

SOLUTIONS

1 Notion de direction de descente

Les courbes de niveau étant des ellipses, la fonction est soit strictement convexe, soit strictement concave. La direction du gradient en y montre qu'elle est strictement concave (faire une coupe le long de la droite $y + \mathbb{R}\nabla f(y)$). Donc la courbe de niveau encercle une « colline » de f . C'est donc dans le dessin A que d est de descente.

2 Pas optimal pour un critère quadratique

1. Soit $x_{k+1} = x_k + \alpha_k d_k$. Le pas α_k est optimal si

$$0 = f'(x_{k+1}) \cdot d_k = (g_k + \alpha_k A d_k)^\top d_k.$$

On en déduit

$$\alpha_k = -\frac{g_k^\top d_k}{d_k^\top A d_k},$$

qui est > 0 , car d_k est une direction de descente ($g_k^\top d_k < 0$) et A est définie positive.

2. On a

$$\begin{aligned} f_{k+1} &= f_k + \alpha_k g_k^\top d_k + \frac{\alpha_k^2}{2} d_k^\top A d_k, \\ g_{k+1} &= g_k + \alpha_k A d_k. \end{aligned}$$

En multipliant scalairement la seconde équation par $\frac{\alpha_k}{2} d_k$, en observant que $g_{k+1}^\top d_k = 0$ et en retranchant la première équation, on obtient

$$f_{k+1} - f_k = \frac{\alpha_k}{2} g_k^\top d_k,$$

si bien que

$$\alpha_k = \frac{-2(f_k - f_{k+1})}{g_k^\top d_k}, \quad (5.19)$$

qui ne fait plus intervenir A .

Remarque. La formule (5.19) peut donner une estimation du pas à prendre le long de d_k pour une fonction non linéaire quelconque, à condition que l'on ait une idée de la décroissance de f (c'est souvent le cas). Le pas α_k donné par (5.19) est appelé **pas de Fletcher**.

3 Mise en œuvre de la règle de Wolfe

1. Si on passe indéfiniment dans 2.7, on a $\alpha \rightarrow +\infty$ et

$$f(x_k + \alpha d_k) \leq f_k + \omega_1 \alpha g_k^\top d_k,$$

ce qui impliquerait que f n'est pas bornée inférieurement sur $x_k + \mathbb{R}d_k$. Par conséquent, après un nombre fini d'itérations, $\bar{\alpha} < +\infty$.

2. A chaque itération, $|\bar{\alpha} - \underline{\alpha}|$ est divisé par deux, donc les suites croissante ($\underline{\alpha}$) et décroissante ($\bar{\alpha}$) convergent vers la même valeur $\hat{\alpha}$.
3. On passe à la limite dans

$$f(x_k + \bar{\alpha}d_k) > f_k + \bar{\alpha}\omega_1 g_k^\top d_k \quad (5.20)$$

et dans

$$f(x_k + \underline{\alpha}d_k) \leq f_k + \underline{\alpha}\omega_1 g_k^\top d_k.$$

4. Par passage à la limite dans

$$g(x_k + \underline{\alpha}d_k)^\top d_k \geq \omega_2 g_k^\top d_k.$$

5. En retranchant (5.3) de (5.20), en divisant le résultat par $(\bar{\alpha} - \hat{\alpha})$ et en passant à la limite, on trouve le résultat.
6. De (5.5) et $\omega_1 < \omega_2$, on déduit que

$$g(x_k + \hat{\alpha}d_k)^\top d_k > \omega_2 g_k^\top d_k,$$

ce qui contredit (5.4).

Remarque. En pratique, aux étapes 2.2 et 2.8, on utilise des formules d'interpolation avec sauvegardes plutôt que de diviser l'intervalle $[\underline{\alpha}, \bar{\alpha}]$ par 2.

4 Acceptation du pas unité dans les méthodes de Newton et de quasi-Newton

1. On utilise le développement de Taylor de $f(x_k + d_k)$ autour de x_k :

$$\begin{aligned} & f(x_k + d_k) \\ &= f(x_k) + g_k^\top d_k + \frac{1}{2} d_k^\top \nabla^2 f(x_k) d_k + \frac{1}{2} \int_0^1 (f''(x_k + td_k) - f''(x_k)) \cdot d_k^2 dt \\ &= f(x_k) + g_k^\top d_k + \frac{1}{2} d_k^\top \nabla^2 f(x_k) d_k + o(\|d_k\|^2). \end{aligned}$$

Alors

$$\begin{aligned} & f(x_k + d_k) - f(x_k) - \omega_1 g_k^\top d_k \\ &= (1 - \omega_1) g_k^\top d_k + \frac{1}{2} d_k^\top \nabla^2 f(x_k) d_k + o(\|d_k\|^2) \\ &= -\left(\frac{1}{2} - \omega_1\right) d_k^\top M_k d_k + \frac{1}{2} d_k^\top (\nabla^2 f(x_k) - M_k) d_k + o(\|d_k\|^2) \\ &\leq -\left(\frac{1}{2} - \omega_1\right) d_k^\top M_k d_k + o(\|d_k\|^2) \\ &\leq 0, \end{aligned}$$

pour k assez grand. On a utilisé le fait qu'il existe une constante $C > 0$ telle que $d_k^\top M_k d_k \geq C\|d_k\|^2$, ce que l'on voit en utilisant (5.6) et le fait que $\nabla^2 f(x_*)$ est définie positive.

