TD 01 : Décomposition en valeurs singulières

Fondements du Machine Learning, L3 IM2D

Septembre 2024

Exercice 1.1: Valeurs propres et valeurs singulières

Soit $\boldsymbol{U\Sigma V}^{\mathrm{T}}$ une *SVD* de $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ avec $\boldsymbol{U} = [\boldsymbol{u}_1 \dots \boldsymbol{u}_m]$, $\boldsymbol{V} = [\boldsymbol{v}_1 \dots \boldsymbol{v}_n]$, et $\sigma_i = [\boldsymbol{\Sigma}]_{ii}$ pour tout $1 \leq i \leq \min(m, n)$.

a) Montrer que $\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$ et $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i$. pour tout $1 \leq i \leq \min(m, n)$.

b) En déduire un lien entre les valeurs propres de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ et les valeurs singulières de $\boldsymbol{A}$.

c) Si $\mathrm{rang}(\boldsymbol{A}) = r$, que peut-on dire des valeurs singulières ?

Exercice 1.2: Carrés de matrices

Soit $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ une matrice définie positive (donc symétrique). Soit $\boldsymbol{P\Lambda P}^{T}$ une décomposition en valeurs propres de $\boldsymbol{A}$, avec $\boldsymbol{P}$ orthogonale et $\boldsymbol{\Lambda}$ diagonale à coefficients diagonaux ordonnés par ordre croissant $\boldsymbol{\Lambda}_{11} \geq \boldsymbol{\Lambda}_{22} \geq \cdots \geq \boldsymbol{\Lambda}_{nn}$.

a) Donner une décomposition en valeurs singulières de $\boldsymbol{A}$ basée sur la décomposition en valeurs propres ci-dessus.

b) Soit $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ telle que $\boldsymbol{B}^2 = \boldsymbol{A}$ et $\boldsymbol{B} \succ 0$. Donner une décomposition en valeurs propres et une décomposition en valeurs singulières de $\boldsymbol{B}$ en fonction de celles de $\boldsymbol{A}$.

c) Que vaut $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$? Retrouver alors le lien entre les valeurs propres de $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ et les valeurs singulières de $\boldsymbol{B}$.

Exercice 1.3: SVD et factorisation QR

On appelle *factorisation QR* d'une matrice $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ toute décomposition de la forme $\boldsymbol{X} = \boldsymbol{QR}$, où $\boldsymbol{Q} \in \mathbb{R}^{m\times m}$ est orthogonale et $\boldsymbol{R}$ est triangulaire supérieure, c'est-à-dire que $[\boldsymbol{R}]_{ij} = 0$ si $i > j$.

a) Pour toute matrice orthogonale $\boldsymbol{C} \in \mathbb{R}^{m\times m}$, donner une décomposition en valeurs singulières de $\boldsymbol{CX}$ en fonction de la SVD de $\boldsymbol{X}$.

b) En déduire une décomposition en valeurs singulières de $\boldsymbol{R}$ en fonction de la SVD de $\boldsymbol{X}$ de la question a).

Exercice 1.4: Normes et valeurs singulières

On définit sur $\mathbb{R}^{m\times n}$ la norme matricielle induite et la norme de Frobenius :

$$\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, \begin{cases} \|\boldsymbol{A}\| & := & \max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq \boldsymbol{0}}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\ \|\boldsymbol{A}\|_F & := & \sqrt{\sum_{\substack{1\leq i\leq m \\ 1\leq j\leq n}} [\boldsymbol{A}]_{ij}^2}. \end{cases}$$

a) En utilisant une décomposition en valeurs singulières de $\boldsymbol{A}$, montrer que $\|\boldsymbol{A}\| = \sigma_1$, où σ_1 désigne la plus grande valeur singulière de la matrice A.
Indication: Utiliser le fait que si $\boldsymbol{Q} \in \mathbb{R}^{\ell\times\ell}$ est une matrice orthogonale, on a $\|\boldsymbol{Q}\boldsymbol{x}\| = \|\boldsymbol{x}\|$ pour tout $\boldsymbol{x} \in \mathbb{R}^\ell$.

b) Si $\boldsymbol{A}$ est carrée et symétrique, que vaut $\|\boldsymbol{A}\|$?

c) Avec les notations de la question a), montrer également que

$$\|\boldsymbol{A}\|_F = \sqrt{\sum_{i=1}^{\min(n,m)} \sigma_i^2}$$

où $\sigma_1 \geq \cdots \geq \sigma_{\min(n,m)}$ sont les valeurs singulières de $\boldsymbol{A}$.

