Notes du cours d'Analyse Numérique Matricielle

Pour Master Mathématiques - EDP -

Chapitre 1

Rappels et compléments d'algèbre linéaire

1.1 Quelques définitions

Soit m et n deux entiers positifs et soit K un corps commutatif. Par la suite seuls les cas $K = \mathbb{R}$ et $K = \mathbb{C}$ seront considérés, et les éléments de K seront de ce fait appelés *scalaires*. Une *matrice* à m lignes et n colonnes est un ensemble de mn scalaires (dits *éléments de la matrice*) indexés par les éléments du produit cartésien $I \times J$ avec $I := [\![1, m]\!]$ et $J := [\![1, n]\!]$: $a_{ij} \in K$, $1 \leq i \leq m$, $1 \leq j \leq n$. Les indices i et j sont dits, respectivement *ligne* et *colonne* de a_{ij}. En effet, on peut interpréter le couple d'indices (i, j) comme les coordonnées de l'élément a_{ij} dans le tableau suivant :

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

Par la suite on utilisera la notation $A = (a_{ij})$ pour indiquer que l'élément générique de la matrice A est noté a_{ij} (les intervalles I et J ne seront précisés si on peut les déduire du contexte). La notation $(A)_{ij}$ pourra également être utilisée pour indiquer l'élément d'indices (i, j). Pour tout $1 \leq i \leq m$ on définit le *vecteur ligne* i de A (noté $A_{i:}$) comme suit :

$$A_{i:} := \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix} \in K^n.$$

De manière analogue, pour tout $1 \leq j \leq m$ on définit le *vecteur colonne* j de A (noté $A_{:j}$) par

$$A_{:j} := \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in K^m.$$

On peut identifier chaque ligne i (resp. colonne j) de A avec le vecteur ligne $A_{i:}$ (resp. vecteur colonne $A_{:j}$) correspondant, et on parlera alors tout simplement de lignes et de colonnes de A. Une matrice qui a autant de lignes que de colonnes est dite *carrée*. Les matrices non carrées sont dites *rectangulaires*.

Définition 1.1 (Matrice triangulaire, strictement triangulaire et diagonale)**.** *Une matrice* $A \in \mathbb{C}^{n,n}$ *est dite* triangulaire supérieure *(resp.* inférieure*) si, pour tout* $1 \le i, j \le n$, $\big(i > j \implies a_{ij} = 0\big)$ *(resp.* $\big(i < j \implies a_{ij} = 0\big)$*). Si les inégalité strictes sont remplacées par des inégalités simples, on parle alors de matrice* strictement triangulaire supérieure *(resp.* inférieure*). Une matrice (strictement) triangulaire inférieure ou supérieure est dite (strictement) triangulaire. La matrice A est dite* diagonale *si* $i \neq j \implies a_{ij} = 0$*. pour tout* $1 \le i, j \le n$*.*

Exercice 1.2 (Matrice triangulaire, strictement triangulaire et diagonale)**.** *Dire si les matrices suivantes sont triangulaires, strictement triangulaires ou diagonales :*

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 2 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Définition 1.3 (Matrice de Hessenberg)**.** *Une matrice* $A \in \mathbb{C}^{n+1,n}$ *est dite* de Hessenberg supérieure *si* $i > j+1 \implies a_{ij} = 0$.

Par la suite nous omettrons d'indiquer les éléments nuls des matrices triangulaires, diagonales ou de Hessenberg.

Définition 1.4 (Transposée et transconjuguée d'une matrice)**.** *Soit* $A = (a_{ij}) \in \mathbb{C}^{m,n}$*. On définit la* transposée *et la* transconjuguée *(ou* matrice adjointe*) de A respectivement par*

$$A^T := (a_{ji}) \in \mathbb{C}^{n,m}, \qquad A^H := (\overline{a}_{ji}) \in \mathbb{C}^{n,m}.$$

Example 1.5 (Transposée et transconjuguée)**.** *On considère la matrice suivante :*

$$A = \begin{pmatrix} 2+3i & 3+4i \\ 1+5i & 3+7i \end{pmatrix}.$$

Sa transposée et transconjuguée sont données, respectivement, par

$$A^T = \begin{pmatrix} 2+3i & 1+5i \\ 3+4i & 3+7i \end{pmatrix}, \qquad A^H = \begin{pmatrix} 2-3i & 1-5i \\ 3-4i & 3-7i \end{pmatrix}.$$

Exercice 1.6 (Propriété de la transconjuguée)**.** *Vérifier que*

(i) Pour toutes matrices $A, B \in \mathbb{C}^{n,m}$*,*

$$(A+B)^H = A^H + B^H; \tag{1.1}$$

(ii) Pour tout scalaire $\alpha \in \mathbb{C}$ *et toute matrice* $A \in \mathbb{C}^{n,m}$*,*

$$(\alpha A)^H = \overline{\alpha} A^H.$$

Soient $l, m, n \in \mathbb{N}_*$, $A = (a_{ij}) \in \mathbb{C}^{n,m}$, $B = (b_{ij}) \in \mathbb{C}^{m,l}$. On définit le *produit matriciel* $AB \in \mathbb{C}^{n,l}$ comme la matrice telle que

$$(AB)_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} \qquad \forall i \in [\![1, n]\!], \forall j \in [\![1, l]\!].$$

Le produit matriciel est illustré dans la Figure 1.1.

Proposition 1.7 (Propriétés du produit matriciel)**.** *Le produit matriciel est (i)* associatif, *à savoir, pour tout A, B, et C pour lesquelles les produits ont un sens, on a* $(AB)C = A(BC) = ABC$ *; (ii)* distributif par rapport à addition, *à savoir, pour toutes matrices A, B, C pour lesquelles l'écriture $A(B+C)$ a un sens on a $A(B+C) = AB + AC$.*

Il est important de retenir que le produit matriciel *n'est pas commutatif.*

Exercice 1.8 (Produit matriciel)**.** *Soient $A \in \mathbb{R}^{2,3}$ et $B \in \mathbb{R}^{3,3}$ définies comme suit :*

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \qquad B := \begin{pmatrix} 2 & 3 & 1 \\ 7 & 3 & 2 \\ 8 & 1 & 2 \end{pmatrix}.$$

On a

$$AB = \begin{pmatrix} 40 & 12 & 11 \\ 28 & 16 & 9 \end{pmatrix}.$$

Précisez si les produits BA et BA^T sont définis et, si c'est le cas, les calculer.

Proposition 1.9 (Produit de deux matrices triangulaires)**.** *Soient A, $B \in \mathbb{R}^{n,n}$ deux matrices triangulaires inférieures (resp. supérieures). Alors, AB est triangulaire inférieure (resp. supérieure).*

Démonstration. On détaille la preuve pour le cas triangulaire inférieure, l'autre cas pouvant se traiter de manière analogue. Par définition nous avons

$$(i < k \implies a_{ik} = 0) \text{ et } \big(k < j \implies b_{kj} = 0\big).$$

Par conséquent,

$$i < j \implies (AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \sum_{k=1}^{i} a_{ik} \underbrace{b_{kj}}_{=0 \text{ car } k \le i < j} + \sum_{k=i+1}^{n} \underbrace{a_{ik}}_{=0 \text{ car } i < k} b_{kj} = 0. \qquad \square$$

Example 1.10 (Produit de deux matrices triangulaires)**.** *On considère les deux matrices triangulaires inférieures*

$$A = \begin{pmatrix} 1 & & \\ 2 & 3 & \\ 2 & 4 & 5 \end{pmatrix}, \qquad B = \begin{pmatrix} 8 & & \\ 2 & 4 & \\ 5 & 9 & 2 \end{pmatrix}.$$

On a

$$AB = \begin{pmatrix} 8 & & \\ 22 & 12 & \\ 49 & 61 & 10 \end{pmatrix}.$$

La preuve de la proposition suivante est laissée en exercice.

Proposition 1.11 (Transposé et transconjugué d'un produit)**.** *Pour tout $A \in \mathbb{C}^{n,m}$ et $B \in \mathbb{C}^{m,l}$ on a*

$$(AB)^H = B^H A^H, \qquad (AB)^T = B^T A^T.$$

Exercice 1.12 (Transposé et transconjugué d'un produit)**.** *Reprendre les matrices de l'Exercice 1.8 et vérifier que* $(AB)^T = B^T A^T$.

Définition 1.13 (Matrice hermitienne et symétrique)**.** *Une matrice complexe* $A \in \mathbb{C}^{n,n}$ *est dite* hermitienne *si* $A^H = A$. *Une matrice réelle* $A \in \mathbb{R}^{n,n}$ *est dite* symétrique *si* $A^T = A$.

Example 1.14 (Matrice hermitienne et symétrique)**.** *Les matrices suivantes sont, respectivement, hermitienne et symétrique.*

$$A = \begin{pmatrix} 1 & 1+i & 2+i \\ 1-i & 2 & 3+i \\ 2-i & 3-i & 3 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 3 \end{pmatrix}.$$

Que peut-on dire des éléments diagonaux d'une matrice hermitienne ?

Définition 1.15 (Matrice normale, unitaire et orthogonale)**.** *Soit* $A \in \mathbb{C}^{n,n}$. *On dit que* A *est* normale *si* $AA^H = A^H A$. *Si, de plus,* $AA^H = A^H A = I_n$ *on dit que* A *est* unitaire. *Si* $A \in \mathbb{R}^{n,n}$, *on dit que* A *est* orthogonale *si* $AA^T = A^T A = I_n$.

Il est utile de rappeler que les produits scalaires canoniques sur $\mathbb{R}^n$ et $\mathbb{C}^n$ sont définis, respectivement, par

$$(x, y)_{\mathbb{R}^n} := y^T x, \quad \forall x, y \in \mathbb{R}^n, \qquad (x, y)_{\mathbb{C}^n} := y^H x, \quad \forall x, y \in \mathbb{C}^n.$$

Quand le contexte rende la notation non ambiguë, nous allons omettre l'indice pour le produit interne de $\mathbb{R}^n$.

Remarque 1.16 (Matrices unitaires et norme 2)**.** *La dénomination des matrices unitaires se justifie par la remarque suivante. Soit* $\|\cdot\|_2$ *la norme vectorielle sur* $\mathbb{C}^n$ *définie par* $\|x\|_2^2 := (x, x)_{\mathbb{C}^n} = x^H x$. *Alors, pour toute matrice unitaire* $U \in \mathbb{C}^{n,n}$ *et tout vecteur* $x \in \mathbb{C}^n$,

$$\|Ux\|_2^2 = (Ux, Ux)_{\mathbb{C}^n} = (Ux)^H Ux = x^H U^H U x = x^H x = \|x\|_2^2,$$

à savoir, la multiplication par une matrice unitaire ne modifie pas la norme d'un vecteur. *Plus généralement, pour tout* $x, y \in \mathbb{C}^n$, *on a*

$$(Ux, Uy)_{\mathbb{C}^n} = (Uy)^H Ux = y^H U^H U x = y^H x = (x, y)_{\mathbb{C}^n},$$

à savoir, les matrices unitaires préservent le produit scalaire.

Example 1.17 (Matrice élémentaire unitaire)**.** *Soit* $w \in \mathbb{C}^n$ *tel que* $(w, w)_{\mathbb{C}^n} = w^H w = 1$. *On définit*

$$A := I_n - 2ww^H. \tag{1.2}$$

Une matrice de la forme (1.2) *est dite* élémentaire. *On remarquera que* A *est hermitienne. En effet, en utilisant* (1.1) *suivi par la Proposition 1.11, il vient*

$$A^H = (I_n - 2ww^H)^H = I_n^H - 2(w^H)^H w^H = I_n - 2ww^H = A.$$

On vérifie ensuite que A est unitaire. En effet,

$$\begin{aligned}
A^H A &= (I_n - 2ww^H)^H (I_n - 2ww^H) \\
&= (I_n - 2ww^H)(I_n - 2ww^H) && \textit{A hermitienne} \\
&= I_n - 4ww^H + 4(ww^H)(ww^H) && \textit{Proposition 1.7, distributivité} \\
&= I_n - 4ww^H + 4w\underbrace{(w^H w)}_{=1} w^H && \textit{Proposition 1.7, associativité} \\
&= I_n - 4ww^H + 4ww^H = I_n.
\end{aligned}$$

Définition 1.18 (Inverse d'une matrice)**.** *Soit $A \in \mathbb{C}^{n,n}$. On dit que A est* inversible *s'il existe $B \in \mathbb{C}^{n,n}$ telle que $AB = BA = I_n$. B est dit* inverse *de la matrice A et elle est notée A^{-1}.*

On vérifie aisément que l'inverse d'une matrice, si elle existe, est unique. Pour s'en convaincre, soit $A \in \mathbb{C}^{n,n}$ une matrice inversible et $B, C \in \mathbb{C}^{n,n}$ deux matrices telles que $AB = BA = AC = CA = I_n$. De par la Proposition 1.7 on a

$$AB = BA \implies CAB = CBA \implies (CA)B = C(BA) \implies I_n B = CI_n \implies B = C.$$

Par définition, en outre, toute matrice $A \in \mathbb{C}^{n,n}$ (resp. $A \in \mathbb{R}^{n,n}$) unitaire (resp. orthogonale) est inversible avec $A^{-1} = A^H$ (resp. $A^{-1} = A^T$).

Il est utile de considérer l'inverse de la transposée d'une matrice inversible. On a, par définition,

$$I_n = (A^T)^{-1} A^T \iff I_n^T = \left[(A^T)^{-1} A^T\right]^T \iff I_n = A\left[(A^T)^{-1}\right]^T = AA^{-1},$$

à savoir $(A^{-1})^T = (A^T)^{-1}$. Cette remarque suggère la définition suivante.

Définition 1.19 (Inverse transposée)**.** *Soit $A \in \mathbb{C}^{n,n}$ inversible. On définit la matrice* inverse transposée *de A par $A^{-T} := (A^{-1})^T = (A^T)^{-1}$.*

Proposition 1.20 (Inverse d'un produit)**.** *Pour tout $A, B \in \mathbb{C}^{n,n}$ inversibles telles que AB est inversible on a*

$$(AB)^{-1} = B^{-1} A^{-1}.$$

Démonstration. On a que

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1} I_n B = B^{-1} B = I_n,$$

où nous avons utilisé l'associativité du produit matriciel (voir Proposition 1.7) dans la première égalité. D'autre part, en procédant de façon similaire, on a que

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_n A^{-1} = AA^{-1} = I_n.$$

La conclusion s'ensuit par Définition 1.18 de l'inverse. □

1.2 Valeurs et vecteurs propres, rayon spectral déterminant

Définition 1.21 (Valeurs et vecteurs propres)**.** *Soit $A \in \mathbb{C}^{n,n}$. Un scalaire $\lambda \in \mathbb{C}$ est dit* valeur propre *de A s'il existe un vecteur $x \in \mathbb{C}^n$* non nul *tel que*

$$Ax = \lambda x.$$

On dit dans ce cas que $x \in \mathbb{C}^n$ est un vecteur propre *associé à la valeur propre λ. L'ensemble des valeurs propres d'une matrice A, noté* $\mathrm{Sp}(A)$*, est dit* spectre *de A.*

Les valeurs propres sont par définition les solutions de l'*équation caractéristique*

$$p_A(\lambda) := \det(A - \lambda I_n) = 0.$$

Comme p_A est un polynôme de degré n, il admet précisément n racines complexes (non nécessairement distinctes).

Proposition 1.22 (Valeurs propres de l'inverse d'une matrice)**.** *Pour toute matrice $A \in \mathbb{C}^{n,n}$ inversible on a*

$$\lambda \in \mathrm{Sp}(A) \implies \lambda^{-1} \in \mathrm{Sp}(A^{-1}). \tag{1.3}$$

Démonstration. On verra plus loin (Théorème 1.33) que l'inversibilité de A implique que toutes ses valeurs propres sont non nulles. Comme $\lambda \in \mathrm{Sp}(A)$, il existe $x \in \mathbb{C}^n$ non nul tel que

$$Ax = \lambda x \iff x = A^{-1}Ax = \lambda A^{-1}x \iff \lambda^{-1}x = A^{-1}x,$$

ce qui prouve que $\lambda^{-1} \in \mathrm{Sp}(A^{-1})$. □

Définition 1.23 (Matrice HDP et SDP)**.** *Une matrice $A \in \mathbb{C}^{n,n}$ est HDP si elle est* hermitienne *et* définie positive*, à savoir,*

$$(Ax, x)_{\mathbb{C}^n} > 0 \qquad \forall x \in \mathbb{C}^n \setminus \{0 \in \mathbb{C}^n\}.$$

Une matrice $A \in \mathbb{R}^{n,n}$ est SDP si elle est symétrique *et* définie positive.

Remarque 1.24 (Matrices définies positives non symétrique)**.** *Considérons, par simplicité, le cas de matrices réelles. On peut se poser la question s'il existe des matrices définies positives et non symétriques. Soit $A \in \mathbb{R}^{n,n}$. On commence par remarquer que l'on peut décomposer A en la somme d'une partie symétrique A_{s} et une anti-symétrique A_{ss},*

$$A = \frac{1}{2}\left(A + A^T\right) + \frac{1}{2}\left(A - A^T\right) := A_{\mathrm{s}} + A_{\mathrm{ss}}.$$

Par définition de matrice définie positive on a pour tout $x \in \mathbb{R}^n$, $x \neq 0$,

$$0 < (Ax, x) = (A_{\mathrm{s}}x, x) + (A_{\mathrm{ss}}x, x). \tag{1.4}$$

On a

$$(A_{\mathrm{ss}}x, x) = x^T A_{\mathrm{ss}} x = (x^T A_{\mathrm{ss}} x)^T = x^T A_{\mathrm{ss}}^T x = -x^T A_{\mathrm{ss}} x = -(A_{\mathrm{ss}}x, x),$$

où nous avons utilisé, respectivement, la définition du produit scalaire de $\mathbb{R}^n$*, le fait que la transposée d'un scalaire est le scalaire même, la Proposition 1.11 deux fois, le fait que* $A_{ss} = -A_{ss}^T$ *(car* A_{ss} *est antisymétrique), et encore une fois la définition du produit scalaire de* $\mathbb{R}^n$*. Ceci montre que* $(A_{ss}x, x) = 0$*, à savoir, la partie anti-symétrique d'une matrice ne contribue pas à son caractère défini positif. De plus, on peut déduire de* (1.4) *qu'une matrice est positive définie si et seulement si sa partie symétrique l'est.*

Example 1.25 (Matrice définie positive non symétrique)**.** *On considère la matrice* $A \in \mathbb{R}^{2,2}$ *suivante*

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

Vérifier que A est définie positive et montrer que $\mathrm{Sp}(A) = \{1-i, 1+i\}$*. Cet exercise montre, au passage, qu'une matrice définie positive* mais non symétrique *n'a pas forcement des valeurs propres réelles positives.*

Proposition 1.26 (Valeurs propres d'une matrice HDP/SDP)**.** *Soit* $A \in \mathbb{C}^{n,n}$ *(resp.* $A \in \mathbb{R}^{n,n}$*) une matrice HDP (resp. SDP). Alors, toutes les valeurs propres de A sont réelles et strictement positives.*

Démonstration. On détaille la preuve uniquement pour le cas HDP, l'autre étant similaire. Soit λ une valeur propre de A et $x \in \mathbb{C}^n \setminus \{0 \in \mathbb{C}^n\}$ un vecteur propre associé. Alors,

$$Ax = \lambda x \implies \mathbb{R}_*^+ \ni x^H Ax = \lambda x^H x = \lambda \|x\|_2^2 \implies \lambda = \frac{x^H Ax}{\|x\|_2^2} \in \mathbb{R}_*^+. \qquad \square$$

Nous admettrons le résultat suivant.