2. On utilise le développement de Taylor de $g(x_k + d_k)$ autour de x_k :

$$\begin{aligned} g(x_k + d_k) &= g_k + \nabla^2 f(x_k) d_k + \int_0^1 (\nabla^2 f(x_k + t d_k) - \nabla^2 f(x_k)) d_k dt \\ &= g_k + \nabla^2 f(x_k) d_k + o(\|d_k\|). \end{aligned}$$

On en déduit

$$\begin{aligned} g(x_k + d_k)^\top d_k - \omega_2 g_k^\top d_k &= -\omega_2 g_k^\top d_k + g_k^\top d_k + d_k^\top \nabla^2 f(x_k) d_k + o(\|d_k\|^2) \\ &= \omega_2 d_k^\top M_k d_k - d_k^\top (M_k - \nabla^2 f(x_k)) d_k + o(\|d_k\|^2) \\ &\geq \omega_2 C \|d_k\|^2 + o(\|d_k\|^2) \\ &\geq 0, \end{aligned}$$

pour k assez grand. On a utilisé le fait qu'il existe une constante $C > 0$ telle que $d_k^\top M_k d_k \geq C \|d_k\|^2$, ce que l'on voit en utilisant (5.7) et le fait que $\nabla^2 f(x_*)$ est définie positive.

5 Problème de moindres-carrés linéaire

1. On peut donner beaucoup de démonstrations. Commençons par les plus longues !

- – C'est clair si A est injective car alors, avec $\lambda = \lambda_{\min}(A^\top A) > 0$, on a

$$\begin{aligned} f(x) &= \frac{1}{2} x^\top A^\top A x - b^\top A x + \frac{1}{2} \|b\|^2 \\ &\geq \frac{\lambda}{2} \|x\|^2 - \|A^\top b\| \|x\| + \frac{1}{2} \|b\|^2, \end{aligned}$$

si bien que $f(x) \rightarrow +\infty$ lorsque $\|x\| \rightarrow \infty$.

- Si $A = 0$, tout point de \mathbb{R}^n est solution.
- Si A est non nulle, mais n'est pas injective, on décompose \mathbb{R}^n en $\mathcal{N}(A)$ (dans lequel f ne varie pas) et le sous-espace vectoriel supplémentaire $\mathcal{R}(A^\top) \neq \{0\}$. Si les colonnes de Y forment une base de $\mathcal{R}(A^\top)$ et les colonnes de Z forment une base de $\mathcal{N}(A)$, on a

$$\mathbb{R}^n = \mathcal{R}(Y) \oplus \mathcal{R}(Z).$$

Donc x s'écrit $Y u + Z v$ et le problème revient à minimiser en u

$$\frac{1}{2} \|A Y u - b\|^2.$$

Mais $A Y$ est injective (car $A Y u = 0$ implique que $Y u \in \mathcal{R}(A^\top) \cap \mathcal{N}(A) = \{0\}$ et donc $u = 0$ car Y est injective). Dès lors ce problème a une solution (par la première partie), soit u_* , et les solutions du problème sont de la forme

$$x_* + \mathcal{N}(A),$$

où $x_* = Y u_*$ (cela provient de l'unicité de u_* , voir plus loin).

- Au lieu d'utiliser la démonstration « topologique » ci-dessus, on utilise une démonstration « calculatoire », en écrivant la condition d'optimalité et en montrant que celle-ci a toujours une solution. Elle s'écrit

$$A^T Ax = A^T b. \quad (5.21)$$

Elle a une solution car $A^T b \in \mathcal{R}(A^T) = \mathcal{R}(A^T A)$ (en effet, on vérifie facilement que $\mathcal{N}(A) = \mathcal{N}(A^T A)$ et on « dualise » cette relation – on prend l'orthogonal). L'équation (5.21) est appelée l'*équation normale* du problème de moindres-carrés linéaire. Elle est donc équivalente à ce problème.

- On observe que le problème consiste à projeter b sur $\mathcal{R}(A)$ (qui est un convexe fermé non vide!). Il existe une unique projection. Celle-ci est de la forme Ax_* (le x_* n'est pas nécessairement unique cependant) et x_* est solution du problème de moindres-carrés.
 - On utilise l'exercice 9 de la séance 1, en notant que le critère est quadratique et borné inférieurement.
2. Si A est injective, f est strictement convexe et le problème a une solution unique. Inversement, si A n'est pas injective et si x_* est solution, tout point de $x_* + \mathcal{N}(A)$ est encore solution.
 3. On a montré ci-dessus que l'ensemble des solutions est de la forme $x_* + \mathcal{N}(A)$, où x_* est une solution arbitraire (dont l'existence a été démontrée).

