Chapitre 15

La décomposition en valeurs singulières

15.1 Introduction

"Today, singular value decomposition has spread through many branches of science, in particular psychology and sociology, climate and atmospheric science, and astronomy. It is also extremely useful in machine learning and in both descriptive and predictive statistics. "

Peter Mills (lien web)

"Eigenvalues and eigenvectors are restricted to square matrices. But data comes in rectangular matrices. "

Gilbert Strang (lien web)

Si la diagonalisation a permis de comprendre la nature géométrique de certaines applications linéaires, elle exige malheureusement que l'application considérée se prête à cette analyse (qu'elle soit *diagonalisable* justement), et surtout : elle ne s'applique qu'à des matrices carrées.

La décomposition en valeurs singulières

La décomposition en valeurs singulières (SVD=Singular Value Decomposition) est une méthode très générale de factorisation qui donne une nouvelle interprétation géométrique de n'importe quelle application linéaire $T : \mathbb{R}^n \to \mathbb{R}^m$. Elle consiste à factoriser une matrice quelconque A $(m \times n)$ en un produit,

$$A = U\Sigma V^T\,,$$

où

a) U est $m \times m$, orthogonale : $U^TU = UU^T = I_m$,

b) Σ est $m \times n$, **diagonale** dans le sens où $\Sigma_{ij} = 0$ si $i \neq j$. De plus, $\Sigma_{ij} \geqslant 0$.

c) V est $n \times n$, orthogonale : $V^TV = VV^T = I_n$.

On sait d'une part que les matrices orthogonales représentent des *isométries*, c'est-à-dire des transformations rigides de l'espace, et doivent être comprises essentiellement comme des *rotations*. D'autre part, une matrice diagonale $m \times n$ a pour effet de stretcher certaines directions (avec un changement de dimension, voir plus bas). Donc la décomposition en valeurs singulières permet de décomposer l'application $T : \mathbb{R}^n \to \mathbb{R}^m$ définie par $T(\mathbf{x}) = A\mathbf{x}$ en trois parties :

$$\mathbb{R}^n \overset{\overset{V^T}{\text{(rotation)}}}{\longrightarrow} \mathbb{R}^n \overset{\overset{\Sigma}{\text{(stretch)}}}{\longrightarrow} \mathbb{R}^m \overset{\overset{U}{\text{(rotation)}}}{\longrightarrow} \mathbb{R}^m$$

Il est important d'insister sur le fait que la décomposition en valeurs singulières ne suppose rien sur A ; elle est toujours possible. En particulier, elle s'applique à des matrices qui ne sont pas forcément carrées.

Structure

Dans la section suivante, nous établirons rigoureusement la décomposition en valeurs singulières. Pour l'instant, supposons qu'une décomposition

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

soit donnée, et voyons ce que cela dit déjà sur les matrices U, Σ et V.

Nommons les colonnes de V, U et Σ :

$$\begin{aligned} U &= [\mathbf{u}_1 \cdots \mathbf{u}_m] \qquad & \mathbf{u}_k &\in \mathbb{R}^m \\ \Sigma &= [\sigma_1 \cdots \sigma_n] \qquad & \sigma_i &\in \mathbb{R}^m \\ V &= [\mathbf{v}_1 \cdots \mathbf{v}_n] \qquad & \mathbf{v}_j &\in \mathbb{R}^n \,. \end{aligned}$$

Puisque U et V sont orthogonales, les familles $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathbb{R}^m$ et $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{R}^n$ sont orthogonales.

La matrice Σ représente une application $\mathbb{R}^n \to \mathbb{R}^m$ dont la simplicité rappelle celle des matrices diagonales *carrées*. Nous noterons σ_i les éléments diagonaux de Σ. Notons que si $m > n$ (resp. $m < n$), alors certaines lignes (resp. colonnes) de Σ sont nulles.

Exemple 15.1. Si $m = 7$ et $n = 4$, alors les 3 dernières lignes de Σ sont nulles :

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Notons $\mathcal{B}_{\text{can}}$ et $\mathcal{B}'_{\text{can}}$ les bases canoniques de $\mathbb{R}^4$ et $\mathbb{R}^7$:

$$\begin{aligned} \mathcal{B}_{\text{can}} &= (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4)\,, \\ \mathcal{B}'_{\text{can}} &= (\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3, \mathbf{e}'_4, \mathbf{e}'_5, \mathbf{e}'_6, \mathbf{e}'_7)\,. \end{aligned}$$

L'application $\mathbf{x} \mapsto \Sigma\mathbf{x}$ représente des "stretches" pour les 4 vecteurs de $\mathcal{B}_{\text{can}}$,

$$\Sigma\mathbf{e}_1 = \sigma_1\mathbf{e}'_1\,, \quad \Sigma\mathbf{e}_2 = \sigma_2\mathbf{e}'_2\,, \quad \Sigma\mathbf{e}_3 = \sigma_3\mathbf{e}'_3\,, \quad \Sigma\mathbf{e}_4 = \sigma_4\mathbf{e}'_4\,.$$

◇

Exemple 15.2. Si $m = 3$, $n = 5$, alors les 2 dernières colonnes de Σ sont nulles :

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \end{pmatrix}$$

Notons $\mathcal{B}_{\text{can}}$ et $\mathcal{B}'_{\text{can}}$ les bases canoniques de $\mathbb{R}^5$ et $\mathbb{R}^3$:

$$\begin{aligned} \mathcal{B}_{\text{can}} &= (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5)\,, \\ \mathcal{B}'_{\text{can}} &= (\mathbf{e}'_1, \mathbf{e}'_2, \mathbf{e}'_3)\,. \end{aligned}$$

On a des "stretches" pour les 3 premiers vecteurs de $\mathcal{B}_{\text{can}}$,

$$\Sigma\mathbf{e}_1 = \sigma_1\mathbf{e}'_1\,, \quad \Sigma\mathbf{e}_2 = \sigma_2\mathbf{e}'_2\,, \quad \Sigma\mathbf{e}_3 = \sigma_3\mathbf{e}'_3\,,$$

mais les deux derniers sont tous envoyés sur le vecteur nul :

$$\Sigma\mathbf{e}_4 = \Sigma\mathbf{e}_5 = \mathbf{0}\,.$$

◇

Remarque 15.3. Les relations ci-dessus, "$\Sigma\mathbf{e}_j = \sigma_j\mathbf{e}'_j$", rappellent celles du type "$A\mathbf{v} = \lambda\mathbf{v}$". La grande différence ici est que $\mathbf{e}_j$ et $\mathbf{e}'_j$ vivent dans des espaces différents ! ◇

Pour comprendre les relations entre les $\mathbf{u}_k$, les $\mathbf{v}_j$ et la matrice Σ, on multiplie A par sa transposée pour obtenir une matrice $n \times n$ donnée par

$$\begin{aligned} A^T A &= (U\Sigma V^T)^T(U\Sigma V^T) \\ &= V\Sigma^T U^T U\Sigma V^T \\ &= V(\Sigma^T\Sigma)V^T \end{aligned}$$