Proposition 1.27 (Valeurs propres d'une matrice triangulaire)**.** *Les valeurs propres d'une matrice triangulaire sont ses éléments diagonaux.*

Définition 1.28 (Rayon spectral)**.** *Soit* $A \in \mathbb{C}^{n,n}$*. On définit son* rayon spectral *comme la plus grande valeur absolue de ses valeurs propres, à savoir,*

$$\rho(A) := \max_{\lambda \in \mathrm{Sp}(A)} |\lambda|.$$

Le rayon spectral n'est pas une norme matricielle (voir Section 1.5), car on peut avoir $\rho(A) = 0$ sans que A soit nulle (il suffit de prendre une matrice triangulaire non nulle avec diagonale nulle pour s'en convaincre).

Définition 1.29 (Déterminant)**.** *Soit* $A \in \mathbb{C}^{n,n}$*. Le* déterminant *de A est donné par*

$$\det(A) := \prod_{\lambda \in \mathrm{Sp}(A)} \lambda.$$

Le déterminant d'une matrice permet de décider de son inversibilité. Plus précisément, une matrice est inversible si et seulement si son déterminant est non nul. De par la Proposition 1.22 on a, en outre, que

$$\det(A^{-1}) = \prod_{\lambda \in \mathrm{Sp}(A^{-1})} \lambda = \prod_{\tilde{\lambda} \in \mathrm{Sp}(A)} \frac{1}{\tilde{\lambda}} = \frac{1}{\det(A)}.$$

1.3 Noyau et image d'une matrice

Soit $A \in \mathbb{R}^{m,n}$ et $\mathscr{C}_A$ le sous-espace vectoriel de $\mathbb{R}^m$ engendré par ses colonnes,

$$\mathscr{C}_A := \operatorname{span}(A_{:j})_{1\le j\le n}.$$

L'*image* de A est le sous-espace vectoriel de $\mathbb{R}^m$ défini par

$$\operatorname{range}(A) := \{Ax \,:\, x \in \mathbb{R}^n\} \in \mathbb{R}^m.$$

Le *noyau* de A est la préimage du vecteur nul de $\mathbb{R}^m$ par l'application linéaire associée à A,

$$\ker(A) := \{x \in \mathbb{R}^n \,:\, Ax = 0 \in \mathbb{R}^m\} \in \mathbb{R}^n.$$

Proposition 1.30 (Image d'une matrice)**.** *On a* $\mathscr{C}_A = \operatorname{range}(A)$.

Démonstration. Il suffit de remarquer que le résultat de la multiplication (à droite) d'une matrice $A \in \mathbb{R}^{m,n}$ par un vecteur $x \in \mathbb{R}^n$ est la combinaison linéaire des colonnes de la matrice avec coefficients donnés par les composantes de x,

$$\left((Ax)_i = \sum_{j=1}^n A_{ij}x_j \quad \forall i \in [\![1,m]\!]\right) \iff Ax = \sum_{j=1}^n A_{:j}x_j. \qquad \square$$

Définition 1.31 (Rang d'une matrice)**.** *On définit le* rang *d'une matrice comme la dimension de son image,*

$$\operatorname{rank}(A) := \dim(\operatorname{range}(A)).$$

Proposition 1.32 (Propriétés du rang d'une matrice)**.** *Soit* $A \in \mathbb{R}^{m,n}$*. Nous avons* $\operatorname{rank}(A) = \operatorname{rank}(A^T)$ *et*

$$\operatorname{rank}(A) + \dim(\ker(A)) = n.$$

Le résultat suivant, que l'on admettra, établit un lien entre le déterminant, le rang, et le noyau d'une matrice et son inversibilité.

Théorème 1.33 (Caractérisation des matrices inversibles)**.** *Soit* $A \in \mathbb{R}^{n,n}$*. Les propriétés suivantes sont équivalentes : (i) A est inversible; (ii)* $\det(A) \neq 0$*; (iii)* $\ker(A) = \{0 \in \mathbb{R}^n\}$*; (iv)* $\operatorname{rank}(A) = n$ *; (v) les colonnes et les lignes de A sont linéairement indépendantes.*

Les résultats de cette section s'étendent sans difficulté au cas de matrices complexes.

1.4 Décompositions d'une matrice

Il est souvent utile de décomposer une matrice en un produit de matrices ayant des propriétés favorables. Dans cette section nous allons rappeler quelques décompositions importantes dans les applications numériques.

1.4.1 Matrices diagonalisables

Définition 1.34 (Matrice diagonalisable)**.** *Une matrice $A \in \mathbb{C}^{n,n}$ est* diagonalisable *s'il existe une matrice $Q \in \mathbb{C}^{n,n}$ inversible telle que $A = Q^{-1}\Lambda Q$ avec $\Lambda \in \mathbb{C}^{n,n}$ matrice diagonale.*

Proposition 1.35 (Valeurs et vecteurs propres d'une matrice diagonalisable)**.** *Soit $A \in \mathbb{C}^{n,n}$ une matrice diagonalisable. Alors, en reprenant la notation de la Définition 1.34, Λ contient les valeurs propres de A et les colonnes de Q^{-1} sont des vecteurs propres associés.*

Démonstration. En multipliant à droite par Q^{-1} la relation $A = Q^{-1}\Lambda Q$ on obtient $AQ^{-1} = Q^{-1}\Lambda$, ou, de façon équivalente,

$$AQ^{-1}_{:j} = Q^{-1}\Lambda_{:j} = \lambda_j Q^{-1}_{:j} \qquad \forall 1 \le j \le n,$$

où nous avons noté $Q^{-1}_{:j}$ la j-ème colonne de Q^{-1} et $\lambda_j = (\Lambda)_{jj}$. □

La Proposition 1.35 suggère le résultat suivant, que nous admettrons.

Théorème 1.36 (Caractérisation d'une matrice diagonalisable)**.** *Une matrice $A \in \mathbb{C}^{n,n}$ (resp. $A \in \mathbb{R}^{n,n}$) est diagonalisable si et seulement si on peut construire une base de $\mathbb{C}^n$ (resp. $\mathbb{R}^n$) formée de vecteurs propres de A.*

Dans certains cas, on peut prouver des propriétés additionnelles pour la matrice Q qui apparaît dans la Définition 1.34. Un exemple à retenir est donné dans le lemme suivant, qui affirme que les matrices normales sont les seules matrices unitairement semblables à des matrices diagonales.

Lemme 1.37 (Diagonalisation d'une matrice normale)**.** *La matrice $A \in \mathbb{C}^{n,n}$ est normale (c.-à-d., $AA^H = A^H A$) si et seulement s'il existe une matrice unitaire $U \in \mathbb{C}^{n,n}$ (c.-à-d., $UU^H = U^H U = I_n$) telle que*

$$U^{-1}AU = U^H AU = \Lambda := \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

avec $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ valeurs propres de A.

1.4.2 Décomposition de Schur

Pour une matrice générique, on peut prouver uniquement qu'elle est unitairement semblable à une matrice triangulaire, comme l'affirme le résultat suivant.

Théorème 1.38 (Décomposition de Schur (DS))**.** *Pour toute $A \in \mathbb{C}^{n,n}$ il existe $U \in \mathbb{C}^{n,n}$ unitaire telle que*

$$U^{-1}AU = U^H AU = \begin{pmatrix} \lambda_1 & b_{12} & \cdots & b_{1n} \\ & \lambda_2 & & b_{2n} \\ & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix} =: T,$$

où $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ sont les valeurs propres de A.

On pourra remarquer que, si T et une matrice carrée triangulaire supérieure, sa décomposition de Schur s'obtient en posant $U = I_n$. Ceci implique, en particulier, que les valeurs propres de T sont ses éléments diagonaux. Ce résultat s'applique également aux matrices triangulaires inférieures.

Le corollaire suivant montre une autre conséquence importante de la DS, à savoir, toute matrice hermitienne est diagonalisable.

Corollaire 1.39 (Valeurs propres et DS des matrices complexes hermitienne et réelles symétriques)**.** *Si la matrice $A \in \mathbb{C}^{n,n}$ est hermitienne, T est diagonale, $\lambda_i \in \mathbb{R}$ pour tout $1 \le i \le n$, et les colonnes de U (ou, de manière équivalente, les lignes de $U^H = U^{-1}$) sont des vecteurs propres de A.*

D'une même manière, si la matrice $A \in \mathbb{R}^{n,n}$ est symétrique, T est diagonale, $\lambda_i \in \mathbb{R}$ pour tout $1 \le i \le n$, U est à valeurs réelles et orthogonale, et ses colonnes (ou, de manière équivalent, les lignes de $U^T = U^{-1}$) sont des vecteurs propres de A.

Démonstration. On se limite à prouver le premier point. Comme A est hermitienne on a

$$T^H = (U^H A U)^H = U^H A^H U = U^H A U = T,$$

ce qui implique $T = \text{diag}(\lambda_1, \ldots, \lambda_n)$ et, pour tout $1 \le i \le n$, $\lambda_i = \overline{\lambda_i} \iff \Im(\lambda_i) = 0 \iff \lambda_i \in \mathbb{R}$. Pour prouver que les lignes de U sont des vecteurs propres on procède comme dans la preuve de la Proposition 1.35 en observant que $U^{-1} = U^H$. □

1.4.3 Décomposition en valeurs singulières

Toute matrice $A \in \mathbb{C}^{m,n}$ peut être transformée en une matrice diagonale rectangulaire à l'aide de *deux* matrices unitaires.

Théorème 1.40 (Décomposition en valeurs singulières (DVS))**.** *Pour toute $A \in \mathbb{C}^{m,n}$ il existe deux matrices unitaires $U \in \mathbb{C}^{m,m}$ et $V \in \mathbb{C}^{n,n}$ telles que*

$$U^H A V = \Sigma = \text{diag}(\sigma_1, \ldots \sigma_p) \in \mathbb{R}^{m,n} \qquad p := \min(m, n).$$

Les réels $0 \le \sigma_1 \le \ldots \le \sigma_p$ sont dits valeurs singulières *de A. Si, de plus, $A \in \mathbb{R}^{m,n}$, on a $U \in \mathbb{R}^{m,m}$ et $V \in \mathbb{R}^{n,n}$ et on peut écrire $U^T A V = \Sigma$.*

Proposition 1.41 (Caractérisation des valeurs singulières)**.** *Soit $A \in \mathbb{C}^{n,n}$. On a*

$$\sigma_i(A) = \sqrt{\lambda_i(A^H A)} \qquad \forall 1 \le i \le n.$$

Démonstration. On a

$$\Sigma^H \Sigma = (U^H A V)^H U^H A V = V^H A^H U^H U A V = V^H A^H A V = V^{-1} A^H A V$$

où nous avons utilisé le fait que U et V sont unitaires pour conclure $U^H U = I_n$ et $V^H = V^{-1}$. La matrice diagonale $\Sigma^H \Sigma$ contient les valeurs propres de $A^H A$ et les colonnes de V des vecteurs propres correspondants. Pour s'en convaincre, il suffit de procéder comme dans la preuve de la Proposition 1.35. Comme $\Sigma^H \Sigma = \text{diag}(\sigma_i^2(A))_{1 \le i \le n}$ et $\sigma_i(A) \ge 0$ pour tout $1 \le i \le n$, on a donc

$$\lambda_i(A^H A) = \sigma_i(A)^2 \iff \sigma_i(A) = \sqrt{\lambda_i(A^H A)}. \qquad \square$$

Corollaire 1.42 (Valeurs singulières et rayon spectral d'une matrice hermitienne)**.** *Si $A \in \mathbb{C}^{n,n}$ est une matrice hermitienne, on a*

$$\sigma_i(A) = \sqrt{\lambda_i(A)^2} = |\lambda_i(A)|, \qquad \rho(A) = \max_{1\le i\le n} \sigma_i(A) = \|A\|_2.$$

Démonstration. Comme A est hermitienne on a $A^H A = A^2$. Or, soit λ une valeur propre de A et $x \in \mathbb{C}^n$ un vecteur propre associé. Alors

$$Ax = \lambda x \implies AAx = \lambda Ax = \lambda^2 x,$$

à savoir, $\lambda_i(A^H A) = \lambda_i(A^2) = \lambda_i(A)^2 = \sigma_i(A)^2$ pour tout $1 \le i \le n$. Par la Définition 1.28 de rayon spectrale, on a de plus

$$\max_{1\le i\le n} \sigma_i(A) = \max_{1\le i\le n} |\lambda_i(A)| = \rho(A).$$

La conclusion s'ensuit de la Proposition 1.48 plus bas. □

1.5 Normes matricielles

Définition 1.43 (Norme matricielle)**.** *Soit K un corps commutatif. Une norme matricielle sur $K^{n,n}$ est une application $\|\cdot\|$ de $K^{n,n}$ dans $\mathbb{R}$ telle que, pour tout $A \in K^{n,n}$,*

(i) $\|A\| > 0$ *si* $A \neq 0$ *et* $\|A\| = 0$ *si et seulement si* $A = 0 \in K^{n,n}$ *;*

(ii) $\|\alpha A\| = |\alpha|\|A\|$ *pour tout* $\alpha \in K$ *;*

(iii) $\|A + B\| \le \|A\| + \|B\|$ *pour tout* $B \in K^{n,n}$*.*

*La dernière propriété est connue sous le nom d'*inégalité triangulaire.

Remarque 1.44 (Norme sous-multiplicative)**.** *Afin d'identifier des normes matricielles d'intérêt pratique, on ajoute souvent la propriété suivante : pour toutes matrices $A \in \mathbb{R}^{n,m}$ et $B \in \mathbb{R}^{m,q}$,*

$$\|AB\| \le \|A\|\|B\|.$$

Une norme matricielle qui satisfait cette propriété s'appelle sous-multiplicative. *On verra dans la Proposition 1.49 que les normes subordonnées définies plus bas sont sous-multiplicatives.*

Dans le reste de cette section nous allons supposer $A \in \mathbb{R}^{n,n}$ sans nécessairement le préciser à chaque fois.

Définition 1.45 (Norme matricielle subordonnée à une norme vectorielle)**.** *Soit $\|\cdot\|$ une norme vectorielle sur $\mathbb{R}^n$. On définit la* norme matricielle subordonné à la norme vectorielle *$\|\cdot\|$ par*

$$\forall A \in \mathbb{R}^{n,n}, \qquad \|A\| := \sup_{x\in\mathbb{R}^n,\, x\neq 0} \frac{\|Ax\|}{\|x\|}.$$

Remarque 1.46 (Définitions alternatives d'une norme subordonnée)**.** *Soit $\|\cdot\|$ une norme vectorielle sur $\mathbb{R}^n$. La norme matricielle subordonnée à $\|\cdot\|$ vérifie pour tout $A \in \mathbb{R}^{n,n}$,*

$$\|A\| = \sup_{x\in\mathbb{R}^n,\, \|x\|=1} \|Ax\| = \sup_{x\in\mathbb{R}^n,\, \|x\|\le 1} \|Ax\|.$$

Des normes matricielles particulièrement importantes sont les normes p subordonnées aux normes vectorielles définies par

$$\forall x \in \mathbb{R}^n, \qquad \|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \tag{1.5}$$

avec $p \in \mathbb{N}_*$ et

$$\|x\|_\infty := \max_{1 \le i \le n} |x_i|. \tag{1.6}$$

Cette définition s'étend à p réel ≥ 1, mais les valeurs entières de p sont les plus souvent utilisées en pratique.

Proposition 1.47 (Norme d'un produit matrice-vecteur)**.** *Soit* $\|\cdot\|$ *une norme matricielle subordonnée à la norme vectorielle* $\|\cdot\|$*. Alors pour toute matrice* $A \in \mathbb{R}^{n,n}$ *et tout vecteur* $x \in \mathbb{R}^n$*,*

$$\|Ax\| \le \|A\| \|x\|.$$

Démonstration. Si $x = 0 \in \mathbb{R}^n$, le résultat est trivial. Si $x \neq 0 \in \mathbb{R}^n$, on a par définition

$$\|A\| := \sup_{y \in \mathbb{R}^n,\, y \neq 0} \frac{\|Ay\|}{\|y\|} \ge \frac{\|Ax\|}{\|x\|},$$

où la deuxième inégalité vient de la définition de supremum. □

Proposition 1.48 (Normes 1, 2 et ∞)**.** *On a*

$$\|A\|_2 = \|A^T\|_2 = \max_{1 \le i \le n} \sigma_i(A), \tag{1.7}$$

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|, \tag{1.8}$$

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|. \tag{1.9}$$

Démonstration. Soit $A \in \mathbb{R}^{n,n}$.