Comme $Ax_* = P_{\mathcal{R}(A)} b$ (projection orthogonale de b sur l'image de A), la valeur optimale s'écrit $\text{val}(P) = \frac{1}{2} \|(I - P_{\mathcal{R}(A)})b\|^2$. On note que $I - P_{\mathcal{R}(A)} = P_{\mathcal{R}(A)^\perp}$, si bien que

$$\text{val}(P) = \frac{1}{2} \|P_{\mathcal{R}(A)^\perp} b\|^2.$$

4. L'équation d'optimalité du premier ordre s'écrit

$$A^T Ax = A^T b.$$

Elle porte le nom d'**équation normale**. Comme le problème est convexe l'ensemble des solutions du problème de moindres-carré est l'ensemble des solutions de cette équation.

- (a) On projette 0 sur un sous-espace affine, qui est certainement un convexe fermé non vide!
- (b) La problème (5.9) est convexe et les contraintes sont qualifiées (car linéaires). Alors \hat{x} est solution de (5.9) si, et seulement si, il existe un $\lambda \in \mathbb{R}^m$ (non nécessairement unique) tel que

$$\begin{cases} \hat{x} + A^T A \lambda = 0 \\ A^T A \hat{x} = A^T b \end{cases} \quad \text{ou} \quad \begin{pmatrix} I & A^T A \\ A^T A & 0 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ A^T b \end{pmatrix}. \quad (5.22)$$

Comme \hat{x} est déterminé de manière unique par cette équation (pas λ si A n'est pas injective), il est fonction linéaire du second membre, lui-même fonction linéaire de b . Donc \hat{x} dépend linéairement de b .

(c) Par les conditions d'optimalité, $\hat{x} \in \mathcal{R}(A^\top)$, clairement.

Inversement, x est solution de (5.8), il vérifie l'équation normale $A^\top Ax = A^\top b$. Si, de plus, $x \in \mathcal{R}(A^\top)$, alors il existe un $\lambda \in \mathbb{R}^n$ tel que $x = -A^\top A \lambda$ (car $\mathcal{R}(A^\top) = \mathcal{R}(A^\top A)$). Alors (x, λ) vérifie les conditions d'optimalité de (5.9), qui sont suffisantes pour affirmer que x est la solution de norme minimale.

- (d) i. Si $A = 0$, $A^\dagger = 0$ (car $\hat{x} = 0$ quel que soit b).
 ii. Comme $\hat{x} \in \mathcal{R}(A^\top)$, on peut écrire $\hat{x} = Yu$, ce qui donne (on a vu ou on montre facilement que $Y^\top A^\top AY$ est inversible – on a besoin que Y existe ici et donc que $A \neq 0$)

$$u = (Y^\top A^\top AY)^{-1} Y^\top A^\top b$$

et donc

$$\hat{x} = Y(Y^\top A^\top AY)^{-1} Y^\top A^\top b.$$

Forcément

$$\boxed{A^\dagger = Y(Y^\top A^\top AY)^{-1} Y^\top A^\top.}$$

- iii. On vérifie d'abord que toute solution x de l'équation normale vérifie $U^\top x = \Sigma^{-1} V^\top b$. Quant à \hat{x} il est dans $\mathcal{R}(A^\top) = \mathcal{R}(U)$, si bien que $\hat{x} = Uu$. On déduit de la première équation de (5.22) que $u = \Sigma^{-1} V^\top b$ et donc que $\hat{x} = U \Sigma^{-1} V^\top b$. Dès lors

$$\boxed{A^\dagger = U \Sigma^{-1} V^\top.}$$

6 Problème de moindres-carrés non linéaire : méthode de Gauss-Newton

1. On a

$$\nabla f(x) = J(x)^\top r(x) \quad \text{et} \quad \nabla^2 f(x) = J(x)^\top J(x) + \sum_{i=1}^p r_i(x) \nabla^2 r_i(x).$$

Comme le coût de (5.10) s'écrit

$$d \in \mathbb{R} \mapsto \nabla f(x_k)^\top d + \frac{1}{2} d^\top J(x_k)^\top J(x_k) d,$$

on voit que l'on passe de l'algorithme de Newton à l'algorithme de Gauss-Newton en négligeant le terme $\sum_{i=1}^p r_i(x_k) \nabla^2 r_i(x_k)$ du hessien de f en x_k , ce qui permet de ne pas devoir calculer les dérivées secondes de r .

