ALGORITMOS EN ÁLGEBRA LINEAL

Notas de curso (UBA - 2do cuatrimestre de 2005)

http://atlas.mat.ub.es/personals/sombra/curso.html

Michelle Schatzman

Martín Sombra

Institut Camille Jordan (Mathématiques), Université de Lyon 1 ; 43 Bd. du 11 Novembre 1918, 69622 Villeurbanne Cedex, Francia

Universitat de Barcelona, Departament d'Àlgebra i Geometria ; Gran Via 585, 08007 Barcelona, España

CHAPITRE 1

Valeurs singulières des matrices

La plupart des choses dans ce chapitre se trouvent aussi dans le classique [**2**] ou dans le futur classique [**1**].

1. Rappel des résultats d'algèbre linéaire

Soit $\mathbb{K}$ le corps de base ; dans ce qui suit on considérera le corps des réels $\mathbb{K} = \mathbb{R}$ ou des complexes $\mathbb{K} = \mathbb{C}$. Cependant pour les algorithmes exacts on pourra (et voudra) considérer d'autres corps (les rationnels $\mathbb{Q}$, les corps finis $\mathbb{F}_q$).

Soit $f : V \to W$ un opérateur linéaire entre des $\mathbb{K}$-espaces vectoriels V et W de dimension n et m respectivement. Soient $\mathcal{B} = (v_1, \ldots, v_n)$ et $\mathcal{C} = (w_1, \ldots, w_m)$ des bases de V et de W respectivement. La *matrice de f dans les bases $\mathcal{B}$ et $\mathcal{C}$* est la matrice

$$A = [f]_{\mathcal{B},\mathcal{C}} \in \mathbb{K}^{m\times n}$$

dont les colonnes sont l'image des vecteurs de $\mathcal{B}$ par rapport à la base $\mathcal{C}$: si $f(v_j) = \sum_{i=1}^{m} a_{i,j} w_i$ alors

$$\mathrm{col}_j(A) = \begin{bmatrix} a_{1,j} \\ \vdots \\ a_{m,j} \end{bmatrix}.$$

Typiquement $V = \mathbb{K}^n$ et $W = \mathbb{K}^m$ sont munis des bases standard. On identifie une forme linéaire $\ell : \mathbb{K}^n \to \mathbb{K}$ avec un vecteur *ligne* et un point $x \in \mathbb{R}^m$ avec un vecteur *colonne.*

1.1. Décomposition en blocs d'opérateurs et matrices.

La décomposition en blocs apparaît naturellement lors des discrétisations des EDPs en plus d'une variable. Considérons les décompositions

$$V = \oplus_{j=1}^{N} V_j \quad , \quad W = \oplus_{i=1}^{M} W_i.$$

Pour $1 \leq j \leq N$ et $1 \leq i \leq M$ posons

$$\begin{aligned} J_j : V_j \hookrightarrow V \quad &, \quad x_j \mapsto (0, \ldots, 0, x_j, 0, \ldots, 0), \\ Q_i : W \to W_i \quad &, \quad y = (y_1, \ldots, y_M) \mapsto y_i \end{aligned}$$

pour les inclusion et projection canoniques. L'opérateur $f_{i,j} = Q_i \circ f \circ J_j : V_j \to W_i$ est défini par le diagramme :

$$\begin{array}{ccc} V & \to^{f} & W \\ \uparrow J_j & & \downarrow Q_i \\ V_j & \dashrightarrow^{f_{i,j}} & W_j \end{array}$$

En d'autres termes, c'est la j-ème composante de la restriction de f au sous-espace V_j. La *décomposition en blocs* de f associée est

$$[f] = \begin{bmatrix} f_{1,1} & \cdots & f_{1,N} \\ \vdots & & \vdots \\ f_{M,1} & \cdots & f_{M,N} \end{bmatrix}$$

Soit $x = \sum_{j=1}^{N} x_j$, alors

$$f(x) = \left(\sum_{j=1}^{N} f_{1,j}(x_j), \ldots, \sum_{j=1}^{N} f_{M,j}(x_j) \right).$$

