Optimisation quadratique

Table des matières

1 Existence et unicité d'un minimum

1.1 Notations

On considère E l'espace vectoriel $\mathbf{R}^n$, k un sous-ensemble non vide de E et J une *fonctionnelle continue* définie sur K à valeur dans $\mathbf{R}$.

1.2 Problème

Trouver $u \in K$, tel que $J(u) = \inf_{v \in K} J(v)$.

1.3 Existence

Définition : Minimum local–global

$u \in K$ est un **point de minimum local** de J sur K si, et seulement si

$$\exists \eta > 0,\ \forall v \in K,\ \|v - u\| < \eta \Rightarrow J(u) \leqslant J(v)$$

$u \in K$ est un **minimum global** de J sur K si, et seulement si

$$\forall v \in K,\ J(u) \leqslant J(v)$$

Définition : Suite minimisante

On dit qu'une suite $(u_k)_{k\in\mathbf{N}}$ d'éléments de K est une **suite minimisante** si et seulement si

$$\lim_{k\to+\infty} J(u_k) = \inf_{v\in K} J(v)$$

Théorème : (rappel)

Si K est *compact* et J *continue* sur K, alors J **atteint ses *extrema***.

Définition :

On dit qu'un fonctionnelle J est **infine à l'infini** si, et seulement si :

$$\text{pour toute suite } (u_n)_n \in K^n,\ \lim_{n\to+\infty} \|v_n\| = +\infty \Rightarrow \lim_{n\to+\infty} J(v_n) = +\infty$$

Théorème : Existence d'un minimum global

Dans le cas où $E = \mathbf{R}^n$, si K est un fermé, et si J est *continue* et *infinie à l'infini* dans K, alors elle admet **un minimum global** sur K. De plus, de toute suite minimisante, on peut extraire une *sous-suite* qui converge vers un point de minimum.

1.4 Convexité et unicité

Définition : Ensemble convexe

On dit qu'un sous-ensemble K de E est convexe si, et seulement si, pour tout couple d'éléments (u, v), le segment $[u, v]$ est inclus dans K :

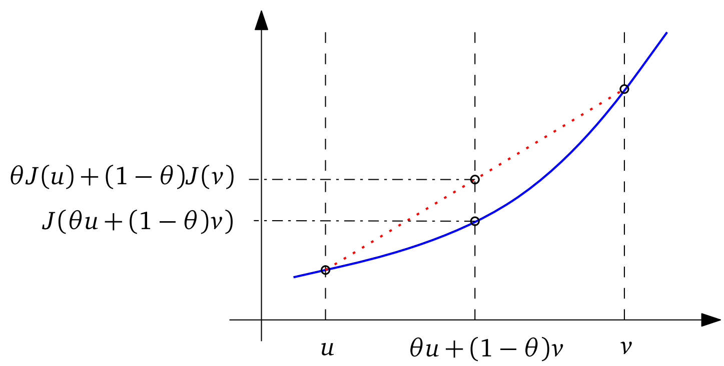
$$\forall (u,v) \in K^2,\ \forall t \in [0,1],\ tv + (1-t)u \in K$$

Définition : Fonctionnelle convexe

Soit J une fonctionnelle définie sur un sous-ensemble convexe non vide K de E, à valeurs dans R. On dit que J est **convexe** si et seulement si :

$$\forall (u,v) \in K^2,\ u \neq v,\ \forall \theta \in]0,1[,\ J(\theta u + (1-\theta v) \leqslant \theta J(u) + (1-\theta) J(v)$$

Dans le cas d'une inégalité stricte, on dit que la fonctionnelle est **strictement** convexe.



Propriété

Soit J une fonctionnelle convexe définie sur un convexe non vide K :

❶ si u et v sont deux points de minimum locaux, alors $J(u) = J(v)$;

❷ si en plus, J est strictement convexe alors $u = v$.

Théorème :

Soit J une fonctionnelle convexe définie sur un convexe non vide K de E :

❶ tout point de minimum local est un point de minimum globale ;

❷ si de plus J est *strictement* convexe, le point de minimum, s'il exite, est *unique*.