1) *Preuve de* (1.7). On commence par remarquer que la matrice $A^T A$ est symétrique. De par le Corollaire 1.39, il existe une matrice orthogonale $U \in \mathbb{R}^{n,n}$ telle que $UA^T AU^T = \Lambda = \mathrm{diag}(\lambda_1(A^T A), \ldots, \lambda_n(A^T A))$ où on a noté $\lambda_i(A^T A) \ge 0$, $1 \le i \le n$, les valeurs propres de $A^T A$ (le fait que ces valeurs propres soient non négatives est une conséquence de la Proposition 1.41). Nous avons alors que

$$\begin{aligned}
\|A\|_2 &= \sup_{x \in \mathbb{R}^n,\, \|x\|_2 = 1} \|Ax\|_2 && \text{Définition 1.45} \\
&= \sup_{x \in \mathbb{R}^n,\, \|x\|_2 = 1} \sqrt{x^T A^T A x} && \text{Éq. (1.5)} \\
&= \sup_{x \in \mathbb{R}^n,\, \|x\|_2 = 1} \sqrt{x^T U^T \Lambda U x} && A^T A = U^T \Lambda U \\
&= \sup_{y \in \mathbb{R}^n,\, \|U^T y\|_2 = 1} \sqrt{y^T \Lambda y} && y := Ux \\
&= \sup_{y \in \mathbb{R}^n,\, \|y\|_2 = 1} \sqrt{y^T \Lambda y} && \text{Remarque 1.16} \\
&= \max_{1 \le i \le n} \sqrt{\lambda_i(A^T A)} = \max_{1 \le i \le n} \sigma_i(A), && \text{Proposition 1.41}
\end{aligned}$$

ce qui prouve le premier point.

2) *Preuve de* (1.8). Pour tout $x \in \mathbb{R}^n$ on a que

$$\|Ax\|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \le \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}||x_j| = \sum_{j=1}^{n} |x_j| \left(\sum_{i=1}^{n} |a_{ij}| \right) \le \left(\max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}| \right) \|x\|_1.$$

Par conséquent,

$$\|A\|_1 = \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 \le \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|. \tag{1.10}$$

Soit maintenant $\hat{j}$ l'indice qui réalise le maximum à droite de l'expression précédente, de telle sorte à avoir :

$$\sum_{i=1}^{n} |a_{i\hat{j}}| = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|.$$

En choisissant $\hat{x} \in \mathbb{R}^n$ tel que $\hat{x}_{\hat{j}} = 1$ et $\hat{x}_j = 0$ pour tout $j \ne \hat{j}$, on a $\|\hat{x}\|_1 = 1$ et

$$\|A\hat{x}\|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} \hat{x}_j \right| = \sum_{i=1}^{n} |a_{i\hat{j}}| = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|. \tag{1.11}$$

En combinant (1.10) et (1.11) nous avons que

$$\max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}| = \|A\hat{x}\|_1 \le \sup_{x \in \mathbb{R}^n, \|x\|_1 = 1} \|Ax\|_1 = \|A\|_1 \le \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|,$$

ce qui prouve (1.8).

3) *Preuve de* (1.9). Pour tout $x \in \mathbb{R}^n$ on a que

$$\begin{aligned} \|Ax\|_\infty &= \max_{1 \le i \le n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \\ &\le \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}||x_j| \\ &\le \max_{1 \le i \le n} \left(\max_{1 \le j \le n} |x_j| \sum_{j=1}^{n} |a_{ij}| \right) \le \left(\max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}| \right) \|x\|_\infty. \end{aligned}$$

(Justifier soigneusement les passages à la deuxième et à la troisième ligne.) Par conséquent,

$$\|A\|_\infty = \sup_{x \in \mathbb{R}^n, \|x\|_\infty = 1} \|Ax\|_\infty \le \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|. \tag{1.12}$$

Notons $\hat{\imath}$ l'indice qui réalise le maximum dans l'expression précédente, et considérons le vecteur $\hat{x} \in \mathbb{R}^n$ tel que, pour tout $1 \le j \le n$, $\hat{x}_j = 1$ si $a_{\hat{\imath} j} \ge 0$, $\hat{x}_j = -1$ sinon. On a clairement $\|\hat{x}\|_\infty = 1$ et

$$\|A\hat{x}\|_\infty = \max_{1 \le i \le n} \left| \sum_{j=1}^{n} a_{ij} \hat{x}_j \right| = \sum_{j=1}^{n} |a_{\hat{\imath} j}|. \tag{1.13}$$

L'équation (1.13) demande une justification détaillée. Définissons les ensembles suivants :

$$\mathscr{A} := \left\{ \left| \sum_{j=1}^{n} a_{ij}\hat{x}_j \right| \,:\, 1 \le i \le n \right\}, \qquad \mathscr{B} := \left\{ \sum_{j=1}^{n} |a_{ij}| \,:\, 1 \le i \le n \right\}.$$

On démontre aisément que pour tout $\alpha \in \mathscr{A}$, il existe $\beta \in \mathscr{B}$ tel que $\alpha \le \beta$. En effet, si i est tel que $\alpha = \left|\sum_{j=1}^{n} a_{ij}\hat{x}_j\right|$, il suffit de prendre $\beta = \sum_{j=1}^{n} |a_{ij}|$ et d'utiliser l'inégalité triangulaire avec le fait que $|\hat{x}_j| = 1$ pour tout $1 \le j \le n$. De plus, par définition de $\hat{\imath}$, on a que

$$\max_{\beta \in \mathscr{B}} \beta = \sum_{j=1}^{n} |a_{\hat{\imath}j}| = \sum_{j=1}^{n} a_{\hat{\imath}j}\hat{x}_j = \left| \sum_{j=1}^{n} a_{\hat{\imath}j}\hat{x}_j \right| \in \mathscr{A},$$

où nous avons utilisé le fait que tous les termes de la somme sont positif dans la troisième égalité. En combinant les résultats précédents on déduit (1.13). Enfin, la rélation (1.9) s'ensuit en combinant (1.12) et (1.13) comme dans le point 2) de cette preuve. □

Proposition 1.49 (Propriétés des normes subordonnées)**.** *Soit* $\|\cdot\|$ *une norme matricielle subordonnée sur* $\mathbb{R}^{n,n}$*. Alors,*

(i) pour toute matrice $A \in \mathbb{R}^{n,n}$ *il existe* $x_A \in \mathbb{R}^n \setminus \{0\}$ *tel que* $\|A\| = \|Ax_A\|/\|x_A\|$ *;*
(ii) en notant I_n *la matrice identité de taille* n*, on a* $\|I_n\| = 1$ *;*
(iii) pour toutes matrices $A, B \in \mathbb{R}^{n,n}$*,* $\|AB\| \le \|A\|\,\|B\|$*.*

Démonstration. (i) La fonction $\|Ax\|$ définie sur $\mathbb{R}^n$ et à valeurs réelles est par définition continue car $\|Ax\| \le \|A\|\|x\|$. Par conséquent, elle atteint son maximum sur le compact $\|x\| = 1$.

(ii) On a par définition $\|I_n\| = \sup_{x\in\mathbb{R}^n, \|x\|=1} \|I_n x\| = 1$.

(iii) On peut écrire :

$$\|AB\| = \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|ABx\|}{\|x\|} \le \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|A\|\|Bx\|}{\|x\|} = \|A\| \sup_{x\in\mathbb{R}^n, x\neq 0} \frac{\|Bx\|}{\|x\|} = \|A\|\|B\|,$$

où nous avons utilisé la Proposition 1.47 dans l'inégalité, sortie $\|A\|$ du sup dans le troisième passage, et utilisé à nouveau la Définition 1.45 de norme matricielle subordonnée pour conclure.

□

On amdettra le résultat suivant qui relie le rayon spectral avec les normes matricielles subordonnées.

Lemme 1.50 (Rayon spectral et normes subordonnées)**.** *On a les résultats suivants :*

(i) Soit $\|\cdot\|$ *une norme matricielle subordonnée. Alors, pour toute -*$A \in \mathbb{R}^{n,n}$*,* $\rho(A) \le \|A\|$ *;*
(ii) Pour toute matrice A *et tout réel* $\epsilon > 0$*, il existe une norme subordonnée* $\|\cdot\|$ *telle que*

$$\|A\| \le \rho(A) + \epsilon.$$

1.6 Exercices

Exercice 1.51 (Matrice SDP)**.** *Soit* $A = (a_{ij}) \in \mathbb{R}^{n,n}$ *une matrice SDP. Prouver que* $a_{kk} > 0$ *pour tout* $1 \le k \le n$.

Comme A est SDP, on a pour tout $x \in \mathbb{R}^n \setminus \{0 \in \mathbb{R}^n\}$

$$x^T Ax = (Ax, x)_{\mathbb{R}^n} > 0.$$

En prénant $x = e^k$ avec e^k k-ème vecteur de la base canonique de $\mathbb{R}^n$ tel que $e_l^k = \delta_{kl}$ pour tout $1 \le l \le n$, nous avons

$$0 < (Ae^k, e^k)_{\mathbb{R}^n} = \sum_{i=1}^{n} e_i^k \sum_{j=1}^{n} a_{ij} e_j^k = \sum_{i=1} e_i^k a_{ik} = a_{kk},$$

qui est le résultat cherché.

Exercice 1.52 (Matrice antisymétrique)**.** *Soit* $A \in \mathbb{R}^{n,n}$ *une matrice* antisymétrique*, à savoir* $A^T = -A$*. On pose*

$$B_{\pm} := I_n \pm A.$$

On supposera B_- *inversible. (i) Vérifier que* B_+ *est inversible et identifier son inverse. (ii) Vérifier que* B_- *est normale. (iii) Montrer que la matrice* $C := B_+ B_-^{-1}$ *est orthogonale, à savoir,* $B^{-1} = B^T$.

Exercice 1.53 (Matrice hermitienne et antihermitienne)**.** *Une matrice* $A \in \mathbb{C}^{n,n}$ *est dite* antihermitienne *si* $A^H = -A$*. Montrer que*

(i) les éléments diagonaux d'une matrice hermitienne sont des réels, tandis que ceux d'une matrice antihermitienne sont des imaginaires purs ;

(ii) montrer que, si une matrice triangulaire est hermitienne ou antihermitienne, elle est diagonale.

(i) Si A est hermitienne on a $a_{ii} = \overline{a_{ii}}$ pour tout $1 \le i \le n$, à savoir, $\Im(a_{ii}) = 0 \implies a_{ii} \in \mathbb{R}$. Si A est antihermitienne, $a_{ii} = -\overline{a_{ii}} \implies \Re(a_{ii}) = 0$ pour tout $1 \le i \le n$.

(ii) Supposons A triangulaire supérieure (resp. inférieure). Si A est hermitienne ou antihermitienne on a $a_{ij} = \pm 0 \implies a_{ij} = 0$ pour tout $j > i$ (resp. $i > j$), à savoir, A est diagonale.

Exercice 1.54 (Matrices de Hilbert)**.** *La* matrice de Hilbert $H(n) = (h_{ij}) \in \mathbb{R}^{n,n}$ *d'ordre* $n \ge 1$ *est la matrice carrée symétrique à valeurs réelles telle que*

$$h_{ij} = \frac{1}{i + j - 1}.$$

Montrer que $H(n)$ *est définie positive.*

Pour tout $1 \le i, j \le n$, on remarque que

$$h_{ij} = \frac{1}{i + j - 1} = \int_0^1 t^{j-1} t^{i-1} \mathrm{d}t. \tag{1.14}$$

Soit maintenant $x = (x_i)_{1 \le i \le n} \in \mathbb{R}^n$ non nul. On a

$$\begin{aligned}
x^T H(n) x &= \sum_{i=1}^n \sum_{j=1}^n x_j h_{ij} x_i \\
&= \sum_{i=1}^n \sum_{j=1}^n \int_0^1 (x_j t^{j-1})(x_i t^{i-1}) \mathrm{d}t \qquad (1.14)\\
&= \int_0^1 \left(\sum_{j=1}^n x_j t^{j-1} \right) \left(\sum_{i=1}^n x_i t^{i-1} \right) \mathrm{d}t \\
&= \int_0^1 w(t)^2 \mathrm{d}t > 0, \qquad w(t) := \sum_{i=1}^n x_i t^{i-1}
\end{aligned}$$

où on a conclu grâce au fait que, par hypothèse, w n'est pas la fonction identiquement nulle.

Exercice 1.55. *Soient A, $B \in \mathbb{R}^{n,n}$ deux matrices inversibles telles que $A+B$ est inversible. Montrer que $A^{-1}+B^{-1}$ est inversible et qu'on a*

$$\left(A^{-1}+B^{-1}\right)^{-1} = B(A+B)^{-1}A = A(A+B)^{-1}B.$$

On observe que

$$A^{-1}+B^{-1} = A^{-1}\left(I_n + AB^{-1}\right) = A^{-1}(B+A)B^{-1}.$$

Or, A, B et $A+B$ étant inversibles, nous avons que

$$B(A+B)^{-1}A = \left[A^{-1}(B+A)B^{-1}\right]^{-1} = (A^{-1}+B^{-1})^{-1},$$

ce qui prouve la première égalité. Pour prouver la deuxième on procède de manière analogue en factorisant A^{-1} à droite et B^{-1} à gauche.

Exercice 1.56 (Valeurs propres d'un polynôme à variable matricielle)**.** *Soit $A \in \mathbb{C}^{n,n}$. Montrer que, si $P(A) := \sum_{k=0}^N c_k A^k$ et $\mathrm{Sp}(A)$ est le spectre de A, on a $\mathrm{Sp}(P(A)) = P(\mathrm{Sp}(A))$.*

Par la suite, l'opérateur diag est utilisé comme la commande Matlab correspondante. De par le Théorème 1.38 (décomposition de Schur), il existe $U \in \mathbb{C}^{n,n}$ unitaire telle que $A = UTU^H$ avec T triangulaire supérieure contenant les valeurs propres de A sur sa diagonale. On peut exprimer ce dernier fait en écrivant (avec un petit abus de notation)

$$\mathrm{diag}(T) = \mathrm{Sp}(A). \qquad (1.15)$$

(i) On montre d'abord que

$$\mathrm{Sp}(P(T)) = P(\mathrm{Sp}(A)), \qquad (1.16)$$

à savoir, les valeurs propres de la matrice $P(T)$ sont de la forme $P(\lambda)$ avec λ valeur propre de A. En combinant la définition de T avec la Proposition 1.27 concernant les valeurs propres d'une matrice triangulaire, nous avons

$$\mathrm{Sp}(T) = \mathrm{Sp}(A).$$

D'autre part, nous avons aussi que (justifier la troisième égalité)

$$P(A) = \sum_{k=0}^{N} c_k A^k = \sum_{k=0}^{N} c_k (UTU^H)^k = \sum_{k=0}^{N} c_k (UT^kU^H) = \sum_{k=0}^{N} U(c_k T^k)U^H = UP(T)U^H.$$

Or, $P(T)$ est triangulaire supérieure car somme de matrices triangulaires supérieures (prouver que T^k est triangulaire supérieure pour tout $0 \le k \le N$), et

$$\mathrm{diag}(\mathrm{diag}(P(T))) = \mathrm{diag}(P(\mathrm{diag}(T))) = \mathrm{diag}(P(\mathrm{Sp}(A))),$$

où nous avons utilisé (1.15) pour conclure. De par la Proposition 1.27, nous avons donc que le spectre de $P(T)$ coïncide avec l'ensemble $\{P(\lambda) : \lambda \in \mathrm{Sp}(A)\}$, ce qui prouve (1.17).

(ii) Pour conclure, il suffit de montrer que

$$\mathrm{Sp}(P(T)) = \mathrm{Sp}(P(A)), \tag{1.17}$$

c.-à-d., que les valeurs propres de $P(T)$ coïncident avec celles de $P(A)$. En effet, en combinant (1.16) avec (1.17) on a immédiatement $\mathrm{Sp}(P(A)) = P(\mathrm{Sp}(A))$, qui est la conclusion cherchée. Soit $\lambda \in \mathrm{Sp}(P(A))$ une valeur propre de $P(A)$ et $x \in \mathbb{C}^n \setminus \{0 \in \mathbb{C}^n\}$ un vecteur propre associé. Nous avons que

$$P(A)x = \lambda x \implies UP(T)U^H x = \lambda x \implies P(T)U^H x = \lambda U^H x,$$

ce qui montre que $\lambda \in \mathrm{Sp}(P(T))$ et $U^H x$ est un vecteur propre associé. Par conséquent, $\mathrm{Sp}(P(A)) \subset \mathrm{Sp}(P(T))$. On prouve de manière similaire que $\mathrm{Sp}(P(T)) \subset \mathrm{Sp}(P(A))$, ce qui conclut l'exercice.

Exercice 1.57 (Valeurs singulières de l'inverse d'une matrice)**.** *Soit $A \in \mathbb{C}^{n,n}$ inversible. Exprimer les valeurs singulières de A^{-1} en fonction de celles de A.*

De par le Théorème 1.40 il existent deux matrices unitaires U, $V \in \mathbb{C}^{n,n}$ telles que $U^H AV = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ avec $(\sigma_i)_{1 \le i \le n}$ valeurs singulières de A. On a donc,

$$(U^H AV)^{-1} = \Sigma^{-1} \iff V^{-1}A^{-1}(U^H)^{-1} = \Sigma^{-1} \iff V^H A^{-1} U = \Sigma^{-1},$$

qui est la DVS de A^{-1}. Par conséquent $\Sigma^{-1} = \mathrm{diag}\left(\frac{1}{\sigma_1}, \ldots \frac{1}{\sigma_n}\right)$ contient les valeurs singulières de A^{-1}.