Il est donc raisonnable de s'attendre à une convergence rapide dans le voisinage d'une solution x_* si $r(x_*)$ est petit ou si les $\nabla^2 r_i(x)$ sont petits (problème qui n'est pas trop non linéaire).

2. (a) Le problème (5.10) a toujours au moins une solution car c'est un problème de moindres-carrés linéaire.
 (b) Il faut ensuite vérifier que d_k est une direction de descente de f en x_k (si x_k n'est pas stationnaire) pour que la recherche linéaire fonctionne. Par optimalité de d_k , on a

$$J_k^\top J_k d_k + J_k^\top r_k = 0,$$

si bien que

$$\nabla f(x_k)^\top d_k = -\|J_k d_k\|^2,$$

qui est < 0 si x_k n'est pas stationnaire (en effet, dans ce cas $J_k d_k \neq 0$).

3. (a) Sous les hypothèses énoncées, $d_k = -(J_k^\top J_k)^{-1} \nabla f_k$ et les suites de matrices

$$\{J_k^\top J_k\} \text{ et } \{(J_k^\top J_k)^{-1}\} \text{ sont bornées.}$$

Il s'en suit que le cosinus de l'angle θ_k entre d_k et $-\nabla f_k$ est uniformément positif.

On peut aussi le montrer directement : en notant $C > 0$ une constante générique, on a $\|J_k^\top r_k\| \leq C \|J_k d_k\|$ et donc

$$\cos \theta_k = \frac{-\nabla f(x_k)^\top d_k}{\|\nabla f(x_k)\| \|d_k\|} = \frac{\|J_k d_k\|^2}{\|J_k^\top r_k\| \|d_k\|} \geq C \frac{\|J_k d_k\|}{\|d_k\|} \geq C,$$

du fait de l'uniforme injectivité de J_k . D'après la condition de Zoutendijk, ceci implique que $\nabla f(x_k) \rightarrow 0$, c'est-à-dire la convergence de l'algorithme.

- (b) D'après l'hypothèse (ii), lorsque x_k est voisin de x_* , $M_k := J_k^\top J_k$ est inversible et d_k est donc donnée par la formule

$$d_k = -M_k^{-1} \nabla f(x_k).$$

Comme, d'après l'hypothèse (i),

$$M_k - \nabla^2 f(x_*) = M_k - J(x_*)^\top J(x_*) = O(\|x_k - x_*\|),$$

et que $\nabla^2 f(x_*) = J(x_*)^\top J(x_*)$ est définie positive, la convergence est quadratique.

Remarque. Cette étude montre que, contrairement à la méthode de Newton, l'algorithme de Gauss-Newton est toujours bien défini et génère toujours des directions de descente. Sa convergence est assurée sous des conditions très souvent vérifiées (la technique des régions de confiance permet de faire face au cas où l'uniforme injectivité de $J(x)$ ne serait pas vérifiée). Par contre, la vitesse de convergence peut être assez lente si le résidu n'est pas nul en la solution ou si r est fortement non linéaire (on s'écarte alors de la méthode de Newton).

7 Approximation de rang un d'une matrice symétrique

1. *Première méthode.* Le critère est continu et tend vers l'infini à l'infini. Pour ce dernier point, on peut utiliser l'une des minoration suivantes du critère $x \mapsto f(x)$:

$$\begin{aligned} f(x) &\geq \frac{1}{2} (\|M\|_F - \|xx^\top\|_F)^2 = \frac{1}{2} (\|M\|_F - \|x\|^2)^2, \\ f(x) &= \frac{1}{2} \|M\|_F^2 - x^\top M x + \frac{1}{2} \|x\|^4 \geq \frac{1}{2} \|M\|_F^2 - \lambda_1 \|x\|^2 + \frac{1}{2} \|x\|^4. \end{aligned}$$

Chaque fonction minorante tend vers l'infini lorsque $\|x\|$ tend vers l'infini.

Deuxième méthode. On réécrit le problème comme un problème sur l'espace vectoriel \mathcal{S}^n :

$$\inf_{\substack{X \in \mathcal{S}_+^n \\ \text{rg}(X) \leq 1}} \frac{1}{2} \|M - X\|_F^2,$$

où $\text{rg}(X)$ désigne le rang de X . Ce problème a une solution car son ensemble admissible est fermé (le rang est s.c.i.) non vide (la matrice nulle est admissible) et son critère tend vers l'infini à l'infini.

2. On a pour tout $h \in \mathbb{R}^n$:

$$f'(x) \cdot h = -2x^\top Mh + 2(x^\top h) \|x\|^2.$$

La condition d'optimalité du premier ordre s'écrit $f'(x) \cdot h = 0$ pour tout $h \in \mathbb{R}^n$, c'est-à-dire

$$Mx = \|x\|^2 x.$$

Donc une solution x est soit le vecteur nul, soit un vecteur propre de M (donc non nul) dont la valeur propre est $\lambda = \|x\|^2$.