On a, dans le terme de droite, trois matrices $n \times n$. Puisque $\Sigma^T\Sigma$ est diagonale, et puisque V^T est l'inverse de V (car cette dernière est orthogonale), on voit que ce produit de trois matrices carrées représente une *diagonalisation de la matrice symétrique* $A^T A$. En particulier, **les colonnes de V sont des vecteurs propres orthonormés de** $A^T A$, associés à des valeurs propres qui sont les éléments diagonaux de $\Sigma^T\Sigma$, à savoir σ_i^2 :

$$(A^T A)\mathbf{v}_j = \sigma_j^2\mathbf{v}_j \qquad \forall j = 1, \dots, n\,.$$

De même pour AA^T : c'est une matrice $m \times m$, et

$$AA^T = U(\Sigma\Sigma^T)U^T\,,$$

qui implique que **les colonnes de U sont des vecteurs propres orthonormés de** AA^T, associés à des valeurs propres qui sont les éléments diagonaux de $\Sigma\Sigma^T$:

$$AA^T\mathbf{u}_k = \sigma_k^2\mathbf{u}_k \qquad \forall k = 1, \dots, m\,.$$

Remarquons encore que dans les deux cas, les matrices diagonales $\Sigma^T\Sigma$ et $\Sigma\Sigma^T$ ont des coefficients diagonaux donnés par les carrés σ_i^2.

Cette discussion montre que si une décomposition en valeurs singulières existe, alors les matrices U et V se calculent en diagonalisant AA^T et $A^T A$. (On verra comment simplifier un peu ce procédé par la suite.)

Ce qui n'est pas du tout démontré par l'argument ci-dessus, c'est si la décomposition existe effectivement ; nous le démontrerons dans la section suivante.

Matrices définies par blocs

Dans ce chapitre, nous définirons et manipulerons des matrices définies **par blocs**, ce qui signifie définies comme composées de sous-matrices. On utilisera l'indice "□" pour indiquer qu'une matrice est définie par blocs.

Exemple 15.4. Avec

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix},$$

on peut définir

$$\begin{bmatrix} A & B \end{bmatrix}_\square = \begin{pmatrix} a & b & c & 1 & 2 \\ d & e & f & 3 & 4 \end{pmatrix}.$$

◇

Les blocs qui composent une matrice par blocs doivent avoir des dimensions compatibles.

Plus généralement, si on possède quatre matrices,

$$A\,(m \times k), \quad B\,(m \times l), \quad C\,(h \times k), \quad D\,(h \times l),$$

on peut définir

⋆ la matrice $m \times (k+l)$:
$$\begin{bmatrix} A & B \end{bmatrix}_\square,$$

⋆ la matrice $(m+h) \times k$:
$$\begin{bmatrix} A \\ C \end{bmatrix}_\square.$$

⋆ la matrice $(m+h) \times (k+l)$:
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}_\square.$$

15.2 Existence

Dans cette section, on montre que toute matrice possède une décomposition en valeurs singullières :

Théorème 15.5. *(Existence d'une SVD) Toute matrice $A\,(m \times n)$ peut s'écrire comme un produit,*

$$A = U \Sigma V^T,$$

où

a) U est $m \times m$, orthogonale : $U^T U = U U^T = I_m$; ses colonnes sont appelées ***left singular vectors*** *de A.*

b) Σ est $m \times n$, diagonale, les coefficients situés sur sa diagonale sont $\geqslant 0$, et sont appelés les ***valeurs singulières*** *de A.*

c) V est $n \times n$, orthogonale : $V^T V = V V^T = I_n$, ses colonnes sont appelées ***right singular vectors*** *de A.*

Les matrices $A^T A$ et AA^T

Notre point de départ :

Lemme 22. *Pour une matrice A ($m \times n$) quelconque,*

- ⋆ *$A^T A$ ($n \times n$) est symétrique,*
- ⋆ *AA^T ($m \times m$) est symétrique.*

Preuve: Par les propriétés de la transposée,

$$(A^T A)^T = A^T (A^T)^T = A^T A\,,$$
$$(AA^T)^T = (A^T)^T A^T = AA^T\,.$$

□

Exemple 15.6. Si $A = \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix}$, alors

$$A^T A = \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 5 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix} = \begin{pmatrix} 14 & 13 \\ 13 & 29 \end{pmatrix}\,,$$

et

$$AA^T = \begin{pmatrix} -1 & 2 \\ 2 & 0 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 5 & -2 & 7 \\ -2 & 4 & 6 \\ 7 & 6 & 34 \end{pmatrix}$$

◇

Étant symétriques, le **Théorème Spectral** (lien web) garantit que $A^T A$ et AA^T sont diagonalisables :

- ⋆ Il existe une matrice orthogonale V ($n \times n$) et une matrice diagonale D ($n \times n$) telle que
$$A^T A = VDV^T\,.$$
- ⋆ Il existe une matrice orthogonale U ($m \times m$) et une matrice diagonale D' ($m \times m$) telle que
$$AA^T = UD'U^T\,.$$

On sait que les éléments diagonaux de D (resp. D') sont les valeurs propres de $A^T A$ (resp. AA^T), avec éventuellement des répétitions selon les dimensions des espaces propres associés. Or ces valeurs propres ont des propriétés particulières :

Lemme 23. *Pour toute matrice A,*

a) Un scalaire $\lambda \neq 0$ est valeur propre de $A^T A$ si et seulement si il est également valeur propre de AA^T.

b) Si λ est valeur propre de $A^T A$ ou de AA^T, alors $\lambda \geqslant 0$.

Preuve: 1. Supposons que $\lambda \neq 0$ est valeur propre de $A^T A$. Alors il existe $\mathbf{v} \in \mathbb{R}^n$, non-nul, tel que

$$(A^T A)\mathbf{v} = \lambda \mathbf{v}\,.$$

Remarquons que $A\mathbf{v} \neq \mathbf{0}$ puisque $A^T A\mathbf{v} \neq \mathbf{0}$.

Ensuite, en multipliant les deux côtés de l'identité de dessus par A, on obtient

$$AA^T(A\mathbf{v}) = \lambda(A\mathbf{v}),$$

qui signifie que λ est aussi valeur propre de AA^T, associée au vecteur propre $A\mathbf{v} \in \mathbb{R}^m$ (qui est non-nul comme on a dit). Le même argument montre que toute valeur propre non-nulle de AA^T est également valeur propre de A^TA.

