Optimisation

MATH 3163

Ibrahima Dione

Hiver 2024

Optimisation avec contraintes

Sommaire

1 Optimisation sur des ensembles simples

Nous commençons l'étude des problèmes d'optimisation avec contraintes par le "plus simple" d'entre eux, à savoir le problème (P) suivant

$$(P) \quad \min_{x \in S} f(x)$$

où S est un sous-ensemble de $\mathbb{R}^n$ dont la "structure est simple". Pour être plus précis, nous considérerons ces quelques exemples de ce sous-ensemble :

$$S = \{x \in \mathbb{R}^n;\ Ax = b\},$$
$$S = \{x \in \mathbb{R}^n;\ \|x\| \leqslant r\},$$
$$S = \{x \in \mathbb{R}^n;\ a_i \leqslant x_i \leqslant b_i,\ i = 1, \ldots, n\}.$$

1.1 Etude théorique

Théorème 1.1

Condition nécessaire du 1^er ordre

Soit x^* un mininum local de f. On suppose que S est convexe et que f est de classe

C^1 sur S. Alors, on a

$$(\nabla f(x^*), x - x^*) \geqslant 0, \forall\, x \in S.$$

Preuve :

- Supposons qu'il existe un point $x^0 \in S$, tel que

$$\left(\nabla f(x^*), x^0 - x^*\right) < 0.$$

  ✓ Alors, prenons $x = x^* + \alpha\left(x^0 - x^*\right) \in S$, pour $0 \leqslant \alpha \leqslant 1$, on a

$$f(x) = f(x^*) + \alpha \underbrace{\left(\nabla f(x^*), x^0 - x^*\right)}_{<0} + o(\alpha) < f(x^*)$$

  ✓ Et pour $\alpha > 0$ suffisamment petit, alors on obtient

$$f(x) < f(x^*)$$

  ✓ Ce qui contredit le fait que x^* est un minimum local. □

Contrairement à un problème d'optimisation sans contraintes, une condition suffisante d'optimalité pour le problème (P) peut être formulée à partir du gradient de la fonction non convexe f.

Théorème 1.2

Soit f une fonction de classe C^1 sur S et soit x^* un point de S. On suppose que S est convexe et f satisfait la condition :

$$(\nabla f(x^*), x - x^*) \geqslant \alpha \|x - x^*\|, \ \alpha > 0, \text{ pour tout } x \in S \text{ et voisin de } x^*.$$

Alors x^* est un minimum local de $f(x)$ sur S.

Preuve :

- Soit le voisinage $V_\varepsilon(x^*) \cap S$ de x^*, pour lequel

$$(\nabla f(x^*), x - x^*) \geqslant \alpha \|x - x^*\|, \ \forall\, x \in V_\varepsilon(x^*) \cap S.$$

  ✓ Soit $\varepsilon_1 > 0$, avec $\varepsilon_1 \leqslant \varepsilon$, alors nous avons

$$|f(x^* + y) - f(x^*) - (\nabla f(x^*), y)| \leqslant (\alpha/2)\|y\|$$

  pour tout y tel que $\|y\| \leqslant \varepsilon_1$.

✓ En effet, par un développement de Taylor d'ordre 1, on a

$$f(x^* + y) = f(x^*) + (\nabla f(x^*), y) + o(\|y\|)$$

où $o(\|y\|) = \|y\|\varepsilon(y)$, avec $\varepsilon(y) \underset{y \to 0}{\longrightarrow} 0$. Cette limite signifie

$$\forall\, \eta_1 > 0,\ \exists\, \eta_2 > 0 : \forall\, y,\ \|y\| \leqslant \eta_2 \Rightarrow |\varepsilon(y)| \leqslant \eta_1$$

Prenons dans ce cas la valeur $\eta_1 = \alpha/2$, il existe alors un $\eta_2 > 0$ tel que

$$\forall\, y,\ \|y\| \leqslant \eta_2 \Rightarrow |\varepsilon(y)| \leqslant \alpha/2$$

En particulier, il suffit de prendre $\varepsilon_1 = \inf(\eta_2, \varepsilon)$, et alors on obtient

$$|f(x^* + y) - f(x^*) - (\nabla f(x^*), y)| = \|y\||\varepsilon(y)| \leqslant \alpha/2\|y\|,$$

pour tout y tel que $\|y\| \leqslant \varepsilon_1$.