Dès lors, si $\lambda_1 \leq 0$ (c'est-à-dire $M \preceq 0$), $x = 0$ (sinon M aurait une valeur propre $\lambda = \|x\|^2 > 0$).

3. La condition du second ordre s'écrit $f''(x) \cdot h^2 \geq 0$ pour tout $h \in \mathbb{R}^n$. Or on a

$$f''(x) \cdot h^2 = -2h^\top Mh + 2\|x\|^2 \|h\|^2 + 4(x^\top h)^2.$$

Alors, x ne peut pas être nul (car $\lambda_1 > 0$ montre que l'on ne peut avoir $f''(0) \cdot h^2 \geq 0$ pour tout h) et donc x est un vecteur propre de M de valeur propre $\lambda = \|x\|^2$ (point 2). Pour $h \perp x$, on trouve

$$f''(x) \cdot h^2 = -2h^\top Mh + 2\|x\|^2 \|h\|^2.$$

On doit donc avoir $h^\top Mh \leq \|x\|^2 \|h\|^2 = \lambda \|h\|^2$ pour tout $h \perp x$. Donc $\lambda = \lambda_1$.

4. Soit \bar{x} une solution du problème, qui est donc un vecteur propre de M associé à la valeur propre maximale λ_1 et de norme $\lambda_1^{1/2}$. Calculons la valeur optimale du problème. On note $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ les valeurs propres de M et v_i des vecteurs propres orthonormaux associés. Sans perte de généralité, on peut supposer que $v_1 \parallel \bar{x}$ et donc $\bar{x} = \pm \lambda_1^{1/2} v_1$. En utilisant la décomposition spectrale $M = \sum_i \lambda_i v_i v_i^\top$, on voit que le coût optimal vaut

$$\text{val}(P) = \frac{1}{2} \left\| \sum_{i=1}^n \lambda_i v_i v_i^\top - \bar{x} \bar{x}^\top \right\|_F^2 = \frac{1}{2} \sum_{i=2}^n \lambda_i^2.$$

Ce coût optimal est obtenu par tout vecteur propre x de valeur propre maximale λ_1 et de norme $\lambda_1^{1/2}$ (on adapte la décomposition spectrale à ce vecteur propre x).

8 Valeurs singulières

1. Il s'agit de maximiser une fonction continue sur un compact (fermé borné) de $\mathbb{E} \times \mathbb{F}$.
2. Si l'on note $c(x, y) = (\frac{1}{2}(\|x\|^2 - 1), \frac{1}{2}(\|y\|^2 - 1))$ la contrainte, on a pour $(\alpha, \beta) \in \mathbb{R}^2$:

$$c'(x, y) \cdot (h, k) = \begin{pmatrix} \langle x, h \rangle \\ \langle y, k \rangle \end{pmatrix} \quad \text{et} \quad c'(x, y)^*(\alpha, \beta) = \begin{pmatrix} \alpha x \\ \beta y \end{pmatrix}.$$

Comme en un point admissible, $x \neq 0$ et $y \neq 0$, $c'(x, y)^*$ est injective. On en déduit que $c'(x, y)$ est surjective et donc que la contrainte est qualifiée.

On peut aussi voir directement que $c'(x, y)$ est surjective, car on peut atteindre (α, β) par $c'(x, y)$ en l'appliquant à $(h, k) = (\alpha x, \beta y)$.

3. [\Rightarrow] On introduit la lagrangien

$$\ell(x, y, \lambda, \mu) = -\langle Ax, y \rangle + \frac{\lambda}{2}(\|x\|^2 - 1) + \frac{\mu}{2}(\|y\|^2 - 1).$$

Les conditions de Lagrange s'écrivent

$$\begin{cases} A^*y = \lambda x \\ Ax = \mu y \\ \|x\|^2 = 1 \\ \|y\|^2 = 1. \end{cases} \quad (5.23)$$

En multipliant la première équation par x , la seconde par y et en tenant compte des contraintes, on voit que $\lambda = \mu$. Donc, avec $z = (x, y)$, on a $\hat{A}z = \lambda z$, ce qui montre que z est un vecteur propre de \hat{A} , de valeur propre λ .

En ce point stationnaire, le critère vaut $\langle Ax, y \rangle = \mu\|y\|^2 = \lambda$.

[\Leftarrow] Si $z = (x, y)$ est un vecteur propre de \hat{A} , de valeur propre λ , vérifiant $\|x\| = \|y\| = 1$, on a $A^*y = \lambda x$ et $Ax = \lambda y$, ce qui montre que les conditions (5.23) sont vérifiées et donc que (x, y) est un point stationnaire de (P) .

4. Si (\bar{x}, \bar{y}) est solution de (P) et $\bar{\lambda} := \langle A\bar{x}, \bar{y} \rangle$, la qualification des contraintes (point 2) et le point 3 montrent que (\bar{x}, \bar{y}) est un vecteur propre de \hat{A} de valeur propre $\bar{\lambda} = \langle A\bar{x}, \bar{y} \rangle$.

