

La décroissance de $A \succ 0 \mapsto A^{-1}$ par plusieurs faces[†]

par Jean-Baptiste Hiriart-Urruty

Université Paul Sabatier de Toulouse <http://www.math.univ-toulouse.fr/~jbhu/>

RÉSUMÉ. *On démontre que si deux matrices réelles symétriques A et B vérifient $B \succ A \succ 0$, alors $A^{-1} \succ B^{-1}$; on le fait de manières différentes, en utilisant les ressources de l'algèbre linéaire ou bilinéaire, en suivant la voie de l'optimisation, en utilisant les résultats de calcul différentiel. Chaque manière de faire a son propre intérêt.*

ABSTRACT. The decrease of $A \succ 0 \mapsto A^{-1}$ in different ways

We prove that if two symmetric real matrices A and B satisfy $B \succ A \succ 0$, then $A^{-1} \succ B^{-1}$; this is done in different ways, using the resources of linear or bilinear algebra, following the path of optimization, using the results of differential calculus. Every way of doing things has its own interest.

MOTS-CLÉS : *matrice définie positive, inversion de matrice.*

Introduction

Commençons par décrire le contexte de travail. Nous désignons par $\mathcal{S}_n(\mathbb{R})$ l'ensemble des matrices symétriques réelles de taille $n \times n$. Cet espace vectoriel devient euclidien grâce au

produit scalaire naturel $\langle\langle U, V \rangle\rangle := \text{tr}(UV) = \sum_{i,j=1}^n u_{ij}v_{ij}$.

Une matrice $A \in \mathcal{S}_n(\mathbb{R})$ est dite *définie positive*, ce qu'on note $A \succ 0$, lorsque $\langle Ax, x \rangle > 0$ pour tout $x \neq 0$ dans \mathbb{R}^n ; $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel dans \mathbb{R}^n . L'ensemble des matrices définies positives, noté $\mathcal{S}_n^>(\mathbb{R})$, constitue un cône convexe ouvert de sommet (ou pointe) l'origine. Dit schématiquement, $\mathcal{S}_n^>(\mathbb{R})$ est à $\mathcal{S}_n(\mathbb{R})$ ce que le cône convexe ouvert $\mathbb{R}_{>}^n = \{x = (x_1, \dots, x_n); \forall i \in \{1, \dots, n\} \quad x_i > 0\}$ est à \mathbb{R}^n .

Plusieurs opérations matricielles préservent la définie positivité, mentionnons succinctement : l'addition ($A + B$ est définie positive dès que A et B le sont); le passage à l'inverse (A est

[†]2010 Mathematics Subject Classification : 15B48

définie positive si et seulement si A^{-1} l'est) ; l'addition parallèle ($A//B = (A^{-1} + B^{-1})^{-1}$ est définie positive lorsque A et B le sont). Parmi ces opérations, c'est certainement l'addition parallèle qui est la moins connue ; elle fut l'objet d'un problème de concours que nous avons concocté il y a quelques années ([4]).

L'adhérence de $\mathcal{S}_n^>(\mathbb{R})$ est l'ensemble des matrices *positives*, que l'on appelle aussi **semi-définies positives** ; ce sont les matrices $A \in \mathcal{S}_n(\mathbb{R})$ qui vérifient $\langle Ax, x \rangle \geq 0$ pour tout x dans \mathbb{R}^n . À son tour, ce cône convexe fermé, noté $\mathcal{S}_n^{\geq}(\mathbb{R})$, est à $\mathcal{S}_n(\mathbb{R})$ ce que le cône $\{x = (x_1, \dots, x_n) ; \forall i \in \{1, \dots, n\} \ x_i \geq 0\}$, appelé parfois « orthant positif », est à \mathbb{R}^n . On notera la positivité de A par $A \succcurlyeq 0$. L'ensemble $\mathcal{S}_n(\mathbb{R})$, muni de \succcurlyeq , est un espace vectoriel (partiellement) ordonné.