Exercice 1.58 (Inégalité de Kantorovitch)**.** *Soit $A \in \mathbb{R}^{n,n}$ SDP, et soit $\lambda_{\min}$ et $\lambda_{\max}$ respectivement la plus petite et la plus grande valeur propre de A. Alors, pour tout $x \in \mathbb{R}^n$,*

$$(Ax, x)(A^{-1}x, x) \le \frac{1}{4}\left(\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}\right)^2 \|x\|_2^4.$$

De par le Corollaire 1.39, il existe une matrice orthogonale U telle que

$$A = UDU^T = UDU^{-1}, \quad D = \mathrm{diag}(\lambda)_{\lambda \in \mathrm{Sp}(A)}.$$

Soit $t := \sqrt{\lambda_{\min}\lambda_{\max}}$. On a

$$\frac{1}{t}A + tA^{-1} = U\left(\frac{1}{t}D + tD^{-1}\right)U^{-1} := U\Delta U^{-1}, \tag{1.18}$$

avec $\Delta = \mathrm{diag}(\lambda/t + t/\lambda)_{\lambda\in\mathrm{Sp}(A)}$. La fonction $f(\xi) := \xi/t + t/\xi$ est convexe sur $\mathbb{R}_*^+$ et elle atteint son minimum en $\xi = t$, comme on le voit en résolvant

$$0 = f'(\xi) = \frac{1}{t} - \frac{t}{\xi^2} \iff \left(\xi^2 - t^2 = 0 \text{ et } t > 0\right) \iff \xi = t.$$

De plus, on a

$$\max_{\xi\in[\lambda_{\min},\lambda_{\max}]} f = f(\lambda_{\min}) = f(\lambda_{\max}) = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}.$$

Par conséquent, pour tout $\xi \in [\lambda_{\min}, \lambda_{\max}]$,

$$f(\xi) \le \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}. \tag{1.19}$$

On a donc, pour tout $x \in \mathbb{R}^n$,

$$\begin{aligned}
\sqrt{(Ax,x)(A^{-1}x,x)} &\le \frac{1}{2}\left(\frac{1}{t}(Ax,x) + t(A^{-1}x,x)\right) && \sqrt{ab} \le \tfrac{a}{2\epsilon} + \tfrac{b\epsilon}{2} \quad \forall a, b \ge 0, \forall \epsilon > 0 \\
&= \frac{1}{2}\left(\left(\frac{1}{t}A + tA^{-1}\right)x, x\right) \\
&= \frac{1}{2}(U\Delta U^{-1}x, x) && \text{Éq. (1.18)} \\
&= \frac{1}{2}(\Delta y, y) && y := U^T x,\ U^{-1} = U^T \\
&\le \frac{1}{2}\left(\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}\right)\|y\|_2^2 && \text{Éq. (1.19)} \\
&\le \frac{1}{2}\left(\sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} + \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}\right)\|x\|_2^2, && \text{Remarque (1.16)}
\end{aligned}$$

ce qui permet de conclure.

Exercice 1.59. *Soit A une matrice HDP et $\alpha \in \mathbb{R}_*^+$. Montrer que la matrice $I_n + \alpha A$ est inversible, que le rayon spectral de la matrice $B := (I_n - \alpha A)(I_n + \alpha A)^{-1}$ est < 1 et que la matrice B est hermitienne.*

La matrice A est hermitienne donc diagonalisable (voir Corollaire 1.39), à savoir, il existe une matrice unitaire $Q \in \mathbb{C}^{n,n}$ telle que $A = Q^{-1}\Lambda Q$, avec $\Lambda \in \mathbb{R}^{n,n}$ matrice diagonale contenant les valeurs propres de A (le fait que les valeurs propres de A sont des réels strictement positifs est une conséquence de la Proposition 1.26). Nous avons donc

$$I_n \pm \alpha A = Q^{-1}Q \pm \alpha Q^{-1}\Lambda Q = Q^{-1}(I_n \pm \alpha\Lambda)Q,$$

ce qui permet de conclure que la matrice $I_n \pm \alpha A$ est diagonalisable et ses valeurs propres sont de la forme $1 \pm \alpha\lambda$, $\lambda \in \mathrm{Sp}(A)$. De par le Théorème 1.33, la matrice $I_n + \alpha A$ est inversible si et seulement si toutes ses valeurs propres sont non nulles, à savoir $\alpha \ne -1/\lambda$ pour tout $\lambda \in \mathrm{Sp}(A)$, ce qui est toujours vérifié car $\alpha > 0$ par hypothèse. De plus,

$$(I_n + \alpha A)^{-1} = \left[Q^{-1}(I_n + \alpha\Lambda)Q\right]^{-1} = Q^{-1}(I_n + \alpha\Lambda)^{-1}Q,$$

et $(I_n+\alpha\Lambda)^{-1}=\mathrm{diag}\left(\frac{1}{1+\alpha\lambda}\right)_{\lambda\in\mathrm{Sp}(A)}$. De par les résultats précédents, on a

$$B=(I_n-\alpha A)(I_n+\alpha A)^{-1}=Q^{-1}(I_n-\alpha\Lambda)QQ^{-1}(I_n+\alpha\Lambda)^{-1}Q=Q^{-1}\tilde{\Lambda}Q, \tag{1.20}$$

avec $\tilde{\Lambda}=\mathrm{diag}(\tilde{\lambda})_{\tilde{\lambda}\in\mathrm{Sp}(B)}=\mathrm{diag}\left(\frac{1-\alpha\lambda}{1+\alpha\lambda}\right)_{\lambda\in\mathrm{Sp}(A)}$ matrice diagonale contenant les valeurs propres de B. Comme les valeurs propres de A sont >0, on vérifie aisément que $|\tilde{\lambda}|<1$ pour tout $\tilde{\lambda}\in\mathrm{Sp}(B)$ si $\alpha>0$ et, par conséquent, $\rho(B):=\max_{\tilde{\lambda}\in\mathrm{Sp}(B)}|\tilde{\lambda}|<1$. Pour prouver que B est hermitienne, on utilise la décomposition (1.20) pour écrire

$$B^H=(Q^H\tilde{\Lambda}Q)^H=Q^H\tilde{\Lambda}^HQ=Q^H\tilde{\Lambda}Q=B,$$

où nous avons utilisé le fait que $\tilde{\Lambda}$ est une matrice diagonale à valeurs réells pour conclure $\tilde{\Lambda}^H=\tilde{\Lambda}$.

Exercice 1.60 (Pséudo-inverse de Moore–Penrose)**.** *Soient m et n deux entiers avec $m\geq n$, et soit $A\in\mathbb{R}^{m,n}$ une matrice de rang n. On admettra par la suite que $A^TA\in\mathbb{R}^{n,n}$ est définie positive et donc inversible. On définit la* pseudo-inverse de Moore–Penrose *par*

$$A^\dagger:=(A^TA)^{-1}A^T.$$

Prouver les propriétés suivantes :

$$A^\dagger A=I_n,\qquad A^\dagger AA^\dagger=A^\dagger,\qquad \text{si } m=n,\ A^\dagger=A^{-1};$$

Soit $b\in\mathbb{R}^m$. On considère le problème dit aux moindres carrés

$$\min_{y\in\mathbb{R}^n}\left\{J(y):=\|Ay-b\|_2^2\right\}.$$

Ce problème admet une unique solution x caractérisée par la propriété $\nabla J(y)=0$. Montrer que $x=A^\dagger b$.

On a $A^\dagger A=(A^TA)^{-1}(A^TA)=I_n$, d'où, en multipliant à droite par $A^\dagger$, $A^\dagger AA^\dagger=I_nA^\dagger=A^\dagger$. Pour prouver le deuxième point on observe que, pour tout $y\in\mathbb{R}^n$,

$$\begin{aligned}J(y)&=(Ay-b)^T(Ay-b)\\&=(Ay)^TAy-(Ay)^Tb-b^TAy+b^Tb\\&=y^T(A^TA)y-y^TA^Tb-b^TAy+b^Tb.\\&=y_j(A^TA)_{ij}y_i-y_i(A^Tb)_i-(b^TA)_iy_i+b_ib_i\end{aligned}$$

où par brévité nous avons utilisé la notation de Einstein qui sous-entend les sommes sur les indices repétés. Par conséquent, pour tout $1\leq k\leq n$,

$$\frac{\partial J(y)}{\partial y_k}=(A^TA)_{ik}y_i+y_j(A^TA)_{kj}-(A^Tb)_k-(b^TA)_k,$$

à savoir,

$$\nabla J(y)=2(A^TA)y-2A^Tb.$$

La condition d'optimalité donne

$$\nabla J(x)=2(A^TA)x-2A^Tb=0 \iff x=(A^TA)^{-1}A^Tb=A^\dagger b.$$

Exercice 1.61 (Valeurs singulières d'une matrice normale)**.** *Soit $A \in \mathbb{C}^{n,n}$ une matrice normale. Montrer que les valeurs singulières de A sont les modules de ses valeurs propres.*

Exercice 1.62 (Inégalité triangulaire)**.** *Soit $\|\cdot\|$ une norme matricielle subordonnée. Prouver que pour tout A, $B \in \mathbb{C}^{n,n}$,*

$$\|A + B\| \le \|A\| + \|B\|.$$

Montrer que toute norme subordonnée satisfait également les deux premières propriétés caractérisant une norme matricielle (voir Definition 1.43).

On a

$$\begin{aligned}\|A+B\| &:= \sup_{x\in\mathbb{C}^n,\,\|x\|=1} \|(A+B)x\| = \sup_{x\in\mathbb{C}^n,\,\|x\|=1} \|Ax+Bx\| \\ &\le \sup_{x\in\mathbb{C}^n,\,\|x\|=1} (\|Ax\|+\|Bx\|) \\ &\le \sup_{x\in\mathbb{C}^n,\,\|x\|=1} \|Ax\| + \sup_{x\in\mathbb{C}^n,\,\|x\|=1} \|Bx\| := \|A\|+\|B\|,\end{aligned}$$

où nous avons utilisé l'inégalité triangulaire pour la norme vectorielle pour passer à la deuxième ligne. Le second point est laissé au lecteur.

Exercice 1.63 (Norme d'une matrice unitaire)**.** *Soit $U \in \mathbb{C}^{n,n}$ une matrice unitaire. Prouver que $\|U\|_2 = 1$ et que pour tout $A \in \mathbb{C}^{n,n}$ on a $\|AU\|_2 = \|UA\|_2 = \|A\|_2$.*

Soit $x \in \mathbb{C}^n$. Par définition on a $\|Ux\|_2^2 = (Ux)^H(Ux) = x^H U^H U x = x^H x = \|x\|_2^2$. Par conséquent,

$$\|U\|_2 = \sup_{x\in\mathbb{C}^n,\,\|x\|_2=1} \|Ux\|_2 = 1.$$

Un raisonnement similaire montre que $\|U^H\|_2 = 1$. D'autre part,

$$\|AU\|_2 = \sup_{x\in\mathbb{C}^n,\,\|x\|_2=1} \|AUx\|_2 = \sup_{y\in\mathbb{C}^n,\,\|U^{-1}y\|_2=1} \|Ay\|_2 = \sup_{y\in\mathbb{C}^n,\,\|y\|_2=1} \|Ay\|_2,$$

où nous avons utilisé le point précédent pour conclure $\|U^{-1}y\|_2 = \|U^H y\|_2 = \|y\|_2$.

Exercice 1.64 (Norme de Frobenius)**.** *Soit $A = (a_{ik}) \in \mathbb{R}^{n,n}$. La norme de Frobenius est définie par*

$$\|A\|_{\mathrm{F}} := \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2}.$$

Cette norme n'est pas subordonnée à une norme vectorielle. Prouver que

$$\frac{1}{n}\|A\|_1 \le \|A\|_{\mathrm{F}} \le \sqrt{n}\|A\|_1, \qquad \frac{1}{n}\|A\|_\infty \le \|A\|_{\mathrm{F}} \le \sqrt{n}\|A\|_\infty. \tag{1.21}$$

Calculer $\|I_n\|_{\mathrm{F}}$.

(i) Équivalence entre la norme de Frobenius et la norme 1. Il est utile de rappeler que

$$\|A\|_{\mathrm{F}}^2 := \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2, \quad \|A\|_1 := \max_{1\le j\le n} \sum_{i=1}^n |a_{ij}|, \quad \|A\|_\infty := \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}|.$$

Or,

$$\|A\|_1 = \max_{1\le j\le n}\sum_{i=1}^n |a_{ij}| \le \sum_{j=1}^n\sum_{i=1}^n 1\times|a_{ij}| \le \left(\sum_{j=1}^n\sum_{i=1}^n 1^2\right)^{\frac{1}{2}} \times \left(\sum_{j=1}^n\sum_{i=1}^n |a_{ij}|^2\right)^{\frac{1}{2}} = n\|A\|_{\mathrm{F}},$$

ou nous avons utilisé l'inégalité de Cauchy–Schwarz dans le deuxième passage. Pour prouver la borne supérieure, on observe que

$$\|A\|_{\mathrm{F}}^2 = \sum_{i=1}^n\sum_{j=1}^n |a_{ij}|^2 \le n\max_{1\le j\le n}\sum_{i=1}^n |a_{ij}|^2 \le n\max_{1\le j\le n}\left(\sum_{i=1}^n |a_{ij}|\right)^2 = n\|A\|_1^2,$$

d'où $\|A\|_{\mathrm{F}} \le \sqrt{n}\|A\|_1$.

(ii) Équivalence entre la norme de Frobenius et la norme ∞. On commence par remarquer que, pour toute matrice $A\in\mathbb{R}^{n,n}$, on a $\|A\|_{\mathrm{F}} = \|A^T\|_{\mathrm{F}}$ et $\|A\|_\infty = \|A^T\|_1$. En utilisant la réponse donnée au point précédent on a donc

$$\frac{1}{n}\|A^T\|_1 \le \|A^T\|_{\mathrm{F}} = \|A\|_{\mathrm{F}} \le \sqrt{n}\|A^T\|_1 \implies \frac{1}{n}\|A\|_\infty \le \|A\|_{\mathrm{F}} \le \sqrt{n}\|A\|_\infty.$$

(iii) Norme de Frobenius de la matrice identité. On a $\|I_n\|_{\mathrm{F}} = \sqrt{n}$.

Exercice 1.65 (Rayon spectral)**.** *Soit $A\in\mathbb{R}^{n,n}$ une matrice SDP. Montrer que $\rho(A)<1$ si et seulement s'il existe une matrice $Q\in\mathbb{R}^{n,n}$ SDP telle que $B := Q - A^TQA$ est SDP.*

(i) Preuve de $\rho(A)<1 \implies (\exists Q\in\mathbb{R}^{n,n}, B$ SDP$)$. Il suffit de prendre $Q = I_n$. En effet, avec ce choix on a pour tout $x\in\mathbb{R}^n\setminus\{0\in\mathbb{R}^n\}$,

$$\begin{aligned}
x^TBx &= x^T\left(I_n - A^TI_nA\right)x \\
&= \|x\|_2^2 - \|Ax\|_2^2 \\
&\ge \|x\|_2^2 - \|A\|_2^2\|x\|_2^2 && \text{Proposition 1.47} \\
&= \left(1-\rho(A)^2\right)\|x\|_2^2 > 0, && \text{Corollaire 1.42}
\end{aligned}$$

ce qui montre le caractère défini positif de B.

(ii) Preuve de $(\exists Q\in\mathbb{R}^{n,n}, B$ SDP$) \implies \rho(A)<1$. Soit $x\in\mathbb{R}^n\setminus\{0\in\mathbb{R}^n\}$ et posons $y := Ax$. On a par hypothèse

$$x^TBx>0 \implies x^T(Q-A^TQA)x>0 \implies x^TQx - x^TA^TQAx > 0 \implies y^TQy < x^TQx. \quad (1.22)$$

Soit $\overline{\lambda}$ la plus grande valeur propre (en valeur absolue) de A, $\overline{x}\in\mathbb{R}^n$ un vecteur propre associé avec $\|\overline{x}\|_2 = 1$, et $\overline{y} := A\overline{x} = \overline{\lambda}\overline{x}$. De par la rélation (1.22) on a

$$\overline{x}^TQ\overline{x} > \overline{y}^TQ\overline{y} = \overline{\lambda}^2\overline{x}^TQ\overline{x},$$

ce qui prouve $\overline{\lambda}<1$ et, par conséquent, $\rho(A)<1$.

2.1.2 Conditionnement

Proposition 2.2. *Soit $A \in \mathbb{C}^{n,n}$ telle que $\rho(A) < 1$. Alors, la matrice $I_n - A$ est inversible et on a pour toute norme matricielle subordonnée telle que $\|A\| < 1$,*

$$\frac{1}{1+\|A\|} \leq \|(I_n - A)^{-1}\| \leq \frac{1}{1-\|A\|}. \tag{2.3}$$

Démonstration. La preuve procède en trois étapes.

(i) On commence par prouver que $I_n - A$ est inversible. Les valeurs propres de A sont les solutions dans $\mathbb{C}$ de l'équation caractéristique

$$\det(A - \lambda I_n) = 0.$$

En ajoutant et soustrayant I_n à l'argument du déterminant, on voit que λ est solution de cette équation si et seulement si

$$\det((A - I_n) + (1-\lambda) I_n) = 0,$$

ce qui montre que

$$\mathrm{Sp}(A - I_n) = \{1 - \lambda \,:\, \lambda \in \mathrm{Sp}(A)\}.$$

Or, on a que

$$|1-\lambda| \geq 1 - |\lambda| \geq 1 - \varrho(A) > 0,$$

où nous avons utilisé la Définition 1.28 du rayon spectral dans la deuxième inégalité et le fait que $\varrho(A) \leq \|A\| < 1$ (cf. Lemma 1.50) pour conclure.

