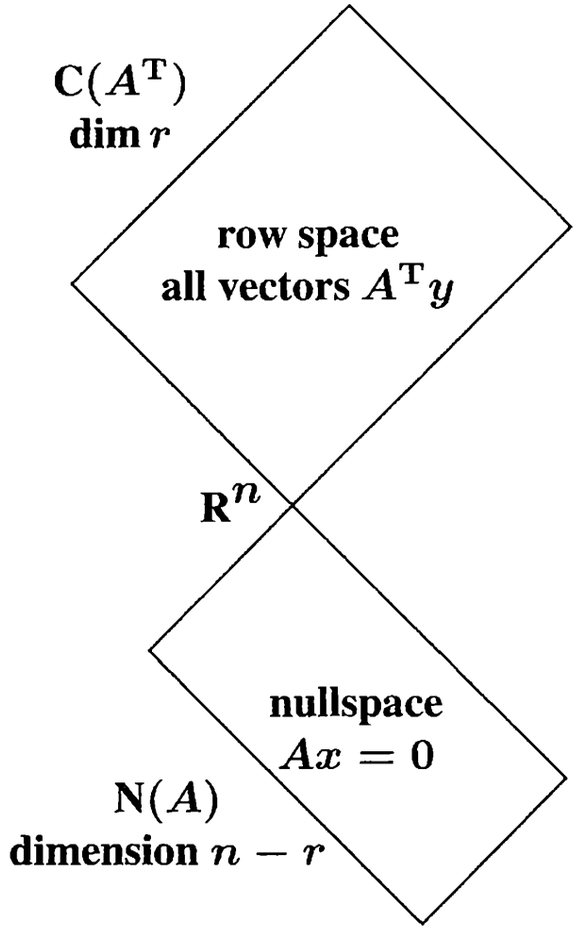
Décomposition en valeurs singulières

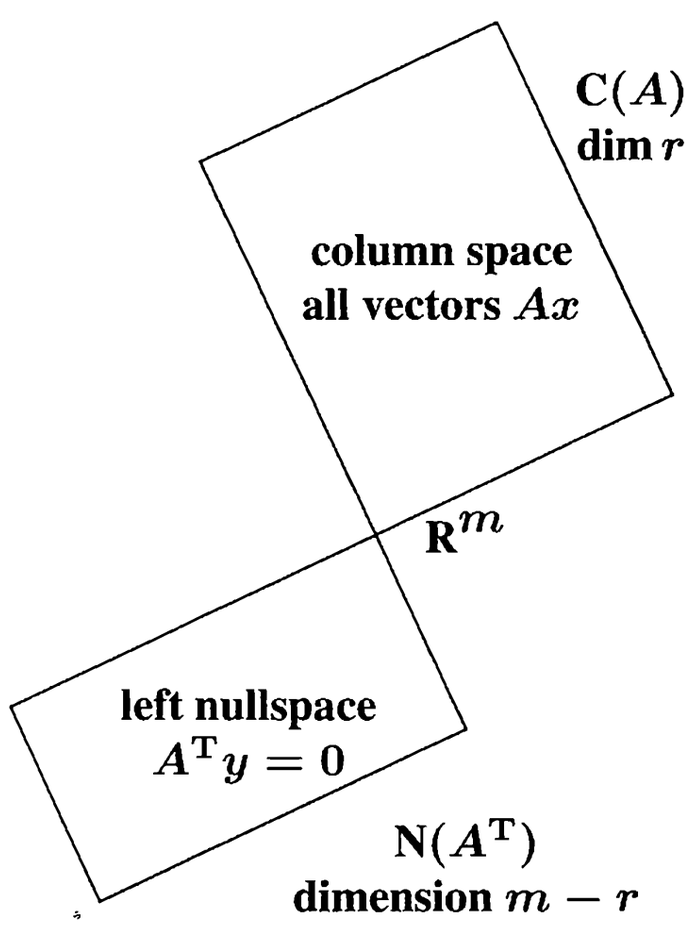
$$AV = U\Sigma \qquad A \begin{bmatrix} & & \\ v_1 & .. & v_r \\ & & \end{bmatrix} .. v_n = \begin{bmatrix} & & \\ u_1 & .. & u_r \\ & & \end{bmatrix} .. u_m \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$



Espaces associés à une matrice

$$AV = U\Sigma \qquad A\begin{bmatrix} v_1 \,..\, v_r \,..\, v_n \end{bmatrix} = \begin{bmatrix} u_1 \,..\, u_r \,..\, u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array}\right]$$





Theorème spectral

Si S est une matrice symétrique, réelle de taille $m \times m$, alors il existe une matrice orthogonale réelle Q de taille $m \times m$ et une matrice diagonale réelle Λ de taille $m \times m$ telles que $S = Q\Lambda Q^T$.

Une matrice symétrique réelle est dite définie positive, noté $S \succ 0$ ssi pour toute matrice colonne u, $u^T Su > 0$.

Une matrice symétrique réelle est dite semi-définie positive, $S \succeq 0$, ssi pour toute matrice colonne u, $u^T Su \geq 0$.

Relation d'ordre sur les matrices symétriques réelles:
$S_1 \prec S_2$ ssi $S_2 - S_1 \succ 0$
et
$S_1 \preceq S_2$ ssi $S_2 - S_1 \succeq 0$

Diagonalisation

Soit A une matrice à coefficients réels de taille $m \times m$. $\lambda \in \mathbf{C}$ est une valeur propre de A ssi il existe $v \in \mathbf{C}^m$, non-nul, tel que $Av = \lambda v$, i.e. l'image de v par A est dans la même direction que v. Un tel v est appelé un vecteur propre de A associé à la valeur propre λ.

Diagonalisation

suppose $v_1, \ldots, v_n$ is a *linearly independent* set of eigenvectors of $A \in \mathbf{R}^{n \times n}$:

$$Av_i = \lambda_i v_i, \quad i = 1, \ldots, n$$

express as

$$A \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

define $T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ and $\Lambda = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$, so

$$AT = T\Lambda$$

and finally

$$T^{-1}AT = \Lambda$$

Forme canonique de Jordan

any matrix $A \in \mathbf{R}^{n \times n}$ can be put in *Jordan canonical form* by a similarity transformation, $i.e.$

$$T^{-1}AT = J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix}$$

where

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix} \in \mathbf{C}^{n_i \times n_i}$$

is called a *Jordan block* of size n_i with eigenvalue λ_i (so $n = \sum_{i=1}^{q} n_i$)

Determinant et trace

A matrice carrée $m \times m$

$$\det(A) = \sum_{\sigma \in S_n} \left(\operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i} \right),$$

$\operatorname{sgn}(\sigma)$ est la parité du nombre d'élement dans une décomposition de σ en une séquence de transpositions (échange de deux éléments).

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

$$\operatorname{Tr}(A) = \sum_{i=1}^{m} a_{i,i}$$

$$\operatorname{Tr}(\mathrm{AB}) = \operatorname{Tr}(\mathrm{BA})$$

Crédits

Certaines figures de

- Strang, Gilbert. *Linear algebra and learning from data*. Cambridge: Wellesley-Cambridge Press, 2019

- Boyd, Stephen EE263 - Introduction to Linear Dynamical Systems (https://see.stanford.edu/Course/EE263)

- Wikipedia