La multiplication respecte la structure en blocs : soit

$$X = \oplus_{h=1}^{L} X_h$$

et $g : W \to X$ un autre opérateur linéaire avec

$$[g] = \begin{bmatrix} g_{1,1} & \cdots & g_{1,M} \\ \vdots & & \vdots \\ g_{L,M} & \cdots & g_{L,M} \end{bmatrix}$$

la décomposition en blocs correspondante, alors

$$(g \circ f)_{h,j} = \sum_{i=1}^{M} g_{h,i} \circ f_{i,j}.$$

Par exemple

$$\begin{bmatrix} B_{1,1} & B_{1,2} \end{bmatrix} \begin{bmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \end{bmatrix} = \begin{bmatrix} B_{1,1}A_{1,1} + B_{1,2}A_{2,1} & B_{1,1}A_{1,2} + B_{1,2}A_{2,2} & B_{1,1}A_{1,3} + B_{1,2}A_{2,3} \end{bmatrix}.$$

1.2. Produits scalaires, adjonction, opérateurs hermitiens. Le *produit scalaire canonique* est

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j y_j = x^T y \quad , \quad \text{pour } x, y \in \mathbb{R}^n$$

$$\langle x, y \rangle = \sum_{j=1}^{n} x_j \overline{y_j} = x^T \overline{y} \quad , \quad \text{pour } x, y \in \mathbb{C}^n.$$

Soit $f : \mathbb{K}^n \to \mathbb{K}^m$, l'*opérateur adjoint* est l'unique $f^* : \mathbb{K}^m \to \mathbb{K}^n$ tel que

$$\langle x, f^*(y) \rangle_{\mathbb{K}^n} = \langle f(x), y \rangle_{\mathbb{K}^m}$$

pour tout $x \in \mathbb{K}^n$ et $y \in \mathbb{K}^m$. Si $A = [f] \in \mathbb{K}^{m \times n}$ est la matrice de f, alors

$$[f^*] = A^* := \overline{A}^T \in \mathbb{K}^{n \times m};$$

c'est-à-dire $(A^*)_{j,i} = \overline{A_{i,j}}$.

Une matrice carrée $A \in \mathbb{K}^{n \times n}$ est *hermitienne* ou *auto-adjointe* si $A^* = A$. Le résultat fondamental est que toute matrice hermitienne est diagonalisable dans une base unitaire, et ses valeurs propres sont réels. Explicitement

$$A = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^*$$

avec $\lambda_j \in \mathbb{R}$ et $U \in \mathbb{K}^{n \times n}$ telle que $U^* U = \mathbf{1}_n$ (c'est-à-dire $U \in O(n)$ si $\mathbb{K} = \mathbb{R}$ et $U \in U(n)$ si $\mathbb{K} = \mathbb{C}$).

Une matrice hermitienne est *positive* (resp. *semi-positive*) si tous ses valeurs propres sont positifs (resp. positifs ou nuls).

1.3. Normes de vecteurs, opérateurs et matrices. Les normes sont utilisées pour mesurer les erreurs dans les calculs matriciels ; on a donc besoin d'apprendre à les manipuler.

Soit V un $\mathbb{K}$-espace vectoriel. Une *norme* sur V est une fonction $|\cdot| : V \to \mathbb{R}$ satisfaisant

(1) $|x| \geq 0$ pour tout $x \in V$, et $|x| = 0$ si et seulement si $x = 0$;

(2) $|\lambda \cdot x| = |\lambda| \cdot |x|$ pour $\lambda \in \mathbb{K}$ et $x \in V$;

(3) inégalité du triangle : $|x + y| \leq |x| + |y|$ pour $x, y \in V$.

Exemples typiques : soit $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$; la *norme* ℓ^p est

$$|x|_p = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \qquad \text{pour } 1 \leq p < \infty;$$

$$|x|_\infty = \max_{1 \leq j \leq n} |x_j|, \qquad \text{pour } p = \infty.$$

Une autre famille d'exemples : norme indexée par une matrice hermitienne définie positive :

$$|x|_A := (x^T A \overline{x})^{1/2}.$$

Toutes les normes sont équivalentes sur $\mathbb{R}^n$ ou sur $\mathbb{C}^d$. Mais les constantes dépendent de la dimension ! :

$$\begin{aligned} |x|_2 &\leq |x|_1 \leq \sqrt{n}|x|_2, \\ |x|_\infty &\leq |x|_2 \leq \sqrt{n}|x|_\infty, \\ |x|_\infty &\leq |x|_1 \leq n|x|_\infty. \end{aligned}$$

