République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A. Mira - Béjaia
Faculté des Sciences Exactes
Département de Mathématiques

Cours

Intitulé

Optimisation non linéaire en dimension finie avec contraintes

Troisième année Licence de Mathématiques

Chargé de cours : BOURAINE M.

Année universitaire 2020/2021

Table des matières

Chapitre 1

Minimisation Avec Contraintes

1.1 Généralités

On s'intéresse au problème de minimisation d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ comportant des contraintes. Donc des problèmes de la forme :

$$(P) \quad \begin{cases} min\ f(x), \\ x\ \in C,\ C \subset \mathbb{R}^n \end{cases}$$

où C est l'ensemble des contraintes, C domaine non vide et fermé de $\mathbb{R}^n$.

Le problème (P) s'écrit aussi sous la forme générale suivante :

$$(P) \quad \begin{cases} min\ f(x), & \\ h_i(x) = 0,\ i = 1, \ldots, p & , \\ g_j(x) \leq 0,\ j = 1, \ldots, q & , \\ x\ C \subset \mathbb{R}^n & . \end{cases}$$

- Les fonctions $f,\ h_i,\ i = \overline{1,p}$ et $g_j,\ j = \overline{1,q}$ définies de C dans $\mathbb{R}$ sont de classe $\mathcal{C}^1$;
- Les conditions $h_i(x) = 0,\ i = \overline{1,p},\ g_j(x) \leq 0,\ j = \overline{1,q}$ et $x \in \mathbb{R}^n$ sont appelées **contraintes** du problème (P).

Définition 1.1. (Solution réalisable)

On appelle solution **réalisable** ou **admissible** tout vecteur x vérifiant les contraintes du problème (P).

L'ensemble $C = \{x \in \mathbb{R}^n : h_i(x) = 0, \forall i = \overline{1,p}$ et $g_j(x) \leq 0, \forall j = \overline{1,q}\ \}$ est appelé **ensemble des solutions réalisables** ou **admissibles**.

Définition 1.2. **(Solution optimale)**

On appelle **solution optimale** du problème (P) (ou minimum global) une solution réalisable x^* qui minimise $f(x)$ sur l'ensemble de toutes les solutions réalisables, c.à.d :

$$\forall x \in C, \ f(x^*) \leq f(x).$$

Définition 1.3. **(Optimum local)**

On dit qu'un point x^* est un **optimum (minimum) local** de f si, et seulement si, il existe un voisinage $V(x^*)$ de x^* tel que x^* soit un min global de f sur $V(x^*)$:

$$\forall x \in V(x^*) \cap C, \ f(x^*) \leq f(x).$$

Définition 1.4. **(Minimum strict)**

Un **minimum** est dit **strict** si les inégalités dans les définitions précédentes sont strictes.

Définition 1.5. **(Contrainte active)**

Si pour $x \in C$ et pour $j \in \{1, 2, \ldots, q\}$ on a $g_j(x) = 0$, on dit que la contrainte g_j est **saturée** ou **active** en x.

Une contrainte qui n'est pas active est dite inactive. On note $I(x)$ l'ensemble des indices j correspondants aux contraintes actives en x :

$$I(x) = \{j \in \{1, 2, \ldots, q\} : \ g_j(x) = 0\}.$$

Définition 1.6. **(Direction admissible)**

On dit qu'un vecteur $d \in \mathbb{R}^n \backslash \{0\}$ est une **direction admissible** en $x^0 \in C$ si :

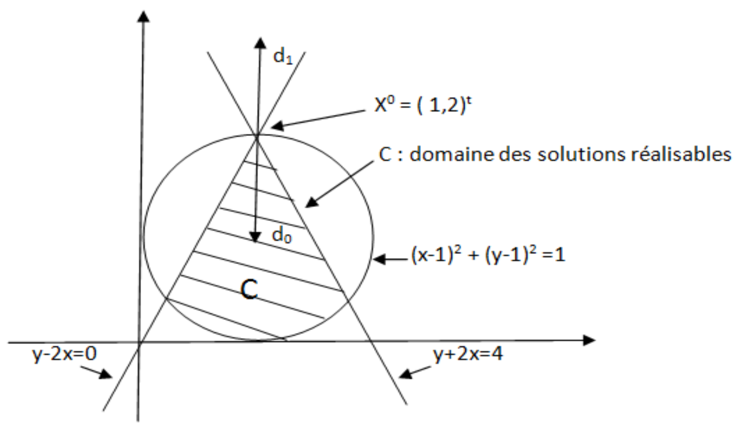
- pour $i \in \{1, 2, \ldots, p\}, \ d^t \cdot \nabla h_i(x^0) = 0,$
- pour $j \in \{1, 2, \ldots, q\}, \ g_j(x^0) = 0 \ \text{ alors } \ d^t \cdot \nabla g_j(x^0) \leq 0.$

Suivre une direction admissible à partir d'un point de C permet de rester dans C ou de le quitter tangentiellement.

Exemple 1.1. *Considérons le problème ci-dessous :*

$$(Po) \quad \begin{cases} min \ f(x,y), \\ g_1(x,y) = (x-1)^2 + (y-1)^2 - 1 \leq 0, \\ g_2(x,y) = y - 2x \leq 0, \\ g_3(x,y) = y + 2x - 4 \leq 0, \\ (x,y) \in \mathbb{R}^2. \end{cases}$$

Considérons les directions $d_0 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ *et* $d_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ *et le point* $X^0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.



1. *Quelles sont les contraintes actives en* X^0 *?*

   $g_1(X^0) = g_1(1,2) = (1-1)^2 + (2-1)^2 - 1 = 0$, *alors* g_1 *est active.*

   $g_2(X^0) = g_2(1,2) = 2 - 2 = 0$, *alors* g_2 *est active.*

   $g_3(X^0) = g_3(1,2) = 2 + 2 - 4 = 0$, *alors* g_3 *est active.*

   Donc $I(X^0) = I(1,2) = \{1,2,3\}$.

2. d_0 *est-elle une direction admissible pour* f *en* X^0 *?*

   Il faut avoir $\langle \nabla g_j(X^0), d_0 \rangle \leq 0, \forall j = \overline{1,3}$ *car* g_1, g_2 *et* g_3 *sont actives.*

   $\nabla g_1(x,y) = \begin{pmatrix} 2(x-1) \\ 2(y-1) \end{pmatrix} \ \Rightarrow \ \nabla g_1(1,2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$,

   $\nabla g_2(x,y) = \nabla g_2(1,2) = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ *et* $\nabla g_3(x,y) = \nabla g_3(1,2) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

   $\langle \nabla g_1(1,2), d_0 \rangle = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = -2 \leq 0$;

   $\langle \nabla g_2(1,2), d_0 \rangle = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = -1 \leq 0$;

$\langle \nabla g_3(1,2), d_0 \rangle = \begin{pmatrix} 0 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = -1 \leq 0.$

Donc d_0 est une direction admissible pour f en X^0.

3. *d_1 est-elle une direction admissible pour f en X^0 ?*

   $\langle \nabla g_1(1,2), d_1 \rangle = \begin{pmatrix} 0 & 11 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = 2 > 0\,;$

   Donc d_1 n'est pas une direction admissible pour f en X^0.

1.2 Résultats d'existence et d'unicité

Théorème 1.1. ***(Existence)***

Supposons que f est continue, que C est un sous-ensemble fermé non vide de $\mathbb{R}^n$ et que l'une des conditions suivantes est satisfaite :

(a) Soit C est borné ;

(b) Soit f est coercive.

Alors le problème (P) admet au moins une solution.

Preuve. Analogue à celle dans le cas sans contraintes.

On montre que la suite minimisante est bornée soit parce qu'elle est dans C qui est borné, soit parce que la fonction f est coercive.

La limite de la sous-suite extraite est alors dans C puisque cet ensemble est fermé.

C'est donc une solution de (P). □

Théorème 1.2. ***(Existence et unicité)***

Soit f une fonction continue et strictement convexe et soit C est un sous-ensemble non vide, convexe et fermé de $\mathbb{R}^n$.

Si C est borné ou si f est coercive, alors il existe un unique $x^* \in C$ solution de (P).

Preuve. Evidente, à faire en exercice. □

1.3 Conditions d'optimalité du 1^{er} ordre

1.3.1 Conditions d'optimalité du 1^{er} ordre générales

Théorème 1.3. ***(Condition nécessaire du 1^{er} ordre)***

Si f est une fonction gâteaux différentiable et si C est un convexe et fermé de $\mathbb{R}^n$, alors toute solution x^* du problème (P) vérifie la propriété suivante :

$$\forall\ x \ \in \ C, \quad \langle \nabla f(x^*), x - x^* \rangle \geq 0. \tag{1.1}$$

Preuve. Soient x^* une solution du problème (P) et $x \in C$, alors :

$$\forall x \in C,\ f(x^*) \leq f(x).$$

C est convexe, alors

$$x^* + t(x - x^*) \ \in C,\ \forall t \in [0, 1].$$

Donc

$$\forall x \in C,\ f\left(x^* + t(x - x^*)\right) - f(x^*) \geq 0.$$

D'où

$$\forall x \in C,\ \forall t \in [0, 1],\ \frac{f(x^* + t(x - x^*)) - f(x^*)}{t} \geq 0.$$

Alors

$$\forall x \in C,\ \forall t \in [0, 1],\ \lim_{t \to 0^+} \frac{f(x^* + t(x - x^*)) - f(x^*)}{t} \geq 0.$$

Par conséquent,

$$\forall\ x \ \in \ C, \langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

□

Théorème 1.4. ***(C.N.S. du 1^{er} ordre dans le cas convexe)***

Soient f une fonction convexe et gâteaux différentiable et C est un convexe et fermé de $\mathbb{R}^n$. Soit x^* un élément quelconque de C.

La conditions (1.1) est nécessaire et suffisante pour que x^* soit solution du problème (P).

Preuve. La condition est nécessaire (voir Théorème 1.3). Il reste à montrer qu'elle est suffisante.

Soit x^* un élément de C, comme f est convexe, alors :

$$\forall x \in C,\ f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle.$$

Puisque $\langle \nabla f(x^*), x - x^* \rangle \geq 0$, alors $\forall x \in C,\ f(x) \geq f(x^*)$.

D'où x^* est solution du problème (P). $\square$

1.3.2 Contraintes en égalité

Lorsqu'on a que des contraintes d'égalité, le problème (P) se réduit à la forme suivante :

$$(P_e) \quad \begin{cases} min\ f(x), \\ h_i(x) = 0,\ i = 1, \ldots, p \\ x \in \mathbb{R}^n. \end{cases},$$

Théorème 1.5. ***(C.N. du*** 1^{er} ***ordre - Contraintes en égalité)***

On suppose que :

- f, h_i *pour* $i = \overline{1,p}$, *sont de classe* $\mathcal{C}^1$ *dans* $\mathbb{R}^n$ *;*
- *Le problème* (P_e) *admet une solution* x^* *;*
- *Les* p *vecteurs de* $\mathbb{R}^n$ *:* $\nabla h_1(x^*), \nabla h_2(x^*), \ldots, \nabla h_p(x^*)$ *sont linéairement indépendants (et donc* $p \leq n$*).*

Alors il existe p *réels* $\lambda_1^*,\ \lambda_2^*, \ldots,\ \lambda_p^*$ *tels que*

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^*\ \nabla h_i(x^*) = 0. \tag{1.2}$$

Preuve. Analogue à celle du théorème de Karush-Kuhn-Tucker qui sera démontré par la suite. $\square$

Définition 1.7. (Multiplicateurs de Lagrange)

Les réels $\lambda_1^*,\ \lambda_2^*, \ldots,\ \lambda_p^*$ du théorème 1.5 sont appelés **multiplicateurs de Lagrange**.