1.4.1 Projection sur un convexe fermé

Définition :

Pour tout w de E, on définit sa projection sur un convexe fermé K $P_K(w)$comme l'élément u tel que :

$$\|w-u\| = \min_{v\in K} \|w-v\|$$

Propriété

Si K est un convexe fermé, alors tout élément w de E admet une projection *unique* $P_K(w)$ sur K. DE lus l'application $w \longmapsto P_K(w)$ est *contractante*.

1.4.2 Propriété des fonctions convexes

Théorème :

Soit J une fonctionnelle différentiable sur un sous-ensemble K *convexe non vide*.
Les assertions suivantes sont équivalentes :

❶ J est *convexe* sur K.

❷ $\forall (u, v) \in K^2,\ u \neq v$:
$$J(v) \geqslant J(u) + \langle \nabla J(u),\, v-u\rangle.$$

❸ $\forall (u, v) \in K^2,\ u \neq v$:
$$\langle \nabla J(u) - \nabla J(v),\, u-v\rangle \geqslant 0.$$

Pour la stricte convexité, on a des inégalité strictes.

Théorème : classe $\mathscr{C}^2$

Soit J une fonctionnelle de E dans $\mathbf{R}$ de classe $\mathscr{C}^2$. Alors J est *convexe* si, et seulement si,

$$\forall (u, d) \in E^2,\ \left\langle \nabla^2 J(u) d,\, d\right\rangle \geqslant 0.$$

2 Conditions d'optimalité

2.1 Cas général

Définition : Cône des directions admissibles

Soit K un sous-ensemble non vide de E, et v un point de K. On appelle cône des directions admissibles l'ensemble $\mathcal{T}_K$ des tangentes en v aux chemins inclus dans K et commençants en v.

Théorème :

Soient K un sous-ensemble non vide de E, u un point de K et J une fonctionnelle de K dans $\mathbf{R}$. On suppose que J est Frechet-*différentiable* en u. Si u est un point de minimum local de J sur K, on a nécessairement :

$$\forall w \in \mathcal{T}_K, \quad \langle \nabla J(u), w \rangle \geqslant 0$$

Propriété

Si $K = E$ ou si le minimum u est intérieur à K alors l'inéquation ci-dessus devient :

$$\nabla J(u) = 0$$

2.2 Cas convexe

Ici onse place dans le cas où K est convexe.

Théorème : Inéquation d'Euler

Soient K un sous-ensemble convexe non vide de E, u un point de K et J une fonctionnelle de K dans $\mathbf{R}$. On suppose que J est différentiable en u. Si u est un *point de minimum local* de J sur K, on a nécessairement

$$\forall v \in K, \quad \langle \nabla J(u), v - u \rangle \geqslant 0$$

Théorème : J convexe

Soient K un sous-ensemble convexe non vide de E, u un point de K et J une fonctionnelle de K dans $\mathbf{R}$. On suppose que J est **convexe** et différentiable en u. Si u est un *point de minimum local* de J sur K **si, et seulement si**

$$\forall v \in K, \quad \langle \nabla J(u), v - u \rangle \geqslant 0$$

2.3 Contraintes d'égalité affines

On suppose ici que l'ensemble K est défini par :

$$K = \{v \in \mathbf{R}^n \,/\, Cv = f\}$$

où C est une matrice $p \times n$ et f un élément de $\mathbf{R}^p$ (on suppose $K \neq 0$). On remarque que K est un *convexe* non vide.

Théorème : (Karush, Kuhn et Tucker)

Soit $C \in \mathcal{M}_{p,n}(\mathbf{R})$ et $f \in \mathbf{R}^p$, K le sous-espace défini par $K = \{v \in \mathbf{R}^n \,/\, Cv = f\}$, u un point de K et J une fonctionnelle différentiable en u.
Si u est un point de *minimum local* de J sur K, on a **nécessairement** :

$$\begin{cases} \exists \lambda \in \mathbf{R}^p, \quad \nabla J(u) + {}^t C\lambda = 0 \\ Cu - f = 0 \end{cases}$$

2.4 Le lagrangien

Définition : Le lagrangien

On définit la fonctionnelle (appelée le *lagrangien* du système) :

$$\forall (v,\mu) \in \mathrm{e} \times \mathbf{R}^p, \quad \mathscr{L}(v,u) = J(v) + \langle \mu C v - f \rangle$$

Les élément μ de $\mathbf{R}^p$ sont appelés les **multiplicateurs de Lagrange**.