On note que $\bar{\lambda} \geq 0$. En effet, si $A = 0$ alors $\bar{\lambda} := \langle A\bar{x}, \bar{y} \rangle = 0$. Sinon, on peut trouver un x de norme 1 tel que $Ax \neq 0$; en prenant $y = Ax/\|Ax\|$, on obtient $\langle Ax, y \rangle = \|Ax\| \neq 0$ et donc la valeur optimale $\bar{\lambda}$ de (P) est positive.

Soit à présent (x, y) un vecteur propre de \hat{A} de valeur propre λ ; il suffit de montrer que $\lambda \leq \bar{\lambda}$.

- Si $x = 0$, alors $\lambda = 0$ (car $Ax = \lambda y$ et $(x, y) \neq 0$) et donc $\lambda \leq \bar{\lambda}$, car on a vu que $\bar{\lambda} \geq 0$.
- On raisonne de même si $y = 0$.
- Maintenant, si $x \neq 0$ et $y \neq 0$, on peut supposer que $\|x\| = \|y\| = 1$. Par le point 3, $\lambda = \langle Ax, y \rangle$, si bien que λ est inférieur à la valeur optimale de (P) qui est $\bar{\lambda}$.

9 Formule de BFGS

1. On a vu à la séance 1 que $\text{ld} \in \overline{\text{Conv}}(\mathcal{S}^n)$; donc aussi $\psi \in \overline{\text{Conv}}(\mathcal{S}^n)$. Alors $M \in \mathcal{S}_{++}^n$ minimise ψ si, et seulement si, $\nabla\psi(M) = 0$. On a pour $H \in \mathcal{S}^n$:

$$\psi'(M) \cdot H = \text{tr } H - \langle M^{-1}, H \rangle = \langle I - M^{-1}, H \rangle.$$

Donc $\nabla\psi(M) = I - M^{-1}$ et l'unique minimiseur de ψ est $M = I$.

2. Si le problème (5.11) a une solution M celle-ci est dans \mathcal{S}_{++}^n et satisfait la contrainte $y = Ms$. Alors $y^T s = s^T Ms > 0$, car $s \neq 0$.
3. Une contrainte affine est toujours qualifiée.

Si on note $c(M) = Ms - y$, on peut aussi vérifier que la jacobienne $c'(M)$ est surjective. Pour $H \in \mathcal{S}^n$, on a $c'(M) \cdot H = Hs$.

Même dans ce cas particulier, il n'est pas si simple de construire explicitement une matrice $H \in \mathcal{S}^n$ (elle doit être symétrique puisque $c : \mathcal{S}^n \rightarrow \mathbb{R}$) telle que $c'(M) \cdot H = y$, un vecteur donné dans \mathbb{R}^n . On peut prendre (c'est la formule SR1) :

$$H = \begin{cases} \frac{yy^\top}{y^\top s} & \text{si } y^\top s \neq 0 \\ I - \frac{(y-s)(y-s)^\top}{\|s\|_2^2} & \text{sinon.} \end{cases}$$

Il est en général plus aisé d'utiliser le procédé non constructif consistant à vérifier que $c'(M)^*$ est injective, ce qui se déduit immédiatement du fait que, pour $z \in \mathbb{R}^n$: $c'(M)^*z = (zs^\top + sz^\top)/2$.

4. Par sa convexité, sa régularité et la qualification de sa contrainte, le problème (5.11) a une solution M si, et seulement si, M vérifie ses conditions d'optimalité du premier ordre. On note

$$W_0 = M_0^{-1}.$$

Le lagrangien du problème est la fonction $\ell : \mathcal{S}^n \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ définie par

$$\ell(M, \lambda) = \psi(W_0^{1/2} M W_0^{1/2}) + \lambda^\top (y - Ms).$$

Calculons-en le gradient par rapport à M en $M \in \mathcal{S}_{++}^n$. Sachant que $\nabla \psi(M) = I - M^{-1}$, on obtient

$$\begin{aligned} \ell'_M(M, \lambda) &= \langle I - (W_0^{1/2} M W_0^{1/2})^{-1}, W_0^{1/2} H W_0^{1/2} \rangle - \lambda^\top H s \\ &= \langle W_0 - M^{-1}, H \rangle - \left\langle \frac{s\lambda^\top + \lambda s^\top}{2}, H \right\rangle. \end{aligned}$$