1.3.3 Contraintes en égalité et en inégalité

Considérons un problème d'optimisation sous forme générale :

$$(P) \quad \begin{cases} min\ f(x), \\ h_i(x) = 0,\ i = 1, \ldots, p \quad, \\ g_j(x) \leq 0,\ j = 1, \ldots, q \quad, \\ x \in \mathbb{R}^n. \end{cases}$$

Théorème 1.6. ***(Conditions d'optimalité non qualifiées)***

On suppose que f, h_i pour $i = \overline{1, p}$ et g_j pour $j = \overline{1, q}$ sont de classe $\mathcal{C}^1$. Soit x^* une solution de (P).
Alors il existe $\lambda^* = (\lambda_1^*,\ \lambda_2^*, \ldots,\ \lambda_p^*) \in \mathbb{R}^p$, $\mu^* = (\mu_1^*,\ \mu_2^*, \ldots,\ \mu_q^*) \in \mathbb{R}_+^q$ et $\mu_0 \in \mathbb{R}_+$ tels que :

- $\forall j \in \{0,\ 1,\ \ldots, q\},\ \mu_j^* \geq 0;$ (1.3)
- $h_i(x^*) = 0$ *pour* $i = 1, \ldots, p$ *et* $g_j(x^*) \leq 0$ *pour* $j = 1, \ldots, q;$ (1.4)
- $\forall j \in \{1,\ \ldots, q\} :\ \mu_j^* \cdot g_j(x^*) = 0;$ (1.5)
- $\mu_0^* \cdot \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \cdot \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \cdot \nabla g_j(x^*) = 0.$ (1.6)

Preuve. La preuve se fait par une méthode de pénalisation. Cette méthode consiste à remplacer le problème (P) avec contraintes par la suite de problèmes sans contraintes :

$$(P_k) \quad \begin{cases} min\ f_k(x) = f(x) + k\ \alpha(x), \\ x \in \mathbb{R}^n. \end{cases}$$

où $\alpha : \mathbb{R}^n \to \mathbb{R}$ est une fonction de pénalisation des contraintes et $k > 0$. α est choisie de façon à avoir (P) et (P_k) équivalents (c.à.d. ayant les mêmes solutions).

Pour tout entier k, considérons le problème :

$$(P_k) \quad \begin{cases} min\ f_k(x), \\ x \in B(x^*, \rho). \end{cases}$$

où

$$f_k(x) = f(x) + \frac{k}{2} \sum_{i=1}^{p} [h_i(x)]^2 + \frac{k}{2} \sum_{j=1}^{q} \left[g_j^+(x)\right]^2 + \|x - x^*\|^2, \tag{1.7}$$

avec $g_j^+(x) = max(0, g_j(x))$ et $B(x^*, \rho)$ est une boule fermée (compact) centrée en x^* et de rayon $\rho > 0$.

Remarque 1.1. Les fonction g_j sont de classe $\mathcal{C}^1$, alors $(g_j^+)^2$ est différentiable et

$$\frac{d(g_j^+)^2}{dx}(x) = 2\ \frac{dg_j}{dx}(x)\ g_j^+(x).$$

• **Le problème $(\mathbf{P_k})$ a au moins une solution $\mathbf{x_k}$**

En effet, f_k est continue, elle atteint donc son minimum sur le compact $B(x^*, \rho)$.

• **La suite $(\mathbf{x_k})$ converge vers $\mathbf{x}^*$**

La suite (x_k) est dans le compact $B(x^*, \rho)$ et on peut en extraire une sous-suite, noté (x_k), qui converge vers $\tilde{x} \in B(x^*, \rho)$.

Comme $f_k(x_k) \leq f_k(x^*) = f(x^*) < +\infty$ car x_k est le mimnimum de f_k sur $B(x^*, \rho)$, nous avons alors

$$\sum_{i=1}^{p} [h_i(x_k)]^2 + \sum_{j=1}^{q} \left[g_j^+(x_k)\right]^2 \leq \frac{2}{k}\left[f(x^*) - f(x_k) - \|x_k - x^*\|^2\right]. \tag{1.8}$$

$[f(x^*) - f(x_k) - \|x_k - x^*\|^2]$ étant borné, donc

$$\forall i = \overline{1, p},\ \lim_{k\to+\infty} h_i(x_k) = h_i(\tilde{x}) = 0,$$

et

$$\forall j = \overline{1, q},\ \lim_{k\to+\infty} g_j^+(x_k) = g_j^+(\tilde{x}) = 0.$$

Alors $\tilde{x}$ est une solution réalisable.

D'autre part :

$$f(x_k) + \|x_k - x^*\|^2\ \leq f_k(x_k)\ \leq f(x^*)$$

car $\frac{k}{2}\,[h_i(x_k)]^2 + \frac{k}{2}\sum\limits_{j=1}^{q}\left[g_j^+(x_k)\right]^2 \geq 0$ (c'est une somme de carrés).

Comme x^* est une solution du problème (P), alors

$$f(\tilde{x}) + \|\tilde{x} - x^*\|^2\ \leq f(x^*)\ \leq f(\tilde{x}).$$

Donc $\|\tilde{x} - x^*\|^2 = 0$. Par conséquent, $\tilde{x} = x^*$.

Ce raisonnement peut être fait pour toute valeur d'adhérence de la suite (x_k).

Donc la suite (x_k) converge vers x^*.

• **Condition d'optimalité pour $(\mathbf{P_k})$**

Comme (x_k) converge vers x^*, elle est dans $B(x^*, \rho)$ à partir d'un certain rang et donc x_k est un minimum local (sans contraintes) de f_k. Par conséquent, $\nabla f_k(x_k) = 0$ pour k assez grand, c.à.d.

$$\nabla f(x_k) + k \sum_{i=1}^{p} [h_i(x_k)\ \nabla h_i(x_k)] + k \sum_{j=1}^{q} \left[g_j^+(x_k)\ \nabla g_j(x_k)\right] + 2(x_k - x^*) = 0. \tag{1.9}$$

Posons

$$S_k = \left(1 + k^2 \sum_{i=1}^{p} [h_i(x_k)]^2 + k^2 \sum_{j=1}^{q} \left[g_j^+(x_k)\right]^2\right)^{\frac{1}{2}},$$

$$\mu_0^k = \frac{1}{S_k}, \quad \lambda_i^k = \frac{k\ h_i(x_k)}{S_k}, \text{ pour } i = \overline{1,p} \text{ et } \mu_j^k = \frac{k\ g_j^+(x_k)}{S_k}, \text{ pour } j = \overline{1,q}.$$

La relation (1.9) devient :

$$\mu_0^k\ \nabla f(x_k) + \sum_{i=1}^{p} \left[\lambda_i^k\ \nabla h_i(x_k)\right] + \sum_{j=1}^{q} \left[\mu_j^k\ \nabla g_j(x_k)\right] + \frac{2}{S_k}(x_k - x^*) = 0. \tag{1.10}$$

Comme le vecteur $(\lambda_1^k, \ \ldots, \ \lambda_p^k, \mu_0^k, \mu_1^k, \ldots, \mu_q^k) \in \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q$ est de norme 1, on peut en extraire une sous-suite convergente vers $(\lambda_1^*, \ \ldots, \ \lambda_p^*, \mu_0^*, \mu_1^*, \ldots, \mu_q^*) \neq 0$ et passer à la limite dans (1.10). Alors, on aura :

$$\mu_0^*\ \nabla f(x^*) + \sum_{i=1}^{p} [\lambda_i^*\ \nabla h_i(x^*)] + \sum_{j=1}^{q} \left[\mu_j^*\ \nabla g_j(x^*)\right] = 0 \tag{1.11}$$

car toutes les fonctions considérées sont continues et $S_k \geq 1$. Notons que $\mu_j^* \geq 0, \ \forall j = \overline{0,q}$.

• **Relation (1.5)**

Si $g_j(x^*) < 0$, alors $g_j(x_k) < 0$ à partir d'un certain rang et $\mu_j^k = 0$. Par conséquent, en passant à la limite $\mu_j^* = 0$. □

Remarque 1.2. Quelques remarques importantes :

1. Les réels λ_i^* et μ_j^* sont les multiplicateurs de Lagrange.

2. La relation (1.4) est une relation de **réalisabilité**, c.à.d. que tout point x vérifiant (1.4) est une solution réalisable.

3. La relation (1.5) est une relation de **complémentarité**.

4. Les conditions du théorème (1.6) sont dites **non qualifiées** car le réel μ_0^* peut être nul et on n'a pas de renseignements sur le minimum de f puisqu'elle n'apparaît nulle part dans la relation d'optimalité. Donc il est important de donner des conditions qui permettent d'assurer que μ_0^* soit non nul.

De telles conditions sont dites **conditions de qualification** ou de **régularité**. Lorsqu'elles sont vérifiées le problème est dit **qualifié**.

Définition 1.8. **(Point régulier ou condition de qualification 1)**

On dit qu'un élément $x^* \in \mathbb{R}^n$ est régulier pour les contraintes h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$:

- s'il est réalisable : $h_i(x^*) = 0$ pour $i = \overline{1,p}$ et $g_j(x^*) \leq 0$ pour $j = \overline{1,q}$:
- si les vecteurs $\nabla h_i(x^*)$ pour $i = \overline{1,p}$ sont linéairement indépendants
- et si on peut trouver une direction $d \in \mathbb{R}\backslash\{0\}$ tel que

$$\langle \nabla h_i(x^*), d \rangle = 0,\ \ i = \overline{1,p}\ \ \text{ et } \langle \nabla g_j(x^*), d \rangle < 0,\ \ \forall j \in I(x^*).$$

On dit aussi que x^* vérifie la condition de qualification de Mangasarian-Fromowitz que nous noterons $(CQ1)$.

La condition $(CQ2)$ suivante est plus forte que la condition $(CQ1)$.

Définition 1.9. **(Condition de qualification 2)**

On dit qu'un élément $x^* \in \mathbb{R}^n$ est régulier pour les contraintes h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$:

- s'il est réalisable,
- et si les vecteurs $\nabla h_i(x^*)$, $\nabla g_j(x^*)$ pour $i = \overline{1,p}$ et $j = \overline{1,q}$ sont linéairement indépendants

Théorème 1.7. ***(Conditions de Karush-Kuhn-Tucker)***

On suppose que f, h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$ sont de classe $\mathcal{C}^1$. Soit x^* une solution de (P).

On suppose que x^* est un point régulier pour les contraintes $h_i,\ i=\overline{1,p}$ et $g_j,\ j=\overline{1,q}$.
Alors il existe $\lambda^*=(\lambda_1^*,\ \lambda_2^*,\dots,\ \lambda_p^*)\in\mathbb{R}^p$ et $\mu^*=(\mu_1^*,\ \mu_2^*,\dots,\ \mu_q^*)\in\mathbb{R}_+^q$ tels que :

- $\forall j\in\{1,\ \dots,q\},\ \mu_j^*\geq 0;$ (1.12)
- $h_i(x^*)=0$ *pour* $i=\overline{1,p}$ *et* $g_j(x^*)\leq 0$ *pour* $j=\overline{1,q};$ (1.13)
- $\forall j\in\{1,\ \dots,q\}:\ \mu_j^*\ g_j(x^*)=0;$ (1.14)
- $\nabla f(x^*)+\sum_{i=1}^{p}\lambda_i^*\ \nabla h_i(x^*)+\sum_{j=1}^{q}\mu_j^*\ \nabla g_j(x^*)=0.$ (1.15)

Preuve. On montre que sous les hypothèses de régularité (CQ1), le réel μ_0^* donné dans le théorème (1.6) est non nul.
On a $(\lambda_1^*,\ \lambda_2^*,\dots,\ \lambda_p^*,\ \mu_0^*,\ \mu_1^*,\ \mu_2^*,\dots,\ \mu_q^*)\ \neq 0$. On suppose que $\mu_0^*=0$ et supposons aussi que $\mu_j^*=0,\ \forall j\in I(x^*)$.
Alors le vecteur $\lambda^*=(\lambda_1^*,\ \lambda_2^*,\dots,\ \lambda_p^*)\neq 0$, on a alors $\sum\limits_{i=1}^{p}\lambda_i^*\ \nabla h_i(x^*)=0$. Ce résultat contredit l'indépendances linéaire des $\nabla h_i(x^*),\ i=\overline{1,p}$.
Donc il existe $j_0\in I(x^*)$ tel que $\mu_{j_0}^*\neq 0$.
Nous avons, en prenant la direction d donnée par (CQ1) :

$$0=\sum_{i=1}^{p}\lambda_i^*\langle\ \nabla h_i(x^*),d\rangle+\sum_{j=1}^{q}\mu_j^*\ \langle\nabla g_j(x^*),d\rangle<\mu_{j_0}^*\ \langle\nabla g_{j_0}(x^*),d\rangle<0,$$

d'où la contradiction.
L'ensemble des contraintes (1.12)-(1.15) sont appelées **conditions de Karush-Kuhn-Tucker**, notées **K.K.T.** □

Définition 1.10. **(Lagrangien)**

On appelle **lagrangien** du problème (P) la fonction définie de $\mathbb{R}^n\times\mathbb{R}^p\times\mathbb{R}^q$ dans $\mathbb{R}$ par :

$$\mathcal{L}(x,\lambda,\mu)=f(x)+\sum_{i=1}^{p}\lambda_i\ h_i(x)+\sum_{j=1}^{q}\mu_j\ g_j(x).$$

Remarque 1.3. La relation (1.15) s'écrit

$$\nabla_x\mathcal{L}(x^*,\lambda^*,\mu^*)=0,$$

où ∇_x désigne le gradient du lagrangien par rapport à la première variable.