(ii) On prouve ensuite la première inégalité de (2.3). On a que

$$\begin{aligned}
1 &= \|I_n\| && \text{Proposition 1.49(ii)}\\
&= \|(I_n - A)(I_n - A)^{-1}\| && \text{Définition 1.18}\\
&\leq \|I_n - A\| \; \|(I_n - A)^{-1}\| && \text{Proposition 1.49(iii)}\\
&\leq (\|I_n\| + \|A\|) \; \|(I_n - A)^{-1}\| && \text{Inégalité triangulaire}\\
&= (1 + \|A\|) \; \|(I_n - A)^{-1}\| && \text{Proposition 1.49(ii)}
\end{aligned}$$

à savoir,

$$\frac{1}{1+\|A\|} \leq \|(I_n - A)^{-1}\|.$$

(iii) Pour prouver la deuxième inégalité de (2.3), il suffit d'observer que

$$(I_n - A)^{-1} = (I_n - A)^{-1} I_n = (I_n - A)^{-1} (I_n - A + A) = I_n + (I_n - A)^{-1} A,$$

d'où, en passant à la norme et en utilisant l'inégalité triangulaire et à nouveau le point (ii) de la Proposition 1.49,

$$\begin{aligned}
\|(I_n - A)^{-1}\| &\leq \|I_n\| + \|(I_n - A)^{-1} A\|\\
&\leq 1 + \|(I_n - A)^{-1}\| \; \|A\| \iff \|(I_n - A)^{-1}\| \leq \frac{1}{1-\|A\|}. \quad \square
\end{aligned}$$

Théorème 2.3 (Estimation d'erreur)**.** *Soit $A \in \mathbb{R}^{n,n}$ une matrice inversible et $\delta A \in \mathbb{R}^{n,n}$ une perturbation telle que pour une norme matricielle subordonnée $\|\cdot\|$,*

$$\|A^{-1}\|\ \|\delta A\| < 1. \tag{2.4}$$

On note $x \in \mathbb{R}^n$ la solution du système linéaire $Ax = b$ avec $b \in \mathbb{R}^n \setminus \{0\}$ et, pour une perturbation du seconde membre $\delta b \in \mathbb{R}^n$, soit $\delta x \in \mathbb{R}^n$ tel que

$$(A+\delta A)(x+\delta x) = b+\delta b.$$

Alors,

$$\frac{\|\delta x\|}{\|x\|} \le \frac{\mathrm{cond}(A)}{1-\mathrm{cond}(A)\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta b\|}{\|b\|}+\frac{\|\delta A\|}{\|A\|}\right). \tag{2.5}$$

Remarque 2.4 (Hypothèse (2.4))**.** *L'hypothèse* (2.4) *a une interprétation naturelle dans la mesure où elle implique $\|A^{-1}\delta A\| < 1$, à savoir, les erreurs d'arrondi sur la matrice A sont supposées petites par rapport aux valeurs des entrées de A.*

Démonstration. Comme $b \neq 0$ et A est inversible, $x \neq 0 \implies \|x\| \neq 0$. De plus, la matrice $I_n + A^{-1}\delta A$ est inversible car, par hypothèse, $\rho(A^{-1}\delta A) \le \|A^{-1}\delta A\| \le \|A^{-1}\|\|\delta A\| < 1$ (la première inégalité est une conséquence du Lemme 1.50). On peut donc appliquer le résultat de la Proposition 2.2 à la matrice $-A^{-1}\delta A$, obtenant ainsi

$$\begin{aligned}
& (A+\delta A)(x+\delta x) = b+\delta b && \text{définition de } \delta x\\
\implies & \cancel{Ax} + \delta A x + A(I_n + A^{-1}\delta A)\delta x = \cancel{b} + \delta b && Ax=b,\ A+\delta A = A(I_n+A^{-1}\delta A)\\
\implies & \frac{\delta x}{\|x\|} = (I+A^{-1}\delta A)^{-1}A^{-1}\left(\frac{\delta b}{\|x\|} - \frac{\delta A x}{\|x\|}\right) && \text{mult. à gauche par } \frac{(I_n+A^{-1}\delta A)^{-1}A^{-1}}{\|x\|}\\
\implies & \frac{\|\delta x\|}{\|x\|} \le \|(I+A^{-1}\delta A)^{-1}\|\ \|A^{-1}\|\left(\frac{\|\delta b\|}{\|x\|} + \|\delta A\|\right) && \|\delta A x\| \le \|\delta A\|\ \|x\|\\
\implies & \frac{\|\delta x\|}{\|x\|} \le \|(I+A^{-1}\delta A)^{-1}\|\,\mathrm{cond}(A)\left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right) && \|A\|\ \|x\| \ge \|Ax\| = \|b\|\\
\implies & \frac{\|\delta x\|}{\|x\|} \le \frac{\mathrm{cond}(A)}{1-\|A^{-1}\delta A\|}\left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right) && \text{Proposition 2.2}\\
\implies & \frac{\|\delta x\|}{\|x\|} \le \frac{\mathrm{cond}(A)}{1-\mathrm{cond}(A)\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta b\|}{\|b\|} + \frac{\|\delta A\|}{\|A\|}\right). && (2.4)
\end{aligned}$$

□

Remarque 2.5. *Si les perturbations δb et δA sont telles que $\|\delta b\| = \gamma\|b\|$ et $\|\delta A\| = \gamma\|A\|$ avec $\gamma \in \mathbb{R}^+_*$ et $\gamma\,\mathrm{cond}(A) < 1$, l'estimation* (2.5) *se reduit à*

$$\frac{\|\delta x\|}{\|x\|} \le \frac{2\gamma\,\mathrm{cond}(A)}{1-\gamma\,\mathrm{cond}(A)}.$$

En pratique cette hypothèse est raisonnable, et le paramètre γ depend de l'epsilon machine, à savoir, le plus petit nombre positif ϵ tel que $1+\epsilon > 1$ en arithmétique flottante.

2.6 Exercices

Exercice 2.20 (Conditionnement)**.** *Soit* $A \in \mathbb{C}^{n,n}$ *une matrice inversible. Prouver que*

(i) *si* $\|\cdot\|$ *est une norme matricielle subordonnée, alors* $\mathrm{cond}(A) = \mathrm{cond}(A^{-1}) \geq 1$, $\mathrm{cond}(\alpha A) = \mathrm{cond}(A)$ *pour tout* $\alpha \in \mathbb{R}^*$ *;*

(ii) $\mathrm{cond}_2(A) = \sigma_{\max}(A)/\sigma_{\min}(A)$ *où* $\sigma_{\min}(A)$ *et* $\sigma_{\max}(A)$ *sont respectivement la plus petite et la plus grande valeur singulière de* A *;*

(iii) *si* A *est normale,* $\mathrm{cond}_2(A) = |\lambda_{\max}(A)|/|\lambda_{\min}(A)|$ *où* $\lambda_{\min}(A)$ *et* $\lambda_{\max}(A)$ *sont respectivement la plus petite et la plus grande valeur propre en module de* A *;*

(iv) *pour toute matrice unitaire* U, $\mathrm{cond}_2(U) = 1$ *et* $\mathrm{cond}_2(AU) = \mathrm{cond}_2(UA) = \mathrm{cond}_2(A)$.

(i) On commence par remarquer que $\mathrm{cond}\, A^{-1} = \|(A^{-1})^{-1}\|\, \|A^{-1}\| = \|A\|\, \|A^{-1}\| = \mathrm{cond}\, A$. Ensuite, puisque $\|\cdot\|$ est subordonnée, nous avons

$$1 = \|I_n\| = \|A^{-1}A\| \leq \|A^{-1}\|\|A\|_2 = \mathrm{cond}(A),$$

à savoir, $\mathrm{cond}(A) \geq 1$. D'autre part, comme

$$\|\alpha A\| = \sup_{x \in \mathbb{C}^n, \|x\|=1} \|\alpha A x\| = |\alpha| \sup_{x \in \mathbb{C}^n, \|x\|=1} \|Ax\| = |\alpha|\|A\|,$$

et que $(\alpha A)^{-1} = \alpha^{-1}A^{-1}$, nous avons d'après l'Exercice 1.63,

$$\mathrm{cond}(\alpha A) = \|\alpha A\|\|(\alpha A)^{-1}\| = |\alpha|\|A\|\frac{1}{|\alpha|}\|A^{-1}\| = \mathrm{cond}(A).$$

(ii) On peut diagonaliser A à l'aide de deux matrices unitaires U, $V \in \mathbb{C}^{n.n}$ (cf. Theorem 1.40) comme suit :

$$U^H AV = \Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_n).$$

Comme U et V sont unitaires nous avons que $\|U^H AV\|_2 = \|A\|_2$ et $\|(U^H AV)^{-1}\|_2 = \|V^H A^{-1} U\|_2 = \|A^{-1}\|_2$. Par conséquent,

$$\mathrm{cond}_2(\Sigma) = \mathrm{cond}_2(U^H AV) = \|U^H AV\|_2\|(U^H AV)^{-1}\|_2 = \|A\|_2\|A^{-1}\|_2 = \mathrm{cond}_2(A).$$

D'autre part, puisque Σ est diagonal, $\mathrm{cond}_2(\Sigma) = \sigma_{\max}(A)/\sigma_{\min}(A)$, ce qui conclut la preuve.

(iii) Puisque A est normale, de par le Lemme 1.37 elle admet une décomposition de la forme

$$W^H AW = \Lambda = \mathrm{diag}(\lambda_1(A), \ldots, \lambda_n(A)),$$

avec W matrice unitaire. Par conséquent, $A = W\Lambda W^H$ et

$$AA^H = A^H A = (W\Lambda W^H)^H (W\Lambda W^H) = W\Lambda^H W^H W \Lambda W^H = W\Lambda^H \Lambda W^H,$$

à savoir, $\lambda_i(A^H A) = \overline{\lambda}_i(A)\lambda_i(A) = |\lambda_i(A)|^2$. Pour s'en convaincre, il suffit de multiplier à droite l'inégalité ci-dessus par W et utiliser le fait que $W^H W = I_n$, ce qui donne

$$AA^H W = W\,\mathrm{diag}(|\lambda_i(A)|^2, \ldots, |\lambda_n(A)|^2).$$

On remarquera aussi que les colonnes de W sont des vecteurs propres de AA^H. De par la Proposition 1.41 on a, de plus, que $\sigma_i(A)^2 = \lambda_i(A^H A) = |\lambda_i(A)|^2$, ce qui permet de conclure.

(iv) Conséquence de l'Exercice 1.63.

Exercice 2.21 (Conditionnement de la matrice de l'Exemple 2.19)**.** *L'objectif de cet exercice est d'étudier le conditionnement de la matrice A définie dans l'Exemple 2.19 en fonction de sa taille.*

(i) Spectre. *Vérifier que les valeurs propres de la matrice sont*

$$\lambda_k = \frac{4}{h^2}\sin^2\left(\frac{k\pi}{2(N+1)}\right) \qquad 1 \le k \le N,$$

avec vecteurs propres

$$u^k = (u_i^k)_{1\le i\le N} = \left(\sin\left(\frac{ik\pi}{N+1}\right)\right)_{1\le i\le N}.$$

(ii) Conditionnement. *Préciser le comportement asymptotique du conditionnement de la matrice A en norme* $\|\cdot\|_2$ *pour h petit.*

(i) Spectre. On a pour tout $1 \le k \le N$ et tout $1 \le i \le N$, en rappelant que nous avons posé par convention $u_0^k = \sin(0) = 0$ (relation utile pour $i = 1$) et $u_{N+1}^k = \sin(k\pi) = 0$ (relation utile pour $i = N$),

$$\begin{aligned}(Au^k)_i &= \frac{1}{h^2}\left(-u_{i-1}^k + 2u_i^k - u_{i+1}^k\right)\\ &= \frac{1}{h^2}\left(-\sin\left(\frac{(i-1)k\pi}{N+1}\right) + 2\sin\left(\frac{ik\pi}{N+1}\right) - \sin\left(\frac{(i+1)k\pi}{N+1}\right)\right).\end{aligned}$$

Remarquons, maintenant, que

$$\begin{aligned}\sin\left(\frac{(i-1)k\pi}{N+1}\right) + \sin\left(\frac{(i+1)k\pi}{N+1}\right) &= \sin\left(\frac{ik\pi}{N+1} - \frac{k\pi}{N+1}\right) + \sin\left(\frac{ik\pi}{N+1} + \frac{k\pi}{N+1}\right)\\ &= 2\sin\left(\frac{ik\pi}{N+1}\right)\cos\left(\frac{k\pi}{N+1}\right),\end{aligned}$$

où nous avons utilisé l'identité trigonométrique $\sin(\alpha+\beta)+\sin(\alpha-\beta) = 2\sin(\alpha)\cos(\beta)$. Ceci nous permet d'écrire

$$(Au^k)_i = \frac{2}{h^2}\left(1-\cos\left(\frac{k\pi}{N+1}\right)\right)\sin\left(\frac{ik\pi}{N+1}\right) = \frac{4}{h^2}\sin^2\left(\frac{k\pi}{2(N+1)}\right)\sin\left(\frac{ik\pi}{N+1}\right) = \lambda_k u_i^k,$$

où nous avons utilisé l'identité trigonométrique $\sin^2(\alpha) = (1-\cos(2\alpha))/2$. Nous avons donc vérifié que λ_k est valeur propre de la matrice A et que u^k est un vecteur propre associé. Puisque nous avons trouvé N valeurs propres distinctes, nous avons trouvé l'intégralité du spectre de A.

(ii) Conditionnement. La fonction sinus est strictement croissante dans l'intervalle $[0,\pi/2]$. Comme pour tout entier $1 \le k \le N$ nous avons $\frac{k}{N+1}\frac{\pi}{2} \in (0,\pi/2)$, on peut conclure que $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_N$. De plus, comme la matrice A est symétrique et définie positive (car ses valeurs propres sont des réels strictement positifs), nous avons (voir l'Exercice 2.20) :

$$\operatorname{cond}_2(A) = \frac{\lambda_N}{\lambda_1} = \frac{\sin^2\left((1-h)\frac{\pi}{2}\right)}{\sin^2\left(h\frac{\pi}{2}\right)} \sim \frac{1}{h^2}.$$

Chapitre 2

Méthodes itératives

Dans ce chapitre nous allons présenter quelques exemples de méthode *itératives* pour la résolution du système linéaire

$$Ax = b. \tag{3.1}$$

Nous allons nous restreindre au cas où A est une matrice carrée (réelle ou complexe) inversible et b un vecteur non nul. Contrairement aux méthodes directes, les méthodes itératives ne fournissent pas en général la solution exacte du système linéaire : elles consistent à engendrer une suite $(x_n)_{n\in\mathbb{N}}$ de solutions approchées qui tend vers x. Pour que les méthodes itératives soient intéressantes, deux conditions doivent être remplies :

(i) la génération d'un nouvel élément de la suite à partir du (des) précédent(s) est significativement moins coûteuse que résoudre le système linéaire par une méthode directe ;

(ii) le nombre d'itérations n nécessaires pour arriver à un élément x_n qui approche x à la précision souhaitée est «petit».

3.1 Méthodes de point fixe

La première famille de méthodes que nous allons considérer repose sur la caractérisation de la solution du système linéaire (3.1) comme point fixe d'une fonction obtenue à partir de la matrice A et du seconde membre b.

3.1.1 Formulation abstraite basée sur une décomposition régulière

Soit $A \in \mathbb{C}^{n,n}$ une matrice inversible, $b \in \mathbb{C}^n$ non nul, et P, $N \in \mathbb{C}^{n,n}$ deux matrices telles que

$$A = P - N \text{ et } P \text{ est facilement inversible.} \tag{3.2}$$

Une décomposition de la forme (3.2) est dite *régulière*. On vérifie aisément que la solution du système est point fixe de la fonction $\Phi : y \to P^{-1}Ny + P^{-1}b$. En effet,

$$Ax = (P - N)x = b \iff x = P^{-1}Nx + P^{-1}b = \Phi(x). \tag{3.3}$$

Cette remarque nous conduit à considérer la méthode itérative suivante : Pour tout $k \geq 0$,

$$x^{(k+1)} = P^{-1}Nx^{(k)} + P^{-1}b = \Phi(x^{(k)}), \tag{3.4}$$

avec $x^{(0)} \in \mathbb{C}^n$ estimation initiale. A chaque itération de la méthode (3.4) on résout un système linéaire de matrice P, ce qui justifie l'hypothèse que P soit facile à inverser. Par exemple, si on choisit P diagonale ou triangulaire, le coût lié à l'inversion sera de l'ordre de n ou n^2 opérations respectivement.

Remarque 3.1 (Hypothèse sur P)**.** *Nous avons supposé dans* (3.2) P facilement *inversible. En effet, le choix $A = P$ (qui est toujours possible car A est inversible) donne une méthode itérative qui converge en une itération, mais dont la complexité équivaut à celle de la résolution du système linéaire $Ax = b$. Le choix de P doit donc représenter un compromis qui garantisse que le coût d'inversion de P à chaque itération reste raisonnable tout en évitant que le nombre d'itérations pour arriver à convergence explose.*

Le résultat suivant est une conséquence naturelle de l'interprétation de la méthode (3.4) comme recherche d'un point fixe.

Lemme 3.2 (Condition nécessaire et suffisante pour la convergence de la méthode (3.4))**.** *La méthode* (3.4) *converge pour toute estimation initiale $x^{(0)}$ vers l'unique solution x du système linéaire* (3.1) *si et seulement si $\rho(P^{-1}N) < 1$.*

Démonstration. Pour tout $k \geq 0$, soit $e^{(k)} := x^{(k)} - x$ l'erreur associée à l'estimation $x^{(k)}$. De par (3.3) et (3.4) nous avons pour $k \geq 1$,

$$e^{(k)} = x^{(k)} - x = \Phi(x^{(k-1)}) - \Phi(x) = P^{-1}N(x^{(k-1)} - x) = P^{-1}Ne^{(k-1)} = (P^{-1}N)^k e^{(0)}. \tag{3.5}$$

(i) $\rho(P^{-1}N) < 1 \implies$ **convergence pour tout** $x^{(0)}$**.** Comme $\rho(P^{-1}N) < 1$, il existe une norme matricielle subordonnée $\|\cdot\|$ telle que $\|P^{-1}N\| < 1$ (il s'agit d'une conséquence du Lemme 1.50). En prenant la norme de (3.5) on obtient

$$\|e^{(k)}\| = \|(P^{-1}N)^k e^{(0)}\| \leq \|(P^{-1}N)^k\| \|e^{(0)}\| \leq \|P^{-1}N\|^k \|e^{(0)}\|, \tag{3.6}$$

où nous avons utilisé la Proposition 1.47 dans la première inégalité et le point (iii) de la Proposition 1.49 pour écrire $\|(P^{-1}N)^k\| \leq \|P^{-1}N\|^k$. Si $\|P^{-1}N\| < 1$, le seconde membre de (3.6) tend vers 0 lorsque $k \to +\infty$ et donc la méthode converge pour tout $x^{(0)}$.

(ii) Convergence pour tout $x^{(0)} \implies \rho(P^{-1}N) < 1$**.** Supposons maintenant que la méthode converge pour toute estimation initiale $x^{(0)}$. Soit $\|\cdot\|$ une norme vectorielle et $\|\cdot\|$ la norme matricielle subordonnée. Soit λ la plus grande valeur propre en valeur absolue de $P^{-1}N$ et x_λ vecteur propre associé avec $\|x_\lambda\| = 1$. On considère la méthode avec estimation initiale $x^{(0)}$ telle que $e^{(0)} := x^{(0)} - x = x_\lambda$. De par (3.5), nous avons que

$$e^{(k)} = (P^{-1}N)^k e^{(0)} = (P^{-1}N)^k x_\lambda = \lambda^k x_\lambda.$$

Prenant la norme de l'expression précédente,

$$\|e^{(k)}\| = |\lambda|^k.$$

Comme la méthode converge, $\|e^{(k)}\| \to 0$ lorsque $k \to +\infty$, ce qui implique $\rho(P^{-1}N) = |\lambda| < 1$. □

Remarque 3.3 (Choix de la norme dans la preuve du Lemme 3.2)**.** *Dans la preuve du Lemme 3.2 nous avons utilisé une norme* $\|\cdot\|$ *subordonnée satisfaisant la condition* $\|P^{-1}N\| < 1$. *Cependant, comme que* $e^{(k)} \to 0 \in \mathbb{R}^n$ *pour* $k \to +\infty$, *pour toute norme* $\|\cdot\|_*$ *sur* $\mathbb{R}^n$, *nous avons* $\|e^{(k)}\|_* \to 0$ *lorsque* $k \to +\infty$.

