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

Niveaux

Master 1 Probabilités-Statistique et Applications
Troisième année Licence de Mathématiques

Chargé de cours : BOURAINE M.

Année universitaire 2019/2020

Quelques remarques

- Ce cours vous est présenté pour parer à une situation particulière. Il ne remplace pas un cours fait avec les étudiants où il y'a un débat continu ;
- Ce cours vous sera posté chapitre par chapitre ;
- Ce cours est accompagné d'une série de TD qui englobe les trois chapitres du programme ;
- Cette façon de faire nous permettra un gain de temps considérable après la reprise des enseignements ;
- Si vous avez des remarques ou si vous relevez des erreurs, merci de les signaler ;
- Si vous avez des questions n'hésitez pas à me les communiquer.

Quelques recommandations

Je vous recommande d'être méthodique dans votre travail. Il faut réserver trois à quatre séances d'une heure trente à deux heures par semaine. Je vous propose le plan suivant :

- Traiter les pages 3 à 5 du chapitre 1 ;
- Traiter les exercices 1, 2, 3 et 4 de la série de TD ;
- Traiter les pages 6 à 16 du chapitre 1 ;
- Traiter les exercices 5, 6, 7 et 8 de la série de TD ;
- Traiter les pages 17 et 18 du chapitre 1 ;
- Un ou deux exercices vont compléter la série de TD concernant cette partie.

Si vous avez des questions vous pouvez les poser à travers la plate forme e-leanrning ou par mail à l'adresse suivante : mohbouraine@gmail.com

Travaillez bien et bonne santé à tous.

Table des matières

1 Minimisation Avec Contraintes **3**

1.1 Généralités . . . 3

1.2 Résultats d'existence et d'unicité . . . 6

1.3 Conditions d'optimalité du 1$^{\text{er}}$ ordre . . . 6

1.3.1 Conditions d'optimalité du 1$^{\text{er}}$ ordre générales . . . 6

1.3.2 Contraintes en égalité . . . 8

1.3.3 Contraintes en égalité et en inégalité . . . 8

1.4 Conditions d'optimalité du second ordre . . . 17

2 Applications **19**

Introduction . . . 20

2.1 Projection sur un convexe fermé . . . 20

2.2 Régression linéaire avec contraintes . . . 26

Bibliographie **28**

Chapitre 1

Minimisation Avec Contraintes

1.1 Généralités

On s'intéresse au problème de minimisation d'une fonction $f : \mathbb{R}^n \to \mathbb{R}$ 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 :

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

- 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 :

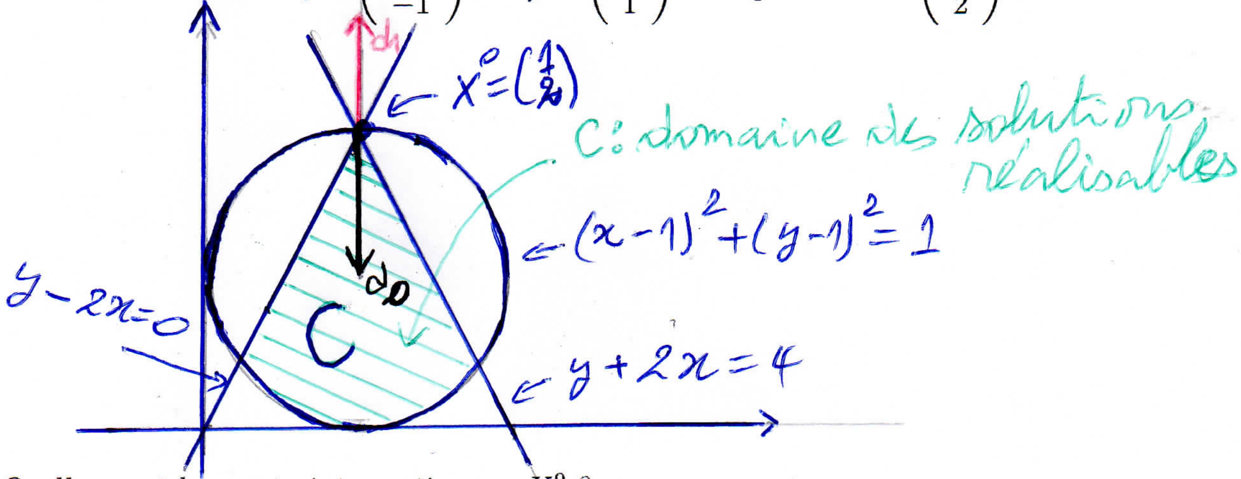
- pour $i \in \{1, 2, \ldots, p\},\ d^t \nabla h_i(x^0) = 0,$
- pour $j \in \{1, 2, \ldots, q\},\ g_j(x^0) = 0$ alors $d^t \nabla g_j(x^0) \leq 0.$