2. Maintenant avec une valeur propre λ de A^TA, et un vecteur propre $\mathbf{v} \neq \mathbf{0}$, $A^TA\mathbf{v} = \lambda\mathbf{v}$, on peut écrire

$$\begin{aligned}\lambda\|\mathbf{v}\|^2 = \lambda(\mathbf{v}\cdot\mathbf{v}) &= \mathbf{v}\cdot(\lambda\mathbf{v})\\ &= \mathbf{v}\cdot(A^TA\mathbf{v})\\ &= (A\mathbf{v})\cdot(A\mathbf{v})\\ &= \|A\mathbf{v}\|^2 \geqslant 0\,.\end{aligned}$$

Puisque $\|\mathbf{v}\| > 0$, on en déduit que $\lambda \geqslant 0$. □

Les deux lemmes ci-dessus impliquent :

Corollaire 9. *Pour toute matrice A ($m \times n$), il existe au plus $\min\{m, n\}$ valeurs propres non-nulles communes de A^TA et AA^T.*

Preuve: On sait que A^TA est $n \times n$ et possède donc au maximum n valeurs propres non-nulles, et AA^T est $m \times m$ et possède donc au maximum m valeurs propres non-nulles. Comme ces matrices ont les mêmes valeurs propres non-nulles, le nombre de ces valeurs propres non-nulles est plus petit que n et que m. □

Preuve du théorème :

(La preuve ci-dessous est tirée de **wikipedia** (lien web).) Considérons la diagonalisation de A^TA :

$$A^TA = VDV^T\,.$$

En multipliant à gauche par V^T puis à droite par V,

$$V^T(A^TA)V = D\,.$$

Par le lemme ci-dessus, toutes les valeurs propres de A^TA, sur la diagonale de D, sont $\geqslant 0$.

Sans perte de généralité, on peut supposer que la valeur propre nulle (éventuellement répétée) apparaît en bas de la diagonale :

$$D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_l, 0, \ldots, 0)\,,$$

avec $\lambda_1 > 0, \ldots, \lambda_l > 0$. Distinguons ensuite la sous-matrice diagonale de D qui contient les valeurs propres strictement positives, en écrivant :

$$D = \begin{bmatrix} D_* & 0 \\ 0 & 0 \end{bmatrix}_{\square},$$

où $D_* = \mathrm{diag}(\lambda_1, \ldots, \lambda_l)$ est $l \times l$, et les "0" sont des matrices nulles.

À l'ordre fixé par les valeurs propres dans D correspond un ordre des colonnes dans la matrice de changement de base V :

$$V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}_{\square},$$

où

- ⋆ V_1 est une matrice $n \times l$ dont les colonnes forment une famille libre de vecteurs propres associés aux valeurs propres non-nulles $\lambda_1, \ldots, \lambda_l$.
- ⋆ V_2 est une matrice $n \times (n-l)$ dont les colonnes forment une famille libre de vecteurs propres associés à la valeur propre $\lambda = 0$.

L'orthonormalité des colonnes de V implique que

$$V_1^T V_1 = I_l\,, \qquad V_2^T V_2 = I_{n-l}\,,$$

mais la relation $VV^T = I_n$ implique aussi que

$$I_n = VV^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix}_\square \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_\square = V_1 V_1^T + V_2 V_2^T\,.$$

Utilisons ces matrices V_1 et V_2 pour récrire la diagonalisation $A^T A = VDV^T$, qui est équivalente à $V^T(A^T A)V = D$. Comme

$$\begin{aligned} V^T(A^T A)V &= \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_\square A^T A \begin{bmatrix} V_1 & V_2 \end{bmatrix}_\square \\ &= \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_\square \begin{bmatrix} A^T A V_1 & A^T A V_2 \end{bmatrix}_\square \\ &= \begin{bmatrix} V_1^T A^T A V_1 & V_1^T A^T A V_2 \\ V_2^T A^T A V_1 & V_2^T A^T A V_2 \end{bmatrix}_\square\,, \end{aligned}$$

et comme cette matrice est égale à

$$D = \begin{bmatrix} D_* & 0 \\ 0 & 0 \end{bmatrix}_\square\,,$$

ceci implique que

$$V_1^T A^T A V_1 = D_* \quad (l \times l)$$

et

$$V_2^T A^T A V_2 = 0 \quad ((n-l) \times (n-l))$$

De cette dernière, on tire que $(AV_2)^T(AV_2) = 0$, qui implique que

$$AV_2 = 0\,.$$

(En effet, on sait que pour toute matrice M, $M^T M$ contient tous les produits scalaires possibles entre les colonnes de M, en particulier, sur sa diagonale, les carrés des normes des colonnes. Si $M^T M = 0$, cela implique que la norme de chaque colonne de M est nulle, et donc que M est la matrice nulle.)

Définissons maintenant la matrice $m \times l$:

$$U_1 := AV_1 D_*^{-1/2}\,,$$

où

$$D_*^{-1/2} := \mathrm{diag}(1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_l})$$

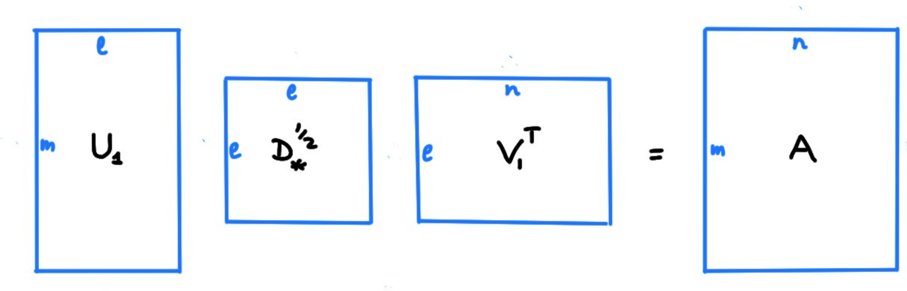
est bien définie puisque $\lambda_k > 0$ pour tout $k = 1, \ldots, l$, et n'est rien d'autre que l'inverse de

$$D_*^{1/2} := \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_l}).$$

On remarque maintenant que les colonnes de U_1 forment une famille orthonormale, puisque

$$\begin{aligned} U_1^T U_1 &= (AV_1 D_*^{-1/2})^T AV_1 D_*^{-1/2} \\ &= D_*^{-1/2}(V_1^T A^T AV_1) D_*^{-1/2} \\ &= D_*^{-1/2} D_* D_*^{-1/2} \\ &= I_l \,. \end{aligned}$$

Montrons que U_1, D_* et V_1 fournissent déjà une première factorisation de A :



En effet,

$$\begin{aligned} U_1 D_*^{1/2} V_1^T &= AV_1 \underbrace{D_*^{-1/2} D_*^{1/2}}_{=I_l} V_1^T \\ &= AV_1 V_1^T \\ &= A(I_n - V_2 V_2^T) \\ &= A - \underbrace{(AV_2)}_{=0} V_2^T \\ &= A \end{aligned}$$

Cette première factorisation constitue la base de l'argument; il reste maintenant à modifier le produit $U_1 D_*^{1/2} V_1^T$, en augmentant les tailles des matrices, de façon à ce qu'il devienne $U \Sigma V^T$.