Les sciences de l'ingénierie et les applications des mathématiques (via les matrices de variance-covariance en Statistique, les fonctions de *Lyapounov* en Automatique, les noyaux en Apprentissage automatique (*Machine learning*), etc.) montrent sans ambiguïté que les matrices positives (resp. définies positives) sont la généralisation (matricielle) la plus importante de la notion de réel positif ou de vecteur à coefficients positifs (resp. strictement positifs). L'article de synthèse ([5]) passe en revue les propriétés utiles, en insistant sur les caractérisations et les propriétés géométriques des cônes $\mathcal{S}_n^>(\mathbb{R})$ et $\mathcal{S}_n^{\geq}(\mathbb{R})$. Une caractérisation inhabituelle de la définie positivité de A fait même répartir les contributions de A et A^{-1} sur un sous-espace vectoriel H et son orthogonal H^\perp (cf. [3]). Par ailleurs, l'Optimisation, notamment celle qui apparaît en Statistique, est gourmande de l'utilisation de matrices définies positives ; nous avons eu l'occasion d'en proposer plusieurs exemples dans la rubrique Questions-Réponses de cette revue.

L'objet de la présente note est la transformation $A \mapsto A^{-1}$ de $\mathcal{S}_n^>(\mathbb{R})$ dans lui-même, plus particulièrement sa décroissance stricte : si $B \succ A$ (c'est-à-dire si $B - A \succ 0$), alors $A^{-1} \succ B^{-1}$. Comme une montagne qu'on aborde par plusieurs faces, cette propriété, certes connue, est démontrée de quatre manières différentes, en s'appuyant chaque fois sur des ressources (résultats, techniques) mathématiques différentes, ce qui en fait leur intérêt.

Première méthode. Points d'appui : algèbre linéaire, algèbre bilinéaire.

C'est certainement la manière de procéder la plus classique, celle qui vient à l'esprit en premier. Puisqu'il n'y a que deux matrices symétriques A et B en jeu, et que l'une est définie positive, on peut penser à utiliser la technique de réduction simultanée par congruence. La forme que nous en prenons est la suivante ([6, Theorem 7.6.4]) : il existe une matrice inversible P telle que

$$P^T A P = \text{diag}(a_1, \dots, a_n) \text{ et } P^T B P = \text{diag}(b_1, \dots, b_n). \quad (1)$$

En clair, modulo le changement de variables $y = P^{-1}x$, les deux formes quadratiques en jeu sont réduites à

$$\langle \text{diag}(a_1, \dots, a_n)y, y \rangle = \sum_{i=1}^n a_i y_i^2 \text{ et } \langle \text{diag}(b_1, \dots, b_n)y, y \rangle = \sum_{i=1}^n b_i y_i^2.$$

Ici, comme A et B sont définies positives, les a_i et b_i sont des réels strictement positifs. Avoir $B \succ A$ se traduit facilement en $\text{diag}(b_1, \dots, b_n) \succ \text{diag}(a_1, \dots, a_n)$, c'est-à-dire $b_i > a_i$ pour tout $i = 1, \dots, n$. Par suite, $\frac{1}{b_i} < \frac{1}{a_i}$ pour tout i , soit $\text{diag}\left(\frac{1}{a_1}, \dots, \frac{1}{a_n}\right) \succ \text{diag}\left(\frac{1}{b_1}, \dots, \frac{1}{b_n}\right)$, ce qui, en faisant le cheminement matriciel inverse au précédent, conduit à $A^{-1} \succ B^{-1}$.

Remarques. \triangleright Dans cette première méthode, nous avons délibérément contourné l'utilisation de la racine carrée $C^{1/2}$ d'une matrice positive C ; mais c'est aussi une voie possible, avec des variantes d'ailleurs.

\triangleright On peut facilement se faire piéger en pensant aux valeurs propres de A et B . Si $B \succcurlyeq A \succ 0$, et que les valeurs propres de B et A sont rangées en ordre décroissant, $\lambda_1 \geq \dots \lambda_n > 0$ et $\mu_1 \geq \dots \mu_n > 0$ respectivement, certes on a $\lambda_i \geq \mu_i$ pour tout i dès que $B \succcurlyeq A$, mais ceci n'est pas une caractérisation de $B \succcurlyeq A$.

Deuxième méthode ([7]). *Points d'appui : propriétés du produit scalaire, calcul matriciel.*

On se contente ici de démontrer que si $B \succcurlyeq A \succ 0$, alors $A^{-1} \succcurlyeq B^{-1}$.