- Par suite, pour $x \in S$, tel que $\|x - x^*\| \leqslant \varepsilon_1$ (où $y = x - x^*$), nous avons

$$\begin{aligned} f(x) - f(x^*) - (\nabla f(x^*), x - x^*) \geqslant -(\alpha/2)\|x - x^*\| \\ f(x) &\geqslant f(x^*) + (\nabla f(x^*), x - x^*) - (\alpha/2)\|x - x^*\| \\ f(x) &\geqslant f(x^*) + (\alpha/2)\|x - x^*\| \\ f(x) &\geqslant f(x^*), \end{aligned}$$

pour tout $x \in S$ et $\|x - x^*\| \leq \varepsilon$, i.e., x^* est un minimum local. □

Note : La condition $(\nabla f(x^*), x - x^*) \geqslant \alpha\|x - x^*\|, \alpha > 0$ dans le Théorème 1.2 ne peut être vérifiée si x^* est un point de l'intérieur de S. Alors, sous les conditions du Théorème 1.2 le minimum est nécessairement atteint sur le bord de S.

On peut aussi montrer le résultat suivant :

Théorème 1.3

Soit f une fonction convexe de classe C^1 sur S avec S convexe. Soit x^* un point de S. Alors la condition suivante

$$(\nabla f(x^*), x - x^*) \geqslant 0,\ \forall\, x \in S,$$

est une condition nécessaire et suffisante pour que x^* soit un minimum global de $f(x)$ sur S.

Preuve :

- Condition nécessaire :

✓ C'est une conséquence immédiate du Théorème 1.1,

✓ et du fait que tout optimum local est aussi global pour les problèmes convexes.

- Condition suffisante :

✓ D'après le Théorème 4.1 du chapitre I, on a :

$$f(x) \geqslant f(x^*) + (\nabla f(x^*), \ x - x^*)$$

car f est convexe et de classe C^1.

✓ Et donc $f(x) \geqslant f(x^*), \ \forall x \in S$. □

Exemple 1.1 : Soit $a \in \mathbb{R}^n$ et S un convexe de $\mathbb{R}^n$. Soit la fonction $f(x) = \|x - a\|^2$ et on cherche à résoudre le problème

$$\min_{x \in S} f(x).$$

- Tout d'abord on a $\nabla f(x) = 2(x - a)$ et donc la matrice Hessienne $H(x) = 2I$. Donc $f(x)$ est strictement convexe.
- Alors, d'après le Théorème 1.3, une condition nécessaire et suffisante pour que x^* soit un minimum global de $f(x)$ sur S est

$$(\nabla f(x^*), x - x^*) \geqslant 0, \ \forall x \in S, \ \text{c'est-à-dire } (x^* - a, x - x^*) \geqslant 0, \ \forall x \in S.$$

- Ce qui veut dire, en d'autres termes, que x^* est la projection du vecteur $a \in \mathbb{R}^n$ sur le convexe S (notée $x^* = P_S(a)$, où P_S est l'opérateur de projection). De plus cette solution est unique car $f(x)$ est strictement convexe.

Note :

- Soit S un convexe de $\mathbb{R}^n$. $P_S(x)$ est le point de S qui réalise le minimum de $\|x - y\|$ sur l'ensemble des $y \in S$. On notera par :

$$P_S(x) = \underset{y \in S}{\operatorname{argmin}} \|x - y\|.$$

- La caractérisation de la projection sur un ensemble convexe S est donnée par :

$$(x - P_S(x), y - P_S(x)) \leqslant 0, \ \forall y \in S.$$

- Il est évident que si $x \in S$, alors $P_S(x) = x$.