Soit $f : V \to W$ un opérateur linéaire entre des espaces munis des normes N et M respectivement. La *norme d'opérateurs* subordonnée à N et M est

$$||f||_{N,M} = \max_{x \in V \setminus \{0\}} \frac{M(f(x))}{N(x)} = \max_{N(x)=1} M(f(x)).$$

En particulier, c'est une norme sur l'espace vectoriel $\mathrm{Hom}_{\mathbb{K}}(\mathbb{K}^n, \mathbb{K}^m)$. Par exemple, pour $A \in \mathbb{K}^{m \times n}$ on pose

$$\|A\|_{p,q} = \sup_{x \neq 0} \frac{\|Ax\|_q}{\|x\|_p},$$

$$\|A\|_p = \|A\|_{p,p}.$$

Une fonction norme $||\cdot|| : \mathbb{K}^{n \times n} \to \mathbb{R}$ est dite *matricielle* si pour tout $A, B \in \mathbb{K}^{n \times n}$ elle vérifie

$$||A \cdot B|| \leq ||A|| \cdot ||B||.$$

Toute norme d'opérateurs est matricielle, pourvu que $V = W$. Un autre exemple de norme matricielle est la *norme de Frobenius* définie par

$$\|A\|_F = (\mathrm{Tr}(A^* A))^{1/2} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

pour $A \in \mathbb{K}^{n \times n}$. La norme de Frobenius n'est pas subordonnée à une norme vectorielle. Ceci se voit facilement en remarquant que $||\mathbf{1}_n||_F = \sqrt{n}$ tandis que $||\mathbf{1}_n|| = 1$ pour toute norme d'opérateurs $||\cdot||$.

La norme qui est vraiment utile est la norme 2 (c'est-à-dire $||\cdot||_2$) mais elle est difficile à calculer. Par contre on peut mieux manipuler la norme de Frobenius, et elle nous fournit des estimations pour cette norme :

$$||A||_2 \leq ||A||_F \leq \sqrt{n}||A||_2.$$

Convention : si $m > n$, $\mathbb{K}^n$ s'injecte dans $\mathbb{K}^m$, et la norme $\|x\|_*$ de $x \in \mathbb{K}^n$ est égale à la norme de $\|X\|_*$ défini par

$$X = \begin{bmatrix} x \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

La convention est symétrique si $m < n$. Dans le cas $m > n$, on pose

$$\mathbf{1}_{m\times n} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

et dans le cas $m \leq n$, on pose

$$\mathbf{1}_{m\times n} = \begin{bmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}.$$

Alors $\|\mathbf{1}_{m\times n}\|_F = \sqrt{\min(m,n)}$ et

$$\sup_{x\neq 0} \frac{\|\mathbf{1}_{m\times n}x\|_*}{\|x\|_*} = 1,$$

puisque $\mathbf{1}x$ s'identifie à X.

Voici quelques inégalités supplémentaires : soit $A \in \mathbb{K}^{m\times n}$, alors

$$\max_{\substack{1\leq i\leq m\\1\leq j\leq n}} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{\substack{1\leq i\leq m\\1\leq j\leq n}} |a_{ij}|, \tag{1a}$$