On rajoute d'abord à V_1^T le bloc V_2^T, ce qui donne

$$V^T = \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_{\square}$$

Passons à U. Si $l = m$, alors on peut prendre $U = U_1$. Mais, si $l < m$, U_1 n'est pas carrées : ses colonnes forment une base orthonormée de $\mathrm{Col}(U_1) \subset \mathbb{R}^m$, mais pas de $\mathbb{R}^m$, on peut donc compléter cette base en une base de $\mathbb{R}^m$, et même, via un procédé de Gram-Schmidt si nécessaire, la compléter en une base orthonormée de $\mathbb{R}^n$. Les $m - l$ vecteurs rajoutés peuvent être rangés dans une matrice U_2, $m \times (m - l)$, qui permet de définir la matrice $m \times m$ orthogonale

$$U := \begin{bmatrix} U_1 & U_2 \end{bmatrix}_{\square} \,.$$

Finalement, Σ ($m \times n$) est construite à partir de $D_*^{1/2}$ ($l \times l$), en rajoutant des blocs nuls, si nécessaire (rappelons que $l \leqslant \min\{m, n\}$) :

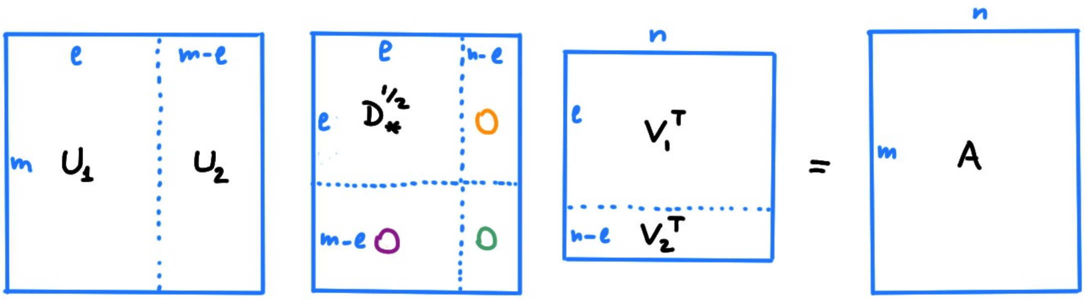
$$\Sigma := \begin{bmatrix} D_*^{1/2} & 0 \\ 0 & 0 \end{bmatrix}_{\square} \,.$$

où

- ⋆ 0 est $l \times (n-l)$,
- ⋆ 0 est $(m-l) \times l$,
- ⋆ 0 est $(m-l) \times (n-l)$.

Remarquons que ceci peut a priori faire apparaître des valeurs singulières nulles sur la diagonale de Σ.

Montrons que l'on a ce qu'on voulait :



En effet,

$$\begin{aligned} U\Sigma V^T &= \begin{bmatrix} U_1 & U_2 \end{bmatrix}_\square \begin{bmatrix} D_*^{1/2} & 0 \\ 0 & 0 \end{bmatrix}_\square \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}_\square \\ &= \begin{bmatrix} U_1 & U_2 \end{bmatrix}_\square \begin{bmatrix} D_*^{1/2} V_1^T \\ 0 \end{bmatrix}_\square \\ &= U_1 D_*^{1/2} V_1^T \\ &= A\,. \end{aligned}$$

Ceci termine la preuve de l'existence d'une SVD.

Quelques remarques au vu de la preuve que l'on vient de donner :

- ⋆ Les valeurs singulières non-nulles de A sont **les racines carrées des valeurs propres de** $A^T A$ (et AA^T) :
$$\sigma_j = \sqrt{\lambda_j}\,.$$
Les valeurs singulières nulles sont possibles (elles apparaissent au moment où on complète $D_*^{1/2}$ avec des blocs de zéros), et sont liées au fait que $A^T A$ ou AA^T peuvent posséder $\lambda = 0$ comme valeur propre.
- ⋆ On l'a dit, les colonnes de V forment une base orthonormée de $\mathbb{R}^n$, formée de vecteurs propres de $A^T A$, et les colonnes de U forment une base orthonormale de $\mathbb{R}^m$, formée de vecteurs propres de AA^T.
- ⋆ La définition du bloc $U_1 = [\mathbf{u}_1 \cdots \mathbf{u}_l]$ peut aussi s'exprimer comme suit :
$$\begin{aligned} U_1 &= AV_1 D_*^{-1/2} \\ &= [A\mathbf{v}_1 \cdots A\mathbf{v}_l] D_*^{-1/2} \\ &= [\tfrac{1}{\sqrt{\lambda_1}} A\mathbf{v}_1 \cdots \tfrac{1}{\sqrt{\lambda_l}} A\mathbf{v}_l]\,. \end{aligned}$$
On a donc toujours un moyen direct de calculer les l premières colonnes de U_1 :
$$\mathbf{u}_j := \frac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j\,, \qquad j = 1, 2, \ldots, l\,,$$
où $\mathbf{v}_j$ est la j-ème colonne de V_1, correspondant au vecteur propre orthonormé de $A^T A$, associé à la j-ème valeur propre $\lambda_j > 0$.

⋆ Comme la décomposition QR, la décomposition en valeurs singulières existe toujours, mais n'est pas unique. En effet, le choix des vecteurs propres, dans la construction de V, peut toujours se faire de multiples façons, menant à autant de SVD différentes.

15.3 Exemples

La preuve de la section précédente a montré clairement quelles sont les étapes menant à une décomposition singulière d'une matrice A :

a) Diagonaliser $A^T A$: distinguer ses valeurs propres positives, $\lambda_1 > 0, \dots, \lambda_l > 0$, qui donnent les valeurs singulières de A

$$\sigma_j = \sqrt{\lambda_j}\,, \qquad j = 1, 2, \dots, l\,.$$

Les vecteurs propres de $A^T A$ donnent $V = [\mathbf{v}_1, \dots, \mathbf{v}_n]$, les $n - l$ dernières colonnes étant associées à la valeur propre nulle (si besoin est).

b) En principe, les colonnes de U sont les vecteurs propres de AA^T, mais pour faire plus simple, on peut d'abord calculer les l premières colonnes de U par

$$\mathbf{u}_j := \tfrac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j\,, \qquad j = 1, 2, \dots, l\,,$$

et ensuite, si nécessaire, déterminer les $m - l$ dernières en complétant $(\mathbf{u}_1, \dots, \mathbf{u}_l)$ en une base orthonormale $(\mathbf{u}_1, \dots, \mathbf{u}_l, \mathbf{v}_{l+1}, \dots, \mathbf{v}_n)$ (via Gram-Schmidt + normalisation).

c) La matrice Σ s'obtient à partir des valeurs singulières non-nulles $\sigma_1, \dots, \sigma_l$, en rajoutant les blocs de zéros nécessaires.

Remarque 15.7. Le calcul des premiers $\mathbf{u}_j$ peut également se faire comme suit :

$$\mathbf{u}_j := \tfrac{1}{\|A\mathbf{v}_j\|} A\mathbf{v}_j\,, \qquad j = 1, 2, \dots, l\,.$$

En effet, par un calcul que l'on a déjà fait,

$$\begin{aligned}
\|A\mathbf{v}_j\|^2 &= (A\mathbf{v}_j) \cdot (A\mathbf{v}_j) \\
&= \mathbf{v}_j \cdot (A^T A\mathbf{v}_j) \\
&= \mathbf{v}_j \cdot (\lambda_j \mathbf{v}_j) \\
&= \lambda_j \|\mathbf{v}_j\|^2 \\
&= \lambda_j\,,
\end{aligned}$$

et donc $\|A\mathbf{v}_j\| = \sqrt{\lambda_j}$. ◇

Informel 15.8. Remarquons que le travail nécessaire pour diagonaliser $A^T A$ et AA^T peut être très différent, étant donné que ces matrices sont a priori de tailles différentes !

