

Chapitre II: Optimisation

Introduction

Soit $N \in \mathbb{N}^*$. L'objectif de ce chapitre est de présenter des arguments théoriques et des méthodes pratiques afin de résoudre le problème de la minimisation d'une fonctionnelle F sur l'espace $\mathbb{R}^N$, ou sur un sous-ensemble de cet espace.

La méthode générale de résolution d'un tel problème consiste d'abord à vérifier que celui-ci fait sens, à savoir que la fonctionnelle F est bien définie et minorée sur l'ensemble X sur lequel nous cherchons à la minorer. Il s'agit ensuite d'étudier l'existence d'un possible minimiseur x_* , soit d'un élément de X tel que :

$$\forall x \in X, F(x_*) \leq F(x).$$

Il est étonnamment permis de s'intéresser à l'unicité de ce minimiseur, mais, en pratique, il s'agit surtout de déterminer la valeur de ce minimiseur. Cela passe par l'écriture des équations d'Euler-Lagrange pour ce minimiseur, puis par leur résolution théorique et numérique.

Dans la suite, nous présentons des résultats théoriques et numériques pour ces différentes étapes :

- la construction des minimiseurs,
- leur caractérisation différentielle,
- leur calcul pratique.

En particulier, nous insistons sur les méthodes de gradient qui permettent de résoudre des équations linéaires ou non linéaires en se basant sur l'appareil précédent.

I Construction de minimiseurs

Soit X , un sous-ensemble de $\mathbb{R}^N$ et $E: X \rightarrow \mathbb{R}$, une fonctionnelle définie sur l'ensemble X . Considérons le problème de minimisation suivant:

$$I = \inf_{x \in X} (E(x)).$$

La résolution de ce problème passe par les étapes suivantes:

- établir que l'infimum I est fini,
- considérer une suite minimisante $(x_n)_{n \in \mathbb{N}}$, soit une suite telle que:

$$\forall n \in \mathbb{N}, x_n \in X \text{ et } E(x_n) \xrightarrow{n \rightarrow +\infty} I,$$

et établir l'existence d'une sous-suite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ et d'un élément $x_* \in X$ tel que:

$$x_{\varphi(n)} \xrightarrow{n \rightarrow +\infty} x_*.$$

- vérifier que:

$$E(x_*) \leq \lim_{n \rightarrow +\infty} E(x_{\varphi(n)}),$$

auquel cas x_* sera un minimiseur de la fonctionnelle E sur l'ensemble X .

Cette stratégie est très générale. Elle s'applique également en dimension infinie même si certaines de ces étapes en particulier la preuve de l'existence d'une limite x_* à une sous-suite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ est alors plus compliquée.

Chaque étape peut de plus poser problème:

- il existe des fonctionnelles E non minorées,
- il existe des suites $(x_n)_{n \in \mathbb{N}}$ sans valeurs d'adhérence dans l'ensemble X ,
- il existe des fonctionnelles E qui ne sont pas semi-continues inférieures.

Il n'y a donc aucune garantie de succès à appliquer cette démarche.

Il est néanmoins facile de fournir des conditions sous lesquelles elle va fonctionner.

Définition: Soit $x_0 \in X$. La fonctionnelle F est semi-continue inférieure au point x_0 si: $F(x_0) \leq \liminf_{x \rightarrow x_0} F(x)$.

Exemples: (i) Si F est continue en un point $x_0 \in X$, alors, elle est semi-continue inférieure en ce point.

(ii) Soit $\forall x \in \mathbb{R}_+$, $f(x) = 1$ si $x > 0$, 0 sinon. La fonction f est semi-continue inférieure sur $\mathbb{R}_+$.

Théorème: Soit X , un sous-ensemble compact non vide de $\mathbb{R}^N$ et $F: X \rightarrow \mathbb{R}$, une fonctionnelle semi-continue inférieure sur X . Alors, il existe un élément $x_* \in X$ tel que:

$$F(x_*) = \min_{x \in X} F(x)$$

Preuve:

Soit $I = \inf_{x \in X} F(x)$, et $(x_n)_{n \in \mathbb{N}}$, une suite de X telle que:

$$F(x_n) \rightarrow I.$$

Par compacité de l'ensemble X , suite à extraire, il existe $x_* \in X$ tel que:

$$x_n \rightarrow x_*$$

Par semi-continuité inférieure, il vient:

$$F(x_*) \leq \liminf_{n \rightarrow +\infty} F(x_n) = I \Rightarrow F(x_*) = I.$$

Ce résultat demeure vrai pour des fonctionnelles convexes lorsque l'ensemble X n'est plus nécessairement compact.

Définition: Soit $F: X \rightarrow \mathbb{R}$. La fonctionnelle F est coercive sur l'ensemble X si: $\lim_{\|x\| \rightarrow +\infty, x \in X} F(x) = +\infty$.

Théorème: Soit X , un sous-ensemble fermé non vide de $\mathbb{R}^N$ et $F: X \rightarrow \mathbb{R}$, une fonctionnelle semi-continue inférieure et coercive sur X . Alors, il existe un élément $x_* \in X$ tel que:

$$F(x_*) = \min_{x \in X} F(x).$$

Preuve:

Soit $I = \inf_{x \in X} F(x)$. Comme $F(x) \xrightarrow{\|x\|_2 \rightarrow +\infty} +\infty$, il existe un nombre

$R > 0$ tel que:

$$\forall x \in X, \|x\| > R \Rightarrow F(x) \geq I + 1,$$

De sorte que:

$$I = \inf_{x \in X \cap B_F(0, R)} F(x).$$

Comme $X \cap B_F(0, R)$ est compact, le théorème précédent permet de conclure.

Exemple: Problème des moindres carrés

Soit $n \geq N$. Considérons une matrice $A \in \mathcal{M}_{n, N}(\mathbb{R})$ et un vecteur $b \in \mathbb{R}^n$. Le problème des moindres carrés associé à la matrice A et au vecteur b consiste à minimiser sur l'ensemble $\mathbb{R}^N$ la fonctionnelle:

$$\forall x \in \mathbb{R}^N, J(x) = \|Ax - b\|_2^2,$$

où $\|\cdot\|_2$ désigne la norme euclidienne canonique de $\mathbb{R}^n$.

Théorème: Soit $A \in \mathcal{M}_{n, N}(\mathbb{R})$ et $b \in \mathbb{R}^n$. Si $\text{Ker}(A) = \{0\}$, alors:

$$\exists x_* \in \mathbb{R}^N \text{ c.q. } J(x_*) = \inf_{x \in \mathbb{R}^N} J(x) = \|A(b)\|_2^2.$$

Preuve:

La fonctionnelle J est bien définie et continue, donc semi-continue inférieurement sur $\mathbb{R}^N$. De plus, elle s'écrit:

$$\forall x \in \mathbb{R}^N, J(x) = \langle {}^t A A x, x \rangle_2 = 2 \langle {}^t A(b), x \rangle_2 + \|b\|_2^2.$$

Si $\text{Ker}(A) = \{0\}$, la matrice ${}^t A A$ est symétrique définie positive de sorte qu'il existe un nombre $\lambda > 0$ tel que:

$$\forall x \in \mathbb{R}^N, \langle {}^t A A x, x \rangle_2 \geq \lambda \|x\|_2^2$$

Il vient alors:

$$\forall x \in \mathbb{R}^N, J(x) \geq \lambda \|x\|_2^2 - 2 \|{}^t A(b)\|_2 \|x\|_2 + \|b\|_2^2$$

$$\Rightarrow J(x) \xrightarrow{\|x\|_2 \rightarrow +\infty} +\infty.$$

La fonctionnelle J est donc coercive, et l'existence du x_* résulte du théorème précédent.

Remarque: Ce résultat reste vrai lorsque $\text{Ker}(A) \neq \{0\}$.

II Caractérisation différentielle d'un minimiseur

Soit X , un sous-ensemble de $\mathbb{R}^N$ et $F: X \rightarrow \mathbb{R}$, une fonctionnelle définie sur l'ensemble X .

1. Caractérisation différentielle d'un minimum local

Définition: Soit $x_* \in X$. La fonctionnelle F a un minimum local au point x_* , ou le point x_* est un minimiseur local de la fonctionnelle F ssi il existe un nombre strictement positif ρ tel que:

- (i) $\forall y \in \mathbb{R}^N, \|x_* - y\| < \rho \Rightarrow y \in X,$
- (ii) $\forall y \in \mathbb{R}^N, \|x_* - y\| < \rho \Rightarrow F(y) \geq F(x_*).$

Exemple: La fonction cosinus a des minima locaux en tout point de la forme $(2h+1)\pi$, $h \in \mathbb{Z}$.