$$\|A\|_2 \leq \sqrt{\|A\|_1\|A\|_\infty}, \tag{1b}$$

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|, \tag{1c}$$

$$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|, \tag{1d}$$

$$n^{-1/2}\|A\|_\infty \leq \|A\|_2 \leq m^{1/2}\|A\|_\infty, \tag{1e}$$

$$m^{-1/2}\|A\|_1 \leq \|A\|_2 \leq n^{1/2}\|A\|_1. \tag{1f}$$

Exercice 1.1. ◁ Montrer les inégalités (1a) à (1f). ▷

2. Décomposition en valeurs singulières (DVS)

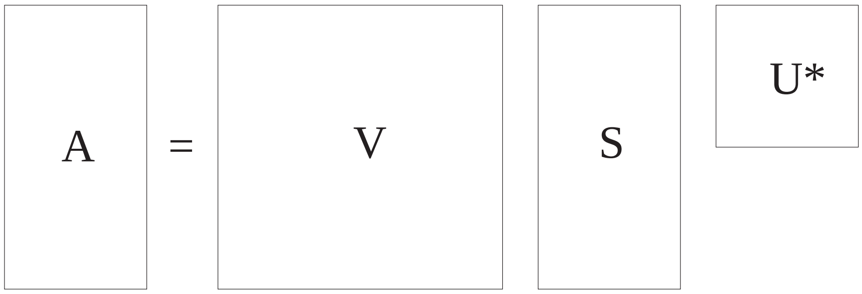
Théorème 2.1 (DVS). *Soit $A \in \mathbb{K}^{m\times n}$, alors il existe U matrice unitaire dans $\mathbb{K}^{n\times n}$, V matrice unitairee dans $\mathbb{K}^{m\times m}$ et $\ell = \min(m,n)$ nombres positifs $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_\ell \geq 0$ tels que*

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

avec $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_\ell) \in \mathbb{K}^{m\times n}$.

REMARQUE 2.2. Attention à l'abus de notations ! La matrice $\mathrm{diag}(\sigma_1, \ldots, \sigma_\ell)$ est une matrice à m lignes et n colonnes dont le bloc supérieur gauche $\ell \times \ell$ est la matrice diagonale dont les valeurs diagonales sont les σ_i.

Graphiquement



Si on pense à la matrice A comme un opérateur $\mathbb{K}^n \to \mathbb{K}^m$, ce résultat peut s'énoncer géométriquement comme que modulo des changements unitaires des coordonnées de $\mathbb{K}^n$ (les colonnes de U) et de $\mathbb{K}^m$ (les colonnes de V), l'opérateur devient diagonal ≥ 0.

DÉMONSTRATION. On supposera s.p.d.g. $m \geq n$ et on fait récurrence sur m et n. Pour $n = 1$ et m quelconque on pose

$$V = \frac{1}{||A||_2} A \quad , \quad \Sigma = ||A||_2 \quad , \quad U = 1.$$

Maintenant soit $m \geq n \geq 2$ et supposons le résultat vrai pour $m-1$ et $n-1$. Soit $u \in \mathbb{K}^n$ un vecteur unitaire tel que $\sigma_1 := |Au|_2 = ||A||_2$.

Si $A = 0$, le résultat est évident. On suppose donc $A \neq 0$, donc $\sigma_1 > 0$. Soient

$$\widetilde{U} \in \mathbb{K}^{n\times(n-1)} \quad , \quad \widetilde{V} \in \mathbb{K}^{m\times(m-1)}$$

des matrices telles que

$$\widehat{U} = \begin{bmatrix} u & \widetilde{U} \end{bmatrix} \in U(n) \quad , \quad \widehat{V} = \begin{bmatrix} v & \widetilde{V} \end{bmatrix} \in U(m)$$

soient des matrices unitaires, et calculons $\widehat{V}^* A \widehat{U}$:

$$\widehat{V}^* A \widehat{U} = \begin{bmatrix} v^* \\ \widetilde{v} \end{bmatrix} A \begin{bmatrix} u & \widetilde{U} \end{bmatrix} = \begin{bmatrix} v^*Au & v^*A\widetilde{U} \\ \widetilde{V}^*Au & \widetilde{V}^*A\widetilde{U} \end{bmatrix} = \begin{bmatrix} v^*Au & v^*A\widetilde{U} \\ \widetilde{V}^*Au & \widetilde{A} \end{bmatrix}.$$

On a $v^*Au = |Au|_2 v^*v = \sigma_1$ et $\widetilde{V}^*Au = \widetilde{V}^*v = 0$ puisque les colonnes de $\widetilde{V}$ sont orthogonales à v par construction.

Estimons la norme de $\widehat{V}^* A \widehat{U}$, de façon à montrer que $w := v^*A\widetilde{U} = 0$: d'une part :

$$\|\widehat{V}^* A \widehat{U}\|_2 = \|A\|_2 = \sigma_1$$

car $\widehat{V}$ et $\widehat{U}$ sont unitaires ; d'autre part,

$$\|\widehat{V}^* A \widehat{U}\|_2 = \|\widehat{U} A^* \widehat{V}^*\|_2 = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{|\widehat{U} A^* \widehat{V}^* x|_2}{|x|_2}.$$