Dans le cas convexe, nous avons le résultat suivant :

Théorème 1.8. ***(C.N.S. dans le cas convexe)***

On suppose que f, h_i pour $i = \overline{1,p}$ et g_j pour $j = \overline{1,q}$ sont de classe $\mathcal{C}^1$ et que f et g_j, $j = \overline{1,q}$ sont convexes et h_i, $i = \overline{1,p}$ sont affines. On suppose aussi que x^* est un point régulier. Alors,

$$(\ x^* \text{ est solution de } (P)\ ) \ \Leftrightarrow\ (\ \text{les conditions } (1.12) - (1.15) \text{ sont satisfaites }).$$

Preuve. On montre que si les conditions (1.12)-(1.15) sont satisfaites alors x^* est une solution du problème (P).

De (1.13), on déduit que x^* est réalisable. Comme toutes les fonctions sont convexes, alors le lagrangien est convexe par rapport à la variable x et la condition (1.15) est équivalente à dire que x^* est un minimum de $\mathcal{L}(x, \lambda^*, \mu^*)$. On obtient donc :

$$\forall\ x \in \mathbb{R}^n : \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*).$$

Si $x \in C$, alors $h_i(x) = 0$ et $g_j(x) \leq 0$ de sorte que

$$\sum_{i=1}^{p} \lambda_i^*\ h_i(x) + \sum_{j=1}^{q} \mu_j^*\ g_j(x) = \sum_{j=1}^{q} \mu_j^*\ g_j(x) \leq 0,$$

puisque $\mu_j^* \geq 0$ d'après (1.12).

De plus avec la relation de complémentarité (1.14), on voit que $\mathcal{L}(x^*, \lambda^*, \mu^*) = f(x^*)$.

Finalement, on obtient

$$f(x^*) = \mathcal{L}(x^*, \lambda^*, \mu^*) \leq \mathcal{L}(x, \lambda^*, \mu^*) = f(x^*) \leq f(x).$$

Ce qui prouve que x^* est solution du problème (P). $\square$

Exemple 1.2. *Considérons le problème suivant :*

$$(P') \quad \begin{cases} \min\ f(x,y) = 2x^2 + 2xy + y^2 - 10x - 10y, \\ g_1(x,y) = x^2 + y^2 - 5 \leq 0, \\ g_2(x,y) = 3x + y - 6 \leq 0, \\ (x,y) \in \mathbb{R}^2. \end{cases}$$

1. *Ce problème admet-il une solution ? Est-elle unique ?*

- *f est continue,*
- *f est coercive car c'est une fonction quadratique dont la matrice associée* $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ *est définie positive,*
- *Le domaine des contraintes est fermé car son complémentaire est un ouvert,*

Donc le problème (P') admet au moins une solution.

- *f est strictement convexe car c'est une fonction quadratique de matrice associée définie positive,*
- *Le domaine des solutions réalisables est convexe car g_1 et g_2 sont convexes.*

Par conséquent, le problème (P') admet une solution unique.

2. $X = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

a. *Quelles sont les contraintes actives en X ?*

$g_1(1,2) = 1 + 4 - 5 = 0$, *g_1 est donc active,*

$g_2(1,2) = 3 + 2 - 6 = -1$, *$g_2$ est donc inactive, d'où $I(1,2) = \{1\}$.*

b. *X est-il un point régulier ?*

- *X est réalisable :* $g_1(1,2) = 0 \leq 0$ *et* $g_2(1,2) = -1 \leq 0$,
- $\exists ? d \in \mathbb{R}^2 \backslash \{0\} : \ \langle \nabla g_1(X), d \rangle < 0$.

  $\nabla g_1(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \ d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$.

  Donc $\langle \nabla g_1(1,2), d \rangle = 2d_1 + 4d_2 < 0$, *d'où* $d_1 < -2d_2$.

  Il existe une infinité de solutions. Il suffit de prendre $d_2 = 1$ et $d_1 = -3$.

c. *X vérifie-t-il les conditions de K.K.T. ? Conclure.*

Les conditions de K.K.T. donnent :

- $\mu_1 \geq 0$ *et* $\mu_2 \geq 0$, *(1)*
- $g_1(1,2) \leq 0$ *et* $g_2(1,2) \leq 0$, *(2)*
- $\mu_1 \ g_1(1,2) = 0$ *et* $\mu_2 \ g_2(1,2) = 0$, *(3)*
- $\nabla f(1,2) + \mu_1 \ \nabla g_1(1,2) + \mu_2 \ \nabla g_2(1,2) = 0$. *(4)*

$\nabla f(x,y) = \begin{pmatrix} 4x + 2y - 10 \\ 2x + 2y - 10 \end{pmatrix} \ \Rightarrow \ \nabla f(1,2) = \begin{pmatrix} -2 \\ -4 \end{pmatrix}$.

(3) $\Rightarrow \ \mu_2 = 0$ *car g_2 est inactive.*

(3) et (4) $\Rightarrow \begin{cases} -2 + 2\mu_1 = 0 \\ -4 + 4\mu_1 = 0 \end{cases} \quad \Rightarrow \ \mu_1 = 1 \geq 0.$

D'où $X = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ *vérifie les conditions de K.K.T.*

Conclusion : X vérifie les conditions de K.K.T., alors c'est l'unique minimum de (P').

Remarque 1.4. Dans l'exemple précédent nous avons utilisé les conditions d'optimalité pour vérifier si une solution réalisable particulière d'un problème d'optimisation est une solution optimale. Par contre, dans l'exemple qui suit on utilisera les conditions d'optimalité pour trouver la solution optimale d'un problème d'optimisation.

Exemple 1.3. *Considérons le problème suivant :*

$$(Pc) \quad \begin{cases} min\ f(x,y,z) = x^2 + y^2 + z^2, \\ h(x,y,z) = 2y - 2z - 1 = 0, \\ g(x,y,z) = 1 - x \leq 0, \\ (x,y,z) \in \mathbb{R}^3. \end{cases}$$

Il est facile de montrer que le problème (Pc) *admet une solution unique. Déterminons alors cette solution.*

Les conditions de K.K.T. donnent :

- $\lambda \in \mathbb{R}$ *et* $\mu \in \mathbb{R}^+$,
- $2y - 2z - 1 = 0$ *et* $1 - x \leq 0$,
- $\mu\ (1 - x) = 0$,
- $\nabla f(x,y,z) + \lambda \nabla h(x,y,z) + \mu \nabla g(x,y,z) = 0.$

$$\nabla f(x,y,z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix},\ \nabla h(x,y,z) = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix} \text{ et } \nabla g(x,y,z) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

On distingue deux cas : g inactive ou g active.

1er ***cas :*** *g est inactive* $\Rightarrow\ \mu = 0$.

Le système à résoudre est alors :

$$\begin{cases} 2x = 0 \\ 2y + 2\lambda = 0 \\ 2z - 2\lambda = 0 \\ 2y - 2z = 0 \end{cases} \Leftrightarrow \begin{cases} x = 0 \\ y = -\lambda \\ z = \lambda \\ -2\lambda - 2\lambda = 1 \end{cases} \Leftrightarrow \begin{cases} x = 0, \\ y = \frac{1}{4}, \\ z = -\frac{1}{4}, \\ \lambda = -\frac{1}{4} \end{cases}$$

$X = \begin{pmatrix} 0 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$ *est une solution rejetée car elle n'est pas réalisable.*

2nd ***cas :*** *g est active* $\Leftrightarrow\ 1 - x = 0\ \Leftrightarrow\ x = 1$ *et* $\mu = 0$.

Le système à résoudre est alors :

$$\begin{cases} 2x - \mu = 0 \\ 2y + 2\lambda = 0, \\ 2z - 2\lambda = 0, \\ 2y - 2z = 0, \\ 1 - x = 0 \end{cases} \Leftrightarrow \begin{cases} x = 1, \\ y = \frac{1}{4}, \\ z = -\frac{1}{4}, \\ \lambda = -\frac{1}{4} \in \mathbb{R}, \\ \mu = 2 \geq 0, \end{cases}$$

Donc $X^* = \begin{pmatrix} 1 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$, $\lambda = -\frac{1}{4}$ *et* $\mu = 2$.

X^* *est la solution du problème* (Pc) *si c'est un point régulier.*

- X^* *est réalisable car* $h(X^*) = 0$ *et* $g(X^*) = 0 \leq 0$.
- $\nabla h(X^*) = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix} \neq 0$ *donc libre.*
- *Soit* $d = (d_1, d_2, d_3)^t \in \mathbb{R}^3 \backslash \{0\}$.
  $$\begin{cases} \langle \nabla h(X^*), d \rangle = 0 \\ \langle \nabla g(X^*), d \rangle < 0 \end{cases} \Leftrightarrow \begin{cases} 2d_2 - 2d_3 = 0 \\ -d_1 < 0 \end{cases} \Leftrightarrow \begin{cases} d_2 = d_3 \\ d_1 > 0 \end{cases}$$
  Il suffit de prendre $d_1 = d_2 = d_3 = 1$.

  Donc X^* *est un point régulier.*

Conclusion : $X^* = \begin{pmatrix} 1 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$ *est la solution du problème* (Pc).

1.4 Conditions d'optimalité du second ordre

Les conditions d'optimalité du premier ordre permettent de déterminer les bons condidats à la solution de (P). Les Conditions d'optimalité du second ordre vont permettre dans un premier temps de restreindre encore le nombre de candidats.

Théorème 1.9. ***(Condition nécessaire du second ordre)***

On suppose que f, h_i, $i = \overline{1,p}$ *et* g_j, $j = \overline{1,q}$ *sont de classe* $\mathcal{C}^2$, *que* x^* *est un minimum de* f *sur* C *et que la condition (CQ1) est vérifiée.*

Alors il existe $\lambda^* = (\lambda_1^*, \ \lambda_2^*, \ldots, \ \lambda_p^*) \in \mathbb{R}^p$ *et* $\mu^* = (\mu_1^*, \ \mu_2^*, \ldots, \ \mu_q^*) \in \mathbb{R}^q$ *tels que :*

- *Les relations (1.12)-(1.15) de K.K.T. sont satisfaites et*
- *Pour toute direction* $d \in \mathbb{R}^n \backslash \{0\}$ *vérifiant :*

$$\begin{cases} \langle \nabla h_i(x^*), d \rangle = 0, & pour \ i = \overline{1,p}; \\ \langle \nabla g_j(x^*), d \rangle = 0, & pour \ j \in I^+(x^*); \\ \langle \nabla g_j(x^*), d \rangle \leq 0, & pour \ j \in I(x^*) \backslash I^+(x^*); \end{cases} \tag{1.16}$$

où $I^+(x^*) = \{j \in \{1, 2, \ldots, q\} : \ g_j(x^*) = 0 \ et \ \mu_j^* > 0\}$*, on a*

$$\langle \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) \ d, \ d\rangle \geq 0, \tag{1.17}$$

avec

$$\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) = \nabla^2 f(x) + \sum_{i=1}^{p} \lambda_i^* \ \nabla^2 h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \ \nabla^2 g_j(x^*)$$

désigne la seconde dérivée de $\mathcal{L}$ *au point* (x^*, λ^*, μ^*).

Preuve. Voir théorème (1.10) de Hiriart-Urruty J.B., *L'Optimisation que sais-je ?*. P.U.F., Paris, 1996. □

Définition 1.11. (Contraintes fortement actives)

L'ensemble $I^+(x^*)$ est l'ensemble des **contraintes fortement actives**.
Lorsque $I^+(x^*) = I(x^*)$, c.à.d. $g_j(x^*) = 0 \quad \Leftrightarrow \quad \mu_j^* > 0$ on dit qu'il y'a stricte complémentarité.

Le résultat suivant donne une condition suffisante du second ordre.

Théorème 1.10. ***(Condition suffisante du second ordre)***

On suppose que f, h_i, $i = \overline{1,p}$ *et* g_j, $j = \overline{1,q}$ *sont de classe* $\mathcal{C}^2$. *Soit* $x^* \in \mathbb{R}^n$ *vérifiant les conditions de K.K.T. avec les multiplicateurs de Lagrange* λ^*, μ^*.
Si la matrice hessienne du lagrangien au point (x^*, λ^*, μ^*), $\nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*)$, *est définie positive sur le sous-espace*

$$\tau = \{d \in \mathbb{R}^n \backslash \{0\} : \langle \nabla h_i(x^*), d\rangle = 0, \ i = \overline{1,p} \ et \ \langle \nabla g_j(x^*), d\rangle = 0, \ j \in I^+(x^*)\}$$

alors x^* *est un minimum strict de* f *sur* C.