Donc, en notant $W = M^{-1}$,

$$\nabla_M \ell(M, \lambda) = W_0 - W - \frac{s\lambda^\top + \lambda s^\top}{2}.$$

et l'optimalité s'écrit

$$W = W_0 - \frac{s\lambda^\top + \lambda s^\top}{2}$$

Il reste à déterminer λ en utilisant la contrainte $W y = s$, ce qui conduit à

$$s = W_0 y - \frac{y^\top \lambda}{2} s - \frac{y^\top s}{2} \lambda.$$

En prenant le produit scalaire avec y et en utilisant le fait que $y^\top s > 0$, on trouve $y^\top \lambda = y^\top (W_0 y - s)/(y^\top s)$ qui, injecté dans la relation ci-dessus, fournit la valeur du multiplicateur

$$\lambda = \frac{-2}{y^\top s} (s - W_0 y) + \frac{y^\top (s - W_0 y)}{(y^\top s)^2} s.$$

En utilisant celle-ci dans la formule de W ci-dessus, on trouve

$$W = W_0 + \frac{(s - W_0 y) s^\top + s (s - W_0 y)^\top}{y^\top s} - \frac{y^\top (s - W_0 y)}{(y^\top s)^2} s s^\top. \quad (5.24)$$

Par un calcul laborieux, mais mécanique, on vérifie que M donné par (5.12) est bien l'inverse de W donné par (5.24) (il suffit de vérifier que $MW = I$).

10 Méthodes des régions de confiance

1. La fonction f est continue et le domaine admissible est fermé, borné, non vide.
2. On va écrire les conditions de Kuhn et Tucker. Pour cela, il faut montrer que la contrainte est qualifiée en \bar{x} . Seul le cas où $\|\bar{x}\| = \Delta$ est à considérer. Alors, $\bar{x} \neq 0$ et donc $\nabla c(\bar{x}) = \bar{x} \neq 0$. La contrainte est donc qualifiée.

Il existe alors un multiplicateur $\bar{\lambda}$ tel que les conditions de Kuhn et Tucker soient vérifiées. En remplaçant la contrainte de (5.13) par la contrainte équivalente (mais différentiable) $c(x) \leq 0$, on trouve les conditions (5.14), (5.15) et (5.16).

3. Si $\|\bar{x}\| < \Delta$, \bar{x} est dans l'intérieur du domaine admissible et est donc minimum local du problème sans contrainte. La condition nécessaire du second ordre dit que $\nabla^2 f(\bar{x}) = A$ est semi-définie positive.

D'autre part, (5.16) implique que $\bar{\lambda} = 0$. On a donc démontré (5.17).

4. (i) Si $\|\bar{x}\| = \Delta$, le cône critique (définition 1.20, remarque 1.21) contient les $h \in \mathbb{R}^n$ tels que $c'(\bar{x}) \cdot h = \bar{x}^\top h = 0$. Pour ces h , on a donc (Théorème 1.22)

$$h^\top \nabla_{xx}^2 \ell(\bar{x}, \bar{\lambda}) h = h^\top (A + \bar{\lambda} I) h \geq 0.$$

(ii) C'est clair, car si $\|x\| = \|\bar{x}\|$, alors x est admissible et donc $f(\bar{x}) \leq f(x)$.

(iii) En utilisant (5.14), on a

$$\begin{aligned} f(\bar{x}) &= \frac{1}{2} \bar{x}^\top A \bar{x} - b^\top \bar{x} \\ &= \frac{1}{2} \bar{x}^\top A \bar{x} - \bar{x}^\top A \bar{x} - \bar{\lambda} \|\bar{x}\|^2 \\ &= -\frac{1}{2} \bar{x}^\top A \bar{x} - \bar{\lambda} \|\bar{x}\|^2. \end{aligned}$$

Pour $\|x\| = \|\bar{x}\|$, on a alors d'après (ii), la relation ci-dessus et (5.14)

$$\begin{aligned} -\frac{1}{2} \bar{x}^\top A \bar{x} - \bar{\lambda} \|\bar{x}\|^2 - \frac{\bar{\lambda}}{2} \|x\|^2 &= -\frac{1}{2} \bar{x}^\top A \bar{x} - \bar{\lambda} \|\bar{x}\|^2 \\ &= f(\bar{x}) \\ &\leq f(x) \\ &= \frac{1}{2} x^\top A x - b^\top x \\ &= \frac{1}{2} x^\top A x - \bar{x}^\top (A + \bar{\lambda} I) x. \end{aligned}$$

La relation entre les membres extrêmes donnent le résultat.

- (iv) C'est une équation du second degré dont le discriminant vaut $4(\bar{x}^\top h)^2 > 0$. Il y a donc deux racines réelles distinctes.
- (v) Soient $h \in \mathbb{R}^n$ tel que $\bar{x}^\top h \neq 0$ et α la racine non nulle de l'équation de (iv). On prend $x = \bar{x} + \alpha h$. On a $\|x\| = \|\bar{x}\|$ et en appliquant (iii), on trouve le résultat ($\alpha \neq 0$).