Choix particulier de x :

$$x = \begin{bmatrix} \sigma_1 \\ w^* \end{bmatrix};$$

alors

$$\widehat{U}A^*\widehat{V}^*x = \begin{bmatrix} \sigma_1 x_1 + ww^* \\ \widetilde{A}w^* \end{bmatrix},$$

et donc

$$|x|_2 = (\sigma_1^2 + |w|_2^2)^{1/2} \quad , \quad |\widehat{U}A^*\widehat{V}^*x|_2 = \left(\sigma_1^2 + |w|_2^2 + |\widetilde{A}w^*|_2^2\right)^{1/2},$$

et par conséquent

$$\frac{|\widehat{U}A^*\widehat{V}^*x|_2}{|x|_2} \geq \frac{\left(\sigma_1^2 + |w|_2^2 + |\widetilde{A}w^*|_2^2\right)^{1/2}}{(\sigma_1^2 + |w|_2^2)^{1/2}} \geq (\sigma_1^2 + |w|_2^2)^{1/2}.$$

L'inégalité

$$\sigma_1 \geq (\sigma_1^2 + |w|_2^2)^{1/2}$$

implique immédiatement $w = 0$.

Finalement on applique l'hypothèse de récurrence :

$$\widetilde{A} := \widetilde{V}^* A \widetilde{U} = V_1 \Sigma_1 U_1^*$$

et donc

$$A = V \begin{bmatrix} \sigma_1 & \\ & \widetilde{A} \end{bmatrix} U^* = V \begin{bmatrix} 1 & \\ & V_1 \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \Sigma_1 \end{bmatrix} \begin{bmatrix} 1 & \\ & U_1^* \end{bmatrix} \widehat{U}^*;$$

le théorème est donc démontré. □

On remarque que

$$A^*A = (V\Sigma U^*)^*(V\Sigma U^*) = U\Sigma^2 U^*,$$

donc σ_i^2 est la i-ème valeur propre de A^*A. On peut démontrer facilement que $||A||_2 = \sigma_1$ et on en déduit

$$||A||_2 = \sqrt{\rho(A^*A)} = \sqrt{||A^*A||_2};$$

cette dernière égalité due à ce que $A^*A \geq 0$.

Avec la DVS on trouve une expression de A comme somme de matrices de rang 1 :

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^*.$$

Pour $0 \leq k \leq n$ posons

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^* \tag{2}$$

qui est donc une matrice de rang au plus k.

Soit $M_k(m,n) \subset \mathbb{C}^{m\times n}$ l'ensemble des matrices de rang au plus k. Ceci est une variété algébrique *déterminantale* (c'est-à-dire définie par l'annulation d'un nombre fini de déterminants) :

$$M_k(m,n) = Z\left(\det(B_{i_1,\dots,i_{k+1};j_1,\dots,j_{k+1}}) : 0 \leq i_1 < \cdots < i_{k+1} \leq m, 0 \leq j_1 < \cdots < j_{k+1} \leq n\}\right).$$

Théorème 2.3 (Théorème d'approximation). *La distance pour la norme 2 de A à l'ensemble des matrices $m \times n$ de rang au plus k est*

$$\mathrm{dist}_{||\cdot||_2}(A, M_k(m,n)) = ||A - A_k||_2 = \sigma_{k+1}.$$

DÉMONSTRATION. On a

$$A'_k = V \operatorname{diag}(\sigma_1, \ldots, \sigma_k, 0, \ldots, 0) U^*$$

et

$$\|A - A_k\|_2 = \|\operatorname{diag}(\underbrace{0, \ldots, 0}_{k \text{ zéros}}, \sigma_{k+1}, \ldots, \sigma_l)\|_2 = \sigma_{k+1}.$$

Réciproquement, supposons qu'il existe une matrice B de rang $\leq k$ telle que $\|A - B\|_2 < \sigma_k$. Soient $v_1, \ldots, v_{k+1}$ les $k+1$ premiers vecteurs colonne de la matrice V ; comme le noyau de B est au moins de dimension $n-k$, l'espace engendré par $v_1, \ldots, v_{k+1}$ est d'intersection non vide avec le noyau de B ; soit x un élément de norme 1 appartenant à cette intersection ; alors

$$\|A - B\|_2 \geq |Ax - Bx|_2 = |Ax|_2.$$

Si on pose

$$x = \sum_{j=1}^{k+1} x_j v_j,$$

alors

$$Ax = \sum_{j=1}^{k+1} x_j A v_j = \sum_{j=1}^{k+1} x_j \sigma_j u_j,$$

et donc

$$|Ax|_2 = \left(\sum_{j=1}^{k+1} \sigma_j^2 x_i^2 \right)^{1/2} \geq \sigma_{k+1} \, |x|_2 = \sigma_{k+1}.$$