Puisque $\langle Au, u \rangle \geq 0$ pour tout $u \in \mathbb{R}^n$, choisissons $u = A^{-1}x - y$, où x et y sont quelconques dans \mathbb{R}^n . C'est évidemment là que réside l'astuce ! On a :

$$\langle A(A^{-1}x - y), A^{-1}x - y \rangle \geq 0 \text{ pour tout } x \text{ et } y \text{ dans } \mathbb{R}^n.$$

En développant, puis en utilisant le fait que $\langle Ay, A^{-1}x \rangle = \langle x, y \rangle$, cela donne

$$\langle Ay, y \rangle - 2 \langle x, y \rangle + \langle x, A^{-1}x \rangle \geq 0 \text{ pour tout } x \text{ et } y \text{ dans } \mathbb{R}^n.$$

Comme $B \succcurlyeq A$, on en déduit :

$$\langle By, y \rangle - 2 \langle x, y \rangle + \langle x, A^{-1}x \rangle \geq 0 \text{ pour tout } x \text{ et } y \text{ dans } \mathbb{R}^n. \tag{2}$$

Il suffit de choisir $y = B^{-1}x$ dans (2) pour obtenir :

$$\langle x, A^{-1}x \rangle \geq \langle x, B^{-1}x \rangle \text{ pour tout } x \text{ dans } \mathbb{R}^n.$$

Troisième méthode. *Points d'appui : techniques d'optimisation, de calcul différentiel.*

Dois-je commencer par avouer que c'est la méthode que je préfère ? L'idée est de chercher à obtenir $\frac{1}{2} \langle A^{-1}p, p \rangle$ comme valeur optimale dans un problème d'optimisation très simple où apparaît $\frac{1}{2} \langle Ax, x \rangle$.

De manière plus précise, à partir de la fonction quadratique $f : x \mapsto \frac{1}{2} \langle A^{-1}x, x \rangle$, on définit sa transformée de *Legendre-Fenchel* comme suit : pour tout $p \in \mathbb{R}^n$ (p est une pente de droite dans l'interprétation géométrique de cette transformation, p est un prix dans l'utilisation qu'en font les économistes),

$$f^*(p) = \sup_{x \in \mathbb{R}^n} [\langle p, x \rangle - f(x)]. \tag{3}$$

Si, dans le problème de maximisation au-dessus, la fonction f est convexe, différentiable, et vérifie $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$, on peut aisément exprimer $f^*(p)$: c'est $\langle p, x_p \rangle - f(x_p)$, où x_p est solution de l'équation $\nabla f(x) = p$. Lorsque cette solution est unique, ce qui sera notre cas, on la note $(\nabla f)^{-1}(x)$ et on obtient

$$f^*(p) = \left\langle p, (\nabla f)^{-1}(x) \right\rangle - f[(\nabla f)^{-1}(x)]. \tag{4}$$

C'est cette formulation (4) qu'utilise *A.-M. Legendre* dans la transformation qui porte son nom, et qui a été généralisée sous le format (3) à toute fonction par le géomètre *W. Fenchel*. La transformation $f \rightsquigarrow f^*$ est ainsi appelée transformation de *Legendre-Fenchel*. C'est une belle involution dans la classe des fonctions convexes. Dans le cas qui nous préoccupe, elle prend une forme très simple : l'équation d'optimalité $\nabla f(x) = Ax = p$ a une seule solution qui est $p = A^{-1}x$; il en résulte que $f^*(p) = \frac{1}{2} \langle A^{-1}p, p \rangle$.

De par la définition (3), il est clair que $f^* \geq g^*$ dès que $f \leq g$. Par conséquent :

$$\begin{aligned} \frac{1}{2} \langle A^{-1}p, p \rangle &\geq \frac{1}{2} \langle B^{-1}p, p \rangle \text{ pour tout } p \in \mathbb{R}^n \\ &\text{dès que} \\ \frac{1}{2} \langle Ax, x \rangle &\leq \frac{1}{2} \langle Bx, x \rangle \text{ pour tout } x \in \mathbb{R}^n. \end{aligned}$$

Nous avons en fait démontré très rapidement que : $(B \succ A) \Rightarrow (A^{-1} \succcurlyeq B^{-1})$. Une remarque additionnelle simple permet de conclure à l'inégalité stricte $A^{-1} \succ B^{-1}$. En effet, si l'on suppose $B \succ A$,

$$\langle p, x \rangle - \frac{1}{2} \langle Bx, x \rangle < \langle p, x \rangle - \frac{1}{2} \langle Ax, x \rangle \text{ pour tout } x \neq 0.$$

En particulier, pour $p \neq 0$, on peut prendre $\bar{x} = B^{-1}p$ pour lequel le membre de gauche (une fonction de x) atteint son maximum. On en déduit que

$$\sup_{x \in \mathbb{R}^n} \left[\langle p, x \rangle - \frac{1}{2} \langle Bx, x \rangle \right] < \sup_{x \in \mathbb{R}^n} \left[\langle p, x \rangle - \frac{1}{2} \langle Ax, x \rangle \right].$$

D'où le résultat voulu.