Propriété

Soit $C \in \mathscr{M}_{p,n}(\mathbf{R})$ et $f \in \mathbf{R}^p$, K le sous-espace défini par $K = \{v \in \mathbf{R}^n \,/\, Cv = f\}$, u un point de K et J une fonctionnelle différentiable en u.
Si u est un point de *minimum local* de J sur K, on a **nécessairement** :

$$\exists \lambda \in \mathbf{R}^p, \text{ tel que } \quad \begin{cases} \nabla_v \mathscr{L}(u,\lambda) = 0 \\ \nabla_\mu \mathscr{L}(u,\lambda) = 0 \end{cases}$$

2.5 Fonctionnelle quadratique et contraintes d'égalité affines

La fonctionnelle est ici quadratique en $v \in \mathbf{R}^n$:

$$\forall v \in \mathbf{R}^n, \quad J(v) = \frac{1}{2}\langle Av,\, v\rangle - \langle b,\, v\rangle + c$$

où A est une matrica **symétrique** de $\mathbf{R}^{n\times n}$, b un vecteur de $\mathbf{R}^n$ et $c \in \mathbf{R}$.
On considère ses variations sur l'espace affine :

$$K = \{v \in \mathbf{R}^p \,/\, Cv = f\}$$

où $C \in \mathbf{R}^{p\times n}$, et f un éléments de R^p.

D'après les théorèmes précédents, on a si u est un point de minimum de J sur K :

$$\exists \lambda \in \mathbf{R}^p, \text{ tel que } \quad \begin{cases} Au + {}^tC\lambda = b \\ Cu = f \end{cases}$$

Propriété

Soit J et K définis ci-dessus. Si u est un point de minimum de J sur K, alors il existe $\lambda \in \mathbf{R}^p$ tel que le couple (u,λ) de $\mathbf{R}^n \times \mathbf{R}^p$ soit solution du syste=ème linéaire :

$$\begin{pmatrix} A & {}^tC \\ C & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}$$

Théorème :

Si A est **symétrique positive**, le vecteur u est un point de minimum de J sur K **si, et seulement si,** il existe un élémet λ de $\mathbf{R}^p$ tel que le couple (u,λ) soit solution de :

$$\begin{pmatrix} A & {}^tC \\ C & 0 \end{pmatrix} \begin{pmatrix} u \\ \lambda \end{pmatrix} = \begin{pmatrix} b \\ f \end{pmatrix}$$

Si de plus A est **définie-positive** et C est **surjective** alors le système linéaire admet une **solution unique**.

2.6 Contraintes d'inégalité affines

Dans cette section on considère que l'ensemble des **contraintes** K est donné par :

$$K = \{v \in E \,/\, Cv \leqslant f\}$$

Théorème :

On considère K défini ci-dessus et J une fonctionnelle *différentiable* sur E dans $\mathbf{R}$. Si $u \in K$ est un minimum local de J sur K alors :

$$\exists \lambda \in \mathbf{R}^p,\ \lambda \geqslant 0, \quad \begin{cases} \nabla J(u) + {}^t C\lambda = 0 \\ \lambda_i \left[Cu - f\right]_i = 0 \\ Cu \leqslant f \end{cases}$$

De plus si J est *convexe*, alors ceci est une condition *suffisante* de minimalité.

3 Algorithmes pour problèmes sans contraintes, fonctionnelle quadratique

On étudie ici les algorithmes qui permettent de calculer **numériquement** la solution du problème de minimisation :

Trouver $u \in \mathbf{R}^n$ tel que $J(u) = min_{v \in \mathbf{R}^n} J(v)$.