Lemme: Soit $x_* \in X$. Supposons que la fonctionnelle F a un minimum local au point x_* . Alors:

- (i) Si F est de classe $\mathcal{C}^1$ en x_* , alors:
$$dF(x_*) = 0.$$
- (ii) Si F est de classe $\mathcal{C}^2$ en x_* , alors:
$$\forall y \in \mathbb{R}^N, d^2F(x_*)(y, y) \geq 0.$$

Preuve:

- (i) Soit $\rho > 0$ tel que:

$$B_0(x_*, \rho) \subset X \text{ et } \forall y \in B_0(x_*, \rho), F(y) \geq F(x_*).$$

Si $y \in \mathbb{R}^N$ et $0 < t < \rho$, alors,

$$x_* + ty \in B_0(x_*, \rho) \Rightarrow F(x_* + ty) \geq F(x_*) \Rightarrow \frac{F(x_* + ty) - F(x_*)}{t} \geq 0$$

à la limite $t \rightarrow 0$, il vient:

$$dE(x_*) (z) \geq 0.$$

De même, nous avons:

$$dE(x_*) (-z) \geq 0,$$

De sorte que:

$$\forall z \in \mathbb{R}^N, dE(x_*) (z) = 0.$$

(ii) Lorsque la fonctionnelle E est de classe C^2 en x_* , il découle de la formule de Taylor-Young que:

$$E(x_* + tz) = E(x_*) + t dE(x_*) (z) + \frac{t^2}{2} d^2E(x_*) (z, z) + o(t^2),$$

De sorte que:

$$\frac{t^2}{2} d^2E(x_*) (z, z) + o(t^2) \geq 0.$$

à la limite $t \rightarrow 0$, il vient:

$$d^2E(x_*) (z, z) \geq 0.$$

Les conditions nécessaires précédentes ne sont pas suffisantes.

Exemple: Soit $\forall x \in \mathbb{R}^3$, $f(x) = x^3$. La fonction f est de classe C^2 sur $\mathbb{R}$ et elle vérifie: $f'(0) = f''(0) = 0$. Cependant, elle n'a pas de minimum local en 0.

Le théorème suivant donne des conditions suffisantes pour l'existence d'un minimum local.

Théorème: Soit $x_* \in X$. Supposons qu'il existe un nombre strictement positif ρ tel que

$$\forall y \in \mathbb{R}^N, \|y - x_*\| \leq \rho \Rightarrow y \in X$$

Considérons une fonctionnelle E de classe C^2 sur la boule ouverte $B_0(x_*, \rho)$ et supposons que:

$$(i) \forall y \in B_0(x_*, \rho), \forall z \in \mathbb{R}^N, dE(y) (z, z) \geq 0,$$

$$(ii) dE(x_*) = 0.$$

Alors la fonctionnelle E a un minimum local en x_* .

Preuve:

Par le théorème de Taylor avec reste intégral, il vient:

$$\forall y \in B_0(x_*, \rho), \quad E(y) = E(x_*) + \int_0^1 (1-t) d^2E(x_* + t(y-x_*))(y-x_*, y-x_*) dt \\ \geq E(x_*).$$

Corollaire: Soit $x_* \in X$. Supposons qu'il existe un nombre strictement positif ρ tel que

$$\forall y \in \mathbb{R}^n, \quad \|y - x_*\| < \rho \Rightarrow y \in X$$

Considérons une fonctionnelle E de classe C^2 sur la boule ouverte

$B_0(x_*, \rho)$ et supposons que:

$$(i) \quad dE(x_*) = 0,$$

$$(ii) \quad \exists \lambda \in \mathbb{R}_+^*, \text{ t.q. } \forall y \in \mathbb{R}^n, \quad d^2E(x_*)(y, y) \geq \lambda \|y\|^2.$$

Alors la fonctionnelle E a un minimum local en x_* .

Preuve:

Comme la fonctionnelle E est de classe C^2 sur $B_0(x_*, \rho)$, il existe un nombre strictement positif $\delta < \rho$ tel que:

$$\forall y \in B(x_*, \delta), \quad \forall z \in \mathbb{R}^n, \quad d^2E(y)(z, z) \geq 0.$$

Le corollaire découle alors du théorème précédent.

En pratique, les résultats précédents rendent possible le calcul des possibles minima locaux une fois leur existence établie. Et noter que ces résultats ne garantissent en aucun cas que les minima locaux ainsi déterminés sont globaux. Par essence, une information locale telle que des conditions sur les différentielles premières et secondes en un point x_* ne peut fournir une information globale sur la fonctionnelle E .

Exemple: Problème des moindres carrés

Soit $n \in \mathbb{N}^*$. Considérons une matrice $A \in \mathcal{M}_{n,n}(\mathbb{R})$ et un vecteur $b \in \mathbb{R}^n$, et notons J la fonctionnelle définie par:

$$\forall x \in \mathbb{R}^n, \quad J(x) = \|Ax - b\|_2^2.$$

Théorème: Soit $A \in M_{n,n}(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonctionnelle J a un minimiseur x_* sur $\mathbb{R}^n$:

$$\exists x_* \in \mathbb{R}^n \text{ s.t. } J(x_*) = \min_{x \in \mathbb{R}^n} J(x).$$

Preuve:

La fonctionnelle J est bien définie et de classe C^2 sur $\mathbb{R}^n$, et elle vérifie:

$$\forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, \begin{aligned} dJ(x)(y) &= {}^t A A(x) - {}^t A(b), y \\ {}^t [d^2 J(x)](y, y) &= {}^t A A(y), y \end{aligned}$$

En particulier,

$$\forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n, d^2 J(x)(y, y) \geq 0,$$

De sorte que s'il existe un vecteur $x_* \in \mathbb{R}^n$ tel que $dJ(x_*) = 0$, alors, par le théorème précédent, ce vecteur sera un minimiseur de la fonctionnelle J sur $\mathbb{R}^n$. Il s'agit donc de déterminer un vecteur x_* tel que:

$${}^t A A(x_*) = {}^t A(b),$$

autrement dit de prouver que:

$$\text{Im}({}^t A A) \subset \text{Im}({}^t A A)$$

Comme $\text{Im}({}^t A A) \subset \text{Im}({}^t A)$, ces deux ensembles sont égaux s'ils ont la même dimension. Puisque

$$\forall x \in \mathbb{R}^n, \|A(x)\|_2^2 = \langle {}^t A A(x), x \rangle_2,$$

nous savons que:

$$\text{Ker}({}^t A A) = \text{Ker}(A).$$

Il s'ensuit que:

$$\begin{aligned} \dim(\text{Im}({}^t A A)) &= \dim(\text{Im}(A)) = n - \dim(\text{Ker}(A)) = n - \dim(\text{Ker}({}^t A A)) \\ &= \dim(\text{Im}({}^t A A)), \end{aligned}$$

De sorte que $\text{Im}({}^t A A) = \text{Im}({}^t A)$, ce qui achève la preuve de l'existence d'un minimiseur x_* .

2. Brise en compte de la convexité

Supposons désormais que l'ensemble X est convexe c'est-à-dire que :

$$\forall (x, y) \in X^2, \forall t \in [0, 1], tx + (1-t)y \in X.$$

La caractérisation différentielle d'un minimiseur x_* de la fonctionnelle F sur l'ensemble X est alors la suivante.

Lemme : Soit $x_* \in X$. Supposons que la fonctionnelle F est de classe C^1 en x_* et qu'elle a un minimum en x_* . Alors :

$$\forall y \in X, dF(x_*)(y - x_*) \geq 0.$$

Preuve :

Si $y \in X$, alors :

$$\begin{aligned} \forall t \in [0, 1], tx + (1-t)x_* \in X &\Rightarrow F(tx + (1-t)x_*) \geq F(x_*) \\ &\Rightarrow \lim_{t \rightarrow 0} \frac{dF(x_*)(y - x_*)}{t} \geq 0. \end{aligned}$$

Cette condition nécessaire est suffisante dans le cas des fonctionnelles convexes.

Définition : Soit $F: X \rightarrow \mathbb{R}$. La fonctionnelle F est convexe sur l'ensemble convexe X si :

$$\forall (x, y) \in X^2, \forall t \in [0, 1], F(tx + (1-t)y) \leq tF(x) + (1-t)F(y).$$

Elle est strictement convexe sur l'ensemble convexe X si :

$$\forall (x, y) \in X^2, \text{ avec } x \neq y, \forall t \in]0, 1[, F(tx + (1-t)y) < tF(x) + (1-t)F(y).$$

Exemple : Soit $n \in \mathbb{N}^*$, $A \in M_{n,n}(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonctionnelle J définie par :

$$\forall x \in \mathbb{R}^n, J(x) = \|Ax - b\|_2^2,$$

est convexe sur $\mathbb{R}^n$. Elle est strictement convexe si le noyau de la matrice A est réduit au singleton $\{0\}$.

Lemme : Soit $x_* \in X$. Supposons que la fonctionnelle F est de classe C^1 et convexe sur l'ensemble convexe X .

(i) La fonctionnelle F a un minimum en x_* ssi:

$$\forall y \in X, dF(x_*)(y - x_*) \geq 0.$$

(ii) Si la fonctionnelle F est strictement convexe, alors, elle admet au plus un minimiseur x_* , et le minimum associé est strict.

(iii) Si l'ensemble X est ouvert, alors, la fonctionnelle F a un minimum en x_* ssi: $dF(x_*) = 0$.

