

Chapitre 3

Optimisation

3.1 Définitions et rappels

3.1.1 Définition des problèmes d'optimisation

L'objectif de ce chapitre est de rechercher des minima ou des maxima d'une fonction $f \in C(\mathbb{R}^N, \mathbb{R})$ avec ou sans contrainte. Le problème d'optimisation sans contrainte s'écrit :

$$\begin{cases} \text{Trouver } \bar{x} \in \mathbb{R}^N \text{ tel que :} \\ f(\bar{x}) \leq f(y), \quad \forall y \in \mathbb{R}^N. \end{cases} \quad (3.1)$$

Le problème d'optimisation avec contrainte s'écrit :

$$\begin{cases} \text{Trouver } \bar{x} \in K \text{ tel que :} \\ f(\bar{x}) \leq f(y), \quad \forall y \in K. \end{cases} \quad (3.2)$$

où $K \subset \mathbb{R}^N$ et $K \neq \mathbb{R}^N$

Si \bar{x} est solution du problème (3.1), on dit que $\bar{x} \in \arg \min_{\mathbb{R}^N} f$, et si \bar{x} est solution du problème (3.2), on dit que $\bar{x} \in \arg \min_K f$.

3.1.2 Rappels et notations de calcul différentiel

Définition 3.1 Soient E et F des espaces vectoriels normés, f une application de E dans F et $x \in E$. On dit que f est différentiable en x s'il existe $T \in \mathcal{L}(E, F)$ (où $\mathcal{L}(E, F)$ est l'ensemble des applications linéaires de E dans F) telle que $f(x+h) = f(x) + T(h) + \|h\|\varepsilon(h)$ avec $\varepsilon(h) \rightarrow 0$ quand $h \rightarrow 0$. L'application T est alors unique et on note $Df(x) = T \in \mathcal{L}(E, F)$.

On peut remarquer qu'en dimension infinie, T dépend des normes associées à E et F . Voyons maintenant quelques cas particuliers d'espaces E et F :

Cas où $E = \mathbb{R}^N$ et $F = \mathbb{R}^p$ Soit $f : \mathbb{R}^N \rightarrow \mathbb{R}^p$, $x \in \mathbb{R}^N$ et supposons que f est différentiable en x ; alors $Df(x) \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^p)$, et il existe $A(x) \in \mathcal{M}_{p,N}(\mathbb{R})$ telle que $\underbrace{Df(x)(y)}_{\in \mathbb{R}^p} = \underbrace{Ay}_{\in \mathbb{R}^p}$, $\forall y \in \mathbb{R}^N$. On confond alors

l'application linéaire $Df(x) \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^p)$ et la matrice $A(x) \in \mathcal{M}_{p,N}(\mathbb{R})$ qui la représente. On écrit donc :

$$A(x) = Df(x) = (a_{i,j})_{1 \leq i \leq p, 1 \leq j \leq N} \text{ où } a_{i,j} = \partial_j f_i(x),$$

∂_j désignant la dérivée partielle par rapport à la j -ème variable.

Exemple 3.2 Prenons $N = 3$ et $p = 2$, notons $x = (x_1, x_2, x_3)^t$ et considérons $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ définie par :

$$f(x) = \begin{pmatrix} x_1^2 + x_2^3 + x_3^4 \\ 2x_1 - x_2 \end{pmatrix}$$

On vérifiera par le calcul (exercice) que pour $h = (h_1, h_2, h_3)^t$, on a :

$$Df(x)h = \begin{pmatrix} 2x_1h_1 + 3x_2^2h_2 + 4x_3^3h_3 \\ 2h_1 - h_2 \end{pmatrix}$$

et donc, avec les notations précédentes,

$$A(x) = \begin{pmatrix} 2x_1 & 3x_2^2 & 4x_3^3 \\ 2 & -1 & 0 \end{pmatrix}.$$

Cas où $E = \mathbb{R}^N$, $F = \mathbb{R}$ C'est un sous-cas du paragraphe précédent, puisqu'on est ici dans le cas $p = 1$. Soit $x \in \mathbb{R}^N$ et f une fonction de E dans F différentiable en x ; on a donc (avec l'abus de notation signalé dans le paragraphe précédent) $Df(x) \in \mathcal{M}_{1,N}(\mathbb{R})$, et on peut définir le gradient de f en x par $\nabla f(x) = (Df(x))^t \in \mathbb{R}^N$. Pour $(x, y) \in (\mathbb{R}^N)^2$, on a donc

$$Df(x)y = \sum_{j=1}^N \partial_j f(x)y_j = \nabla f(x) \cdot y \text{ où } \nabla f(x) = \begin{bmatrix} \partial_1 f(x) \\ \vdots \\ \partial_N f(x) \end{bmatrix} \in \mathbb{R}^N.$$

Cas où E est un espace de Hilbert et $F = \mathbb{R}$ On généralise ici le cas présenté au paragraphe précédent. Soit $f : E \rightarrow \mathbb{R}$ différentiable en $x \in E$. Alors $Df(x) \in \mathcal{L}(E, \mathbb{R}) = E'$, où E' désigne le dual topologique de E , c.à.d. l'ensemble des formes linéaires continues sur E . Par le théorème de représentation de Riesz, il existe un unique $u \in E$ tel que $Df(x)(y) = (u|y)_E$ pour tout $y \in E$, où $(\cdot|\cdot)_E$ désigne le produit scalaire sur E . On appelle encore gradient de f en x ce vecteur u . On a donc $u = \nabla f(x) \in E$ et pour $y \in E$, $Df(x)(y) = (\nabla f(x)|y)_E$.

Différentielle d'ordre 2, matrice hessienne Revenons maintenant au cas général de deux espaces vectoriels normés E et F , et supposons maintenant que $f \in C^2(E, F)$. Le fait que $f \in C^2(E, F)$ signifie que $Df \in C^1(E, \mathcal{L}(E, F))$. Par définition, on a $D^2f(x) \in \mathcal{L}(E, \mathcal{L}(E, F))$ et donc pour $y \in E$, $D^2f(x)(y) \in \mathcal{L}(E, F)$; en particulier, pour $z \in E$, $D^2f(x)(y)(z) \in F$.

Considérons maintenant le cas particulier $E = \mathbb{R}^N$ et $F = \mathbb{R}$. On a :

$$f \in C^2(\mathbb{R}^N, \mathbb{R}) \Leftrightarrow [f \in C^1(\mathbb{R}^N, \mathbb{R}) \text{ et } \nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)].$$

Soit $g = \nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$, et $x \in \mathbb{R}^N$, alors $Dg(x) \in \mathcal{M}_N(\mathbb{R})$ et on peut définir la matrice hessienne de f , qu'on note H_f , par : $H_f(x) = Dg(x) = D(Df)(x) = (b_{i,j})_{i,j=1\dots N} \in \mathcal{M}_N(\mathbb{R})$ où $b_{i,j} = \partial_{i,j}^2 f(x)$ où $\partial_{i,j}^2$ désigne la dérivée partielle par rapport à la variable i de la dérivée partielle par rapport à la variable j . Notons que par définition, $Dg(x)$ est la matrice jacobienne de g en x .

3.2 Optimisation sans contrainte

3.2.1 Définition et condition d'optimalité

Soit $f \in C(E, \mathbb{R})$ et E un espace vectoriel normé. On cherche soit un minimum global de f , c.à.d. :

$$\bar{x} \in E \text{ tel que } f(\bar{x}) \leq f(y) \quad \forall y \in E, \quad (3.3)$$

ou un minimum local, c.à.d. :

$$\bar{x} \text{ tel que } \exists \alpha > 0 \quad f(\bar{x}) \leq f(y) \quad \forall y \in B(\bar{x}, \alpha). \quad (3.4)$$

Proposition 3.3 (Condition nécessaire d'optimalité)

Soit E un espace vectoriel normé, et soient $f \in C(E, \mathbb{R})$, et $\bar{x} \in E$ tel que f est différentiable en \bar{x} . Si \bar{x} est solution de (3.4) alors $Df(\bar{x}) = 0$.

Démonstration Supposons qu'il existe $\alpha > 0$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in B(\bar{x}, \alpha)$. Soit $z \in E \setminus \{0\}$, alors si $|t| < \frac{\alpha}{\|z\|}$, on a $\bar{x} + tz \in B(\bar{x}, \alpha)$ (où $B(\bar{x}, \alpha)$ désigne la boule ouverte de centre \bar{x} et de rayon α) et on a donc $f(\bar{x}) \leq f(\bar{x} + tz)$. Comme f est différentiable en \bar{x} , on a :

$$f(\bar{x} + tz) = f(\bar{x}) + Df(\bar{x})(tz) + |t|\varepsilon_z(t),$$

où $\varepsilon_z(t) \rightarrow 0$ lorsque $t \rightarrow 0$. On a donc $f(\bar{x}) + tDf(\bar{x})(z) + |t|\varepsilon_z(t) \geq f(\bar{x})$. Et pour $\frac{\alpha}{\|z\|} > t > 0$, on a $Df(\bar{x})(z) + \varepsilon_z(t) \geq 0$. En faisant tendre t vers 0, on obtient que

$$Df(\bar{x})(z) \geq 0, \quad \forall z \in E.$$

On a aussi $Df(\bar{x})(-z) \geq 0 \quad \forall z \in E$, et donc : $-Df(\bar{x})(z) \geq 0 \quad \forall z \in E$.

On en conclut que

$$Df(\bar{x}) = 0.$$

Remarque 3.4 Attention, la proposition précédente donne une condition nécessaire mais non suffisante. En effet, $Df(\bar{x}) = 0$ n'entraîne pas que f atteigne un minimum (ou un maximum) même local, en \bar{x} . Prendre par exemple $E = \mathbb{R}$, $\bar{x} = 0$ et la fonction f définie par : $f(x) = x^3$ pour s'en convaincre.

3.2.2 Résultats d'existence et d'unicité

Théorème 3.5 (Existence) Soit $E = \mathbb{R}^N$ et $f : E \rightarrow \mathbb{R}$ une application telle que

- (i) f est continue,
- (ii) $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$.

Alors il existe $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) \leq f(y)$ pour tout $y \in \mathbb{R}^N$.

Démonstration La condition (ii) peut encore s'écrire

$$\forall A \in \mathbb{R}, \quad \exists R \in \mathbb{R}; \|x\| \geq R \Rightarrow f(x) \geq A. \quad (3.5)$$

On écrit (3.5) avec $A = f(0)$. On obtient alors :

$$\exists R \in \mathbb{R} \text{ tel que } \|x\| \geq R \Rightarrow f(x) \geq f(0).$$

On en déduit que $\inf_{\mathbb{R}^N} f = \inf_{B_R} f$, où $B_R = \{x \in \mathbb{R}^N; |x| \leq R\}$. Or, B_R est un compact de \mathbb{R}^N et f est continue donc il existe $\bar{x} \in B_R$ tel que $f(\bar{x}) = \inf_{B_R} f$ et donc $f(\bar{x}) = \inf_{\mathbb{R}^N} f$.

Remarque 3.6

1. Le théorème est faux si E est de dimension infinie (i.e. si E est espace de Banach au lieu de $E = \mathbb{R}^N$), car si E est de dimension infinie, B_R n'est pas compacte.
2. L'hypothèse (ii) du théorème peut être remplacée par

$$(ii)' \quad \exists b \in \mathbb{R}^N, \exists R > 0 \text{ tel que } \|x\| \geq R \Rightarrow f(x) \geq f(b).$$

3. Sous les hypothèses du théorème il n'y a pas toujours unicité de \bar{x} même dans le cas $N = 1$, prendre pour s'en convaincre la fonction f définie de \mathbb{R} dans \mathbb{R} par $f(x) = x^2(x-1)(x+1)$.

Définition 3.7 (Convexité) Soit E un espace vectoriel et $f : E \rightarrow \mathbb{R}$. On dit que f est convexe si

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \text{ pour tout } (x, y) \in E^2 \text{ t.q. } x \neq y \text{ et } t \in [0, 1].$$

On dit que f est strictement convexe si

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \text{ pour tout } (x, y) \in E^2 \text{ t.q. } x \neq y \text{ et } t \in]0, 1[.$$

Théorème 3.8 (Condition suffisante d'unicité) Soit E un espace vectoriel normé et $f : E \rightarrow \mathbb{R}$ strictement convexe alors il existe au plus un $\bar{x} \in E$ tel que $f(\bar{x}) \leq f(y), \forall y \in E$.

Démonstration

Soit f strictement convexe, supposons qu'il existe \bar{x} et $\bar{\bar{x}} \in E$ tels que $f(\bar{x}) = f(\bar{\bar{x}}) = \inf_{\mathbb{R}^N} f$. Comme f est strictement convexe, si $\bar{x} \neq \bar{\bar{x}}$ alors

$$f\left(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{\bar{x}}\right) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{\bar{x}}) = \inf_{\mathbb{R}^N} f,$$

ce qui est impossible ; donc $\bar{x} = \bar{\bar{x}}$.

Remarque 3.9 Ce théorème ne donne pas l'existence. Par exemple dans le cas $N = 1$ la fonction f définie par $f(x) = e^x$ n'atteint pas son minimum ; en effet, $\inf_{\mathbb{R}^N} f = 0$ et $f(x) \neq 0$ pour tout $x \in \mathbb{R}$, et pourtant f est strictement convexe.

Par contre, si on réunit les hypothèses des théorèmes 3.5 et 3.8, on obtient le résultat d'existence et unicité suivant :

Théorème 3.10 (Existence et unicité) Soit $E = \mathbb{R}^N$, et soit $f : E \rightarrow \mathbb{R}$. On suppose que :

- (i) f continue,
- (ii) $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$,
- (iii) f est strictement convexe ;

alors il existe un unique $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) = \inf_{\mathbb{R}^N} f$.

Remarque 3.11 Le théorème reste vrai (voir cours de maîtrise) si E est un espace de Hilbert ; on a besoin dans ce cas pour la partie existence des hypothèses (i), (ii) et de la convexité de f .

Proposition 3.12 (1ère caractérisation de la convexité) Soit E un espace vectoriel normé (sur \mathbb{R}) et $f \in C^1(E, \mathbb{R})$ alors :

1. f convexe si et seulement si $f(y) \geq f(x) + Df(x)(y - x)$, pour tout couple $(x, y) \in E^2$,
2. f est strictement convexe si et seulement si $f(y) > f(x) + Df(x)(y - x)$ pour tout couple $(x, y) \in E^2$ tel que $x \neq y$.

Démonstration

Démonstration de 1.

(\Rightarrow) Supposons que f est convexe : soit $(x, y) \in E^2$; on veut montrer que $f(y) \geq f(x) + Df(x)(y - x)$. Soit $t \in [0, 1]$, alors $f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x)$ grâce au fait que f est convexe. On a donc :

$$f(x + t(y - x)) - f(x) \leq t(f(y) - f(x)). \quad (3.6)$$

Comme f est différentiable, $f(x + t(y - x)) = f(x) + Df(x)(t(y - x)) + t\varepsilon(t)$ où $\varepsilon(t)$ tend vers 0 lorsque t tend vers 0. Donc en reportant dans (3.6),

$$\varepsilon(t) + Df(x)(y - x) \leq f(y) - f(x), \quad \forall t \in]0, 1[.$$

En faisant tendre t vers 0, on obtient alors :

$$f(y) \geq Df(x)(y - x) + f(x).$$

(\Leftarrow) Montrons maintenant la réciproque : Soit $(x, y) \in E^2$, et $t \in]0, 1[$ (pour $t = 0$ ou $= 1$ on n'a rien à démontrer). On veut montrer que $f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$. On pose $z = tx + (1 - t)y$. On a alors par hypothèse :

$$\begin{aligned} f(y) &\geq f(z) + Df(z)(y - z), \\ \text{et } f(x) &\geq f(z) + Df(z)(x - z). \end{aligned}$$

En multipliant la première inégalité par $1 - t$, la deuxième par t et en les additionnant, on obtient :

$$\begin{aligned} (1 - t)f(y) + tf(x) &\geq f(z) + (1 - t)Df(z)(y - z) + tDf(z)(x - z) \\ (1 - t)f(y) + tf(x) &\geq f(z) + Df(z)((1 - t)(y - z) + t(x - z)). \end{aligned}$$

Et comme $(1 - t)(y - z) + t(x - z) = 0$, on a donc $(1 - t)f(y) + tf(x) \geq f(z) = f(tx + (1 - t)y)$.

Démonstration de 2

(\Rightarrow) On suppose que f est strictement convexe, on veut montrer que $f(y) > f(x) + Df(x)(y - x)$ si $y \neq x$. Soit donc $(x, y) \in E^2$, $x \neq y$. On pose $z = \frac{1}{2}(y - x)$, et comme f est convexe, on peut appliquer la partie 1. du théorème et écrire que $f(x + z) \geq f(x) + Df(x)(z)$. On a donc $f(x) + Df(x)\left(\frac{y-x}{2}\right) \leq f\left(\frac{x+y}{2}\right)$. Comme f est strictement convexe, ceci entraîne que $f(x) + Df(x)\left(\frac{y-x}{2}\right) < \frac{1}{2}(f(x) + f(y))$, d'où le résultat.

(\Leftarrow) La méthode de démonstration est la même que pour le 1.

Proposition 3.13 (Caractérisation des points tels que $f(\bar{x}) = \inf_E f$)

Soit E espace vectoriel normé et f une fonction de E dans \mathbb{R} . On suppose que $f \in C^1(E, \mathbb{R})$ et que f est convexe. Soit $\bar{x} \in E$. Alors :

$$f(\bar{x}) = \inf_E f \Leftrightarrow Df(\bar{x}) = 0.$$

En particulier si $E = \mathbb{R}^N$ alors $f(\bar{x}) = \inf_{x \in \mathbb{R}^N} f(x) \Leftrightarrow \nabla f(\bar{x}) = 0$.

Démonstration

(\Rightarrow) Supposons que $f(\bar{x}) = \inf_E f$ alors on sait (voir Proposition 3.3) que $Df(\bar{x}) = 0$ (la convexité est inutile).

(\Leftarrow) Si f est convexe et différentiable, d'après la proposition 3.12, on a : $f(y) \geq f(\bar{x}) + Df(\bar{x})(y - x)$ pour tout $y \in E$ et comme par hypothèse $Df(\bar{x}) = 0$, on en déduit que $f(y) \geq f(\bar{x})$ pour tout $y \in E$. Donc $f(\bar{x}) = \inf_E f$.

Proposition 3.14 (2ème caractérisation de la convexité) Soit $E = \mathbb{R}^N$ et $f \in C^2(E, \mathbb{R})$. Soit $H_f(x)$ la hessienne de f au point x , i.e. $(H_f(x))_{i,j} = \partial_{i,j}^2 f(x)$. Alors

1. f est convexe si et seulement si $H_f(x)$ est symétrique et positive pour tout $x \in E$ (c.à.d. $H_f(x)^t = H_f(x)$ et $H_f(x)y \cdot y \geq 0$ pour tout $y \in \mathbb{R}^N$)
2. f est strictement convexe si $H_f(x)$ est symétrique définie positive pour tout $x \in E$. (Attention la réciproque est fausse.)

Démonstration

Démonstration de 1.

(\Rightarrow) Soit f convexe, on veut montrer que $H_f(x)$ est symétrique positive. Il est clair que $H_f(x)$ est symétrique car $\partial_{i,j}^2 f = \partial_{j,i}^2 f$ car f est C^2 . Par définition, $H_f(x) = D(\nabla f(x))$ et $\nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$. Soit $(x, y) \in E^2$, comme f est convexe et de classe C^1 , on a, grâce à la proposition 3.12 :

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x). \quad (3.7)$$

Soit $\varphi \in C^2(\mathbb{R}, \mathbb{R})$ définie par $\varphi(t) = f(x + t(y - x))$. Alors :

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt = [\varphi'(t)(t-1)]_0^1 - \int_0^1 \varphi''(t)(t-1) dt,$$

c'est-à-dire : $f(y) - f(x) = \varphi'(0) + \int_0^1 \varphi''(t)(1-t) dt$. Or $\varphi'(t) = \nabla f(x + t(y-x)) \cdot (y-x)$, et

$$\varphi''(t) = D(\nabla f(x + t(y-x)))(y-x) \cdot (y-x) = H_f(x + t(y-x))(y-x) \cdot (y-x).$$

On a donc :

$$f(y) - f(x) = \nabla f(x)(y-x) + \int_0^1 H_f(x + t(y-x))(y-x) \cdot (y-x)(1-t) dt. \quad (3.8)$$

Les inégalités (3.7) et (3.8) entraînent : $\int_0^1 H_f(x + t(y-x))(y-x) \cdot (y-x)(1-t) dt \geq 0 \forall x, y \in E$. On a donc :

$$\int_0^1 H_f(x + tz)z \cdot z(1-t) dt \geq 0 \quad \forall x, \forall z \in E. \quad (3.9)$$

En fixant $x \in E$, on écrit (3.9) avec $z = \varepsilon y$, $\varepsilon > 0$, $y \in \mathbb{R}^N$. On obtient :

$$\varepsilon^2 \int_0^1 H_f(x + t\varepsilon y)y \cdot y(1-t) dt \geq 0 \quad \forall x, y \in E, \quad \forall \varepsilon > 0, \text{ et donc :}$$

$$\int_0^1 H_f(x + t\varepsilon y)y \cdot y(1-t) dt \geq 0 \quad \forall \varepsilon > 0.$$

Pour $(x, y) \in E^2$ fixé, $H_f(x + t\varepsilon y)$ tend vers $H_f(x)$ uniformément lorsque $\varepsilon \rightarrow 0$, pour $t \in [0, 1]$. On a donc :

$$\int_0^1 H_f(x)y \cdot y(1-t) dt \geq 0, \text{ c.à.d. } \frac{1}{2}H_f(x)y \cdot y \geq 0.$$

Donc pour tout $(x, y) \in (\mathbb{R}^N)^2$, $H_f(x)y \cdot y \geq 0$ donc $H_f(x)$ est positive.

(\Leftarrow) Montrons maintenant la réciproque : On suppose que $H_f(x)$ est positive pour tout $x \in E$. On veut démontrer que f est convexe ; on va pour cela utiliser la proposition 3.12 et montrer que : $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ pour tout $(x, y) \in E^2$. Grâce à (3.8), on a :

$$f(y) - f(x) = \nabla f(x) \cdot (y - x) + \int_0^1 H_f(x + t(y - x))(y - x) \cdot (y - x)(1 - t) dt.$$

Or $H_f(x + t(y - x))(y - x) \cdot (y - x) \geq 0$ pour tout couple $(x, y) \in E^2$, et $1 - t \geq 0$ sur $[0, 1]$. On a donc $f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$ pour tout couple $(x, y) \in E^2$. La fonction f est donc bien convexe.

Démonstration de 2.

(\Leftarrow) On suppose que $H_f(x)$ est strictement positive pour tout $x \in E$, et on veut montrer que f est strictement convexe. On va encore utiliser la caractérisation de la proposition 3.12. Soit donc $(x, y) \in E^2$ tel que $y \neq x$. Alors :

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \int_0^1 \underbrace{H_f(x + t(y - x))(y - x) \cdot (y - x)}_{>0 \text{ si } x \neq y} \underbrace{(1 - t)}_{\neq 0 \text{ si } t \in]0, 1[} dt.$$

Donc $f(y) > f(x) + \nabla f(x)(y - x)$ si $x \neq y$, ce qui prouve que f est strictement convexe. \blacksquare

Contre-exemple Pour montrer que la réciproque de 2. est fautive, on propose le contre-exemple suivant : Soit $N = 1$ et $f \in C^2(\mathbb{R}, \mathbb{R})$, on a alors $H_f(x) = f''(x)$. Si f est la fonction définie par $f(x) = x^4$, alors f est strictement convexe car $f''(x) = 12x^2 \geq 0$, mais $f''(0) = 0$.

Cas d'une fonctionnelle quadratique Soient $A \in \mathcal{M}_N(\mathbb{R})$, $b \in \mathbb{R}^N$, et f la fonction de \mathbb{R}^N dans \mathbb{R}^N définie par $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$. Alors $f \in C^\infty(\mathbb{R}^N, \mathbb{R})$. Le calcul du gradient de f et de sa hessienne font l'objet de l'exercice 68 : on montre que

$$\nabla f(x) = \frac{1}{2}(Ax + A^t x) - b.$$

Donc si A est symétrique $\nabla f(x) = Ax - b$. Le calcul de la hessienne de f donne :

$$H_f(x) = D(\nabla f(x)) = \frac{1}{2}(A + A^t).$$

On en déduit que si A est symétrique, $H_f(x) = A$. On peut montrer en particulier (voir exercice 68) que si A est symétrique définie positive alors il existe un unique $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$, et que ce \bar{x} est aussi l'unique solution du système linéaire $Ax = b$.

3.3 Algorithmes d'optimisation sans contrainte

Soit $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$. On suppose qu'il existe $\bar{x} \in E$ tel que $f(\bar{x}) = \inf_E f$. On cherche à calculer \bar{x} (si f est de classe C^1 , on a nécessairement $\nabla f(\bar{x}) = 0$). On va donc maintenant développer des algorithmes (ou méthodes de calcul) du point \bar{x} qui réalise le minimum de f .

3.3.1 Méthodes de descente

Définition 3.15 Soient $f \in C(E, \mathbb{R})$ et $E = \mathbb{R}^N$.

1. Soit $x \in E$, on dit que $w \in E \setminus \{0\}$ est une direction de descente en x s'il existe $\rho_0 > 0$ tel que

$$f(x + \rho w) \leq f(x) \quad \forall \rho \in [0, \rho_0]$$

2. Soit $x \in E$, on dit que $w \in E \setminus \{0\}$ est une direction de descente stricte en x si s'il existe $\rho_0 > 0$ tel que

$$f(x + \rho w) < f(x) \quad \forall \rho \in]0, \rho_0[.$$

3. Une "méthode de descente" pour la recherche de \bar{x} tel que $f(\bar{x}) = \inf_E f$ consiste à construire une suite $(x_n)_n$ de la manière suivante :

- (a) Initialisation $x_0 \in E$;
 (b) Itération n : on suppose $x_0 \dots x_n$ connus ($n \geq 0$) ;
 i. On cherche w_n direction de descente stricte de x_n
 ii. On prend $x_{n+1} = x_n + \rho_n w_n$ avec $\rho_n > 0$ "bien choisi".

Proposition 3.16 Soient $E = \mathbb{R}^N$, $f \in C^1(E, \mathbb{R})$, $x \in E$ et $w \in E \setminus \{0\}$; alors

- si w direction de descente en x alors $w \cdot \nabla f(x) \leq 0$
- si $\nabla f(x) \neq 0$ alors $w = -\nabla f(x)$ est une direction de descente stricte en x .

Démonstration

- Soit $w \in E \setminus \{0\}$ une direction de descente en x alors par définition,

$$\exists \rho_0 > 0 \text{ tel que } f(x + \rho w) \leq f(x), \quad \forall \rho \in [0, \rho_0].$$

Soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par : $\varphi(\rho) = f(x + \rho w)$. On a $\varphi \in C^1(\mathbb{R}, \mathbb{R})$ et $\varphi'(\rho) = \nabla f(x + \rho w) \cdot w$. Comme w est une direction de descente, on peut écrire : $\varphi(\rho) \leq \varphi(0), \forall \rho \in [0, \rho_0]$, et donc

$$\forall \rho \in]0, \rho_0[, \quad \frac{\varphi(\rho) - \varphi(0)}{\rho} \leq 0;$$

en passant à la limite lorsque ρ tend vers 0, on déduit que $\varphi'(0) \leq 0$, c.à.d. $\nabla f(x) \cdot w \leq 0$.

- Soit $w = -\nabla f(x) \neq 0$. On veut montrer qu'il existe $\rho_0 > 0$ tel que si $\rho \in]0, \rho_0]$ alors $f(x + \rho w) < f(x)$ ou encore que $\varphi(\rho) < \varphi(0)$ où φ est la fonction définie en 1 ci-dessus. On a : $\varphi'(0) = \nabla f(x) \cdot w = -|\nabla f(x)|^2 < 0$. Comme φ' est continue, il existe $\rho_0 > 0$ tel que si $\rho \in [0, \rho_0]$ alors $\varphi'(\rho) < 0$. Si $\rho \in]0, \rho_0]$ alors $\varphi(\rho) - \varphi(0) = \int_0^\rho \varphi'(t) dt < 0$, et on a donc bien $\varphi(\rho) < \varphi(0)$ pour tout $\rho \in]0, \rho_0]$, ce qui prouve que w est une direction de descente stricte en x .

■

Algorithme du gradient à pas fixe Soient $f \in C^1(E, \mathbb{R})$ et $E = \mathbb{R}^N$. On se donne $\rho > 0$.

$$\left\{ \begin{array}{l} \text{Initialisation : } x_0 \in E, \\ \text{Itération } n : \quad x_n \text{ connu, } (n \geq 0) \\ \quad \quad \quad w_n = -\nabla f(x_n), \\ \quad \quad \quad x_{n+1} = x_n + \rho w_n. \end{array} \right. \quad (3.10)$$

Théorème 3.17 (Convergence du gradient à pas fixe) Soient $E = \mathbb{R}^N$ et $f \in C^1(E, \mathbb{R})$ On suppose que :

- $\exists \alpha > 0$ tel que $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha \|x - y\|^2, \forall (x, y) \in E^2$,
- $\exists M > 0$ tel que $\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|, \forall (x, y) \in E^2$,

alors :

- f est strictement convexe,
- $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$,
- il existe un et un seul $\bar{x} \in E$ tel que $f(\bar{x}) = \inf_E f$ (conséquence de 1. et 2.),
- si $0 < \rho < \frac{2\alpha}{M^2}$ alors la suite $(x_n)_{n \in \mathbb{N}}$ construite par (3.10) converge vers \bar{x} lorsque $n \rightarrow +\infty$.

La démonstration de ce théorème fait l'objet de l'exercice 70.

Algorithme du gradient à pas optimal L'idée de l'algorithme du gradient à pas optimal est d'essayer de calculer à chaque itération le paramètre qui minimise la fonction dans la direction de descente donnée par le gradient. Soient $f \in C^1(E, \mathbb{R})$ et $E = \mathbb{R}^N$, cet algorithme s'écrit :

$$\left\{ \begin{array}{l} \text{Initialisation : } x_0 \in \mathbb{R}^N. \\ \text{Itération } n : \\ \quad x_n \text{ connu.} \\ \quad \text{On calcule } w_n = -\nabla f(x_n). \\ \quad \text{On choisit } \rho_n \geq 0 \text{ tel que} \\ \quad \quad f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n) \quad \forall \rho \geq 0. \\ \quad \text{On pose } x_{n+1} = x_n + \rho_n w_n. \end{array} \right. \quad (3.11)$$

Les questions auxquelles on doit répondre pour s'assurer du bien fondé de ce nouvel algorithme sont les suivantes :

1. Existe-t-il ρ_n tel que $f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n), \forall \rho \geq 0$?
2. Comment calcule-t-on ρ_n ?
3. La suite $(x_n)_{n \in \mathbb{N}}$ construite par l'algorithme converge-t-elle ?

La réponse aux questions 1. et 3. est apportée par le théorème suivant :

Théorème 3.18 (Convergence du gradient à pas optimal)

Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$ telle que $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$. Alors :

1. La suite $(x_n)_{n \in \mathbb{N}}$ est bien définie par (3.11). On choisit $\rho_n > 0$ tel que $f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n) \quad \forall \rho \geq 0$ (ρ_n existe mais n'est pas nécessairement unique).
2. La suite $(x_n)_{n \in \mathbb{N}}$ est bornée et si $(x_{n_k})_{k \in \mathbb{N}}$ est une sous suite convergente, i.e. $x_{n_k} \rightarrow x$ lorsque $k \rightarrow +\infty$, on a nécessairement $\nabla f(x) = 0$. De plus si f est convexe on a $f(x) = \inf_{\mathbb{R}^N} f$.
3. Si f est strictement convexe on a alors $x_n \rightarrow \bar{x}$ quand $n \rightarrow +\infty$, avec $f(\bar{x}) = \inf_{\mathbb{R}^N} f$.

La démonstration de ce théorème fait l'objet de l'exercice 72. On en donne ici les idées principales.

1. On utilise l'hypothèse $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$ pour montrer que la suite $(x_n)_{n \in \mathbb{N}}$ construite par (3.11) existe : en effet, à x_n connu,
 - 1er cas : si $\nabla f(x_n) = 0$, alors $x_{n+1} = x_n$ et donc $x_p = x_n \quad \forall p \geq n$,
 - 2ème cas : si $\nabla f(x_n) \neq 0$, alors $w_n = -\nabla f(x_n)$ est une direction de descente stricte.

Dans ce deuxième cas, il existe donc ρ_0 tel que

$$f(x_n + \rho w_n) < f(x_n), \quad \forall \rho \in]0, \rho_0]. \quad (3.12)$$

De plus, comme $w_n \neq 0$, $|x_n + \rho w_n| \rightarrow +\infty$ quand $\rho \rightarrow +\infty$ et donc $f(x_n + \rho w_n) \rightarrow +\infty$ quand $\rho \rightarrow +\infty$. Il existe donc $M > 0$ tel que si $\rho > M$ alors $f(x_n + \rho w_n) \geq f(x_n)$. On a donc :

$$\inf_{\rho \in \mathbb{R}_+^*} f(x_n + \rho w_n) = \inf_{\rho \in [0, M]} f(x_n + \rho w_n).$$

Comme $[0, M]$ est compact, il existe $\rho_n \in [0, M]$ tel que $f(x_n + \rho_n w_n) = \inf_{\rho \in [0, M]} f(x_n + \rho w_n)$. De plus on a grâce à (3.12) que $\rho_n > 0$.