On en conclut $\|A - B\|_2 \geq \sigma_{k+1}$. □

En particulier la distance pour la norme 2 de A à l'ensemble des systèmes mal conditionnés $M_{n-1}(m,n)$ est la dernière valeur singulière σ_n.

La DVS s'applique à la compression d'images. Une image est juste une matrice A dont le coefficient $a_{i,j}$ donne l'intensité du pixel (i,j). A la place des mn coefficients, on peut juste stocker A_k, qui selon le théorème d'approximation ci-dessus est la meilleure approximation de A dans $M_k(m,n)$, à partir de laquelle on peut reconstruire partiellement l'image en question. La matrice A_k est représentée par $(m+n)k$ coefficients (*voir* l'expression (2) ci-dessus) ; le taux de compression est

$$\frac{(m+n)k}{mn}.$$

Pour des exemples concrets, *voir* [**1**, § 3.2.3].

La DVS permet de calculer facilement l'*inverse de Moore-Penrose* d'une matrice rectangulaire $A \in \mathbb{K}^{m \times n}$ $(n \leq m)$ de rang maximal n :

$$A^+ := (A^* A)^{-1} A^* = U \Sigma^{-1} V^*.$$

PROPOSITION 2.4. *Soit $S^{n-1} = \{x \in \mathbb{R}^n : |x|_2 = 1\}$ la sphère unité de $\mathbb{R}^n$, alors $A(S^{n-1})$, l'image de cette sphère sous l'application A, est l'ellipsoïde de $\mathbb{R}^m$ centré en $\mathbf{0}$ d'axes principaux $\sigma_j v_j$ pour $j = 1, \ldots, m$.*

Par exemple pour

$$A = \begin{bmatrix} 1 & 2 \\ -2 & -1 \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \right),$$

l'ensemble $A(S^2)$ est l'ellipse d'axes principaux $3(-1/\sqrt{2}, 1/\sqrt{2})$ et $(1/\sqrt{2}, 1/\sqrt{2})$.

DÉMONSTRATION. Puisque U est unitaire on a $U^*(S^{n-1}) = S^{n-1}$. Ensuite, $\Sigma(S^{n-1})$ est l'ellipsoïde d'équation $\sum_{i=1}^m (w_i/\sigma_i)^2 = 1$, d'axes principaux $\sigma_i e_i$, où e_i est le i-ème vecteur de la base standard de $\mathbb{R}^m$. Finalement la multiplication par V a l'effet de changer les e_is en les v_is. $\square$

Les origines de la DVS sont expliquées dans l'article [**3**]. La DVS (théorème 2.1) fut indépendamment découverte par Eugenio Beltrami en 1873, et par Camille Jordan un an après. Le théorème 2.3 d'approximation est du à Erhard Schmid (celui du procédé d'orthogonalisation de Gram-Schmid et un étudiant de Hilbert) en 1907 et peut se considérer comme le résultat fondamental de la DVS. Ce résultat est souvent (et à tort) appelé théorème de Eckart-Young, qui l'ont redécouvert 29 ans après.

2.1. Pour ceux qui connaissent un peu d'analyse fonctionnelle. On note $\mathcal{L}(H)$ l'ensemble des opérateurs bornés d'un espace de Hilbert H dans lui-même. Les valeurs singulières d'un opérateur $A \in \mathcal{L}(H)$ sont les racines carrées des valeurs propres de A^*A, rangées par ordre décroissant :

$$\sigma_1(A) \geq \sigma_2(A) \geq \cdots$$

On rappelle que si $B \in \mathcal{L}(H)$ est un opérateur autoadjoint ≥ 0 il possède une unique racine carrée autoadjointe ≥ 0, qui sera notée $\sqrt{B}$. La norme de $x \in H$ est notée $|x|$ et la norme d'opérateur subordonnée à la norme de H est notée $\|A\| = \sup\{|Ax| : |x| \leq 1\}$.