Preuve. Voir théorème (1.11) de Hiriart-Urruty J.B., *L'Optimisation que sais-je ?*. P.U.F., Paris, 1996. □

Chapitre 2

Applications

Introduction

Dans ce chapitre nous donnons deux problèmes dont la résolution nécessite l'utilisation de problèmes d'optimisation avec contraintes.
Le premier problème est théorique. Il s'agit de la projection d'un point sur un convexe fermé. Il y'a un double objectif derrière ce choix. Le premier est de montrer que les problèmes d'optimisation nous permettent de résoudre des problèmes théoriques. Le deuxième est de maitriser cette notion car on aura à l'utiliser pour développer la méthode du gradient projeté qui sera présentée au troisième chapitre.
Le deuxième problème est appliqué. Il s'agit du problème de régression linéaire. Ce dernier a été abordé au semestre 1 dans le cas sans contraintes. Dans ce chapitre nous traitons le cas où le problème de régression présente des contraintes sur les variables.

La norme, $\|\cdot\|$, utilisée dans le cadre de ce cours désigne la norme euclidienne de $\mathbb{R}^n$.

2.1 Projection sur un convexe fermé

Soit C un sous-ensemble de $\mathbb{R}^n$ et soit $x \in \mathbb{R}^n,\ x \notin C$. On s'intéresse à la distance de ce point à l'ensemble C.

Théorème 2.1. *Soit C sous-ensemble convexe et fermé de $\mathbb{R}^n$ et soit x un élément quelconque de $\mathbb{R}^n$. Alors le problème*

$$\begin{cases} min\ \|x-y\|^2 = \sum_{i=1}^{n}(x_i-y_i)^2, \\ y \in C, \end{cases}$$

a une solution unique $x^* \in C$ qui est caractérisée par :

$$\forall\ y \in C :\ \langle x - x^*, y - x^* \rangle \leq 0.$$

Preuve. Le problème s'écrit :

$$\begin{cases} min\ f(y) = \sum_{i=1}^{n}(x_i-y_i)^2, \\ y \in C, \end{cases}$$

où C est un sous-ensemble non vide, convexe et fermé de $\mathbb{R}^n$.

- La fonction f est continue, coercive et strictement convexe.
- C est un convexe et fermé de $\mathbb{R}^n$.

Alors ce problème admet une solution unique (cf. Théorème 1.2).

Du théorème 1.3 on obtient la caractérisation de x^* :

$$\forall\ y \in C :\ \langle \nabla f(x^*), y - x^* \rangle \geq 0.$$

Comme $\nabla f(x^*) = -2 \begin{pmatrix} x_1 - x_1^* \\ x_2 - x_2^* \\ \vdots \\ x_n - x_n^* \end{pmatrix} = -2(x - x^*)$, alors

$$\langle -2(x - x^*), y - x^* \rangle \geq 0,\ \forall y \in C.$$

Donc

$$\langle x - x^*, y - x^* \rangle \leq 0,\ \forall y \in C. \tag{2.1}$$

□

Corollaire 2.1. *Sous les hypothèses du théorème précédent on peut caractériser le point $x^* \in C$ par :*

$$\langle x^* - y, y - x \rangle \leq 0,\ \forall y \in C. \tag{2.2}$$

Preuve. On montre d'abord la condition nécessaire puis la condition suffisante.

• Condition nécessaire

On a

$$\begin{aligned} \langle x^* - y, y - x\rangle &= \langle x^* - y, y - x + x^* - x^*\rangle,\ \forall y \in C, \\ &= \langle x^* - y, (x^* - x) + (y - x^*)\rangle,\ \forall y \in C, \\ &= \langle x^* - y, x^* - x\rangle - \|x^* - y\|^2,\ \forall y \in C. \end{aligned}$$

Comme $\langle x^* - y, x^* - x\rangle\ \leq 0$ et $\|x^* - y\|^2 \geq 0,\ \forall y \in C$, alors

$$\langle x^* - y, y - x\rangle \leq 0,\ \forall y \in C.$$

• Condition suffisante

Soit $y \in C$ et $z = x^* + \lambda(y - x^*),\ \lambda \in]0, 1[$. $z \in C,\ \forall \lambda \in]0, 1[$ car C est convexe.

La relation (2.2) donne

$$\forall \lambda \in]0, 1[:\ \langle x^* - z, z - x\rangle \leq 0$$

d'où

$$\forall \lambda \in]0, 1[:\ \langle x^* - z, z - x\rangle = -\lambda\langle y - x^*, x^* - x + \lambda(y - x^*)\rangle \leq 0$$

alors

$$\forall \lambda \in]0, 1[:\ \langle y - x^*, x^* - x + \lambda(y - x^*)\rangle \geq 0.$$

On fait tendre λ vers 0^+ pour obtenir :

$$\langle y - x^*, x^* - x\rangle \geq 0, \forall y \in C\ \Rightarrow\ \langle x - x^*, y - x^*\rangle \leq 0,\ \forall y \in C.$$

$\square$

Définition 2.1. (Projeté - Projection)

Le point x^* est **le projeté** du point x sur C.

L'application

$$\begin{aligned} \pi_C :\quad & \mathbb{R}^n \to C \\ & x \mapsto x^* = \pi_C(x) \end{aligned}$$

qui à x associe son **projeté** x^* est **la projection** sur C.

Définition 2.2. **(Distance d'un point à un ensemble)**

On définit la **fonction distance** d'un point x à l'ensemble C par :

$$d(x, C) = \inf_{y \in C} \|x - y\|.$$

Remarque 2.1. Dans le cas où C est un ensemble convexe fermé, on vient de montrer que

$$d(x, C) = \|x - \pi_C(x)\|.$$

Proposition 2.1. *On a*

$$\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n : \ \|\pi_C(x) - \pi_C(y)\| \ \leq \ \|x - y\|.$$

Preuve. Soient x_1 et x_2 deux éléments quelconques de $\mathbb{R}^n$. On applique (2.1) à :
$x = x_1,\ x^* = \pi_C(x_1)$ et $y = \pi_C(x_2) \in C$, puis à : $x = x_2,\ x^* = \pi_C(x_2)$ et $y = \pi_C(x_1) \in C$. On aura

$$\langle x_1 - \pi_C(x_1), \pi_C(x_2) - \pi_C(x_1) \rangle \leq 0,$$

et

$$\langle x_2 - \pi_C(x_2), \pi_C(x_1) - \pi_C(x_2) \rangle \leq 0 \ \Rightarrow \ \langle -x_2 + \pi_C(x_2), \pi_C(x_2) - \pi_C(x_1) \rangle \leq 0.$$

La somme des deux inégalités donne :

$$\langle x_1 - x_2 - \pi_C(x_1) + \pi_C(x_2), \pi_C(x_2) - \pi_C(x_1) \rangle \leq 0$$

alors

$$\langle x_1 - x_2, \pi_C(x_2) - \pi_C(x_1) \rangle + \langle \pi_C(x_2) - \pi_C(x_1), \pi_C(x_2) - \pi_C(x_1) \rangle \leq 0$$

donc

$$\langle x_1 - x_2, \pi_C(x_2) - \pi_C(x_1) \rangle + \|\pi_C(x_2) - \pi_C(x_1)\|^2 \leq 0$$

d'où

$$\|\pi_C(x_2) - \pi_C(x_1)\|^2 \leq \langle x_2 - x_1, \pi_C(x_2) - \pi_C(x_1) \rangle.$$

Avec l'inégalité de Cauchy-Schwartz on aura :

$$\|\pi_C(x_2) - \pi_C(x_1)\|^2 \leq \langle x_2 - x_1, \pi_C(x_2) - \pi_C(x_1) \rangle \leq \|x_2 - x_1\| \cdot \|\pi_C(x_2) - \pi_C(x_1)\|.$$

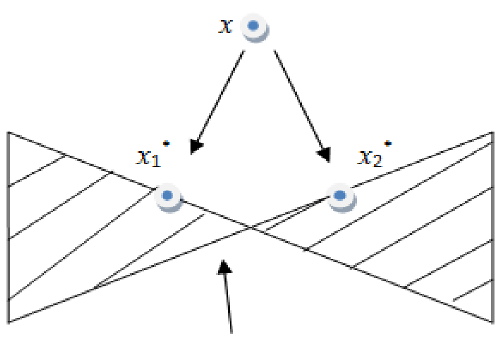
- Si $\pi_C(x_2) = \pi_C(x_1)$ alors la relation cherchée est évidente.
- Si $\pi_C(x_2) \neq \pi_C(x_1)$ alors on divise par $\|\pi_C(x_2) - \pi_C(x_1)\|$ pour avoir

$$\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n : \ \|\pi_C(x_2) - \pi_C(x_1)\| \leq \|x_2 - x_1\|.$$

□

Remarque 2.2. Nous avons les remarques suivantes :

1. Si $x \in C \Rightarrow \pi_C(x) = x$.
2. Si $C = \mathbb{R}^n \Rightarrow \pi_C = Id_{\mathbb{R}^n}$.
3. Le théorème 2.1 est faux si C n'est pas connexe (problème d'unicité). Nous illustrons cette situation dans l'exemple qui suit :



C : domaine des solutions réalisables
x_1^* 1[er] projeté de x sur C et x_2^* 2[ème] projeté de x sur C

4. Le théorème 2.1 est également faux si C n'est pas fermé (problème d'existence). Nous pouvons avoir une idée sur ce résultat à travers l'exemple suivant :

$$C = \{(x, y) \in \mathbb{R}^2 : \ x^2 + y^2 < 4\}$$

est le disque ouvert de centre $(0,0)$ est de rayon 2.
Il n'y a pas de point de C réalisant la distance de $(2,3)$ à C. Le seul point possible se situe sur le cercle de centre $(0,0)$ est de rayon 2, mais ce dernier n'appartient pas à C.

Exemple 2.1. *On calcule la distance de l'origine au domaine C donné par*

$$C = \{X = (x, y, z) \in \mathbb{R}^3 : \ 2y - 2z = 1 \ \textit{et} \ 1 - x \leq 0\}.$$

On détermine d'abord le projeté de l'origine $X^0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ *sur le domaine C.*

Ce problème se modélise sous forme d'un problème d'optimisation avec contraintes. Le problème à résoudre est comme suit :

$$\begin{cases} \textit{min} \ f(x, y, z) = \|X - X^0\|^2 = (x-0)^2 + (y-0)^2 + (z-0)^2 = x^2 + y^2 + z^2, \\ h(x, y, z) = 2y - 2z - 1 = 0, \\ g(x, y, z) = 1 - x \leq 0. \end{cases}$$

Soit $\lambda \in \mathbb{R}$ et $\mu \in \mathbb{R}$. Les conditions de K.K.T. donnent

- $\lambda \in \mathbb{R}, \ \mu \geq 0,$
- $2y - 2z - 1 = 0$ *et* $1 - x \leq 0,$
- $\mu(1 - x) = 0,$
- $\nabla f(X) + \lambda \ \nabla h(X) + \mu \ \nabla g(X) = 0.$

$$\nabla f(X) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \ \nabla h(X) = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix} \ \textit{et} \ \nabla g(X) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}.$$

On distingue deux cas :

1er *cas* : *g est inactive* $\Rightarrow \ \mu = 0.$

Le système à résoudre est alors

$$\nabla f(X) + \lambda \ \nabla h(X) = 0 \ \Leftrightarrow \begin{cases} 2x = 0, \\ 2y + 2\lambda = 0, \\ 2z - 2\lambda = 0, \\ 2y - 2z = 1. \end{cases} \Leftrightarrow \begin{cases} x = 0, \\ y = -\lambda, \\ z = \lambda, \\ -2\lambda - 2\lambda = 1. \end{cases} \Leftrightarrow \begin{cases} x = 0, \\ y = \frac{1}{4}, \\ z = -\frac{1}{4}, \\ \lambda = -\frac{1}{4}. \end{cases}$$

$X = \begin{pmatrix} 0 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$ *est une solution rejetée car elle est non réalisable.*

2ème *cas* : *g est active* $\Rightarrow \ 1 - x = 0$ *et* $\mu \geq 0.$

Le système à résoudre est alors

$$\nabla f(X) + \lambda \ \nabla h(X) = +\mu \ \nabla g(X) = 0 \ \Leftrightarrow \begin{cases} 2x - \mu = 0, \\ 2y + 2\lambda = 0, \\ 2z - 2\lambda = 0, \\ 2y - 2z = 1, \\ 1 - x = 0. \end{cases} \Leftrightarrow \begin{cases} x = 1, \\ y = \frac{1}{4}, \\ z = -\frac{1}{4}, \\ \lambda = -\frac{1}{4}, \\ \mu = 2. \end{cases}$$

Donc $X^* = \begin{pmatrix} 0 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$, $\lambda = -\frac{1}{4} \in \mathbb{R}$ *et* $\mu = 2 \geq 0$.