2. Le point 2. découle du fait que la suite $(f(x_n))_{n \in \mathbb{N}}$ est décroissante, donc la suite $(x_n)_{n \in \mathbb{N}}$ est bornée (car $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$). On montre ensuite que si $x_{n_k} \rightarrow x$ lorsque $k \rightarrow +\infty$ alors $\nabla f(\bar{x}) = 0$ (ceci est plus difficile, les étapes sont détaillées dans l'exercice 72).

Reste la question du calcul de ρ_n . Soit φ la fonction de \mathbb{R}_+ dans \mathbb{R} définie par : $\varphi(\rho) = f(x_n + \rho w_n)$. Comme $\rho_n > 0$ et $\varphi(\rho_n) \leq \varphi(\rho)$ pour tout $\rho \in \mathbb{R}_+$, on a nécessairement $\varphi'(\rho_n) = \nabla f(x_n + \rho_n w_n) \cdot w_n = 0$. Considérons le cas d'une fonctionnelle quadratique, i.e. $f(x) = \frac{1}{2} Ax \cdot x - b \cdot x$, A étant une matrice symétrique définie positive. Alors $\nabla f(x_n) = Ax_n - b$, et donc $\nabla f(x_n + \rho_n w_n) \cdot w_n = (Ax_n + \rho_n Aw_n - b) \cdot w_n = 0$. On a ainsi dans ce cas une expression explicite de ρ_n :

$$\rho_n = \frac{(b - Ax_n) \cdot w_n}{Aw_n \cdot w_n},$$

(en effet, $Aw_n \cdot w_n \neq 0$ car A est symétrique définie positive).

Dans le cas d'une fonction f générale, on n'a pas en général de formule explicite pour ρ_n . On peut par exemple le calculer en cherchant le zéro de f' par la méthode de la sécante ou la méthode de Newton...

L'algorithme du gradient à pas optimal est donc une méthode de minimisation dont on a prouvé la convergence. Cependant, cette convergence est lente (en général linéaire), et de plus, l'algorithme nécessite le calcul du paramètre ρ_n optimal.

Algorithme du gradient à pas variable Dans ce nouvel algorithme, on ne prend pas forcément le paramètre optimal pour ρ , mais on lui permet d'être variable d'une itération à l'autre. L'algorithme s'écrit :

$$\left\{ \begin{array}{l} \text{Initialisation : } x_0 \in \mathbb{R}^N. \\ \text{Itération : } \quad \text{On suppose } x_n \text{ connu ; soit } w_n = -\nabla f(x_n) \text{ où } : w_n \neq 0 \\ \quad \quad \quad \text{(si } w_n = 0 \text{ l'algorithme s'arrête).} \\ \quad \quad \quad \text{On prend } \rho_n > 0 \text{ tel que } f(x_n + \rho_n w_n) < f(x_n). \\ \quad \quad \quad \text{On pose } x_{n+1} = x_n + \rho_n w_n. \end{array} \right. \quad (3.13)$$

Théorème 3.19 (Convergence du gradient à pas variable)

Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$ une fonction telle que $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$, alors :

1. On peut définir une suite $(x_n)_{n \in \mathbb{N}}$ par (3.13).
2. La suite $(x_n)_{n \in \mathbb{N}}$ est bornée. Si $x_{n_k} \rightarrow x$ quand $k \rightarrow +\infty$ et si $\nabla f(x_{n_k}) \rightarrow 0$ quand $n \rightarrow +\infty$ alors $\nabla f(x) = 0$. Si de plus f est convexe on a $f(x) = \inf_{\mathbb{R}^N} f$.
3. Si $\nabla f(x_n) \rightarrow 0$ quand $n \rightarrow +\infty$ et si f est strictement convexe alors $x_n \rightarrow \bar{x}$ et $f(\bar{x}) = \inf_{\mathbb{R}^N} f$.

Démonstration : Elle est facile à partir de la démonstration du théorème précédent : reprendre en l'adaptant l'exercice 72.

3.3.2 Algorithmes du gradient conjugué

La méthode du gradient conjugué a été découverte en 1952 par Hestenes et Steifel pour la minimisation de fonctionnelles quadratiques, c'est-à-dire de fonctionnelles de la forme

$$f(x) = \frac{1}{2}Ax \cdot x - b \cdot x,$$

où $A \in \mathcal{M}_N(\mathbb{R})$ est une matrice symétrique définie positive et $b \in \mathbb{R}^N$. On rappelle (voir section (3.2.2) et exercice (68)) que $f(\bar{x}) = \inf_{\mathbb{R}^N} f \Leftrightarrow A\bar{x} = b$.

Définition 3.20 (Vecteurs conjugués) Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive,

1. Deux vecteurs v et w de $\mathbb{R}^N \setminus \{0\}$ sont dits A -conjugués si $Av \cdot w = w \cdot Av = 0$.
2. Une famille $(w^{(1)}, \dots, w^{(p)})$ de $\mathbb{R}^N \setminus \{0\}$ est dite A -conjuguée si $w^{(i)} \cdot Aw^{(j)} = 0$ pour tout couple $(i, j) \in \{1, \dots, p\}^2$ tel que $i \neq j$.

Proposition 3.21 Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive, $(w^{(1)}, \dots, w^{(p)})$ une famille de \mathbb{R}^N , alors :

1. si la famille $(w^{(1)}, \dots, w^{(p)})$ est A -conjuguée alors elle est libre ;
2. dans le cas où $p = N$, si la famille $(w^{(1)}, \dots, w^{(N)})$ est A -conjuguée alors c'est une base de \mathbb{R}^N .

Démonstration : Le point 2. est immédiat dès qu'on a démontré le point 1. Supposons donc que $(w^{(1)}, \dots, w^{(p)})$ est une famille A -conjuguée, i.e. $w^{(i)} \neq 0, \forall i$ et $w^{(i)} \cdot Aw^{(j)} = 0$ si $i \neq j$; soit $(\alpha_i)_{i=1, \dots, p} \subset \mathbb{R}$, supposons que $\sum_{i=1}^p \alpha_i w^{(i)} = 0$, on a donc $\sum_{i=1}^p \alpha_i w^{(i)} \cdot Aw^{(j)} = 0$ et donc $\alpha_j w^{(j)} \cdot Aw^{(j)} = 0$. Or $w^{(j)} \cdot Aw^{(j)} \neq 0$ car $w^{(j)} \neq 0$ et A est symétrique définie positive. On en déduit que $\alpha_j = 0$ pour $j = 1, \dots, p$. La famille $(w^{(1)}, \dots, w^{(p)})$ est donc libre.

Proposition 3.22 Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive, $b \in \mathbb{R}^N$ et f une fonction définie de \mathbb{R}^N dans \mathbb{R} par $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$. On suppose que la suite $(x^{(n)})_n$ est définie par :

- Initialisation $x^{(0)} \in \mathbb{R}^N$
 Itération n $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$ où
- 1) $w^{(n)} \neq 0$ est une direction de descente stricte en $x^{(n)}$
 - 2) ρ_n est optimal dans la direction $w^{(n)}$.

Si la famille $(w^{(0)}, \dots, w^{(N-1)})$ est une famille A -conjuguée alors $x^{(N)} = \bar{x}$ avec $A\bar{x} = b$.

Démonstration Soit $w^{(n)}$ direction de descente stricte en $x^{(n)}$ et ρ_n optimal dans la direction $w^{(n)}$; alors $\rho_n > 0$ et $\nabla f(x^{(n+1)}) \cdot w^{(n)} = 0$, c'est-à-dire

$$(Ax^{(n+1)} - b) \cdot w^{(n)} = 0 \quad (3.14)$$

On va montrer que

$$(Ax^{(N)} - b) \cdot w^{(p)} = 0, \forall p \in \{0, \dots, N-1\}.$$

Comme $(w^{(0)}, \dots, w^{(N-1)})$ est une base de \mathbb{R}^N , on en déduit alors que $Ax^{(N)} = b$, c'est-à-dire $x^{(N)} = \bar{x}$. Remarquons d'abord grâce à (3.14) que $(Ax^{(N)} - b) \cdot w^{(N-1)} = 0$. Soit maintenant $p < N-1$. On a :

$$Ax^{(N)} - b = A(x^{(N-1)} + \rho_{N-1}w^{(N-1)}) - b = Ax^{(N-1)} - b + \rho_{N-1}Aw^{(N-1)}.$$

On a donc en itérant,

$$Ax^{(N)} - b = Ax^{(p+1)} - b + \rho_{N-1}Aw^{(N-1)} + \dots + \rho_{p+1}Aw^{(p+1)}, \forall p \geq 1$$

. On en déduit que

$$(Ax^{(N)} - b) \cdot w^{(p)} = (Ax^{(p+1)} - b) \cdot w^{(p)} + \sum_{j=p+1}^{N-1} (\rho_j Aw_j \cdot w^{(p)}).$$

Comme les directions w_i sont conjuguées, on a donc $(Ax^{(N)} - b) \cdot w^{(p)} = 0$ pour tout $p = 0 \dots, N-1$ et donc $Ax^{(N)} = b$. ■

Le résultat précédent suggère de rechercher une méthode de minimisation de la fonction quadratique f selon le principe suivant : Pour $x^{(0)} \dots x^{(n)}$ connus, $w^{(0)}, \dots, w^{(n-1)}$ connus, on cherche $w^{(n)}$ tel que :

1. $w^{(n)}$ soit une direction de descente stricte en $x^{(n)}$,
2. $w^{(n)}$ soit A -conjugué avec $w^{(p)}$ pour tout $p < n$.

Si on arrive à trouver $w^{(n)}$ on prend alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$ avec ρ_n optimal dans la direction $w^{(n)}$. La propriété précédente donne $x^{(N)} = \bar{x}$ avec $A\bar{x} = b$.

Définition 3.23 (Méthode du gradient conjugué) Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive, $b \in \mathbb{R}^N$ et $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$.

Initialisation

Soit $x^{(0)} \in \mathbb{R}^N$, et soit $r^{(0)} = b - Ax^{(0)} = -\nabla f(x^{(0)})$.

- 1) Si $r^{(0)} = 0$, alors $Ax^{(0)} = b$ et donc $x^{(0)} = \bar{x}$, auquel cas l'algorithme s'arrête.
- 2) Si $r^{(0)} \neq 0$, alors on pose $w^{(0)} = r^{(0)}$, et on choisit ρ_0 optimal dans la direction $w^{(0)}$.
On pose alors $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$.

Itération $1 \leq n \leq N-1$:

On suppose $x^{(0)}, \dots, x^{(n)}$ et $w^{(0)}, \dots, w^{(n-1)}$ connus et on pose $r^{(n)} = b - Ax^{(n)}$.

- 1) Si $r^{(n)} = 0$ on a $Ax^{(n)} = b$ donc $x^{(n)} = \bar{x}$ auquel cas l'algorithme s'arrête.
- 2) Si $r^{(n)} \neq 0$, alors on pose $w^{(n)} = r^{(n)} + \lambda_{n-1}w^{(n-1)}$ avec λ_{n-1} tel que $w^{(n)} \cdot Aw^{(n-1)} = 0$, et on choisit ρ_n optimal dans la direction $w^{(n)}$;
On pose alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$.

(3.15)

Théorème 3.24 Soit A une symétrique définie positive, $A \in \mathcal{M}_N(\mathbb{R})$, $b \in \mathbb{R}^N$ et $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$ alors (3.15) définit une suite $(x^{(n)})_{n=0, \dots, p}$ avec $p \leq N$ telle que $x^{(N)} = \bar{x}$ avec $A\bar{x} = b$.

Démonstration

Initialisation Si $r^{(0)} = 0$, alors $Ax^{(0)} = b$ et donc $x^{(0)} = \bar{x}$ auquel cas $p = 0$. Si $r^{(0)} \neq 0$, comme $w^{(0)} = r^{(0)} = b - Ax^{(0)} = -\nabla f(x^{(0)})$, $w^{(0)}$ est une direction de descente stricte ; il existe donc ρ_0 qui minimise la fonction φ définie de \mathbb{R} dans \mathbb{R} par $\varphi(\rho) = f(x^{(0)} + \rho w^{(0)})$. La valeur de ρ_0 est obtenue en demandant que $\varphi'(\rho) = 0$, ce qui donne : $\rho_0 = \frac{r^{(0)} \cdot w^{(0)}}{Aw^{(0)} \cdot w^{(0)}}$. L'élément $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$ est donc bien défini. Notons que $r^{(1)} = Ax^{(1)} - b = r^{(0)} - \rho_0 Aw^{(0)}$, et donc $r^{(1)} \cdot w^{(0)} = 0$.

Itération n

On suppose $x^{(0)}, \dots, x^{(n)}$ et $w^{(0)}, \dots, w^{(n)}$ connus, et on pose $r^{(n)} = b - Ax^{(n)}$.

Si $r^{(n)} = 0$ alors $Ax^{(n)} = b$ et donc $x^{(n)} = \bar{x}$ auquel cas l'algorithme s'arrête et $p = n$.

Si $r^{(n)} \neq 0$, on pose $w^{(n)} = r^{(n)} + \lambda_{n-1} w^{(n-1)}$. Comme $w^{(n-1)} \neq 0$, on peut choisir λ_{n-1} tel que $w^{(n)} \cdot Aw^{(n-1)} = 0$, c.à.d. $(r^{(n)} + \lambda_{n-1} w^{(n-1)}) \cdot Aw^{(n-1)} = 0$, en prenant

$$\lambda_{n-1} = -\frac{r^{(n)} \cdot Aw^{(n-1)}}{w^{(n-1)} \cdot Aw^{(n-1)}}.$$

Montrons maintenant que $w^{(n)}$ est une direction de descente stricte en $x^{(n)}$. On a :

$$\begin{aligned} w^{(n)} \cdot (-\nabla f(x^{(n)})) &= (r^{(n)} + \lambda_{n-1} w^{(n-1)}) \cdot (-\nabla f(x^{(n)})) \\ &= (-\nabla f(x^{(n)}) + \lambda_{n-1} w_{n-1}) \cdot (-\nabla f(x_n)) \\ &= |\nabla f(x^{(n)})|^2 - \lambda_{n-1} w^{(n-1)} \cdot \nabla f(x^{(n)}). \end{aligned}$$

Or $w^{(n-1)} \cdot \nabla f(x^{(n)}) = 0$ car ρ_{n-1} est le paramètre de descente optimal en $x^{(n-1)}$ dans la direction $w^{(n-1)}$, on a donc :

$$-w^{(n)} \cdot \nabla f(x^{(n)}) = |\nabla f(x^{(n)})|^2 = |r^{(n)}|^2 > 0$$

ceci donne que $w^{(n)}$ est une direction de descente stricte en $x^{(n)}$. On peut choisir $\rho_n > 0$ optimal en $x^{(n)}$ dans la direction $w^{(n)}$, et le calcul de ρ_n (similaire à celui de l'étape d'initialisation) donne

$$\rho_n = \frac{r^{(n)} \cdot w^{(n)}}{Aw^{(n)} \cdot w^{(n)}}. \quad (3.16)$$

On peut donc bien définir $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$. Remarquons que ce choix de ρ_n entraîne que

$$r^{(n)} \cdot w^{(n-1)} = 0. \quad (3.17)$$

Pour pouvoir appliquer la proposition 3.22, il reste à montrer que la famille $w^{(0)}, \dots, w^{(n)}$ est A -conjuguée. Ceci est l'objet de la proposition 3.26 qui suit. Grâce à cette proposition, on obtient que si $r^{(n)} \neq 0$, $n = 0, \dots, N-1$, la famille $(w^{(0)}, \dots, w^{(N-1)})$ est donc A -conjuguée, et $w^{(n)}$ est une direction de descente stricte en $x^{(n)}$ pour tout $n \leq N-1$. On en déduit par la proposition 3.22 que $x^{(N)} = \bar{x}$. ■

La démonstration de la proposition 3.26 que l'on vient d'utiliser se fait par récurrence, et nécessite les petits résultats préliminaires énoncés dans le lemme suivant :

Lemme 3.25 *Sous les hypothèses et notations de la définition 3.23, on a :*

$$\rho_n = \frac{r^{(n)} \cdot r^{(n)}}{w^{(n)} \cdot Aw^{(n)}}, \quad (3.18)$$

$$r^{(n)} = r^{(n-1)} + \rho_{n-1} Aw^{(n-1)}, \quad (3.19)$$

$$r^{(n)} \cdot r^{(n-1)} = 0, \quad (3.20)$$

$$\lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}}, \quad (3.21)$$

Démonstration :

1. Comme ρ_n est le paramètre optimal dans la direction $w^{(n)}$, on sait (voir (3.16)) que

$$\rho_n = \frac{r^{(n)} \cdot w^{(n)}}{Aw^{(n)} \cdot w^{(n)}}.$$

Or par définition, $w^{(n)} = r^{(n)} + \lambda_{n-1}w^{(n-1)}$, et donc $w^{(n)} \cdot r^{(n)} = r^{(n)} \cdot r^{(n)} + \lambda_{n-1}w^{(n-1)} \cdot r^{(n)}$. Il ne reste plus à remarquer que $w^{(n-1)} \cdot r^{(n)} = 0$ en raison de l'optimalité de ρ_{n-1} (voir (3.17)). On en déduit que

$$\rho_n = \frac{r^{(n)} \cdot r^{(n)}}{w^{(n)} \cdot Aw^{(n)}}.$$

2. Par définition, $x^{(n)} = x^{(n-1)} + \rho_{n-1}w^{(n-1)}$, donc $Ax^{(n)} = Ax^{(n-1)} + \rho_{n-1}Aw^{(n-1)}$, ce qui entraîne $r^{(n)} = r^{(n-1)} + \rho_{n-1}Aw^{(n-1)}$.

3. Par définition, et grâce à (3.19), on a :

$$r^{(n)} \cdot r^{(n-1)} = r^{(n-1)} \cdot r^{(n-1)} + \rho_{n-1}Aw^{(n-1)} \cdot r^{(n-1)}.$$

Or $w^{(n-1)} = r^{(n-1)} + \lambda_{n-1}w^{(n-2)}$, et donc $r^{(n-1)} = w^{(n-1)} - \lambda_{n-1}w^{(n-2)}$. On en déduit que

$$r^{(n)} \cdot r^{(n-1)} = r^{(n-1)} \cdot r^{(n-1)} - \rho_{n-1}Aw^{(n-1)} \cdot w^{(n-1)} - \rho_{n-1}\lambda_{n-1}Aw^{(n-1)} \cdot w^{(n-2)}.$$

Or $Aw^{(n-1)} \cdot w^{(n-2)} = 0$ et par (3.18), on a $r^{(n-1)} \cdot r^{(n-1)} - \rho_{n-1}Aw^{(n-1)} \cdot w^{(n-1)} = 0$.

4. Par définition,

$$\lambda_{n-1} = -\frac{r^{(n)} \cdot Aw^{(n-1)}}{w^{(n-1)} \cdot Aw^{(n-1)}}.$$

Or par (3.19), on a :

$$Aw^{(n-1)} = \frac{1}{\rho_{n-1}}(r^{(n-1)} - r^{(n)}).$$

On conclut grâce à (3.20) et (3.18). ■

Proposition 3.26 *Sous les hypothèses et notations de la définition 3.23, soit $n \in \mathbb{N}$ tel que $1 \leq n \leq N$, si $r^{(q)} \neq 0$ pour $0 \leq q \leq n$, les propriétés suivantes sont vérifiées :*

1. $r^{(n)} \cdot w^{(q)} = 0, \forall q = 0, \dots, n-1$,
2. $\text{Vect}(r^{(0)}, \dots, r^{(n)}) = \text{Vect}(r^{(0)}, \dots, A^n r^{(0)})$,
3. $\text{Vect}(w^{(0)}, \dots, w^{(n)}) = \text{Vect}(r^{(0)}, \dots, A^n r^{(0)})$,
4. $w^{(n)} \cdot Aw^{(q)} = 0, \forall q = 0, \dots, n-1$,
5. $r^{(n)} \cdot r^{(q)} = 0, \forall q = 0, \dots, n-1$,

où $\text{Vect}(w^{(0)}, \dots, w^{(n)})$ désigne l'espace vectoriel engendré par les vecteurs $w^{(0)}, \dots, w^{(n)}$. En particulier, la famille $(w^{(0)}, \dots, w^{(N-1)})$ est A -conjuguée.

L'espace $\text{Vect}(r^{(0)}, \dots, A^n r^{(0)})$ est appelé espace de Krylov.

Démonstration :

On démontre les propriétés 1. à 5 par récurrence.

Etudions tout d'abord le cas $n = 1$. Remarquons que $r^{(1)} \cdot w^{(0)} = 0$ en vertu de (3.17) (on rappelle que cette propriété découle du choix optimal de ρ_0).

On a grâce à (3.19) :

$$r^{(1)} = r^{(0)} - \rho_0 Aw^{(0)} = r^{(0)} - \rho_0 Ar^{(0)},$$

car $w^{(0)} = r^{(0)}$. On a donc $\text{Vect}(r^{(0)}, r^{(1)}) = \text{Vect}(r^{(0)}, Ar^{(0)})$.

De plus, comme $w^{(0)} = r^{(0)}$, et $w^{(1)} = r^{(1)} + \lambda_1 w^{(0)}$, on a

$$\text{Vect}(r^{(0)}, r^{(1)}) = \text{Vect}(w^{(0)}, w^{(1)}).$$

On en déduit que 2. et 3. sont vraies pour $n = 1$.

Enfin, on a bien $w^{(1)} \cdot Aw^{(0)} = 0$ car $w^{(0)}$ et $w^{(1)}$ sont conjuguées, et $r^{(0)} \cdot r^{(1)} = 0$ en vertu de (3.20).

On a ainsi montré que les propriétés 1. à 5. sont vérifiées au rang $n = 1$. Supposons maintenant que ces propriétés soient vérifiées jusqu'au rang n , et démontrons qu'elles le sont encore au rang $n + 1$.

1. En vertu de (3.19), et par les hypothèses de récurrence 1. et 4., on a :

$$r^{(n+1)} \cdot w^{(q)} = r^{(n)} \cdot w^{(q)} - \rho_n Aw^{(n)} \cdot w^{(q)} = 0, \forall q \leq n - 1.$$

De plus, (3.20) entraîne $r^{(n+1)} \cdot w^{(n)} = 0$

2. Montrons que $Vect(r^{(0)}, r^{(1)} \dots, r^{(n+1)}) = Vect(r^{(0)}, Ar^{(0)}, \dots, A^{(n+1)}r^{(0)})$. Pour ce faire, commençons par remarquer que

$$r^{(n+1)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^{(n+1)}r^{(0)}).$$

En effet, en vertu de (3.19), on a : $r^{(n+1)} = r^{(n)} - \rho_n Aw^{(n)}$, et par hypothèse de récurrence, on a

$$r^{(n)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^n r^{(0)}), \text{ et } w^{(n)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^n r^{(0)}).$$

Montrons maintenant que $A^{n+1}r^{(0)} \in Vect(r^{(0)}, r^{(1)} \dots, r^{(n+1)})$. Comme $r^{(n+1)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^{(n+1)}r^{(0)})$, il existe une famille $(\alpha_k)_{k=0, \dots, n+1}$ telle que

$$r^{(n+1)} = \sum_{k=0}^{n+1} \alpha_k A^k r^{(0)} = \sum_{k=0}^n \alpha_k A^k r^{(0)} + \alpha_{n+1} A^{n+1} r^{(0)}.$$

Or grâce à la propriété 1. on sait que $r^{(n+1)} \cdot w^{(q)} = 0, \forall q \leq n$, et donc $r^{(n+1)} \notin Vect(w^{(0)}, w^{(1)} \dots, w^{(n)})$. On a donc $\alpha_{n+1} \neq 0$, et on peut donc écrire

$$A^{n+1}r^{(0)} = \frac{1}{\alpha_{n+1}} (r^{(n+1)} - \sum_{k=0}^n \alpha_k A^k r^{(0)}) \in Vect(r^{(0)}, r^{(1)} \dots, r^{(n+1)}),$$

par hypothèse de récurrence.

3. Montrons maintenant que

$$Vect(w^{(0)}, w^{(1)} \dots, w^{(n+1)}) = Vect(r^{(0)}, Ar^{(0)} \dots, A^{n+1}r^{(0)}).$$

On a : $w^{(n+1)} = r^{(n+1)} + \lambda_n w^{(n)}$. Or on vient de montrer que

$$r^{(n+1)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^{n+1}r^{(0)}),$$

et par hypothèse de récurrence, $w^{(n)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^n r^{(0)})$. On a donc bien $w^{(n+1)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^{n+1}r^{(0)})$. Montrons que réciproquement, $A^{n+1}r^{(0)} \in Vect(w^{(0)}, w^{(1)} \dots, w^{(n+1)})$. On a montré en 2. que

$$A^{n+1}r^{(0)} = \frac{1}{\alpha_{n+1}} (r^{(n+1)} - \sum_{k=0}^n \alpha_k A^k r^{(0)}).$$

Or $r^{(n+1)} = w^{(n+1)} - \lambda_n w^{(n)} \in Vect(w^{(0)}, w^{(1)} \dots, w^{(n+1)})$, et

$$\sum_{k=0}^n \alpha_k A^k r^{(0)} \in Vect(r^{(0)}, r^{(1)} \dots, r^{(n)}) = Vect(w^{(0)}, w^{(1)} \dots, w^{(n)}),$$

par hypothèse de récurrence. On en déduit que

$$A^{n+1}r^{(0)} \in Vect(w^{(0)}, w^{(1)} \dots, w^{(n)}).$$

4. On veut maintenant montrer que $w^{(n+1)} \cdot Aw^{(q)} = 0, \forall q \leq n$. Pour $q = n$, cette propriété est vérifiée en raison du choix de $w^{(n+1)}$ (conjuguée avec $w^{(n)}$). Pour $q < n$, on calcule :

$$w^{(n+1)} \cdot Aw^{(q)} = r^{(n+1)} \cdot Aw^{(q)} + \lambda_n w^{(n)} \cdot Aw^{(q)}. \quad (3.22)$$

Or $w^{(n)} \cdot Aw^{(q)} = 0$ pour tout $q \leq n - 1$ par hypothèse de récurrence. De plus, toujours par hypothèse de récurrence, $w^{(q)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^q r^{(0)})$, et donc

$$Aw^{(q)} \in Vect(r^{(0)}, Ar^{(0)} \dots, A^{q+1}r^{(0)}) = Vect(w^{(0)}, w^{(1)} \dots, w^{(q+1)}).$$

On a montré en 1. que $r^{(n+1)} \cdot w^{(k)} = 0$ pour tout $k \leq n$, on a donc $r^{(n+1)} \cdot Aw^{(q)} = 0$, et en reportant dans (3.22), on obtient donc que $w^{(n+1)} \cdot Aw^{(q)} = 0$ pour tout $q \leq n$.

5. Il reste à montrer que $r^{(n+1)} \cdot r^{(q)} = 0$ pour tout $q \leq n$. Pour $q = n$, on l'a démontré dans le lemme 3.25. Pour $q \leq n - 1$, on a

$$r^{(n+1)} \cdot r^{(q)} = (r^{(n)} - \lambda_n Aw^{(n)}) \cdot r^{(q)} = r^{(n)} \cdot r^{(q)} - \lambda_n Aw^{(n)} \cdot r^{(q)}.$$

Or $r^{(n)} \cdot r^{(q)} = 0$ par hypothèse de récurrence, et $Aw^{(n)} \cdot r^{(q)} = w^{(n)} \cdot Ar^{(q)}$; or $Ar^{(q)} \in Vect(r^{(0)}, \dots, r^{(q)})$ et $w^{(n)} \cdot r^{(k)} = 0$ pour tout $k \leq n - 1$ par hypothèse de récurrence 1. On en déduit que $r^{(n+1)} \cdot r^{(q)} = 0$.

Ceci termine la démonstration de la proposition (3.26). ■

Remarque 3.27 (Gradient conjugué préconditionné)

1. On a vu que $\lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}}$ et que $\rho_n = \frac{r^{(n)} \cdot r^{(n)}}{w^{(n)} \cdot Aw^{(n)}}$.

On peut calculer le nombre d'opérations nécessaires pour calculer \bar{x} (c.à.d. pour calculer $x^{(N)}$), sauf dans le cas miraculeux où $x^{(N)} = \bar{x}$ pour $n < N$) et montrer (exercice) que :

$$N_{gc} = 2N^3 + \mathcal{O}(N^2)$$

On rappelle que le nombre d'opérations pour Choleski est $\frac{N^3}{6}$ donc la méthode n'est pas intéressante comme méthode directe car elle demande 12 fois plus d'opérations que Choleski.

2. On peut alors se demander si la méthode est intéressante comme méthode itérative, c.à.d. si on peut espérer que $x^{(n)}$ soit "proche de \bar{x} " pour " $n \ll N$ ". Malheureusement, si la dimension N du système est grande, ceci n'est pas le cas en raison de l'accumulation des erreurs d'arrondi. Il est même possible de devoir effectuer plus de N itérations pour se rapprocher de \bar{x} . Cependant, dans les années 80, des chercheurs se sont rendus compte que ce défaut pouvait être corrigé à condition d'utiliser un "préconditionnement". Donnons par exemple le principe du préconditionnement dit de "Choleski incomplet".

On calcule une "approximation" de la matrice de Choleski de A c.à.d. qu'on cherche L triangulaire inférieure inversible telle que A soit "proche" de LL^t , en un sens à définir. Si on pose $y = L^t x$, alors le système $Ax = b$ peut aussi s'écrire $L^{-1}A(L^t)^{-1}y = L^{-1}b$, et le système $(L^t)^{-1}y = x$ est facile à résoudre car L^t est triangulaire supérieure. Soit $B \in \mathcal{M}_N(\mathbb{R})$ définie par $B = L^{-1}A(L^t)^{-1}$, alors

$$B^t = ((L^t)^{-1})^t A^t (L^{-1})^t = L^{-1}A(L^t)^{-1} = B$$

et donc B est symétrique. De plus,

$$Bx \cdot x = L^{-1}A(L^t)^{-1}x \cdot x = A(L^t)^{-1}x \cdot (L^t)^{-1}x,$$

et donc $Bx \cdot x > 0$ si $x \neq 0$. La matrice B est donc symétrique définie positive. On peut donc appliquer l'algorithme du gradient conjugué à la recherche du minimum de la fonction f définie par

$$f(y) = \frac{1}{2}By \cdot y - L^{-1}b \cdot y.$$

On en déduit l'expression de la suite $(y^{(n)})_{n \in \mathbb{N}}$ et donc $(x^{(n)})_{n \in \mathbb{N}}$.

On peut alors montrer (voir exercice 77) que l'algorithme du gradient conjugué préconditionné ainsi obtenu peut s'écrire directement pour la suite $(x^{(n)})_{n \in \mathbb{N}}$, de la manière suivante :

Itération n On pose $r^{(n)} = b - Ax^{(n)}$,

on calcule $s^{(n)}$ solution de $LL^t s^{(n)} = r^{(n)}$.

On pose alors $\lambda_{n-1} = \frac{s^{(n)} \cdot r^{(n)}}{s^{(n-1)} \cdot r^{(n-1)}}$ et $w^{(n)} = s^{(n)} + \lambda_{n-1}w^{(n-1)}$.

Le paramètre optimal ρ_n a pour expression : $\rho_n = \frac{s^{(n)} \cdot r^{(n)}}{Aw^{(n)} \cdot w^{(n)}}$, et on pose alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$.

Le choix de la matrice L peut se faire par exemple dans le cas d'une matrice creuse, en effectuant une factorisation "LL^t" incomplète, qui consiste à ne remplir que certaines diagonales de la matrice L pendant la factorisation, et laisser les autres à 0.

On peut généraliser le principe de l'algorithme du gradient conjugué à une fonction f non quadratique. Pour cela, on reprend le même algorithme que (3.15), mais on adapte le calcul de λ_{n-1} et ρ_n .

Itération n :

A $x^{(0)}, \dots, x^{(n)}$ et $w^{(0)}, \dots, w^{(n-1)}$ connus, on calcule $r^{(n)} = -\nabla f(x^{(n)})$.

Si $r^{(n)} = 0$ alors $Ax^{(n)} = b$ et donc $x^{(n)} = \bar{x}$ auquel cas l'algorithme s'arrête.

Si $r^{(n)} \neq 0$, on pose $w^{(n)} = r^{(n)} + \lambda_{n-1}w^{(n-1)}$ où λ_{n-1} peut être choisi de différentes manières :

1ère méthode (Fletcher–Reeves)

$$\lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}},$$

2ème méthode (Polak–Ribière)

$$\lambda_{n-1} = \frac{(r^{(n)} - r^{(n-1)}) \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}}.$$

On pose alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$, où ρ_n est choisi, si possible, optimal dans la direction $w^{(n)}$.

La démonstration de la convergence de l'algorithme de Polak–Ribière fait l'objet de l'exercice 78 page 143.