Exemple 15.9. Calculons la décomposition en valeurs singulières de

$$A = \begin{pmatrix} \frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}} \end{pmatrix}$$

Comme A est 2×2, sa décomposition $A = U\Sigma V^T$ sera un produit de trois matrices 2×2.

On commence par calculer V, qui on le rappelle est formée de vecteurs propres de A^TA. Or

$$A^TA = \frac{1}{100}\begin{pmatrix} 153 & 96 \\ 96 & 97 \end{pmatrix},$$

et on sait (voir exercices) que cette dernière possède deux valeurs propres, $\lambda_1 = 9/4$, $\lambda_2 = 1/4$, et que les espaces propres associés sont

$$E_{9/4} = \mathrm{Vect}\left\{\begin{pmatrix} 4 \\ 3 \end{pmatrix}\right\} \qquad E_{1/4} = \mathrm{Vect}\left\{\begin{pmatrix} -3 \\ 4 \end{pmatrix}\right\}$$

qui donne, après normalisation,

$$E_{9/4} = \mathrm{Vect}\left\{\begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix}\right\} \qquad E_{1/4} = \mathrm{Vect}\left\{\begin{pmatrix} -3/5 \\ 4/5 \end{pmatrix}\right\}$$

On peut donc prendre

$$V = \begin{pmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \end{pmatrix},$$

qui correspond à une rotation d'angle $\theta = \mathrm{Arccos}(4/5)$. Ainsi, $V^T = V^{-1}$ correspond à une rotation de $-\theta$.

Étant connues les valeurs propres de A^TA, les valeurs singulières de A sont données par

$$\sigma_1 = \sqrt{\lambda_1} = \frac{3}{2}, \qquad \sigma_2 = \sqrt{\lambda_2} = \frac{1}{2},$$

ce qui donne

$$\Sigma = \begin{pmatrix} 3/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Ensuite, U a pour colonnes des vecteurs propres de AA^T, or

$$AA^T = \begin{pmatrix} 5/4 & 1 \\ 1 & 5/4 \end{pmatrix},$$

qui possède comme valeurs propres $\lambda_1 = 9/4$ et $\lambda_2 = 1/4$ (comme on sait, les mêmes que A^TA !). Ses espaces propres correspondants sont donnés par

$$E_{9/4} = \mathrm{Vect}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\} \qquad E_{1/4} = \mathrm{Vect}\left\{\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\},$$

ou encore, après normalisation :

$$E_{9/4} = \mathrm{Vect}\left\{\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\} \qquad E_{1/4} = \mathrm{Vect}\left\{\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}\right\},$$

On a donc

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix},$$

qui n'est autre qu'une rotation de $\phi = \frac{\pi}{4}$.

Remarquons qu'on aurait aussi pu trouver les colonnes de U en faisant

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{3/2}\begin{pmatrix} \frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}} \end{pmatrix}\begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

pareil pour $\mathbf{u}_2$.

On a donc la décomposition en valeurs singulières de A, qui permet de voir la transformation

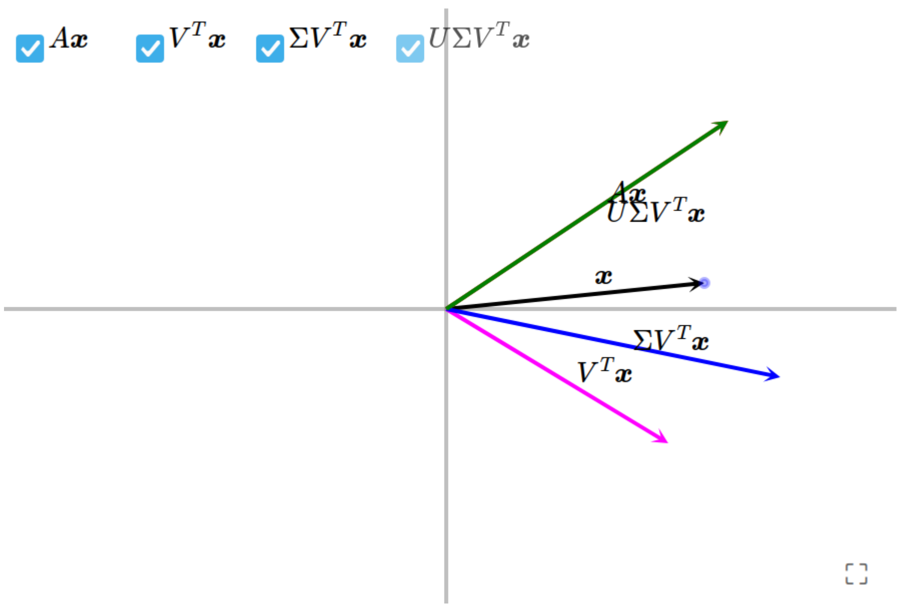
$$\mathbf{x} \mapsto A\mathbf{x} = \underbrace{\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \underbrace{\begin{pmatrix} 3/2 & 0 \\ 0 & 1/2 \end{pmatrix} \underbrace{\begin{pmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{pmatrix} \mathbf{x}}_{V^T\mathbf{x}}}_{\Sigma V^T\mathbf{x}}}_{U\Sigma V^T\mathbf{x}} .$$

comme une composition

a) d'une rotation d'angle $-\theta \simeq -36.9^o$, suivie

b) d'un stretch le long des axes de coordonnées ($3/2$ selon $\mathbf{e}_1$, $1/2$ selon $\mathbf{e}_2$), suivi

c) d'une rotation d'angle $\phi = +45^o$.

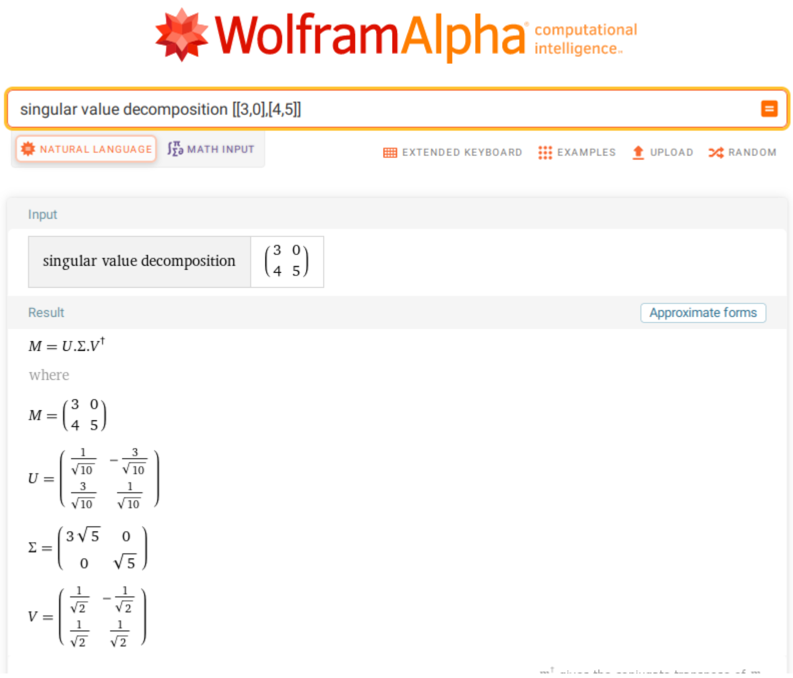


◇

Remarque 15.10. **Wolfram Alpha** (lien web) peut donner une décomposition en valeurs singulières de n'importe quelle matrice. Par exemple, pour obtenir la décomposition de