Exercice 1.5: Preuve constructive de la SVD

Soit une matrice $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ non nulle.

a) En utilisant la définition de $\|\boldsymbol{X}\|$, trouver $\boldsymbol{v}_1 \in \mathbb{R}^n$ et $\boldsymbol{u}_1 \in \mathbb{R}^m$ tels que $\|\boldsymbol{u}_1\| = 1, \|\boldsymbol{v}_1\| = 1$ et $\boldsymbol{X}\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$, où $\sigma_1 = \|\boldsymbol{X}\|$.

b) Prouver alors qu'il existe $\boldsymbol{U}_1$ et $\boldsymbol{V}_1$ telles que

$$\boldsymbol{X} = \boldsymbol{U}_1\boldsymbol{X}_1\boldsymbol{V}_1^{\mathrm{T}}, \qquad \boldsymbol{X}_1 = \left[\begin{array}{c|c} \sigma_1 & \boldsymbol{w}^{\mathrm{T}} \\ \hline 0 & \boldsymbol{Y}_1 \end{array}\right].$$

en précisant les dimensions de $\boldsymbol{U}_1, \boldsymbol{V}_1, \boldsymbol{w}, \boldsymbol{Y}_1$.

c) Montrer que $\left\|\boldsymbol{X}_1 \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix}\right\| \geq (\sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w})$. En utilisant la valeur de $\|\boldsymbol{X}_1\|$, en déduire que $\boldsymbol{w} = \boldsymbol{0}$.

d) Conclure.

Exercice 1.6: SVD tronquée

Soit $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ de rang r, $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ une *SVD* réduite de $\boldsymbol{X}$ avec $\sigma_1 > \cdots > \sigma_r > 0$.

a) On considère la *SVD* tronquée $\boldsymbol{X}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_{k,k}\boldsymbol{V}_k^{\mathrm{T}}$. Evaluer $\|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2$.

b) Application : Supposons que $\sigma_i^2 = \frac{1}{2^i}$ $\forall i = 1, \ldots, m$. Pour tout $\epsilon > 0$, déterminer le plus petit k tel que $\|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 \leq \epsilon^2$.

Solutions

Solution de l'exercice 1.1: Valeurs propres et valeurs singulières

a) Comme $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$ et $\boldsymbol{V}$ est orthogonale, on a $\boldsymbol{AV} = \boldsymbol{U\Sigma}$. Par conséquent, pour tout i entre 1 et $\min(m,n)$, $\boldsymbol{Av}_i$ est la i-ème colonne de $\boldsymbol{AV}$, et donc de $\boldsymbol{U\Sigma}$. Cette i-ème colonne est obtenue par le produit

$$[\boldsymbol{u}_1 \cdots \boldsymbol{u}_i \cdots \boldsymbol{u}_m] \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_i \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_i[\boldsymbol{u}_i]_1 \\ \vdots \\ \vdots \\ \vdots \\ \sigma_i[\boldsymbol{u}_i]_m. \end{bmatrix} = \sigma_i \boldsymbol{u}_i,$$

d'où le résultat. La seconde propriété $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i$ s'obtient de manière identique en utilisant $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{V\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}$.