Vérifions que X^* *est un point régulier.*

- X^* *est réalisable car* $h(X^*) = 0$ *et* $g(X^*) \leq 0$
- $\nabla h(X^*) = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix} \neq 0$ *donc libre*
- $\exists$? $d = (d_1, d_2, d_3)^t \in \mathbb{R}^3 \setminus \{0\} : \begin{cases} \langle \nabla h(X^*), d \rangle = 0, \\ \langle \nabla h(X^*), d \rangle < 0. \end{cases}$

  $\begin{cases} \langle \nabla h(X^*), d \rangle = 0, \\ \langle \nabla h(X^*), d \rangle < 0. \end{cases} \Leftrightarrow \begin{cases} 2d_2 - 2d_3 = 0, \\ -d_1 < 0, \end{cases} \Leftrightarrow \begin{cases} d_1 > 0, \\ d_2 = d_3. \end{cases}$

  Il suffit de prendre $d_1 = 1,\ d_2 = d_3 = 1$. *Donc* X^* *est un point régulier.*

Conclusion : $X^* = \begin{pmatrix} 0 \\ \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$ *est la projection du point* $X^0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ *sur le domaine* C.

On détermine la distance l'origine $X^0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ *au domaine* C.

$$d(X^0, C) = \|X^* - X^0\| = \|X^*\| = \sqrt{1 + (\frac{1}{4})^2 + (-\frac{1}{4})^2} = \frac{3}{4}\sqrt{2}.$$

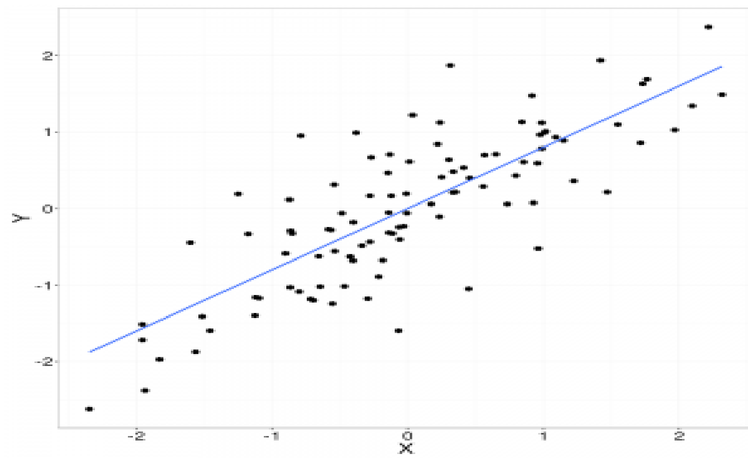
2.2 Régression linéaire avec contraintes

Considérons un nuage de n points de $\mathbb{R}^2$

$$N = \{(x_i, y_i) : x_i \in \mathbb{R} \text{ et } y_i \in \mathbb{R}, i = \overline{1, n}\}.$$

Ces données sont souvent le résultat de mesures. On cherche souvent à décrire le comportement global de ce nuage de points. En général ces points ne sont pas alignés, mais dans le cas particulier où la tendance est linéaire, on cherche la droite qui approche au mieux ce nuage de points.

On cherche donc la droite de régression d'équation $y = ax + b$, ce qui conduit à minimiser, par la méthode des moindres carrés, $\sum_{i=1}^{n} (y_i - ax_i - b)^2$ sous la contrainte par exemple $b \geq 0$.



Le problème est alors

$$(P_{rl}) \begin{cases} min\ f(a,b) = \sum\limits_{i=1}^{n} (y_i - ax_i - b)^2, \\ g(a,b) = -b \leq 0, \\ (a,b) \in \mathbb{R}^2, \end{cases}$$

$\star$ Le problème de régression linéaire (P_{rl}) admet-il une solution ? est-elle unique ?

Calculons d'abord ∇f, $\nabla^2 f$ et ∇g en tout point $(a,b) \in \mathbb{R}^2$.

$$\frac{\partial f}{\partial a}(a,b) = \sum_{i=1}^{n} 2(ax_i + b - y_i)x_i = 2a \sum_{i=1}^{n} x_i^2 + 2b \sum_{i=1}^{n} x_i - 2 \sum_{i=1}^{n} x_i y_i$$

$$\frac{\partial f}{\partial b}(a,b) = \sum_{i=1}^{n} 2(ax_i + b - y_i) = 2a \sum_{i=1}^{n} x_i + 2nb - 2 \sum_{i=1}^{n} y_i$$

Posons $S_{x^2} = \sum\limits_{i=1}^{n} x_i^2$; $S_x = \sum\limits_{i=1}^{n} x_i$; $S_y = \sum\limits_{i=1}^{n} y_i$ et $S_{xy} = \sum\limits_{i=1}^{n} x_i y_i$. Alors

$$\nabla f(a,b) = \begin{pmatrix} 2S_{x^2}\ a + 2S_x\ b - 2S_{xy} \\ 2S_x\ a + 2n\ b - 2S_y \end{pmatrix},\ \nabla^2 f(a,b) = 2 \begin{pmatrix} S_{x^2} & S_x \\ S_x & n \end{pmatrix} \text{ et } \nabla g(a,b) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

- f est continue (fonction polynômiale),
- f est strictement convexe et coercive. En effet, le premier mineur principal de la matrice hessienne $\nabla^2 f(a,b)$, $\Delta_1 = 2S_{x^2} > 0$, et le deuxième mineur principal $\Delta_2 = 4(nS_{x^2} - S_x^2) > 0$ (en réalité $\Delta_2 \geq 0$, mais le cas $\Delta_2 = 0$ est trivial, il correspond au cas où toutes les observations sont égales). Donc la matrice hessienne est définie positive, d'où f est coercive et strictement convexe.

– l'ensemble $C = \{(a,b) \in \mathbb{R}^2 : b \geq 0\}$ est non vide, fermé (son complémentaire est un ouvert) est convexe (demi-plan).

Le problème admet donc une solution unique.

$*$ Quels sont les points réguliers ?

Tout point (a,b) est régulier. En effet, il existe $d = (d_1, d_2)^t \in \mathbb{R}^2 \backslash \{0\}$, tel que
$\langle \nabla g(a,b), d\rangle = (d_1, d_2)\begin{pmatrix} 0 \\ -1 \end{pmatrix} = -d_2 < 0$, donc $d_1 \in \mathbb{R}$ et $d_2 > 0$.
Il suffit de prendre $d_1 = 1$ et $d_2 = 1$.

$*$ Quelle est la solution optimale ?

Les conditions de K.K.T. donnent :

- $\exists\ \mu \geq 0$,
- $-b \leq 0$,
- $\mu\ b = 0$,
- $\nabla f(a,b) + \mu\ \nabla g(a,b) = 0$.

Donc

- $\mu \geq 0,\ \ b \geq 0$,
- $\mu\ b = 0$,
- $a\ S_{x^2} + b\ S_x = S_{xy}$,
- $a\ S_x + b\ n - \mu = S_y$.

Si $b > 0 \Rightarrow \mu = 0$, on résoud alors le système ci-dessus. Alors

1. Si la solution b trouvée est strictement positive, on garde cette solution, puis on calcule a.
2. Sinon $b = 0$, alors $a = \frac{S_{xy}}{S_{x^2}}$.

Travail à faire : Résoudre l'exercice suivant

On considère le nuage de points $N = \{M_i = (x_i, y_i), i = 1, \ldots, n\}$. Le problème consiste à ajuster ce nuage de points par le modèle linéaire $y = ax + b$ sous la condition $a \geq 1$.

1. En utilisant la méthode des moindres carrés, modéliser ce problème sous forme d'un problème de minimisation.
2. Déterminer les valeurs de a et b qui réalisent le minimum.

3. Application numérique :

x_i	0	1	2	3	4	5	6	7	8	9	10
y_i	0	2	2	4	3	5	5	6	9	8	12

Chapitre 3

Algorithmes

Introduction

Les méthodes d'optimisation non linéaires avec contraintes se divisent en deux familles :

– méthodes directes ou primales ;

– méthodes utilisant la notion de dualité.

Les méthodes primales opèrent directement sur le problème donné. Elles engendrent une suite de solutions réalisables en assurant une décroissance monotone de la fonction objectif à minimiser. Leur avantage, c'est qu'elles présentent une solution approchée satisfaisant les contraintes si le processus itératif est interrompu. Leurs inconvénients c'est la difficulté de mise au point et la propriété de convergence globale est souvent difficile à obtenir.

Les méthodes duales sont plus robustes et la convergence globale est souvent plus facile à obtenir ; par contre, elles présentent l'inconvénient de ne fournir une solution primale réalisable qu'en fin de convergence.

Le principe des méthodes d'optimisation avec contraintes est le même que dans le cas sans contraintes. Il s'agit donc de construire une suite minimisante qui converge vers la solution optimale qui appartient bien sûr au domaine des solutions réalisables.

Nous aborderons dans ce chapitre la méthode du gradient projeté, la méthode de Lagrange-Newton pour des contraintes en égalité, les méthodes de pénalisations, les méthodes SQP et la méthode d'Uzawa.

3.1 Méthode du gradient projeté

Cette méthode s'inspire de la méthode du gradient dans le cas sans contraintes. Néanmoins, lorsqu'en minimise sur un ensemble de contraintes C et qu'à l'itération k, $x^k \in C$, on n'est pas sûr que l'itéré suivant $x^{k+1} = x^k - \rho_k \nabla f(x^k)$ soit dans C. Il faut donc le ramener dans C grâce à une projection sur C.
La méthode du gradient projeté est résumée dans l'algorithme suivant :

Algorithme : Gradient Projeté

1. **Initialisation** $(\mathbf{k = 0})$

   Choix de x^0, $\rho_0 > 0$ et $\varepsilon > 0$.

2. **Itération k**

   $x^{k+1} = \pi_C \left(x^k - \rho_k \nabla f(x^k)\right)$.

3. **Critère d'arrêt**

   - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop et fin ;

   - Sinon $k = k + 1$ et retourner à l'étape (2).

La convergence de la méthode est donnée dans le résultat ci-après :

Théorème 3.1. *Soit f de classe $\mathcal{C}^1$ de $\mathbb{R}^n$ dans $\mathbb{R}$. On suppose que f est elliptique et de dérivée lipschitzienne :*

$$\exists\ \alpha > 0,\ \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,\ \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \alpha \|x - y\|^2;$$

$$\exists\ M > 0,\ \forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n,\ \|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|.$$

Alors, si on choisit la pas ρ_k dans un intervalle $[\beta_1, \beta_2]$ tel que $0 < \beta_1 < \beta_2 < \frac{2\alpha}{M^2}$, la suite $(x^k)_{k \in \mathbb{N}}$ définie par la méthode du gradient projeté converge vers la solution du problème (P).

Preuve. f étant lipschitzienne, alors elle est coercive et strictement convexe (voir cours optimisation sans contraintes). Dans ce cas le problème (P) admet une solution unique notée x^*. Alors

$$x^* = \pi_C \left(x^* - \rho \nabla f(x^*)\right),\ \forall\ \rho > 0. \tag{3.1}$$

Du théorème 1.3, on obtient

$$\forall x \in C, \langle [x^* - \rho \ \nabla f(x^*)] - x^*, x - x^* \rangle \leq 0$$

et comme

$$x^{k+1} = \pi_C \left(x^k - \rho_k \ \nabla f(x^k) \right)$$

et en soustrayant (3.1) avec $\rho = \rho_k$, on obtient :

$$x^{k+1} - x^* = \pi_C \left(x^k - \rho_k \ \nabla f(x^k) \right) - \pi_C \left(x^* - \rho_k \ \nabla f(x^*) \right).$$

Comme la projection est contractante, on aura :

$$\begin{array}{rcl} \|x^{k+1} - x^*\|^2 & = & \|\pi_C \left(x^k - \rho_k \ \nabla f(x^k) \right) - \pi_C \left(x^* - \rho_k \ \nabla f(x^*) \right) \|^2 \\ & \leq & \|x^k - \rho_k \ \nabla f(x^k) - x^* + \rho_k \ \nabla f(x^* \|^2 \\ & \leq & \|x^k - x^*\|^2 + \rho_k^2 \|\nabla f(x^k) - \nabla f(x^* \|^2 - 2\rho_k^2 \langle x^k - x^*, \nabla f(x^k) - \nabla f(x^*) \rangle. \end{array}$$

Comme f est elliptique et de dérivée lipschitzienne, alors

$$\|x^{k+1} - x^*\|^2 \leq (1 + M^2 \rho_k^2 - 2\alpha\rho_k) \|x^k - x^*\|^2.$$

Si on choisit le pas ρ_k dans un intervalle $[\beta_1, \beta_2]$ tel que $0 < \beta_1 < \beta_2 < \frac{2\alpha}{M^2}$, on obtient :

$$\|x^{k+1} - x^*\| \leq K \|x^k - x^*\|,$$

où K est une constante de l'intervalle $]0, 1[$ indépendante de k.
La suite (x^k) converge donc vers x^*.