$$M = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix} ,$$

il suffit d'entrer CODE$>singularvaluedecomposition[[3,0],[4,5]]<$CODE



◇

Exemple 15.11. Calculons la décomposition en valeurs singulières de

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$$

On commence par

$$A^T A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix},$$

qui possède deux valeurs propres, $\lambda_1 = 4$ et $\lambda_2 = 0$. Ainsi, A possède une seule valeur singulière non-nulle : $\sigma_1 = \sqrt{\lambda_1} = 2$. On trouve un vecteur propre unitaire pour chaque valeur propre, par exemple :

$$\mathbf{v}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

qui donne déjà

$$V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix},$$

Pour calculer U, on peut soit passer par l'étude de AA^T, ou alors commencer par obtenir une de ses colonnes en prenant

$$\mathbf{u}_1 = \frac{1}{\sqrt{\lambda_1}} A\mathbf{v}_1 = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}$$

On doit maintenant trouver deux colonnes $\mathbf{u}_2$ et $\mathbf{u}_3$, de façon à ce que $U = [\mathbf{u}_1\, \mathbf{u}_2\, \mathbf{u}_3]$ soit orthogonale. On peut par exemple prendre

$$\mathbf{u}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Il reste à produire Σ. Puisque A n'a qu'une seule valeur singulière non-nulle, et que Σ doit être 3×2, on rajoute des blocs appropriés :

$$\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} .$$

On a donc une SVD pour A :

$$\begin{aligned} A &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}^T , \end{aligned}$$

◇

Exemple 15.12. Étudions la SVD de

$$A = \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} ,$$

pour laquelle on aura U (2×2), Σ (2×3), V (3×3). Commençons par

$$A^T A = \begin{pmatrix} 10 & 0 & 2 \\ 0 & 10 & 4 \\ 2 & 4 & 2 \end{pmatrix} ,$$

qui a pour polynôme caractéristique

$$P_{A^T A}(\lambda) = -\lambda(\lambda - 10)(\lambda - 12) .$$

On a donc les valeurs propres, en ordre décroissant, $\lambda_1 = 12$, $\lambda_2 = 10$, $\lambda_3 = 0$. On a donc deux valeurs singulières strictement positives, $\sigma_1 = \sqrt{12} = 2\sqrt{3}$, $\sigma_2 = \sqrt{10}$.

Les espaces propres sont :

$$E_{12} = \ker(A^T A - 12 I_3) = \text{Vect}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\} ,$$

$$E_{10} = \ker(A^T A - 10 I_3) = \text{Vect}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \right\} ,$$

$$E_0 = \ker(A^T A) = \ker(A) = \text{Vect}\left\{ \begin{pmatrix} -1 \\ -2 \\ 5 \end{pmatrix} \right\} .$$

On peut donc normaliser et obtenir

$$V = [\mathbf{v}_1\, \mathbf{v}_2\, \mathbf{v}_3] = \begin{pmatrix} 1/\sqrt{6} & -2/\sqrt{5} & -1/\sqrt{30} \\ 2/\sqrt{6} & 1/\sqrt{5} & -2/\sqrt{30} \\ 1/\sqrt{6} & 0 & 5/\sqrt{30} \end{pmatrix}$$

La matrice $U = [\mathbf{u}_1\, \mathbf{u}_2]$ s'obtient par

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{2\sqrt{3}} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

Finalement, les deux valeurs singulières positives permettent d'écrire

$$\Sigma = \begin{pmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix}.$$

On a donc une SVD pour A :

$$\begin{aligned} A &= \begin{pmatrix} 3 & 1 & 1 \\ -1 & 3 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2\sqrt{3} & 0 & 0 \\ 0 & \sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} & -2/\sqrt{5} & -1/\sqrt{30} \\ 2/\sqrt{6} & 1/\sqrt{5} & -2/\sqrt{30} \\ 1/\sqrt{6} & 0 & 5/\sqrt{30} \end{pmatrix}^T \end{aligned}$$

◇

15.4 Rang et représentation

Soit A une matrice $m \times n$ dont une décomposition en valeurs singulières est donnée :

$$A = U \Sigma V^T.$$

Comme nous avons fait avec la décomposition spectrale, nous allons profiter de la SVD pour écrire A comme une combinaison linéaire de matrices plus simples.

Pour commencer, remarquons que l'on peut toujours imposer, dans une SVD, que les valeurs singulières apparaissent sur la diagonale de Σ en ordre décroissant :

$$\sigma_1 \geqslant \sigma_1 \geqslant \dots$$

En nommant les colonnes de U et de V :

$$U = \begin{bmatrix} \mathbf{u}_1 \cdots \mathbf{u}_m \end{bmatrix}, \qquad V = \begin{bmatrix} \mathbf{v}_1 \cdots \mathbf{v}_n \end{bmatrix}.$$

Définissons alors l'indice de la plus petite valeur singulière strictement positive,

$$r := \max\{1 \leqslant k \leqslant l \colon \sigma_k > 0\}.$$

et procédons comme on l'a fait pour la décomposition spectrale en écrivant, pour tout

$\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned}
A\mathbf{x} = U\Sigma V^T\mathbf{x} = \begin{bmatrix}\mathbf{u}_1 \cdots \mathbf{u}_n\end{bmatrix}\Sigma \begin{pmatrix}\mathbf{u}_1^T\mathbf{x} \\ \vdots \\ \mathbf{u}_n^T\mathbf{x}\end{pmatrix} \\
&= \begin{bmatrix}\sigma_1\mathbf{u}_1 \cdots \sigma_r\mathbf{u}_r\,\mathbf{0}\cdots\mathbf{0}\end{bmatrix} \begin{pmatrix}\mathbf{u}_1^T\mathbf{x} \\ \vdots \\ \mathbf{u}_n^T\mathbf{x}\end{pmatrix} \\
&= \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^T \mathbf{x} \\
&= \left(\sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) \mathbf{x}\,.
\end{aligned}$$

On a donc pu écrire A comme une somme :

$$\boxed{A = \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \mathbf{v}_k^T\,,}$$