Exemple 1.2 : Soit l'ensemble convexe $S = \{x \in \mathbb{R}^n;\ \|x\| \leq \rho\}$. Alors, on a

$$P_S(x) = \begin{cases} x, & \text{si } \|x\| \leqslant \rho, \\ \rho \frac{x}{\|x\|}, & \text{si } \|x\| > \rho. \end{cases} \tag{1}$$

Concernant l'unicité de la solution du problème (P), on peut énoncer les résultats suivants :

Théorème 1.4

Sous les hypothèses du Théorème 1.2, x^* est l'unique minimum local.

Preuve :

- D'après la démonstration du Théorème 1.2, on a vu que

$$f(x) \geqslant f(x^*) + \alpha/2 \|x - x^*\|,$$

pour tout x dans un voisinage de x^*.

- Supposons qu'il existe un autre minimum local ξ^*. Alors on a aussi

$$f(x) \geqslant f(\xi^*) + \alpha/2 \|x - \xi^*\|. \tag{2}$$

- D'où, on obtient à partir de (2)

$$f(x^*) \geqslant f(\xi^*) + \alpha/2 \|x^* - \zeta^*\|.$$

- Du fait que ξ^* est un autre minimum local, alors on a

$$f(x^*) = f(\zeta^*) \ \text{ et donc } \ \|x^* - \zeta^*\| \leq 0 \ \Rightarrow \ x^* = \zeta^*. \qquad \square$$

L'unicité de la solution de (P) peut être garantie par la stricte convexité de f. Cependant, d'autres conditions peuvent être imposées sur S pour assurer l'unicité du minimum. Commençons par donner une définition.

Définition 1.1

On dit que S est stictement convexe si pour tout $x_1, x_2 \in S$, $x_1 \neq x_2$ et tout $\lambda \in]0,\ 1[$, le point $\lambda x_1 + (1-\lambda)x_2$ est un point de l'intérieur de S.

Exemple 1.3 :

- La boule $B_\rho = \{x \in \mathbb{R}^n; \|x\| \leqslant \rho\}$ est strictement convexe.
- Par contre le parallélipéde $S = \{x \in \mathbb{R}^n, a \leq x \leq b\}$ n'est pas strictement convexe.

Nous avons aussi une propriété qui relie la stricte convexité de f et la stricte convexité de $S = \{x \in \mathbb{R}^n; f(x) \leqslant \alpha\}$.

Proposition 1.1

Si f est une fonction strictement convexe sur $\mathbb{R}^n$, alors l'ensemble $S_\alpha = \{x \in \mathbb{R}^n; f(x) \leq \alpha\}$ est strictement convexe pour tout α.

Théorème 1.5

Soit f une fonction convexe de classe C^1 sur $\mathbb{R}^n$. Soit S un sous-ensemble de $\mathbb{R}^n$ strictement convexe. Si $\|\nabla f(x)\| \geqslant \varepsilon > 0$ pour tout $x \in S$, alors le minimum de $f(x)$ sur S est unique.

Exemple 1.4 :

1 $x^* = \underset{a \leq x \leq b}{\operatorname{argmin}}(c, x)$, où (c, x) est convexe et $[a, b]$ est convexe donc C.N.

$$(\nabla f(x^*), x - x^*) \geqslant 0, \ \forall\, x \in [a, b]$$

- ✓ Or $\nabla f(x^*) = c$, alors

$$\begin{aligned} (c, x - x^*) &\geqslant 0, \ \forall\, x \in [a, b], \\ c_i (x_i - x_i^*) &\geqslant 0, \ \forall\, i \in \{1, \ldots, n\}, \forall\, x_i \in [a_i, b_i] \end{aligned}$$

- ✓ si $c_i > 0$ alors $x_i^* = a_i$
- ✓ si $c_i < 0$ alors $x_i^* = b_i$
- ✓ si $c_i = 0$ abord $x_i^* = \lambda \in [a_i, b_i]$.
- ✓ Alors une solution est :

$$x_i^* = \begin{cases} a_i & \text{si } c_i > 0 \\ b_i & \text{si } c_i < 0 \\ \lambda, & a_i \leqslant \lambda \leqslant b_i, \ \text{si } c_i = 0. \end{cases}$$

2 $x^* = \underset{\|x\| \leq \rho}{\operatorname{argmin}}(c, x)$ avec $c \neq 0$. La solution ici est donnée par $x^* = -\rho \frac{c}{\|c\|}$.