Preuve:

(i) Il suffit par le théorème précédent de démontrer l'implication réciproque. Comme F est convexe, nous savons que:

$$\forall (x, y) \in X^2, \forall t \in [0, 1], F(x(1-t) + ty) \leq (1-t)F(x) + tF(y)$$

$$\stackrel{t \rightarrow 0}{\Rightarrow} F(y) - F(x) \geq \frac{1}{t} (F(x + t(y-x)) - F(x))$$

$$\stackrel{t \rightarrow 0}{\Rightarrow} F(y) \geq F(x) + (F'(x))(y-x)$$

Sous la condition ci-dessus, il vient donc:

$$\forall y \in X, F(y) \geq F(x_*),$$

De sorte que F a un minimum en x_* .

(ii) Si F a deux minimiseurs x_* et y_* sur X , alors,

$$F\left(\frac{x_* + y_*}{2}\right) \leq \frac{1}{2} (F(x_*) + F(y_*)) = \text{dit-on } F(x_*),$$

ce qui est absurde! Les minimiseurs associés sont donc stricts.

(iii) Si X est un ouvert, alors, il existe un nombre strictement positif ρ tel que:

$$B_0(x_*, \rho) \subset X$$

En particulier, il vient:

$$\forall y \in B_0(x_*, \rho), dF(x_*)(y - x_*) \geq 0$$

$$\Rightarrow dF(x_*)(x_* - y) \geq 0$$

$$\Rightarrow dF(x_*)(y - x_*) = 0$$

De sorte que par linéarité:

$$dF(x_*) = 0$$

Exemple: Problème des moindres carrés

Soit $n \geq N$. Considérons une matrice $A \in \mathcal{M}_{n,N}(\mathbb{R})$ et un vecteur $b \in \mathbb{R}^n$, et notons J la fonctionnelle définie par:

$$\forall x \in \mathbb{R}^N, J(x) = \|Ax - b\|_2^2.$$

Théorème: Soit $A \in \mathcal{M}_{n,N}(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonctionnelle J a un unique minimum $x_* \in \mathbb{R}^N$ si: $\text{Ker}(A) = \{0\}$.

Preuve:

Immédiate

3. Caractérisation différentielle d'un minimum sous contraintes

Supposons maintenant que l'ensemble X est défini par un ensemble de contraintes. Plus précisément, considérons un sous-ensemble ouvert Ω de $\mathbb{R}^N$ et des fonctions $(\varphi_i)_{1 \leq i \leq n}$ définies sur l'ouvert Ω et à valeurs réelles (avec $n \leq N$), et introduisons l'ensemble des contraintes $(\varphi_i)_{1 \leq i \leq n}$ comme l'ensemble:

$$X := \{x \in \Omega \text{ t.q. } \forall 1 \leq i \leq n, \varphi_i(x) = 0\}.$$

Définition: Soit $x_* \in X$. La fonctionnelle F a un minimum sous les contraintes

$(\varphi_i)_{1 \leq i \leq n}$ au point x_* si:

$$\forall y \in X, F(y) \geq F(x_*).$$

Exemple: La fonctionnelle F définie par:

$$\forall (x_1, x_2) \in \mathbb{R}^2, F(x_1, x_2) = x_1^2 + x_2^2,$$

a un minimum sous la contrainte:

$$\varphi(x_1, x_2) = x_1^2 + x_2^2 - 2,$$

au point $x_* = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, et la valeur de ce minimum est égale à $-\sqrt{2}$. En effet, la fonctionnelle G définie par:

$$\forall x_2 \in (-2, 2), G(x_2) = x_2 - \sqrt{2 - x_2^2},$$

atteint son minimum $-\sqrt{2}$ en $x_* = -\frac{1}{\sqrt{2}}$.

Théorème des extrema liés: Soit $x_0 \in X$. Supposons que les fonctions $(f_i)_{1 \leq i \leq n}$ sont de classe C^1 sur l'ouvert Ω et que les différentielles $(df_i(x_0))_{1 \leq i \leq n}$ sont linéairement indépendantes. Si la fonctionnelle F a un minimum sous les contraintes $(f_i)_{1 \leq i \leq n}$ au point x_0 et si elle est de classe C^1 en ce point, alors, il existe des nombres réels $(\lambda_i)_{1 \leq i \leq n}$ tels que:

$$dF(x_0) = \sum_{j=1}^n \lambda_j df_j(x_0).$$

Les nombres $(\lambda_j)_{1 \leq j \leq n}$ sont les multiplicateurs de Lagrange associés au minimiseur sous contraintes x_0 .

Preuve:

Soit $\forall x \in \Omega$, $F(x) = (f_1(x), \dots, f_n(x))$. La fonction F est de classe C^1 sur l'ouvert Ω et sa matrice jacobienne $J(F)$ est égale à:

$$\forall x \in \Omega, J(F)(x) = (df_i f(x))_{1 \leq i \leq n}$$

Comme les différentielles $(df_i(x_0))_{1 \leq i \leq n}$ sont linéairement indépendantes, la matrice $J(F)(x_0)$ est de rang n , et quitte à permuter les variables $(x_j)_{1 \leq j \leq n}$, nous pouvons supposer que la matrice:

$$A = (df_j f_i(x_0))_{1 \leq i, j \leq n}$$

est inversible. Comme $F(x_0) = 0$, il résulte alors du théorème des fonctions implicites qu'il existe un voisinage V_1 de $x_0^1 = ((x_0)_1, \dots, (x_0)_n)$ dans $\mathbb{R}^n$ et un voisinage V_2 de $x_0^2 = ((x_0)_{n+1}, \dots, (x_0)_N)$ dans $\mathbb{R}^{N-n}$, et une fonction $\Theta \in C^1(V_1, \mathbb{R}^{N-n})$ telle que:

$$\forall x = (x^1, x^2) \in V_1 \times V_2, F(x, x^2) = 0 \Leftrightarrow x^2 = \Theta(x^1).$$

Notons en particulier que:

$$\forall x^1 \in V_1, F(\Theta(x^1), x^1) = 0,$$

ce qui implique qu'en point x_0 ,

$$\forall 1 \leq h \leq N-n, d_h F(x_0) + \sum_{j=1}^n d_h \Theta_j(x_0^1) df_j f(x_0) = 0,$$

ce qui s'écrit aussi sous la forme:

$$\forall 1 \leq h \leq N-n, d_h F(x_0) = -A(d_h \Theta(x_0^1)),$$

ce qui induit que:

$$\forall 1 \leq h \leq N-n, d_h \Theta(x_0^1) = -A^{-1}(d_h F(x_0)).$$

D'autre part, il vient :

$$\forall x = (x_1, x_2) \in X \cap U_1 \times U_2, x_2 = \theta(x_1).$$

Si la fonctionnelle F a un minimum sous les contraintes $(\varphi_i)_{1 \leq i \leq n}$ au point x_* , alors :

$$\forall y_2 \in U_2, F(\theta(y_2), y_2) \geq F(\theta(x_1^*), x_1^*).$$

L'application G définie par :

$$\forall y_2 \in U_2, G(y_2) = F(\theta(y_2), y_2),$$

a un minimum local au point x_*^2 . Comme G est de classe $\mathcal{C}^1$ en x_*^2 , il vient :

$$dG(x_*^2) = 0 \Leftrightarrow \forall 1 \leq h \leq N-1, \partial_{h+1} F(x_*) + \sum_{j=1}^n \lambda_j \partial_j \theta(x_*^2) \partial_j F(x_*) = 0.$$

Si nous notons :

$$A^{-1} = (\alpha_{ij})_{1 \leq i, j \leq n},$$

alors, il vient :

$$\begin{aligned} \forall 1 \leq h \leq n-N, \partial_{h+1} F(x_*) &= \sum_{j=1}^n \sum_{c=1}^n \alpha_{j,c} \partial_{h+1} \Phi_c(x_*) \partial_j F(x_*) \\ &= \sum_{c=1}^n \lambda_c \partial_{h+1} \Phi_c(x_*), \end{aligned}$$

où $\forall 1 \leq c \leq n, \lambda_c = \sum_{j=1}^n \alpha_{j,c} \partial_j F(x_*)$. Notons enfin que :

$$\begin{aligned} \forall 1 \leq h \leq n, \sum_{c=1}^n \lambda_c \partial_h \Phi_c(x_*) &= \sum_{j=1}^n \sum_{c=1}^n \alpha_{j,c} \partial_h \Phi_c(x_*) \partial_j F(x_*) \\ &= \sum_{j=1}^n \sum_{c=1}^n \alpha_{j,c} \alpha_{c,h} \partial_j F(x_*) \\ &= \partial_h F(x_*), \end{aligned}$$

De sorte que nous avons bien établi que :

$$dF(x_*) = \sum_{j=1}^n \lambda_j d\Phi_j(x_*).$$

Exemple : Soit $\forall (x_1, x_2) \in \mathbb{R}^2, F(x_1, x_2) = x_1 + x_2$ et $\varphi(x_1, x_2) = x_1^2 + x_2^2 - 1$.