5. La condition (5.17) est vérifiée grâce aux points 3, 4(i) et 4(v).

6. (i) D'après (5.14), \bar{x} vérifie la condition d'optimalité $\nabla_x \ell(\bar{x}, \bar{\lambda}) = 0$ du problème. D'autre part, l'application $x \mapsto \ell(x, \bar{\lambda})$ est convexe, car d'après (5.17) $\nabla_{xx}^2 \ell(x, \bar{\lambda}) = A + \bar{\lambda}I$ est semi-définie positive. Alors \bar{x} est solution du problème (la condition d'optimalité du premier ordre est une CNS lorsque le problème est convexe).

(ii) Pour les x tels que $\|x\| \leq \Delta$, on a

$$\begin{aligned} f(\bar{x}) &= \ell(\bar{x}, \bar{\lambda}) && \text{[par (5.16)]} \\ &\leq \ell(x, \bar{\lambda}) && \text{[par 6(i)]} \\ &\leq f(x) && \text{[car } \|x\| \leq \Delta \text{ et } \bar{\lambda} \geq 0]. \end{aligned}$$

Ceci démontre le résultat.

11 Conditions d'optimalité du second ordre

Il faut déterminer les multiplicateurs optimaux λ associés à la contrainte en chaque point stationnaire. L'unique contrainte c est qualifiée en tout point admissible x , car $\nabla c(x) = x \neq 0$. Le lagrangien s'écrit

$$\ell(x, \lambda) = \frac{1}{3}(x_1^3 + x_2^3) + \frac{\lambda}{2}(x_1^2 + x_2^2 - 1)$$

et les conditions de Lagrange sont

$$\begin{cases} x_1(x_1 + \lambda) = 0 \\ x_2(x_2 + \lambda) = 0 \\ x_1^2 + x_2^2 = 1. \end{cases}$$

Les multiplicateurs associés λ sont donc

$$\begin{aligned} x = \pm(0, 1) &\longrightarrow \lambda = \mp 1, \\ x = \pm(1, 0) &\longrightarrow \lambda = \mp 1, \\ x = \pm(1/\sqrt{2}, 1/\sqrt{2}) &\longrightarrow \lambda = \mp 1/\sqrt{2}. \end{aligned}$$

Le hessien du lagrangien s'écrit

$$L \equiv \nabla_{xx}^2 \ell(x, \lambda) = \begin{pmatrix} 2x_1 + \lambda & 0 \\ 0 & 2x_2 + \lambda \end{pmatrix}$$

et, par la qualification de la contrainte, l'espace tangent à la contrainte est

$$T_x X := \{d : x_1 d_1 + x_2 d_2 = 0\}.$$

Considérons chaque point stationnaire.

- Point $x = \pm(0, 1)$. L'espace tangent est $T_x X := \{d : d_2 = 0\}$ et $d^T L d = \lambda d_1^2 = \mp d_1^2$, si bien que
 - $x = (0, 1)$ est un maximum local strict (critère valant $1/3$),
 - $x = (0, -1)$ est un minimum local strict (critère valant $-1/3$),
- Point $x = \pm(1, 0)$. L'espace tangent est $T_x X := \{d : d_1 = 0\}$ et $d^T L d = \lambda d_2^2 = \mp d_2^2$, si bien que

- $x = (1, 0)$ est un maximum local strict (critère valant $1/3$),
- $x = (-1, 0)$ est un minimum local strict (critère valant $-1/3$),
- Point $x = \pm(1/\sqrt{2}, 1/\sqrt{2})$. L'espace tangent est $T_x X := \{d : d_1 + d_2 = 0\}$ et $d^T L d = -\lambda \|d\|_2^2 = \pm(1/\sqrt{2}) \|d\|_2^2$, si bien que
 - $x = (1/\sqrt{2}, 1/\sqrt{2})$ est un minimum local strict (critère valant $1/(3\sqrt{2})$),
 - $x = -(1/\sqrt{2}, 1/\sqrt{2})$ est un maximum local strict (critère valant $-1/(3\sqrt{2})$),

Le dernier cas peut paraître contre-intuitif si l'on ne regarde que les deux points du cas (pas si l'on compare avec les valeurs du critère aux points des autres cas), car la valeur du critère est plus grande au minimum local qu'au maximum local. En réalité les courbes de niveau du critère sont comme à la figure 6 et la théorie ne nous a pas trahi. Un minimum

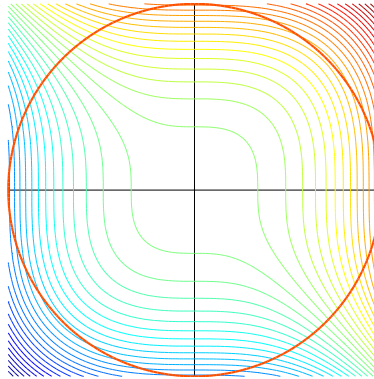


Fig. 6: Courbes de niveau du critère et contrainte pour le problème (5.18).

global est donc atteint en $x = (0, -1)$ et $x = (-1, 0)$.