- ✓ En effet

$$|(c, x)| \leqslant \|c\| \|x\| \leqslant \rho \|c\|, \forall\, x \ \text{ tel que } \ \|x\| \leqslant \rho$$

donc $(c, x) \geqslant -\rho \|c\|$.

- ✓ Si on trouve un point x^* tel que $(c, x^*) = -\rho\|c\|$ et de norme plus petit que ρ, alors il sera une solution.
- ✓ Or $(c, x^*) = -\rho\|c\| \ \Rightarrow \ (c, x^*) = \left(-\rho \frac{c}{\|c\|}, c\right)$ et donc il suffit de prendre

$$x^* = -\rho \frac{c}{\|c\|}.$$

✓ De plus c'est l'unique solution de ce problème car $B_\rho = \{x \in \mathbb{R}^n; \|x\| \leq \rho\}$ est strictement convexe et la fonction $x \longrightarrow (c, x)$ est convexe aussi avec

$$\|\nabla f(x)\| = \|c\| > 0, \text{ (pour } c \neq 0).$$

1.2 Méthodes numériques de base

Les méthodes que nous allons décrire sont des méthodes similaires à celles déjà étudiées dans le cadre d'optimisation sans contraintes.

• 1.2.1 - Méthode du gradient projeté

Cette méthode est une généralisation directe de la méthode de gradient. Puisque, en général, les itérés sont en dehors de l'ensemble S, il est possible d'ajouter l'opération de projection sur S. Nous aboutissons à la méthode :

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

où P_S est l'opérateur de projection sur S. Donnons quelques exemples de projections !

Exemple 1.5 :

1. Soit $S = \{x \in \mathbb{R}^n; x \geqslant 0\}$. Pour tout $x \in \mathbb{R}^n$, on définit la projection sur S par

$$P_S(x) = x_+, \quad \text{où} \quad x_+ \overset{\text{def.}}{=} (\sup(x_i, 0))_{1 \leqslant i \leqslant n}.$$

Dans ce cas, la méthode du gradient projeté prend la forme

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

2. Soit $S = \{x \in \mathbb{R}^n; a \leq x \leq b\}$. Considérons la notation suivante :

$$\text{si } \lambda, \alpha, \beta \in \mathbb{R}, \text{ on pose : } (\lambda)_\alpha^\beta = \begin{cases} \lambda & \text{si} \quad \alpha \leq \lambda \leq \beta \\ \beta & \text{si} \quad \lambda > \beta \\ \alpha & \text{si} \quad \lambda < \alpha \end{cases}$$

Alors pour tout $x \in \mathbb{R}^n$, on définit la projection sur S par

$$P_S(x) = (x)_a^b \overset{\text{def.}}{=} \left((x_i)_{a_i}^{b_i}\right)_{1 \leq i \leq n},$$

et la méthode du gradient projeté s'écrit

$$x^{k+1} = \left(x^k - \gamma \nabla f\left(x^k\right)\right)_a^b$$

3. Si $S = \{x \in \mathbb{R}^n; \|x\| \leq \rho\}$, alors on avait défini en (1) la projection $P_S(x)$. La méthode du gradient projeté s'écrit donc

$$x^{k+1} = \begin{cases} x^k - \gamma \nabla f\left(x^k\right), & \text{si} \quad \left\|x^k - \nabla f\left(x^k\right)\right\| \leq \rho, \\ \rho \frac{x^k - \gamma \nabla f\left(x^k\right)}{\left\|x^k - \gamma \nabla f\left(x^k\right)\right\|}, & \text{si} \quad \left\|x^k - \nabla f\left(x^k\right)\right\| > \rho. \end{cases}$$