En résumé, la méthode du gradient conjugué est très efficace dans le cas d'une fonction quadratique à condition de l'utiliser avec préconditionnement. Dans le cas d'une fonction non quadratique, le préconditionnement n'existe pas et il vaut donc mieux la réserver au cas "N petit".

3.3.3 Méthodes de Newton et Quasi–Newton

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R})$ et $g = \nabla f \in C^1(\mathbb{R}^N, \mathbb{R}^N)$. On a dans ce cas :

$$f(x) = \inf_{\mathbb{R}^N} f \Rightarrow g(x) = 0.$$

Si de plus f est convexe alors on a $g(x) = 0 \Rightarrow f(x) = \inf_{\mathbb{R}^N} f$. Dans ce cas d'équivalence, on peut employer la méthode de Newton pour minimiser f en appliquant l'algorithme de Newton pour chercher un zéro de $g = \nabla f$. On a $D(\nabla f) = H_f$ où $H_f(x)$ est la matrice hessienne de f en x . La méthode de Newton s'écrit dans ce cas :

$$\begin{cases} \text{Initialisation} & x^{(0)} \in \mathbb{R}^N, \\ \text{Itération } n & H_f(x^{(n)})(x^{(n-1)} - x^{(n)}) = -\nabla f(x^{(n)}). \end{cases} \quad (3.23)$$

Remarque 3.28 La méthode de Newton pour minimiser une fonction f convexe est une méthode de descente. En effet, si $H_f(x_n)$ est inversible, on a $x^{(n+1)} - x^{(n)} = [H_f(x^{(n)})]^{-1}(-\nabla f(x^{(n)}))$ soit encore $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$ où $\rho_n = 1$ et $w^{(n)} = [H_f(x^{(n)})]^{-1}(-\nabla f(x^{(n)}))$. Si f est convexe, H_f est une matrice symétrique positive (déjà vu). Comme on suppose $H_f(x^{(n)})$ inversible par hypothèse, la matrice $H_f(x^{(n)})$ est donc symétrique définie positive.

Donc $w^{(n)}$ est alors une direction de descente stricte si $w^{(n)} \neq 0$ (donc $\nabla f(x^{(n)}) \neq 0$). On en déduit que

$$-w^{(n)} \cdot \nabla f(x^{(n)}) = [H_f(x^{(n)})]^{-1} \nabla f(x^{(n)}) \cdot \nabla f(x^{(n)}) > 0$$

ce qui est une condition suffisante pour que $w^{(n)}$ soit une direction de descente stricte.

La méthode de Newton est donc une méthode de descente avec $w^{(n)} = -H_f(x^{(n)})(\nabla f(x^{(n)}))$ et $\rho_n = 1$.

On peut aussi remarquer, en vertu du théorème 2.16 page 83, que si $f \in C^3(\mathbb{R}^N, \mathbb{R})$, si \bar{x} est tel que $\nabla f(\bar{x}) = 0$ et si $H_f(\bar{x}) = D(\nabla f)(\bar{x})$ est inversible alors il existe $\varepsilon > 0$ tel que si $x_0 \in B(\bar{x}, \varepsilon)$, alors la suite $(x^{(n)})_n$ est bien définie par (3.23) et $x^{(n)} \rightarrow \bar{x}$ lorsque $n \rightarrow +\infty$. De plus, d'après la proposition 2.14, il existe $\beta > 0$ tel que $|x^{(n+1)} - \bar{x}| \leq \beta |x^{(n)} - \bar{x}|^2$ pour tout $n \in \mathbb{N}$.

Remarque 3.29 (Sur l'implantation numérique) La convergence de la méthode de Newton est très rapide, mais nécessite en revanche le calcul de $H_f(x)$, qui peut s'avérer impossible ou trop coûteux.

On va maintenant donner des variantes de la méthode de Newton qui évitent le calcul de la matrice hessienne.

Proposition 3.30 Soient $f \in C^1(\mathbb{R}^N, \mathbb{R})$, $x \in \mathbb{R}^N$ tel que $\nabla f(x) \neq 0$, et soit $B \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive ; alors $w = -B\nabla f(x)$ est une direction de descente stricte en x .

Démonstration On a : $w \cdot \nabla f(x) = -B \nabla f(x) \cdot \nabla f(x) < 0$ car B est symétrique définie positive et $\nabla f(x) \neq 0$ donc w est une direction de descente stricte en x . En effet, soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par $\varphi(\rho) = f(x + \rho w)$. Il est clair que $\varphi \in C^1(\mathbb{R}, \mathbb{R})$, $\varphi'(\rho) = \nabla f(x + \rho w) \cdot w$ et $\varphi'(0) = \nabla f(x) \cdot w < 0$. Donc $\exists \rho_0 > 0$ tel que $\varphi'(\rho) < 0$ si $\rho \in]0, \rho_0[$. Par le théorème des accroissements finis, $\varphi(\rho) < \varphi(0) \forall \rho \in]0, \rho_0[$ donc w est une direction de descente stricte. ■

Méthode de Broyden La première idée pour construire une méthode de type quasi Newton est de prendre comme direction de descente en $x^{(n)}$ le vecteur $w^{(n)} = -(B^{(n)})^{-1}(\nabla f(x^{(n)}))$ où la matrice $B^{(n)}$ est censée approcher $H_f(x^{(n)})$ (sans calculer la dérivée seconde de f). On suppose $x^{(n)}, x^{(n-1)}$ et $B^{(n-1)}$ connus. Voyons comment on peut déterminer $B^{(n)}$. On peut demander par exemple que la condition suivante soit satisfaite :

$$\nabla f(x^{(n)}) - \nabla f(x^{(n-1)}) = B^{(n)}(x^{(n)} - x^{(n-1)}). \quad (3.24)$$

Ceci est un système à N équations et $N \times N$ inconnues, et ne permet donc pas de déterminer entièrement la matrice $B^{(n)}$ si $N > 1$. Voici un moyen possible pour déterminer entièrement $B^{(n)}$, dû à Broyden. On pose $s^{(n)} = x^{(n)} - x^{(n-1)}$, on suppose que $s^{(n)} \neq 0$, et on pose $y^{(n)} = \nabla f(x^{(n)}) - \nabla f(x^{(n-1)})$. On choisit alors $B^{(n)}$ telle que :

$$\begin{cases} B^{(n)}s^{(n)} = y^{(n)} \\ B^{(n)}s = B^{(n-1)}s, \forall s \perp s^{(n)} \end{cases} \quad (3.25)$$

On a exactement le nombre de conditions qu'il faut avec (3.25) pour déterminer entièrement $B^{(n)}$. Ceci suggère la méthode suivante :

Initialisation Soient $x^{(0)} \in \mathbb{R}^N$ et $B^{(0)}$ une matrice symétrique définie positive. On pose $w^{(0)} = (B^{(0)})^{-1}(-\nabla f(x^{(0)}))$; alors $w^{(0)}$ est une direction de descente stricte sauf si $\nabla f(x^{(0)}) = 0$.

On pose alors $x^{(1)} = x^{(0)} + \rho^{(0)}w^{(0)}$, où $\rho^{(0)}$ est optimal dans la direction $w^{(0)}$.

Itération n On suppose $x^{(n)}, x^{(n-1)}$ et $B^{(n-1)}$ connus, ($n \geq 1$), et on calcule $B^{(n-1)}$ par (3.25). On pose $w^{(n)} = -(B^{(n-1)})^{-1}(\nabla f(x^{(n)}))$. On choisit $\rho^{(n)}$ optimal en $x^{(n)}$ dans la direction $w^{(n)}$, et on pose $x^{(n+1)} = x^{(n)} + \rho^{(n)}w^{(n)}$.

Le problème avec cet algorithme est que si la matrice est $B^{(n-1)}$ symétrique définie positive, la matrice $B^{(n)}$ ne l'est pas forcément, et donc $w^{(n)}$ n'est pas forcément une direction de descente stricte. On va donc modifier cet algorithme dans ce qui suit.

Méthode de BFGS La méthode DFGS (de Broyden¹, Fletcher², Goldfarb³ et Shanno⁴) cherche à construire $B^{(n)}$ proche de $B^{(n-1)}$, telle que $B^{(n)}$ vérifie (3.24) et telle que si $B^{(n-1)}$ est symétrique définie positive alors $B^{(n)}$ est symétrique définie positive. On munit $\mathcal{M}_N(\mathbb{R})$ d'une norme induite par un produit scalaire, par exemple si $A \in \mathcal{M}_N(\mathbb{R})$ et $A = (a_{i,j})_{i,j=1,\dots,N}$ on prend $\|A\| = \left(\sum_{i,j=1}^N a_{i,j}^2\right)^{1/2}$. $\mathcal{M}_N(\mathbb{R})$ est alors un espace de Hilbert.

On suppose $x^{(n)}, x^{(n-1)}, B^{(n-1)}$ connus, et on définit

$$\mathcal{C}_n = \{B \in \mathcal{M}_N(\mathbb{R}) \mid B \text{ symétrique, vérifiant (3.24)}\},$$

qui est une partie de $\mathcal{M}_N(\mathbb{R})$ convexe fermée non vide. On choisit alors $B^{(n)} = P_{\mathcal{C}_n} B^{(n-1)}$ où $P_{\mathcal{C}_n}$ désigne la projection orthogonale sur \mathcal{C}_n . La matrice $B^{(n)}$ ainsi définie existe et est unique; elle est symétrique d'après le choix de \mathcal{C}_n . On peut aussi montrer que si $B^{(n-1)}$ symétrique définie positive alors $B^{(n)}$ l'est aussi.

Avec un choix convenable de la norme sur $\mathcal{M}_N(\mathbb{R})$, on obtient le choix suivant de $B^{(n)}$ si $s^{(n)} \neq 0$ et $\nabla f(x^{(n)}) \neq 0$ (sinon l'algorithme s'arrête) :

$$B^{(n)} = B^{(n-1)} + \frac{y^{(n)}(y^{(n)})^t}{(s^{(n)})^t \cdot y^{(n)}} - \frac{B^{(n-1)}s^{(n)}(s^{(n)})^t B^{(n-1)}}{(s^{(n)})^t B^{(n-1)}s^{(n)}}. \quad (3.26)$$

1. Broyden, C. G., The Convergence of a Class of Double-rank Minimization Algorithms, *Journal of the Institute of Mathematics and Its Applications* 1970, 6, 76-90

2. Fletcher, R., A New Approach to Variable Metric Algorithms, *Computer Journal* 1970, 13, 317-322

3. Goldfarb, D., A Family of Variable Metric Updates Derived by Variational Means, *Mathematics of Computation* 1970, 24, 23-26

4. Shanno, D. F., Conditioning of Quasi-Newton Methods for Function Minimization, *Mathematics of Computation* 1970, 24, 647-656

L'algorithme obtenu est l'algorithme de BFGS.

Algorithme de BFGS

$$\left\{ \begin{array}{l}
 \text{Initialisation} \quad \text{On choisit } x^{(0)} \in \mathbb{R}^N \text{ et} \\
 \quad B^{(0)} \text{ symétrique définie positive} \\
 \quad (\text{par exemple } B^{(0)} = Id) \text{ et on pose} \\
 \quad w^{(0)} = -B^{(0)} \nabla f(x^{(0)}) \\
 \quad \text{si } \nabla f(x^{(0)}) \neq 0, \text{ on choisit } \rho^{(0)} \text{ optimal} \\
 \quad \text{dans la direction } w^{(0)}, \text{ et donc} \\
 \quad w^{(0)} \text{ est une direction de descente stricte.} \\
 \quad \text{On pose } x^{(1)} = x^{(0)} + \rho^{(0)} w^{(0)}. \\
 \text{Itération } n \quad \text{A } x^{(n)}, x^{(n-1)} \text{ et } B_{n-1} \text{ connus } (n \geq 1) \\
 \quad \text{On suppose} \\
 \quad s^{(n)} = x^{(n)} - x^{(n-1)} \quad y^{(n)} = \nabla f(x) - \nabla f(x^{(n-1)}) \\
 \quad \text{si } s^{(n)} \neq 0 \text{ et } \nabla f(x^{(n)}) \neq 0, \\
 \quad \text{on choisit } B^{(n)} \text{ vérifiant (3.26)} \\
 \quad \text{On calcule } w^{(n)} = -(B^{(n)})^{-1} (\nabla f(x^{(n)})) \\
 \quad (\text{direction de descente stricte en } x^{(n)}). \\
 \quad \text{On calcule } \rho^{(n)} \text{ optimal dans la direction } w^{(n)} \\
 \quad \text{et on pose } x^{(n+1)} = x^{(n)} + \rho^{(n)} w^{(n)}.
 \end{array} \right. \quad (3.27)$$

On donne ici sans démonstration le théorème de convergence suivant :

Théorème 3.31 (Fletcher, 1976) Soit $f \in C^2(\mathbb{R}^N, \mathbb{R})$ telle que $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$. On suppose de plus que f est strictement convexe (donc il existe un unique $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) = \inf_{\mathbb{R}^N} f$) et on suppose que la matrice hessienne $H_f(\bar{x})$ est symétrique définie positive.

Alors si $x^{(0)} \in \mathbb{R}^N$ et si $B^{(0)}$ est symétrique définie positive, l'algorithme BFGS définit bien une suite $x^{(n)}$ et on a $x^{(n)} \rightarrow \bar{x}$ quand $n \rightarrow +\infty$

De plus, si $x^{(n)} \neq \bar{x}$ pour tout n , la convergence est super linéaire i.e.

$$\left| \frac{x^{(n+1)} - \bar{x}}{x^{(n)} - \bar{x}} \right| \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Pour éviter la résolution d'un système linéaire dans BFGS, on peut choisir de travailler sur $(B^{(n)})^{-1}$ au lieu de $B^{(n)}$.

$$\left\{ \begin{array}{l}
 \text{Initialisation} \quad \text{Soit } x^{(0)} \in \mathbb{R}^N \text{ et } K^{(0)} \text{ symétrique définie positive} \\
 \quad \text{telle que } \rho_0 \text{ soit optimal dans la direction } -K^{(0)} \nabla f(x^{(0)}) = w^{(0)} \\
 \quad x^{(1)} = x^{(0)} + \rho_0 w^{(0)} \\
 \text{Itération } n : \quad \text{A } x^{(n)}, x^{(n-1)}, K^{(n-1)} \text{ connus, } n \geq 1, \\
 \quad \text{on pose } s^{(n)} = x^{(n)} - x^{(n-1)}, y^{(n)} = \nabla f(x^{(n)}) - \nabla f(x^{(n-1)}) \\
 \quad \text{et } K^{(n)} = P_{C_n} K^{(n-1)}. \\
 \quad \text{On calcule } w^{(n)} = -K^{(n)} \nabla f(x^{(n)}) \text{ et on choisit } \rho_n \\
 \quad \text{optimal dans la direction } w^{(n)}. \\
 \quad \text{On pose alors } x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}.
 \end{array} \right. \quad (3.28)$$

Remarquons que le calcul de la projection de $P_{C_n} K^{(n-1)}$ peut s'effectuer avec la formule (3.26) où on a remplacé $B^{(n-1)}$ par $K^{(n-1)}$. Malheureusement, on obtient expérimentalement une convergence nettement moins bonne pour l'algorithme de quasi-Newton modifié (3.28) que pour l'algorithme de BFGS (3.26).

3.3.4 Résumé sur les méthodes d'optimisation

Faisons le point sur les avantages et inconvénients des méthodes qu'on a vues sur l'optimisation sans contrainte.

Méthodes de gradient : Ces méthodes nécessitent le calcul de $\nabla f(x^{(n)})$. Leur convergence est linéaire (donc lente).

Méthode de gradient conjugué : Si f est quadratique (c.à.d. $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$ avec A symétrique définie positive), la méthode est excellente si elle est utilisée avec un préconditionnement (pour N grand). Dans le cas général, elle n'est efficace que si N n'est pas trop grand.

Méthode de Newton : La convergence de la méthode de Newton est excellente (convergence localement quadratique) mais nécessite le calcul de $H_f(x^{(n)})$ (et de $\nabla f(x^{(n)})$). Si on peut calculer $H_f(x^{(n)})$, cette méthode est parfaite.

Méthode de quasi Newton : L'avantage de la méthode de quasi Newton est qu'on ne calcule que $\nabla f(x^{(n)})$ et pas $H_f(x^{(n)})$. La convergence est super linéaire. Par rapport à une méthode de gradient où on calcule $w^{(n)} = -\nabla f(x^{(n)})$, la méthode BFGS nécessite une résolution de système linéaire : $w^{(n)} = (B^{(n)})^{-1}(-\nabla f(x^{(n)}))$.

Quasi-Newton modifié :

Pour éviter la résolution de système linéaire dans BFGS, on peut choisir de travailler sur $(B^{(n)})^{-1}$ au lieu de $B^{(n)}$, pour obtenir l'algorithme de quasi Newton (3.28). Cependant, on perd alors en vitesse de convergence.

Comment faire si on ne veut (ou peut) pas calculer $\nabla f(x^{(n)})$? On peut utiliser des "méthodes sans gradient", c.à.d. qu'on choisit *a priori* les directions $w^{(n)}$. Ceci peut se faire soit par un choix déterministe, soit par un choix stochastique.

Un choix déterministe possible est de calculer $x^{(n)}$ en résolvant N problèmes de minimisation en une dimension d'espace. Pour chaque direction $i = 1, \dots, N$, on prend $w^{(n,i)} = e_i$, où e_i est le i -ème vecteur de la base canonique, et pour $i = 1, \dots, N$, on cherche $\theta \in \mathbb{R}$ tel que :

$$f(x_1^{(n)}, x_2^{(n)}, \dots, \theta, \dots, x_N^{(n)}) \leq f(x_1^{(n)}, x_2^{(n)}, \dots, t, \dots, x_N^{(n)}), \forall t \in \mathbb{R}.$$

Remarquons que si f est quadratique, on retrouve la méthode de Gauss Seidel.

3.4 Optimisation sous contraintes

3.4.1 Définitions

Soit $E = \mathbb{R}^N$, soit $f \in C(E, \mathbb{R})$, et soit K un sous ensemble de E . On s'intéresse à la recherche de $\bar{u} \in K$ tel que :

$$\begin{cases} \bar{u} \in K \\ f(\bar{u}) = \inf_K f \end{cases} \quad (3.29)$$

Ce problème est un problème de minimisation avec contrainte (ou "sous contrainte") au sens où l'on cherche u qui minimise f en astreignant u à être dans K . Voyons quelques exemples de ces contraintes (définies par l'ensemble K), qu'on va expliciter à l'aide des p fonctions continues, $g_i \in C(E, \mathbb{R})$ $i = 1 \dots p$.

- Contraintes égalités.** On pose $K = \{x \in E, g_i(x) = 0 \ i = 1 \dots p\}$. On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème des multiplicateurs de Lagrange (voir théorème 3.38).
- Contraintes inégalités.** On pose $K = \{x \in E, g_i(x) \leq 0 \ i = 1 \dots p\}$. On verra plus loin que le problème de minimisation de f peut alors être résolu grâce au théorème de Kuhn-Tucker (voir théorème 3.42).
 - *Programmation linéaire.* Avec un tel ensemble de contraintes K , si de plus f est linéaire, c'est-à-dire qu'il existe $b \in \mathbb{R}^N$ tel que $f(x) = b \cdot x$, et les fonctions g_i sont affines, c'est-à-dire qu'il existe $b_i \in \mathbb{R}^N$ et $c_i \in \mathbb{R}$ tels que $g_i(x) = b_i \cdot x + c_i$, alors on dit qu'on a affaire à un problème de "programmation linéaire". Ces problèmes sont souvent résolus numériquement à l'aide de l'algorithme de Dantzig, inventé vers 1950.
 - *Programmation quadratique.* Avec le même ensemble de contraintes K , si de plus f est quadratique, c'est-à-dire si f est de la forme $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$, et les fonctions g_i sont affines, alors on dit qu'on a affaire à un problème de "programmation quadratique".
- Programmation convexe.** Dans le cas où f est convexe et K est convexe, on dit qu'on a affaire à un problème de "programmation convexe".

3.4.2 Existence – Unicité – Conditions d’optimalité simple

Théorème 3.32 (Existence) Soit $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$.

1. Si K est un sous-ensemble fermé borné de E , alors il existe $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$.
2. Si K est un sous-ensemble fermé de E , et si f est croissante à l’infini, c’est-à-dire que $f(x) \rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$, alors $\exists \bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$

Démonstration

1. Si K est un sous-ensemble fermé borné de E , comme f est continue, elle atteint ses bornes sur K , d’où l’existence de \bar{x} .
2. Si f est croissante à l’infini, alors il existe $R > 0$ tel que si $\|x\| > R$ alors $f(x) > f(0)$; donc $\inf_K f = \inf_{K \cap B_R} f$, où B_R désigne la boule de centre 0 et de rayon R . L’ensemble $K \cap B_R$ est compact, car intersection d’un fermé et d’un compact. Donc, par ce qui précède, il existe $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_{K \cap B_R} f = \inf_{B_R} f$. ■

Théorème 3.33 (Unicité) Soit $E = \mathbb{R}^N$ et $f \in C(E, \mathbb{R})$. On suppose que f est strictement convexe et que K est convexe. Alors il existe au plus un élément \bar{x} de K tel que $f(\bar{x}) = \inf_K f$.

Démonstration

Supposons que \bar{x} et $\bar{\bar{x}}$ soient deux solutions du problème (3.29), avec $\bar{x} \neq \bar{\bar{x}}$

Alors $f(\frac{1}{2}\bar{x} + \frac{1}{2}\bar{\bar{x}}) < \frac{1}{2}f(\bar{x}) + \frac{1}{2}f(\bar{\bar{x}}) = \inf_K f$. On aboutit donc à une contradiction. ■

Des théorèmes d’existence 3.32 et d’unicité 3.33 on déduit immédiatement le théorème d’existence et d’unicité suivant :

Théorème 3.34 (Existence et unicité) Soient $E = \mathbb{R}^N$, $f \in C(E, \mathbb{R}^N)$ une fonction strictement convexe et K un sous ensemble convexe fermé de E . Si K est borné ou si f est croissante à l’infini, c’est-à-dire si $f(x) \Rightarrow +\infty$ quand $\|x\| \rightarrow +\infty$, alors il existe un unique élément \bar{x} de K solution du problème de minimisation (3.29), i.e. tel que $f(\bar{x}) = \inf_K f$

Remarque 3.35 On peut remplacer $E = \mathbb{R}^N$ par E espace de Hilbert de dimension infinie dans le dernier théorème, mais on a besoin dans ce cas de l’hypothèse de convexité de f pour assurer l’existence de la solution (voir cours de maîtrise).

Proposition 3.36 (Condition simple d’optimalité) Soient $E = \mathbb{R}^N$, $f \in C(E, \mathbb{R})$ et $\bar{x} \in K$ tel que $f(\bar{x}) = \inf_K f$. On suppose que f est différentiable en \bar{x}

1. Si $\bar{x} \in \overset{\circ}{K}$ alors $\nabla f(\bar{x}) = 0$.
2. Si K est convexe, alors $\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq 0$ pour tout $x \in K$.

Démonstration

1. Si $\bar{x} \in \overset{\circ}{K}$, alors il existe $\varepsilon > 0$ tel que $B(\bar{x}, \varepsilon) \subset K$ et $f(\bar{x}) \leq f(x) \forall x \in B(\bar{x}, \varepsilon)$. Alors on a déjà vu (voir preuve de la Proposition 3.3 page 112) que ceci implique $\nabla f(\bar{x}) = 0$.
2. Soit $x \in K$. Comme \bar{x} réalise le minimum de f sur K , on a : $f(\bar{x} + t(x - \bar{x})) = f(tx + (1-t)\bar{x}) \geq f(\bar{x})$ pour tout $t \in]0, 1]$, par convexité de K . On en déduit que

$$\frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t} \geq 0 \text{ pour tout } t \in]0, 1].$$

En passant à la limite lorsque t tend vers 0 dans cette dernière inégalité, on obtient : $\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq 0$. ■

3.4.3 Conditions d'optimalité dans le cas de contraintes égalité

Dans tout ce paragraphe, on considèrera les hypothèses et notations suivantes :

$$\begin{aligned} f &\in C^1(\mathbb{R}^N, \mathbb{R}), \quad g_i \in C^1(\mathbb{R}^N, \mathbb{R}), \quad i = 1 \dots p; \\ K &= \{u \in \mathbb{R}^N, \quad g_i(u) = 0 \quad \forall i = 1 \dots p\}; \\ g &= (g_1, \dots, g_p)^t \in C^1(\mathbb{R}^N, \mathbb{R}^p) \end{aligned} \quad (3.30)$$

Remarque 3.37 (Quelques rappels de calcul différentiel)

Comme $g \in C^1(\mathbb{R}^N, \mathbb{R}^p)$, si $u \in \mathbb{R}^N$, alors $Dg(u) \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^p)$, ce qui revient à dire, en confondant l'application linéaire $Dg(u)$ avec sa matrice, que $Dg(u) \in \mathcal{M}_{p,N}(\mathbb{R})$. Par définition, $\text{Im}(Dg(u)) = \{Dg(u)z, z \in \mathbb{R}^N\} \subset \mathbb{R}^p$, et $\text{rang}(Dg(u)) = \dim(\text{Im}(Dg(u))) \leq p$. On rappelle de plus que

$$Dg(u) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_N} \\ \dots & \ddots & \dots \\ \frac{\partial g_p}{\partial x_1} & \dots & \frac{\partial g_p}{\partial x_N} \end{pmatrix},$$

et que $\text{rang}(Dg(u)) \leq \min(N, p)$. De plus, si $\text{rang}(Dg(u)) = p$, alors les vecteurs $(Dg_i(u))_{i=1 \dots p}$ sont linéairement indépendants dans \mathbb{R}^N .

Théorème 3.38 (Multipliateurs de Lagrange) Soit $\bar{u} \in K$ tel que $f(\bar{u}) = \inf_K f$. On suppose que f est différentiable en \bar{u} et $\dim(\text{Im}(Dg(\bar{u}))) = p$ (ou $\text{rang}(Dg(\bar{u})) = p$), alors :

$$\text{il existe } (\lambda_1, \dots, \lambda_p)^t \in \mathbb{R}^p \text{ tels que } \nabla f(\bar{u}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{u}) = 0.$$

(Cette dernière égalité a lieu dans \mathbb{R}^N)

Démonstration Pour plus de clarté, donnons d'abord une idée "géométrique" de la démonstration dans le cas $N = 2$ et $p = 1$. On a dans ce cas $f \in C^1(\mathbb{R}^2, \mathbb{R})$ et $K = \{(x, y) \in \mathbb{R}^2, g(x, y) = 0\}$, et on cherche $u \in K$ tel que $f(u) = \inf_K f$. Traçons dans le repère (x, y) la courbe $g(x, y) = 0$, ainsi que les courbes de niveau de f .

Si on se "promène" sur la courbe $g(x, y) = 0$, en partant du point P_0 vers la droite (voir figure 3.1), on rencontre les courbes de niveau successives de f et on se rend compte sur le dessin que la valeur minimale que prend f sur la courbe $g(x, y) = 0$ est atteinte lorsque cette courbe est tangente à la courbe de niveau de f : sur le dessin, ceci correspond au point P_1 où la courbe $g(x, y) = 0$ est tangente à la courbe $f(x, y) = 3$. Une fois qu'on a passé ce point de tangence, on peut remarquer que f augmente.

On utilise alors le fait que si φ est une fonction continûment différentiable de \mathbb{R}^2 dans \mathbb{R} , le gradient de φ est orthogonal à toute courbe de niveau de φ , c'est-à-dire toute courbe de la forme $\varphi(x, y) = c$, où $c \in \mathbb{R}$. (En effet, soit $(x(t), y(t))$, $t \in \mathbb{R}$ un paramétrage de la courbe $g(x, y) = c$, en dérivant par rapport à t , on obtient : $\nabla g(x(t), y(t)) \cdot (x'(t), y'(t))^t = 0$). En appliquant ceci à f et g , on en déduit qu'au point de tangence entre une courbe de niveau de f et la courbe $g(x, y) = 0$, les gradients de f et g sont colinéaires. Et donc si $\nabla g(u) \neq 0$, il existe $\lambda \neq 0$ tel que $\nabla f(u) = \lambda \nabla g(u)$.

Passons maintenant à la démonstration rigoureuse du théorème dans laquelle on utilise le théorème des fonctions implicites⁵.

Par hypothèse, $Dg(\bar{u}) \in \mathcal{L}(\mathbb{R}^N, \mathbb{R}^p)$ et $\text{Im}(Dg(\bar{u})) = \mathbb{R}^p$. Donc il existe un sous espace vectoriel F de \mathbb{R}^N de dimension p , tel que $Dg(\bar{u})$ soit bijective de F dans \mathbb{R}^p . En effet, soit $(e_1 \dots e_p)$ la base canonique de \mathbb{R}^p , alors pour tout $i \in \{1, \dots, p\}$, il existe $y_i \in \mathbb{R}^N$ tel que $Dg(\bar{x})y_i = e_i$. Soit F le sous espace engendré par la famille $\{y_1 \dots y_p\}$; on remarque que cette famille est libre, car si $\sum_{i=1}^p \lambda_i y_i = 0$, alors $\sum_{i=1}^p \lambda_i e_i = 0$, et donc $\lambda_i = 0$ pour tout $i = 1, \dots, p$. On a ainsi montré l'existence d'un sous espace F de dimension p telle que $Dg(\bar{x})$ soit bijective (car surjective) de F dans \mathbb{R}^p .

5. **Théorème des fonctions implicites** Soient p et q des entiers naturels, soit $h \in C^1(\mathbb{R}^q \times \mathbb{R}^p, \mathbb{R}^p)$, et soient $(\bar{x}, \bar{y}) \in \mathbb{R}^q \times \mathbb{R}^p$ et $c \in \mathbb{R}^p$ tels que $h(\bar{x}, \bar{y}) = c$. On suppose que la matrice de la différentielle $D_2 h(\bar{x}, \bar{y}) \in \mathcal{M}_p(\mathbb{R})$ est inversible. Alors il existe $\varepsilon > 0$ et $\nu > 0$ tels que pour tout $x \in B(\bar{x}, \varepsilon)$, il existe un unique $y \in B(\bar{y}, \nu)$ tel que $h(x, y) = c$. on peut ainsi définir une application ϕ de $B(\bar{x}, \varepsilon)$ dans $B(\bar{y}, \nu)$ par $\phi(x) = y$. On a $\phi(\bar{x}) = \bar{y}$, $\phi \in C^1(\mathbb{R}^q, \mathbb{R}^p)$ et $D\phi(x) = -[D_2 h(x, \phi(x))]^{-1} \cdot D_1 h(x, \phi(x))$.

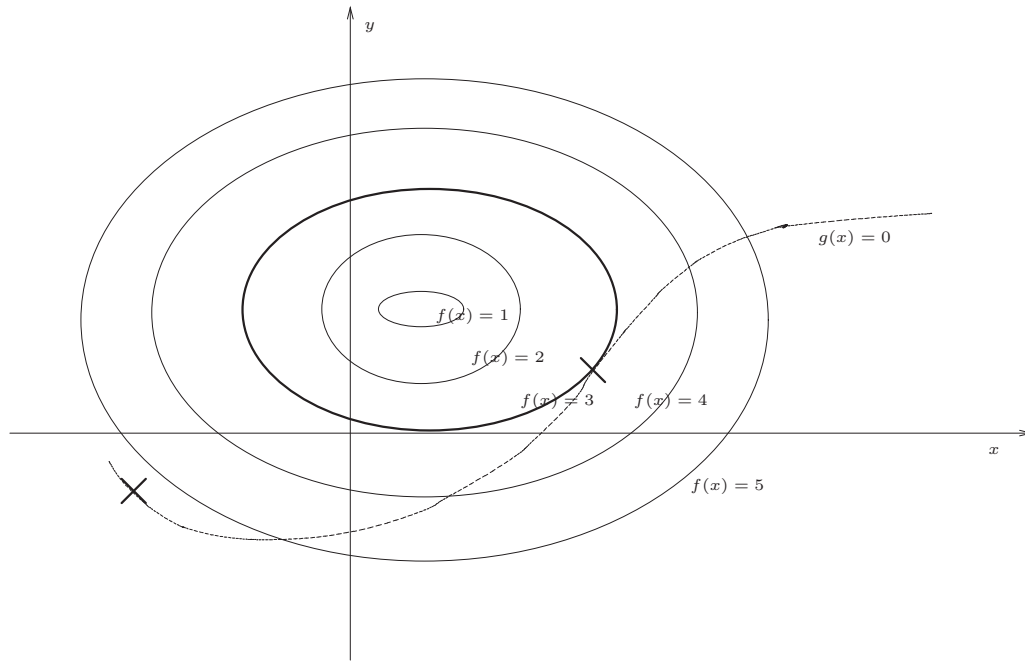


FIGURE 3.1 – Interprétation géométrique des multiplicateurs de Lagrange

Il existe un sous espace vectoriel G de \mathbb{R}^N , tel que $\mathbb{R}^N = F \oplus G$. Pour $v \in F$ et $w \in G$; on pose $\bar{g}(w, v) = g(v+w)$ et $\bar{f}(w, v) = f(v+w)$. On a donc $\bar{f} \in C^1(G \times F, \mathbb{R})$ et $\bar{g} \in C^1(G \times F, \mathbb{R})$. De plus, $D_2\bar{g}(w, v) \in \mathcal{L}(F, \mathbb{R}^p)$, et pour tout $z \in F$, on a $D_2\bar{g}(w, v)z = Dg(v+w)z$.