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 *?*

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

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

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

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

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 X^0 *est une direction admissible pour* f *en* X^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). $\square$

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. $\square$

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 :

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

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

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

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). □

1.3.2 Contraintes en égalité

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

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

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 $\mathbb{R}^n$: $\nabla h_1(x^*), \nabla h_2(x^*), \dots, \nabla h_p(x^*)$ sont linéairement indépendants (et donc $p \leq n$).*

Alors il existe p réels $\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*$ 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. □

Définition 1.7. **(Multiplicateurs de Lagrange)**

Les réels $\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*$ 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 :

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

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

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 $\frac{k}{2}\left[h_i(x_k)\right]^2 + \frac{k}{2}\sum_{j=1}^{q}\left[g_j^+(x_k)\right]^2 \geq 0$ (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 $(\lambda_1^k,\ \dots,\ \lambda_p^k, \mu_0^k, \mu_1^k, \dots, \mu_q^k) \in \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^q$ est de norme 1, on peut en extraire une sous-suite convergente vers $(\lambda_1^*,\ \dots,\ \lambda_p^*, \mu_0^*, \mu_1^*, \dots, \mu_q^*) \neq 0$ 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

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

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 $(\lambda_1^*,\ \lambda_2^*, \ldots,\ \lambda_p^*,\ \mu_0^*,\ \mu_1^*,\ \mu_2^*, \ldots,\ \mu_q^*)\ \neq 0$. On suppose que $\mu_0^* = 0$ et supposons aussi que $\mu_j^* = 0,\ \forall j \in I(x^*)$.

Alors le vecteur $\lambda^* = (\lambda_1^*,\ \lambda_2^*, \ldots,\ \lambda_p^*) \neq 0$, 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 $\mu_{j0}^* \neq 0$.

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). □

Exemple 1.2. *Considérons le problème suivant :*

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

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

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 prmiers 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 $\lambda^* = (\lambda_1^*,\ \lambda_2^*, \dots,\ \lambda_p^*) \in \mathbb{R}^p$ *et* $\mu^* = (\mu_1^*,\ \mu_2^*, \dots,\ \mu_q^*) \in \mathbb{R}^q$ *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 :*

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

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

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

avec

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

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. $\square$

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^*), $\nabla^2_{xx}\mathcal{L}(x^*, \lambda^*, \mu^*)$, est définie positive sur le sous-espace

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

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. $\square$

Chapitre 2

Applications

Important :

Avant de commencer le chapitre 2, je vous invite à revoir l'exemple 1.1 du chapitre 1.

– J'ai ajouté une figure à l'exemple.

– Il faut corriger la deuxième et la troisième question : d_0 est-elle une direction admissible pour f en X^0 ? au lieu de : d_0 est-elle une direction de descente pour f en X^0 ?

Ramadhan moubarek et bonne santé à tous.

Introduction

Dans ce chapitre nous donnons deux problèmes dont la résolution nécessite l'utilisation de problèmes d'optimisation avec contraintes.
Le premier problème est un problème 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 un problème 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, $\|\cdot\|$, utilisé 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

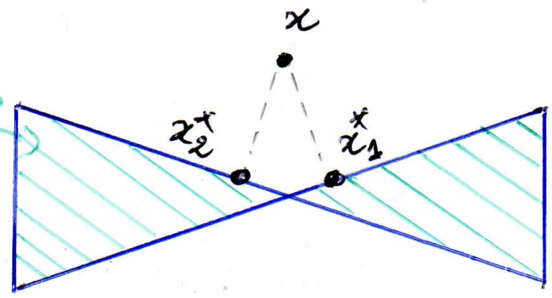
$$\forall (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n : \ \|\pi_C(x_2) - \pi_C(x_1)\| \leq \|x_2 - x_1\|.$$

□

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^* : 1er projeté de x sur C ;
x_2^* : 2ème 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*

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

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 :

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

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* $1 - x \leq 0,$
- $\mu(1-x) = 0,$
- $\nabla f(X) + \lambda\ \nabla h(X) + \mu\ \nabla g(X) = 0.$

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

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*.*

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

Travail à faire : Résoudre les exercices 9 et 10 de la série de TD.

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, $\sum\limits_{i=1}^{n}(y_i - ax_i - b)^2$ 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

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**.