La même définition (3) a d'autres conséquences immédiates, comme $f^*(p) + f(x) \geq \langle p, x \rangle$ pour tout $(p, x) \in \mathbb{R}^n \times \mathbb{R}^n$. Dans le cas de notre exemple, cela donne une inégalité bien connue :

$$A + A^{-1} \succeq 2I_n.$$

Remarque. Une autre voie, toujours dans le même esprit, celui d'exprimer $\langle A^{-1}p, p \rangle$ comme valeur optimale dans un problème d'optimisation simple où apparaît $\langle Ax, x \rangle$ est comme suit ([2, Exercice 7.15]) :

$$\text{pour tout } p \neq 0, \frac{1}{\langle A^{-1}p, p \rangle} = \inf_{x \in \mathbb{R}^n, \langle p, x \rangle \neq 0} \frac{\langle Ax, x \rangle}{(\langle p, x \rangle)^2}. \quad (5)$$

La valeur optimale dans le membre de droite de (5) est obtenue pour $x = A^{-1}p$.

Quatrième méthode. *Points d'appui : techniques de calcul matriciel, de calcul différentiel.*

Soit I un intervalle de \mathbb{R} et $t \in I \mapsto A(t) \in \mathcal{S}_n(\mathbb{R})$ une application dérivable. Il est clair que la matrice $A'(t)$, limite de matrices symétriques, est à son tour symétrique.

Supposons $A'(t) \succ 0$ pour tout $t \in I$, et voyons ce qu'on peut en tirer.

▷ Première conséquence. L'application $t \in I \mapsto A(t)$ est strictement croissante, c'est-à-dire :

$$t_1 < t_2 \implies A(t_2) \succ A(t_1). \quad (6)$$

Démonstration. Soit, pour $x \neq 0$ dans \mathbb{R}^n , la fonction $q_x : t \in I \mapsto q_x(t) = \langle A(t)x, x \rangle$. Alors, q_x est dérivable sur I avec $q'_x(t) = \langle A'(t)x, x \rangle$. Comme $q'_x(t) > 0$ pour tout $t \in I$ (cf. hypothèse faite sur $A'(t)$), q_x est strictement croissante sur I . Ainsi,

$$t_1 < t_2 \implies \langle A(t_2)x, x \rangle > \langle A(t_1)x, x \rangle.$$

Ceci étant assuré pour tout $x \neq 0$, on a bien (6). **cqfd**

▷ Deuxième conséquence. On suppose, de plus, que $A(t)$ est inversible pour tout $t \in I$. Alors, l'application $\mathcal{A}^{-1} : t \in I \mapsto [A(t)]^{-1}$ est strictement décroissante, c'est-à-dire :

$$t_1 < t_2 \implies [A(t_1)]^{-1} \succ [A(t_2)]^{-1}.$$

Démonstration. L'application \mathcal{A}^{-1} est dérivable, comme composée d'applications dérivables, et l'on a l'égalité $(\mathcal{A}^{-1})'(t) = -[A(t)]^{-1} A'(t) [A(t)]^{-1}$. Comme $A'(t) \succ 0$, l'application $t \mapsto [A(t)]^{-1} A'(t) [A(t)]^{-1}$ vérifie elle aussi

$$\left\langle [A(t)]^{-1} A'(t) [A(t)]^{-1} x, x \right\rangle = \left\langle A'(t) [A(t)]^{-1} x, [A(t)]^{-1} x \right\rangle > 0.$$

Comme ci-dessus, on en déduit que $t \in I \mapsto [A(t)]^{-1}$ est strictement décroissante. **cqfd**