• 1.2.2 - La méthode de Newton

Pour construire la méthode de Newton pour le problème (P), on peut utiliser la même idée d'approximation quadratique de $f(x)$ comme dans le cas d'optimisation sans contraintes. La seule différence est qu'il est nécessaire de trouver l'approximation du minimum sur S au lieu de tout l'espace $\mathbb{R}^n$. Ces arguments mènent à la méthode

$$x^{k+1} = \underset{x \in S}{\operatorname{argmin}} f_k(x)$$

où $f_k(x) = f\left(x^k\right) + \left(\nabla f\left(x^k\right), x - x^k\right) + (1/2)\left(H\left(x^k\right)\left(x - x^k\right), x - x^k\right).$

• 1.2.3 - La méthode du gradient conjugué : cas quadratique

On considère une fonction quadratique $f(x)$. Soit $S = \{x \in \mathbb{R}^n; Cx = 0\}$ où C est une matrice (m, n) de rang m $(m \leq n)$. Si $m = n$, alors $S = \{0\}$. On a la projection

$$P_S(x) = \left(I - C^\top \left(CC^\top\right)^{-1} C\right) x.$$

Dans ces conditions, la méthode du gradient conjugué s'écrit :

Algorithme : L'algorithme gradient conjugué s'écrit :

1. Étant donné $x^0 \in S$ et $d^0 = -P_S\left(\nabla f\left(x^0\right)\right).$
2. Déterminer à l'étape k :
   - $\alpha_k = \operatorname{argmin} f\left(x^k + \alpha d^k\right)$
   - $x^{k+1} = x^k + \alpha_k d^k$
   - $d^{k+1} = -P_S\left(\nabla f\left(x^{k+1}\right)\right) + \beta_{k+1} d^k$
   - $\beta_{k+1} = \frac{\left\|P_S(\nabla f(x^{k+1}))\right\|^2}{\left\|P_S(\nabla f(x^k))\right\|^2}$

Note :

- On montre que si $f(x)$ est quadratique, $f(x) = \frac{1}{2}(Ax, x) + (b, x) + c$, et si $(Ax, x) \geqslant \alpha\|x\|^2$, $\alpha > 0$, pour tout $x \in S$, alors la méthode du gradient conjugué (qu'on vient de décrire) converge au plus en $(n - m)$ itérations.
- Remarquons aussi que si $m = 0$ (ce qui entraine que $C = 0$), alors $S = \mathbb{R}^n$ et on obtient le même résultat de convergence que dans le cas d'optimisation sans contraintes puisque A est en particulier définie positive.

2 Contraintes d'égalité

Nous nous intéressons ici au problème

$$(P)\ \begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x) \\ g_i(x)=0,\ i=1,\ldots,m \end{cases} \tag{3}$$

où f et g_i sont des fonctions assez régulières. C'est un cas particulier du problème général de la programmation mathématique. Nous étudierons ce problème en détail puisque les ideés qui se dégageront de cette étude, seront utilisables dans le cas général.

2.1 Conditions nécessaires du 1^{er} et du $2^{\text{ième}}$ ordre

Soit $S=\{x\in\mathbb{R}^n;\ g_i(x)=0, i=1,\ldots,m\}$. Les points $x\in S$ sont dits admissibles. Le point x^* est appelé minimum (local) pour le problème (P), s'il est admissible et si

$$f(x^*)\leqslant f(x),$$

pour tout x admissible voisin de x^*.