b) Pour tout $i \leq \min(m,n)$, on a

$$\boldsymbol{A}^\top \boldsymbol{A v}_i = \sigma_i \boldsymbol{A}^\top \boldsymbol{u}_i = \sigma_i^2 \boldsymbol{v}_i,$$

d'où l'on obtient que σ_i^2 est une valeur propre de $\boldsymbol{A}^\top\boldsymbol{A}$ (et $\boldsymbol{v}_i$ un vecteur propre associé). Comme les $\boldsymbol{v}_i$ forment une base de l'espace $\mathbb{R}^n$ ($\boldsymbol{V}$ est orthogonale), on peut conclure que l'ensemble des $\{\sigma_i^2\}_i$ est *exactement* l'ensemble des valeurs propres de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

c) D'après les résultats précédents, si $\mathrm{rang}(\boldsymbol{A}) = \mathrm{rang}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = r$, alors il y a exactement r valeurs singulières de $\boldsymbol{A}$ non nulles.

Solution de l'exercice 1.2: Carrés de matrices

a) On a $\boldsymbol{A} = \boldsymbol{P\Lambda P}^{\mathrm{T}}$. Une décomposition en valeurs singulières de $\boldsymbol{A}$ est de la forme $\boldsymbol{U\Sigma V}^{\mathrm{T}}$, où $\boldsymbol{U}$ et $\boldsymbol{V}$ doivent être des matrices orthogonales de taille $n \times n$, et $\boldsymbol{\Sigma}$ doit être de taille $n \times n$ diagonale (car carrée ici). Pour obtenir une décomposition qui convienne à cette définition, il suffit de poser :

$$\boldsymbol{U} = \boldsymbol{P}, \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}, \quad \boldsymbol{V} = \boldsymbol{P}.$$

b) La matrice $\boldsymbol{B}$ est définie positive, donc symétrique. Elle admet donc une décomposition en valeurs propres, qui s'écrit $\boldsymbol{QDQ}^{\mathrm{T}}$, où $\boldsymbol{Q}$ est orthogonale et $\boldsymbol{D}$ diagonale. Comme $\boldsymbol{B}^2 = \boldsymbol{A}$, on a alors :

$$\boldsymbol{B}^2 = \boldsymbol{QDQ}^{\mathrm{T}}\boldsymbol{QDQ}^{\mathrm{T}} = \boldsymbol{QD}^2\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{P\Lambda P}^{\mathrm{T}} = \boldsymbol{A}.$$

En posant $\boldsymbol{Q} = \boldsymbol{P}$ et $\boldsymbol{D} = \boldsymbol{\Lambda}^{1/2}$ (matrice diagonale dont les coefficients diagonaux sont $\sqrt{\boldsymbol{\Lambda}_{11}}$, $\sqrt{\boldsymbol{\Lambda}_{22}}, \ldots, \sqrt{\boldsymbol{\Lambda}_{nn}}$ dans cet ordre), on obtient bien le résultat.

Par le même raisonnement qu'à la question a), on en déduit que la décomposition $\boldsymbol{P\Lambda}^{1/2}\boldsymbol{P}^{\mathrm{T}}$ est également une décomposition en valeurs singulières de $\boldsymbol{B}$.

c) En utilisant la décomposition de $\boldsymbol{B}$ trouvée dans la question précédente, on a :

$$\begin{aligned}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} &= (\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}} \\ &= \boldsymbol{Q}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}} \\ &= \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}} \\ &= \boldsymbol{Q}\boldsymbol{D}^2\boldsymbol{Q}^{\mathrm{T}} \\ &= \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{\mathrm{T}}.\end{aligned}$$

On a donc $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{A}$: les valeurs propres de $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ sont donc celles de $\boldsymbol{A}$. Celles-ci sont les carrés des valeurs propres de $\boldsymbol{B}$, qui sont aussi les valeurs singulières de $\boldsymbol{B}$. On retrouve donc sur cet exemple la propriéte que les valeurs propres de $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$ sont les carrés des valeurs singulières de $\boldsymbol{B}$, qui est vraie pour toute matrice réelle.