Une application. Choisissons $B \succ A \succ 0$ et considérons $t \in [0, 1] \mapsto A(t) = A + t(B - A)$. Cette application est dérivable, avec $A'(t) = B - A \succ 0$ pour tout $t \in [0, 1]$. De plus, la matrice $A(t)$ est définie positive (comme barycentre convexe de deux matrices définies positives), donc inversible pour tout $t \in [0, 1]$. En conséquence, vu ce qui a été démontré plus haut, l'application $\mathcal{A}^{-1} : t \in [0, 1] \mapsto [A(t)]^{-1} = [A + t(B - A)]^{-1}$ est strictement décroissante ; en particulier

$$[A(0)]^{-1} \succ [A(1)]^{-1}, \text{ soit } A^{-1} \succ B^{-1}.$$

Conclusion

L'analogie de $\mathcal{S}_n(\mathbb{R})$ avec \mathbb{R} , celle de $\mathcal{S}_n^{\geq}(\mathbb{R})$ avec \mathbb{R}_{\geq} , celle des opérations sur les matrices avec celles correspondantes sur les réels, sont tentantes... mais il faut parfois s'en méfier. C'est l'occasion de signaler quelques ressemblances et dissemblances, pour $n \geq 2$, ([1],[6]).

▷ Le théorème de la borne supérieure comme de la borne inférieure tombent en défaut dans l'ensemble ordonné $(\mathcal{S}_n(\mathbb{R}), \succ)$.

▷ Toute suite croissante majorée dans $\mathcal{S}_n(\mathbb{R})$ converge.

▷ L'application $M \mapsto M^2$ de $\mathcal{S}_n^{\geq}(\mathbb{R})$ dans lui-même n'est pas croissante.

▷ L'application $M \mapsto M^{1/2}$ de $\mathcal{S}_n^{\geq}(\mathbb{R})$ dans lui-même est croissante.

▷ La décroissance stricte de l'application de $M \mapsto M^{-1}$ de $\mathcal{S}_n^>(\mathbb{R})$ dans lui-même est vraie, c'était l'objet de cette note que de la montrer par quelques voies différentes.

▷ La fonction $A \succ 0 \mapsto \ln(\det A)$, de $\mathcal{S}_n^>(\mathbb{R})$ dans \mathbb{R} , est la cousine matricielle de la fonction logarithme usuelle $a > 0 \mapsto \ln a$. Elle est strictement concave et son gradient en A , dans l'espace euclidien $(\mathcal{S}_n(\mathbb{R}), \langle \langle \cdot, \cdot \rangle \rangle)$, est A^{-1} ; belle analogie avec $\ln'(a) = \frac{1}{a}$. Toutefois, contrairement au cas où $n = 1$, cette stricte concavité ne permet d'accéder qu'à une « décroissance sur les produits scalaires », c'est-à-dire : pour deux matrices A et B de $\mathcal{S}_n^>(\mathbb{R})$, comparables ou pas, on a

$$\langle \langle B^{-1} - A^{-1}, B - A \rangle \rangle \leq 0. \quad (7)$$

Signalons pour terminer que la deuxième et troisième manières de faire sont valables en dimension infinie : si A et B sont deux applications linéaires continues, auto-adjointes, bijectives, d'un espace de Hilbert H dans lui-même, alors : $(B \succ A) \iff (A^{-1} \succ B^{-1})$.

Références

- [1] Exbrayat J.-M. *Réponse P23 RMS* n° 106, p. 766-768. (avril 1996)

- [2] Hiriart-Urruty J.-B. et Plusquellec Y. *Exercices d'algèbre linéaire et bilinéaire (Bac + 2)*. Cepadues-Éditions. (1988)
- [3] Hiriart-Urruty J.-B. *Une nouvelle caractérisation des formes quadratiques définies positives sur \mathbb{R}^n* . RMS n° 97, p. 247-249 (1986-87)
- [4] Hiriart-Urruty J.-B. *Première composition de mathématiques (Filière PC)*. Concours d'admission à l'École Polytechnique. (1999)
- [5] Hiriart-Urruty J.-B. et Malick J. *A Fresh Variational-Analysis Look at the Positive Semidefinite Matrices World*. Journal of Optimization Theory and Applications n° 153, p. 551-577. (2012)
- [6] Horn R.A. and Johnson C.R. *Matrix Analysis*. Cambridge University Press. (1996)
- [7] Toda A.A. *Operator reverse monotonicity of the inverse*. American Math. Monthly n° 118, p. 82-83. (2011)