Les fonctionnelles F et φ sont de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ et elles vérifient

$$\nabla F(x_1, x_2) = e_1 + e_2 \text{ et } \nabla \varphi(x_1, x_2) = 2x_1 e_1 + 2x_2 e_2,$$

$$\text{où } e_1 = (1, 0) \text{ et } e_2 = (0, 1).$$

Si la fonctionnelle F a un minimum sous la contrainte φ au point

$x_* = (x_*^1, x_*^2)$, alors, il existe un nombre réel λ tel que :

$$\nabla F(x_*) = \lambda \nabla \varphi(x_*) \Leftrightarrow \begin{cases} 1 = 2\lambda x_*^1 \\ 1 = 2\lambda x_*^2 \end{cases}$$

Le multiplicateur de Lagrange λ est donc non nul, et :

$$x^* = \left(\frac{1}{2\lambda}, \frac{1}{2\lambda} \right).$$

Il vient donc :

$$\varphi(x^*) = \frac{1}{2\lambda^2} - 1.$$

Puisque $\varphi(x^*) = 0$, il s'ensuit que $\lambda = \pm \frac{1}{\sqrt{2}}$; le point $x^* = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)$ est donc le seul candidat à la minimisation de la fonctionnelle E sous la contrainte φ . Pour conclure, il s'agit alors de vérifier que ce point est bien un minimiseur de la fonctionnelle E sous la contrainte φ , ce que nous avons fait plus haut.

La condition nécessaire précédente n'est pas suffisante. Dans l'exemple précédent, le point $y_* = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right)$ n'est pas un minimum de la fonctionnelle E sous la contrainte φ , mais son maximum. Cette condition ne détecte que les points critiques de la fonctionnelle E sous la contrainte φ , qui peuvent être un max, des minima (locaux) pour cette fonctionnelle.

En pratique, elle rend possible le calcul des possibles minima locaux une fois leur existence établie. Il note que ces résultats ne garantissent en aucun cas que les minima locaux ainsi déterminés sont globaux. Une information locale telle qu'une condition sur des différentielles premières en x_* ne peut pas essence fournir une information globale sur la fonctionnelle E .

Exemple: Fonctionnelle quadratique avec contrainte linéaire.

Soit $1 \leq n < N$. Considérons des matrices $A \in M_n(\mathbb{R})$ et $C \in M_{n,n}(\mathbb{R})$, et des vecteurs $b \in \mathbb{R}^N$ et $d \in \mathbb{R}^n$, et intéressons-nous au problème de minimisation de la fonctionnelle quadratique E définie par :

$$\forall x \in \mathbb{R}^N, E(x) = \frac{1}{2} \langle A(x), x \rangle_{\mathbb{R}} - \langle b, x \rangle_{\mathbb{R}},$$

sous la contrainte linéaire φ donné par :

$$\forall x \in \mathbb{R}^N, \varphi(x) = C(x) - d;$$

lorsque la matrice C est de rang n , le théorème des extrema

liés s'applique, et tout possible minimiseur x_* de la fonctionnelle F sous la contrainte Ψ satisfait l'équation:

$$\nabla F(x_*) = \sum_{j=1}^n \lambda_j \nabla \Psi_j(x_*),$$

$$\text{ou } \lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n. \text{ Lemme}$$

$$\forall x \in \mathbb{R}^n, \quad \nabla F(x) = \frac{1}{2} (A + {}^t A)(x) - b,$$

Et

$$\forall 1 \leq j \leq n, \quad \forall x \in \mathbb{R}^n, \quad \nabla \Psi_j(x) = (c_j, 0)_{x \in \mathbb{R}^n},$$

il s'agit donc de résoudre le système d'équations:

$$\begin{cases} \frac{1}{2} (A + {}^t A)(x^*) = b + {}^t L(\lambda), \\ {}^t L(\lambda) = d, \end{cases}$$

puis de vérifier si les points x^* ainsi obtenus sont un (des) minimiseur(s) de la fonctionnelle F sous la contrainte Ψ . Notons ici que ces deux fonctionnelles sont convexes, de sorte qu'en cas d'existence d'une solution x^* , ce sera bien un minimiseur de la fonctionnelle F sous la contrainte Ψ .

III Méthodes du gradient

Soit $F: \mathbb{R}^n \rightarrow \mathbb{R}$, une fonctionnelle de classe C^1 sur $\mathbb{R}^n$. Considérons le problème de minimisation de cette fonctionnelle sur $\mathbb{R}^n$ et supposons que ce problème a (au moins) une (unique) solution, à savoir qu'il existe un point $x_* \in \mathbb{R}^n$ tel que:

$$\forall y \in \mathbb{R}^n, \quad F(y) \geq F(x_*).$$

Il s'agit maintenant de déterminer la valeur du minimiseur x_* de manière au moins approchée. Les méthodes du gradient fournissent une solution à ce problème.

Ce sont des méthodes itératives. À partir d'un point initial $x_0 \in \mathbb{R}^n$, il s'agit de construire une suite de points $(x_n)_{n \in \mathbb{N}} \in (\mathbb{R}^n)^{\mathbb{N}}$ telle que:

$$\forall n \in \mathbb{N}, \quad F(x_{n+1}) \leq F(x_n),$$

l'espoir étant que la suite décroissante $(E(x_n))_{n \in \mathbb{N}}$ finisse par converger vers le minimum $E(x_*)$ et que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers le point x_* .

Afin de garantir la décroissance de la suite $(E(x_n))_{n \in \mathbb{N}}$, il est nécessaire de déterminer au voisinage de chaque point x_n au moins une direction de descente $d_n \in \mathbb{R}^N$ telle que, pour un pas $\alpha_n \in \mathbb{R}_+^*$, nous ayons l'inégalité, si possible stricte,

$$E(x_n + \alpha_n d_n) \leq E(x_n).$$

Le point x_{n+1} est alors choisi égal à :

$$x_{n+1} = x_n + \alpha_n d_n.$$

Les méthodes du gradient sont un cas particulier des méthodes de descente ainsi définies. Lorsque le pas α_n est assez petit, la différentiabilité de la fonction E au voisinage du point x_n induit que :

$$E(x_n + \alpha_n d_n) = E(x_n) + \alpha_n \langle \nabla E(x_n), d_n \rangle + o(\alpha_n).$$

Par l'inégalité de Cauchy-Schwarz, il vient :

$$|\langle \nabla E(x_n), d_n \rangle| \leq \|d_n\|_2 \|\nabla E(x_n)\|_2,$$

et cette inégalité est une égalité si :

$$d_n = \pm \nabla E(x_n).$$

La direction optimale de descente est donc donnée par l'opposé du gradient de la fonctionnelle E au point x_n . Les méthodes du gradient reposent sur ce choix de direction de descente :

$$\forall n \in \mathbb{N}, d_n = -\nabla E(x_n).$$

Elles varient cependant suivant la manière de choisir le pas α_n à chaque itération.

Néanmoins que la convergence de ces méthodes est en général sujette à caution. L'une des difficultés tient au choix du pas α_n , mais de manière générale, la principale difficulté résulte de la possible existence de minima locaux vers lesquels ces méthodes peuvent converger suivant le choix du point

initial x_0 . Notons qu'en pratique le calcul itéré du gradient de la fonctionnelle F peut aussi s'avérer fastidieux.

Voici néanmoins certaines conditions sous lesquelles les méthodes du gradient à pas constant, du gradient à pas optimal et du gradient conjugué sont convergentes.

1. Minimisation des fonctionnelles elliptiques

Définition: Une fonctionnelle $F \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R})$ est elliptique ssi il existe un nombre strictement positif λ tel que:

$$\forall (x, y) \in [\mathbb{R}^N]^2, \quad \langle \nabla F(y) - \nabla F(x), y - x \rangle_2 \geq \lambda \|y - x\|_2^2.$$

Exemple: Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$ et $b \in \mathbb{R}^N$. La fonctionnelle quadratique F définie par $\forall x \in \mathbb{R}^N, \quad F(x) = \frac{1}{2} \langle A(x), x \rangle_2 - \langle b, x \rangle_2$ est elliptique. En effet, si $\lambda_1 \in \mathbb{R}_+^*$ désigne la plus petite valeur propre de la matrice A , alors:

$$\begin{aligned} \forall (x, y) \in [\mathbb{R}^N]^2, \quad \langle \nabla F(x) - \nabla F(y), x - y \rangle_2 &= \langle A(x - y), x - y \rangle_2 \\ &\geq \lambda_1 \|x - y\|_2^2. \end{aligned}$$

Lemme: Soit $F \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R})$. Si F est une fonctionnelle elliptique, alors:

- (i) $\forall (x, y) \in [\mathbb{R}^N]^2, \quad F(y) - F(x) \geq \langle \nabla F(x), y - x \rangle_2 + \frac{\lambda}{2} \|y - x\|_2^2$, où $\lambda \in \mathbb{R}_+^*$ est le paramètre d'ellipticité de F .
- (ii) La fonctionnelle F est strictement convexe et concave.

Preuve:

(i) La formule de Taylor avec reste intégral permet en effet d'écrire:

$$\begin{aligned} \forall (x, y) \in [\mathbb{R}^N]^2, \quad F(y) - F(x) &= \int_0^1 \langle \nabla F(x + t(y-x)), y-x \rangle_2 dt \\ &= \langle \nabla F(x), y-x \rangle_2 + \int_0^1 \langle \nabla F(x + t(y-x)) - \nabla F(x), y-x \rangle_2 dt \\ &\geq \langle \nabla F(x), y-x \rangle_2 + \int_0^1 \lambda t \|y-x\|_2^2 dt, \end{aligned}$$

D'où la formule précédente.