□

Avantage : Méthode simple.

Inconvénient : Difficile à appliquer puisqu'à chaque itération il faut calculer le projeté d'un vecteur de $\mathbb{R}^n$ sur C.

3.2 Méthode de Lagrange-Newton pour des contraintes en égalité

On distingue deux cas : le cas quadratique et le cas général.
Le cas quadratique est très particulier. Il concerne les problèmes d'optimisation où la fonction à minimiser est quadratique et les contraintes sont affines. Sa résolution revient à résoudre un système d'équations linéaires.
Le cas général concerne les problèmes d'optimisation où la fonction à minimiser et les contraintes sont quelconques. Sa résolution se ramène à résoudre une suite de problèmes quadratiques.

3.2.1 Cas quadratique

Considérons le problème quadratique suivant avec des contraintes en égalités affines :

$$(P_Q) \quad \begin{cases} min f(x) = \frac{1}{2}x^tQx - C^tx = \frac{1}{2}\langle Qx, x\rangle - \langle C, x\rangle, \\ Ax = b \;\Leftrightarrow\; h(x) = Ax - b = 0, \end{cases}$$

où

- Q matrice carrée d'ordre n ;
- x et C deux vecteurs de $\mathbb{R}^n$;
- A matrice de type $p \times n$;
- b vecteur de $\mathbb{R}^p$;
- h une fonction définie de $\mathbb{R}^n$ dans $\mathbb{R}^p$.

Les conditions de K.K.T. pour le problème (P_Q) donnent

- $\exists\ \lambda \in \mathbb{R}^p$;
- $h(x) = Ax - b = 0$;
- $\nabla\ f(x) + \lambda^t\ \nabla(Ax - b) = 0$.

Ceci implique :

$$\begin{cases} Qx + A^t\lambda = C \\ Ax \qquad\quad = b \end{cases} \quad \Leftrightarrow \quad \left[\begin{array}{c|c} Q & A^t \\ \hline A & 0 \end{array}\right] \left[\begin{array}{c} x \\ \hline \lambda \end{array}\right] = \left[\begin{array}{c} C \\ \hline b \end{array}\right].$$

Ainsi la solution optimale (x^*, λ^*) est solution de ce système linéaire.

Si $M = \left[\begin{array}{ccc} Q & | & A^t \\ -- & -- & -- \\ A & | & 0 \end{array}\right]$ est inversible, ce système admet une solution unique.

Exemple 3.1. *Considérons le problème ci-dessous :*

$$(P'_Q) \quad \begin{cases} min f(x,y) = x^2 + xy + 2y^2 \\ h_1(x,y) = x + y - 1 = 0 \\ h_2(x,y) = -2x + y = 0 \end{cases}$$

$f(x,y) = \frac{1}{2}(x,y)\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - (0,0)\begin{pmatrix} x \\ y \end{pmatrix}$, *donc* $Q = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$ *et* $C = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\begin{cases} h_1(x,y) = 0 \\ h_2(x,y) = 0 \end{cases} \Leftrightarrow \begin{cases} x + y = 1 \\ -2x + y = 0 \end{cases} \Leftrightarrow \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$,

ainsi $A = \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix}$ *et* $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Soient λ_1 *et* λ_2 *dans* $\mathbb{R}$, *on obtient donc le système linéaire suivant :*

$$\left[\begin{array}{ccccc} 2 & 1 & | & 1 & -2 \\ 1 & 4 & | & 1 & 1 \\ -- & -- & | & -- & -- \\ 1 & 1 & | & 0 & 0 \\ -2 & 1 & | & 0 & 0 \end{array}\right] \left[\begin{array}{c} x \\ y \\ - \\ \lambda_1 \\ \lambda_2 \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ - \\ 1 \\ 0 \end{array}\right] \Leftrightarrow \begin{cases} 2x + y + \lambda_1 - 2\lambda_2 = 0 & (1) \\ x + 4y + \lambda_1 + \lambda_2 = 0 & (2) \\ x + y = 1 & (3) \\ -2x + y = 0 & (4) \end{cases}$$

$(4) \Leftrightarrow y = 2x$, *de* (3) *on obtient* $3x = 1$. *Alors* $x = 1/3$ *et* $y = 2/3$.

(1) *et* (2) *donnent alors le système suivant :* $\begin{cases} \lambda_1 - 2\lambda_2 = -4/3 \\ \lambda_1 + \lambda_2 = -9/3 \end{cases}$

D'où $\lambda_1 = -5/9 \in \mathbb{R}$ *et* $\lambda_2 = -22/9 \in \mathbb{R}$.

Donc $X^* = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$ *est la solution optimale du problème* (P'_Q).

3.2.2 Cas général

Dans le cas général, nous avons le problème suivant :

$$(P_e) \begin{cases} min\ f(x) \\ h(x) = 0 \end{cases}$$

où $h = (h_1, h_2, \ldots, h_p)$ et $h_i : \mathbb{R}^n \to \mathbb{R}$ pour $i = \overline{1,p}$.

Les conditions de K.K.T. pour le problème (P_e) donnent :

- $\exists\ \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p) \in \mathbb{R}^p$
- $\nabla_x\ \mathcal{L}(x,\lambda) = 0$, avec $\mathcal{L}(x,\lambda) = f(x) + \lambda^t h(x)$

- $h(x) = 0$

D'où

$$\begin{cases} \nabla\ f(x) + \sum\limits_{i=1}^{p} \lambda_i \nabla\ h_i(x) = 0 \\ h_i(x) = 0,\ \forall i = \overline{1,p} \end{cases} \tag{3.2}$$

On linéarise les deux expressions de (3.2) à partir d'un point (x^k, λ^k) au voisinage de x^k pour (x^{k+1}, λ^{k+1}) la solution du système linéarisé.

La linéarisation d'une fonction F au voisinage de (x_0, y_0) est donnée par

$$F(x,y) = F(x_0,y_0) + (x - x_0)\frac{\partial F}{\partial x}(x_0,y_0) + (y - y_0)\frac{\partial F}{\partial y}(x_0,y_0).$$

$$\begin{cases} \nabla f(x^k) + \sum\limits_{i=1}^{p} \lambda_i^k \nabla h_i(x^k) + \left[\nabla^2 f(x^k) + \sum\limits_{i=1}^{p} \lambda_i^k \nabla^2 h_i(x^k)\right](x^{k+1} - x^k) + \sum\limits_{i=1}^{p}(\lambda_i^{k+1} - \lambda_i^k)\nabla h_i(x^k) = 0 \\ h_i(x^k) + \langle \nabla\ h_i(x^k), x^{k+1} - x^k\rangle = 0,\ \forall i = \overline{1,p} \end{cases}$$

Donc

$$\begin{cases} \left[\nabla^2 f(x^k) + \sum\limits_{i=1}^{p} \lambda_i^k \nabla^2 h_i(x^k)\right](x^{k+1} - x^k) + \sum\limits_{i=1}^{p}(\lambda_i^{k+1} - \lambda_i^k)\nabla h_i(x^k) = -\left[\nabla f(x^k) + \sum\limits_{i=1}^{p} \lambda_i^k \nabla h_i(x^k)\right] \\ \langle \nabla\ h_i(x^k), x^{k+1} - x^k\rangle = -h_i(x^k),\ \forall i = \overline{1,p} \end{cases}$$

Comme $\mathcal{L}(x,\lambda) = f(x) + \sum\limits_{i=1}^{p} \lambda_i h_i(x)$, alors nous avons le système suivant :

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x^k,\lambda^k) & | & \nabla h_1(x^k) & \dots & \nabla h_p(x^k) \\ ----- & | & ---- & -- & --- \\ \nabla h_1^t(x^k) & | & & & \\ \vdots & | & & 0 & \\ \nabla h_p^t(x^k) & | & & & \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ ---- \\ \lambda^{k+1} - \lambda^k \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}(x^k,\lambda^k) \\ ----- \\ -h(x^k) \end{bmatrix} \tag{3.3}$$

Donc

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x^k,\lambda^k) & | & \nabla h^t(x^k) \\ ------ & - & --- \\ \nabla h(x^k) & | & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ ---- \\ \lambda^{k+1} - \lambda^k \end{bmatrix} = \begin{bmatrix} -\nabla f(x^k) - \langle \lambda, \nabla h(x^k)\rangle \\ ---------- \\ -h(x^k) \end{bmatrix} \tag{3.4}$$

Remarque 3.1. Nous donnons ci-après quelques remarques concernant cette méthode :

1. Cette méthode converge de façon quadratique vers (x^*, λ^*) (c.f. Ortega et Rheinboldt1970, Powel 1978).

2. Pour mettre en oeuvre cette méthode , on a besoin à chaque étape de calculer :

– le gradient de f et $h_i, i = \overline{1,p}$;

– Le hessien en x du lagrangien $\mathcal{L}(x, \lambda)$.

3. Si on ajoute dans les n premières équations le terme $\langle \nabla h(x^k), \lambda^k \rangle$, on aura :

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x^k,\lambda^k) & | & \nabla h^t(x^k) \\ ----- & - & ---- \\ \nabla h(x^k) & | & 0 \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ ---- \\ \lambda^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) \\ ---- \\ h(x^k) \end{bmatrix} \tag{3.5}$$

Algorithme : Lagrange-Newton

1. **Initialisation** $(\mathbf{k = 0})$

   Choix de $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^p$ et $\varepsilon > 0$.

2. **Itération k**

   On connaît (x^k, λ^k) et résoudre

$$\begin{bmatrix} \nabla^2_{xx}\mathcal{L}(x^k,\lambda^k) & | & \nabla h^t(x^k) \\ ------ & | & --- \\ \nabla h(x^k) & | & 0 \end{bmatrix} \begin{bmatrix} d_k \\ -- \\ y_k \end{bmatrix} = - \begin{bmatrix} \nabla \mathcal{L}^t(x^k,\lambda^k) \\ ---- \\ h(x^k) \end{bmatrix}$$

$$x^{k+1} = x^k + d_k \quad \text{et } \lambda^{k+1} = \lambda^k + y_k$$

3. **Critère d'arrêt**

   - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop et fin ;

   - Sinon $k = k + 1$ et retourner à l'étape (2).

Remarque 3.2. Les inconvénients de cette méthode et les moyens pour y remédier :

1. Inconvénients de cette méthode :

   – Les relations (3.2) sont vérifiées non seulement pour le minimum mais aussi pour le maximum avec contraintes.

   – La suite (x^k, λ^k) peut ne pas converger si (x^0, λ^0) est choisi très éloigné de la solution optimale (x^*, λ^*).

2. Pour remédier à ce problème, on peut recherché une bonne approximation de (x^*, λ^*) par d'autres méthodes, par la suite, on applique la méthode de Lagrange Newton à partir de ce point.