Soit $(\bar{v}, \bar{w}) \in F \times G$ tel que $\bar{u} = \bar{v} + \bar{w}$. Alors $D_2\bar{g}(\bar{w}, \bar{v})z = Dg(\bar{u})z$ pour tout $z \in F$. L'application $D_2\bar{g}(\bar{w}, \bar{v})$ est une bijection de F sur \mathbb{R}^p , car, par définition de F , $Dg(\bar{u})$ est bijective de F sur \mathbb{R}^p .

On rappelle que $K = \{u \in \mathbb{R}^N : g(u) = 0\}$ et on définit $\bar{K} = \{(w, v) \in G \times F, \bar{g}(w, v) = 0\}$. Par définition de \bar{f} et de \bar{g} , on a

$$\begin{cases} (\bar{w}, \bar{v}) \in K \\ \bar{f}(\bar{w}, \bar{v}) \leq f(w, v) \quad \forall (w, v) \in \bar{K} \end{cases} \quad (3.31)$$

D'autre part, le théorème des fonctions implicites (voir note de bas de page 130) entraîne l'existence de $\varepsilon > 0$ et $\nu > 0$ tels que pour tout $w \in B_G(\bar{w}, \varepsilon)$ il existe un unique $v \in B_F(\bar{v}, \nu)$ tel que $\bar{g}(w, v) = 0$. On note $v = \phi(w)$ et on définit ainsi une application $\phi \in C^1(B_G(\bar{w}, \varepsilon), B_F(\bar{v}, \nu))$.

On déduit alors de (3.31) que :

$$\bar{f}(\bar{w}, \phi(\bar{w})) \leq \bar{f}(w, \phi(w)), \quad \forall w \in B_G(\bar{w}, \varepsilon),$$

et donc

$$f(\bar{u}) = f(\bar{w} + \phi(\bar{w})) \leq f(w + \phi(w)), \quad \forall w \in B_G(\bar{w}, \varepsilon).$$

En posant $\psi(w) = \bar{f}(w, \phi(w))$, on peut donc écrire

$$\psi(\bar{w}) = \bar{f}(\bar{w}, \phi(\bar{w})) \leq \psi(w), \quad \forall w \in B_G(\bar{w}, \varepsilon).$$

On a donc, grâce à la proposition 3.36,

$$D\psi(\bar{w}) = 0. \quad (3.32)$$

Par définition de ψ , de \bar{f} et de \bar{g} , on a :

$$D\psi(\bar{w}) = D_1\bar{f}(\bar{w}, \phi(\bar{w})) + D_2\bar{f}(\bar{w}, \phi(\bar{w}))D\phi(\bar{w}).$$

D'après le théorème des fonctions implicites,

$$D\phi(\bar{w}) = -[D_2\bar{g}(\bar{w}, \phi(\bar{w}))]^{-1}D_1\bar{g}(\bar{w}, \phi(\bar{w})).$$

On déduit donc de (3.32) que

$$D_1\bar{f}(\bar{w}, \phi((\bar{w})))w - [D_2\bar{g}(\bar{w}, \phi((\bar{w})))^{-1}D_1\bar{g}(\bar{w}, \phi((\bar{w})))w = 0, \text{ pour tout } w \in G. \quad (3.33)$$

De plus, comme $D_2\bar{g}(\bar{w}, \phi((\bar{w})))^{-1}D_2\bar{g}(\bar{w}, \phi((\bar{w}))) = Id$, on a :

$$D_2\bar{f}(\bar{w}, \phi((\bar{w})))z - D_2\bar{f}(\bar{w}, \phi((\bar{w})))D_2\bar{g}(\bar{w}, \phi((\bar{w})))^{-1}D_2\bar{g}(\bar{w}, \phi((\bar{w})))z = 0, \forall z \in F. \quad (3.34)$$

Soit $x \in \mathbb{R}^N$, et $(z, w) \in F \times G$ tel que $x = z + w$. En additionnant (3.33) et (3.34), et en notant $\Lambda = -D_2\bar{f}(\bar{w}, \phi((\bar{w})))D_2\bar{g}(\bar{w}, \phi((\bar{w})))^{-1}$, on obtient :

$$Df(\bar{u})x + \Lambda Dg(\bar{u})x = 0,$$

ce qui donne, en transposant : $Df(\bar{u}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{u}) = 0$, avec $\Lambda = (\lambda_1, \dots, \lambda_p)$. ■

Remarque 3.39 (Utilisation pratique du théorème de Lagrange) Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$, $g = (g_1, \dots, g_p)^t$ avec $g_i \in C(\mathbb{R}^N, \mathbb{R})$ pour $i = 1, \dots, p$, et soit $K = \{u \in \mathbb{R}^N, g_i(u) = 0, i = 1, \dots, p\}$. Le problème qu'on cherche à résoudre est le problème de minimisation (3.29) qu'on rappelle ici :

$$\begin{cases} \bar{u} \in K \\ f(\bar{u}) = \inf_K f \end{cases}$$

D'après le théorème des multiplicateurs de Lagrange, si \bar{u} est solution de (3.29) et $Im(Dg(\bar{u})) = \mathbb{R}^p$, alors il existe $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tel que \bar{u} est solution du problème

$$\begin{cases} \frac{\partial f}{\partial x_j}(\bar{u}) + \sum_{i=1}^p \lambda_i \frac{\partial g_i}{\partial x_j} = 0, j = 1, \dots, N, \\ g_i(\bar{u}) = 0, i = 1, \dots, p. \end{cases} \quad (3.35)$$

Le système (3.35) est un système non linéaire de $(N+p)$ équations et à $(N+p)$ inconnues $(\bar{x}, \dots, \bar{x}_N, \lambda_1, \dots, \lambda_p)$. Ce système sera résolu par une méthode de résolution de système non linéaire (Newton par exemple).

Remarque 3.40 On vient de montrer que si \bar{x} solution de (3.29) et $Im(Dg(\bar{x})) = \mathbb{R}^p$, alors \bar{x} solution de (3.35). Par contre, si \bar{x} est solution de (3.35), ceci n'entraîne pas que \bar{x} est solution de (3.29).

Des exemples d'application du théorème des multiplicateurs de Lagrange sont donnés dans les exercices 83 page 145 et 84 page 146.

3.4.4 Contraintes inégalités

Soit $f \in C(\mathbb{R}^N, \mathbb{R})$ et $g_i \in C^1(\mathbb{R}^N, \mathbb{R})$ $i = 1, \dots, p$, on considère maintenant un ensemble K de la forme : $K = \{x \in \mathbb{R}^N, g_i(x) \leq 0 \forall i = 1 \dots p\}$, et on cherche à résoudre le problème de minimisation (3.29) qui s'écrit :

$$\begin{cases} \bar{x} \in K \\ f(\bar{x}) \leq f(x), \forall x \in K. \end{cases}$$

Remarque 3.41 Soit \bar{x} une solution de (3.29) et supposons que $g_i(\bar{x}) < 0$, pour tout $i \in \{1, \dots, p\}$. Il existe alors $\varepsilon > 0$ tel que si $x \in B(\bar{x}, \varepsilon)$ alors $g_i(x) < 0$ pour tout $i = 1, \dots, p$.

On a donc $f(\bar{x}) \leq f(x) \forall x \in B(\bar{x}, \varepsilon)$. On est alors ramené à un problème de minimisation sans contrainte, et si f est différentiable en \bar{x} , on a donc $\nabla f(\bar{x}) = 0$.

On donne maintenant sans démonstration le théorème de Kuhn-Tucker qui donne une caractérisation de la solution du problème (3.29).

Théorème 3.42 (Kuhn-Tucker) Soit $f \in C(\mathbb{R}^N, \mathbb{R})$, soit $g_i \in C^1(\mathbb{R}^N, \mathbb{R})$, pour $i = 1, \dots, p$, et soit $K = \{x \in \mathbb{R}^N, g_i(x) \leq 0 \forall i = 1 \dots p\}$. On suppose qu'il existe \bar{x} solution de (3.29), et on pose $I(\bar{x}) = \{i \in \{1, \dots, p\}; g_i(\bar{x}) = 0\}$. On suppose que f est différentiable en \bar{x} et que la famille (de \mathbb{R}^N) $\{\nabla g_i(\bar{x}), i \in I(\bar{x})\}$ est libre. . Alors il existe une famille $(\lambda_i)_{i \in I(\bar{x})} \subset \mathbb{R}_+$ telle que

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Remarque 3.43 1. Le théorème de Kuhn-Tucker s'applique pour des ensembles de contrainte de type inégalité. Si on a une contrainte de type égalité, on peut évidemment se ramener à deux contraintes de type inégalité en remarquant que $\{h(x) = 0\} = \{h(x) \leq 0\} \cap \{-h(x) \leq 0\}$. Cependant, si on pose $g_1 = h$ et $g_2 = -h$, on remarque que la famille $\{\nabla g_1(\bar{x}), \nabla g_2(\bar{x})\} = \{\nabla h(\bar{x}), -\nabla h(\bar{x})\}$ n'est pas libre. On ne peut donc pas appliquer le théorème de Kuhn-Tucker sous la forme donnée précédemment dans ce cas (mais on peut il existe des versions du théorème de Kuhn-Tucker permettant de traiter ce cas, voir Bonans-Sagiez).

2. Dans la pratique, on a intérêt à écrire la conclusion du théorème de Kuhn-Tucker (i.e. l'existence de la famille $(\lambda_i)_{i \in I(\bar{x})}$) sous la forme du système de $N + p$ équations et $2p$ inéquations à résoudre suivant :

$$\begin{cases} \nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0, \\ \lambda_i g_i(\bar{x}) = 0, \quad \forall i = 1, \dots, p, \\ g_i(\bar{x}) \leq 0, \quad \forall i = 1, \dots, p, \\ \lambda_i \geq 0, \quad \forall i = 1, \dots, p. \end{cases}$$

$$i = 1 \dots p \quad g_i(\bar{x}) \leq 0 \quad i = 1 \dots p \\ \lambda_i \geq 0$$

3.5 Algorithmes d'optimisation sous contraintes

3.5.1 Méthodes de gradient avec projection

On rappelle le résultat suivant de projection sur un convexe fermé :

Proposition 3.44 (Projection sur un convexe fermé) Soit E un espace de Hilbert, muni d'une norme $\|\cdot\|$ induite par un produit scalaire (\cdot, \cdot) , et soit K un convexe fermé non vide de E . Alors, tout $x \in E$, il existe une unique $x_0 \in K$ tel que $\|x - x_0\| \leq \|x - y\|$ pour tout $y \in K$. On note $x_0 = p_K(x)$ la projection orthogonale de x sur K . On a également :

$$x_0 = p_K(x) \text{ si et seulement si } (x - x_0, x_0 - y) \geq 0, \quad \forall y \in K.$$

Dans le cadre des algorithmes de minimisation avec contraintes que nous allons développer maintenant, nous considérerons $E = \mathbb{R}^N$, $f \in C^1(\mathbb{R}^N, \mathbb{R})$ une fonction convexe, et K fermé convexe non vide. On cherche à calculer une solution approchée de \bar{x} , solution du problème (3.29).

Algorithme du gradient à pas fixe avec projection sur K (GPFK) Soit $\rho > 0$ donné, on considère l'algorithme suivant :

Algorithme (GPFK)

Initialisation : $x_0 \in K$

Itération :

$$x_n \text{ connu} \quad x_{n+1} = p_K(x_n - \rho \nabla f(x_n))$$

où p_K est la projection sur K définie par la proposition 3.44.

Lemme 3.45 Soit $(x_n)_n$ construite par l'algorithme (GPFK). On suppose que $x_n \rightarrow x$ quand $n \rightarrow \infty$. Alors x est solution de (3.29).

Démonstration :

Soit $p_K : \mathbb{R}^N \rightarrow K \subset \mathbb{R}^N$ la projection sur K définie par la proposition 3.44. Alors p_K est continue. Donc si $x_n \rightarrow x$ quand $n \rightarrow \infty$ alors $x = p_K(x - \rho \nabla f(x))$ et $x \in K$ (car $x_n \in K$ et K est fermé).

La caractérisation de $p_K(x - \rho \nabla f(x))$ donnée dans la proposition 3.44 donne alors :

$(x - \rho \nabla f(x) - x/x - y) \geq 0$ pour tout $y \in K$, et comme $\rho > 0$, ceci entraîne $(\nabla f(x)/x - y)$ pour tout $y \in K$. Or f est convexe donc $f(y) \geq f(x) + \nabla f(x)(y - x)$ pour tout $y \in K$, et donc $f(y) \geq f(x)$ pour tout $y \in K$, ce qui termine la démonstration. ■

Théorème 3.46 (Convergence de l'algorithme GPFK)

Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$, et K convexe fermé non vide. On suppose que :

1. il existe $\alpha > 0$ tel que $(\nabla f(x) - \nabla f(y)|x - y) \geq \alpha|x - y|^2$, pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$,
2. il existe $M > 0$ tel que $|\nabla f(x) - \nabla f(y)| \leq M|x - y|$ pour tout $(x, y) \in \mathbb{R}^N \times \mathbb{R}^N$,

alors :

1. il existe un unique élément $\bar{x} \in K$ solution de (3.29),
2. si $0 < \rho < \frac{2\alpha}{M^2}$, la suite (x_n) définie par l'algorithme (GPFK) converge vers \bar{x} lorsque $n \rightarrow +\infty$.

Démonstration :

1. La condition 1. donne que f est strictement convexe et que $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$. Comme K est convexe fermé non vide, il existe donc un unique \bar{x} solution de (3.29).
2. On pose, pour $x \in \mathbb{R}^N$, $h(x) = p_K(x - \rho \nabla f(x))$. On a donc $x_{n+1} = h(x_n)$. Pour montrer que la suite $(x_n)_{n \in \mathbb{N}}$ converge, il suffit donc de montrer que h est strictement contractante dès que

$$0 < \rho < \frac{2\alpha}{M^2}. \quad (3.36)$$

Grâce au lemme 3.47 démontré plus loin, on sait que p_K est contractante. Or h est définie par :

$$h(x) = p_K(\bar{h}(x)) \quad \text{où } \bar{h}(x) = x - \rho \nabla f(x).$$

On a déjà vu que \bar{h} est strictement contractante si la condition (3.36) est vérifiée (voir théorème 3.17 page 117 et exercice 70 page 139), et plus précisément :

$$|\bar{h}(x) - \bar{h}(y)| \leq (1 - 2\alpha\rho + M^2\rho^2)|x - y|^2.$$

On en déduit que :

$$|h(x) - h(y)|^2 \leq |p_K(\bar{h}(x)) - p_K(\bar{h}(y))|^2 \leq |\bar{h}(x) - \bar{h}(y)|^2 \leq (1 - 2\alpha\rho + \rho^2 M^2)|x - y|^2.$$

L'application h est donc strictement contractante dès que $0 < \frac{2\alpha}{M^2}$. La suite $(x_n)_{n \in \mathbb{N}}$ converge donc bien vers $x = \bar{x}$

Lemme 3.47 (Propriété de contraction de la projection orthogonale) Soit E un espace de Hilbert, $\|\cdot\|$ la norme et (\cdot, \cdot) le produit scalaire, K un convexe fermé non vide de E et p_K la projection orthogonale sur K définie par la proposition 3.44, alors $\|p_K(x) - p_K(y)\| \leq \|x - y\|$ pour tout $(x, y) \in E^2$.

Démonstration Comme E est un espace de Hilbert,

$$\|p_K(x) - p_K(y)\|^2 = (p_K(x) - p_K(y)|p_K(x) - p_K(y)).$$

On a donc

$$\begin{aligned} \|p_K(x) - p_K(y)\|^2 &= (p_K(x) - x + x - y + y - p_K(y)|p_K(x) - p_K(y)) \\ &= (p_K(x) - x|p_K(x) - p_K(y))_E + (x - y|p_K(x) - p_K(y)) + \\ &\quad (y - p_K(y)|p_K(x) - p_K(y)). \end{aligned}$$

Or $(p_K(x) - x|p_K(x) - p_K(y)) \geq 0$ et $(y - p_K(y)|p_K(x) - p_K(y))$, d'où :

$$\|p_K(x) - p_K(y)\| \leq (x - y|p_K(x) - p_K(y)),$$

et donc, grâce à l'inégalité de Cauchy-Schwarz,

$$\|p_K(x) - p_K(y)\| \leq \|x - y\| \|p_K(x) - p_K(y)\| \leq \|x - y\|.$$

■

Algorithme du gradient à pas optimal avec projection sur K (GPOK)

L'algorithme du gradient à pas optimal avec projection sur K s'écrit :

Initialisation $x_0 \in K$

Itération x_n connu

$w_n = -\nabla f(x_n)$; calculer ρ_n optimal dans la direction w_n

$x_{n+1} = p_K(x_n + \rho_n w_n)$

La démonstration de convergence de cet algorithme se déduit de celle de l'algorithme à pas fixe.

Remarque 3.48 On pourrait aussi utiliser un algorithme de type Quasi-Newton avec projection sur K .

Les algorithmes de projection sont simples à décrire, mais ils soulèvent deux questions :

1. Comment calcule-t-on p_K ?
2. Que faire si K n'est pas convexe ?

On peut donner une réponse à la première question dans les cas simples :

1er cas On suppose ici que $K = C^+ = \{x \in \mathbb{R}^N, x = (x_1, \dots, x_N)^t, x_i \geq 0 \forall i\}$.

Si $y \in \mathbb{R}^N, y = (y_1 \dots y_N)^t$, on peut montrer (exercice 3.6 page 147) que

$$(p_K(y))_i = y_i^+ = \max(y_i, 0), \quad \forall i \in \{1, \dots, N\}$$

2ème cas Soit $(\alpha_i)_{i=1, \dots, N} \subset \mathbb{R}^N$ et $(\beta_i)_{i=1, \dots, N} \subset \mathbb{R}^N$ tels que $\alpha_i \leq \beta_i$ pour tout $i = 1, \dots, N$. Si

$$K = \prod_{i=1, N} [\alpha_i, \beta_i],$$

alors

$$(p_K(y))_i = \max(\alpha_i, \min(y_i, \beta_i)), \quad \forall i = 1, \dots, N$$

Dans le cas d'un convexe K plus "compliqué", ou dans le cas où K n'est pas convexe, on peut utiliser des méthodes de dualité introduites dans le paragraphe suivant.

3.5.2 Méthodes de dualité

Supposons que les hypothèses suivantes sont vérifiées :

$$\begin{cases} f \in C^1(\mathbb{R}^N, \mathbb{R}), \\ g_i \in C^1(\mathbb{R}^N, \mathbb{R}), \\ K = \{x \in \mathbb{R}^N, g_i(x) \leq 0 \ i = 1, \dots, p\}, \text{ et } K \text{ est non vide.} \end{cases} \quad (3.37)$$

On définit un problème "primal" comme étant le problème de minimisation d'origine, c'est-à-dire

$$\begin{cases} \bar{x} \in K, \\ f(\bar{x}) \leq f(x), \text{ pour tout } x \in K, \end{cases} \quad (3.38)$$

On définit le "lagrangien" comme étant la fonction L définie de $\mathbb{R}^N \times \mathbb{R}^p$ dans \mathbb{R} par :

$$L(x, \lambda) = f(x) + \lambda \cdot g(x) = f(x) + \sum_{i=1}^p \lambda_i g_i(x), \quad (3.39)$$

avec $g(x) = (g_1(x), \dots, g_p(x))^t$ et $\lambda = (\lambda_1(x), \dots, \lambda_p(x))^t$.

On note C^+ l'ensemble défini par

$$C^+ = \{\lambda \in \mathbb{R}^p, \lambda = (\lambda_1, \dots, \lambda_p)^t, \lambda_i \geq 0 \text{ pour tout } i = 1, \dots, p\}.$$

Remarque 3.49 Le théorème de Kuhn-Tucker entraîne que si \bar{x} est solution du problème primal (3.38) alors il existe $\lambda \in C^+$ tel que $D_1 L(\bar{x}, \lambda) = 0$ (c'est-à-dire $Df(\bar{x}) + \lambda \cdot Dg(\bar{x}) = 0$) et $\lambda \cdot g(\bar{x}) = 0$.

On définit alors l'application M de \mathbb{R}^p dans \mathbb{R} par :

$$M(\lambda) = \inf_{x \in \mathbb{R}^N} L(x, \lambda), \text{ pour tout } \lambda \in \mathbb{R}^p. \quad (3.40)$$

On peut donc remarquer que $M(\lambda)$ réalise le minimum (en x) du problème sans contrainte, qui s'écrit, pour $\lambda \in \mathbb{R}^p$ fixé :

$$\begin{cases} x \in \mathbb{R}^N \\ L(x, \lambda) \leq L(y, \lambda) \text{ pour tout } y \in \mathbb{R}^N, \end{cases} \quad (3.41)$$

Lemme 3.50 *L'application M de \mathbb{R}^p dans \mathbb{R} définie par (3.40) est concave (ou encore l'application $-M$ est convexe), c'est-à-dire que pour tous $\lambda, \mu \in \mathbb{R}^p$ et pour tout $t \in]0, 1[$ on a $M(t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu)$*

Démonstration :

Soit $\lambda, \mu \in \mathbb{R}^p$ et $t \in]0, 1[$; on veut montrer que $M(t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu)$.

Soit $x \in \mathbb{R}^N$, alors :

$$\begin{aligned} L(x, t\lambda + (1-t)\mu) &= f(x) + (t\lambda + (1-t)\mu)g(x) \\ &= tf(x) + (1-t)f(x) + (t\lambda + (1-t)\mu)g(x). \end{aligned}$$

On a donc $L(x, t\lambda + (1-t)\mu) = tL(x, \lambda) + (1-t)L(x, \mu)$. Par définition de M , on en déduit que pour tout $x \in \mathbb{R}^N$,

$$L(x, t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu)$$

Or, toujours par définition de M ,

$$M(t\lambda + (1-t)\mu) = \inf_{x \in \mathbb{R}^N} L(x, t\lambda + (1-t)\mu) \geq tM(\lambda) + (1-t)M(\mu).$$

■

On considère maintenant le problème d'optimisation dit "dual" suivant :

$$\begin{cases} \mu \in C^+, \\ M(\mu) \geq M(\lambda) \quad \forall \lambda \in C^+. \end{cases} \quad (3.42)$$

Définition 3.51 *Soit $L : \mathbb{R}^N \times \mathbb{R}^p \rightarrow \mathbb{R}$ et $(x, \mu) \in \mathbb{R}^N \times C^+$. On dit que (x, μ) est un point selle de L sur $\mathbb{R}^N \times C^+$ si*

$$L(x, \lambda) \leq L(x, \mu) \leq L(y, \mu) \text{ pour tout } y \in \mathbb{R}^N \text{ et pour tout } \lambda \in C^+.$$

Proposition 3.52 *Sous les hypothèses (3.37), soit L définie par $L(x, \lambda) = f(x) + \lambda g(x)$ et $(x, \mu) \in \mathbb{R}^N \times C^+$ un point selle de L sur $\mathbb{R}^N \times C^+$.*

alors

1. \bar{x} est solution du problème (3.38),
2. μ est solution de (3.42),
3. \bar{x} est solution du problème (3.41) avec $\lambda = \mu$.

On admettra cette proposition.

Réciproquement, on peut montrer que (sous des hypothèses convenables sur f et g), si μ est solution de (3.42), et si \bar{x} solution de (3.41) avec $\lambda = \mu$, alors (\bar{x}, μ) est un point selle de L , et donc \bar{x} est solution de (3.38).

De ces résultats découle l'idée de base des méthodes de dualité : on cherche μ solution de (3.42). On obtient ensuite une solution \bar{x} du problème (3.38), en cherchant \bar{x} comme solution du problème (3.41) avec $\lambda = \mu$ (qui est un problème de minimisation sans contraintes). La recherche de la solution μ du problème dual (3.42) peut se faire par exemple par l'algorithme très classique d'Uzawa, que nous décrivons maintenant.

Algorithme d'Uzawa L'algorithme d'Uzawa consiste à utiliser l'algorithme du gradient à pas fixe avec projection (qu'on a appelé "GPFK", voir page 133) pour résoudre de manière itérative le problème dual (3.42). On cherche donc $\mu \in C^+$ tel que $M(\mu) \geq M(\lambda)$ pour tout $\lambda \in C^+$. On se donne $\rho > 0$, et on note p_{C^+} la projection sur le convexe C^+ (voir proposition 3.44 page 133). L'algorithme (GPFK) pour la recherche de μ s'écrit donc :

Initialisation : $\mu_0 \in C_+$

Itération : $\mu_{n+1} = p_{C^+}(\mu_n + \rho \nabla M(\mu_n))$

Pour définir complètement l'algorithme d'Uzawa, il reste à préciser les points suivants :

1. Calcul de $\nabla M(\mu_n)$,
2. calcul de $p_{C^+}(\lambda)$ pour λ dans \mathbb{R}^N .

On peut également s'intéresser aux propriétés de convergence de l'algorithme.

La réponse au point 2 est simple (voir exercice 3.6 page 147) : pour $\lambda \in \mathbb{R}^N$, on calcule $p_{C^+}(\lambda) = \gamma$ avec $\gamma = (\gamma_1, \dots, \gamma_p)^t$ en posant $\gamma_i = \max(0, \lambda_i)$ pour $i = 1, \dots, p$, où $\lambda = (\lambda_1, \dots, \lambda_p)^t$.

La réponse au point 1. est une conséquence de la proposition suivante (qu'on admettra ici) :

Proposition 3.53 *Sous les hypothèses (3.37), on suppose que pour tout $\lambda \in \mathbb{R}^N$, le problème (3.41) admet une solution unique, notée x_λ et on suppose que l'application définie de \mathbb{R}^p dans \mathbb{R}^N par $\lambda \mapsto x_\lambda$ est différentiable. Alors $M(\lambda) = L(x_\lambda, \lambda)$, M est différentiable en λ pour tout λ , et $\nabla M(\lambda) = g(x_\lambda)$.*

En conséquence, pour calculer $\nabla M(\lambda)$, on est ramené à chercher x_λ solution du problème de minimisation sans contrainte (3.41). On peut donc maintenant donner le détail de l'itération générale de l'algorithme d'Uzawa :

Itération de l'algorithme d'Uzawa. Soit $\mu_n \in C^+$ connu ;

1. On cherche $x_n \in \mathbb{R}^N$ solution de $\begin{cases} x_n \in \mathbb{R}^N, \\ L(x_n, \mu_n) \leq L(x, \mu_n), \forall x \in \mathbb{R}^N \end{cases}$ (On a donc $x_n = x_{\mu_n}$)
2. On calcule $\nabla M(\mu_n) = g(x_n)$
3. $\bar{\mu}_{n+1} = \mu_n + \rho \nabla M(\mu_n) = \mu_n + \rho g(x_n) = ((\bar{\mu}_{n+1})_1, \dots, (\bar{\mu}_{n+1})_p)^t$
4. $\mu_{n+1} = p_{C^+}(\bar{\mu}_{n+1})$, c'est-à-dire $\mu_{n+1} = ((\mu_{n+1})_1, \dots, (\mu_{n+1})_p)^t$ avec $(\mu_{n+1})_i = \max(0, (\bar{\mu}_{n+1})_i)$ pour tout $i = 1, \dots, p$.

On a alors le résultat suivant de convergence de l'algorithme :

Proposition 3.54 (Convergence de l'algorithme d'Uzawa) *Sous les hypothèses (3.37), on suppose de plus que :*

1. *il existe $\alpha > 0$ tel que $(\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2$ pour tout $(x, y) \in (\mathbb{R}^N)^2$,*
2. *il existe $M_f > 0$ $|\nabla f(x) - \nabla f(y)| \leq M_f |x - y|$ pour tout $(x, y) \in (\mathbb{R}^N)^2$,*
3. *pour tout $\lambda \in C^+$, il existe un unique $x_\lambda \in \mathbb{R}^N$ tel que $L(x_\lambda, \lambda) \leq L(x, \lambda)$ pour tout $x \in \mathbb{R}^N$.*

Alors si $0 < \rho < \frac{2\alpha}{M_f^2}$, la suite $((x_n, \mu_n))_n \in \mathbb{R}^N \times C^+$ donnée par l'algorithme d'Uzawa vérifie :

1. $x_n \rightarrow \bar{x}$ quand $n \rightarrow +\infty$, où \bar{x} est la solution du problème (3.38),
2. $(\mu_n)_{n \in \mathbb{N}}$ est bornée.

Remarque 3.55 (Sur l'algorithme d'Uzawa)

1. *L'algorithme est très efficace si les contraintes sont affines : (i.e. si $g_i(x) = \alpha_i \cdot x + \beta_i$ pour tout $i = 1, \dots, p$, avec $\alpha_i \in \mathbb{R}^N$ et $\beta_i \in \mathbb{R}$).*
2. *Pour avoir l'hypothèse 3 du théorème, il suffit que les fonctions g_i soient convexes. (On a dans ce cas existence et unicité de la solution x_λ du problème (3.41) et existence et unicité de la solution \bar{x} du problème (3.38).)*

3.6 Exercices

Exercice 66 (Convexité et continuité) *Suggestions en page 149.*

- Soit $f : \mathbb{R} \rightarrow \mathbb{R}$. On suppose que f est convexe.
 - Montrer que f est continue.
 - Montrer que f est localement lipschitzienne.
- Soit $N \geq 1$ et $f : \mathbb{R}^N \rightarrow \mathbb{R}$. On suppose que f est convexe.
 - Montrer f est bornée supérieurement sur les bornés (c'est-à-dire : pour tout $R > 0$, il existe m_R t.q. $f(x) \leq m_R$ si la norme de x est inférieure ou égale à R).
 - Montrer que f est continue.
 - Montrer que f est localement lipschitzienne.
 - On remplace maintenant \mathbb{R}^N par E , e.v.n. de dimension finie. Montrer que f est continue et que f est localement lipschitzienne.
- Soient E un e.v.n. de dimension infinie et $f : E \rightarrow \mathbb{R}$. On suppose que f est convexe.
 - On suppose, dans cette question, que f est bornée supérieurement sur les bornés. Montrer que f est continue.
 - Donner un exemple d'e.v.n. (noté E) et de fonction convexe $f : E \rightarrow \mathbb{R}$ t.q. f soit non continue.

Exercice 67 (Maximisation)

Suggestions en page 149

Soit E un espace vectoriel normé et $f : E \rightarrow \mathbb{R}$.

- Donner une condition suffisante d'existence de $\bar{x} \in E$ tel que $x = \sup_{x \in E} f(x)$.
- Donner une condition suffisante d'unicité de $\bar{x} \in E$ tel que $x = \sup_{x \in E} f(x)$.
- Donner une condition suffisante d'existence et unicité de $\bar{x} \in E$ tel que $x = \sup_{x \in E} f(x)$.

Exercice 68 (Minimisation d'une fonctionnelle quadratique) *Suggestions en page 3.7. Corrigé détaillé en page 151*

Soient $A \in \mathcal{M}_N(\mathbb{R})$, $b \in \mathbb{R}^N$, et f la fonction de \mathbb{R}^N dans \mathbb{R} définie par $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$.

- Montrer que $f \in C^\infty(\mathbb{R}^N, \mathbb{R})$ et calculer le gradient et la matrice hessienne de f en tout point.
- Montrer que si A est symétrique définie positive alors il existe un unique $\bar{x} \in \mathbb{R}^N$ qui minimise f , et que ce \bar{x} est l'unique solution du système linéaire $Ax = b$.

Exercice 69 (Complément de Schur)

Soient n et p deux entiers naturels non nuls. Dans toute la suite, si u et v sont deux vecteurs de \mathbb{R}^k , $k \geq 1$, le produit scalaire de u et v est noté $u \cdot v$. Soient A une matrice carrée d'ordre n , symétrique définie positive, soit B une matrice $n \times p$, C une matrice carrée d'ordre p symétrique, et soit $f \in \mathbb{R}^n$ et $g \in \mathbb{R}^p$. On considère le système linéaire suivant :

$$M \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}, \text{ avec } M = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}. \quad (3.43)$$

- On suppose dans cette question seulement que $n = p = 1$, et $A = [a]$, $B = [b]$, $C = [c]$
 - Donner une condition nécessaire et suffisante sur a , b , et c pour que M soit inversible.
 - Donner une condition nécessaire et suffisante sur a , b , et c pour que M soit symétrique définie positive.
- On définit la matrice $S = C - B^t A^{-1} B$, qu'on appelle "complément de Schur".
 - Calculer S dans le cas $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
 - Montrer qu'il existe une unique solution au problème (3.43) si et seulement si la matrice S est inversible. Est-ce le cas dans la question (a) ?