La condition du Lemme 3.2 est optimale mais difficile à vérifier en pratique. Une condition suffisante souvent plus simple à prouver est identifiée dans le lemme suivant.

Lemme 3.4 (Condition suffisante pour la convergence de la méthode (3.4))**.** *Soit* $A \in \mathbb{C}^{n,n}$ *hermitienne définie positive. On considère une décomposition régulière de la forme* (3.2) *avec* $P \neq A$ *et* $P^H + N$ *définie positive. Alors la méthode* (3.4) *converge vers la solution de* (3.1).

Démonstration. On commence par remarquer que

$$P^H + N = P^H + P - A. \tag{3.7}$$

Considérons maintenant la norme matricielle $\|\cdot\|_A$ subordonnée à la norme vectorielle définie par le produit scalaire $(x, y)_A := (Ax, y)_{\mathbb{C}^n}$ (que l'on notera toujours $\|\cdot\|_A$). De par la Proposition 1.49, il existe $y \in \mathbb{C}^n$ tel que $\|y\|_A = (Ay, y)_{\mathbb{C}^n} = 1$ et $\|P^{-1}N\|_A = \|P^{-1}Ny\|_A$. Nous avons alors

$$\begin{aligned}
&\|P^{-1}N\|_A^2 \\
&\quad = \|P^{-1}Ny\|_A^2 \\
&\quad = (AP^{-1}Ny, P^{-1}Ny)_{\mathbb{C}^n} \\
&\quad = (AP^{-1}(P-A)y, P^{-1}(P-A)y)_{\mathbb{C}^n} && (N = P - A) \\
&\quad = ((A - AP^{-1}A)y, (I_n - P^{-1}A)y)_{\mathbb{C}^n} \\
&\quad = (Ay, y)_{\mathbb{C}^n} - (AP^{-1}Ay, y)_{\mathbb{C}^n} - (Ay, P^{-1}Ay)_{\mathbb{C}^n} + (AP^{-1}Ay, P^{-1}Ay)_{\mathbb{C}^n} \\
&\quad = 1 - (Az, A^{-1}Pz)_{\mathbb{C}^n} - (Pz, z)_{\mathbb{C}^n} + (Az, z)_{\mathbb{C}^n} && (z := P^{-1}Ay) \\
&\quad = 1 - (P^H z, z)_{\mathbb{C}^n} - (Pz, z)_{\mathbb{C}^n} + (Az, z)_{\mathbb{C}^n} && (A^H = A) \\
&\quad = 1 - ((P^H + P - A)z, z)_{\mathbb{C}^n} = 1 - ((P^H + N)z, z)_{\mathbb{C}^n} < 1, && (3.7)
\end{aligned}$$

où nous avons utilisé le fait que $P^{-1}A$ est une matrice de rang plein et $y \neq 0$ pour conclure que $z \neq 0$ et le caractére défini positif de $P^H + N$ pour en déduire $((P^H + N)z, z)_{\mathbb{C}^n} > 0$. □

On conclut cette section en observant qu'une généralisation de (3.4) consiste à considérer des méthodes de la forme

$$x^{(k+1)} = Bx^{(k)} + f,$$

où $B \in \mathbb{C}^{n,n}$ est dite *matrice d'itération* et f est obtenu à partir du second membre b. Dans ce cas, la consistance de la méthode est garantie si en posant $x^{(k)} = x$ on trouve $x^{(k+1)} = x$, à savoir,

$$x = Bx + f, \tag{3.8}$$

et la solution exacte est un point fixe de $\Phi(y) = By + f$. Pour une méthode consistante on a

$$e^{(k)} = x^{(k)} - x = (Bx^{(k-1)} - f) - (Bx - f) = B(x^{(k-1)} - x) = Be^{(k-1)} = B^k e^{(0)}. \tag{3.9}$$

La condition nécessaire et suffisante pour la convergence de la méthode est donc

$$\rho(B) < 1. \tag{3.10}$$

3.1.2 Les méthodes de Jacobi et Gauss–Seidel

Les méthodes de Jacobi et Gauss–Seidel sont obtenues à partir de la décomposition suivante de la matrice A (*décomposition DEL*) :

$$A = D - E - L, \tag{3.11}$$

où D est la diagonale de A, $-E$ sa portion strictement triangulaire supérieure, et $-L$ sa portion strictement triangulaire inférieure.

Example 3.5 (Décomposition DEL)**.** *La décomposition DEL de la matrice suivante :*

$$A = \begin{pmatrix} 2 & 3 & 1 \\ 7 & 1 & 2 \\ 8 & 1 & 2 \end{pmatrix}$$

est

$$D = \begin{pmatrix} 2 & & \\ & 1 & \\ & & 2 \end{pmatrix}, \qquad E = \begin{pmatrix} & -3 & -1 \\ & & -2 \\ & & \end{pmatrix}, \qquad L = \begin{pmatrix} & & \\ -7 & & \\ -8 & -1 & \end{pmatrix}.$$

La *méthode de Jacobi* est obtenue avec $P = D$ et $N = E + L$. A l'itération $k+1$ la mise à jour est effectuée en posant

$$x^{(k+1)} = D^{-1}(E+L)x^{(k)} + D^{-1}b. \tag{3.12}$$

La matrice d'itération correspondante est

$$B_{\mathrm{J}} = D^{-1}(E+L) = I - D^{-1}A. \tag{3.13}$$

Une condition suffisante pour la convergence de la méthode de Jacobi simple à vérifier est énoncée dans l'Exercice 3.15. Une variante de la méthode de Jacobi consiste à définir $x^{(k+1)}$ comme une moyenne pondérée de $x^{(k)}$ et de la valeur donnée par l'expression (3.12). Plus précisément, pour $\omega \in (0,1]$ on pose

$$x^{(k+1)} = (1-\omega)x^{(k)} + \omega\left[D^{-1}(E+L)x^{(k)} + D^{-1}b\right]. \tag{3.14}$$

Ce choix correspond à la méthode de surrélaxation JOR (*Jacobi Over-Relaxation*). La matrice d'itération correspondante est

$$B_{\mathrm{JOR}} = (1-\omega)I_n + \omega B_{\mathrm{J}}. \tag{3.15}$$

Lemme 3.6 (Convergence de la méthode JOR)**.** *Si $A = (a_{ij}) \in \mathbb{R}^{n,n}$ est SDP la méthode JOR converge pour toute estimée initiale x_0 si et seulement si le paramètre $\omega \in (0,1]$ satisfait la condition suivante :*

$$0 < \omega < \frac{2}{\rho(D^{-1}A)}.$$

Démonstration. On a

$$B_{\mathrm{JOR}} = I_n - \omega D^{-1} D + \omega D^{-1}(E+L) = I_n - \omega D^{-1} A.$$

Comme A est SDP, $a_{kk} > 0$ pour tout $1 \le k \le n$ et $(D^{-1})_{kk} = 1/a_{kk} > 0$, c.-à-d., D^{-1} est définie positive. En utilisant le résultat de l'Exercice 1.56, on peut conclure que

$$\mathrm{Sp}(B_{\mathrm{JOR}}) = \mathrm{Sp}(I_n - \omega D^{-1} A) = \left\{1 - \omega\lambda \,:\, \lambda \in \mathrm{Sp}(D^{-1}A)\right\}.$$

Par conséquent,

$$\rho(I_n - \omega D^{-1} A) = \max_{\lambda \in \mathrm{Sp}(D^{-1}A)} |1 - \omega\lambda|.$$

La condition (3.10) est donc vérifiée si $\omega \in (0,1]$ est tel que

$$\left(\forall \lambda \in \mathrm{Sp}(D^{-1}A), \quad |1-\omega\lambda| < 1\right) \iff \left(\forall \lambda \in \mathrm{Sp}(D^{-1}A), \quad \omega < \frac{2}{\lambda}\right) \iff \omega < \frac{2}{\rho(D^{-1}A)}. \quad \square$$

La *méthode de Gauss–Seidel* consiste à choisir $P = D - L$ et $N = E$. A l'itération $k+1$ la mise à jour est effectuée en posant

$$x^{(k+1)} = (D-L)^{-1} E\, x^{(k)} + (D-L)^{-1} b. \tag{3.16}$$

La matrice d'itération correspondante est

$$B_{\mathrm{GS}} = (D-L)^{-1} E.$$

Dans ce cas aussi on peut pondérer la valeur (3.16) par $x^{(k)}$, obtenant ainsi la méthode SOR (*Successive Over-Relaxation*). La mise à jour à l'itération $k+1$ revient donc à poser pour $\omega \in (0,1)$,

$$x^{(k+1)} = (1-\omega) x^{(k)} + \omega\left[(D-L)^{-1} E\, x^{(k)} + (D-L)^{-1} b\right]. \tag{3.17}$$

La matrice d'itération correspondante est

$$B_{\mathrm{SOR}} = (1-\omega) I_n + \omega B_{\mathrm{GS}}.$$

3.1.3 La méthode du gradient

Le dernier algorithme de point fixe que nous allons examiner admet une interprétation qui permet de dépasser le cadre de la résolution de systèmes linéaires, et qui le rend applicable à des problèmes plus généraux d'optimisation. Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP. On définit la fonctionnelle quadratique $J : \mathbb{R}^n \to \mathbb{R}$ telle que

$$J(y) := \frac{1}{2} y^T A y - b^T y, \tag{3.18}$$

en on considère le problème de minimisation libre

$$\min_{y \in \mathbb{R}^n} J(y).$$

On peut prouver que ce problème admet une et une seule solution $x \in \mathbb{R}^n$ (dite *minimiseur global* de J sur $\mathbb{R}^n$) telle que

$$J(x) \le J(y) \qquad \forall y \in \mathbb{R}^n,$$

et caractérisée par

$$\nabla J(x) = Ax - b = 0 \in \mathbb{R}^n.$$

Ce résultat est précisé dans le lemme suivant.

Lemme 3.7 (Minimisation d'une fonctionnelle quadratique et systèmes linéaires). *Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP. Alors, pour tout $b \in \mathbb{R}^n$,*

$$Ax = b \iff J(x) = \min_{y \in \mathbb{R}^n} J(y).$$

Démonstration. Pour tout $y, z \in \mathbb{R}^n$ et tout $t \in \mathbb{R}$ on a

$$\begin{aligned}
&J(z+ty) \\
&= \frac{1}{2}(z+ty)^T A(z+ty) - b^T(z+ty) \\
&= \underbrace{\frac{1}{2}z^T Az - b^T z}_{=J(z)} + \frac{t}{2}\left(z^T Ay + y^T Az\right) + \frac{t^2}{2}y^T Ay - tb^T y && \text{Proposition 1.7} \qquad (3.19)\\
&= J(z) + ty^T(Az-b) + \frac{t^2}{2}y^T Ay. && \text{Symétrie de } A
\end{aligned}$$

(i) Implication $J(x) = \min_{y\in\mathbb{R}^n} J(y) \implies Ax = b$. Puisque x est minimiseur global de J, l'identité (3.19) avec $z = x$ implique, pour tout $t \in \mathbb{R}$ et tout $y \in \mathbb{R}^n$,

$$J(x+ty) \geq J(x) \iff ty^T(Ax-b) + \frac{t^2}{2}y^T Ay \geq 0.$$

En prenant $t > 0$, en divisant par t, et en faisant tendre $t \to 0^+$ il vient $y^T(Ax-b) \geq 0$, d'où $Ax - b = 0 \in \mathbb{R}^n$ car $y \in \mathbb{R}^n$ est générique.

(ii) Implication $Ax = b \implies J(x) = \min_{y\in\mathbb{R}^n} J(y)$. Pour tout $w \in \mathbb{R}^n$, en utilisant encore l'identité (3.19) avec $z = x$, $t = 1$, et $y = w - x$ il vient

$$J(w) = J(x) + t(w-x)^T \underbrace{(Ax-b)}_{=0\in\mathbb{R}^n} + \frac{1}{2}(w-x)^T A(w-x).$$

La matrice A étant SDP, le dernier terme du membre de droite est non négatif (positif si $w \neq x$). Par suite, x est minimiseur global de J sur V. □

L'idée de la *méthode du gradient* se base sur la remarque suivante : pour tout $y \in \mathbb{R}^n$ avec $\nabla J(y) \neq 0$, on peut réduire *au moins localement* la valeur de $J(y)$ en se déplaçant dans la direction $-\nabla J(y)$. En effet, $\nabla J(y)$ correspond à la direction de plus grande pente (positive) de J en y. Fixons une estimation initiale $x_0 \in \mathbb{R}^n$. Pour $k \geq 1$ et jusqu'à convergence on pose

$$r^{(k)} := -\nabla J(x^{(k)}) = b - Ax^{(k)}, \qquad x^{(k+1)} = x^{(k)} + \alpha r^{(k)},$$

où le réel $\alpha > 0$ est lié au module du déplacement dans la direction de la plus profonde descente. La direction de descente $r^{(k)}$ est le *résidu* du système linéaire correspondant à l'approximation $x^{(k)}$ (on a $r^{(k)} = 0$ si $x^{(k)} = x$).

On peut ensuite se demander de combien faut-il avancer dans la direction $r^{(k)}$. Pour répondre à cette question, on cherche à reformuler la méthode du gradient sous la forme (3.4). On a

$$x^{(k+1)} = x^{(k)} + \alpha(b - Ax^{(k)}) = (I_n - \alpha A)x^{(k)} + \alpha b =: B_G x^{(k)} + f_G.$$

Afin de maximiser la vitesse de convergence (c.-à-d., réduire le plus possible l'erreur à chaque itération), le paramètre α doit être choisi de façon à minimiser le rayon spectral de la matrice d'itération B_G, comme le montre le lemme suivant.

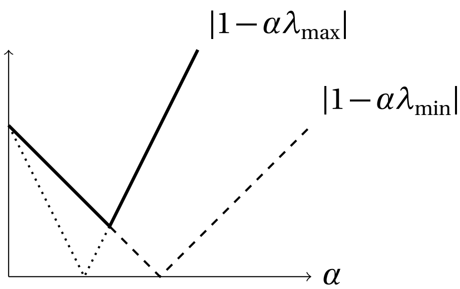
Lemme 3.8 (Vitesse de convergence de la méthode du gradient)**.** *Soit $A \in \mathbb{R}^{n,n}$ SDP et on note respectivement $\lambda_{\max}$ et $\lambda_{\min}$ la plus grande et la plus petite valeur propre de A. En choisissant $\alpha = \frac{2}{\lambda_{\max}+\lambda_{\min}}$, la méthode du gradient converge pour toute estimation initiale $x^{(0)} \in \mathbb{R}^n$ et on a pour $k \geq 1$,*

$$\|e^{(k+1)}\|_2 \leq \frac{\operatorname{cond}_2(A)-1}{\operatorname{cond}_2(A)+1}\|e^{(k)}\|_2.$$

Démonstration. On a par définition

$$\operatorname{Sp}(B_G) = 1 - \alpha\operatorname{Sp}(A) \implies \rho(B_G) = \max_{\lambda\in\operatorname{Sp}(()A)} |1-\alpha\lambda|.$$

Le graphe de la fonction $\varphi(\alpha) := \max_{\lambda\in\operatorname{Sp}(A)} |1-\alpha\lambda|$ est représenté en trait plein dans la figure suivante.



Le minimum de $\varphi(\alpha)$ est donc atteint pour α tel que $|1-\alpha\lambda_{\min}| = |1-\alpha\lambda_{\max}|$, à savoir, $\alpha = \frac{2}{\lambda_{\max}+\lambda_{\min}}$. De plus, on a

$$\varphi\left(\frac{2}{\lambda_{\min}+\lambda_{\max}}\right) = \rho(B_G) = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} = \frac{\cancel{\lambda_{\min}}(\operatorname{cond}_2(A)-1)}{\cancel{\lambda_{\min}}(\operatorname{cond}_2(A)+1)} < 1,$$

où nous avons utilisé (2.2) et le fait que A est SDP pour conclure $\operatorname{cond}_2(A) = \lambda_{\max}/\lambda_{\min}$. Par conséquent, avec le choix proposé pour α, nous avons $\varrho(B_G) < 1$. Comme la méthode est consistante, ceci implique sa convergence pour tout choix initial (voir la discussion à la fin de la Section 3.1.1 à ce sujet).

Passant à la norme $\|\cdot\|_2$ dans (3.9), on trouve pour tout $k \geq 0$,

$$\|e^{(k+1)}\|_2 = \|B_G e^{(k)}\|_2 \leq \|B_G\|_2\|e^{(k)}\|_2 = \varrho(B_G)\|e^{(k)}\|_2,$$

où nous avons utilisé la Proposition 1.47 dans la majoration et le Corollaire 1.42 pour écrire $\|B_G\|_2 = \varrho(B_G)$. □

Example 3.9 (Convergence lente de la méthode du gradient)**.** *Le Lemme 3.8 montre que convergence de la méthode du gradient peut devenir lente lorsque la matrice A est mal conditionnée : lorsque* $\operatorname{cond}_2 A$ *est grand, on a en effet* $\rho(B_G) = \frac{\operatorname{cond}_2(A)-1}{\operatorname{cond}_2(A)+1} \approx 1$.