(ii) Comme

$$F(x) + \langle \nabla F(x), y-x \rangle_2 + \frac{\lambda}{2} \|y-x\|_2^2 \geq F(x) - \|\nabla F(x)\|_2 \|y-x\|_2 + \frac{\lambda}{2} \|y-x\|_2^2$$

$\xrightarrow{\|y-x\|_2 \rightarrow +\infty} +\infty,$

il est clair que:

$$F(y) \xrightarrow{\|y-x\|_2 \rightarrow +\infty} +\infty,$$

On note que la fonctionnelle F est convexe. De plus, si $(x, y) \in (\mathbb{R}^n)^2$, avec $x \neq y$, et $0 \leq t \leq 1$, alors :

$$\begin{aligned} F(x) &\geq F(tx + (1-t)y) + (1-t) \langle \nabla F(tx + (1-t)y), x-y \rangle_2 + \frac{\lambda}{2} (1-t)^2 \|x-y\|_2^2 \\ \begin{cases} F(y) &\geq F(tx + (1-t)y) - t \langle \nabla F(tx + (1-t)y), x-y \rangle_2 + \frac{\lambda}{2} t^2 \|x-y\|_2^2 \end{cases} \\ \Rightarrow t F(x) + (1-t) F(y) &\geq F(tx + (1-t)y) + \frac{\lambda}{2} [t(1-t)^2 + (1-t)t^2] \|x-y\|_2^2 \\ &\geq F(tx + (1-t)y), \end{aligned}$$

D'où la stricte convexité de la fonctionnelle F .

Théorème: Soit X , un sous-ensemble convexe, fermé et non vide de $\mathbb{R}^n$. Si $F \in C^1(\mathbb{R}^n, \mathbb{R})$ est elliptique, alors, la fonctionnelle F possède un unique minimum $x_F^* \in X$ sur l'ensemble X qui est caractérisé par l'inéquation :

$$\forall y \in X, \langle \nabla F(x_F^*), y - x_F^* \rangle_2 \geq 0.$$

En particulier, la fonctionnelle F possède un unique minimum global x_F^* sur $\mathbb{R}^n$ qui est caractérisée par l'équation :

$$\nabla F(x_F^*) = 0.$$

Preuve:

Elle découle de manière immédiate des énoncés précédents.

Les fonctionnelles elliptiques sont donc particulièrement adaptées aux méthodes de gradient, et ce sont ainsi pour ces fonctionnelles que nous pourrions établir la convergence de ces méthodes.

En pratique, il est facile de déterminer si une fonctionnelle est elliptique grâce au lemme suivant.

Lemme: Soit $F \in \mathcal{C}^2(\mathbb{R}^N, \mathbb{R})$. La fonctionnelle F est elliptique sur $\mathbb{R}^N$ si il existe un nombre strictement positif λ tel que :

$$\forall (x, y) \in (\mathbb{R}^N)^2, \quad d^2F(x)(y, y) \geq \lambda \|y\|_2^2.$$

Preuve:

Si la fonctionnelle F est elliptique, alors, il existe un nombre strictement positif λ tel que :

$$\forall (x, y) \in (\mathbb{R}^N)^2, \forall t \in \mathbb{R}, \langle \nabla F(x+ty) - \nabla F(x), ty \rangle \geq \lambda t^2 \|y\|_2^2$$

$$\xrightarrow{t \rightarrow 0} d^2F(x)(y, y) \geq \lambda \|y\|_2^2.$$

Réciproquement, si

$$\forall (x, y) \in (\mathbb{R}^N)^2, \quad d^2F(x)(y, y) \geq \lambda \|y\|_2^2,$$

alors,

$$\langle \nabla F(y) - \nabla F(x), y-x \rangle_2 = \int_0^1 d^2F(x+t(y-x))(y-x, y-x) dt$$

$$\geq \lambda \|y-x\|_2^2.$$

Exemple: Soit $A \in \mathcal{M}_n(\mathbb{R})$ et $b \in \mathbb{R}^n$. Soit fonctionnelle quadratique F définie par :

$$\forall x \in \mathbb{R}^n, \quad F(x) = \frac{1}{2} \langle A(x), x \rangle_2 - \langle b, x \rangle_2,$$

est elliptique si : $A+A^t \in \mathcal{S}_n^{++}(\mathbb{R})$. En effet, elle est de classe $\mathcal{C}^2$ sur $\mathbb{R}^n$ et :

$$\forall x \in \mathbb{R}^n, \quad d^2F(x) = \frac{1}{2} (A+A^t),$$

De sorte que l'inégalité précédente est vraie si : $A+A^t \in \mathcal{S}_n^{++}(\mathbb{R})$.

2. Méthode du gradient à pas constant

Afin de résoudre le problème de minimisation de la fonctionnelle F sur l'espace $\mathbb{R}^n$, la méthode de gradient à pas constant consiste à fixer un pas α strictement positif, puis à construire une suite $(x_n)_{n \in \mathbb{N}}$ suivant les formules de récurrence :

$$x_0 \in \mathbb{R}^n \text{ et } \forall n \in \mathbb{N}, \quad x_{n+1} = x_n - \alpha \nabla F(x_n).$$

Dans le cas où cette suite est convergente, sa limite x_* vérifiera l'équation d'Euler-Lagrange:

$$\nabla F(x_*) = 0,$$

Et il sera alors permis d'espérer que la limite x_* soit un minimiseur de la fonctionnelle F sur $\mathbb{R}^N$.

Le résultat suivant garantit que cette stratégie fonctionne dans le cas où la fonctionnelle F est elliptique, et lorsque le pas α est bien choisi.

Théorème: Soit $F \in C^2(\mathbb{R}^N, \mathbb{R})$ une fonctionnelle elliptique c'est-à-dire telle qu'il existe un nombre strictement positif λ pour lequel :

$$\forall (x, y) \in (\mathbb{R}^N)^2, \langle \nabla F(y) - \nabla F(x), y - x \rangle \geq \lambda \|x - y\|_2^2.$$

Alors $x_* \in \mathbb{R}^N$ l'unique minimiseur de la fonctionnelle F sur $\mathbb{R}^N$ et supposons qu'il existe un nombre strictement positif η tel que:

$$\forall (x, y) \in (\mathbb{R}^N)^2, \|\nabla F(y) - \nabla F(x)\|_2 \leq \eta \|x - y\|_2.$$

Si le pas α vérifie la condition:

$$0 < \alpha \leq \frac{2\lambda}{\eta^2},$$

alors, il existe un nombre strictement positif $\beta < 1$ tel que toute suite $(x_n)_{n \in \mathbb{N}}$ définie par le choix d'un point $x_0 \in \mathbb{R}^N$ et la formule de récurrence:

$$\forall n \in \mathbb{N}, x_{n+1} = x_n - \alpha \nabla F(x_n),$$

vérifie:

$$\forall n \in \mathbb{N}, \|x_n - x_*\|_2 \leq \beta^n \|x_0 - x_*\|_2.$$

En particulier, la suite $(x_n)_{n \in \mathbb{N}}$ converge vers le unique minimiseur x_* de la fonctionnelle F sur $\mathbb{R}^N$.

Preuve:

Comme $\nabla F(x_*) = 0$, il vient:

$$\forall n \in \mathbb{N}, x_{n+1} - x_* = x_n - x_* - \alpha (\nabla F(x_n) - \nabla F(x_*)),$$

De sorte que:

$$\begin{aligned}\|x_{n+2} - x_*\|_2^2 &= \|x_n - x_*\|_2^2 - 2\alpha \langle x_n - x_*, \nabla F(x_n) - \nabla F(x_*) \rangle + \alpha^2 \|\nabla F(x_n) - \nabla F(x_*)\|_2^2 \\ &\leq (1 - 2\alpha \lambda + \alpha^2 \mu^2) \|x_n - x_*\|_2^2.\end{aligned}$$

Puisque $0 < \alpha < \frac{2\lambda}{\mu^2}$, il est clair que:

$$1 - 2\alpha \lambda + \alpha^2 \mu^2 = 1 - \alpha(2\lambda + \alpha \mu^2) < 1,$$

① on l'existence d'un nombre $0 < \beta < 1$ tel que:

$$\forall n \in \mathbb{N}, \|x_{n+2} - x_*\|_2 \leq \beta \|x_n - x_*\|_2 \leq \beta^{n+2} \|x_0 - x_*\|_2.$$

et la limite $n \rightarrow +\infty$, il vient en particulier:

$$x_n \xrightarrow[n \rightarrow +\infty]{} x_*.$$

Dans le cas d'une fonctionnelle elliptique F qui satisfait aux hypothèses du théorème précédent, la méthode du gradient a donc une convergence rapide lorsque le pas α est suffisamment petit. Notons que, dans ce cas, le choix du pas α peut aussi être variable, et que le choix de toute suite $(\alpha_n)_{n \in \mathbb{N}}$ de pas qui satisfait la condition:

$$\forall n \in \mathbb{N}, 0 < \alpha_n < \frac{2\lambda}{\mu^2},$$

conduit au même résultat. Cette remarque conduit naturellement à se poser la question du choix optimal du pas, et donc à la méthode du gradient à pas optimal du paragraphe suivant.