3. On suppose dans cette question que C est symétrique.
- Vérifier que M est symétrique.
 - Soient $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$ et $z = (x, y) \in \mathbb{R}^{n+p}$. Calculer $Mz \cdot z$ en fonction de A, B, C, x et y .
 - On fixe maintenant $y \in \mathbb{R}^p$, et on définit la fonction F de \mathbb{R}^n dans \mathbb{R} par : $x \mapsto Ax \cdot x + 2By \cdot x + Cy \cdot y$. Calculer $\nabla F(x)$, et calculer $x_0 \in \mathbb{R}^n$ tel que $\nabla F(x_0) = 0$
 - Montrer que la fonction F définie en 3(b) admet un unique minimum, et calculer la valeur de ce minimum.
 - En déduire que M est définie positive si et seulement si S est définie positive (où S est la matrice définie à la question 1).
4. On suppose dans cette question que C est la matrice (carrée d'ordre p) nulle.
- Montrer que la matrice $\tilde{S} = -S$ est symétrique définie positive si et seulement si $p \leq n$ et $\text{rang}(B)=p$. On supposera que ces deux conditions sont vérifiées dans toute la suite de la question.
 - En déduire que la matrice $P = \begin{bmatrix} A & 0 \\ 0 & \tilde{S} \end{bmatrix}$ est symétrique définie positive.
 - Calculer les valeurs propres de la matrice $T = P^{-1}M$ (il peut être utile de distinguer les cas $\text{Ker}B^t = \{0\}$ et $\text{Ker}B^t \neq \{0\}$).

Exercice 70 (Convergence de l'algorithme du gradient à pas fixe)

Suggestions en page 149, corrigé détaillé en page 151

Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$ ($N \geq 1$). On suppose que f vérifie :

$$\exists \alpha > 0 \text{ t.q. } (\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2, \forall x, y \in \mathbb{R}^N, \quad (3.44)$$

$$\exists M > 0 \text{ t.q. } |\nabla f(x) - \nabla f(y)| \leq M|x - y|, \forall x, y \in \mathbb{R}^N. \quad (3.45)$$

1. Montrer que

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x) + \frac{\alpha}{2}|y - x|^2, \forall x, y \in \mathbb{R}^N.$$

- Montrer que f est strictement convexe et que $f(x) \rightarrow \infty$ quand $|x| \rightarrow \infty$. En déduire qu'il existe un et un seul $\bar{x} \in \mathbb{R}^N$ t.q. $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$.
- Soient $\rho \in]0, (2\alpha/M^2)[$ et $x_0 \in \mathbb{R}^N$. Montrer que la suite $(x_n)_{n \in \mathbb{N}}$ définie par $x_{n+1} = x_n - \rho \nabla f(x_n)$ (pour $n \in \mathbb{N}$) converge vers \bar{x} .

Exercice 71 (Mise en oeuvre de GPF et GPO) Corrigé en page 152.

On considère la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x_1, x_2) = 2x_1^2 + x_2^2 - x_1x_2 - 3x_1 - x_2 + 4$.

- Montrer qu'il existe un unique $\bar{x} \in \mathbb{R}^2$ tel que $\bar{x} = \min_{x \in \mathbb{R}^2} f(x)$ admet un unique minimum, et le calculer.
- Calculer le premier itéré donné par l'algorithme du gradient à pas fixe (GPF) et du gradient à pas optimal (GPO), en partant de $(x_1^{(0)}, x_2^{(0)}) = (0, 0)$, pour un pas de $\rho = .5$ dans le cas de GPF.

Exercice 72 (Convergence de l'algorithme du gradient à pas optimal) Suggestions en page 150. Corrigé détaillé en page 153

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R})$ t.q. $f(x) \rightarrow \infty$ quand $|x| \rightarrow \infty$. Soit $x_0 \in \mathbb{R}^N$. On va démontrer dans cet exercice la convergence de l'algorithme du gradient à pas optimal.

- Montrer qu'il existe $R > 0$ t.q. $f(x) > f(x_0)$ pour tout $x \notin B_R$, avec $B_R = \{x \in \mathbb{R}^N, |x| \leq R\}$.
- Montrer qu'il existe $M > 0$ t.q. $|H(x)y \cdot y| \leq M|y|^2$ pour tout $y \in \mathbb{R}^N$ et tout $x \in B_{R+1}$ ($H(x)$ est la matrice hessienne de f au point x , R est donné à la question 1).
- (Construction de "la" suite $(x_n)_{n \in \mathbb{N}}$ de l'algorithme du gradient à pas optimal.) On suppose x_n connu ($n \in \mathbb{N}$). On pose $w_n = -\nabla f(x_n)$. Si $w_n = 0$, on pose $x_{n+1} = x_n$. Si $w_n \neq 0$, montrer qu'il existe $\bar{\rho} > 0$ t.q. $f(x_n + \bar{\rho}w_n) \leq f(x_n + \rho w_n)$ pour tout $\rho \geq 0$. On choisit alors un $\rho_n > 0$ t.q. $f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n)$ pour tout $\rho \geq 0$ et on pose $x_{n+1} = x_n + \rho_n w_n$.
On considère, dans les questions suivantes, la suite $(x_n)_{n \in \mathbb{N}}$ ainsi construite.
- Montrer que (avec R et M donnés aux questions précédentes)

- (a) la suite $(f(x_n))_{n \in \mathbb{N}}$ est une suite convergente,
 (b) $x_n \in B_R$ pour tout $n \in \mathbb{N}$,
 (c) $f(x_n + \rho w_n) \leq f(x_n) - \rho|w_n|^2 + (\rho^2/2)M|w_n|^2$ pour tout $\rho \in [0, 1/|w_n|]$.
 (d) $f(x_{n+1}) \leq f(x_n) - |w_n|^2/(2M)$, si $|w_n| \leq M$.
 (e) $-f(x_{n+1}) + f(x_n) \geq |w_n|^2/(2\bar{M})$, avec $\bar{M} = \sup(M, \tilde{M})$,
 $\tilde{M} = \sup\{|\nabla f(x)|, x \in B_R\}$.
5. Montrer que $\nabla f(x_n) \rightarrow 0$ (quand $n \rightarrow \infty$) et qu'il existe une sous suite $(n_k)_{k \in \mathbb{N}}$ t.q. $x_{n_k} \rightarrow x$ quand $k \rightarrow \infty$ et $\nabla f(x) = 0$.
6. On suppose qu'il existe un unique $\bar{x} \in \mathbb{R}^N$ t.q. $\nabla f(\bar{x}) = 0$. Montrer que $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$ et que $x_n \rightarrow \bar{x}$ quand $n \rightarrow \infty$.

Exercice 73 (Algorithme du gradient à pas optimal)

Soit $A \in \mathcal{M}_N(\mathbb{R})$ et J la fonction définie de \mathbb{R}^N dans \mathbb{R} par $J(x) = e^{\|Ax\|^2}$, où $\|\cdot\|$ désigne la norme euclidienne sur \mathbb{R}^N .

1. Montrer que J admet un minimum (on pourra le calculer...).
2. On suppose que la matrice A est inversible, montrer que ce minimum est unique.
3. Ecrire l'algorithme du gradient à pas optimal pour la recherche de ce minimum. [On demande de calculer le paramètre optimal ρ_n en fonction de A et de x_n .] A quelle condition suffisante cet algorithme converge-t-il?

Exercice 74 (Fonction non croissante à l'infini) *Suggestions en page 150.*

Soient $N \geq 1$, $f \in C^2(\mathbb{R}^N, \mathbb{R})$ et $a \in \mathbb{R}$. On suppose que $A = \{x \in \mathbb{R}^N; f(x) \leq f(a)\}$ est un ensemble borné de \mathbb{R}^N et qu'il existe $M \in \mathbb{R}$ t.q. $|H(x)y \cdot y| \leq M|y|^2$ pour tout $x, y \in \mathbb{R}^N$ (où $H(x)$ désigne la matrice hessienne de f au point x).

1. Montrer qu'il existe $\bar{x} \in A$ t.q. $f(\bar{x}) = \min\{f(x), x \in \mathbb{R}^N\}$ (noter qu'il n'y a pas nécessairement unicité de \bar{x}).
2. Soit $x \in A$ t.q. $\nabla f(x) \neq 0$. On pose $T(x) = \sup\{\rho \geq 0; [x, x - \rho \nabla f(x)] \subset A\}$. Montrer que $0 < T(x) < +\infty$ et que $[x, x - T(x)\nabla f(x)] \subset A$ (où $[x, x - T(x)\nabla f(x)]$ désigne l'ensemble $\{tx + (1-t)(x - T(x)\nabla f(x)), t \in [0, 1]\}$).
3. Pour calculer une valeur approchée de \bar{x} (t.q. $f(\bar{x}) = \min\{f(x), x \in \mathbb{R}^N\}$), on propose l'algorithme suivant :

Initialisation : $x_0 \in A$,

Itérations : Soit $k \geq 0$. Si $\nabla f(x_k) = 0$, on pose $x_{k+1} = x_k$. Si $\nabla f(x_k) \neq 0$, On choisit $\rho_k \in [0, T(x_k)]$ t.q. $f(x_k - \rho_k \nabla f(x_k)) = \min\{f(x_k - \rho \nabla f(x_k)), 0 \leq \rho \leq T(x_k)\}$ (La fonction T est définie à la question 2) et on pose $x_{k+1} = x_k - \rho_k \nabla f(x_k)$.

- (a) Montrer que, pour tout $x_0 \in A$, l'algorithme précédent définit une suite $(x_k)_{k \in \mathbb{N}} \subset A$ (c'est-à-dire que, pour $x_k \in A$, il existe bien au moins un élément de $[0, T(x_k)]$, noté ρ_k , t.q. $f(x_k - \rho_k \nabla f(x_k)) = \min\{f(x_k - \rho \nabla f(x_k)), 0 \leq \rho \leq T(x_k)\}$).
 - (b) Montrer que cet algorithme n'est pas nécessairement l'algorithme du gradient à pas optimal. [on pourra chercher un exemple avec $N = 1$.]
 - (c) Montrer que $f(x_k) - f(x_{k+1}) \geq \frac{|\nabla f(x_k)|^2}{2M}$, pour tout $k \in \mathbb{N}$.
4. On montre maintenant la convergence de la suite $(x_k)_{k \in \mathbb{N}}$ construite à la question précédente.
- (a) Montrer qu'il existe une sous suite $(x_{k_n})_{n \in \mathbb{N}}$ et $x \in A$ t.q. $x_{k_n} \rightarrow x$, quand $n \rightarrow \infty$, et $\nabla f(x) = 0$.
 - (b) On suppose, dans cette question, qu'il existe un et un seul élément $z \in A$ t.q. $\nabla f(z) = 0$. Montrer que $x_k \rightarrow z$, quand $k \rightarrow \infty$, et que $f(z) = \min\{f(x), x \in A\}$.

Exercice 75 (Méthode de relaxation) *Corrigé détaillé en page 155*

Soit f une fonction continûment différentiable de $E = \mathbb{R}^N$ dans \mathbb{R} vérifiant l'hypothèse (3.44) :

1. Justifier l'existence et l'unicité de $\bar{x} \in \mathbb{R}^N$ tel que $f(\bar{x}) = \inf_{x \in \mathbb{R}^N} f(x)$.

On propose l'algorithme de recherche de minimum de f suivant :

$$\left\{ \begin{array}{l} \text{Initialisation : } x^{(0)} \in E, \\ \text{Itération } n : \quad x^{(n)} \text{ connu, } (n \geq 0) \\ \quad \text{Calculer } x_1^{(n+1)} \text{ tel que, pour tout } \xi \in \mathbb{R}, \\ \quad f(x_1^{(n+1)}, x_2^{(n)}, x_3^{(n)}, \dots, x_N^{(n)}) \leq f(\xi, x_2^{(n)}, x_3^{(n)}, \dots, x_N^{(n)}), \\ \quad \text{Calculer } x_2^{(n+1)} \text{ tel que, pour tout } \xi \in \mathbb{R}, \\ \quad f(x_1^{(n+1)}, x_2^{(n+1)}, x_3^{(n)}, \dots, x_N^{(n)}) \leq f(x_1^{(n+1)}, \xi, x_3^{(n)}, \dots, x_N^{(n)}), \\ \quad \dots \\ \quad \text{Calculer } x_k^{(n+1)} \text{ tel que, pour tout } \xi \in \mathbb{R}, \\ \quad f(x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, x_k^{(n+1)}, x_{k+1}^{(n)}, \dots, x_N^{(n)}) \\ \quad \leq f(x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, \xi, x_{k+1}^{(n)}, \dots, x_N^{(n)}), \\ \quad \dots \\ \quad \text{Calculer } x_N^{(n+1)} \text{ tel que, pour tout } \xi \in \mathbb{R}, \\ \quad f(x_1^{(n+1)}, x_2^{(n+1)}, \dots, x_{N-1}^{(n+1)}, x_N^{(n+1)}) \leq f(x_1^{(n+1)}, \dots, x_{N-1}^{(n+1)}, \xi). \end{array} \right. \quad (3.46)$$

2. Pour $n \in \mathbb{N}$ et $1 \leq k \leq N$, soit $\varphi_k^{(n+1)}$ la fonction de \mathbb{R} dans \mathbb{R} définie par :

$$\varphi_k^{(n+1)}(s) = f(x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, s, x_{k+1}^{(n)}, \dots, x_N^{(n)}).$$

Montrer qu'il existe un unique élément $\bar{s} \in \mathbb{R}$ tel que

$$\varphi_k^{(n+1)}(\bar{s}) = \inf_{s \in \mathbb{R}} \varphi_k^{(n+1)}(s).$$

En déduire que la suite $(x^{(n)})_{n \in \mathbb{N}}$ construite par l'algorithme (3.46) est bien définie.

Dans toute la suite, on note $\|\cdot\|$ la norme euclidienne sur \mathbb{R}^N et (\cdot, \cdot) le produit scalaire associé. Pour $i = 1, \dots, N$, on désigne par $\partial_i f$ la dérivée partielle de f par rapport à la i -ème variable.

3. Soit $(x^{(n)})_{n \in \mathbb{N}}$ la suite définie par l'algorithme (3.46). Pour $n \geq 0$, on définit $x^{(n+1,0)} = x^{(n)} = (x_1^{(n)}, \dots, x_N^{(n)})^t$, et pour $1 \leq k \leq N$, $x^{(n+1,k)} = (x_1^{(n+1)}, \dots, x_k^{(n+1)}, x_{k+1}^{(n)}, \dots, x_N^{(n)})^t$ (de sorte que $x^{(n+1,N)} = x^{(n+1)}$).

(a) Soit $n \in \mathbb{N}$. Pour $1 \leq k \leq N$, montrer que $\partial_k f(x^{(n+1,k)}) = 0$, pour $k = 1, \dots, N$. En déduire que

$$f(x^{(n+1,k-1)}) - f(x^{(n+1,k)}) \geq \frac{\alpha}{2} \|x^{(n+1,k-1)} - x^{(n+1,k)}\|^2.$$

(b) Montrer que la suite $(x^{(n)})_{n \in \mathbb{N}}$ vérifie

$$f(x^{(n)}) - f(x^{(n+1)}) \geq \frac{\alpha}{2} \|x^{(n)} - x^{(n+1)}\|^2.$$

En déduire que $\lim_{n \rightarrow +\infty} \|x^{(n)} - x^{(n+1)}\| = 0$ et que, pour $1 \leq k \leq N$, $\lim_{n \rightarrow +\infty} \|x^{(n+1,k)} - x^{(n+1)}\| = 0$.

4. Montrer que

$$\|x^{(n+1)} - \bar{x}\| \leq \frac{1}{\alpha} \left(\sum_{k=1}^N |\partial_k f(x^{(n+1)})|^2 \right)^{\frac{1}{2}}.$$

5. Montrer que les suites $(x^{(n)})_{n \in \mathbb{N}}$, et $(x^{(n+1,k)})_{n \in \mathbb{N}}$, pour $k = 1, \dots, N$, sont bornées.

Montrer que

$$|\partial_k f(x^{(n+1)})| \rightarrow 0 \text{ lorsque } n \rightarrow +\infty.$$

(On rappelle que $\partial_k f(x^{(n+1,k)}) = 0$.)

Conclure quant à la convergence de la suite $(x^{(n)})_{n \in \mathbb{N}}$ lorsque $n \rightarrow +\infty$.

6. On suppose dans cette question que $f(x) = \frac{1}{2}(Ax|x) - (b|x)$. Montrer que dans ce cas, l'algorithme (3.46) est équivalent à une méthode itérative de résolution de systèmes linéaires qu'on identifiera.

7. On suppose dans cette question que $N = 2$. Soit g la fonction définie de \mathbb{R}^2 dans \mathbb{R} par : $g(x) = x_1^2 + x_2^2 - 2(x_1 + x_2) + 2|x_1 - x_2|$, avec $x = (x_1, x_2)^t$.

(a) Montrer qu'il existe un unique élément $\bar{x} = (\bar{x}_1, \bar{x}_2)^t$ de \mathbb{R}^2 tel que $g(\bar{x}) = \inf_{x \in \mathbb{R}^2} g(x)$.

(b) Montrer que $\bar{x} = (1, 1)^t$.

(c) Montrer que si $x^{(0)} = (0, 0)^t$, l'algorithme (3.46) appliqué à g ne converge pas vers \bar{x} . Quelle est l'hypothèse mise en défaut ici ?

Exercice 76 (Gradient conjugué pour une matrice non symétrique) *Corrigé détaillé en page 157*

Soit $N \in \mathbb{N}$, $N \geq 1$. On désigne par $\|\cdot\|$ la norme euclidienne sur \mathbb{R}^N , et on munit l'ensemble $\mathcal{M}_N(\mathbb{R})$ de la norme induite par la norme $\|\cdot\|$, $\|\cdot\|$. Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice inversible. On définit $M \in \mathcal{M}_N(\mathbb{R})$ par $M = A^t A$. On se donne un vecteur $b \in \mathbb{R}^N$, et on s'intéresse à la résolution du système linéaire

$$Ax = b; . \quad (3.47)$$

1. Montrer que $x \in \mathbb{R}^N$ est solution de (1.63) si et seulement si x est solution de

$$Mx = A^t b; . \quad (3.48)$$

2. On rappelle que le conditionnement d'une matrice $C \in \mathcal{M}_N(\mathbb{R})$ inversible est défini par $\text{cond}(C) = \|C\| \|C^{-1}\|$ (et dépend donc de la norme considérée ; on rappelle qu'on a choisi ici la norme induite par la norme euclidienne).

(a) Montrer que les valeurs propres de la matrice M sont toutes strictement positives.

(b) Montrer que $\text{cond}(A) = \sqrt{\frac{\lambda_N}{\lambda_1}}$, où λ_N (resp. λ_1) est la plus grande (resp. plus petite) valeur propre de M .

3. Ecrire l'algorithme du gradient conjugué pour la résolution du système (3.48), en ne faisant intervenir que les matrices A et A^t (et pas la matrice M) et en essayant de minimiser le nombre de calculs. Donner une estimation du nombre d'opérations nécessaires et comparer par rapport à l'algorithme du gradient conjugué écrit dans le cas d'une matrice carré d'ordre N symétrique définie positive.

Exercice 77 (Gradient conjugué préconditionné par LL^t)

Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive, et $b \in \mathbb{R}^N$. Soit L une matrice triangulaire inférieure inversible, soit $B = L^{-1}A(L^t)^{-1}$ et $\tilde{b} = L^{-1}b$.

1. Montrer que B est symétrique définie positive.

2. Justifier l'existence et l'unicité de $x \in \mathbb{R}^N$ tel que $Ax = b$, et de $y \in \mathbb{R}^N$ tel que $By = \tilde{b}$. Ecrire x en fonction de y .

Soit $y^{(0)} \in \mathbb{R}^N$ fixé. On pose $\tilde{r}^{(0)} = \tilde{w}^{(0)} = \tilde{b} - By^{(0)}$. Si $\tilde{r}^{(0)} \neq 0$, on pose alors $y^{(1)} = y^{(0)} + \rho_0 \tilde{w}^{(0)}$, avec $\rho_0 = \frac{\tilde{r}^{(0)} \cdot \tilde{r}^{(0)}}{\tilde{w}^{(0)} \cdot A \tilde{w}^{(0)}}$.

Pour $n > 1$, on suppose $y^{(0)}, \dots, y^{(n)}$ et $\tilde{w}^{(0)}, \dots, \tilde{w}^{(n-1)}$ connus, et on pose : $\tilde{r}^{(n)} = \tilde{b} - By^{(n)}$. Si $\tilde{r}^{(n)} \neq 0$, on calcule : $\tilde{w}^{(n)} = \tilde{r}^{(n)} + \lambda_{n-1} \tilde{w}^{(n-1)}$ avec $\lambda_{n-1} = \frac{\tilde{r}^{(n)} \cdot \tilde{r}^{(n)}}{\tilde{r}^{(n-1)} \cdot \tilde{r}^{(n-1)}}$ et on pose alors : $y^{(n+1)} = y^{(n)} + \rho_n \tilde{w}^{(n)}$ avec $\rho_n = \frac{\tilde{r}^{(n)} \cdot \tilde{r}^{(n)}}{\tilde{w}^{(n)} \cdot B \tilde{w}^{(n)}}$.

3. En utilisant le cours, justifier que la famille $y^{(n)}$ ainsi construite est finie. A quoi est égale sa dernière valeur ?

Pour $n \in \mathbb{N}$, on pose : $x^{(n)} = L^{-t} y^{(n)}$ (avec $L^{-t} = (L^{-1})^t = (L^t)^{-1}$), $r^{(n)} = b - Ax^{(n)}$, $w^{(n)} = L^{-t} \tilde{w}^{(n)}$ et $s^{(n)} = (LL^t)^{-1} r^{(n)}$.

4. Soit $n > 0$ fixé. Montrer que :

$$(a) \quad \lambda_{n-1} = \frac{s^{(n)} \cdot r^{(n)}}{s^{(n-1)} \cdot r^{(n-1)}}, \quad (b) \quad \rho_n = \frac{s^{(n)} \cdot r^{(n)}}{w^{(n)} \cdot Aw^{(n)}},$$

$$(c) \quad w^{(n)} = s^{(n)} + \lambda_n w^{(n-1)}, \quad (d) \quad x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}.$$

5. On suppose que la matrice LL^t est une factorisation de Choleski incomplète de la matrice A . Ecrire l'algorithme du gradient conjugué préconditionné par cette factorisation, pour la résolution du système $Ax = b$.

Exercice 78 (Méthode de Polak-Ribière) Suggestions en page 150, corrigé en page 158

Dans cet exercice, on démontre la convergence de la méthode de Polak-Ribière (méthode de gradient conjugué pour une fonctionnelle non quadratique) sous des hypothèses "simples" sur f .

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R})$. On suppose qu'il existe $\alpha > 0, \beta \geq \alpha$ t.q. $\alpha|y|^2 \leq H(x)y \cdot y \leq \beta|y|^2$ pour tout $x, y \in \mathbb{R}^N$. ($H(x)$ est la matrice hessienne de f au point x .)

1. montrer que f est strictement convexe, que $f(x) \rightarrow \infty$ quand $|x| \rightarrow \infty$ et que le spectre $\mathcal{VP}(H(x))$ de $H(x)$ est inclus dans $[\alpha, \beta]$ pour tout $x \in \mathbb{R}^N$.

On note \bar{x} l'unique point de \mathbb{R}^N t.q. $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$ (l'existence et l'unicité de \bar{x} est donné par la question précédente). On cherche une approximation de \bar{x} en utilisant l'algorithme de Polak-Ribière :

initialisation. $x^{(0)} \in \mathbb{R}^N$. On pose $g^{(0)} = -\nabla f(x^{(0)})$. Si $g^{(0)} = 0$, l'algorithme s'arrête (on a $x^{(0)} = \bar{x}$). Si $g^{(0)} \neq 0$, on pose $w^{(0)} = g^{(0)}$ et $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$ avec ρ_0 "optimal" dans la direction $w^{(0)}$.

itération. $x^{(n)}, w^{(n-1)}$ connus ($n \geq 1$). On pose $g^{(n)} = -\nabla f(x^{(n)})$. Si $g^{(n)} = 0$, l'algorithme s'arrête (on a $x^{(n)} = \bar{x}$). Si $g^{(n)} \neq 0$, on pose $\lambda_{n-1} = [g^{(n)} \cdot (g^{(n)} - g^{(n-1)})] / [g^{(n-1)} \cdot g^{(n-1)}]$, $w^{(n)} = g^{(n)} + \lambda_{n-1} w^{(n-1)}$ et $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$ avec ρ_n "optimal" dans la direction w_n . (Noter que ρ_n existe bien.)

On suppose dans la suite que $g^{(n)} \neq 0$ pour tout $n \in \mathbb{N}$.

2. Montrer (par récurrence sur n) que $g^{(n+1)} \cdot w^{(n)} = 0$ et $g^{(n)} \cdot g^{(n)} = g^{(n)} \cdot w^{(n)}$, pour tout $n \in \mathbb{N}$.
3. On pose

$$J^{(n)} = \int_0^1 H(x^{(n)} + \theta \rho_n w^{(n)}) d\theta.$$

Montrer que $g^{(n+1)} = g^{(n)} + \rho_n J^{(n)} w^{(n)}$ et que $\rho_n = (-g^{(n)} \cdot w^{(n)}) / (J^{(n)} w^{(n)} \cdot w^{(n)})$ (pour tout $n \in \mathbb{N}$).

4. Montrer que $|w^{(n)}| \leq (1 + \beta/\alpha)|g^{(n)}|$ pour tout $n \in \mathbb{N}$. [Utiliser, pour $n \geq 1$, la question précédente et la formule donnant λ_{n-1} .]
5. Montrer que $x^{(n)} \rightarrow \bar{x}$ quand $n \rightarrow \infty$.

Exercice 79 (Algorithme de quasi Newton)

Corrigé détaillé en page 162

Soit $A \in \mathcal{M}_N(\mathbb{R})$ une matrice symétrique définie positive et $b \in \mathbb{R}^N$. On pose $f(x) = (1/2)Ax \cdot x - b \cdot x$ pour $x \in \mathbb{R}^N$. On rappelle que $\nabla f(x) = Ax - b$. Pour calculer $\bar{x} \in \mathbb{R}^N$ t.q. $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$, on va utiliser un algorithme de quasi Newton, c'est-à-dire :

initialisation. $x^{(0)} \in \mathbb{R}^N$.

itération. $x^{(n)}$ connu ($n \geq 0$). On pose $x^{(n+1)} = x^{(n)} - \rho_n K^{(n)} g^{(n)}$ avec $g^{(n)} = \nabla f(x^{(n)})$, $K^{(n)}$ une matrice symétrique définie positive à déterminer et ρ_n "optimal" dans la direction $w^{(n)} = -K^{(n)} g^{(n)}$. (Noter que ρ_n existe bien.)

Partie 1. Calcul de ρ_n . On suppose que $g^{(n)} \neq 0$.

1. Montrer que $w^{(n)}$ est une direction de descente stricte en $x^{(n)}$ et calculer la valeur de ρ_n (en fonction de $K^{(n)}$ et $g^{(n)}$).
2. On suppose que, pour un certain $n \in \mathbb{N}$, on a $K^{(n)} = (H(x^{(n)}))^{-1}$ (où $H(x)$ est la matrice hessienne de f en x , on a donc ici $H(x) = A$ pour tout $x \in \mathbb{R}^N$). Montrer que $\rho_n = 1$.
3. Montrer que la méthode de Newton pour calculer \bar{x} converge en une itération (mais nécessite la résolution du système linéaire $A(x^{(1)} - x^{(0)}) = b - Ax^{(0)}$...)

Partie 2. Méthode de Fletcher-Powell. On prend maintenant $K^{(0)} = Id$ et

$$K^{(n+1)} = K^{(n)} + \frac{s^{(n)}(s^{(n)})^t}{s^{(n)} \cdot y^{(n)}} - \frac{(K^{(n)} y^{(n)})(K^{(n)} (y^{(n)})^t)}{K^{(n)} y^{(n)} \cdot y^{(n)}}, \quad n \geq 0, \quad (3.49)$$

avec $s^{(n)} = x^{(n+1)} - x^{(n)}$ et $y^{(n)} = g^{(n+1)} - g^{(n)} = A s^{(n)}$.

On va montrer que cet algorithme converge en au plus N itérations (c'est-à-dire qu'il existe $n \leq N + 1$ t.q. $x_{N+1} = \bar{x}$.)

1. Soit $n \in \mathbb{N}$. On suppose, dans cette question, que $s^{(0)}, \dots, s^{(n-1)}$ sont des vecteurs A -conjugués et non-nuls et que $K^{(0)}, \dots, K^{(n)}$ sont des matrices symétriques définies positives t.q. $K^{(j)}As^{(i)} = s^{(i)}$ si $0 \leq i < j \leq n$ (pour $n = 0$ on demande seulement $K^{(0)}$ symétrique définie positive).

- (a) On suppose que $g^{(n)} \neq 0$. Montrer que $s^{(n)} \neq 0$ (cf. Partie I) et que, pour $i < n$,

$$s^{(n)} \cdot As^{(i)} = 0 \Leftrightarrow g^{(n)} \cdot s^{(i)} = 0.$$

Montrer que $g^{(n)} \cdot s^{(i)} = 0$ pour $i < n$. [On pourra remarquer que $g^{(i+1)} \cdot s^{(i)} = g^{(i+1)} \cdot w^{(i)} = 0$ et $(g^{(n)} - g^{(i+1)}) \cdot s^{(i)} = 0$ par l'hypothèse de conjugaison de $s^{(0)}, \dots, s^{(n-1)}$.] En déduire que $s^{(0)}, \dots, s^{(n)}$ sont des vecteurs A -conjugués et non-nuls.

- (b) Montrer que $K^{(n+1)}$ est symétrique.
 (c) Montrer que $K^{(n+1)}As^{(i)} = s^{(i)}$ si $0 \leq i \leq n$.
 (d) Montrer que, pour tout $x \in \mathbb{R}^N$, on a

$$K^{(n+1)}x \cdot x = \frac{(K^{(n)}x \cdot x)(K^{(n)}y^{(n)} \cdot y^{(n)}) - (K^{(n)}y^{(n)} \cdot x)^2}{K^{(n)}y^{(n)} \cdot y^{(n)}} + \frac{(s^{(n)} \cdot x)^2}{As^{(n)} \cdot s^{(n)}}.$$

En déduire que $K^{(n+1)}$ est symétrique définie positive. [On rappelle (inégalité de Cauchy-Schwarz) que, si K est symétrique définie positive, on a $(Kx \cdot y)^2 \leq (Kx \cdot x)(Ky \cdot y)$ et l'égalité a lieu si et seulement si x et y sont colinéaires.]

2. On suppose que $g^{(n)} \neq 0$ si $0 \leq n \leq N-1$. Montrer (par récurrence sur n , avec la question précédente) que $s^{(0)}, \dots, s^{(N-1)}$ sont des vecteurs A -conjugués et non-nuls et que $K^{(N)}As^{(i)} = s^{(i)}$ si $i < N$. En déduire que $K^{(N)} = A^{-1}$, $\rho_N = 1$ et $x^{(N+1)} = A^{-1}b = \bar{x}$.

Exercice 80 (Méthodes de Gauss-Newton et de quasi-linéarisation) *Corrigé détaillé en page 161*

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R}^P)$, avec $N, P \in \mathbb{N}^*$. Soit $C \in \mathcal{M}_P(\mathbb{R})$ une matrice réelle carrée d'ordre P , symétrique définie positive, et $d \in \mathbb{R}^P$. Pour $x \in \mathbb{R}^N$, on pose

$$J(x) = (f(x) - d) \cdot C(f(x) - d).$$

On cherche à minimiser J .

I Propriétés d'existence et d'unicité

- (a) Montrer que J est bornée inférieurement.
 (b) Donner trois exemples de fonctions f pour lesquels les fonctionnelles J associées sont telles que l'on ait :
- i. existence et unicité de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J , pour le premier exemple.
 - ii. existence et non unicité de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J , pour le second exemple.
 - iii. non existence de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J , pour le troisième exemple.

(On pourra prendre $N = P = 1$.)

II Un peu de calcul différentiel

- (a) On note Df et D_2f les différentielles d'ordre 1 et 2 de f . A quels espaces appartiennent $Df(x)$, $D_2f(x)$ (pour $x \in \mathbb{R}^N$), ainsi que Df et D_2f ? Montrer que pour tout $x \in \mathbb{R}^N$, il existe $M(x) \in \mathcal{M}_{P,N}(\mathbb{R})$, où $\mathcal{M}_{P,N}(\mathbb{R})$ désigne l'ensemble des matrices réelles à P lignes et N colonnes, telle que $Df(x)(y) = M(x)y$ pour tout $y \in \mathbb{R}^N$.
 (b) Pour $x \in \mathbb{R}^N$, calculer $\nabla J(x)$.
 (c) Pour $x \in \mathbb{R}^N$, calculer la matrice hessienne de J en x (qu'on notera $H(x)$). On suppose maintenant que M ne dépend pas de x ; montrer que dans ce cas $H(x) = 2M(x)^t C M(x)$.

III Algorithmes d'optimisation

Dans toute cette question, on suppose qu'il existe un unique $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J , qu'on cherche à calculer de manière itérative. On se donne pour cela $x_0 \in \mathbb{R}^N$, et on cherche à construire une suite $(x_n)_{n \in \mathbb{N}}$ qui converge vers \bar{x} .