Le choix du paramètre d'accélération α suggéré par le Lemme 3.8 est souvent remplacé en pratique par d'autres expressions plus simples à calculer. Un choix qui est *localement optimal* est obtenu en posant, à chaque itération k, $\alpha = \alpha^{(k)}$ avec $\alpha^{(k)}$ unique solution du problème de minimisation suivant :

$$\min_{\alpha\in\mathbb{R}} \left\{ J(x^{(k+1)}) = J(x^{(k)} + \alpha r^{(k)}) \right\}. \tag{3.20}$$

On peut montrer (voir Exercice 3.19) que la solution de ce problème est

$$\alpha^{(k)} = \frac{\|r^{(k)}\|_2^2}{\|r^{(k)}\|_A^2}. \tag{3.21}$$

3.2 Méthode du gradient conjugué

Comme nous l'avons vu dans la section précédente, la méthode du gradient devient assez inefficace lorsque la matrice est mal conditionnée. Le problème vient du fait que la direction de descente est choisie en utilisant uniquement des informations locales. L'idée des méthodes considérées dans cette section consiste à se déplacer dans une direction obtenue *en tenant compte des directions déjà explorées au cours des itérations précédentes.*

3.2.1 Vecteurs A-conjuguées

Définition 3.10 (Vecteurs A-conjugués)**.** *Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP. Deux vecteurs* non nuls $x, y \in \mathbb{R}^n$ *sont dits A-conjugués si $y^T A x = x^T A y = 0$. Une famille $(x_i)_{1\le i\le p}$ de vecteurs de $\mathbb{R}^n$* non nuls *est dite A-conjuguée si*

$$\forall 1 \le i, j \le p \qquad \Big(i \ne j \implies x_i^T A x_j = x_j^T A x_i = 0 \Big).$$

Lemme 3.11 (Famille de vecteurs A-conjuguée)**.** *Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP et $\mathscr{F} := (x_i)_{1\le i\le p}$, une famille de vecteurs A-conjuguée. Alors $\mathscr{F}$ est libre et $p \le n$. Si $p = n$, alors $\mathscr{F}$ est une base de $\mathbb{R}^n$.*

Démonstration. On cherche les combinaisons linéaires des vecteurs de $\mathscr{F}$ telles que

$$\sum_{j=1}^{p} \lambda_j x_j = \lambda_1 x_1 + \ldots + \lambda_p x_p = 0 \quad \text{avec} \quad \lambda_j \in \mathbb{R} \quad \forall 1 \le j \le p.$$

Soit $1 \le i \le p$. En multipliant à gauche par $x_i^T A$ l'égalité ci-dessus on obtient

$$x_i^T A \sum_{j=1}^{p} \lambda_j x_j = \sum_{j=1}^{p} \lambda_j x_i^T A x_j = \lambda_i \|x_i\|_A^2 = 0.$$

Comme $x_i \ne 0$ par définition, ceci implique $\lambda_i = 0$, à savoir, la famille $\mathscr{F}$ est libre et, par conséquent, $p \le n$. Si $p = n$ on a une famille libre de cardinalité égale à la dimension de l'espace $\mathbb{R}^n$, et il s'agit donc d'une base de cet espace. $\square$

3.2.2 La méthode du gradient conjugué

Une méthode qui utilise des directions conjuguées est dite *conjuguée*. Nous allons montrer dans cette section comment construire une méthode conjuguée pour la solution du système linéaire (3.1) avec $A \in \mathbb{R}^{n,n}$ SDP. Par brévité dans ce qui suit on notera $(\cdot,\cdot)$ le produit scalaire canonique de $\mathbb{R}^n$ et on posera $(\cdot,\cdot)_A := (A\cdot,\cdot)$ et $\|v\|_A := \sqrt{(v,v)_A}$ pour tout $v \in \mathbb{R}^n$. Soit $x^{(0)} \in \mathbb{R}^n$ une estimation initial de la solution et définissons le résidu $r^{(0)} := b - Ax^{(0)}$. Les directions de descente $w^{(k)}$, $k = 0, \ldots$ sont obtenues comme suit : $w^{(0)} = r^{(0)}$ et

$$w^{(k)} = r^{(k)} - \beta^{(k)} w^{(k-1)} \qquad k = 1, \ldots \quad \text{avec } \beta^{(k)} \in \mathbb{R} \text{ à préciser.} \tag{3.22}$$

Une fois identifiée une direction $w^{(k)}$ appropriée, la mise à jour de la solution se fait en posant

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} w^{(k)} \quad \text{avec } \alpha^{(k+1)} \in \mathbb{R} \text{ à préciser.} \tag{3.23}$$

Ceci implique, en particulier,

$$r^{(k+1)} = b - Ax^{(k+1)} = b - Ax^{(k)} - \alpha^{(k+1)} Aw^{(k)} = r^{(k)} - \alpha^{(k+1)} Aw^{(k)}. \tag{3.24}$$

Pour déterminer les réels $\alpha^{(k)}$ et $\beta^{(k)}$ nous demandons à ce que, pour tout $0 \le j \le k-1$,

$$(w^{(k)}, w^{(j)})_A = 0, \tag{CG1}$$

$$(r^{(k)}, w^{(j)}) = 0. \tag{CG2}$$

Les conditions (CG1) et (CG2) correspondent à la méthode du *gradient conjugué* (CG) et expriment le fait que la nouvelle direction $w^{(k)}$ et le résidu $r^{(k)}$ sont respectivement A-orthogonale et orthogonal aux directions précédentes. Ces deux conditions montrent une caractéristique importante de la méthode CG : la mise à jour n'est pas uniquement basée sur une information locale (direction de plus grande pente), mais elle tient compte également des itérations précédentes. Comme on le verra plus loin, on peut assurer les deux conditions (CG1)–(CG2) *sans mémoriser* les directions $w^{(j)}$, $1 \le j < k-1$. Ce point est très important en pratique, car il implique que la mise en œuvre de la méthode CG ne demande pas significativement plus de mémoire que les autres méthodes itératives que l'on a étudié jusqu'à là.

Le lemme suivant montre de façon constructive l'existence d'une famille de directions de descente $(w^{(k)})_{k=0,\ldots}$ qui remplissent les conditions (CG1)–(CG2) et permet de préciser l'expression des paramètres $\alpha^{(k)}$ et $\beta^{(k)}$.

Lemme 3.12 (Existence des directions $(w^{(k)})_{k=0,\ldots}$)**.** *Pour tout $x^{(0)} \in \mathbb{R}^n$ il existe des valeurs des paramètres $\alpha^{(k)}$ et $\beta^{(k)}$, $k = 1, \ldots$, et une famille de vecteurs $(w^{(k)})_{k=0,\ldots}$ telle que les conditions* (CG1)–(CG2) *sont satisfaites pour l'itération définie par* (3.22) *et* (3.23).

Démonstration. La preuve procède par récurrence.

(i) Preuve du résultat pour $k = 1$. Soit $r^{(0)} := b - Ax^{(0)}$ le résidu correspondant à l'estimation initiale. Si $r^{(0)} = 0$, $x^{(0)} = x$ et nous n'avons rien de plus à faire. Si $r^{(0)} \neq 0$, posons

$w^{(0)} = r^{(0)}$. La nouvelle estimation de la solution est $x^{(1)} = x^{(0)} + \alpha^{(1)} w^{(0)}$, avec $\alpha^{(1)}$ choisi de tel sorte à assurer la condition (CG2). De par (3.24) on a

$$0 = (w^{(0)}, r^{(1)}) = (r^{(0)}, r^{(0)} - \alpha^{(1)} A w^{(0)}) = \|r^{(0)}\|_2^2 - \alpha^{(1)} \|w^{(0)}\|_A^2 \iff \boxed{\alpha^{(1)} = \frac{(w^{(0)}, r^{(0)})}{\|w^{(0)}\|_A^2}.}$$

La direction de descente suivante est donnée par la formule

$$w^{(1)} = r^{(1)} - \beta^{(1)} w^{(0)},$$

avec $\beta^{(1)}$ tel que (CG1) soit vérifiée, c.-à-d.,

$$0 = (w^{(0)}, w^{(1)})_A = (w^{(0)}, r^{(1)} - \beta^{(1)} w^{(0)})_A \iff \boxed{\beta^{(1)} = \frac{(w^{(0)}, r^{(1)})_A}{\|w^{(0)}\|_A^2}.}$$

(ii) Preuve du résultat pour l'indice $k+1$ en le supposant vrai pour l'indice k. Supposons maintenant (CG1)–(CG2) vérifiées pour $k \geq 1$ et prouvons l'existence de $\alpha^{(k+1)}$, $\beta^{(k+1)}$ et $w^{(k+1)}$ tels les deux conditions restent vraies. En utilisant l'expression (3.24) pour $r^{(k+1)}$ on a

$$(w^{(k)}, r^{(k+1)}) = (w^{(k)}, r^{(k)} - \alpha^{(k+1)} A w^{(k)}) = (w^{(k)}, r^{(k)}) - \alpha^{(k+1)} \|w^{(k)}\|_A^2,$$

à savoir, (CG2) pour $j = k$ est vérifiée pour

$$\boxed{\alpha^{(k+1)} = \frac{(w^{(k)}, r^{(k)})}{\|w^{(k)}\|_A^2}.} \tag{3.25}$$

Prouvons maintenant (CG2) pour tout $0 \leq j \leq k-1$. Il suffit d'observer que, pour tout $0 \leq j \leq k-1$,

$$\begin{aligned}
(w^{(j)}, r^{(k+1)}) &= (w^{(j)}, r^{(k)} - \alpha^{(k+1)} A w^{(k)}) && \text{Éq. (3.24)} \\
&= (w^{(j)}, r^{(k)}) - \alpha^{(k+1)} (w^{(j)}, w^{(k)})_A = 0. && \text{Récurrence (CG1)–(CG2)}
\end{aligned}$$

Venons maintenant à (CG1). Nous avons

$$\begin{aligned}
(w^{(k)}, w^{(k+1)})_A &= (w^{(k)}, r^{(k+1)} - \beta^{(k+1)} w^{(k)})_A && \text{Éq. (3.22)} \\
&= (w^{(k)}, r^{(k+1)})_A - \beta^{(k+1)} \|w^{(k)}\|_A^2,
\end{aligned}$$

et (CG1) pour $j = k$ est donc vérifiée pour

$$\boxed{\beta^{(k+1)} = \frac{(w^{(k)}, r^{(k+1)})_A}{\|w^{(k)}\|_A^2}.}$$

Il ne reste plus qu'à prouver (CG1) pour $0 \leq j \leq k-1$. Puisque $w^{(0)} = r^{(0)}$ et chaque nouvelle direction $w^{(k)}$ est obtenue à partir de $r^{(k)}$ et des directions $w^{(j)}$, $0 \leq j \leq k-1$, on a

$$V_{k+1} := \operatorname{span}(w^{(0)}, \ldots, w^{(k)}) = \operatorname{span}(r^{(0)}, \ldots, r^{(k)}).$$

La condition (CG2) pour $0 \le j \le k$ se traduit comme suit :

$$r^{(k+1)} \in V_{k+1}^{\perp} := \{w \in \mathbb{R}^n : (w, v) = 0 \quad \forall v \in V_{k+1}\}. \tag{3.26}$$

D'autre part, pour tout $0 \le j \le k-1$, on a d'après (3.24),

$$\begin{aligned} Aw^{(j)} = \frac{1}{\alpha^{(j+1)}}\left(r^{(j)} - r^{(j+1)}\right) \in V_{k+1} &\implies Aw^{(j)} \in V_{k+1} \\ &\implies (Aw^{(j)}, r^{(k+1)}) = (w^{(j)}, r^{(k+1)})_A = 0, \end{aligned} \tag{3.27}$$

où nous avons conclu grâce à (3.26). En utilisant les remarques précédentes on trouve

$$\begin{aligned} (w^{(j)}, w^{(k+1)})_A &= (w^{(j)}, r^{(k+1)} - \beta^{(k+1)} w^{(k)})_A && (3.22) \\ &= (Aw^{(j)}, r^{(k+1)}) - \beta^{(k+1)}(w^{(j)}, w^{(k)})_A = 0. && (3.27), \text{ récurrence (CG1)} \end{aligned}$$

Ceci conclut la preuve. □

Compte tenu du lemme précédent, la méthode CG est définie comme suit : Pour une estimation initiale $x^{(0)} \in \mathbb{R}^n$, poser $r^{(0)} = b - Ax^{(0)}$, $w^{(0)} = r^{(0)}$ et, pour $k = 0, \ldots,$

$$\alpha^{(k+1)} = \frac{(w^{(k)}, r^{(k)})}{\|w^{(k)}\|_A^2}, \tag{3.28a}$$

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} w^{(k)}, \tag{3.28b}$$

$$r^{(k+1)} = r^{(k)} - \alpha^{(k+1)} Aw^{(k)}, \tag{3.28c}$$

$$\beta^{(k+1)} = \frac{(w^{(k)}, r^{(k+1)})_A}{\|w^{(k)}\|_A^2}, \tag{3.28d}$$

$$w^{(k+1)} = r^{(k+1)} - \beta^{(k+1)} w^{(k)}. \tag{3.28e}$$

Théorème 3.13 (Convergence de la méthode CG)**.** *Soit $A \in \mathbb{R}^{n,n}$ SDP et $b \in \mathbb{R}^n$. La méthode CG pour la solution du système linéaire $Ax = b$ converge en au plus n itérations.*

Démonstration. La famille $\mathscr{F} := (w^{(0)}, \ldots, w^{(n-1)})$ de cardinalité n est libre d'après le Lemme 3.11, et elle est donc une base de l'espace $\mathbb{R}^n$. Comme $r^{(n)}$ est orthogonal à tout vecteur de $\mathscr{F}$ on doit avoir $r^{(n)} = 0$, à savoir, $x^{(n)} = x$. □

Le Théorème 3.13 montre que la méthode du gradient conjugué n'est pas une méthode itérative au sens strict, car elle converge après un nombre fini d'itérations. Cependant, son intérêt est lié au fait que, dans de nombreux cas pratiques, le résidu décroît rapidement, et le critère $\|r^{(k)}\| \le \epsilon$ est vérifié après un petit nombre d'itérations. En effet, le résultat suivant montre que la valeur de la fonctionnelle J définie par (3.18) décroît à chaque itération de la méthode (de par le Lemme 3.7, on sait que résoudre le système linéaire $Ax = b$ avec A SDP équivaut à minimiser J). Dans le jargon de l'optimisation on dit que $w^{(k)}$ est une *direction de descente stricte* de J en $x^{(k-1)}$.

Proposition 3.14 (Monotonicité des itérations CG)**.** *Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP, $b \in \mathbb{R}^n$ et J définie par* (3.18). *On considère l'itération* $(k+1)$ *de la méthode du gradient définie par* (3.28). *Alors, si $w^{(k)} \neq 0$ et $\alpha^{(k+1)} \neq 0$,*

$$J(x^{(k+1)}) < J(x^{(k)}).$$

Si $\alpha^{(k+1)} = 0$, $x^{(k)}$ est le minimiseur de J et $Ax^{(k)} = b$.

Démonstration. On commence par remarquer que

$$
\begin{aligned}
(w^{(k)}, r^{(k)}) &= (w^{(k)}, r^{(k+1)} + \alpha^{(k+1)} A w^{(k)}) && (3.24)\\
&= \underbrace{(w^{(k)}, r^{(k+1)})}_{=0} + \alpha^{(k+1)} \|w^{(k)}\|_A^2 > 0, && \text{(CG2)}
\end{aligned}
$$

où l'inégalité est stricte car $w^{(k)} \neq 0$. Nous avons alors

$$
\begin{aligned}
J(x^{(k+1)}) &= \frac{1}{2}(x^{(k)} + \alpha^{(k+1)} w^{(k)}, x^{(k)} + \alpha^{(k+1)} w^{(k)})_A - (b, x^{(k)} + \alpha^{(k+1)} w^{(k)}) && \text{(3.18) et (3.23)}\\
&= \underbrace{\frac{1}{2}(x^{(k)}, x^{(k)})_A - (b, x^{(k)})}_{=J(x^{(k)})} + \alpha^{(k+1)} \underbrace{(Ax^{(k)} - b, w^{(k)})}_{=-(r^{(k)}, w^{(k)})} + \underbrace{\frac{(\alpha^{(k+1)})^2}{2} \|w^{(k)}\|_A^2}_{=\frac{\alpha^{(k+1)}}{2}(w^{(k)}, r^{(k)})} && (3.25)\\
&= J(x^{(k)}) - \frac{\alpha^{(k+1)}}{2}(w^{(k)}, r^{(k)}) < J(x^{(k)}),
\end{aligned}
$$

où nous avons conclu grâce au fait que $(w^{(k)}, r^{(k)}) > 0$ prouvé précédemment. □

Il est important de retenir que la méthode CG accompagnée d'un bon préconditionneur est la méthode de choix pour la résolution de systèmes linéaires caractérisés par une matrice SDP.

3.3 Exercices

Exercice 3.15 (Matrice à diagonale dominante)**.** *Soit $A = (a_{ij}) \in \mathbb{R}^{n,n}$ une matrice à diagonale dominante par lignes, à savoir*

$$|a_{ii}| > \sum_{j=1,\, j\neq i}^{n} |a_{ij}| \qquad \forall 1 \le i \le n.$$

Soit, de plus, $b \in \mathbb{R}^n$, $b \neq 0$. Montrer que la méthode de Jacobi pour la résolution du système linéaire $Ax = b$ est convergente.

Exercice 3.16 (Méthode des directions alternées)**.** *Soit $A \in \mathbb{R}^{n,n}$ une matrice SDP décomposée en $A = A_1 + A_2$ avec A_1 et A_2 matrices SDP et $b \in \mathbb{R}^n$, $b \neq 0$. Soient $\alpha_1, \alpha_2 \in \mathbb{R}^+_*$. On considère la méthode itérative suivante pour résoudre le système linéaire $Ax = b$:*

$$(I_n + \alpha_1 A_1)x^{(k+1/2)} = (I_n - \alpha_1 A_2)x^{(k)} + \alpha_1 b, \tag{3.29a}$$

$$(I_n + \alpha_2 A_2)x^{(k+1)} = (I_n - \alpha_2 A_1)x^{(k+1/2)} + \alpha_2 b. \tag{3.29b}$$

On souhaite analyser la convergence de cette méthode. A ce propos,

(i) prouver que, pour toute matrice SDP $M \in \mathbb{R}^{n,n}$ et tout $\xi \in \mathbb{R}^+_*$, la matrice $I_n + \xi M$ est définie positive. En déduire que les matrices $I_n + \alpha_i A_i$, $i \in \{1,2\}$, sont définies positives;

(ii) prouver que, étant donné $x^{(k)}$, $x^{(k+1/2)}$ et $x^{(k+1)}$ sont uniques;

(iii) montrer la méthode est consistante, à savoir, si $x^{(k)} = x$, alors $x^{(k+1/2)} = x^{(k+1)} = x$;

(iv) identifier la matrice B et le vecteur f permettant d'interpreter (3.29) *comme une méthode de point fixe de la forme*

$$x^{(k+1)} = Bx^{(k)} + f;$$

(v) On admettra l'estimation suivante du rayon spectrale de la matrice B trouvée au point précédent :

$$\rho(B) \le \max_{\lambda \in \mathrm{Sp}(A_1)} \left| \frac{1-\alpha_2\lambda}{1+\alpha_1\lambda} \right| \times \max_{\lambda \in \mathrm{Sp}(A_2)} \left| \frac{1-\alpha_1\lambda}{1+\alpha_2\lambda} \right|. \tag{3.30}$$

Proposer une condition sur α_1 et α_2 garantissant la convergence de la méthode. B.