Théorème 2.1

Condition nécessaire du 1^{er} ordre

Soit x^* un minimum du problème (P). Supposons que $f(x), g_i(x)$ sont de classe C^1 dans un voisinage de x^*. Alors, il existe $\lambda_0,\lambda_1,\ldots,\lambda_m$ des nombres réels non tous nuls, tels que

$$\lambda_0\nabla f(x^*)+\sum_{i=1}^{m}\lambda_i\nabla g_i(x^*)=0 \tag{4}$$

Note : Soit $g=\begin{pmatrix} g_1\\ \vdots \\ g_m\end{pmatrix}$ où $g_i:\mathbb{R}^n\longrightarrow\mathbb{R}, i=1,\ldots,m$. Le jacobien de g est la matrice notée ∇g de type (m,n) définie par

$$\nabla g(x)=\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_m} \\ \vdots & & & \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_m}\end{bmatrix}=\begin{bmatrix}\nabla g_1^\top \\ \vdots \\ \nabla g_m^\top\end{bmatrix}$$

Alors, la condition (4) du Théorème 2.1 s'écrit aussi comme suit

$$\lambda_0\nabla f(x^*)+\nabla g(x^*)^\top\lambda=0,\ \text{ où le vecteur }\ \lambda=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_m\end{pmatrix} \tag{5}$$

Définition 2.1

Un point x^* est dit régulier si $f(x), g_i(x)$ sont de classe C^1 dans un voisinage de x^* et $\nabla g_i\left(x^*\right), i=1,\ldots,m$, sont linéairement indépendants.

Théorème 2.2

La règle des multiplicateurs de Lagrange

Si x^* est un minimum régulier du problème (P), alors il existe $\lambda_1,\ldots,\lambda_m$ tels que

$$\nabla f\left(x^*\right)+\nabla g\left(x^*\right)^\top \lambda=0, \text{ où } \lambda=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_m\end{pmatrix}. \tag{6}$$

Les coefficients $\lambda_1,\ldots,\lambda_m$ dans (6) sont appelés multiplicateurs de Lagrange.

- Le Théorème 2.2 est une conséquence du Théorème 2.1.

Preuve :

En effet, supposons que $\lambda_0=0$ alors à partir de (4) on a

$$\sum_{i=1}^{m}\lambda_i\nabla g_i\left(x^*\right)=0,$$

avec au moins un des coefficients λ_i non nul, ceci contredit le fait que $\nabla g_i\left(x^*\right)$ sont linéairement indépendants. Alors, $\lambda_0\neq 0$ et donc on peut diviser (5) par λ_0 et ceci donne (6). $\square$

- Dans le Théorème 2.2 il est nécessaire que le point x^* soit régulier pour avoir la formule (6).

Exemple 2.1 : En effet, considérons le problème (P) dans $\mathbb{R}^2$ avec les fonctionnelles

$$f(x,y)=y,\ g_1(x,y)=(x-1)^2+y^2-1 \text{ et } g_2(x,y)=(x+1)^2+y^2-1.$$

- Dans ce cas, l'ensemble admissible est donné par

$$S=\left\{(x,y)\in\mathbb{R}^2;\ g_i(x,y)=0, i=1,2\right\}=\{(0,0)\}$$

et donc la seule solution de ce problème est $(x^*,y^*)=(0,0)$.

- Cependant, le point (x^*, y^*) n'est pas un point régulier car

$$\nabla g_1(x,y) = \begin{pmatrix} 2(x-1) \\ 2y \end{pmatrix}, \nabla g_2(x,y) = \begin{pmatrix} 2(x+1) \\ 2y \end{pmatrix},$$

  et les deux vecteurs $\nabla g_1(x^*, y^*) = \begin{pmatrix} -2 \\ 0 \end{pmatrix}, \nabla g_2(x^*, y^*) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ ne sont pas linéairement indépendants.
- Et s'il existe λ_1, λ_2 tels que

$$\nabla f(x^*, y^*) + \lambda_1 \nabla g_1(x^*, y^*) + \lambda_2 \nabla g_2(x^*, y^*) = 0,$$

  cela implique en particulier que $1 + \lambda_1 \times 0 + \lambda_2 \times 0 = 0 \Rightarrow 1 = 0$ ce qui est absurde.