Solution de l'exercice 1.3: SVD et factorisation QR

a) Une décomposition valide en valeurs singulières de $\boldsymbol{C}\boldsymbol{X}$ est $\boldsymbol{U}_C\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ avec $\boldsymbol{U}_C = \boldsymbol{C}\boldsymbol{U} \in \mathbb{R}^{m\times m}$, qui est bien une matrice orthogonale.

b) On utilise la relation $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R} \Leftrightarrow \boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{X}$ (qui est valide car $\boldsymbol{Q}$ est orthogonale par hypothèse). En appliquant le résultat de la question a) avec $\boldsymbol{C} = \boldsymbol{Q}^{\mathrm{T}}$, on obtient $\boldsymbol{R} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U})\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$, où les matrices $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{U}$, $\boldsymbol{\Sigma}$ et $\boldsymbol{V}$ forment les facteurs de la décomposition en valeurs singulières.

Solution de l'exercice 1.4: Normes et valeurs singulières

a) Soit $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ une décomposition en valeurs singulières de $\boldsymbol{A}$, avec $\boldsymbol{\Sigma}$ à coefficients diagonaux $\sigma_1 \geq \cdots \geq \sigma_{\min(m,n)}$.

Le raisonnement se base sur les propriétés des matrices $\boldsymbol{U}$ et $\boldsymbol{V}^{\mathrm{T}}$. La matrice $\boldsymbol{U}$ est orthogonale par définition de la *SVD* : on a donc $\|\boldsymbol{U}\boldsymbol{y}\| = \|\boldsymbol{y}\|$ pour tout $\boldsymbol{y} \in \mathbb{R}^m$. La matrice $\boldsymbol{V}$ est également orthogonale, ce qui signifie que $\boldsymbol{V}^T = \boldsymbol{V}^{-1}$ l'est aussi. On a donc $\|\boldsymbol{x}\| = \|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|$ pour tout $\boldsymbol{x} \in \mathbb{R}^n$; de plus, l'inversibilité de $\boldsymbol{V}^T$ garantit que

$$\left\{ \boldsymbol{z} \in \mathbb{R}^n \;\middle|\; \boldsymbol{z} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{x},\ \boldsymbol{x} \in \mathbb{R}^n \right\} = \mathbb{R}^n.$$

En conséquence,

$$\begin{aligned}\max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq 0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} &= \max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq 0}} \frac{\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\ &= \max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq 0}} \frac{\|\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\ &= \max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq 0}} \frac{\|\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|}{\|\boldsymbol{V}^{\mathrm{T}}\boldsymbol{x}\|} \\ &= \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n \\ \boldsymbol{z}\neq 0}} \frac{\|\boldsymbol{\Sigma}\boldsymbol{z}\|}{\|\boldsymbol{z}\|}.\end{aligned}$$

On en déduit donc que $\|\boldsymbol{A}\| = \|\boldsymbol{\Sigma}\|$. Le calcul explicite de cette seconde norme est possible, car on a :

$$\begin{aligned}
\max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \boldsymbol{z}\neq 0}} \frac{\|\boldsymbol{\Sigma z}\|}{\|\boldsymbol{z}\|} &= \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \|\boldsymbol{\Sigma z}\| \\
&= \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \left\| \begin{matrix} \sigma_1 z_1 \\ \vdots \\ \sigma_{\min(m,n)} z_{\min(m,n)} \\ 0 \\ \vdots \\ 0 \end{matrix} \right\| \\
&= \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2 z_i^2} \\
&\leq \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_1^2 z_i^2} \\
&= \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \sigma_1 \sqrt{\sum_{i=1}^{\min(m,n)} z_i^2} \\
&\leq \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n\\ \|\boldsymbol{z}\|=1}} \sigma_1 \|\boldsymbol{z}\| = \sigma_1.
\end{aligned}$$

Il suffit ensuite de noter que le choix $\boldsymbol{z} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ vérifie $\|\Sigma \boldsymbol{z}\| = \sigma_1$ pour justifier que cette inégalité est en fait une égalité.