EXERCICE 1.2. $\triangleleft$ Soit $A \in \mathcal{L}(H)$. Montrer qu'on a

$$\sigma_{n+1}(A) = \min\{\|A\,|_{E^\perp}\| : \ E \text{ de dimension } n\}.$$

Montrer que ce minimum est atteint en prenant comme E l'espace engendré par les n plus grandes valeurs propres comptées avec leur multiplicité de $|A| = \sqrt{A^*A}$.

Indication : utiliser la caractérisation maximin des valeurs propres d'un opérateur autoadjoint borné B dans un espace de Hilbert, dont les valeurs propres sont notées $\lambda_1(B) \geq \lambda_2(B) \geq \cdots$ avec leur multiplicité :

$$\lambda_k(B) = \min\left\{\max\left\{\frac{|Bx|}{|x|} : x \in V\right\}, \dim V = k\right\}.$$

$\triangleright$

EXERCICE 1.3. $\triangleleft$ Montrer qu'on a également

$$\sigma_{n+1}(A) = \inf\{\|A - X\| : X \in R_n\}, \tag{3}$$

où R_n désigne l'ensemble des opérateurs de $\mathcal{L}(H)$ de rang au plus n. $\triangleright$

EXERCICE 1.4. $\triangleleft$ Déduire de la caractérisation (3) que, pour tout n, pour tous A et B dans $\mathcal{L}(H)$:

$$|\sigma_n(A) - \sigma_n(B)| \leq \|A - B\|.$$

$\triangleright$

EXERCICE 1.5. $\triangleleft$ Déduire de l'inclusion $R_n + R_m \subset R_{m+n}$ l'inégalité

$$\sigma_{m+n+1}(A + B) \leq \sigma_{m+1}(A) + \sigma_{n+1}(B).$$

$\triangleright$

EXERCICE 1.6. $\triangleleft$ Montrer également qu'on a l'inégalité

$$\sigma_{m+n+1}(AB) \leq \sigma_{m+1}(A)\sigma_{n+1(B)},$$

et en déduire que pour tout n

$$\sigma_n(AB) \leq \sigma_n(A)\|B\|, \quad \sigma_n(AB) \leq \sigma_n(B)\|A\|.$$

Indication : utiliser $X = AX_2 + X_1B - X_1X_2$, avec $X_1 \in R_m$ et $X_2 \in R_n$. ▷

EXERCICE 1.7. ◁Montrer qu'un opérateur A est compact si et seulement si

$$\lim_{n\to\infty} \sigma_n(A) = 0.$$

▷

EXERCICE 1.8. ◁ Montrer que l'ensemble $\mathcal{K}$ des opérateurs compacts de H dans lui-même est un idéal bilatère de $\mathcal{L}(H)$, c'est à dire :

$$A \in \mathcal{K}, B \in \mathcal{L}(H) \Longrightarrow AB \in \mathcal{K}, BA \in \mathcal{K}.$$

▷

EXERCICE 1.9. ◁ Soit $\mathcal{L}^p(H)$ l'ensemble des opérateurs A dans $\mathcal{L}(H)$ tels que

$$\|A\|_p = \left(\sum_{j=1}^{\infty} \sigma_n(A)^p\right)^{1/p} < \infty.$$

soit fini.

Montrer que pour tout $p \in [1, \infty)$, $\mathcal{L}^p(H)$ est un espace vectoriel, et que de plus, c'est un idéal bilatère d'opérateurs compacts. Dans le cas $p = 1$, l'espace $\mathcal{L}^1(H)$ est l'espace des opérateurs de la classe de trace ; dans le cas $p = 2$, montrer qu'on retrouve les opérateurs de Hilbert-Schmidt. ▷

EXERCICE 1.10. ◁ On pose

$$s_N(A) = \sum_{n=1}^{N} \sigma_n(A).$$

Montrer que

$$s_N(A) = \sup\{\|AP_E\|_1 : \dim E = N\},$$

avec P_E la projection orthogonale sur E. En déduire que pour tout $N \geq 1$, s_N est une norme sur E. ▷

Bibliographie

[1] James W. Demmel. *Applied numerical linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[2] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.

[3] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Rev.*, 35(4) :551–566, 1993.