On retrouve bien, pour tout j,

$$\begin{aligned}
A\mathbf{v}_j &= \sum_{k=1}^{r} \sigma_k \mathbf{u}_k \underbrace{\mathbf{v}_k^T\mathbf{v}_j}_{=\delta_{kj}} \\
&= \sigma_j \mathbf{u}_j\,.
\end{aligned}$$

Remarquons qu'à l'inverse de la décomposition spectrale, une matrice $\mathbf{u}_k\mathbf{v}_k^T$ ne représente *pas* une projection puisqu'elle transforme un vecteur de $\mathbb{R}^n$ en un vecteur de $\mathbb{R}^m$. Son seul point commun avec une projection est que

$$\mathrm{rang}(\mathbf{u}_k\mathbf{v}_k^T) = 1\,.$$

SVD fournit donc une représentation d'une matrice comme une somme de matrices de rang égal à 1. Aussi, comme on n'a gardé que les valeurs singulières $\sigma_k > 0$, et comme les $\mathbf{u}_k$ sont indépendants, on a montré :

Théorème 15.13. *Le rang de A est égal au nombre de valeurs singulières non-nulles.*

Approximation optimale par une matrice de rang fixé

Définissons, pour tout $k \leqslant r$,

$$A(k) := \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^T\,.$$

Alors $A(k)$ est la matrice de rang k qui approxime le mieux A, dans le sens suivant (c'est le Théorème d'*Eckart-Young*) :

Théorème 15.14. *Soit A ($m \times n$) de rang $l \leqslant n$. Pour tout $1 \leqslant k \leqslant l$, $A(k)$ est la matrice de rang k qui approxime le mieux A :*

$$\min_B \|A - B\| = \|A - A(k)\|\,,$$

où le minimum est pris sur toutes les matrices $m \times n$ de rang au plus égal à k.

Preuve: □

15.5 Élongations et ellipsoïdes

Dans cette section nous utiliserons SVD pour répondre à deux questions géométrique naturelles à propos d'une application linéaire $T : \mathbb{R}^n \to \mathbb{R}^m$ définie par une matrice A, $T(\mathbf{x}) = A\mathbf{x}$:

a) Comment se transforme **la sphère unité**, définie par

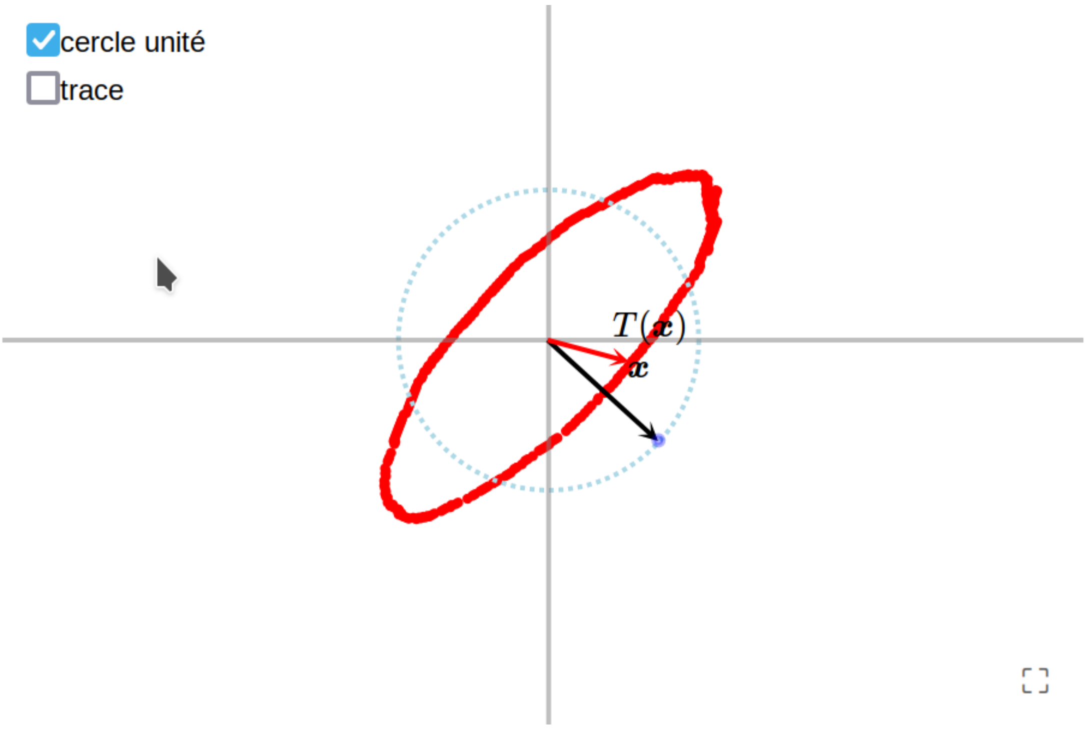
$$\begin{aligned}\mathcal{S} &:= \{\mathbf{x} \in \mathbb{R}^n \,:\, \|\mathbf{x}\| = 1\} \\ &= \{\mathbf{x} \in \mathbb{R}^n \,:\, x_1^2 + \cdots + x_n^2 = 1\}\,,\end{aligned}$$

sous l'action de T ? (En $d = 2$, $\mathcal{S}$ est le cercle de rayon 1 centré à l'origine.)

b) Parmi les vecteurs $\mathbf{x}$ situés sur cette sphère, quels sont ceux qui subissent une élongation maximale/minimale, à savoir ceux pour lesquels $\|A\mathbf{x}\|$ est maximal/minimal ?

Ces deux questions pourront être étudiées simultanément.

Exemple 15.15. Sur l'animation suivante, on observe que l'application linéaire $T : \mathbb{R}^2 \to \mathbb{R}^2$ donnée transforme le cercle $\mathcal{S}$ en *ellipse*. Les axes de cette ellipse doivent donner les directions d'élongation maximale (grand axe) et minimale (petit axe) :



◇

Soit $A = U\Sigma V^T$ une décomposition en valeurs singulières de A, dans laquel on suppose, comme précédemment, que les valeurs singulières sur la diagonale de Σ sont arrangées en ordre décroissant :

$$\sigma_1 \geqslant \sigma_2 \geqslant \ldots$$

On rappelle qu'avec cet ordre, $\sigma_j = \sqrt{\lambda_j}$, où $\lambda_j > 0$ est la j-ème plus grande valeur propre de $A^T A$.