On a donc prouvé que $\|\boldsymbol{A}\| = \|\boldsymbol{\Sigma}\| = \sigma_1$.

b) Si $\boldsymbol{A}$ est carrée diagonalisable, on a $\boldsymbol{A} = \boldsymbol{P\Lambda P}^T$ avec $\boldsymbol{P}$ orthogonale et $\boldsymbol{\Lambda}$ est diagonale. On a alors $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{P\Lambda}^2\boldsymbol{P}^T$, et on peut donc définir $\lambda_1^2 \geq \cdots \geq \lambda_n^2$ comme les valeurs propres de cette matrice. On sait que les valeurs singulières de $\boldsymbol{A}$ sont les racines carrées des valeurs propres de $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, et on a donc $\sigma_1 = |\lambda_1|$. On retrouve ainsi le fait que la norme matricielle $\|\cdot\|$ est égale à la plus grande valeur propre (en valeur absolue), ce qu'on appelle également le rayon spectral.

c) En utilisant la même décomposition qu'en question a), et les propriétés d'orthogonalité de $\boldsymbol{U}$ et $\boldsymbol{V}$, on a $\|\boldsymbol{Ux}\| = \|\boldsymbol{x}\|$ pour tout $\boldsymbol{x} \in \mathbb{R}^m$, et $\|\boldsymbol{zV}^{\mathrm{T}}\| = \|\boldsymbol{z}\|$ pour tout $\boldsymbol{z} \in \mathbb{R}^n$. Or, pour toute matrice $\boldsymbol{M} \in \mathbb{R}^{m\times n}$, la définition de $\|\boldsymbol{M}\|_F$ donne

$$\|\boldsymbol{M}\|_F = \sqrt{\|\boldsymbol{m}_{1\bullet}\|^2 + \cdots + \|\boldsymbol{m}_{m\bullet}\|^2} = \sqrt{\|\boldsymbol{m}_{\bullet 1}\|^2 + \cdots + \|\boldsymbol{m}_{\bullet n}\|^2}$$

où $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{m}_{1\bullet}^{\mathrm{T}} \\ \vdots \\ \boldsymbol{m}_{m\bullet}^{\mathrm{T}} \end{bmatrix} = [\boldsymbol{m}_{\bullet 1} \cdots \boldsymbol{m}_{\bullet n}]$ (la norme de Frobenius d'une matrice au carré est égale à la somme des carrés des normes de ses colonnes *ou* de ses lignes). Par conséquent, on a aussi $\|\boldsymbol{UM}\|_F = \|\boldsymbol{MV}^{\mathrm{T}}\|_F = \|\boldsymbol{M}\|_F$ si $\boldsymbol{U}$ et $\boldsymbol{V}$ sont orthogonales.

On obtient donc :

$$\|\boldsymbol{A}\|_F = \|\boldsymbol{U\Sigma V}^{\mathrm{T}}\|_F = \|\boldsymbol{\Sigma}\|_F = \sqrt{\sum_{i=1}^{\min(m,n)} \sigma_i^2}.$$

Solution de l'exercice 1.5: Preuve constructive de la SVD

a) Par définition de $\|\boldsymbol{X}\| = \max_{\substack{\boldsymbol{z}\in\mathbb{R}^n \\ \|\boldsymbol{z}\|=1}} \|\boldsymbol{Xz}\|$, il existe un vecteur $\boldsymbol{v}_1 \in \mathbb{R}^n$ de norme 1 tel que $\|\boldsymbol{Xv}_1\| = \|\boldsymbol{X}\|$. En posant $\boldsymbol{u}_1 = \frac{\boldsymbol{Xv}_1}{\sigma_1}$ avec $\sigma_1 = \|\boldsymbol{X}\|$, on obtient bien les vecteurs souhaités.