- (a) On cherche à calculer \bar{x} en utilisant la méthode de Newton pour annuler ∇J . Justifier brièvement cette procédure et écrire l'algorithme obtenu.
- (b) L'algorithme dit de "Gauss-Newton" est une modification de la méthode précédente, qui consiste à approcher, à chaque itération n , la matrice jacobienne de ∇J en x_n par la matrice obtenue en négligeant les dérivées secondes de f . Ecrire l'algorithme ainsi obtenu.
- (c) L'algorithme dit de "quasi-linéarisation" consiste à remplacer, à chaque itération $n \in \mathbb{N}$, la minimisation de la fonctionnelle J par celle de la fonctionnelle J_n , définie de \mathbb{R}^N dans \mathbb{R} , et obtenue à partir de J en effectuant un développement limité au premier ordre de $f(x)$ en x_n , c.à.d.

$$J_n(x) = (f(x_n) + Df(x_n)(x - x_n) - d) \cdot C(f(x_n) + Df(x_n)(x - x_n) - d).$$

- i. Soit $n \geq 0$, $x_n \in \mathbb{R}^N$ connu, $M_n = M(x_n) \in \mathcal{M}_{P,N}(\mathbb{R})$, et $x \in \mathbb{R}^N$. On pose $h = x - x_n$. Montrer que

$$J_n(x) = J(x_n) + M_n^t C M_n h \cdot h + 2M_n^t C(f(x_n) - d) \cdot h.$$

- ii. Montrer que la recherche du minimum de J_n est équivalente à la résolution d'un système linéaire dont on donnera l'expression.
- iii. Ecrire l'algorithme de quasi-linéarisation, et le comparer avec l'algorithme de Gauss-Newton.

Exercice 81 (Méthode de pénalisation)

Soit f une fonction continue et strictement convexe de \mathbb{R}^N dans \mathbb{R} , satisfaisant de plus :

$$\lim_{|x| \rightarrow +\infty} f(x) = +\infty.$$

Soit K un sous ensemble non vide, convexe (c'est-à-dire tel que $\forall(x, y) \in K^2$, $tx + (1-t)y \in K$, $\forall t \in]0, 1[$), et fermé de \mathbb{R}^N . Soit ψ une fonction continue de \mathbb{R}^N dans $[0, +\infty[$ telle que $\psi(x) = 0$ si et seulement si $x \in K$. Pour $n \in \mathbb{N}$, on définit la fonction f_n par $f_n(x) = f(x) + n\psi(x)$.

- Montrer qu'il existe au moins un élément $\bar{x}_n \in \mathbb{R}^N$ tel que $f_n(\bar{x}_n) = \inf_{x \in \mathbb{R}^N} f_n(x)$, et qu'il existe un unique élément $\bar{x}_K \in K$ tel que $f(\bar{x}_K) = \inf_{x \in K} f(x)$.
 - Montrer que pour tout $n \in \mathbb{N}$,
- $$f(\bar{x}_n) \leq f_n(\bar{x}_n) \leq f(\bar{x}_K).$$
- En déduire qu'il existe une sous-suite $(\bar{x}_{n_k})_{k \in \mathbb{N}}$ et $y \in K$ tels que $\bar{x}_{n_k} \rightarrow y$ lorsque $k \rightarrow +\infty$.
 - Montrer que $y = \bar{x}_K$. En déduire que toute la suite $(\bar{x}_n)_{n \in \mathbb{N}}$ converge vers \bar{x}_K .
 - Déduire de ces questions un algorithme (dit "de pénalisation") de résolution du problème de minimisation suivant :

$$\begin{cases} \text{Trouver } \bar{x}_K \in K; \\ f(\bar{x}_K) \leq f(x), \forall x \in K, \end{cases}$$

en donnant un exemple de fonction ψ .

Exercice 82 (Sur l'existence et l'unicité) Corrigé en page 164

Etudier l'existence et l'unicité des solutions du problème (3.29), avec les données suivantes : $E = \mathbb{R}$, $f : \mathbb{R} \rightarrow \mathbb{R}$ est définie par $f(x) = x^2$, et pour les quatre différents ensembles K suivants :

$$\begin{aligned} (i) \quad K &= \{|x| \leq 1\}; & (ii) \quad K &= \{|x| = 1\} \\ (iii) \quad K &= \{|x| \geq 1\}; & (iv) \quad K &= \{|x| > 1\}. \end{aligned} \tag{3.50}$$

Exercice 83 (Aire maximale d'un rectangle à périmètre donné)

Corrigé en page 164

1. On cherche à maximiser l'aire d'un rectangle de périmètre donné égal à 2. Montrer que ce problème peut se formuler comme un problème de minimisation de la forme (3.29), où K est de la forme $K = \{x \in \mathbb{R}^2; g(x) = 0\}$. On donnera f et g de manière explicite.

2. Montrer que le problème de minimisation ainsi obtenu est équivalent au problème

$$\begin{cases} \bar{x} = (\bar{x}_1, \bar{x}_2)^t \in \tilde{K} \\ f(\bar{x}_1, \bar{x}_2) \leq f(x_1, x_2), \quad \forall (x_1, x_2)^t \in \tilde{K}, \end{cases} \quad (3.51)$$

où $\tilde{K} = K \cap [0, 1]^2$, K et f étant obtenus à la question 1. En déduire que le problème de minimisation de l'aire admet au moins une solution.

3. Calculer $Dg(x)$ pour $x \in K$ et en déduire que si x est solution de (3.51) alors $x = (1/2, 1/2)$. En déduire que le problème (3.51) admet une unique solution donnée par $\bar{x} = (1/2, 1/2)$.

Exercice 84 (Fonctionnelle quadratique) *Suggestions en page 150, corrigé en page 165*

Soit f une fonction quadratique, i.e. $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$, où $A \in \mathcal{M}_N(\mathbb{R})$ est une matrice symétrique définie positive et $b \in \mathbb{R}^N$. On suppose que la contrainte g est une fonction linéaire de \mathbb{R}^N dans \mathbb{R} , c'est-à-dire $g(x) = d \cdot x - c$ où $c \in \mathbb{R}$ et $d \in \mathbb{R}^N$, et que $d \neq 0$. On pose $K = \{x \in \mathbb{R}^N, g(x) = 0\}$ et on cherche à résoudre le problème de minimisation (3.29).

1. Montrer que l'ensemble K est non vide, fermé et convexe. En déduire que le problème (3.29) admet une unique solution.

2. Montrer que si \bar{x} est solution de (3.29), alors il existe $\lambda \in \mathbb{R}$ tel que $y = (\bar{x}, \lambda)^t$ soit l'unique solution du système :

$$\left[\begin{array}{c|c} A & d \\ \hline d^t & 0 \end{array} \right] \begin{bmatrix} \bar{x} \\ \lambda \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix} \quad (3.52)$$

Exercice 85 (Utilisation du théorème de Lagrange)

- Pour $(x, y) \in \mathbb{R}^2$, on pose : $f(x, y) = -y$, $g(x, y) = x^2 + y^2 - 1$. Chercher le(s) point(s) où f atteint son maximum ou son minimum sous la contrainte $g = 0$.
- Soit $a = (a_1, \dots, a_N) \in \mathbb{R}^N$, $a \neq 0$. Pour $x = (x_1, \dots, x_N) \in \mathbb{R}^N$, on pose : $f(x) = \sum_{i=1}^N |x_i - a_i|^2$, $g(x) = \sum_{i=1}^N |x_i|^2$. Chercher le(s) point(s) où f atteint son maximum ou son minimum sous la contrainte $g = 1$.
- Soient $A \in \mathcal{M}_N(\mathbb{R})$ symétrique, $B \in \mathcal{M}_N(\mathbb{R})$ s.d.p. et $b \in \mathbb{R}^N$. Pour $v \in \mathbb{R}^N$, on pose $f(v) = (1/2)Av \cdot v - b \cdot v$ et $g(v) = Bv \cdot v$. Peut-on appliquer le théorème de Lagrange et quelle condition donne-t-il sur u si $f(u) = \min\{f(v), v \in K\}$ avec $K = \{v \in \mathbb{R}^N; g(v) = 1\}$?

Exercice 86 (Minimisation sans dérivabilité)

Soient $A \in \mathcal{M}_N(\mathbb{R})$ une matrice s.d.p., $b \in \mathbb{R}^N$, $j : \mathbb{R}^N \rightarrow \mathbb{R}$ une fonction continue, convexe, à valeurs positives ou nulles (mais non nécessairement dérivable, par exemple $j(v) = \sum_{j=1}^N \alpha_j |v_j|$, avec $\alpha_i \geq 0$ pour tout $i \in \{1, \dots, N\}$). Soit U une partie non vide, fermée convexe de \mathbb{R}^N . Pour $v \in \mathbb{R}^N$, on pose $J(v) = (1/2)Av \cdot v - b \cdot v + j(v)$.

1. Montrer qu'il existe un et un seul u tel que :

$$u \in U, J(u) \leq J(v), \quad \forall v \in U. \quad (3.53)$$

2. Soit $u \in U$, montrer que u est solution de (3.53) si et seulement si $(Au - b) \cdot (v - u) + j(v) - j(u) \geq 0$, pour tout $v \in U$.

Exercice 87 (Contre exemple aux multiplicateurs de Lagrange)

Soient f et $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, définies par : $f(x, y) = y$, et $g(x, y) = y^3 - x^2$. On pose $K = \{(x, y) \in \mathbb{R}^2; g(x, y) = 0\}$.

1. Calculer le minimum de f sur K et le point (\bar{x}, \bar{y}) où ce minimum est atteint.
2. Existe-t-il λ tel que $Df(\bar{x}, \bar{y}) = \lambda Dg(\bar{x}, \bar{y})$?
3. Pourquoi ne peut-on pas appliquer le théorème des multiplicateurs de Lagrange ?
4. Que trouve-t-on lorsqu'on applique la méthode dite "de Lagrange" pour trouver (\bar{x}, \bar{y}) ?

Exercice 88 (Application simple du théorème de Kuhn-Tucker) *Corrigé en page 166*

Soit f la fonction définie de $E = \mathbb{R}^2$ dans \mathbb{R} par $f(x) = x^2 + y^2$ et $K = \{(x, y) \in \mathbb{R}^2; x + y \geq 1\}$. Justifier l'existence et l'unicité de la solution du problème (3.29) et appliquer le théorème de Kuhn-Tucker pour la détermination de cette solution.

Exercice 89 (Exemple d'opérateur de projection)

Correction en page 166

1. Soit $K = C^+ = \{x \in \mathbb{R}^N, x = (x_1, \dots, x_n)^t, x_i \geq 0, \forall i = 1, \dots, N\}$.

(a) Montrer que K est un convexe fermé non vide.

(b) Montrer que pour tout $y \in \mathbb{R}^N$, on a : $(p_K(y))_i = \max(y_i, 0)$.

2. Soit $(\alpha_i)_{i=1, \dots, N} \subset \mathbb{R}^N$ et $(\beta_i)_{i=1, \dots, N} \subset \mathbb{R}^N$ tels que $\alpha_i \leq \beta_i$ pour tout $i = 1, \dots, N$. Soit $K = \{x = (x_1, \dots, x_N)^t; \alpha_i \leq \beta_i, i = 1, \dots, N\}$.

1. Montrer que K est un convexe fermé non vide.

2. Soit p_K l'opérateur de projection définie à la proposition 3.44 page 133. Montrer que pour tout $y \in \mathbb{R}^N$, on a :

$$(p_K(y))_i = \max(\alpha_i, \min(y_i, \beta_i)), \quad \forall i = 1, \dots, N$$

Exercice 90 (Méthode de relaxation avec Newton problèmes sans contrainte)

On considère le problème :

$$\begin{cases} \bar{x} \in K, \\ f(\bar{x}) \leq f(x), \forall x \in K, \end{cases} \quad (3.54)$$

où $K \subset \mathbb{R}^N$.

(a) On prend ici $K = \prod_{i=1, N} [a_i, b_i]$, où $(a_i, b_i) \in \mathbb{R}^2$ est tel que $a_i \leq b_i$. On considère l'algorithme suivant :

$$\left\{ \begin{array}{l} \text{Initialisation : } x^{(0)} \in E, \\ \text{Itération } n : \quad x^{(n)} \text{ connu, } (n \geq 0) \\ \quad \text{Calculer } x_1^{(n+1)} \in [a_1, b_1] \text{ tel que :} \\ \quad f(x_1^{(n+1)}, x_2^{(n)}, x_3^{(n)}, \dots, x_N^{(n)}) \leq f(\xi, x_2^{(n)}, x_3^{(n)}, \dots, x_N^{(n)}), \text{ pour tout } \xi \in [a_1, b_1], \\ \quad \text{Calculer } x_2^{(n+1)} \in [a_2, b_2] \text{ tel que :} \\ \quad f(x_1^{(n+1)}, x_2^{(n+1)}, x_3^{(n)}, \dots, x_N^{(n)}) \leq f(x_1^{(n+1)}, \xi, x_3^{(n)}, \dots, x_N^{(n)}), \\ \quad \quad \text{pour tout } \xi \in [a_2, b_2], \\ \quad \quad \quad \dots \\ \quad \quad \quad \dots \\ \quad \text{Calculer } x_k^{(n+1)} \in [a_k, b_k], \text{ tel que :} \\ \quad f(x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, x_k^{(n+1)}, x_{k+1}^{(n)}, \dots, x_N^{(n)}) \\ \quad \quad \leq f(x_1^{(n+1)}, \dots, x_{k-1}^{(n+1)}, \xi, x_{k+1}^{(n)}, \dots, x_N^{(n)}), \text{ pour tout } \xi \in [a_k, b_k], \\ \quad \quad \quad \dots \\ \quad \quad \quad \dots \\ \quad \text{Calculer } x_N^{(n+1)} \in [a_N, b_N] \text{ tel que :} \\ \quad f(x_1^{(n+1)}, x_2^{(n+1)}, \dots, x_{N-1}^{(n+1)}, x_N^{(n+1)}) \leq f(x_1^{(n+1)}, \dots, x_{N-1}^{(n+1)}, \xi), \\ \quad \quad \text{pour tout } \xi \in [a_N, b_N]. \end{array} \right. \quad (3.55)$$

Montrer que la suite $x^{(n)}$ construite par l'algorithme (3.55) est bien définie et converge vers \bar{x} lorsque n tend vers $+\infty$, où $\bar{x} \in K$ est tel que $f(\bar{x}) \leq f(x)$ pour tout $x \in K$.

- (b) On prend maintenant $N = 2$, f la fonction de \mathbb{R}^2 dans \mathbb{R} définie par $f(x) = x_1^2 + x_2^2$, et $K = \{(x_1, x_2)^t \in \mathbb{R}^2; x_1 + x_2 \geq 2\}$. Montrer qu'il existe un unique élément $\bar{x} = (\bar{x}_1, \bar{x}_2)^t$ de K tel que $f(\bar{x}) = \inf_{x \in \mathbb{R}^2} f(x)$. Déterminer \bar{x} .

On considère l'algorithme suivant pour la recherche de \bar{x} :

$$\left\{ \begin{array}{l} \text{Initialisation : } x^{(0)} \in E, \\ \text{Itération } n : \quad x^{(n)} \text{ connu, } (n \geq 0) \\ \quad \text{Calculer } x_1^{(n+1)} \geq 2 - x_2^{(n)} \text{ tel que :} \\ \quad f(x_1^{(n+1)}, x_2^{(n)}) \leq f(\xi, x_2^{(n)}), \text{ pour tout } \xi \geq 2 - x_2^{(n)}, \\ \quad \text{Calculer } x_2^{(n+1)} \geq 2 - x_1^{(n)} \text{ tel que :} \\ \quad f(x_1^{(n+1)}, x_2^{(n+1)}) \leq f(x_1^{(n+1)}, \xi), \text{ pour tout } \xi \geq 2 - x_1^{(n)}. \end{array} \right. \quad (3.56)$$

Montrer (éventuellement graphiquement) que la suite construite par l'algorithme ci-dessus ne converge vers \bar{x} que si l'une des composantes de $x^{(0)}$ vaut 1.

Exercice 91 (Convergence de l'algorithme d'Uzawa)

Soient $N, p \in \mathbb{N}^*$. Soit $f \in C^1(\mathbb{R}^N, \mathbb{R})$ ($N \geq 1$) t.q.

$$\exists \alpha > 0, (\nabla f(x) - \nabla f(y)) \cdot (x - y) \geq \alpha |x - y|^2, \forall x, y \in \mathbb{R}^N.$$

Soit $C \in M_{p,N}(\mathbb{R})$ (C est donc une matrice, à éléments réels, ayant p lignes et N colonnes) et $d \in \mathbb{R}^p$. On note $D = \{x \in \mathbb{R}^N, Cx \leq d\}$ et $\mathcal{C}^+ = \{u \in \mathbb{R}^p, u \geq 0\}$.

On suppose $D \neq \emptyset$ et on s'intéresse au problème suivant :

$$x \in D, f(x) \leq f(y), \forall y \in D. \quad (3.57)$$

1. Montrer que $f(y) \geq f(x) + \nabla f(x) \cdot (y - x) + \frac{\alpha}{2} |x - y|^2$ pour tout $x, y \in \mathbb{R}^N$.
2. Montrer que f est strictement convexe et que $f(x) \rightarrow \infty$ quand $|x| \rightarrow \infty$. En déduire qu'il existe une et une seule solution au problème (3.57).

Dans la suite, on note \bar{x} cette solution.

Pour $u \in \mathbb{R}^p$ et $x \in \mathbb{R}^N$, on pose $L(x, u) = f(x) + u \cdot (Cx - d)$.

3. Soit $u \in \mathbb{R}^p$ (dans cette question, u est fixé). Montrer que l'application $x \rightarrow L(x, u)$ est strictement convexe (de \mathbb{R}^N dans \mathbb{R}) et que $L(x, u) \rightarrow \infty$ quand $|x| \rightarrow \infty$ [Utiliser la question 1]. En déduire qu'il existe une et une seule solution au problème suivant :

$$x \in \mathbb{R}^N, L(x, u) \leq L(y, u), \forall y \in \mathbb{R}^N. \quad (3.58)$$

Dans la suite, on note x_u cette solution. Montrer que x_u est aussi l'unique élément de \mathbb{R}^N t.q. $\nabla f(x_u) + C^t u = 0$.

4. On admet que le théorème de Kuhn-Tucker s'applique ici (cf. cours). Il existe donc $\bar{u} \in \mathcal{C}^+$ t.q. $\nabla f(\bar{x}) + C^t \bar{u} = 0$ et $\bar{u} \cdot (C\bar{x} - d) = 0$. Montrer que (\bar{x}, \bar{u}) est un point selle de L sur $\mathbb{R}^N \times \mathcal{C}^+$, c'est-à-dire :

$$L(\bar{x}, v) \leq L(\bar{x}, \bar{u}) \leq L(y, \bar{u}), \forall (y, v) \in \mathbb{R}^N \times \mathcal{C}^+. \quad (3.59)$$

Pour $u \in \mathbb{R}^p$, on pose $M(u) = L(x_u, u)$ (de sorte que $M(u) = \inf\{L(x, u), x \in \mathbb{R}^N\}$). On considère alors le problème suivant :

$$u \in \mathcal{C}^+, M(u) \geq M(v), \forall v \in \mathcal{C}^+. \quad (3.60)$$

5. Soit $(x, u) \in \mathbb{R}^N \times \mathcal{C}^+$ un point selle de L sur $\mathbb{R}^N \times \mathcal{C}^+$ (c'est-à-dire $L(x, v) \leq L(x, u) \leq L(y, u)$, pour tout $(y, v) \in \mathbb{R}^N \times \mathcal{C}^+$). Montrer que $x = \bar{x} = x_u$ (on rappelle que \bar{x} est l'unique solution de (3.57) et x_u est l'unique solution de (3.58)) et que u est solution de (3.60). [On pourra commencer par montrer, en utilisant la première inégalité, que $x \in D$ et $u \cdot (Cx - d) = 0$.]

Montrer que $\nabla f(\bar{x}) + C^t u = 0$ et que $u = P_{C^+}(u + \rho(C\bar{x} - d))$, pour tout $\rho > 0$, où P_{C^+} désigne l'opérateur de projection orthogonale sur C^+ . [on rappelle que si $v \in \mathbb{R}^p$ et $w \in C^+$, on a $w = P_{C^+} v \iff ((v - w) \cdot (w - z) \geq 0, \forall z \in C^+)$.]

6. Dédurre des questions 2, 4 et 5 que le problème (3.60) admet au moins une solution.
7. Montrer que l'algorithme du gradient à pas fixe avec projection pour trouver la solution de (3.60) s'écrit (on désigne par $\rho > 0$ le pas de l'algorithme) :

Initialisation. $u_0 \in C^+$.

Itérations. Pour $u_k \in C^+$ connu ($k \geq 0$). On calcule $x_k \in \mathbb{R}^N$ t.q. $\nabla f(x_k) + C^t u_k = 0$ (montrer qu'un tel x_k existe et est unique) et on pose $u_{k+1} = P_{C^+}(u_k + \rho(Cx_k - d))$.

Dans la suite, on s'intéresse à la convergence de la suite $(x_k, u_k)_{k \in \mathbb{N}}$ donnée par cet algorithme.

8. Soit ρ t.q. $0 < \rho < 2\alpha/\|C\|^2$ avec $\|C\| = \sup\{|Cx|, x \in \mathbb{R}^N \text{ t.q. } |x| = 1\}$. Soit $(\bar{x}, \bar{u}) \in \mathbb{R}^N \times C^+$ un point selle de L sur $\mathbb{R}^N \times C^+$ (c'est-à-dire vérifiant (3.59)) et $(x_k, u_k)_{k \in \mathbb{N}}$ la suite donnée par l'algorithme de la question précédente. Montrer que

$$|u_{k+1} - \bar{u}|^2 \leq |u_k - \bar{u}|^2 - \rho(2\alpha - \rho\|C\|^2)|x_k - \bar{x}|^2, \forall k \in \mathbb{N}.$$

En déduire que $x_k \rightarrow \bar{x}$ quand $k \rightarrow \infty$.

Montrer que la suite $(u_k)_{k \in \mathbb{N}}$ est bornée et que, si \tilde{u} est une valeur d'adhérence de la suite $(u_k)_{k \in \mathbb{N}}$, on a $\nabla f(\bar{x}) + C^t \tilde{u} = 0$. En déduire que, si $\text{rang}(C) = p$, on a $u_k \rightarrow \bar{u}$ quand $k \rightarrow \infty$ et que \bar{u} est l'unique élément de C^+ t.q. $\nabla f(\bar{x}) + C^t \bar{u} = 0$.

3.7 Suggestions

Exercice 66 page 138 (Convexité et continuité)

- (a) Pour montrer la continuité en 0, soit $x \neq 0, |x| < 1$. On pose $a = \text{sgn}(x) (= \frac{x}{|x|})$. Ecrire x comme une combinaison convexe de 0 et a et écrire 0 comme une combinaison convexe de x et $-a$. En déduire une majoration de $|f(x) - f(0)|$.
(b) utiliser la continuité de f et la majoration précédente.
- (a) Faire une récurrence sur N et pour $x = (x_1, y)^t$ avec $-R < x_1 < R$ et $y \in \mathbb{R}^{N-1}$ ($N > 1$), majorer $f(x)$ en utilisant $f(+R, y)$ et $f(-R, y)$.
(b) Reprendre le raisonnement fait pour $N = 1$.
(c) Se ramener à $E = \mathbb{R}^N$.
- (a) reprendre le raisonnement fait pour $E = \mathbb{R}$.
(b) On pourra, par exemple choisir $E = C([0, 1], \mathbb{R}) \dots$

Exercice 67 page 138 (Maximisation)

Appliquer les théorèmes du cours à $-f$.

Exercice 68 page 138 (Minimisation d'une fonctionnelle quadratique)

- Calculer la différentielle de f en formant la différence $f(x+h) - f(x)$ et en utilisant la définition. Calculer la hessienne en formant la différence $\nabla f(x+h) - \nabla f(x)$.
- Utiliser le cours...

Exercice 70 page 139 (Algorithme du gradient à pas fixe)

- Introduire la fonction φ définie (comme d'habitude...) par $\varphi(t) = f(tx + (1-t)y)$, intégrer entre 0 et 1 et utiliser l'hypothèse (3.3.15) sur $\nabla f(x + t(y-x)) - \nabla f(x)$.

2. Utiliser le cours pour la stricte convexité et l'existence et l'unicité de \bar{x} , et la question 1 pour montrer que $f(x) \rightarrow +\infty$ lorsque $|x| \rightarrow +\infty$.
3. Montrer grâce aux hypothèses (3.3.15) et (3.3.16) que $|x_{n+1} - \bar{x}|^2 < |x_n - \bar{x}|^2(1 - 2\alpha\rho + M^2\rho^2)$.

Exercice 72 page 139 (Algorithme du gradient à pas optimal)

2. Utiliser le fait que H est continue.
3. Etudier la fonction $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par $\varphi(\rho) = f(x_n + \rho w_n)$.
4. a. Montrer que f est minorée et remarquer que la suite $(f(x_n))_{n \in \mathbb{N}}$ est décroissante.
4.b se déduit du 4.a
4.c. Utiliser la fonction φ définie plus haut, la question 4.b. et la question 2.
4.d. Utiliser le fait que le choix de ρ_n est optimal et le résultat de 4.c.
4.e. Etudier le polynôme du 2nd degré en ρ défini par : $P_n(\rho) = f(x_n) - \rho|w_n|^2 + \frac{1}{2}M|w_n|^2\rho^2$ dans les cas où $|w_n| \leq M$ (fait à la question 4.c) puis dans le cas $|w_n| \geq M$.
5. utiliser l'inégalité prouvée en 4.e. pour montrer que $|w_n| \rightarrow 0$ lorsque $n \rightarrow +\infty$.
6. Pour montrer que toute la suite converge, utiliser l'argument d'unicité de la limite, en raisonnant par l'absurde (supposer que la suite ne converge pas et aboutir à une contradiction).

Exercice 74 page 140 (Cas où f n'est pas croissante à l'infini)

S'inspirer des techniques utilisées aux exercices 70 et 72 (il faut impérativement les avoir fait avant...).

Exercice 78 page 143 (Méthode de Polak-Ribière)

1. Utiliser la deuxième caractérisation de la convexité. Pour montrer le comportement à l'infini, introduire la fonction φ habituelle... ($\varphi(t) = f(x + ty)$).
2. Pour montrer la concurrence, utiliser le fait que si $w_n \cdot \nabla f(x_n) < 0$ alors w_n est une direction de descente stricte de f en x_n , et que si ρ_n est optimal alors $\nabla f(x_n + \rho_n w_n) = 0$.
3. Utiliser la fonction φ définie par $\varphi(\theta) = \nabla f(x_n + \theta \rho_n w_n)$.
4. C'est du calcul...
5. Montrer d'abord que $-g_n w_n \leq -\gamma|w_n||g_n|$. Montrer ensuite (en utilisant la bonne vieille fonction φ définie par $\varphi(t) = f(x_n + t\rho_n)$), que $g_n \rightarrow 0$ lorsque $n \rightarrow +\infty$.

Exercice 84 page 146 (Fonctionnelle quadratique)

1. Pour montrer que K est non vide, remarquer que comme $d \neq 0$, il existe $\tilde{x} \in \mathbb{R}^N$ tel que $d \cdot \tilde{x} = \alpha \neq 0$. En déduire l'existence de $x \in \mathbb{R}^N$ tel que $d \cdot x = c$.
2. Montrer par le théorème de Lagrange que si \bar{x} est solution de (3.29), alors $y = (\bar{x}, \lambda)^t$ est solution du système (3.52), et montrer ensuite que le système (3.52) admet une unique solution.

3.8 Corrigés

Corrigé de l'exercice 68 page 138 (Minimisation d'une fonctionnelle quadratique)

1. Puisque $f(x) = \frac{1}{2}Ax \cdot x - b \cdot x$, $f \in C^\infty(\mathbb{R}^N, \mathbb{R})$. Calculons le gradient de f :

$$\begin{aligned} f(x+h) &= \frac{1}{2}A(x+h) \cdot (x+h) - b \cdot (x+h) \\ &= \frac{1}{2}Ax \cdot x + \frac{1}{2}Ax \cdot h + \frac{1}{2}Ah \cdot x + \frac{1}{2}Ah \cdot h - b \cdot x - b \cdot h \\ &= f(x) + \frac{1}{2}(Ax \cdot h + Ah \cdot x) - b \cdot h + \frac{1}{2}Ah \cdot h \\ &= f(x) + \frac{1}{2}(Ax + A^t x) \cdot h - b \cdot h + \frac{1}{2}Ah \cdot h. \end{aligned}$$

Et comme $\|Ah \cdot h\| \leq \|A\|_2 \|h\|^2$, on a :

$$\nabla f(x) = \frac{1}{2}(Ax + A^t x) - b. \quad (3.61)$$

Si A est symétrique $\nabla f(x) = Ax - b$. Calculons maintenant la hessienne de f . D'après (3.61), on a :

$$\nabla f(x+h) = \frac{1}{2}(A(x+h) + A^t(x+h)) - b = \nabla f(x) + \frac{1}{2}(Ah + A^t h)$$

et donc $H_f(x) = D(\nabla f(x)) = \frac{1}{2}(A + A^t)$. On en déduit que si A est symétrique, $H_f(x) = A$.

2. Si A est symétrique définie positive, alors f est strictement convexe. De plus, si A est symétrique définie positive, alors $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$. En effet,

$$\begin{aligned} Ah \cdot h &\geq \alpha|h|^2 \text{ où } \alpha \text{ est la plus petite valeur propre de } A, \text{ et } \alpha > 0 \\ f(h) &\geq \frac{\alpha}{2}\|h\|^2 - \|b\|\|h\|; \text{ or } \|bh\| \leq \|b\| \|h\| \\ f(h) &\geq \|h\| \left(\frac{\alpha\|h\|}{2} - b \right) \longrightarrow \infty \text{ quand } h \rightarrow +\infty \end{aligned}$$

On en déduit l'existence et l'unicité de \bar{x} qui minimise f . On a aussi :

$$\nabla f(\bar{x}) = 0 \Leftrightarrow f(\bar{x}) = \inf_{\mathbb{R}^N} f$$

Par la question 1. \bar{x} est donc l'unique solution du système $A\bar{x} = b$.

Corrigé de l'exercice 70 page 139 (Convergence de l'algorithme du gradient à pas fixe)

1. Soit φ la fonction définie de \mathbb{R} dans \mathbb{R}^N par : $\varphi(t) = f(x + t(y-x))$. Alors $\varphi(1) - \varphi(0) = \int_0^1 \nabla f(x + t(y-x)) \cdot (y-x) dt$, et donc :

$$f(y) - f(x) = \int_0^1 \nabla f(x + t(y-x)) \cdot (y-x) dt.$$

On a donc :

$$f(y) - f(x) - \nabla f(x) \cdot (y-x) = \int_0^1 (\nabla f(x + t(y-x)) \cdot (y-x) - \nabla f(x) \cdot (y-x)) dt,$$

c'est-à-dire :

$$f(y) - f(x) - \nabla f(x) \cdot (y-x) = \int_0^1 \underbrace{(\nabla f(x + t(y-x)) - \nabla f(x)) \cdot (y-x)}_{\geq \alpha t(y-x)^2} dt.$$

Grâce à la première hypothèse sur f , ceci entraîne :

$$f(y) - f(x) - \nabla f(x) \cdot (y-x) \geq \alpha \int_0^1 t|y-x|^2 dt = \frac{\alpha}{2}|y-x|^2 > 0 \text{ si } y \neq x. \quad (3.62)$$

2. On déduit de la question 1 que f est strictement convexe. En effet, grâce à la question 1, pour tout $(x, y) \in E^2$, $f(y) > f(x) + \nabla f(x) \cdot (y - x)$; et d'après la première caractérisation de la convexité, voir proposition 3.11 p.47, on en déduit que f est strictement convexe.

Montrons maintenant que $f(y) \rightarrow +\infty$ quand $|y| \rightarrow +\infty$.

On écrit (3.62) pour $x = 0$: $f(y) \geq f(0) + \nabla f(0) \cdot y + \frac{\alpha}{2}|y|^2$.

Comme $\nabla f(0) \cdot y \geq -|\nabla f(0)| |y|$, on a donc

$$f(y) \geq f(0) - |\nabla f(0)| |y| + \frac{\alpha}{2}|y|^2, \text{ et donc :}$$

$$f(y) \geq f(0) + |y| \left(\frac{\alpha}{2}|y| - |\nabla f(0)| \right) \rightarrow +\infty \text{ quand } |y| \rightarrow +\infty.$$

3. On pose $h(x) = x - \rho \nabla f(x)$. L'algorithme du gradient à pas fixe est un algorithme de point fixe pour h .

$$x_{n+1} = x_n - \rho \nabla f(x_n) = h(x_n).$$

Grâce au théorème 2.3 page 77, on sait que h est strictement contractante si $0 < \rho < \frac{2\alpha}{M^2}$.