Proposition 16.

$$\max_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \max_k \sigma_k = \sigma_1 \,,$$
$$\min_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \min_k \sigma_k \,.$$

De plus,

- ⋆ *le maximum est réalisé lorsque* $\mathbf{x}$ *est un vecteur propre unitaire de* A^TA *associé à la plus grande valeur propre de* A^TA *(en l'occurence* λ_1*).*
- ⋆ *le minimum est réalisé lorsque* $\mathbf{x}$ *est un vecteur propre unitaire de* A^TA *associé à la plus petite valeur propre de* A^TA*.*

Nous utiliserons l'entier $r = \mathrm{rang}(A)$, qui implique comme on sait que
$\sigma_r > 0$, $\sigma_{r+1} = 0$.
Preuve: Par l'orthogonalité de U (qui implique $\|U\mathbf{z}\| = \|\mathbf{z}\|$ pour tout $\mathbf{z} \in \mathbb{R}^m$), on peut écrire

$$\|A\mathbf{x}\| = \|U\Sigma V^T\mathbf{x}\| = \|\Sigma V^T\mathbf{x}\| \qquad \forall \mathbf{x} \in \mathbb{R}^n \,.$$

On a donc

$$\begin{aligned}\max_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| &= \max_{\mathbf{x}\in\mathcal{S}} \|\Sigma V^T\mathbf{x}\| \\ &= \max_{\mathbf{y}\in\mathcal{S}} \|\Sigma\mathbf{y}\| \\ &= \max_{\mathbf{y}\in\mathcal{S}} \sqrt{\sigma_1^2y_1^2 + \cdots + \sigma_r^2y_r^2} \,.\end{aligned}$$

Dans la deuxième égalité, on a effectué le changement de variable $\mathbf{y} := V^T\mathbf{x}$ (l'orthogonalité de V^T implique que cette transformation est bijective, et que la condition $\|\mathbf{x}\| = 1$ est préservée puisque $\|V^T\mathbf{x}\| = \|\mathbf{x}\|$).
Ensuite, remarquons que si $\mathbf{y} \in \mathcal{S}$, alors

$$\begin{aligned}\sigma_1^2y_1^2 + \cdots + \sigma_r^2y_r^2 &\leqslant \sigma_1^2y_1^2 + \cdots + \sigma_1^2y_r^2 \\ &= \sigma_1^2(y_1^2 + \cdots + y_r^2) \\ &\leqslant \sigma_1^2\|\mathbf{y}\|^2 \\ &= \sigma_1^2 \,.\end{aligned}$$

Ensuite, soit $\mathbf{z} \in \mathbb{R}^n$ le vecteur qui a toutes ses composantes nulles sauf la première, qui vaut 1. Alors $\mathbf{z} \in \mathcal{S}$, et donc

$$\begin{aligned}\max_{\mathbf{y}\in\mathcal{S}}(\sigma_1^2y_1^2 + \cdots + \sigma_r^2y_r^2) &\geqslant (\sigma_1^2z_1^2 + \cdots + \sigma_r^2z_r^2) \\ &= \sigma_1^2\|\mathbf{z}\|^2 \\ &= \sigma_1^2 \,.\end{aligned}$$

Ceci montre que $\max_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \sigma_1$. Ensuite, on a déjà fait plusieurs fois ce calcul : si $\mathbf{v}_1$ est le vecteur propre unitaire de A^TA associé à λ_1, alors

$$\begin{aligned}\|A\mathbf{v}_1\|^2 &= (A\mathbf{v}_1)\cdot(A\mathbf{v}_1) \\ &= \mathbf{v}_1\cdot(A^TA\mathbf{v}_1) \\ &= \mathbf{v}_1\cdot(\lambda_1\mathbf{v}_1) \\ &= \lambda_1\|\mathbf{v}_1\|^2 \\ &= \lambda_1 \,,\end{aligned}$$

ce qui montre que

$$\max_{\mathbf{x}\in\mathcal{S}} \|A\mathbf{x}\| = \sigma_1 = \sqrt{\lambda_1} = \|A\mathbf{v}_1\| \,.$$

□

Exemple 15.16. On a déjà rencontré la matrice

$$A = \begin{pmatrix} \frac{3}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ \frac{9}{10\sqrt{2}} & \frac{13}{10\sqrt{2}} \end{pmatrix} ,$$

qui possède comme valeurs singulières $\sigma_1 = \frac{3}{2}$ et $\sigma_2 = \frac{1}{2}$. Par le théorème ci-dessus, les vecteur du cercle unité qui subissent l'élongation maximale (d'amplitude $\frac{3}{2}$) sous l'action de A sont

$$\pm \mathbf{v}_1 = \pm \begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix} ,$$

dont l'image est

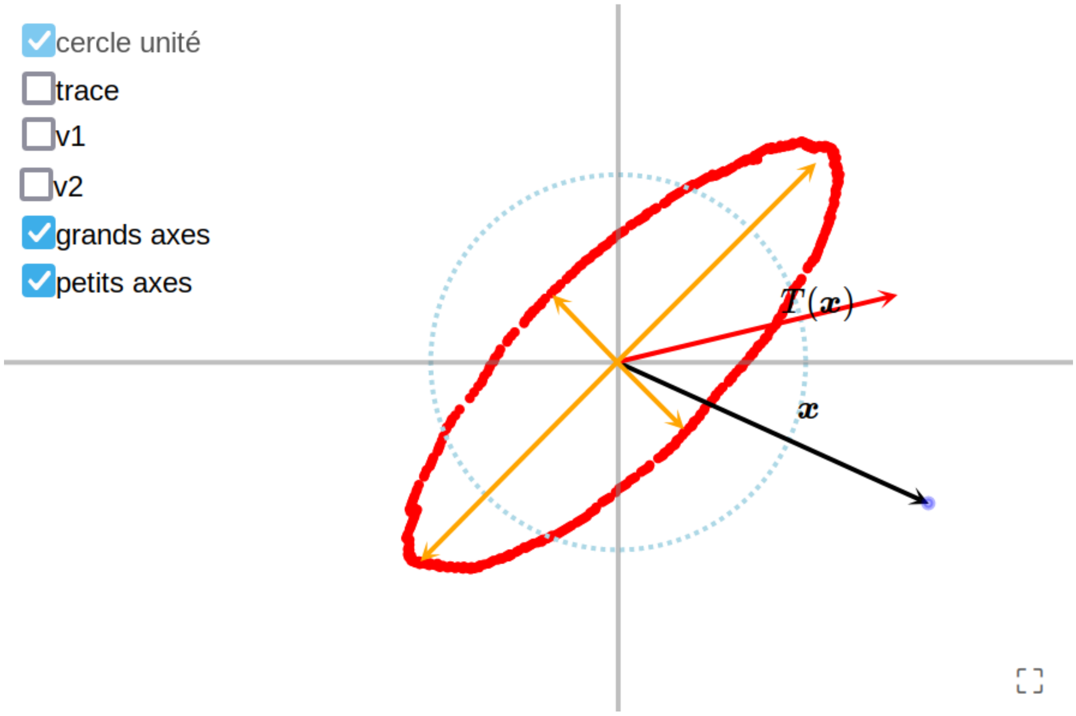
$$\pm A\mathbf{v}_1 = \pm \sigma_1 \mathbf{u}_1 = \pm \frac{3}{2} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} ,$$

et les vecteur du cercle unité qui subissent l'élongation minimale (d'amplitude $\frac{1}{2}$) sous l'action de A sont

$$\pm \mathbf{v}_2 = \pm \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{1} \end{pmatrix} ,$$

dont l'image est

$$\pm A\mathbf{v}_2 = \pm \sigma_2 \mathbf{u}_2 = \pm \frac{1}{2} \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} ,$$



◇