3. Des méthodes quasi-newtoniennes peuvent être utilisées et ceci pour éviter de calculer les dérivés secondes.

Exemple 3.2. *Considérons le problème ci-dessous :*

$$(P_{LN}) \quad \begin{cases} min\ f(X) = f(x,y) = \frac{1}{3}x^3 + xy + \frac{1}{3}y^3 \\ h_1(X) = h_1(x,y) = (x-1)^2 + y^2 - 1 = 0 \\ h_2(X) = h_2(x,y) = x^2 + (y-1)^2 - 1 = 0 \end{cases}$$

- *Soit* $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2$ *;*
- $\nabla h_1(X) = 0$ *et* $\nabla h_2(X) = 0$ *;*
- $\nabla_X \mathcal{L}(X, \lambda) = \nabla f(X) + \sum\limits_{i=1}^{2} \lambda_i\ \nabla h_i(X) = 0.$

On aura alors le système suivant :

$$\begin{cases} \nabla f(X) + \lambda_1\ \nabla h_1(X) + \lambda_2\ \nabla h_2(X) = 0 \\ h_1(X) = 0 \\ h_2(X) = 0 \end{cases}$$

$\nabla f(X) = \begin{pmatrix} x^2+y \\ x+y^2 \end{pmatrix}$; $\nabla^2 f(X) = \begin{pmatrix} 2x & 1 \\ 1 & 2y \end{pmatrix}$; $\nabla h_1(X) = \begin{pmatrix} 2x-2 \\ 2y \end{pmatrix}$;
$\nabla h_2(X) = \begin{pmatrix} 2x \\ 2y-2 \end{pmatrix}$; $\nabla^2 h_1(X) = \nabla^2 h_2(X) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

Donc

$$\begin{aligned} \nabla_X \mathcal{L}(X, \lambda) &= \nabla f(X) + \lambda_1\ \nabla h_1(X) + \lambda_2\ \nabla h_2(X) \\ &= \begin{pmatrix} x^2+y \\ x+y^2 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x-2 \\ 2y \end{pmatrix} + \lambda_2 \begin{pmatrix} 2x \\ 2y-2 \end{pmatrix} \\ &= \begin{pmatrix} x^2+y+2\lambda_1 x + 2\lambda_2 x - 2\lambda_1 \\ x+y^2+2\lambda_1 y + 2\lambda_2 y - 2\lambda_2 \end{pmatrix}. \end{aligned}$$

et

$$\begin{aligned} \nabla^2_{XX} \mathcal{L}(X, \lambda) &= \nabla^2 f(X) + \lambda_1\ \nabla^2 h_1(X) + \lambda_2\ \nabla^2 h_2(X) \\ &= \begin{pmatrix} 2x & 1 \\ 1 & 2y \end{pmatrix} + \lambda_1 \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 2x+2\lambda_1+2\lambda_2 & 1 \\ 1 & 2y+2\lambda_1+2\lambda_2 \end{pmatrix}. \end{aligned}$$

Au point (X^k, λ^k), *avec* $X^k = \begin{pmatrix} x^k \\ y^k \end{pmatrix}$ *et* $\lambda^k = \begin{pmatrix} \lambda_1^k \\ \lambda_2^k \end{pmatrix}$, *on obtient le système suivant :*

$$(S_k) \quad \begin{bmatrix} 2x^k + 2\lambda_1^k + 2\lambda_2^k & 1 & | & 2x^k - 2 & 2x^k \\ 1 & 2y^k + 2\lambda_1^k + 2\lambda_2^k & | & 2y^k & 2y^k - 2 \\ -------- & -------- & - & ---- & ---- \\ 2x^k - 2 & 2y^k & | & 0 & 0 \\ 2x^k & 2y^k - 2 & | & 0 & 0 \end{bmatrix} \begin{bmatrix} d_k \\ -- \\ b_k \end{bmatrix} = -A_k$$

où

$$A_k = \begin{pmatrix} (x^k)^2 + y^k + 2\lambda_1^k x^k + 2\lambda_2^k x^k - 2\lambda_1^k \\ x^k + (y^k)^2 + 2\lambda_1^k y^k + 2\lambda_2^k y^k - 2\lambda_1^k \\ --------------- \\ (x^k - 1)^2 + (y^k)^2 - 1 \\ (x^k)^2 + (y^k - 1)^2 - 1 \end{pmatrix}, \quad X^{k+1} = X^k + d_k \ \ et \ \ \lambda^{k+1} = \lambda^k + b_k.$$

On obtient donc une suite de systèmes linéaires dont la résolution génère une suite de solution (X^k) *qui converge vers la solution optimale du problème* (P_{LN}).

3.3 Méthodes de pénalisation

Introduction

Le principe des méthodes de pénalité réside dans la transformation du problème d'optimisation avec contraintes en une suite de problèmes d'optimisation sans contraintes en ajoutant à la fonction à optimiser une pénalité en cas de violation de celles-ci. Ainsi le problème avec contraintes (P) suivant :

$$(P) \quad \begin{cases} min \ f(x) \\ x \in C, \ \ x \in \mathbb{R}^n \end{cases}$$

sera remplacé par la suite de problèmes sans contraintes suivant :

$$(P_r) \quad \begin{cases} min \ f_r(x) = f(x) + r \ \alpha(x) \\ x \in \mathbb{R}^n \end{cases}$$

où $\alpha : \mathbb{R}^n \to \mathbb{R}$ est une fonction de pénalisation des contraintes et $r > 0$.

Le but est de trouver des fonctions α telles que les problèmes (P) et (P_r) soient équivalents, donc ayant les mêmes solutions.

On distingue deux méthodes de pénalités :

- **Méthode de pénalités intérieures :** Elle consiste à approcher le minimum de l'intérieur du domaine des contraintes C ;
- **Méthode de pénalités extérieures :** Elle consiste à approcher le minimum depuis l'extérieur du domaine des contraintes C.

3.3.1 Méthode de pénalités extérieures

$*$ **La fonction de pénalisation** α

La fonction α de pénalité extérieure est égale à zéro si x est admissible et supérieure à zéro si x ne l'est pas. Ainsi $\begin{cases} \alpha(x) = 0, & \text{si } x \in C, \\ \alpha(x) \geq 0, & \text{si } x \notin C. \end{cases}$
On prend $\alpha : \mathbb{R}^n \to \mathbb{R}$ vérifiant les conditions suivantes :

$$\begin{cases} (a) \;\; \alpha \;\text{ est continue sur }\; \mathbb{R}^n; \\ (b) \;\; \forall x \in \mathbb{R}^n, \; \alpha(x) \geq 0; \\ (c) \;\; \alpha(x) = 0 \Leftrightarrow x \in C. \end{cases} \qquad (C_\alpha)$$

Voici un exemple de fonctions α, utilisées souvent dans la littérature, pour différents types de contraintes :

Contraintes	$x \leq 0$	$h(x) = 0$	$g(x) \leq 0$
Fonction α	$\|x^+\|^2$	$\|h(x)\|^2$	$\|g(x)^+\|^2$

avec $g(x)^+ = max(0, g(x))$ et $x^+ = (x_1^+, x_2^+, \ldots, x_n^+)^t$ où $x_i^+ = max(0, x_i)$.

$*$ **Le paramètre r**

Le paramètre r permet d'amplifier ou de diminuer la valeur de $\alpha(x)$.
Au départ, on choisit une valeur de r pas trop élevée pour éviter d'avoir un point d'accumulation loin de l'optimum ; puis augmenter progressivement celui-ci pour s'approcher de l'ensemble des solutions réalisables du problème (P). Ainsi la suite $(r_k)_k$ utilisée devra être telle que $0 < r_k < r_{k+1}$ avec $r_k \to +\infty$ quand $k \to +\infty$.

$*$ **Résultat de convergence**

Nous avons le résultat de convergence suivant :

Théorème 3.2. *Supposons que f est coercive et continue et soit C un ensemble fermé et non vide.*
On suppose que α vérifie les conditions C_α. Alors :

- $\forall r > 0,\ (P_r)$ *a au moins une solution* x_r *;*
- *La famille* $(x_r)_{r>0}$ *est bornée ;*
- *Toute sous-suite convergente extraite de* $(x_r)_{r>0}$ *converge vers une solution de* (P) *lorsque* $r \to +\infty$.

Preuve. Soient $r > 0$ et x^* une solution de (P).

Comme $f_r(x) = f(x) + r\alpha(x) \geq f(x)$, la coercivité de f entraîne celle de f_r.

De plus la continuité de f et α implique celle de f_r. Par conséquent, (P_r) a au moins une solution de x_r. D'autre part :

$$f(x) \leq f_r(x_r) \leq f_r(x^*) = f(x^*) < \infty, \tag{3.6}$$

car x_r est une solution de (P_r).

Donc $f(x_r)$ est uniformément bornée par rapport à r et comme f_r est coercive, alors la famille $(x_r)_{r>0}$ est bornée.

On peut donc trouver une sous-suite de $(x_r)_{r>0}$, notée $(x_{r_k})_{k>0}$, qui converge vers $\tilde{x}$.

Comme $f(x^*) - f(x_{r_k})$ est borné, alors de la relation (3.6) on obtient :

$$\alpha(x_{r_k}) \leq \frac{1}{r_k}\left[f(x^*) - f(x_{r_k})\right] \Rightarrow 0 \leq \lim_{r_k \to \infty} \alpha(x_{r_k}) = \alpha\left(\lim_{r_k \to \infty} x_{r_k}\right) = \alpha(\tilde{x}) \leq 0 \Rightarrow \alpha(\tilde{x}) = 0.$$

Donc $\alpha(\tilde{x}) \in C$.

On passe à la limite dans (3.6) et on obtient : $f(\tilde{x}) \leq f(x^*) \Rightarrow \tilde{x}$ est une solution de (P) (mais pas nécessairement égal à x^*). □

$*$ **Algorithme de pénalité extérieure**

Algorithme : Pénalité extérieure

1. **Initialisation** $(\mathbf{k = 1})$

   Choix de $x^0 \in \mathbb{R}^n$, $r_1 > 0$ et $\varepsilon > 0$.

2. **Itération k**

   Résoudre le sous-problème

   $$(P_{r_k}) \begin{cases} min\ f_{r_k}(x) = f(x) + r_k\ \alpha(x) \\ x \in \mathbb{R}^n \end{cases}$$

   en prenant x_{k-1} comme point d'initialisation.

3. **Critère d'arrêt**

   - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop et fin ;

   - Sinon $k = k + 1$, on choisit $r_{k+1} > r_k$ et retourner à l'étape (2).

Remarque 3.3. Si la solution du problème (P) n'est pas unique, la suite $(x_k)_{k>0}$ peut osciller entre deux variables d'adhérence.

3.3.2 Méthode de pénalités intérieures

Dans cette méthode, l'optimum est approché par l'intérieur du domaine C. Cette méthode s'applique uniquement aux problèmes d'optimisation dont l'ensemble des contraintes C est défini uniquement par une collection d'inégaltés :

$$(P') \begin{cases} min\ f(x) \\ g_j(x) \leq 0,\ j = \overline{1,q} \end{cases}$$

où f et $g_j(x),\ j = \overline{1,q}$ sont des fonctions continues de $\mathbb{R}^n$ dans $\mathbb{R}$.

L'intérieur de C est défini par :

$$C_I = \{x \in \mathbb{R}^n : g_j(x) < 0,\ \forall j = \overline{1,q}\}.$$

Supposons que l'ensemble des contraintes vérifie :

– l'intérieur de C est non vide $(C_I \neq \emptyset)$;

– Tout point de la frontière de C est limite d'une suite de points appartenant à C_I.

$*$ **La fonction de pénalisation** β

La fonction de pénalité β, appelée aussi fonction barrière vérifie :

- $\beta(x) \geq 0,\ \forall x \in C_I$;
- $\beta(x) \to +\infty$ quand $x \to$ frontière de C.

Le problème consiste donc à minimiser la fonction $\varphi_\varepsilon = f(x) + \varepsilon\beta(x)$ où ε est un réel positif, donc à résoudre :

$$(P_\varepsilon) \begin{cases} min\ \varphi_\varepsilon = f(x) + \varepsilon\beta(x) \\ x \in \mathbb{R}^n \end{cases}$$

Les fonctions de pénalité les plus utilisées dans la littérature sont :

$$\beta_1 = -\sum_{j=1}^{q} \frac{1}{g_j(x)} \quad \text{ou} \quad \beta_2 = -\sum_{j=1}^{q} ln\left(-g_j(x)\right).$$

Remarque 3.4. Si les fonctions $g_j, j = \overline{1, q}$ sont convexes, les fonctions β_1 et β_2 le sont aussi.

$*$ **Le paramètre r**

Le paramètre ε est choisi de façon à ce que la suite $(\varepsilon_k),\ k = 0, 1, \ldots$ vérifie $0 < \varepsilon_{k+1} < \varepsilon_k$ et $\varepsilon_k \to 0$ quand $k \to +\infty$.

$*$ **Résultat de convergence**

Nous avons le résultat de convergence suivant :

Théorème 3.3. *Supposons que f est continue et coercive. Et soit C un ensemble fermé et borné ayant un intérieur non vide et que tout $x \in C$ est limite d'une suite de points appartenant à C_I.*

Soit $\beta : \mathbb{R}^n \to \mathbb{R}$ une fonction de pénalisation intérieure vérifiant :

- $\beta(x) \geq 0,\ \forall x \in C_I$;
- $\beta(x) \to +\infty$ *quand x tend vers la frontière de C ;*
- $\beta(x)$ *est continue sur C_I.*

Alors, lorsque le coefficient de pénalité ε tend vers 0 :

- *la suite $(\widetilde{x_\varepsilon})$ admet au moins un point d'accumulation et tout point d'accumulation de la suite $(\widetilde{x_\varepsilon})$ est un optimum du problème (P_ε).*
- *La quantité $\varepsilon\ \beta(\widetilde{x_\varepsilon}) \to 0$.*

Preuve. Même preuve que pour le théorème 1.2. □

Remarque 3.5. Voiçi quelques remarques concernant cette méthode.