Avec ici, J le fonctionnelle qui à v associe $J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$, avec A une matrice *symétrique définie-positive* de $\mathbf{R}^{n\times n}$ et b un vecteur quelconque de $\mathbf{R}^n$.
Dans ce cas, la solution existe et est *unique*, et elle vérifie le système linéaire :

$$Au = b$$

Les algorithmes suivants consistent tous à choisir une condition initiale u_0 puis à construire une suite $(u_k)_{k\geqslant 1}$. Ces algorithmes doivent vérifier les deux propriétés suivantes :

La convergence de la suite $5u_k)$ est assurée quel que soit le vecteur initiale.
La convergence doit être « suffisament rapide ».

3.1 Taux et vitesse de convergence

Définition :

Soit $\|.\|$ une norme matricielle *induite*. On appelle **conditionnement** d'une matrice réelle *inversible* $A \in \mathbf{R}^{n\times n}$, *relatif à cette norme*, la valeur définie par :

$$\mathrm{cond}(A) = \|A\| \left\|A^{-1}\right\|$$

Définition :

Soit une méthode numérique produisant une suite d'itérés (u_k). Soit $C > 0$ la plus *petite* constante telle que, pour tout $k \geqslant 0$: $\left\|u_{k+1} - u\right\| \leqslant C \left\|u_k - u\right\|$. C est appelé **taux de convergence**. On appellera **vitesse de convergence** $R = -\log C$.

3.2 Méthode de descente

Supposons l'itéré u_k connu, on *choisit* une direction dite de **descente** $d_k \neq 0$, et un **pas de descente** ρ_k. On construit alors l'itéré u_{k+1} par la formule :

$$u_{k+1} = u_k + \rho_k d_k$$

Remarque : Le choix de ρ_k et d_k se fait de manière à assurer que $J(u_{k+1}) < J(u_k)$.

Principe

3.2.1 Pas optimal – fonctionnelle quadratique

Dans le cas où la fonctionnelle est **quadratique**, on obtient une formule du pas optimal :

$$\rho_k = -\frac{\langle \nabla J(u_k), d_k \rangle}{\langle A d_k, d_k \rangle}$$

3.2.2 Relaxation

Définition :

On considère une base orthonormée $(e_i)_{1 \leqslant i \leqslant n}$ de $\mathbf{R}^n$. On choisit la suite des direction de descente $d_0 = e_1, \ldots, d_n = e_1, \ldots$, etc. et ainsi de suite. On choisit le *pas optimal* défini ci dessus. L'algorithme est alors :

Soit $l \geqslant 0$, on choisit $u_o \in \mathbf{R}^n$ et une tolérance ϵ.

Pour $k = nl + i - 1$, on prend :

❶ $d_k = e_i$

❷ $\rho_k = -\dfrac{\langle \nabla J(u_k), d_k \rangle}{\langle A d_k, d_k \rangle}$

❸ $u_{k+1} = u_k + \rho_k e_i$

Propriété

Si A est **symétrique définie positive** alors la méthode de relaxation est **convergente**.

3.2.3 Gradient à pas fixe

Définition :

Ici, on cherche la direction dans laquelle la convergence sera la plus rapide.

Soit $u_0 \in \mathbf{R}^n$ une condition initiale, ρ un pas fixe.
Pour $k \in \mathbf{N}+$

❶ $d_k = -\nabla J(u_k) = b - A u_k$

❷ $u_{k+1} = u_k + \rho d_k$

Propriété

Si A est **symétrique définie positive**, la méthode de gradient à pas fixe est convergente sous réserve que :

$$0 < \rho < \frac{2}{\lambda_{\max}(A)}$$

3.2.4 Méthode du gradient à pas optimal

Définition :

On procède ici comme avec la méthode du gradient à pas fixe, sauf que le pas ρ_k est ici :

$$\rho_k = -\frac{\langle \nabla J(u_k), d_k \rangle}{\langle Ad_k, d_k \rangle}$$

Soit $u_0 \in \mathbf{R}^n$ une condition initiale.
Pour $k \in \mathbf{N}+$

❶ $d_k = -\nabla J(u_k) = b - Au_k$

❷ $u_{k+1} = u_k + \rho_k d_k$

Propriété

Si A est **symétrique définie positive**, la méthode de gradient à pas optimal est convergente.