b) Soient $\boldsymbol{U}_1$ et $\boldsymbol{V}_1$ des matrices orthogonales, respectivement dans $\mathbb{R}^{m\times m}$ et $\mathbb{R}^{n\times n}$, telles que $\boldsymbol{u}_1$ et $\boldsymbol{v}_1$ soient les premières colonnes des matrices $\boldsymbol{U}_1$ et $\boldsymbol{V}_1$, respectivement. Soient $\boldsymbol{u}_1^{(1)}, \ldots, \boldsymbol{u}_{m-1}^{(1)}$ et $\boldsymbol{v}_1^{(1)}, \ldots, \boldsymbol{v}_{n-1}^{(1)}$ les colonnes restantes de $\boldsymbol{U}_1$ et $\boldsymbol{V}_1$. On peut alors considérer la matrice $\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{XV}_1$, dont les coefficients sont donnés par :

$$\left[\begin{array}{c|c} \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Xv}_1 & [\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Xv}_j^{(1)}]_{j=1,\ldots,n-1} \\ \hline [(\boldsymbol{u}_i^{(1)})^{\mathrm{T}}\boldsymbol{Xv}_1]_{i=1,\ldots,m-1} & \left[(\boldsymbol{u}_i^{(1)})^{\mathrm{T}}\boldsymbol{Xv}_j^{(1)}\right]_{\substack{i=1,\ldots,m-1 \\ j=1,\ldots,n-1}} \end{array}\right].$$

Notre but est de montrer que cette matrice est de la forme :

$$\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{XV}_1 = \left[\begin{array}{c|c} \sigma_1 & \boldsymbol{w}^{\mathrm{T}} \\ \hline 0 & \boldsymbol{Y}_1 \end{array}\right],$$

où $\boldsymbol{w} \in \mathbb{R}^{n-1}$ et $\boldsymbol{Y}_1 \in \mathbb{R}^{(m-1)\times(n-1)}$.

On a $\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Xv}_1 = \sigma_1\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_1 = \sigma_1\|\boldsymbol{u}_1\|^2 = \sigma_1$ d'après la première question.

Par ailleurs, pour tout $j = 1, \ldots, n-1$, on a:

$$\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Xv}_j^{(1)} = \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_j^{(1)} = 0$$

car $\boldsymbol{V}_1$ est une matrice orthogonale.

En posant $w = [\boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{Xv}_j^{(1)}]_{j=1,\ldots,n-1}$ et $\boldsymbol{Y}_1 = \left[(\boldsymbol{u}_i^{(1)})^{\mathrm{T}}\boldsymbol{Xv}_j\right]_{\substack{i=1,\ldots,m-1 \\ j=1,\ldots,n-1}}$, on obtient alors la décomposition recherchée.

c) On a:

$$\boldsymbol{X}_1 \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix} = \left[\begin{array}{c|c} \sigma_1 & \boldsymbol{w}^{\mathrm{T}} \\ \hline 0 & \boldsymbol{Y}_1 \end{array}\right] \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \\ \boldsymbol{Y}_1\boldsymbol{w} \end{bmatrix}.$$

Par conséquent,

$$\left\|\boldsymbol{X}_1 \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix}\right\| = \left\|\begin{bmatrix} \sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \\ \boldsymbol{Y}_1\boldsymbol{w} \end{bmatrix}\right\| \geq \left\|\begin{bmatrix} \sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \\ 0 \end{bmatrix}\right\| \geq \sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}$$

où l'on a utilisé la définition de la norme d'un vecteur pour obtenir l'avant-dernière égalité. On remarque que $\sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = \left\| \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix} \right\|^2$.

Il s'agit maintenant d'établir que $\boldsymbol{w} = 0$. Pour cela, on utilise le fait que $\|\boldsymbol{X}_1\| = \|\boldsymbol{X}\| = \sigma_1$, car la norme ne change pas par multiplication avec une matrice orthogonale[1]. Cela signifie que

$$\begin{aligned} \sigma_1 &= \max_{\substack{\boldsymbol{x}\in\mathbb{R}^n \\ \boldsymbol{x}\neq 0}} \frac{\|\boldsymbol{X}_1\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \\ &\geq \frac{\left\| \boldsymbol{X}_1 \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix} \right\|}{\left\| \begin{bmatrix} \sigma_1 \\ \boldsymbol{w} \end{bmatrix} \right\|} \\ &\geq \frac{\sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}{\sqrt{\sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}} \\ &= \sqrt{\sigma_1^2 + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}}. \end{aligned}$$