Exemple: Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonctionnelle quadratique F définie par:

$$\forall x \in \mathbb{R}^n, F(x) = \frac{1}{2} \langle A(x), x \rangle_2 - \langle b, x \rangle_2,$$

est alors elliptique et elle satisfait les deux inégalités:

$$\forall (x, y) \in (\mathbb{R}^n)^2, \begin{cases} \|\nabla F(x) - \nabla F(y)\|_2 \leq \lambda_n \|x - y\|_2, \\ \langle \nabla F(x) - \nabla F(y), x - y \rangle_2 \geq \lambda_1 \|x - y\|_2^2, \end{cases}$$

où λ_1 et λ_n sont la plus petite et la plus grande valeur propre de la matrice A . La méthode du gradient à pas constant

$$x_0 \in \mathbb{R}^n \text{ et } \forall n \in \mathbb{N}, x_{n+2} = x_n - \alpha (A(x_n) - b),$$

converge alors vers l'unique solution $x_* \in \mathbb{R}^n$ de l'équation:

$$A(x_*) = b,$$

dès que le pas α satisfait la condition du théorème précédent à savoir :

$$0 < \alpha < \frac{2}{\lambda_1 + \lambda_2}.$$

En pratique, nous savons que cette méthode itérative de résolution d'un système linéaire convergessi le rayon spectral

$\rho(I_N - \alpha A)$ satisfait la condition :

$$\rho(I_N - \alpha A) < 1.$$

Comme ce rayon vaut :

$$\rho(I_N - \alpha A) = \max \{ |1 - \alpha \lambda_1|, |\alpha \lambda_2| \},$$

la méthode du gradient à pas constant converge (géométriquement) en fait dès que :

$$0 < \alpha < \frac{2}{\lambda_1}.$$

2. Méthode du gradient à pas optimal

La méthode du gradient à pas optimal consiste à optimiser le choix du pas α dans la méthode du gradient à pas constant à chaque étape. Plus précisément, étant donné un point $x_0 \in \mathbb{R}^N$, il s'agit de construire une suite $(x_n)_{n \in \mathbb{N}}$ suivant la formule :

$$\forall n \in \mathbb{N}, x_{n+1} = x_n - \alpha_n \nabla F(x_n), \text{ où :}$$

$$F(x_n - \alpha_n \nabla F(x_n)) = \min \{ F(x_n - \alpha \nabla F(x_n)), \alpha \in \mathbb{R} \}.$$

Dans le cas où cette suite $(x_n)_{n \in \mathbb{N}}$ est bien définie et convergente, et lorsque la suite des pas $(\alpha_n)_{n \in \mathbb{N}}$ est suffisamment bien contrôlée, la limite x_* de la suite $(x_n)_{n \in \mathbb{N}}$ vérifiera l'équation d'Euler-Lagrange :

$$\nabla F(x_*) = 0,$$

Et il sera alors permis d'espérer que cette limite x_* soit un minimiseur de la fonctionnelle F sur $\mathbb{R}^N$.

Le résultat suivant garantit que cette stratégie fonctionne dans le cas où la fonctionnelle F est elliptique.

Chéacine: Soit $F \in C^2(\mathbb{R}^N, \mathbb{R})$ une fonctionnelle elliptique. Notons $x_* \in \mathbb{R}^N$ l'unique minimiseur de la fonctionnelle F sur $\mathbb{R}^N$, et construisons une suite $(x_n)_{n \in \mathbb{N}}$ par le choix d'un point $x_0 \in \mathbb{R}^N$ et la formule de récurrence:

$$\forall n \in \mathbb{N}, x_{n+1} = x_n - \alpha_n \nabla F(x_n),$$

où le pas α_n est choisi de sorte que:

$$F(x_n - \alpha_n \nabla F(x_n)) = \min \{ F(x_n - \alpha \nabla F(x_n)), \alpha \in \mathbb{R}^+ \}.$$

La suite $(x_n)_{n \in \mathbb{N}}$ est alors bien définie et elle satisfait:

$$x_n \xrightarrow[n \rightarrow +\infty]{} x_*.$$

Preuve:

Supposons que la suite $(x_n)_{n \in \mathbb{N}}$ est bien définie jusqu'au rang n , et intéressons-nous à la construction de x_{n+1} . Deux cas se présentent:

(i) Si $\nabla F(x_n) = 0$, alors, la suite $(x_n)_{n \in \mathbb{N}}$ est bien définie et:

$$\forall p \geq n, x_p = x_n.$$

Comme x_* est l'unique point de $\mathbb{R}^N$ tel que $\nabla F(x_*) = 0$, il vient alors:

$$\forall p \geq n, x_p = x_n \xrightarrow[p \rightarrow +\infty]{} x_*.$$

(ii) Si $\nabla F(x_n) \neq 0$, alors, la fonction γ définie par:

$$\forall \alpha \in \mathbb{R}, \gamma(\alpha) = F(x_n - \alpha \nabla F(x_n)),$$

est bien définie et elle vérifie:

$$\forall (\alpha, \beta) \in \mathbb{R}^2, \gamma'(\alpha) - \gamma'(\beta) = - \langle \nabla F(x_n), \nabla F(x_n - \alpha \nabla F(x_n)) - \nabla F(x_n - \beta \nabla F(x_n)) \rangle$$

$$\Rightarrow (\gamma'(\alpha) - \gamma'(\beta))(\alpha - \beta) = \langle \beta \nabla F(x_n) - \alpha \nabla F(x_n), \nabla F(x_n - \alpha \nabla F(x_n)) - \nabla F(x_n - \beta \nabla F(x_n)) \rangle$$

$$= \|\nabla F(x_n - \beta \nabla F(x_n))\|_2^2$$

$$\geq \lambda (\alpha - \beta)^2 \|\nabla F(x_n)\|_2^2,$$

De sorte que'elle est elliptique sur $\mathbb{R}$. En particulier, il existe un unique nombre $\alpha_n \in \mathbb{R}$ tel que:

$$F(x_n - \alpha_n \nabla F(x_n)) = \min \{ F(x_n - \alpha \nabla F(x_n)), \alpha \in \mathbb{R} \}.$$

Dans tous les cas, la suite $(x_n)_{n \in \mathbb{N}}$ est donc bien définie.

De plus, si

$$\forall n \in \mathbb{N}, \quad \nabla F(x_n) \neq 0,$$

alors, le caractère minimiseur du pas α_n induit que:

$$\varphi'(x_n) = 0 \Leftrightarrow \langle \nabla F(x_n), \nabla F(x_n - \alpha_n \nabla F(x_n)) \rangle_L = 0,$$

ce qui implique que:

$$\forall n \in \mathbb{N}, \quad \langle \nabla F(x_n), \nabla F(x_{n+1}) \rangle_L = 0.$$

Comme $\varphi'(0) = -\|\nabla F(x_n)\|_L^2 \neq 0$, il est clair que:

$$\alpha_n \neq 0,$$

De sorte que:

$$\langle \nabla F(x_{n+1}), x_{n+1} - x_n \rangle_L = -\frac{1}{\alpha_n} \langle \nabla F(x_{n+1}), \nabla F(x_n) \rangle_L = 0.$$

L'ellipticité de la fonctionnelle F induit alors que:

$$\begin{aligned} F(x_n) - F(x_{n+1}) &\geq \langle \nabla F(x_{n+1}), x_n - x_{n+1} \rangle_L + \frac{\lambda}{2} \|x_{n+1} - x_n\|_L^2 \\ &= \frac{\lambda}{2} \|x_{n+1} - x_n\|_L^2. \end{aligned}$$

Comme la suite $(F(x_n))_{n \in \mathbb{N}}$ est décroissante et minorée par $F(x_*)$, il existe par ailleurs un nombre réel l tel que:

$$F(x_n) \xrightarrow{n \rightarrow +\infty} l,$$

De sorte que:

$$\|x_{n+1} - x_n\|_L \xrightarrow{n \rightarrow +\infty} 0.$$

Rappelons alors que la fonctionnelle F est coercive. Comme la suite $(F(x_n))_{n \in \mathbb{N}}$ est bornée, il s'ensuit que la suite $(x_n)_{n \in \mathbb{N}}$ est elle-même bornée. Il existe donc un sous-ensemble compact K de $\mathbb{R}^N$ tel que:

$$\forall n \in \mathbb{N}, \quad x_n \in K.$$

Puisque la fonction ∇F est uniformément continue sur K , il vient donc:

$$\|\nabla F(x_{n+1}) - \nabla F(x_n)\|_L \xrightarrow{n \rightarrow +\infty} 0.$$

Enfin, nous savons que:

$$\|\nabla F(x_n)\|_L^2 = \langle \nabla F(x_n), \nabla F(x_n) - \nabla F(x_{n+1}) \rangle_L,$$

De sorte que, par l'inégalité de Cauchy-Schwarz,

$$\|\nabla F(x_n)\|_L \leq \|\nabla F(x_n) - \nabla F(x_{n+1})\|_L \xrightarrow{n \rightarrow +\infty} 0.$$

Il reste alors à établir que:

$$\lambda \|x_n - x_*\|_L^2 \leq \langle \nabla F(x_n) - \nabla F(x_*) , x_n - x_* \rangle_L$$

$$= \langle \nabla F(x_n), x_n - x_* \rangle_{\mathbb{R}^n}$$

$$\Rightarrow \forall \|x_n - x_*\|_2 \leq \|\nabla F(x_n)\|_2 \xrightarrow{n \rightarrow +\infty} 0,$$

Dans certains la preuve du théorème.