Donc $x_n \rightarrow \bar{x}$ unique point fixe de h , c'est-à-dire $\bar{x} = h(\bar{x}) = \bar{x} - \rho \nabla f(\bar{x})$. Ceci entraîne

$$\nabla f(\bar{x}) = 0 \text{ donc } f(\bar{x}) = \inf_E f \text{ car } f \text{ est convexe.}$$

Corrigé de l'exercice 71 page 139 (Mise en oeuvre de GPF et GPO)

1. On a

$$\nabla f(x) = \begin{pmatrix} 4x_1 - x_2 - 3 \\ 2x_2 - x_1 - 1 \end{pmatrix} \text{ et } H_f = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$$

La fonction f vérifie les hypothèses du théorème 3.34 d'existence et d'unicité du minimum. En particulier la hessienne $H_f = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$ est s.d.p., car $H_f x \cdot x = (4x_1 - x_2)x_1 + (-x_1 + 2x_2)x_2 = (x_1 - x_2)^2 + 3x_1^2 + x_2^2 > 0$ sauf pour $x_1 = x_2 = 0$. Le minimum est obtenu pour

$$\partial_1 f(x_1, x_2) = 4x_1 - x_2 - 3 = 0$$

$$\partial_2 f(x_1, x_2) = 2x_2 - x_1 - 1 = 0$$

c'est-à-dire $\bar{x}_1 = 1$ et $\bar{x}_2 = 1$. Ce minimum est $f(\bar{x}_1, \bar{x}_2) = 2$.

2. L'algorithme du gradient à pas fixe s'écrit :

$$\left\{ \begin{array}{l} \text{Initialisation : } x^{(0)} \in E, \\ \text{Itération } n : \quad x^{(n)} \text{ connu, } (n \geq 0) \\ \quad w^{(n)} = -\nabla f(x^{(n)}), \\ \quad x^{(n+1)} = x^{(n)} + \rho w^{(n)}. \end{array} \right.$$

A la première itération, on a $\nabla f(0, 0) = (-3, -1)$ et donc $w_0 = (3, 1)$. On en déduit $x^{(1)} = (3\rho, 2\rho) = (3/2, 1)$ et $f(x^{(1)}) = 5/2$.

L'algorithme du gradient à pas optimal s'écrit :

$$\left\{ \begin{array}{l} \text{Initialisation : } x^{(0)} \in \mathbb{R}^N. \\ \text{Itération } n : \quad x^{(n)} \text{ connu.} \\ \quad \text{On calcule } w^{(n)} = -\nabla f(x^{(n)}). \\ \quad \text{On choisit } \rho_n \geq 0 \text{ tel que} \\ \quad \quad f(x^{(n)} + \rho_n w^{(n)}) \leq f(x^{(n)} + \rho w^{(n)}) \quad \forall \rho \geq 0. \\ \quad \text{On pose } x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}. \end{array} \right.$$

Calculons le ρ_0 optimal à l'itération 0. On a vu précédemment que $w_0 = (3, 2)$. Le ρ_0 optimal minimise la fonction $\rho \mapsto \varphi(\rho) = f(x^{(0)} + \rho w^{(0)}) = f(3\rho, 2\rho)$. On doit donc avoir $\varphi'(\rho_0) = 0$. Calculons $\varphi'(\rho)$. Par le théorème de dérivation des fonctions composées, on a :

$$\varphi'(\rho) = \nabla f(x^{(0)} + \rho w^{(0)}) \cdot w(0) = \begin{bmatrix} 10\rho - 3 \\ \rho - 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 3(10\rho - 3) + (\rho - 1) = 32\rho - 10.$$

On en déduit que $\rho_0 = \frac{5}{16}$. On obtient alors $x^{(1)} = x^{(0)} + \rho_0 w^{(0)} = (\frac{15}{16}, \frac{5}{16})$, et $f(x^{(1)}) = 2.4375$, ce qui est, comme attendu, mieux qu'avec GPF.

Corrigé de l'exercice 72 page 139 (Convergence de l'algorithme du gradient à pas optimal)

1. On sait que $f(x) \rightarrow +\infty$ lorsque $|x| \rightarrow +\infty$. Donc $\forall A > 0, \exists R \in \mathbb{R}_+; |x| > R \Rightarrow f(x) > A$. En particulier pour $A = f(x_0)$ ceci entraîne :

$$\exists R \in \mathbb{R}_+; x \in B_R \Rightarrow f(x) > f(x_0).$$

2. Comme $f \in C^2(\mathbb{R}^N, \mathbb{R})$, sa hessienne H est continue, donc $\|H\|$ atteint son max sur B_{R+1} qui est un fermé borné de \mathbb{R}^N . Soit $M = \max_{x \in B_{R+1}} \|H(x)\|$, on a $H(x)y \cdot y \leq My \cdot y \leq M|y|^2$.
3. Soit $w_n = -\nabla f(x_n)$.
Si $w_n = 0$, on pose $x_{n+1} = x_n$.
Si $w_n \neq 0$, montrons qu'il existe $\bar{\rho} > 0$ tel que

$$f(x_n + \bar{\rho}w_n) \leq f(x_n + \rho w_n) \quad \forall \rho > 0.$$

On sait que $f(x) \rightarrow +\infty$ lorsque $|x| \rightarrow +\infty$.

Soit $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par $\varphi(\rho) = f(x_n + \rho w_n)$. On a $\varphi(0) = f(x_n)$ et $\varphi(\rho) = f(x_n + \rho w_n) \rightarrow +\infty$ lorsque $\rho \rightarrow +\infty$.

En effet si $\rho \rightarrow +\infty$, on a $|x_n + \rho w_n| \rightarrow +\infty$. Donc φ étant continue, φ admet un minimum, atteint en $\bar{\rho}$, et donc $\exists \bar{\rho} \in \mathbb{R}_+; f(x_n + \bar{\rho}w) \leq f(x_n + \rho w_n) \quad \forall \rho > 0$.

4. a) Montrons que la suite $(f(x_n))_{n \in \mathbb{N}}$ est convergente. La suite $(f(x_n))_{n \in \mathbb{N}}$ vérifie

$$f(x_{n+1}) \leq f(x_n).$$

De plus $f(x) \rightarrow +\infty$ lorsque $|x| \rightarrow +\infty$ donc f est bornée inférieurement. On en conclut que la suite $(f(x_n))_{n \in \mathbb{N}}$ est convergente.

- b) Montrons que $x_n \in B_R \quad \forall n \in \mathbb{N}$. On sait que si $x \notin B_R$ alors $f(x) > f(x_0)$. Or la suite $(f(x_n))_{n \in \mathbb{N}}$ est décroissante donc $f(x_n) \leq f(x_0) \quad \forall n$, donc $x_n \in B_R, \quad \forall n \in \mathbb{N}$.
- c) Montrons que $f(x_n + \rho w_n) \leq f(x_n) - \rho|w_n|^2 + \frac{\rho^2}{2}M|w_n|^2, \quad \forall \rho \in [0, \frac{1}{|w_n|}]$. Soit φ définie de \mathbb{R}_+ dans \mathbb{R} par $\varphi(\rho) = f(x_n + \rho w_n)$. On a

$$\varphi(\rho) = \varphi(0) + \rho\varphi'(0) + \frac{\rho^2}{2}\varphi''(\bar{\rho}), \quad \text{où } \bar{\rho} \in]0, \rho[.$$

Or $\varphi'(\rho) = \nabla f(x_n + \rho w_n) \cdot w_n$ et $\varphi''(\rho) = H(x_n + \rho w_n)w_n \cdot w_n$. Donc

$$\varphi(\rho) = \underbrace{\varphi(0)}_0 + \rho \underbrace{\nabla f(x_n) \cdot w_n}_{-|w_n|^2} + \frac{\rho^2}{2}H(x_n + \bar{\rho}w_n)w_n \cdot w_n.$$

Si $\rho \in [0, \frac{1}{|w_n|}]$ on a

$$\begin{aligned} |x_n + \bar{\rho}w_n| &\leq |x_n| + \frac{1}{|w_n|}|w_n| \\ &\leq R + 1, \end{aligned}$$

donc $x_n + \bar{\rho}w_n \in B_{R+1}$ et par la question 2,

$$H(x_n + \bar{\rho}w_n)w_n \cdot w_n \leq M|w_n|^2.$$

On a donc bien

$$\varphi(\rho) = f(x_n + \rho w_n) \leq f(x_n) - \rho|w_n|^2 + \frac{\rho^2}{2}M|w_n|^2.$$

d) Montrons que $f(x_{n+1}) \leq f(x_n) - \frac{|w_n|^2}{2M}$ si $|w_n| \leq M$.

Comme le choix de ρ_n est optimal, on a

$$f(x_{n+1}) = f(x_n + \rho_n w_n) \leq f(x_n + \rho w_n), \quad \forall \rho \in \mathbb{R}_+.$$

donc en particulier

$$f(x_{n+1}) \leq f(x_n + \rho w_n), \quad \forall \rho \in [0, \frac{1}{|w_n|}].$$

En utilisant la question précédente, on obtient

$$f(x_{n+1}) \leq f(x_n) - \rho|w_n|^2 + \frac{\rho^2}{2}M|w_n|^2 = \varphi(\rho), \quad \forall \rho \in [0, \frac{1}{|w_n|}]. \quad (3.63)$$

Or la fonction φ atteint son minimum pour

$$-|w_n|^2 + \rho M|w_n|^2 = 0$$

c'est-à-dire $\rho M = 1$ ou encore $\rho = \frac{1}{M}$ ce qui est possible si $\frac{1}{|w_n|} \geq \frac{1}{M}$ (puisque 3.63 est vraie si $\rho \leq \frac{1}{|w_n|}$).

Comme on a supposé $|w_n| \leq M$, on a donc

$$f(x_{n+1}) \leq f(x_n) - \frac{|w_n|^2}{M} + \frac{|w_n|^2}{2M} = f(x_n) - \frac{|w_n|^2}{2M}.$$

e) Montrons que $-f(x_{n+1}) + f(x_n) \geq \frac{|w_n|^2}{2\bar{M}}$ où $\bar{M} = \sup(M, \tilde{M})$ avec $\tilde{M} = \sup\{|\nabla f(x)|, x \in C_R\}$.

On sait par la question précédente que si

$$|w_n| \leq M, \text{ on a } -f(x_{n+1}) - f(x_n) \geq \frac{|w_n|^2}{2M}.$$

Montrons que si $|w_n| \geq M$, alors $-f(x_{n+1}) + f(x_n) \geq \frac{|w_n|^2}{2\tilde{M}}$. On aura alors le résultat souhaité.

On a

$$f(x_{n+1}) \leq f(x_n) - \rho|w_n|^2 + \frac{\rho^2}{2}M|w_n|^2, \quad \forall \rho \in [0, \frac{1}{|w_n|}].$$

Donc

$$f(x_{n+1}) \leq \min_{[0, \frac{1}{|w_n|}]} \underbrace{[f(x_n) - \rho|w_n|^2 + \frac{\rho^2}{2}M|w_n|^2]}_{P_n(\rho)}$$

– 1er cas si $|w_n| \leq M$, on a calculé ce min à la question c).

– si $|w_n| \geq M$, la fonction $P_n(\rho)$ est décroissante sur $[0, \frac{1}{|w_n|}]$ et le minimum est donc atteint pour $\rho = \frac{1}{|w_n|}$.

$$\begin{aligned} \text{Or } P_n\left(\frac{1}{|w_n|}\right) &= f(x_n) - |w_n| + \frac{M}{2} \leq f(x_n) - \frac{|w_n|}{2} \\ &\leq f(x_n) - \frac{|w_n|^2}{2\tilde{M}}. \end{aligned}$$

5. Montrons que $\nabla f(x_n) \rightarrow 0$ lorsque $n \rightarrow +\infty$. On a montré que $\forall n, |w_n|^2 \leq 2\bar{M}(f(x_n) - f(x_{n+1}))$. Or la suite $(f(x_n))_{n \in \mathbb{N}}$ est convergente. Donc $|w_n| \rightarrow 0$ lorsque $n \rightarrow +\infty$ et $w_n = \nabla f(x_n)$ ce qui prouve le résultat.

La suite $(x_n)_{n \in \mathbb{N}}$ est bornée donc $\exists (n_k)_{k \in \mathbb{N}}$ et $\tilde{x} \in \mathbb{R}^N$; $x_{n_k} \rightarrow \tilde{x}$ lorsque $k \rightarrow +\infty$ et comme $\nabla f(x_{n_k}) \rightarrow 0$, on a, par continuité, $\nabla f(\tilde{x}) = 0$.

6. On suppose $\exists! \bar{x} \in \mathbb{R}^N$ tel que $\nabla f(\bar{x}) = 0$. Montrons que $f(\bar{x}) \leq f(x) \forall x \in \mathbb{R}^N$ et que $x_n \rightarrow \bar{x}$ quand $n \rightarrow +\infty$. Comme f est croissante à l'infini, il existe un point qui réalise un minimum de f , et on sait qu'en ce point le gradient s'annule ; en utilisant l'hypothèse d'unicité, on en déduit que ce point est forcément \bar{x} , et donc $f(\bar{x}) \leq f(x)$ pour tout $x \in \mathbb{R}^N$.

Montrons maintenant que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers \bar{x} . En raison de l'hypothèse d'unicité, on a forcément $\tilde{x} = \bar{x}$, et on sait qu'on a convergence d'une sous-suite de $(x_n)_{n \in \mathbb{N}}$ vers \bar{x} par la question 5. Il reste donc à montrer que c'est toute la suite qui converge. Supposons qu'elle ne converge pas ; alors

$$\exists \varepsilon > 0; \forall k \in \mathbb{N}, \exists n_k \geq k \text{ et } |x_{n_k} - \bar{x}| > \varepsilon \quad (3.64)$$

Mais d'après la question 5), on peut extraire de la suite $(x_{n_k})_{k \in \mathbb{N}}$ une sous-suite qui converge, ce qui contredit (3.64). Donc la suite $(x_n)_{n \in \mathbb{N}}$ converge.

Corrigé de l'exercice 75 page 140 (Méthode de relaxation)

1. On vu à l'exercice 70, questions 1 et 2, que si f vérifie l'hypothèse (3.44) alors f est strictement convexe et tend vers l'infini en l'infini, et donc il existe un unique $\bar{x} \in \mathbb{R}^N$ réalisant son minimum.

2. Ecrivons l'hypothèse (3.44) avec $x = se_k$ et $y = te_k$ où $(s, t) \in \mathbb{R}^2$ et e_k est le k -ième vecteur de la base canonique de \mathbb{R}^N ; en notant $\partial_k f$ la dérivée partielle de f par rapport à la k -ième variable, il vient :

$$(\partial_k f(s) - \partial_k f(t))(s - t) \geq \alpha |s - t|^2.$$

En appliquant à nouveau les résultats de l'exercice 70, questions 1 et 2 au cas $N = 1$, on en déduit l'existence et unicité de \bar{s} tel que

$$\varphi_k^{(n+1)}(\bar{s}) = \inf_{s \in \mathbb{R}} \varphi_k^{(n+1)}(s).$$

Comme l'algorithme (3.46) procède à N minimisations de ce type à chaque itération, on en déduit que la suite $(x^{(n)})_{n \in \mathbb{N}}$ construite par cet algorithme est bien définie.

3.(a) Par définition, $x_k^{(n+1)}$ réalise le minimum de la fonction $\varphi_k^{(n+1)}$ sur \mathbb{R} . Comme de plus, $\varphi_k^{(n+1)} \in C^1(\mathbb{R}, \mathbb{R})$, on a donc $(\varphi_k^{(n+1)})'(x_k^{(n+1)}) = 0$. Or $(\varphi_k^{(n+1)})'(x_k^{(n+1)}) = \partial_k f(x^{(n+1, k)})$, et donc $\partial_k f(x^{(n+1, k)}) = 0$. D'après la question 2 de l'exercice 70, on a

$$\begin{aligned} f(x^{(n+1, k-1)}) - f(x^{(n+1, k)}) &\geq \nabla f(x^{(n+1, k)}) \cdot (x^{(n+1, k-1)} - x^{(n+1, k)}) \\ &\quad + \frac{\alpha}{2} |x^{(n+1, k-1)} - x^{(n+1, k)}|^2. \end{aligned}$$

Or $x^{(n+1, k-1)} - x^{(n+1, k)} = -x_k^{(n+1)} e_k$ et $\nabla f(x^{(n+1, k)}) \cdot e_k = \partial_k f(x^{(n+1, k)}) = 0$. On en déduit que :

$$f(x^{(n+1, k-1)}) - f(x^{(n+1, k)}) \geq \frac{\alpha}{2} |x^{(n+1, k-1)} - x^{(n+1, k)}|^2.$$

3.(b) Par définition de la suite $(x^{(n)})_{n \in \mathbb{N}}$, on a :

$$f(x^{(n)}) - f(x^{(n+1)}) = \sum_{k=1}^N f(x^{(n+1, k-1)}) - f(x^{(n+1, k)}).$$

Par la question précédente, on a donc :

$$f(x^{(n)}) - f(x^{(n+1)}) \geq \frac{\alpha}{2} \sum_{k=1}^N |x^{(n+1, k-1)} - x^{(n+1, k)}|^2.$$

Or $x^{(n+1,k-1)} - x^{(n+1,k)} = -x_k^{(n+1)} e_k$, et $(e_k)_{k \in \text{Ndim}}$ est une base orthonormée. On peut donc écrire que

$$\begin{aligned} \sum_{k=1}^N |x^{(n+1,k-1)} - x^{(n+1,k)}|^2 &= \sum_{k=1}^N |(x_k^{(n)} - x_k^{(n+1)}) e_k|^2 \\ &= \left| \sum_{k=1}^N (x_k^{(n)} - x_k^{(n+1)}) e_k \right|^2 \\ &= \left| \sum_{k=1}^N (x^{(n+1,k-1)} - x^{(n+1,k)}) \right|^2 \\ &= |x^{(n)} - x^{(n+1)}|^2. \end{aligned}$$

On en déduit que

$$f(x^{(n)}) - f(x^{(n+1)}) \geq \frac{\alpha}{2} |x^{(n)} - x^{(n+1)}|^2.$$

La suite $(f(x^{(n)}))_{n \in \mathbb{N}}$ est bornée inférieurement par $f(\bar{x})$; l'inégalité précédente montre qu'elle est décroissante, donc elle converge. On a donc $f(x^{(n)}) - f(x^{(n+1)}) \rightarrow 0$ lorsque $n \rightarrow +\infty$, et donc par l'inégalité précédente,

$$\lim_{n \rightarrow +\infty} |x^{(n)} - x^{(n+1)}| = 0.$$

De plus, pour $1 \leq k \leq N$,

$$\begin{aligned} |x^{(n+1,k)} - x^{(n+1)}|^2 &= \sum_{\ell=k}^N |(x_\ell^{(n)} - x_\ell^{(n+1)}) e_\ell|^2 \\ &= \left| \sum_{\ell=k}^N (x_\ell^{(n)} - x_\ell^{(n+1)}) e_\ell \right|^2 \\ &= \left| \sum_{\ell=k}^N (x^{(n+1,\ell-1)} - x^{(n+1,\ell)}) \right|^2 \\ &\leq |x^{(n)} - x^{(n+1)}|^2. \end{aligned}$$

d'où l'on déduit que $\lim_{n \rightarrow +\infty} |x^{(n+1,k)} - x^{(n+1)}| = 0$.

4. En prenant $x = \bar{x}$ et $y = x^{(n+1)}$ dans l'hypothèse (3.44) et en remarquant que, puisque \bar{x} réalise le minimum de f , on a $\nabla f(\bar{x}) = 0$, on obtient :

$$(-\nabla f(x^{(n+1)}) \cdot (\bar{x} - x^{(n+1)})) \geq \alpha |\bar{x} - x^{(n+1)}|^2,$$

et donc, par l'inégalité de Cauchy Schwarz :

$$|x^{(n+1)} - \bar{x}| \leq \frac{1}{\alpha} \left(\sum_{k=1}^N |\partial_k f(x^{(n+1)})|^2 \right)^{\frac{1}{2}}.$$

5. Par les questions 1 et 2 de l'exercice 70, on sait que la fonction f est croissante à l'infini. Donc il existe $R > 0$ tel que si $|x| > R$ alors $f(x) > f(x_0)$. Or, la suite $(f(x_n))_{n \in \mathbb{N}}$ étant décroissante, on a $f(x_n) \leq f(x_0)$ pour tout n , et donc $|x_n| \leq R$ pour tout n . Par la question 3(b), on sait que pour tout $k \geq 1$, $\lim_{n \rightarrow +\infty} |x^{(n+1,k)} - x^{(n+1)}| = 0$, ce qui prouve que les suites $(x^{(n+1,k)})_{n \in \mathbb{N}}$, pour $k = 1, \dots, N$, sont également bornées.

Comme $\lim_{n \rightarrow +\infty} |x^{(n+1,k)} - x^{(n+1)}| = 0$, on a pour tout $\eta > 0$, l'existence de $N_\eta \in \mathbb{N}$ tel que $|x^{(n+1,k)} - x^{(n+1)}| < \eta$ si $n \geq N_\eta$. Comme $f \in C^1(\mathbb{R}, \mathbb{R})$, la fonction $\partial_k f$ est uniformément continue sur les bornés (théorème de Heine), et donc pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que si $|x - y| < \eta$ alors $|\partial_k f(x) - \partial_k f(y)| \leq \varepsilon$. On a donc, pour $n \geq N_\eta$: $|\partial_k f(x^{(n+1,k)}) - \partial_k f(x^{(n+1)})| \leq \varepsilon$, ce qui démontre que :

$$|\partial_k f(x^{(n+1)})| \rightarrow 0 \text{ lorsque } n \rightarrow +\infty.$$

On en conclut par le résultat de la question 4 que $x^{(n)} \rightarrow \bar{x}$ lorsque $n \rightarrow +\infty$.

6. On a vu à l'exercice 68 que dans ce cas, $\nabla f(x) = \frac{1}{2}(A + A^t)x - b$. L'algorithme 3.46 est donc la méthode de Gauss Seidel pour la résolution du système linéaire $\frac{1}{2}(A + A^t)x = b$.

7 (a) La fonction g est strictement convexe (car somme d'une fonction strictement convexe : $(x_1, x_2) \rightarrow x_1^2 + x_2^2$, d'une fonction linéaire par morceaux : $(x_1, x_2) \mapsto -2(x_1 + x_2) + 2|x_1 - x_2|$. et croissante à l'infini grâce aux termes en puissance 2. Il existe donc un unique élément $\bar{x} = (\bar{x}_1, \bar{x}_2)^t$ de \mathbb{R}^2 tel que $g(\bar{x}) = \inf_{x \in \mathbb{R}^2} g(x)$.

7 (b) Soit $\epsilon > 0$. On a, pour tout $x \in \mathbb{R}$, $\phi_x(\epsilon) = g(x, x + \epsilon) = x^2 + (x + \epsilon)^2 - 4x$, qui atteint (pour tout x) son minimum pour $\epsilon = 0$. Le minimum de g se situe donc sur l'axe $x = y$. Or $\psi(x) = g(x, x) = 2x^2 - 4x$ atteint son minimum en $x = 1$.

7 (c) Si $x^{(0)} = (0, 0)^t$, on vérifie facilement que l'algorithme (3.46) appliqué à g est stationnaire. La suite ne converge donc pas vers \bar{x} . La fonction g n'est pas différentiable sur la droite $x_1 = x_2$.

Corrigé de l'exercice 76 page 142 (Gradient conjugué pour une matrice non symétrique)

1. Comme A est inversible, A^t l'est aussi et donc les systèmes (3.47) et (3.48) sont équivalents.

2 (a) La matrice M est symétrique définie positive, car A est inversible et $M = AA^t$ est symétrique. Donc ses valeurs propres sont strictement positives.

2 (b) On a $\text{cond}(A) = \|A\| \|A^{-1}\|$. Comme la norme est ici la norme euclidienne, on a : $\|A\| = (\rho(A^t A))^{\frac{1}{2}}$ et $\|A^{-1}\| = (\rho((A^{-1})^t A^{-1}))^{\frac{1}{2}} = (\rho(AA^t)^{-1})^{\frac{1}{2}}$. On vérifie facilement que $M = A^t A$ et $A^t A$ ont mêmes valeurs propres et on en déduit le résultat.

3. Ecrivons l'algorithme du gradient conjugué pour la résolution du système (3.48)

Initialisation

Soit $x^{(0)} \in \mathbb{R}^N$, et soit $r^{(0)} = A^t b - A^t A x^{(0)} =$

1) Si $r^{(0)} = 0$, alors $Ax^{(0)} = b$ et donc $x^{(0)} = \bar{x}$,
auquel cas l'algorithme s'arrête.

2) Si $r^{(0)} \neq 0$, alors on pose $w^{(0)} = r^{(0)}$, et on choisit $\rho_0 = \frac{r^{(0)} \cdot r^{(0)}}{A^t A w^{(0)} \cdot w^{(0)}}$.
On pose alors $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$.

Itération $1 \leq n \leq N - 1$:

On suppose $x^{(0)}, \dots, x^{(n)}$ et $w^{(0)}, \dots, w^{(n-1)}$ connus et on pose

$r^{(n)} = A^t b - A^t A x^{(n)}$.

1) Si $r^{(n)} = 0$ on a $Ax^{(n)} = b$ donc $x^{(n)} = \bar{x}$
auquel cas l'algorithme s'arrête.

2) Si $r^{(n)} \neq 0$, alors on pose $w^{(n)} = r^{(n)} + \lambda_{n-1} w^{(n-1)}$

avec $\lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}}$ et on pose $\rho_n = \frac{r^{(n)} \cdot r^{(n)}}{A^t A w^{(n)} \cdot w^{(n)}}$.

On pose alors $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$.

Si on implémente l'algorithme sous cette forme, on a intérêt à calculer d'abord $\tilde{b} = A^t b$ et $M = A^t A$ pour minimiser le nombre de multiplications matrice matrice et matrice vecteur. Au lieu du coût de l'algorithme initial, qui est en $2N^3 + O(N^2)$, on a donc un coût en $3N^3 + O(N^2)$.

Maintenant si on est optimiste, on peut espérer converger en moins de N itérations (en fait, c'est malheureusement

rarement le cas), et dans ce cas il est plus économique d'écrire l'algorithme précédent sous la forme suivante.

$$\left\{ \begin{array}{l}
 \textbf{Initialisation} \\
 \text{Soit } x^{(0)} \in \mathbb{R}^N, \text{ et soit } s^{(0)} = b - Ax^{(0)} \text{ et soit } r^{(0)} = A^t s^{(0)} \\
 \text{1) Si } r^{(0)} = 0, \text{ alors } Ax^{(0)} = b \text{ et donc } x^{(0)} = \bar{x}, \\
 \text{auquel cas l'algorithme s'arrête.} \\
 \text{2) Si } r^{(0)} \neq 0, \text{ alors on pose } w^{(0)} = r^{(0)}, y^{(0)} = Aw^{(0)} \text{ et on choisit } \rho_0 = \frac{r^{(0)} \cdot r^{(0)}}{y^{(0)} \cdot y^{(0)}}. \\
 \text{On pose alors } x^{(1)} = x^{(0)} + \rho_0 w^{(0)}. \\
 \\
 \textbf{Itération } 1 \leq n \leq N - 1 : \\
 \text{On suppose } x^{(0)}, \dots, x^{(n)} \text{ et } w^{(0)}, \dots, w^{(n-1)} \text{ connus et on pose} \\
 s^{(n)} = b - Ax^{(n)} \text{ et } r^{(n)} = A^t s^{(n)}. \\
 \text{1) Si } r^{(n)} = 0 \text{ on a } Ax^{(n)} = b \text{ donc } x^{(n)} = \bar{x} \\
 \text{auquel cas l'algorithme s'arrête.} \\
 \text{2) Si } r^{(n)} \neq 0, \text{ alors on pose } w^{(n)} = r^{(n)} + \lambda_{n-1} w^{(n-1)} \\
 \text{avec } \lambda_{n-1} = \frac{r^{(n)} \cdot r^{(n)}}{r^{(n-1)} \cdot r^{(n-1)}} \text{ et on pose } \rho_n = \frac{r^{(n)} \cdot r^{(n)}}{y^{(n)} \cdot y^{(n)}} \text{ avec } y^{(n)} = Aw^{(n)}. \\
 \text{On pose alors } x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}.
 \end{array} \right.$$

On peut facilement vérifier que dans cette version, on a un produit matrice vecteur en plus à chaque itération, donc le coût est le même pour N itérations, mais il est inférieur si on a moins de N itérations.

Remarque : Cette méthode s'appelle méthode du gradient conjugué appliquée aux équations normales. Elle est facile à comprendre et à programmer. Malheureusement, elle ne marche pas très bien dans la pratique, et on lui préfère des méthodes plus sophistiquées telles que la méthode "BICGSTAB" ou "GMRES".

Corrigé de l'exercice 78 page 143 (Méthode de Polak-Ribière)

1. Montrons que f est strictement convexe et croissante à l'infini. Soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par

$$\varphi(t) = f(x + t(y - x)).$$

On a $\varphi \in C^2(\mathbb{R}, \mathbb{R})$, $\varphi(0) = f(x)$ et $\varphi(1) = f(y)$, et donc :

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t) dt.$$

En intégrant par parties, ceci entraîne :

$$f(y) - f(x) = \varphi'(0) + \int_0^1 (1-t)\varphi''(t) dt. \quad (3.65)$$

Or $\varphi'(t) = \nabla(x + t(y-x)) \cdot (y-x)$ et donc $\varphi''(t) = H(x + t(y-x))(y-x) \cdot (y-x)$. On a donc par hypothèse $\varphi''(t) \geq \alpha|y-x|^2$.

On déduit alors de 3.65 que

$$f(y) \geq f(x) + \nabla f(x) \cdot (y-x) + \frac{\alpha}{2}|y-x|^2. \quad (3.66)$$

L'inégalité 3.66 entraîne la stricte convexité de f et sa croissance à l'infini (voir démonstration de la convergence du gradient à pas fixe, exercice 27).

Il reste à montrer que l'ensemble $\mathcal{VP}(H(x))$ des valeurs propres de $H(x)$ est inclus dans $[\alpha, \beta]$. Comme $f \in C^2(\mathbb{R}, \mathbb{R})$, $H(x)$ est symétrique pour tout $x \in \mathbb{R}$, et donc diagonalisable dans \mathbb{R} . Soit $\lambda \in \mathcal{VP}(H(x))$; il existe donc $y \in \mathbb{R}^N$, $y \neq 0$ tel que $H(x)y = \lambda y$, et donc $\alpha y \cdot y \leq \lambda y \cdot y \leq \beta y \cdot y$, $\forall \lambda \in \mathcal{VP}(H(x))$. On en déduit que $\mathcal{VP}(H(x)) \subset [\alpha, \beta]$.

2. Montrons par récurrence sur n que $g^{(n+1)} \cdot w^{(n)} = 0$ et $g^{(n)} \cdot g^{(n)} = g^{(n)} \cdot w^{(n)}$ pour tout $n \in \mathbb{N}$.

Pour $n = 0$, on a $w^{(0)} = g^{(0)} = -\nabla f(x^{(0)})$.

Si $\nabla f(x^{(0)}) = 0$ l'algorithme s'arrête. Supposons donc que $\nabla f(x^{(0)}) \neq 0$. Alors $w^{(0)} = -\nabla f(x^{(0)})$ est une direction de descente stricte. Comme $x^{(1)} = x^{(0)} + \rho_0 w^{(0)}$ où ρ_0 est optimal dans la direction $w^{(0)}$, on a $g^{(1)} \cdot w^{(0)} = -\nabla f(x^{(1)}) \cdot w^{(0)} = 0$. De plus, on a évidemment $g^{(0)} \cdot w^{(0)} = g^{(0)} \cdot g^{(0)}$.

Supposons maintenant que $g^{(n)} \cdot w^{(n-1)} = 0$ et $g^{(n-1)} \cdot g^{(n-1)} = g^{(n-1)} \cdot w^{(n-1)}$, et montrons que $g^{(n+1)} \cdot w^{(n)} = 0$ et $g^{(n)} \cdot g^{(n)} = 0$.

Par définition, on a :

$$\begin{aligned} w^{(n)} &= g^{(n)} + \lambda_{n-1} w^{(n-1)}, \text{ donc} \\ w^{(n)} \cdot g^{(n)} &= g^{(n)} \cdot g^{(n)} + \lambda_{n-1} w^{(n-1)} \cdot g^{(n)} = g^{(n)} \cdot g^{(n)} \end{aligned}$$

par hypothèse de récurrence. On déduit de cette égalité que $w^{(n)} \cdot g^{(n)} > 0$ (car $g^{(n)} \neq 0$) et donc $w^{(n)}$ est une direction de descente stricte en $x^{(n)}$. On a donc $\nabla f(x^{(n+1)}) \cdot w^{(n)} = 0$, et finalement $g^{(n+1)} \cdot w^{(n)} = 0$.

3. Par définition, $g^{(n)} = -\nabla f(x^{(n)})$; or on veut calculer $g^{(n+1)} - g^{(n)} = -\nabla f(x^{(n+1)}) + \nabla f(x^{(n)})$. Soit φ la fonction de \mathbb{R} dans \mathbb{R} définie par :

$$\varphi(t) = -\nabla f(x^{(n)} + t(x^{(n+1)} - x^{(n)})).$$

On a donc :

$$\begin{aligned} \varphi(1) - \varphi(0) &= g^{(n+1)} - g^{(n)} \\ &= \int_0^1 \varphi'(t) dt. \end{aligned}$$

Calculons φ' : $\varphi'(t) = H(x^{(n)} + t(x^{(n+1)} - x^{(n)}))(x^{(n+1)} - x^{(n)})$. Et comme $x^{(n+1)} = x^{(n)} + \rho_n w^{(n)}$, on a donc :

$$g^{(n+1)} - g^{(n)} = \rho_n J_n w^{(n)}. \quad (3.67)$$

De plus, comme $g^{(n+1)} \cdot w^{(n)} = 0$ (question 1), on obtient par (3.67) que

$$\rho_n = \frac{g^{(n)} \cdot w^{(n)}}{J_n w^{(n)} \cdot w^{(n)}}$$

(car $J_n w^{(n)} \cdot w^{(n)} \neq 0$, puisque J_n est symétrique définie positive).