Théorème 2.3

Condition nécessaire du $2^{\text{ième}}$ ordre

Soit x^* un minimum régulier du problème (P). Soit $\lambda_1, \ldots, \lambda_m$ les multiplicateurs de Lagrange. Supposons que $f(x)$ et $g_i(x), i = 1, \ldots, m$, sont de classe C^2 au voisinage de x^*. Alors la matrice

$$L(x^*) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x^*) + \lambda_1 \frac{\partial^2 g_1}{\partial x_i \partial x_j}(x^*) + \cdots + \lambda_m \frac{\partial^2 g_m}{\partial x_i \partial x_j}(x^*) \right]_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant n}}$$

est semi-définie positive sur le plan tangent $M(x^*) = \{y \in \mathbb{R}^n,\ \nabla g(x^*)\, y = 0\}$, c'est-à-dire $(L(x^*)\, y, y) \geqslant 0,\ \forall\, y \in M(x^*)$.

Dans la définition de $M(x^*)$, $\nabla g(x^*)\, y = 0$ est équivalent à dire

$$(\nabla g_i(x^*),\ y) = 0,\ \forall\, i = 1, \ldots, m.$$

Note : On notera $L(x^*)$ par

$$L(x^*) = F(x^*) + \lambda_1 G_1(x^*) + \cdots + \lambda_m G_m(x^*)$$

où $F(x^*)$ est le Hessian de $f(x)$ au point x^* et $G_i(x^*)$ est le Hessian de $g_i(x)$ au point x^* pour tout $i \in \{1, \ldots, m\}$.

2.2 Condition suffisante du $2^{\text{ième}}$ ordre

Théorème 2.4

Soil x^* un point régulier tel que $g_i(x^*) = 0$, $i = 1, \ldots, m$, et soit $\lambda \in \mathbb{R}^m$ tel que

$$\nabla f(x^*) + \nabla g(x^*)^\top \lambda = 0.$$

Si $L(x^*) = F(x^*) + \lambda_1 G_1(x^*) + \cdots + \lambda_m G_m(x^*)$ est une matrice définie positive sur le plan tangent $M(x^*)$, alors x^* est un minimum de f sous les contraintes $g_i(x^*) = 0, i = 1, \ldots, m$.

Définition 2.2

Un point x^* pour lequel les conditions du Théorème 2.4 sont vérifiées est dit point non singulier.

Note :

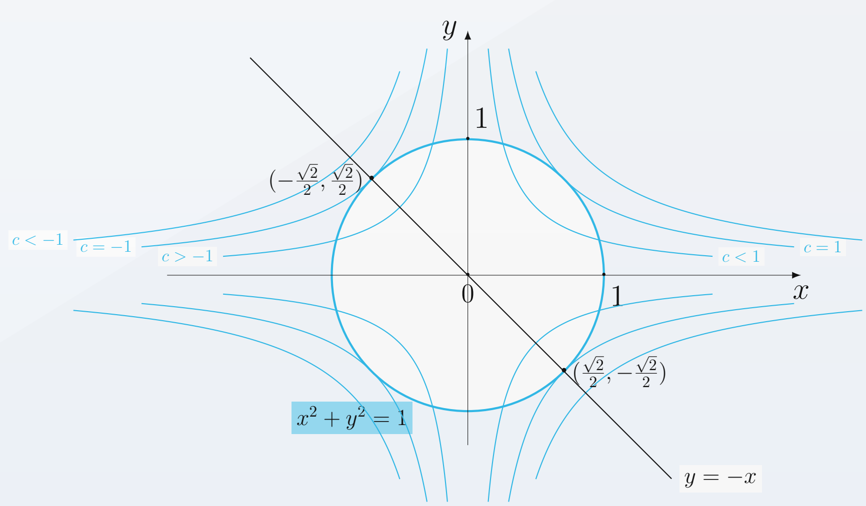
- Un minimum non singulier est unique.
- Pour un minimum régulier, les multiplicateurs de Lagrange sont déterminés de manière unique.