3.2.5 Gradient conjugué

Le principe de la *méthode du gradient conjugué* est de construire une suite de *direction* de descente que l'on garde en mémoire, pour essayer d'éviter les retours.

Si $u_1, \ldots, u_k$ ont déjà été calculés, et si tous les gradients $g_l = \nabla J(u_l)$, on recherche u_{k+1} tel que :

$$J(u_{k+1}) = \min_{v \in u_k + G_k} J(v) \quad \text{avec} \quad G_k = \text{Vect}(g_0, \ldots, g_k)$$

On a :

❶ g_{k+1} est orthogonale à G_k, *i.e* $\forall l \in [\![0, k]\!]$, $\langle g_{k+1}, g_l \rangle = 0$

❷ La méthode du GC converge en au plus n étapes !

❸ (d_l) sont conjugués par rapport à A : $\forall k \neq l$, $\langle Ad_k, d_l \rangle = 0$.

Algorithme

initialisation : $k = 0$, on choisit u_0, on calcul $g_0 = Au_0 - b$ et $d_0 = -g_0$.
tant que : $\|g_k\| > \epsilon$, itérer k

- Calculer $\beta_{k-1} = \dfrac{\langle g_{k-1}, d_{k-1} \rangle}{\langle Ad_{k-1}, d_{k-1} \rangle}$
- Calculer $u_k = u_{k-1} - \beta_{k-1} d_{k-1}$
- Nouvelle direction : $g_k = Au_k - b$ puis $\alpha_k = \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2}$ et enfin $d_k = \alpha_k d_{k-1} - g_k$.

3.2.6 Méthodes itératives

On s'intéresse ici à des méthodes numériques de résolution du système linéaire :

$$J'(u) = Au - b = 0.$$

La solution de ce système est le minimum recherché de la fonctionnelle quadratique J sur $\mathbf{R}^n$.

Principe

❶ On décompose A sous la forme $A = M - N$ avec M *inversible*.

❷ Partant de $u_0 \in \mathbf{R}^n$, on construit $(u_k)_k$ par :

$$Mu_{k+1} = Nu_k - b \quad i.e \quad u_{k+1} = M^{-1}Nu_k - M^{-1}b.$$

Attention : Ces méthodes ne sont intéressante que si le choix de M rend le calcul de u_{k+1} facile.

Propriété

Si $(u_k)_k$ converge alors c'est vers la solution du système linéaire $Au = b$.

Propriété

1. La suite $(u_k)_k$ définie par la méthode iérative est *convergente* si et seulement si le *rayon spectrale* $\rho(M^{-1}N)$ vérifie : $\rho(M^{-1}N) < 1$.
2. Si A est **symétrique définie positive** et si ${}^tM + N$ est aussi **définie positive**, alors

$$\rho(M^{-1}N) < 1$$

On note $D = \mathrm{diag}A$, $-E = \mathrm{tirang}_{\mathrm{inf}}(A)$ et $-F = \mathrm{tirang}_{\mathrm{sup}}(A)$.

Définition : Méthode de Jacobi

Cette méthode consiste à choisir $M = D$ et $N = E + F$. On désigne par *matrice de Jacobi* la matrice $\mathscr{J} = M^{-1}N = D^{-1}(E+F)$. L'algorithme s'écrit alors :

1. On choisit $u_0 \in \mathbf{R}^n$ et $\epsilon > 0$.
2. **Tant que** $\left\|Au_k - b\right\| > \epsilon$, calculer $u_{k+1} = \mathscr{J}u_k + D^{-1}b$.

Définition : Méthode de Gauss-Seidel

Cette méthode consiste à choisir $M = D - E$ et $N = F$. On désigne par *matrice de Gauss-Seidel* la matrice $\mathscr{G} = M^{-1}N = (D-E)^{-1}F$. L'algorithme s'écrit alors :

1. On choisit $u_0 \in \mathbf{R}^n$ et $\epsilon > 0$.
2. **Tant que** $\left\|Au_k - b\right\| > \epsilon$, calculer $u_{k+1} = \mathscr{G}u_k + (D-E)^{-1}b$.