Si $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} > 0$, alors on obtient $\sigma_1 > \sqrt{\sigma_1^2} = \sigma_1$, ce qui est absurde. On en déduit donc que $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} = \|\boldsymbol{w}\|^2 = 0$, donc que $\boldsymbol{w}$ est un vecteur nul.

d) Dans les questions précédentes, on a prouvé que l'on pouvait définir des matrices $\boldsymbol{U}_1$ et $\boldsymbol{V}_1$ telles que

$$\boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{X}\boldsymbol{V}_1 = \left[\begin{array}{c|c} \sigma_1 & 0 \\ \hline 0 & \boldsymbol{Y} \end{array}\right].$$

avec $\sigma_1 = \|\boldsymbol{X}\|$. On peut maintenant appliquer cette technique à $\boldsymbol{Y}$, et obtenir des matrices orthogonales $\boldsymbol{U}_2 \in \mathbb{R}^{(m-1)\times(m-1)}$ et $\boldsymbol{V}_2 \in \mathbb{R}^{(n-1)\times(n-1)}$ telles que

$$\boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{V}_2 = \left[\begin{array}{c|c} \sigma_2 & 0 \\ \hline 0 & \boldsymbol{Z} \end{array}\right].$$

où $\sigma_2 = \|\boldsymbol{Y}\|$ et $\boldsymbol{Z} \in \mathbb{R}^{(m-2)\times(n-2)}$. Soient $\boldsymbol{U}_{12} = [\boldsymbol{u}_1\ \boldsymbol{U}_2]\boldsymbol{U}_1$ et $\boldsymbol{V}_{12} = [\boldsymbol{v}_1\ \boldsymbol{V}_2]\boldsymbol{V}_1$: on peut vérifier que ces matrices sont bien orthogonales comme produits de matrices elles-mêmes orthogonales, et on a alors :

$$\boldsymbol{U}_{12}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{V}_{12} = \left[\begin{array}{c|c|c} \sigma_1 & 0 & 0 \\ \hline 0 & \sigma_2 & 0 \\ \hline 0 & 0 & \boldsymbol{Z} \end{array}\right].$$

En répétant cette opération autant de fois que nécessaire, on obtiendra donc une décomposition qui s'identifie à la décomposition en valeurs singulières.

Solution de l'exercice 1.6: SVD tronquée

a) On a :

$$\|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 = \left\| \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} - \boldsymbol{U} \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{k,k} & 0 \\ \hline 0 & 0 \end{array}\right] \boldsymbol{V}^{\mathrm{T}} \right\|_F^2.$$

[1] Note : contrairement à l'exercice 4, on a *défini* ici $\sigma_1 \stackrel{d}{=} \|\boldsymbol{X}\|$.

La norme de Frobenius est invariante par multiplication avec une matrice orthogonale; on a donc

$$\|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 = \left\| \boldsymbol{\Sigma} - \left[\begin{array}{c|c} \boldsymbol{\Sigma}_{k,k} & 0 \\ \hline 0 & 0 \end{array} \right] \right\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2.$$

b) D'après ce qui précède,

$$\begin{aligned} \|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 &= \sum_{i=k+1}^{r} \sigma_i^2 \\ &= \sum_{i=k+1}^{r} \frac{1}{2^i} \\ &= \frac{1}{2^{k+1}} \frac{1-(1/2)^{r-k}}{1-1/2} = \frac{2^{r-k}-1}{2^r}. \end{aligned}$$

On a donc $\|\boldsymbol{X} - \boldsymbol{X}_k\|_F^2 > \epsilon^2$ si

$$\frac{2^{r-k}-1}{2^r} > \epsilon^2 \ \Leftrightarrow k < r - \log_2\left(1 + 2^r \epsilon^2\right).$$

La valeur cherchée est donc $\left\lceil r - \log_2\left(1 + 2^r\epsilon^2\right)\right\rceil$.