La méthode du gradient à pas optimal converge donc sous des hypothèses un peu moins fortes que celle du gradient à pas constant. Sous les mêmes hypothèses, la vitesse de convergence reste au moins géométrique.

Notons cependant que la méthode du gradient à pas optimal nécessite à chaque itération de calculer le pas optimal, ce qui peut s'avérer coûteux en temps.

En pratique, il s'avère souvent plus simple de déterminer un pas variable à chaque étape pour lequel une condition de descente est assurée.

Dans tous les cas, le calcul du gradient à chaque itération peut également s'avérer très coûteux en temps et d'autres méthodes itératives peuvent être privilégiées, le choix final se faisant en relation avec les caractéristiques numériques de la fonctionnelle F .

Exemple: Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonctionnelle quadratique F définie par:

$$\forall x \in \mathbb{R}^n, F(x) = \frac{1}{2} \langle A(x), x \rangle_{\mathbb{R}^n} - \langle b, x \rangle_{\mathbb{R}^n},$$

est alors elliptique sur $\mathbb{R}^n$. La méthode du gradient à pas optimal

pour cette fonctionnelle consiste alors à choisir un vecteur $x_0 \in \mathbb{R}^n$

et à définir par récurrence les suites $(x_n)_{n \in \mathbb{N}}$ et $(\alpha_n)_{n \in \mathbb{N}}$ par:

$$\alpha_n = \begin{cases} \frac{\|A(x_n) - b\|_{\mathbb{R}^n}^2}{\langle A(A(x_n) - b), A(x_n) - b \rangle_{\mathbb{R}^n}} & \text{si } A(x_n) \neq b \text{ et } x_{n+1} = x_n - \alpha_n (A(x_n) - b), \\ 0 & \text{sinon} \end{cases}$$

Rappelons en effet que le pas α_n doit satisfaire la condition:

$$\langle A(x_n) - b, A(x_n - \alpha_n (A(x_n) - b)) - b \rangle_{\mathbb{R}^n} \stackrel{!}{=} 0,$$

à chaque itération. Cette nouvelle méthode itérative est alors bien

définie et convergente vers la unique solution $x_* \in \mathbb{R}^n$ de l'équation:

$$A(x_*) = b. \text{ Sur le plan numérique, elle est particulièrement avan-}$$

tageuse lorsque le calcul de $A(x_n)$ est rapide, par exemple,

lorsque la matrice A est assez creuse.

3. Méthode du gradient conjugué.

Exemple: Soit $0 < \alpha_1 < \alpha_2$. Considérons la fonctionnelle quadratique F définie par :

$$\forall x = (x_1, x_2) \in \mathbb{R}^2, F(x_1, x_2) = \frac{1}{2}(\alpha_1 x_1^2 + \alpha_2 x_2^2).$$

La fonctionnelle F est elliptique sur $\mathbb{R}^2$ et son unique minimum est le point $x_0 = (0, 0)$. En chaque point $x = (x_1, x_2) \in \mathbb{R}^2 \setminus \{x_0\}$, la direction de descente localement optimale est donnée par l'opposé du vecteur gradient de F , soit le vecteur :

$$-\nabla F(x) = (-\alpha_1 x_1, -\alpha_2 x_2).$$

Cependant, la direction de descente globalement optimale est en fait donnée par le vecteur $-x$ pour lequel la minimisation sur la droite $(x - t x)_{t \in \mathbb{R}}$ convergerait en une seule étape.

Ainsi, lorsque nous choisissons un vecteur $x^0 = (x_1^0, x_2^0) \in \mathbb{R}^2$, avec $x_1^0 \neq 0$ et $x_2^0 \neq 0$, la méthode du gradient à pas opti conduit à construire la suite $(x_n)_{n \geq 0}$, suivant la relation de récurrence :

$$\forall n \in \mathbb{N}, x_{n+1} = \left(\frac{\alpha_1^2 (x_1 - \alpha_1 x_1) - \alpha_2^2 (x_2^0)^2}{\alpha_1^2 (x_1^0)^2 + \alpha_2^2 (x_2^0)^2}, \frac{\alpha_1^2 (x_1 - \alpha_1 x_1) x_2^0 (x_2^0)^2}{\alpha_1^2 (x_1^0)^2 + \alpha_2^2 (x_2^0)^2} \right),$$

laquelle vérifie :

$$\forall n \in \mathbb{N}, x_n \neq x_0,$$

alors même qu'il serait possible d'obtenir cette propriété avec de meilleurs choix de directions de descente.

La méthode du gradient conjugué repose sur cette observation. Elle permet de déterminer le minimum x_0 d'une fonctionnelle quadratique F définie par :

$$\forall x \in \mathbb{R}^N, F(x) = \frac{1}{2} \langle A(x), x \rangle_{\mathbb{R}^N} - \langle b, x \rangle_{\mathbb{R}^N},$$

où $A \in \mathcal{S}_+^N(\mathbb{R})$ et $b \in \mathbb{R}^N$, en au plus N itérations. Une suite $(x_n)_{n \in \mathbb{N}}$ des itérés est construite de la façon suivante. Un vecteur $x_0 \in \mathbb{R}^N$ est choisi, et si les vecteurs $x_1, \dots, x_n$ sont déjà construits et que :

$$\forall 0 \leq k \leq n, \nabla F(x_k) \neq 0,$$

alors, le vecteur x_{n+1} est choisi comme la minimiseur de la fonctionnelle F sur l'ensemble :

$$x_n = \left\{ x_n + \sum_{i=0}^n \lambda_i \nabla F(x_i), \text{ où } (\lambda_0, \dots, \lambda_n) \in \mathbb{R}^{n+1} \right\}.$$

Notons que, puisque la fonctionnelle F est elliptique sur $\mathbb{R}^N$ et que l'ensemble X_n est non vide, convexe et fermé, ce problème de minimisation a bien une unique solution x_{n+1} , qui satisfait de plus les équations:

$$\forall 0 \leq i \leq n, \langle \nabla F(x_{n+1}), \nabla F(x_i) \rangle_{\mathbb{R}^N} = 0.$$

En particulier, si la suite $(x_n)_{n \in \mathbb{N}}$ vérifie la condition:

$$\forall 0 \leq i \in \mathbb{N}, \nabla F(x_i) \neq 0,$$

alors, l'espace vectoriel engendré par les vecteurs $\nabla F(x_0), \dots$, et $\nabla F(x_{N-1})$ est de dimension finie égale à N , et l'ensemble X_{N-1} est donc égal à l'espace vectoriel $\mathbb{R}^N$. Le minimiseur de F sur cet ensemble est donc égal à:

$$x_N = x_*$$

La méthode du gradient conjugué converge donc en au plus N itérations.

Reste à déterminer la complexité de l'algorithme à écrire afin de mettre en œuvre cette méthode. Choisissons un point $x_0 \in \mathbb{R}^N$ et notons:

$$d_0 = \nabla F(x_0) = A(x_0) - b \in \mathbb{R}^N.$$

Si $d_0 = 0$, alors, $\nabla F(x_0) = 0$, ce qui signifie que le point x_0 est égal au minimiseur x_* de la fonctionnelle F sur $\mathbb{R}^N$, et l'algorithme s'arrête à cette étape.

Si non, il s'agit de minimiser la fonctionnelle F sur l'ensemble:

$$X_0 = \{ x_0 + \lambda_0 d_0, \lambda_0 \in \mathbb{R} \}.$$

L'unique solution x_1 de ce problème satisfait l'équation:

$$\langle \nabla F(x_1), d_0 \rangle_{\mathbb{R}^N} = 0,$$

ce qui signifie que:

$$\mu_0 = \frac{\langle \nabla F(x_0), d_0 \rangle_{\mathbb{R}^N}}{\langle A(d_0), d_0 \rangle_{\mathbb{R}^N}} \text{ et } x_1 = x_0 - \mu_0 d_0.$$

Si $\nabla F(x_1) = 0$, alors, le point x_1 est égal au minimiseur x_* , et l'algorithme s'arrête à cette étape.