Théorème :

1. Si A est **symétrique** et **définie-positive**, alors la méthode de Gauss-Seidel **converge**.
2. Si A est à diagonale **strictement dominante** alors les deux méthodes convergent.
3. Si A est **tridiagonale**, alors $\rho(\mathscr{G}) = \rho(\mathscr{J})^2$.

4 Algorithme pour les problèmes contraints

Là encore nous nous intéressons au problème :

> Trouver $u \in K$ tel que $J(u) = \min_{v\in K} J(v)$.

où J est la fonctionnelle qui à $v \in K$ associe $J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle$, $A \in \mathscr{M}_{n\times n}(\mathbf{R})$ est supposée *symétrique définie-positive*, b est un vecteur quelconque de $\mathbf{R}^n$ et K est un *convexe fermé* de $\mathbf{R}^n$.

4.1 Méthode du gradient projeté

Rappel : Pour tout $w \in \mathbf{R}^n$, il existe une unique projection de w sur K, notée $P_K(w) \in K$, solution de :

$$\left\|P_K(w) - w\right\| = \min_{v\in K} \|v - w\|.$$

De plus la projection est caractérisée par :

$$\forall v \in K, \quad \langle P_K(w) - w, P_K(w) - v\rangle \leqslant 0$$

Propriété

Soit K un *convexe fermé non vide* de $\mathbf{R}^n$ et $\rho > 0$. u est solution du problème contraint $J(u) = \min_{v\in K} J(v)$ si et seulement si :

$$u = P_K(u - \rho\nabla J(u)).$$

Algorithme

1. **Initialisation :** $u_0 \in \mathbf{R}^n$, un pas $\rho > 0$ et une précision $\eta > 0$.
2. **Tant que :** $\left\|u_k - u_{k+1}\right\| > \eta$,

$$u_{k+1} = P_K(u_k - \rho\nabla J(u_k))$$

Propriété

Dans les conditions évoquées ci-dessus et si K est *non vide*, si $0 < \rho < \dfrac{2}{\lambda_{\max}(A)}$, alors quel que soit $u_0 \in \mathbf{R}^n$, la suite $(u_k)_k$ définie par le gradient projeté converge vers le minimum u.

Propriété

Si $K = \displaystyle\prod_{i=1}^{n} [a_i, b_i]$, et $y = {}^t(y_1, \ldots, y_n) \in \mathbf{R}^n$, la *projection* $x = P_K(y)$ a pour composantes :

$$x_i = \min\left(\max(a_i, y_i), b_i\right)$$

4.2 Méthode d'Uzawa

Supposons que l'ensemble

$$K = \left\{v \in \mathbf{R}^n \,/\, Cv = f\right\}$$

ou

$$K = \left\{v \in \mathbf{R}^n \,/\, Cv \leqslant f\right\}$$

où C est une matrice $p \times n$ et $f \in \mathbf{R}^p$.
Soit $u \in K$ le minimum de la fonctionnelle J sur K. Alors il existe $\lambda \in \mathbf{F}$ tel que :

$$\left\{\begin{array}{r} \nabla J(u) + {}^tC\lambda = 0 \\ Cu - f \leqslant 0 \\ \langle \lambda Cu - f \rangle = 0 \\ \langle \lambda - \mu, Cu - f \rangle \geqslant 0,\ \forall \mu \in \mathbf{F} \end{array}\right.$$

$\mathbf{F}$ est si il y a contraintes d'égalité égal à $\mathbf{R}^p$, si il y a contraintes d'inégalité égal à $\mathbf{R}^{+p}$.

Propriété

Soit $\lambda \in \mathbf{R}^p$ tel que (u, λ) verifie le système ci-dessus. Pour tout $\rho > 0$, on a :

$$\lambda = P_{\mathbf{F}}(\lambda + \rho(Cu - f))$$