(i) Il suffit de prouver que $I_n + \xi M$ est à son tour SDP. La symétrie est évidente. Soit $x \in \mathbb{R}^n$, $x \neq 0$. Par définition on a

$$x^T(I_n + \xi M)x = x^T x + \xi x^T M x > 0,$$

ce qui prouve que $I_n + \xi M$ est définie positive. Le fait que les matrices $I_n + \alpha_i A_i$, $i \in \{1,2\}$, sont définies positives est une conséquence immédiate.

(ii) Comme les matrices $I_n + \alpha_1 A_1$ et $I_n + \alpha_2 A_2$ sont SDP par le point (i), elles sont aussi inversibles. Par conséquent, les conditions (3.29) définissent $x^{(k+1/2)}$ et $x^{(k+1)}$ de façon unique.

(iii) En remplaçant $x^{(k)}$ par x dans (3.29a) on trouve

$$\begin{aligned}
(I_n+\alpha_1 A_1)x^{(k+1/2)} &= x-\alpha_1 A_2 x+\alpha_1 b \\
&= x-\alpha_1(A-A_1)x+\alpha_1 b && (A_2 = A-A_1) \\
&= x+\alpha_1 A_1 x = (I_n+\alpha_1 A_1)x, && (Ax=b)
\end{aligned}$$

à savoir, $x^{(k+1/2)} = (I_n+\alpha_1 A_1)^{-1}(I_n+\alpha_1 A_1)x = x$.

Pour prouver que $x^{(k+1)} = x$, on procède de façon similaire en remplaçant $x^{(k+1/2)}$ par x dans (3.29b) :

$$\begin{aligned}
(I_n+\alpha_2 A_2)x^{(k+1)} &= (I_n-\alpha_2 A_1)x+\alpha_2 b \\
&= x-\alpha_2(A-A_2)x+\alpha_2 b && (A_2 = A-A_2) \\
&= x+\alpha_2 A_2 x = (I_n+\alpha_2 A_2)x, && (Ax=b)
\end{aligned}$$

et, par conséquent, $x^{(k+1)} = (I_n+\alpha_2 A_2)^{-1}(I_n+\alpha_2 A_2)x = x$, ce qui prouve le résultat souhaité.

(iv) De par (3.29a) on a $x^{(k+1/2)} = (I_n+\alpha_1 A_1)^{-1}(I_n-\alpha_1 A_2)x^{(k)}+\alpha_1(I_n+\alpha_1 A_1)^{-1}b$. En remplaçant cette expression dans (3.29b), on obtient

$$x^{(k+1)} = Bx^{(k)}+f,$$

avec

$$\begin{aligned}
&B := (I_n+\alpha_2 A_2)^{-1}(I_n-\alpha_2 A_1)(I_n+\alpha_1 A_1)^{-1}(I_n-\alpha_1 A_2) \\
&f := \alpha_1(I_n+\alpha_2 A_2)^{-1}(I_n-\alpha_2 A_1)(I_n+\alpha_1 A_1)^{-1}b+\alpha_2(I_n+\alpha_2 A_2)^{-1}b.
\end{aligned}$$

La consistance de la méthode prouvée au point précédent se traduit par la rélation

$$x = Bx+f.$$

(iv) En utilisant l'expression trouvée au point précédent on peut estimer l'erreur $e^{(k)} := x^{(k)}-x$ par récurrence :

$$e^{(k)} = x^{(k)}-x = (Bx^{(k-1)}+f)-(Bx+f) = Be^{(k-1)} = B^k e^{(0)}.$$

La méthode est donc convergente si et seulement si $\rho(B)<1$. Or, comme les valeurs propres de A_1 et A_2 sont >0, il suffit de prendre $\alpha_1=\alpha_2=\alpha>0$ pour que les deux facteurs dans (3.30) soient <1 et, donc, $\rho(B)<1$. Pour s'en convaincre, on résout l'inéquation en α avec $\lambda>0$:

$$\left|\frac{1-\alpha\lambda}{1+\alpha\lambda}\right| < 1.$$

Comme $1+\alpha\lambda>0$, ceci revient à résoudre

$$|1-\alpha\lambda| < 1+\alpha\lambda,$$

à savoir,

$$\begin{cases} 1-\alpha\lambda < 1+\alpha\lambda & \text{si } 0<\alpha\le 1/\lambda, \\ -1+\alpha\lambda < 1+\alpha\lambda & \text{si } \alpha>1/\lambda. \end{cases} \iff \begin{cases} 2\alpha\lambda>0 & \text{si } 0<\alpha\le 1/\lambda, \\ -1<1 & \text{si } \alpha>1/\lambda, \end{cases} \iff \alpha\in\mathbb{R}_+^*,$$

ce qui montre le résultat cherché.

Exercice 3.17 (Analyse d'une méthode itérative). *Pour la résolution du système $Ax = b$ avec*

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 5 \end{pmatrix},$$

on considère la méthode itérative suivante : Pour $k \geq 0$,

$$x^{(k+1)} = B(\theta)x^{(k)} + f(\theta),$$

avec $x^{(0)} \in \mathbb{R}^n$ estimation initiale, $\theta \in \mathbb{R}$, et

$$B = \frac{1}{4}\begin{pmatrix} 2\theta^2+2\theta+1 & -2\theta^2+2\theta+1 \\ -2\theta^2+2\theta+1 & 2\theta^2+2\theta+1 \end{pmatrix}, \qquad f(\theta) = \frac{1}{2}\begin{pmatrix} 1-2\theta \\ 1-2\theta \end{pmatrix}.$$

1) Vérifier que la méthode est consistante pour tout $\theta \in \mathbb{R}$, à savoir, $x = B(\theta)x + f(\theta)$.

2) Préciser ensuite pour quelles valeurs de θ la méthode est convergente;

1) Consistance. La solution du système est $x = (1,1)^T$ et on vérifie aisément pour tout $\theta \in \mathbb{R}$,

$$B(\theta)x + f(\theta) = \frac{1}{4}\begin{pmatrix} \cancel{2\theta^2}+2\theta+1\cancel{-2\theta^2}+2\theta+1 \\ \cancel{-2\theta^2}+2\theta+1+\cancel{2\theta^2}+2\theta+1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1-2\theta \\ 1-2\theta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

ce qui prouve la consistance de la méthode.

2) Spectre de $B(\theta)$ et convergence. Comme la méthode est consistante, condition nécessaire et suffisante pour que la méthode soit convergente est que $\rho(B(\theta)) < 1$. Calculons donc les valeurs propres de B. Il s'agit de résoudre

$$0 = \det(B(\theta) - \lambda I_n) = \frac{1}{4}\begin{vmatrix} 2\theta^2+2\theta+1-4\lambda & -2\theta^2+2\theta+1 \\ -2\theta^2+2\theta+1 & 2\theta^2+2\theta+1-4\lambda \end{vmatrix}$$
$$= \frac{1}{4}\left\{\left(2\theta^2+2\theta+1-4\lambda\right)^2 - \left(-2\theta^2+2\theta+1\right)^2\right\}.$$

On a donc

$$2\theta^2+2\theta+1-4\lambda = \pm(-2\theta^2+2\theta+1),$$

ce qui donne comme valeurs propres

$$\lambda_{1,2} = \frac{2\theta^2+2\theta+1 \pm (-2\theta^2+2\theta+1)}{4} = \begin{cases} \theta + \frac{1}{2}, \\ \theta^2. \end{cases}$$

La condition $\rho(B(\theta)) < 1$ équivaut donc à imposer

$$\max\left(\left|\theta + \frac{1}{2}\right|, \theta^2\right) < 1.$$

Or, on a que (vérifier!)

$$\max\left(\left|\theta + \frac{1}{2}\right|, \theta^2\right) = \begin{cases} \theta^2 & \text{si } \theta \leq \frac{1-\sqrt{3}}{2} \text{ ou } \theta \geq \frac{1+\sqrt{3}}{2}, \\ \theta + \frac{1}{2} & \text{si } \frac{1-\sqrt{3}}{2} < \theta < \frac{1+\sqrt{3}}{2}. \end{cases}$$

Donc, $\varrho(B(\theta)) < 1$ si et seulement si $\theta \in \left(-1, \frac{1}{2}\right)$.

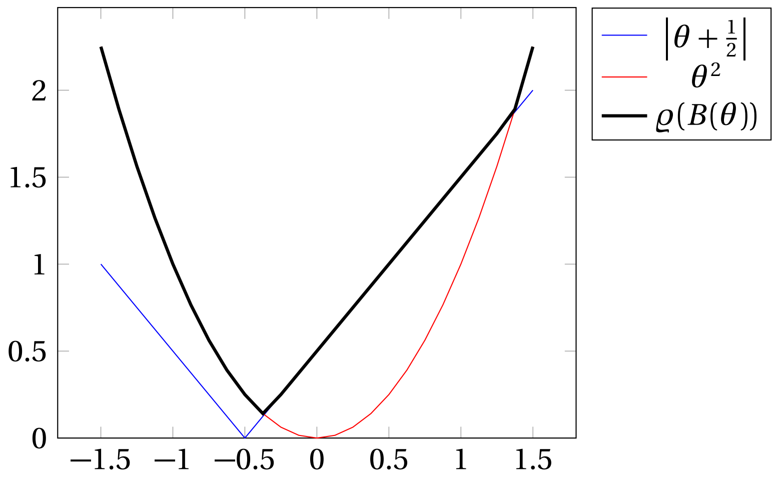


FIGURE 3.1 – Graphe de la fonction $\varrho(B(\theta))$.

Exercice 3.18 (Forte convexité d'une fonctionnelle quadratique)**.** *Montrer que la fonctionnelle J définie par* (3.18) *est* fortement convexe*, à savoir, il existe $\eta \in \mathbb{R}_*^+$ tel que, pour tout $x, y \in \mathbb{R}^n$ et tout $\vartheta \in [0,1]$,*

$$J(\vartheta x+(1-\vartheta)y) \leq \vartheta J(x)+(1-\vartheta)J(y)-\eta\frac{\vartheta(1-\vartheta)}{2}\|x-y\|^2. \tag{3.31}$$

Préciser la valeur de η.

On a

$$\begin{aligned}J(\vartheta x+(1-\vartheta)y) &= \frac{1}{2}(\vartheta x+(1-\vartheta)y)^T A(\vartheta x+(1-\vartheta)y)-b^T(\vartheta x+(1-\vartheta)y)\\ &= \frac{\vartheta^2}{2}x^TAx+\frac{(1-\vartheta)^2}{2}y^TAx+\vartheta(1-\vartheta)y^TAx-\vartheta b^Tx-(1-\vartheta)b^Ty,\end{aligned}$$

où nous avons utilisé la symétrie de A pour conclure $x^TAy = y^TAx$. Pour le troisième terme au seconde membre on observera que

$$y^TAx = \frac{1}{2}x^TAx+\frac{1}{2}y^TAy-(y-x)^TA(y-x).$$

En remplaçant on obtient

$$J(\vartheta x+(1-\vartheta)y) = \vartheta J(x)+(1-\vartheta)J(y)-\frac{\vartheta(1-\vartheta)}{2}(x-y)^TA(x-y). \tag{3.32}$$

Or, en notant $\lambda_{\min}(A) > 0$ la plus petite valeur propre de A, nous avons pour tout $z \in \mathbb{R}^n$,

$$z^TAz \geq \lambda_{\min}(A)\|z\|^2 \iff -z^TAz \leq -\lambda_{\min}(A)\|z\|^2.$$

L'inégalité ci-dessus pour $z = y-x$ donne finalement

$$J(\vartheta x+(1-\vartheta)y) \leq \vartheta J(x)+(1-\vartheta)J(y)-\lambda_{\min}(A)\frac{(1-\vartheta)\vartheta}{2}\|x-y\|^2,$$

qui est (3.31) avec $\eta = \lambda_{\min}(A)$.

Exercice 3.19 (Paramètre d'accéleration pour la méthode du gradient)**.** *Le but de cet exercice est de prouver la formule* (3.21)*. Soit* $\varphi : \mathbb{R} \to \mathbb{R}$ *la fonction définie par* $\varphi(\alpha) := J(x^{(k)} + \alpha r^{(k)})$*. Le problème* (3.20) *équivaut à*

$$\min_{\alpha\in\mathbb{R}} \varphi(\alpha).$$

Prouver que φ *est fortement convexe, à savoir, il existe* $\eta \in \mathbb{R}^+_*$ *tel que pour tout* $\alpha, \beta \in \mathbb{R}$ *et tout* $\vartheta \in [0,1]$,

$$\varphi(\vartheta\alpha + (1-\vartheta)\beta) \le \vartheta\varphi(\alpha) + (1-\vartheta)\varphi(\beta) - \eta\frac{\vartheta(1-\vartheta)}{2}|\alpha-\beta|^2.$$

La forte convexité de φ *assure que le problème* (3.20) *admet unique solution* $\alpha^{(k)}$ *caractérisée par la propriété suivante :*

$$\varphi'(\alpha^{(k)}) = 0.$$

Montrer que cette propriété équivaut à (3.21).

Par brévité, on note $(\cdot,\cdot)$ le produit interne canonique de $\mathbb{R}^n$ et on pose $(\cdot,\cdot)_A := (A\cdot,\cdot)$. On commence par prouver la forte convexité de φ. On a

$$\begin{aligned}
\varphi(\vartheta\alpha + (1-\vartheta)\beta) &= J(x^{(k)} + (\vartheta\alpha + (1-\vartheta)\beta) r^{(k)}) \\
&= J(\vartheta x^{(k)} + (1-\vartheta)x^{(k)} + (\vartheta\alpha + (1-\vartheta)\beta) r^{(k)}) \\
&= J(\vartheta w_\alpha + (1-\vartheta) w_\beta) && (w_\xi := x^{(k)} + \xi r^{(k)}) \\
&= \vartheta J(w_\alpha) + (1-\vartheta) J(w_\beta) - \frac{\vartheta(1-\vartheta)}{2}\|w_\alpha - w_\beta\|_A^2 && \text{Eq. (3.32)} \\
&= \vartheta\varphi(\alpha) + (1-\vartheta)\varphi(\beta) - \|r^{(k)}\|_A^2 \frac{\vartheta(1-\vartheta)}{2}|\alpha-\beta|^2,
\end{aligned}$$

et φ est donc fortement convexe de paramètre $\eta = \|r^{(k)}\|_A^2$. Calculons maintenant la dérivée de φ. Nous avons

$$\begin{aligned}
\varphi'(\alpha) &= \left\{\frac{1}{2}(x^{(k)} + \alpha r^{(k)}, x^{(k)} + \alpha r^{(k)})_A - (b, x^{(k)} + \alpha r^{(k)})\right\}' \\
&= (r^{(k)}, x^{(k)} + \alpha r^{(k)})_A - (b, r^{(k)}) \\
&= (Ax^{(k)}, r^{(k)}) + \alpha\|r^{(k)}\|_A^2 - (b, r^{(k)}) \\
&= -(b - Ax^{(k)}, r^{(k)}) + \alpha\|r^{(k)}\|_A^2 \\
&= -(r^{(k)}, r^{(k)}) + \alpha\|r^{(k)}\|_A^2 && (r^{(k)} = b - Ax^{(k)}) \\
&= -\|r^{(k)}\|^2 + \alpha\|r^{(k)}\|_A^2.
\end{aligned}$$

Par conséquent $\varphi'(\alpha) = 0 \iff \alpha = \frac{\|r^{(k)}\|_2^2}{\|r^{(k)}\|_A^2}$, ce qui prouve (3.21).

Exercice 3.20 (Méthode du gradient conjugué)**.** *Utiliser la méthode CG avec donnée initiale* $x^{(0)} = (0,0,0)^T$ *pour résoudre le système linéaire* $Ax = b$ *avec*

$$A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \qquad b = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}.$$

Par brévité on note $(\cdot,\cdot)$ le produit interne canonique de $\mathbb{R}^n$ et $(\cdot,\cdot)_A = (A\cdot,\cdot)$. On a $r^{(0)} = w^{(0)} = (-1,2,-1)^T$ et

$$\alpha^{(1)} = \frac{3}{10}, \qquad x^{(1)} = \frac{1}{10}\begin{pmatrix}-3\\6\\-3\end{pmatrix}, \qquad r^{(1)} = \frac{1}{5}\begin{pmatrix}1\\1\\1\end{pmatrix}, \qquad \beta^{(1)} = -\frac{1}{50}, \qquad w^{(1)} = \frac{1}{50}\begin{pmatrix}9\\12\\9\end{pmatrix},$$
$$\alpha^{(2)} = \frac{5}{3}, \qquad x^{(2)} = \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad r^{(2)} = \begin{pmatrix}0\\0\\0\end{pmatrix}.$$

La méthode a donc convergé à la deuxième itération.

Exercice 3.21 (Paramètre d'accélération pour la méthode CG)**.** *Avec les techniques de l'Exercice 3.19 montrer que le choix* (3.25) *minimise la fonctionnelle* (3.18) *pour des arguments de la forme $x^{(k)} + \alpha w^{(k)}$ (avec $x^{(k)}$ et $w^{(k)}$ fixés).*

Exercice 3.22 (Mise en œuvre efficace de la méthode CG)**.** *On se place à l'itération k de la méthode CG et on suppose $\alpha^{(k+1)} \neq 0$). Prouver la formule récursive suivante, dont l'intérêt est d'éviter le produit matrice-vecteur $Aw^{(k)}$ pour le calcul de $r^{(k+1)}$ (voir la formule* (3.24)*) :*

$$r^{(k+1)} = -\alpha^{(k+1)} A r^{(k)} + \alpha^{(k+1)}\left(\frac{1}{\alpha^{(k+1)}} - \frac{\beta^{(k)}}{\alpha^{(k)}}\right) r^{(k)} + \alpha^{(k+1)}\frac{\beta^{(k)}}{\alpha^{(k)}} r^{(k-1)}.$$

On pourra remarquer que le résultat du produit matrice-vecteur $Ar^{(k)}$ peut être obtenu sans coût additionnel en mettant en œuvre de manière opportune l'étape (3.28d).