4. Par définition, on a $w^{(n)} = g^{(n)} + \lambda_{n-1} w^{(n-1)}$, et donc

$$|w^{(n)}| \leq |g^{(n)}| + |\lambda_{n-1}| |w^{(n-1)}|. \quad (3.68)$$

Toujours par définition, on a :

$$\lambda_{n-1} = \frac{g^{(n)} \cdot (g^{(n)} - g^{(n-1)})}{g^{(n-1)} \cdot g^{(n-1)}}.$$

Donc, par la question 3, on a :

$$\lambda_{n-1} = \frac{\rho_n g^{(n)} \cdot J^{(n-1)} w^{(n-1)}}{g^{(n-1)} \cdot g^{(n-1)}}.$$

En utilisant la question 2 et à nouveau la question 3, on a donc :

$$\lambda_{n-1} = -\frac{J^{(n-1)} w^{(n-1)} \cdot g^{(n)}}{J^{(n-1)} w^{(n-1)} \cdot w^{(n-1)}},$$

et donc

$$\lambda_{n-1} = \frac{|J^{(n-1)} w^{(n-1)} \cdot g^{(n)}|}{J^{(n-1)} w^{(n-1)} \cdot w^{(n-1)}},$$

car $J^{(n-1)}$ est symétrique définie positive.

De plus, en utilisant les hypothèses sur H , on vérifie facilement que

$$\alpha |x|^2 \leq J^{(n)} x \cdot x \leq \beta |x|^2 \quad \forall x \in \mathbb{R}^N.$$

On en déduit que

$$\lambda_{n-1} \leq \frac{|J^{(n-1)} w^{(n-1)} \cdot g^{(n)}|}{\alpha |w^{(n-1)}|^2}.$$

On utilise alors l'inégalité de Cauchy–Schwarz :

$$\begin{aligned} |J^{(n-1)} w^{(n-1)} \cdot g^{(n)}| &\leq \|J^{(n-1)}\|_2 |w^{(n-1)}| |g^{(n-1)}| \\ &\leq \beta |w^{(n-1)}| |g^{(n-1)}|. \end{aligned}$$

On obtient donc que

$$\lambda_{n-1} \leq \frac{\beta |g^{(n-1)}|}{\alpha |w^{(n-1)}|},$$

ce qui donne bien grâce à (3.68) :

$$|w^{(n)}| \leq |g^{(n)}| \left(1 + \frac{\beta}{\alpha}\right).$$

5. • Montrons d'abord que la suite $(f(x^{(n)}))_{n \in \mathbb{N}}$ converge. Comme $f(x^{(n+1)}) = f(x^{(n)} + \rho_n w^{(n)}) \leq f(x^{(n)} + \rho w^{(n)}) \forall \rho \geq 0$, on a donc en particulier $f(x^{(n+1)}) \leq f(x^{(n)})$. La suite $(f(x^{(n)}))_{n \in \mathbb{N}}$ est donc décroissante. De plus, elle est minorée par $f(\bar{x})$. Donc elle converge, vers une certaine limite $\ell \in \mathbb{R}$, lorsque n tend vers $+\infty$.
- La suite $(x^{(n)})_{n \in \mathbb{N}}$ est bornée : en effet, comme f est croissante à l'infini, il existe $R > 0$ tel que si $|x| > R$ alors $f(x) \geq f(x^{(0)})$. Or $f(x^{(n)}) \geq f(x^{(0)})$ pour tout $n \in \mathbb{N}$, et donc la suite $(x^{(n)})_{n \in \mathbb{N}}$ est incluse dans la boule de rayon R .
- Montrons que $\nabla f(x^{(n)}) \rightarrow 0$ lorsque $n \rightarrow +\infty$.
On a, par définition de $x^{(n+1)}$,

$$f(x^{(n+1)}) \leq f(x^{(n)} + \rho w^{(n)}), \quad \forall \rho \geq 0.$$

En introduisant la fonction φ définie de \mathbb{R} dans \mathbb{R} par $\varphi(t) = f(x^{(n)} + t\rho w^{(n)})$, on montre facilement (les calculs sont les mêmes que ceux de la question 1) que

$$f(x^{(n)} + \rho w^{(n)}) = f(x^{(n)}) + \rho \nabla f(x^{(n)}) \cdot w^{(n)} + \rho^2 \int_0^1 H(x^{(n)} + t\rho w^{(n)}) w^{(n)} \cdot w^{(n)} (1-t) dt,$$

pour tout $\rho \geq 0$. Grâce à l'hypothèse sur H , on en déduit que

$$f(x^{(n+1)}) \leq f(x^{(n)}) + \rho \nabla f(x^{(n)}) \cdot w^{(n)} + \frac{\beta}{2} \rho^2 |w^{(n)}|^2, \quad \forall \rho \geq 0.$$

Comme $\nabla f(x^{(n)}) \cdot w^{(n)} = -g^{(n)} \cdot w^{(n)} = -|g^{(n)}|^2$ (question 2) et comme $|w^{(n)}| \leq |g^{(n)}| \left(1 + \frac{\beta}{\alpha}\right)$ (question 4), on en déduit que :

$$f(x^{(n+1)}) \leq f(x^{(n)}) - \rho |g^{(n)}|^2 + \rho^2 \gamma |g^{(n)}|^2 = \psi_n(\rho), \quad \forall \rho \geq 0,$$

où $\gamma = \frac{\beta^2}{2} + \left(1 + \frac{\beta}{\alpha}\right)^2$. La fonction ψ_n est un polynôme de degré 2 en ρ , qui atteint son minimum lorsque $\psi'_n(\rho) = 0$, i.e. pour $\rho = \frac{1}{2\gamma}$. On a donc, pour $\rho = \frac{1}{2\gamma}$,

$$f(x^{(n+1)}) \leq f(x^{(n)}) - \frac{1}{4\gamma} |g^{(n)}|^2,$$

d'où on déduit que

$$|g^{(n)}|^2 \leq 4\gamma (f(x^{(n)}) - f(x^{(n+1)})) \xrightarrow{n \rightarrow +\infty} 0$$

On a donc $\nabla f(x^{(n)}) \rightarrow 0$ lorsque $n \rightarrow +\infty$.

- La suite $(x^{(n)})_{n \in \mathbb{N}}$ étant bornée, il existe une sous-suite qui converge vers $x \in \mathbb{R}^N$, comme $\nabla f(x^{(n)}) \rightarrow 0$ et comme $\text{Hess} f$ est continue, on a $\nabla f(x) = 0$. Par unicité du minimum (f est croissante à l'infini et strictement convexe) on a donc $x = \bar{x}$.
Enfin on conclut à la convergence de toute la suite par un argument classique (voir question 6 de l'exercice 72 page 139).

Corrigé de l'exercice 80 page 144 (Méthodes de Gauss–Newton et de quasi–linéarisation)

Soit $f \in C^2(\mathbb{R}^N, \mathbb{R}^P)$, avec $N, P \in \mathbb{N}^*$. Soit $C \in \mathcal{M}_P(\mathbb{R})$ une matrice réelle carrée d'ordre P , symétrique définie positive, et $d \in \mathbb{R}^P$. Pour $x \in \mathbb{R}^N$, on pose

$$J(x) = (f(x) - d) \cdot C(f(x) - d).$$

On cherche à minimiser J .

I Propriétés d'existence et d'unicité

- (a) Comme C est symétrique éfinie positive, on a $y \cdot Cy \geq 0$ pour tout $y \in \mathbb{R}^N$, ce qui prouve que $J(x) \geq 0$ pour tout $x \in \mathbb{R}^N$. Donc J est bornée inférieurement.
- (b) Trois exemples
- Si $N = P$ et $f(x) = x$, $J(x) = (x - d) \cdot C(x - d)$ qui est une fonction quadratique pour laquelle on a existence et unicité de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J .
 - Si $f(x) = 0$, $J(x) = d \cdot C$ et J est donc constante. Il y a donc existence et non unicité de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J .
 - Pour $N = P = 1$, si $f(x) = e^x$, avec $c = 1$ et $d = 0$, $J(x) = (e^x)^2$ tend vers 0 en l'infini mais 0 n'est jamais atteint. Il ya donc non existence de $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J .

II Un peu de calcul différentiel

- (a) On note Df et D_2f les différentielles d'ordre 1 et 2 de f . A quels espaces appartiennent $Df(x)$, $D_2f(x)$ (pour $x \in \mathbb{R}^N$), ainsi que Df et D_2f ? Montrer que pour tout $x \in \mathbb{R}^N$, il existe $M(x) \in \mathcal{M}_{P,N}(\mathbb{R})$, où $\mathcal{M}_{P,N}(\mathbb{R})$ désigne l'ensemble des matrices réelles à P lignes et N colonnes, telle que $Df(x)(y) = M(x)y$ pour tout $y \in \mathbb{R}^N$.

$Df(x)$ est la différentielle de f en x et c'est donc une application linéaire de \mathbb{R}^N dans \mathbb{R}^P . Donc il existe $M(x) \in \mathcal{M}_{P,N}(\mathbb{R})$, où $\mathcal{M}_{P,N}(\mathbb{R})$ désigne l'ensemble des matrices réelles à P lignes et N colonnes, telle que $Df(x)(y) = M(x)y$ pour tout $y \in \mathbb{R}^N$.

On a ensuite $D_2f(x) \in \mathcal{L}(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N, \mathbb{R}^P))$.

Enfin, on a $Df \in C^1(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N, \mathbb{R}^P))$ et $D_2f \in \mathcal{L}(\mathbb{R}^N, \mathcal{L}(\mathbb{R}^N, \mathbb{R}^P))$.

- (b) Comme C ne dépend pas de x , on a $\nabla J(x) = M(x)C(f(x) - d) + (f(x) - d)CM(x)$.
- (c) Pour $x \in \mathbb{R}^N$, calculer la matrice hessienne de J en x (qu'on notera $H(x)$). On suppose maintenant que M ne dépend pas de x ; montrer que dans ce cas $H(x) = 2M(x)^tCM(x)$.

III Algorithmes d'optimisation

Dans toute cette question, on suppose qu'il existe un unique $\bar{x} \in \mathbb{R}^N$ qui réalise le minimum de J , qu'on cherche à calculer de manière itérative. On se donne pour cela $x_0 \in \mathbb{R}^N$, et on cherche à construire une suite $(x_n)_{n \in \mathbb{N}}$ qui converge vers \bar{x} .

- (a) On cherche à calculer \bar{x} en utilisant la méthode de Newton pour annuler ∇J . Justifier brièvement cette procédure et écrire l'algorithme obtenu.
- (b) L'algorithme dit de "Gauss-Newton" est une modification de la méthode précédente, qui consiste à approcher, à chaque itération n , la matrice jacobienne de ∇J en x_n par la matrice obtenue en négligeant les dérivées secondes de f . Ecrire l'algorithme ainsi obtenu.
- (c) L'algorithme dit de "quasi–linéarisation" consiste à remplacer, à chaque itération $n \in \mathbb{N}$, la minimisation de la fonctionnelle J par celle de la fonctionnelle J_n , définie de \mathbb{R}^N dans \mathbb{R} , et obtenue à partir de J en effectuant un développement limité au premier ordre de $f(x)$ en x_n , c.à.d.

$$J_n(x) = (f(x_n) + Df(x_n)(x - x_n) - d) \cdot C(f(x_n) + Df(x_n)(x - x_n) - d).$$

- i. Soit $n \geq 0$, $x_n \in \mathbb{R}^N$ connu, $M_n = M(x_n) \in \mathcal{M}_{P,N}(\mathbb{R})$, et $x \in \mathbb{R}^N$. On pose $h = x - x_n$. Montrer que

$$J_n(x) = J(x_n) + M_n^tCM_n h \cdot h + 2M_n^tC(f(x_n) - d) \cdot h.$$

- ii. Montrer que la recherche du minimum de J_n est équivalente à la résolution d'un système linéaire dont on donnera l'expression.
- iii. Ecrire l'algorithme de quasi–linéarisation, et le comparer avec l'algorithme de Gauss-Newton.

Corrigé de l'exercice 79 page 143 (Algorithme de quasi Newton)

Partie 1

1. Par définition de $w^{(n)}$, on a :

$$w^{(n)} \cdot \nabla f(x^{(n)}) = -K^{(n)} \nabla f(x^{(n)}) \cdot \nabla f(x^{(n)}) < 0$$

car K est symétrique définie positive.

Comme ρ_n est le paramètre optimal dans la direction $w^{(n)}$, on a $\nabla f(x^{(n)} + \rho_n w^{(n)}) \cdot w^{(n)} = 0$, et donc $Ax^{(n)} \cdot w^{(n)} + \rho_n Aw^{(n)} \cdot w^{(n)} = b \cdot w^{(n)}$; on en déduit que

$$\rho_n = -\frac{g^{(n)} \cdot w^{(n)}}{Aw^{(n)} \cdot w^{(n)}}.$$

Comme $w^{(n)} = -K^{(n)}g^{(n)}$, ceci s'écrit encore :

$$\rho_n = \frac{g^{(n)} \cdot K^{(n)}g^{(n)}}{AK^{(n)}g^{(n)} \cdot K^{(n)}g^{(n)}}.$$

2. Si $K^{(n)} = A^{-1}$, la formule précédente donne immédiatement $\rho_n = 1$.
3. La méthode de Newton consiste à chercher le zéro de ∇f par l'algorithme suivant (à l'itération 1) :

$$H_f(x^{(0)})(x^{(1)} - x^{(0)}) = -\nabla f(x^{(0)}),$$

(où $H_f(x)$ désigne la hessienne de f au point x) c'est-à-dire

$$A(x^{(1)} - x^{(0)}) = -Ax^{(0)} + b.$$

On a donc $Ax^{(1)} = b$, et comme la fonction f admet un unique minimum qui vérifie $Ax = b$, on a donc $x^{(1)} = x$, et la méthode converge en une itération.

Partie 2 Méthode de Fletcher-Powell.

1. Soit $n \in \mathbb{N}$, on suppose que $g^{(n)} \neq 0$. Par définition, on a $s^{(n)} = x^{(n+1)} - x^{(n)} = -\rho_n K^{(n)}g^{(n)}$, avec $\rho_n > 0$. Comme $K^{(n)}$ est symétrique définie positive elle est donc inversible; donc comme $g^{(n)} \neq 0$, on a $K^{(n)}g^{(n)} \neq 0$ et donc $s^{(n)} \neq 0$.

Soit $i < n$, par définition de $s^{(n)}$, on a :

$$s^{(n)} \cdot As^{(i)} = -\rho_n K^{(n)}g^{(n)} \cdot As^{(i)}.$$

Comme $K^{(n)}$ est symétrique,

$$s^{(n)} \cdot As^{(i)} = -\rho_n g^{(n)} \cdot K^{(n)}As^{(i)}.$$

Par hypothèse, on a $K^{(n)}As^{(i)} = s^{(i)}$ pour $i < n$, donc on a bien que si $i < n$

$$s^{(n)} \cdot As^{(i)} = 0 \Leftrightarrow g^{(n)} \cdot s^{(i)} = 0.$$

Montrons maintenant que $g^{(n)} \cdot s^{(i)} = 0$ pour $i < n$.

- On a

$$\begin{aligned} g^{(i+1)} \cdot s^{(i)} &= -\rho_i g^{(i+1)} \cdot K^{(i)}g^{(i)} \\ &= -\rho_i g^{(i+1)} \cdot w^{(i)}. \end{aligned}$$

Or $g^{(i+1)} = \nabla f(x^{(i+1)})$ et ρ_i est optimal dans la direction $w^{(i)}$. Donc

$$g^{(i+1)} \cdot s^{(i)} = 0.$$

• On a

$$\begin{aligned}
 (g^{(n)} - g^{(i+1)}) \cdot s^{(i)} &= (Ax^{(n)} - Ax^{(i+1)}) \cdot s^{(i)} \\
 &= \sum_{k=i+1}^{n-1} (Ax^{(k+1)} - Ax^{(k)}) \cdot s^{(i)} \\
 &= \sum_{k=i+1}^{n-1} As^{(k)} \cdot s^{(i)}, \\
 &= 0
 \end{aligned}$$

Par hypothèse de A -conjugaison de la famille $(s^{(i)})_{i=1, k-1}$ on déduit alors facilement des deux égalités précédentes que $g^{(n)} \cdot s^{(i)} = 0$. Comme on a montré que $g^{(n)} \cdot s^{(i)} = 0$ si et seulement si $s^{(n)} \cdot As^{(i)} = 0$, on en conclut que la famille $(s^{(i)})_{i=1, \dots, n}$ est A -conjuguée, et que les vecteurs $s^{(i)}$ sont non nuls.

2. Montrons que $K^{(n+1)}$ est symétrique. On a :

$$(K^{(n+1)})^t = (K^{(n)})^t + \frac{(s^{(n)}(s^{(n)})^t)^t}{s^{(n)} \cdot y^{(n)}} - \frac{[(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^t]^t}{K^{(n)}y^{(n)} \cdot y^{(n)}} = K^{(n+1)},$$

car $K^{(n)}$ est symétrique.

3. Montrons que $K^{(n+1)}As^{(i)} = s^{(i)}$ si $0 \leq i \leq n$. On a :

$$K^{(n+1)}As^{(i)} = K^{(n)}As^{(i)} + \frac{s^{(n)}(s^{(n)})^t}{s^{(n)} \cdot y^{(n)}}As^{(i)} - \frac{(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^t}{K^{(n)}y^{(n)} \cdot y^{(n)}}As^{(i)}. \quad (3.69)$$

– Considérons d’abord le cas $i < n$. On a

$$s^{(n)}(s^{(n)})^tAs^{(i)} = s^{(n)}[(s^{(n)})^tAs^{(i)}] = s^{(n)}[s^{(n)} \cdot As^{(i)}] = 0$$

car $s^{(n)} \cdot As^{(i)} = 0$ si $i < n$. De plus, comme $K^{(n)}$ est symétrique, on a :

$$(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^tAs^{(i)} = K^{(n)}y^{(n)}(y^{(n)})^tK^{(n)}As^{(i)}.$$

Or par la question (c), on a $K^{(n)}As^{(i)} = s^{(i)}$ si $0 \leq i \leq n$. De plus, par définition, $y^{(n)} = As^{(n)}$. On en déduit que

$$(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^tAs^{(i)} = K^{(n)}y^{(n)}(As^{(n)})^tAs^{(i)} = K^{(n)}y^{(n)}(s^{(n)})^tAs^{(i)} = 0$$

puisque on a montré en (a) que les vecteurs $s^{(0)}, \dots, s^{(n)}$ sont A -conjugués. On déduit alors de (3.69) que

$$K^{(n+1)}As^{(i)} = K^{(n)}As^{(i)} = s^{(i)}.$$

– Considérons maintenant le cas $i = n$. On a

$$K^{(n+1)}As^{(n)} = K^{(n)}As^{(n)} + \frac{s^{(n)}(s^{(n)})^t}{s^{(n)} \cdot y^{(n)}}As^{(n)} - \frac{(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^t}{K^{(n)}y^{(n)} \cdot y^{(n)}}As^{(n)},$$

et comme $y^{(n)} = As^{(n)}$, ceci entraîne que

$$K^{(n+1)}As^{(n)} = K^{(n)}As^{(n)} + s^{(n)} - K^{(n)}y^{(n)} = s^{(n)}.$$

4. Pour $x \in \mathbb{R}^N$, calculons $K^{(n+1)}x \cdot x$:

$$K^{(n+1)}x \cdot x = K^{(n)}x \cdot x + \frac{s^{(n)}(s^{(n)})^t}{s^{(n)} \cdot y^{(n)}}x \cdot x - \frac{(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^t}{K^{(n)}y^{(n)} \cdot y^{(n)}}x \cdot x.$$

Or $s^{(n)}(s^{(n)})^tx \cdot x = s^{(n)}(s^{(n)} \cdot x) \cdot x = (s^{(n)} \cdot x)^2$, et de même, $(K^{(n)}y^{(n)})(K^{(n)}y^{(n)})^tx \cdot x = (K^{(n)}y^{(n)} \cdot x)^2$. On en déduit que

$$K^{(n+1)}x \cdot x = K^{(n)}x \cdot x + \frac{(s^{(n)} \cdot x)^2}{s^{(n)} \cdot y^{(n)}} - \frac{(K^{(n)}y^{(n)} \cdot x)^2}{K^{(n)}y^{(n)} \cdot y^{(n)}}.$$

En remarquant que $y^{(n)} = As^{(n)}$, et en réduisant au même dénominateur, on obtient alors que

$$K^{(n+1)}x \cdot x = \frac{(K^{(n)}x \cdot x)(K^{(n)}y^{(n)} \cdot y^{(n)}) - (K^{(n)}y^{(n)} \cdot x)^2}{(K^{(n)}y^{(n)} \cdot y^{(n)})} + \frac{(s^{(n)} \cdot x)^2}{As^{(n)} \cdot s^{(n)}}.$$

Montrons maintenant que $K^{(n+1)}$ est symétrique définie positive. Comme $K^{(n)}$ est symétrique définie positive, on a grâce à l'inégalité de Cauchy-Schwarz que $(K^{(n)}y^{(n)} \cdot x)^2 \leq (K^{(n)}x \cdot x)(K^{(n)}y^{(n)} \cdot y^{(n)})$ avec égalité si et seulement si x et $y^{(n)}$ sont colinéaires. Si x n'est pas colinéaire à $y^{(n)}$, on a donc clairement

$$K^{(n+1)}x \cdot x > 0.$$

Si maintenant x est colinéaire à $y^{(n)}$, i.e. $x = \alpha y^{(n)}$ avec $\alpha \in \mathbb{R}_+^*$, on a, grâce au fait que $y^{(n)} = As^{(n)}$,

$$\frac{(s^{(n)} \cdot x)^2}{As^{(n)} \cdot s^{(n)}} = \alpha^2 \frac{(s^{(n)} \cdot As^{(n)})^2}{As^{(n)} \cdot s^{(n)}} > 0, \text{ et donc } K^{(n+1)}x \cdot x > 0.$$

On en déduit que $K^{(n+1)}$ est symétrique définie positive.

5. On suppose que $g^{(n)} \neq 0$ si $0 \leq n \leq N-1$. On prend comme hypothèse de récurrence que les vecteurs $s^{(0)}, \dots, s^{(n-1)}$ sont A-conjugués et non-nuls, que $K^{(j)}As^{(i)} = s^{(i)}$ si $0 \leq i < j \leq n$ et que les matrices $K^{(j)}$ sont symétriques définies positives pour $j = 0, \dots, n$.

Cette hypothèse est vérifiée au rang $n = 1$ grâce à la question 1 en prenant $n = 0$ et $K^{(0)}$ symétrique définie positive.

On suppose qu'elle est vraie au rang n . La question 1 prouve qu'elle est vraie au rang $n + 1$.

Il reste maintenant à montrer que $x^{(N+1)} = A^{-1}b = \bar{x}$. On a en effet $K^{(N)}As^{(i)} = s^{(i)}$ pour $i = 0$ à $N-1$. Or les vecteurs $s^{(0)}, \dots, s^{(N-1)}$ sont A-conjugués et non-nuls : ils forment donc une base. On en déduit que $K^{(N)}A = Id$, ce qui prouve que $K^{(N)} = A^{-1}$, et donc, par définition de $x^{(N+1)}$, que $x^{(N+1)} = A^{-1}b = \bar{x}$.

Exercice 82 page 145 (Sur l'existence et l'unicité)

La fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = x^2$ est continue, strictement convexe, et croissante à l'infini. Etudions maintenant les propriétés de K dans les quatre cas proposés :

(i) L'ensemble $K = \{|x| \leq 1\}$ est fermé borné et convexe. On peut donc appliquer le théorème d'existence et d'unicité 3.34 page 129. En remarquant que $f(x) \geq 0$ pour tout $x \in \mathbb{R}$ et que $f(0) = 0$, on en déduit que l'unique solution du problème (3.29) est donc $\bar{x} = 0$.

(ii) L'ensemble $K = \{|x| = 1\}$ est fermé borné mais non convexe. Le théorème d'existence 3.32 page 129 s'applique donc, mais pas le théorème d'unicité 3.33 page 129. De fait, on peut remarquer que $K = \{-1, 1\}$, et donc $\{f(x), x \in K\} = \{1\}$. Il existe donc deux solutions du problème (3.29) : $\bar{x}_1 = 1$ et $\bar{x}_1 = -1$.

(iii) L'ensemble $K = \{|x| \geq 1\}$ est fermé, non borné et non convexe. Cependant, on peut écrire $K = K_1 \cup K_2$ où $K_1 = [1, +\infty[$ et $K_2 =]-\infty, -1]$ sont des ensembles convexes fermés. On peut donc appliquer le théorème 3.34 page 129 : il existe un unique $\bar{x}_1 \in \mathbb{R}$ et un unique $\bar{x}_2 \in \mathbb{R}$ solution de (3.29) pour $K = K_1$ et $K = K_2$ respectivement. Il suffit ensuite de comparer \bar{x}_1 et \bar{x}_2 . Comme $\bar{x}_1 = -1$ et $\bar{x}_2 = 1$, on a existence mais pas unicité.

(iv) L'ensemble $K = \{|x| > 1\}$ n'est pas fermé, donc le théorème 3.32 page 129 ne s'applique pas. De fait, il n'existe pas de solution dans ce cas, car on a $\lim_{x \rightarrow +\infty} f(x) = +\infty$, et donc $\inf_K f = 1$, mais cet infimum n'est pas atteint.

Exercice 83 page 145 (Maximisation de l'aire d'un rectangle à périmètre donné)

1. On peut se ramener sans perte de généralité au cas du rectangle $[0, x_1] \times [0, x_2]$, dont l'aire est égale à x_1x_2 et de périmètre $2(x_1 + x_2)$. On veut donc maximiser x_1x_2 , ou encore minimiser $-x_1x_2$. Pour $x = (x_1, x_2)^t \in \mathbb{R}^2$, posons $f(x_1, x_2) = -x_1x_2$ et $g(x_1, x_2) = x_1 + x_2$. Définissons

$$K = \{x = (x_1, x_2)^t \in (\mathbb{R}_+)^2 \text{ tel que } x_1 + x_2 = 1\}.$$

Le problème de minimisation de l'aire du rectangle de périmètre donné et égal à 2 s'écrit alors :

$$\begin{cases} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in K \\ f(\bar{x}_1, \bar{x}_2) \leq f(x_1, x_2) \quad \forall (x_1, x_2) \in K \end{cases} \quad (3.70)$$

2. Comme x_1 et x_2 sont tous deux positifs, puisque leur somme doit être égale à 1, ils sont forcément tous deux inférieurs à 1. Il est donc équivalent de résoudre (3.70) ou (3.51). L'ensemble \tilde{K} est un convexe fermé borné, la fonction f est continue, et donc par le théorème 3.32 page 129, il existe au moins une solution du problème (3.51) (ou (3.70)).

3. Calculons $\nabla g : \nabla g(x) = (1, 1)^t$, donc $\text{rang } Dg(x, y) = 1$. Par le théorème de Lagrange, si $x = (x_1, x_2)^t$ est solution de (3.70), il existe $\lambda \in \mathbb{R}$ tel que

$$\begin{cases} \nabla f(\bar{x}, \bar{y}) + \lambda \nabla g(\bar{x}, \bar{y}) = 0, \\ \bar{x} + \bar{y} = 1. \end{cases}$$

Or $\nabla f(\bar{x}, \bar{y}) = (-\bar{x}, -\bar{y})^t$, et $\nabla g(\bar{x}, \bar{y}) = (1, 1)^t$. Le système précédent s'écrit donc :

$$-\bar{y} + \lambda = 0 \quad -\bar{x} + \lambda = 0 \quad \bar{x} + \bar{y} = 1.$$

On a donc

$$\bar{x} = \bar{y} = \frac{1}{2}.$$

Exercice 84 page 146 (Fonctionnelle quadratique)

1. Comme $d \neq 0$, il existe $\tilde{x} \in \mathbb{R}^N$ tel que $d \cdot \tilde{x} = \alpha \neq 0$. Soit $x = \frac{c}{\alpha} \tilde{x}$ alors $d \cdot x = c$. Donc l'ensemble K est non vide. L'ensemble K est fermé car noyau d'une forme linéaire continue de \mathbb{R}^N dans \mathbb{R} , et K est évidemment convexe. La fonction f est strictement convexe et $f(x) \rightarrow +\infty$ quand $|x| \rightarrow +\infty$, et donc par les théorèmes 3.32 et 3.33 il existe un unique \bar{x} solution de (3.29).

2. On veut calculer \bar{x} . On a : $Dg(x)z = d \cdot z$, et donc $Dg(x) = d^t$. Comme $d \neq 0$ on a $\text{rang}(Dg(x)) = 1$, ou encore $\text{Im}(Dg(x)) = \mathbb{R}$ pour tout x . Donc le théorème de Lagrange s'applique. Il existe donc $\lambda \in \mathbb{R}$ tel que $\nabla f(\bar{x}) + \lambda \nabla g(\bar{x}) = 0$, c'est-à-dire $A\bar{x} - b + \lambda d = 0$. Le couple (\bar{x}, λ) est donc solution du problème suivant :

$$\begin{cases} A\bar{x} - b + \lambda d = 0, \\ d \cdot \bar{x} = c \end{cases}, \quad (3.71)$$

qui s'écrit sous forme matricielle : $By = e$, avec $B = \left[\begin{array}{c|c} A & d \\ \hline d^t & 0 \end{array} \right] \in \mathcal{M}_{N+1}(\mathbb{R})$, $y = \begin{bmatrix} \bar{x} \\ \lambda \end{bmatrix} \in \mathbb{R}^{N+1}$ et

$e = \begin{bmatrix} b \\ c \end{bmatrix} \in \mathbb{R}^{N+1}$. Montrons maintenant que B est inversible. En effet, soit $z = \begin{bmatrix} x \\ \mu \end{bmatrix} \in \mathbb{R}^{N+1}$, avec $x \in \mathbb{R}^N$

et $\mu \in \mathbb{R}$ tel que $Bz = 0$. Alors

$$\left[\begin{array}{c|c} A & d \\ \hline d^t & 0 \end{array} \right] \begin{bmatrix} x \\ \mu \end{bmatrix} = 0.$$

Ceci entraîne $Ax - d\mu = 0$ et $d^t x = d \cdot x = 0$. On a donc $Ax \cdot x - (d \cdot x)\mu = 0$. On en déduit que $x = 0$, et comme $d \neq 0$, que $\mu = 0$. On a donc finalement $z = 0$.

On en conclut que B est inversible, et qu'il existe un unique $(x, \lambda)^t \in \mathbb{R}^{N+1}$ solution de (3.71) et et \bar{x} est solution de (3.29).

Exercice 88 page 147 (Application simple du théorème de Kuhn-Tucker)

La fonction f définie de $E = \mathbb{R}^2$ dans \mathbb{R} par $f(x, y) = x^2 + y^2$ est continue, strictement convexe et croissante à l'infini. L'ensemble K qui peut aussi être défini par : $K = \{(x, y) \in \mathbb{R}^2; g(x, y) \leq 0\}$, avec $g(x, y) = 1 - x - y$ est convexe et fermé. Par le théorème 3.34 page 129, il y a donc existence et unicité de la solution du problème (3.29). Appliquons le théorème de Kuhn-Tucker pour la détermination de cette solution. On a :

$$\nabla g(x, y) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \text{ et } \nabla f(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}.$$

Il existe donc $\lambda \in \mathbb{R}_+$ tel que :

$$\begin{cases} 2x - \lambda = 0, \\ 2y - \lambda = 0, \\ \lambda(1 - x - y) = 0, \\ 1 - x - y \leq 0, \\ \lambda \geq 0. \end{cases}$$

Par la troisième équation de ce système, on déduit que $\lambda = 0$ ou $1 - x - y = 0$. Or si $\lambda = 0$, on a $x = y = 0$ par les première et deuxième équations, ce qui est impossible en raison de la quatrième. On en déduit que $1 - x - y = 0$, et donc, par les première et deuxième équations, $x = y = \frac{1}{2}$.

Exercice 3.6 page 147 (Exemple d'opérateur de projection)

2. Soit p_K l'opérateur de projection définie à la proposition 3.44 page 133, il est facile de montrer que, pour tout $i = 1, \dots, N$:

$$\begin{aligned} (p_K(y))_i &= y_i & \text{si } y_i \in [\alpha_i, \beta_i], \\ (p_K(y))_i &= \alpha_i & \text{si } y_i < \alpha_i, \\ (p_K(y))_i &= \beta_i & \text{si } y_i > \beta_i, \end{aligned} \quad \text{ce qui entraîne}$$

$$(p_K(y))_i = \max(\alpha_i, \min(y_i, \beta_i)) \text{ pour tout } i = 1, \dots, N.$$