Si non, il s'agit désormais de minimiser la fonctionnelle F sur l'ensemble:

$$X_1 = \{ x_1 + \lambda_1 \nabla F(x_1) + \lambda_0 \nabla F(x_0), (\lambda_0, \lambda_1) \in \mathbb{R}^2 \}.$$

Notons $x_2 = x_1 - \mu_1 \nabla F(x_1) - \tilde{\mu}_0 \nabla F(x_0)$, l'unique solution de ce problème.

me. Ici $\mu_2 = 0$, alors,

$$x_2 \in X_0 \Rightarrow x_2 = x_0.$$

Puisque le point x_2 satisfait les équations

$$\langle \nabla F(x_2), \nabla F(x_2) \rangle_L = \langle \nabla F(x_2), \nabla F(x_0) \rangle_L = 0,$$

il vient :

$$\|\nabla F(x_2)\|_L^2 = 0,$$

ce qui est absurde ! Il est donc permis d'écrire le point x_2 sous la forme :

$$x_2 = x_0 + \mu_2 d_2, \text{ avec } d_2 = \nabla F(x_0) + \frac{\tilde{\mu}_2}{\mu_2} d_0.$$

Alors que les équations pour le point x_2 deviennent :

$$\|\nabla F(x_2)\|_L^2 = \mu_2 \langle A(d_2), \nabla F(x_2) \rangle_L \text{ et } \langle \nabla F(x_2), \nabla F(x_0) \rangle_L = \mu_2 \langle A(d_2), \nabla F(x_0) \rangle_L$$

Comme nous savons que :

$$\langle \nabla F(x_2), \nabla F(x_0) \rangle_L = 0,$$

il vient :

$$\langle A(d_2), d_2 \rangle_L = 0 \Rightarrow \langle d_2, A(d_2) \rangle_L = 0$$

$$\Rightarrow \langle d_2, \nabla F(x_0) \rangle_L = \langle d_2, \nabla F(x_2) \rangle_L,$$

puisque $\nabla F(x_2) = \nabla F(x_0) - \mu_2 A(d_2)$. Il s'ensuit que :

$$\frac{\tilde{\mu}_2}{\mu_2} = \frac{\|\nabla F(x_2)\|_L^2}{\|\nabla F(x_0)\|_L^2},$$

De sorte que :

$$d_2 = \nabla F(x_2) + \frac{\|\nabla F(x_2)\|_L^2}{\|\nabla F(x_0)\|_L^2} d_0.$$

Il vient alors :

$$\begin{aligned} \langle A(d_2), \nabla F(x_2) \rangle_L &= \langle A(d_2), d_2 \rangle_L - \langle A(d_2), d_0 \rangle_L \frac{\|\nabla F(x_2)\|_L^2}{\|\nabla F(x_0)\|_L^2} \\ &= \langle A(d_2), d_2 \rangle_L \end{aligned}$$

Pendant que :

$$\|\nabla F(x_2)\|_L^2 = \langle \nabla F(x_2), d_2 \rangle_L,$$

De sorte que :

$$\mu_2 = \frac{\langle \nabla F(x_2), d_2 \rangle_L}{\langle A(d_2), d_2 \rangle_L}.$$

Supposons alors avoir construit les suites $(x_k)_{0 \leq k \leq l}$, $(\mu_k)_{0 \leq k \leq l}$ et $(d_k)_{0 \leq k \leq l}$ jusqu'au rang l suivant les formules :

$$x_{k+1} = x_k - \mu_k d_k, \quad d_k = \nabla F(x_k) + \frac{\|\nabla F(x_k)\|_L^2}{\|\nabla F(x_0)\|_L^2} d_{k-1} \text{ et}$$

$$\mu_k = \frac{\langle \nabla F(x_k), d_k \rangle_L}{\langle A(d_k), d_k \rangle_L},$$

Et cherchons à la construire au rang $l+1$ sous l'hypothèse que:

$$DF(x_l) \neq 0.$$

Il faut donc minimiser la fonctionnelle F sur l'ensemble:

$$X_{l+1} = \left\{ x_l - \sum_{j=0}^l \lambda_j DF(x_j), (\lambda_0, \dots, \lambda_l) \in \mathbb{R}^{l+1} \right\}.$$

Notons $x_{l+1} = x_l - \sum_{j=0}^l \tilde{\mu}_j DF(x_j)$, l'unique solution de ce problème. Si $\mu_l = 0$ alors:

$$x_{l+1} \in X_{l-1} \Rightarrow x_{l+1} = x_l.$$

Puisque le point x_{l+1} satisfait les équations:

$$\forall 0 \leq j \leq l, \langle DF(x_{l+1}), DF(x_j) \rangle_L = 0,$$

il vient:

$$\|DF(x_l)\|_L^2 = 0,$$

Ce qui est absurde! Notons alors:

$$x_{l+1} = x_l - \mu_l d_l, \text{ avec } d_l = DF(x_l) + \sum_{j=0}^{l-1} \frac{\tilde{\mu}_j}{\mu_l} DF(x_j).$$

Comme $DF(x_{l+1}) = DF(x_l) - \mu_l A(d_l)$, les équations précédentes deviennent:

$$\forall 0 \leq j \leq l-1, \langle A(d_l), DF(x_j) \rangle_L = 0 \text{ et } \mu_l \langle A(d_l), DF(x_l) \rangle_L = \|DF(x_l)\|_L^2.$$

Par l'hypothèse de récurrence, il vient:

$$\forall 0 \leq j \leq l-1, \langle A(d_l), d_j \rangle_L = 0 \Rightarrow \langle d_l, A(d_j) \rangle_L = 0$$

Comme $DF(x_{j+1}) = DF(x_j) - \mu_j A(d_j)$, il s'ensuit que:

$$\forall 0 \leq j \leq l-1, \langle d_l, DF(x_{j+1}) \rangle_L = \langle d_l, DF(x_j) \rangle_L$$

$$\Rightarrow \begin{cases} \|DF(x_l)\|_L^2 = \frac{\tilde{\mu}_{l-1}}{\mu_l} \|DF(x_{l-1})\|_L^2 \\ \text{et } \forall 0 \leq j \leq l-1, \frac{\tilde{\mu}_{j+1}}{\mu_l} \|DF(x_{j+1})\|_L^2 = \frac{\tilde{\mu}_j}{\mu_l} \|DF(x_j)\|_L^2 \end{cases}$$

De sorte que:

$$d_l = DF(x_l) + \sum_{j=0}^{l-1} \frac{\|DF(x_l)\|_L^2}{\|DF(x_j)\|_L^2} DF(x_j).$$

Par récurrence, il vient alors:

$$d_l = DF(x_l) + \frac{\|DF(x_l)\|_L^2}{\|DF(x_{l-1})\|_L^2} d_{l-1}.$$

Il reste alors à vérifier que:

$$\mu_l = \frac{\|DF(x_l)\|_L^2}{\langle A(d_l), DF(x_l) \rangle_L} = \frac{\langle DF(x_l), d_l \rangle_L}{\langle A(d_l), d_l \rangle_L},$$

Ce qui achève la construction par récurrence des suites $(x_h)_{0 \leq h \leq p}$, $(\mu_h)_{0 \leq h \leq p-1}$

et $(d_h)_{0 \leq h \leq p-1}$.

Nous avons donc établi le théorème suivant.

Théorème: Soit $A \in \mathcal{S}_n^{++}(\mathbb{R})$ et $b \in \mathbb{R}^n$. Considérons un point $x_0 \in \mathbb{R}^n$ et posons:

$$\begin{cases} d_0 = \nabla F(x_0), \quad \mu_0 = \frac{\langle \nabla F(x_0), d_0 \rangle_L}{\langle A(d_0), d_0 \rangle_L}, \quad x_1 = x_0 - \mu_0 d_0 \\ \forall n \in \{1, \dots, N-1\}, \quad d_n = \nabla F(x_n) + \frac{\| \nabla F(x_n) \|_L^2}{\| \nabla F(x_{n-1}) \|_L^2} d_{n-1}, \quad \mu_n = \frac{\langle \nabla F(x_n), d_n \rangle_L}{\langle A(d_n), d_n \rangle_L} \\ \text{et } x_{n+1} = x_n - \mu_n d_n. \end{cases}$$

Les suites $(x_n)_{0 \leq n \leq N-1}$, $(d_n)_{0 \leq n \leq N-1}$ et $(\mu_n)_{0 \leq n \leq N-1}$ sont bien définies jusqu'à un entier $0 \leq l \leq N-1$ tel que:

$$\nabla F(x_l) = 0.$$

L'unique solution $x_* \in \mathbb{R}^n$ de l'équation:

$$A(x_*) = b$$

est alors égale à x_l si $0 \leq l \leq N-1$, à x_N sinon.

Preuve:

Immédiate.

Le nom de méthode du gradient conjugué repose sur le fait que les directions de descente $(d_n)_{0 \leq n \leq N-1}$ sont conjuguées par rapport à la matrice A , c'est-à-dire qu'elles vérifient:

$$\forall 0 \leq j \neq l \leq N-1, \langle A(d_j), d_l \rangle_L = 0.$$

Cette méthode est une méthode directe de résolution d'un système linéaire où la matrice A est symétrique définie positive. Notons que le nombre d'opérations nécessaires à la méthode du gradient conjugué est de l'ordre de n^3 puisqu'il est nécessaire de calculer n fois des produits de la forme $A(d_n)$ qui nécessitent de l'ordre de n^2 opérations élémentaires. Ce n^3 est donc pas mieux que les méthodes de Gauss ou de Cholesky. Cependant, la seule opération coûteuse consiste en le calcul des produits $A(d_n)$, de sorte que si la matrice A est de plus creuse, alors, le temps de calcul est réduit. Notons également qu'en pratique la méthode du gradient conjugué peut très bien converger en moins de n itérations.