1. L'algorithme correspondant à cette méthode est analogue à celui du cas de pénalités extérieures, sauf qu'à l'étape 3 on choisit $\varepsilon_{k+1} < \varepsilon_k$.
2. Pour appliquer cette algorithme, il faut disposer d'un point initial x^0 situé à l'intérieur de C.
3. Les méthodes de pénalités intérieures sont très intéressantes pour traiter des problèmes d'optimisation avec contraintes fortement non linéaires.
4. Les méthodes de pénalités fournissent de bonnes approximations des multiplicateurs de Karush-Kuhn et Tucker optimaux.

5. Ces méthodes sont simples et efficaces pour obtenir rapidement de bonnes solutions approchées d'un problème de type (P') ainsi que de bonnes approximations des multiplicateurs de Karush-Kuhn et Tucker optimaux. Cependant elles ne permettent pas d'obtenir une précision élevée. C'est pourquoi elles sont utilisées en conjonction avec d'autres méthodes.
   Ainsi, on les utilise pour avoir de bonnes approximations de x^* et μ^*, puis mettre en oeuvre une autre méthode (Newton par exemple) pour résoudre les équations de Karush-Kuhn et Tucker.

Exemple 3.3. *On résout le problème suivant en appliquant la méthode de pénalités extérieures :*

$$(P)\ \begin{cases} min\ f(x,y) = x^2 - 3y^2 \\ x + 2y = 1 \end{cases} \qquad \Leftrightarrow \begin{cases} min\ f(x,y) = x^2 - 3y^2 \\ h(x,y) = x + 2y - 1 = 0 \end{cases}$$

On pose $f_r(x,y) = (x^2 - 3y^2) + r(x + 2y - 1)^2$.

La résolution du problème avec contraintes (P) *revient à la résolution de la suite de problèmes sans contraintes* (P_r) *donnée ci-après :*

$$(P_r)\ \begin{cases} min\ f_r(x,y) = (x^2 - 3y^2) + r(x + 2y - 1)^2 \\ (x,y) \in \mathbb{R}^2,\ \ avec\ \ r > 0 \end{cases}$$

Il suffit alors d'appliquer l'algorithme correspondant à cette méthode. Néanmoins, la résolution analytique de ce problème est très simple.

$*$ ***Gradient de*** $\mathbf{f_r}$ *:*

$$\nabla f_r(x,y) = \begin{pmatrix} 2x + 2r(x + 2y - 1) \\ -6y + 4r(x + 2y - 1) \end{pmatrix}$$

$*$ ***Points critiques :***

$$\nabla f_r(x,y) = 0 \Leftrightarrow \begin{cases} 2x + 2r(x + 2y - 1) = 0 & (1) \\ -6y + 4r(x + 2y - 1) = 0 & (2) \end{cases}$$

$2 \cdot (1) - (2) \Leftrightarrow 2x + 3y = 0 \Leftrightarrow x = -\dfrac{3}{2}y.$

$(1) \Rightarrow 2 \cdot \dfrac{-3}{2}y + 2r\left(\dfrac{-3}{2} + 2y - 1\right) = 0 \Leftrightarrow y(r-3) = 2r,\ alors\ y = \dfrac{2r}{r-3},$

d'où $x = \dfrac{-3r}{r-3}$, $r \neq 3$.

Donc $X_r = \begin{pmatrix} \dfrac{-3r}{r-3} \\ \dfrac{2r}{r-3} \end{pmatrix}$, *avec* $r \neq 3$, *est le seul point critique pour* f_r

$*$ ***Hessien de*** $\mathbf{f_r}$ *:*

$$\nabla^2 f_r(x,y) = \begin{pmatrix} 2+2r & 4r \\ 4r & -6+8r \end{pmatrix}.$$

$*$ ***Nature des critiques :***

- $\Delta_1 = 2 + 2r > 0$ *car* $r > 0$.
- $\Delta_2 = (2+2r)(-6+8r) - 16r^2 = 4r - 12 > 0$ *pour* $r > 3$.

Donc X_r *est la solution du problème* (P_r) *pour* $r > 3$.

$*$ ***Conclusion :***

$X^* = \lim\limits_{r \to +\infty} X_r = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ *est la solution du problème* (P).

3.4 Méthodes de programmation quadratique successive

Les méthodes de programmation quadratique successive, en anglais sequential quadratic programming (SQP), sont basées sur les conditions d'optimalité du 1^{er} et du 2^{nd} ordre. On distingue deux cas : cas des contraintes en égalités du cas général.

3.4.1 Cas des contraintes en égalités

Un problème d'optimisation avec des contraintes en égalités est défini comme suit :

$$(P_e) \begin{cases} min f(x) \\ x \in C \end{cases}$$

avec $C = \{x \in \mathbb{R}^n : \; h_i(x) = 0, \; i = \overline{1,p}\}$.

Une solution x^* du problème (P_e) est un point critique du lagrangien $\mathcal{L}$, mais ce n'est pas, en général, un minimum de ce lagrangien.

L'idée consiste à établir une méthode de descente qui vérifie les conditions nécessaires et suffisantes du 1er ordre et de résoudre une succession de problèmes quadratiques avec des contraintes linéaires.

Considérons l'itéré x^k, on cherche

$$x^{k+1} = x^k + \rho_k d_k$$

où $d_k \in \mathbb{R}^n \setminus \{0\}$ est une direction de descente et le pas $\rho_k > 0$.

On fait une approximation des contraintes h_i grâce à la formule de Taylor :

$$h_i(x^k + d) = h_i(x^k) + \langle d, \nabla h_i(x^k)\rangle + o\left(\|d\|^2\right).$$

Si on néglige les termes d'ordre ≥ 2, on définit la direction d_k comme étant la direction permettant d'assurer que $h_i(x^k + d) \simeq 0$. Donc $h_i(x^k) + \langle d, \nabla h_i(x^k)\rangle \simeq 0, \ \forall i = \overline{1,p}$, c.à.d.

$$Dh(x^k)d_k = -h(x^k) \tag{3.7}$$

où $h = (h_1, h_2, \ldots, h_p)$ et $Dh(x^k)$ est le jacobien de h en x^k. Ceci correspond à une linéarisation des contraintes au voisinage de x^k.

Néanmoins, x^{k+1} doit diminuer la valeur du lagrangien

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^t h(x) = f(x) + \sum_{i=1}^{p} \lambda_i \ h_i(x).$$

Elle sera quadratique puisqu'on ne peut pas se contenter d'une condition du premier ordre

$$\mathcal{L}(x^k + d, \lambda) = \mathcal{L}(x^k, \lambda) + \langle \nabla_x \mathcal{L}(x^k, \lambda), d\rangle + \frac{1}{2}\langle \nabla^2_{xx}\mathcal{L}(x^k, \lambda)d, d\rangle + o\left(\|d\|^3\right).$$

Si on néglige les termes d'ordre ≥ 3, on voit qu'il faut minimiser

$$\langle \nabla_x \mathcal{L}(x^k, \lambda), d\rangle + \frac{1}{2}\langle \nabla^2_{xx}\mathcal{L}(x^k, \lambda)d, d\rangle$$

pour espérer minimiser le lagrangien.

On cherche alors d_k comme solution du problème :

$$(QP_e) \ \begin{cases} min\langle \nabla_x f(x^k), d\rangle + \frac{1}{2}\langle \nabla^2_{xx}\mathcal{L}(x^k, \lambda)d, d\rangle \\ Dh(x^k)d + h(x^k) = 0 \end{cases}$$

puisque

$$\begin{aligned}\langle \nabla_x \mathcal{L}(x^k, \lambda), d\rangle &= \langle \nabla_x f(x^k), d\rangle + \langle \lambda, Dh(x^k)d\rangle \\ &= \langle \nabla_x f(x^k), d\rangle + \langle \lambda, h(x^k)\rangle\end{aligned}$$

avec $\langle \lambda, h(x^k)\rangle$ est une constante.

Le choix de ρ_k et des multiplicateurs λ_k offre beaucoup de possibilités qui donnent lieu à autant de variantes de la méthode.

Algorithme : SQP pour contraintes en égalités avec $\rho_k = 1$

1. **Initialisation**

   Choix de $x^0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^p$ et $\varepsilon > 0$.

2. **Itération k**

   Résoudre le sous-problème quadratique :

$$(QP_e) \left\{ \begin{array}{l} min\ \langle \nabla_x f(x^k), d\rangle + \frac{1}{2}\langle \nabla^2_{xx}\mathcal{L}(x^k, \lambda^k)d, d\rangle \\ Dh(x^k)d + h(x^k) = 0 \end{array} \right.$$

3. λ^{k+1} $\mathbb{R}^p$ est le multiplicateur associé à la contrainte (en égalité) de (QP_e) et

$$x^{k+1} = x^k + d_k.$$

4. **Critère d'arrêt**

   - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop et fin ;

   - Sinon $k = k + 1$ et retourner à l'étape (2).

3.4.2 Cas général

On adopte le même principe que précédemment au cas de contraintes en égalité et en inégalité défini ci-après :

$$(QP) \left\{ \begin{array}{l} min\ f(x) \\ x \in C \end{array} \right.$$

$C = \{x \in \mathbb{R}^n :\ h(x) = 0 \text{ et } g(x) \leq 0\} = \{x \in \mathbb{R}^n :\ h_i(x) = 0, i = \overline{1,p} \text{ et } g_j(x) \leq 0, j = \overline{1,q}\}$,

avec $g = (g_1, g_2, \ldots, g_q)$.

Le lagrangien est alors :

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \lambda^t h(x) + \mu^t g(x).$$

On linéarise les contraintes et on fait une approximation quadratique du lagrangien $\mathcal{L}$. Nous obtenons alors l'algorithme suivant :

Algorithme : SQP - Cas général avec $\rho_k = 1$

1. **Initialisation**

   Choix de $x^0 \in \mathbb{R}^n$, $(\lambda^0, \mu^0) \in \mathbb{R}^p \times \mathbb{R}^q$ et $\varepsilon > 0$.

2. **Itération k**

   Résoudre le sous-problème quadratique :

$$(QP) \begin{cases} min \ \langle \nabla_x f(x^k), d \rangle + \frac{1}{2} \langle \nabla^2_{xx} \mathcal{L}(x^k, \lambda^k, \mu^k) d, d \rangle \\ Dh(x^k)d + h(x^k) = 0 \\ Dg(x^k)d + g(x^k) \leq 0 \end{cases}$$

3. λ^{k+1} et μ^{k+1} sont les multiplicateurs de lagrange associés aux contraintes du problème (QP) et

$$x^{k+1} = x^k + d_k.$$

4. **Critère d'arrêt**

   - Si $\|x^{k+1} - x^k\| < \varepsilon$, stop et fin ;

   - Sinon $k = k + 1$ et retourner à l'étape (2).

3.5 Méthode d'Uzawa

L'idée de cette méthode et de considérer le lagrangien $\mathcal{L}$ au lieu de la fonction f car le lagrangien englobe f et les contraintes du problème h et g. Et aussi utiliser la condition nécessaire du premier ordre pour que x^* soit un minimum de f avec contraintes est que x^* soit un point critique du lagrangien $\mathcal{L}$.

L'aspect théorique de cette méthode et son algorithme font l'objet d'un exercice dans le devoir de vacances.

Bibliographie

[1] M. Berguounioux, *Optimisation et contrôle des systèmes linéaires - cours et exercices avec solutions*. Paris, Francis Lefebvre, 2001. **Cote : 519.6/02**.

[2] M. Bierlaire, *Introduction à l'optimisation différentielle*. Presses polytechniques et universitaires romandes, 2006. **Cote : 519.3/14**.

[3] J.C. Culioli, *Introduction à l'optimisation*. Paris : Ellipses, 2012. **Cote : 519/7.6**.

[4] Y. Dodge, *Optimisation appliquée*. Berlin : Springer, 2004. **Cote : 519.6/37**

[5] J. Gauvin, *Leçons de programmation mathématique*. Ecole Polytechnique de Montreal, 1995. **Cote : 519.7/44**.

[6] J.B. Hiriart-Urruty, *Les mathématiques du mieux faire : Premiers pas en optimisation*. Volume 1, Paris : Ellipses, 2007. **Cote : 519.3/22**.

[7] J.B. Hiriart-Urruty, *Optimisation et analyse convexe - exercices et problèmes corrigés, avec rappels de cours*. EDP Sciences, 2009. **Cote : 519.6/53**.

[8] M. Minoux, *Programmation mathématiques : Théorie et algorithmes*. Tome 1, Editions Dunod, 1983. **Cote : 519.7/52**.