Exemple 2.2 :

- Considérons le problème de minimisation suivant

$$\min_{g(x,y)=0} f(x,y), \text{ où } f(x,y) = 2xy \text{ et } g(x,y) = x^2 + y^2 - 1.$$

  - Résolvons le problème géométriquement.
    - ✓ Les courbes de niveau $f(x,y) = c$ sont des hyperboles d'équations $y = \frac{c}{2x}$.



Les solutions sont les points sur les courbes de niveau ayant la plus petite valeur de c et qui sont sur le cercle d'équation $x^2 + y^2 = 1$.

- ✓ Gaphiquement, on voit que les solutions sont les points qui sont en même temps sur la courbe de niveau $2xy = -1$ $(c = -1)$ et sur le cercle.
- ✓ Alors, on obtient analytiquement le système d'équations suivant

$$\begin{cases} x^2 + y^2 = 1 \\ 2xy = -1 \end{cases} \Rightarrow (x+y)^2 = 0 \Rightarrow y = -x$$

- ✓ D'où $2xy = -2x^2 = -1 \Rightarrow x^2 = 1/2 \Rightarrow x = \pm\sqrt{2}/2$. Donc $\min\limits_{g(x,y)=0} f(x,y) = -1$ et il y'a deux points qui réalisent le minimum :

$$\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \text{ et } \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

- Maintenant, vérifions les conditions d'optimalité.
  - ✓ Les gradients de la fonction à minimiser f et de la fonction contrainte g :

$$\nabla f(x,y) = \begin{pmatrix} 2y \\ 2x \end{pmatrix}, \ \nabla g(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

  - ✓ Condition nécessaire du 1^{ier} ordre : $\nabla f(x^*, y^*) + \lambda \nabla g(x^*, y^*) = 0$

$$\begin{pmatrix} 2y^* \\ 2x^* \end{pmatrix} + \lambda \begin{pmatrix} 2x^* \\ 2y^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ i.e. } \begin{cases} y^* + \lambda x^* = 0 \\ x^* + \lambda y^* = 0 \end{cases}$$

Ce qui donne les points suivants, peu importe que λ soit $\lambda = 1$ ou $\lambda = -1$:

$$\left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \text{ et } \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

  - ✓ Condition nécessaire du $2^{\text{ième}}$ ordre : $(L(x^*, y^*)\xi, \xi) \geqslant 0, \ \forall \xi \in M(x^*, y^*)$

$$F(x,y) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}, \ G(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

$$L(x^*, y^*) = F(x^*, y^*) + \lambda G(x^*, y^*) = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \text{ où on a pris } \lambda = 1.$$

$$M(x^*, y^*) = \left\{ (\xi, \eta) \in \mathbb{R}^2; \left(\nabla g(x^*, y^*), \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right) = 0 \right\}.$$

Prenons par exemple, $(x^*, y^*) = \left(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$. Alors

$$\nabla g\,(x^*, y^*) = \begin{pmatrix} -\sqrt{2} \\ \sqrt{2} \end{pmatrix} \Rightarrow \left(\nabla g\,(x^*, y^*), \begin{pmatrix} \xi \\ \eta \end{pmatrix}\right) = -\sqrt{2}(\xi - \eta),$$

et dans ce cas $M\,(x^*, y^*) = \{(\xi, \xi), \xi \in \mathbb{R}\}$.
Nous obtenons le même résultat pour le point $(x^*, y^*) = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$.

✓ Par suite, $L\,(x^*, y^*)$ est définie positive sur $M\,(x^*, y^*)$ car on obtient

$$(\xi, \xi) \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \xi \\ \xi \end{pmatrix} = (\xi, \xi) \cdot \begin{pmatrix} 4\xi \\ 4\xi \end{pmatrix} = 8\xi^2 > 0$$

pour $\xi \neq 0$.

- Considérons le problème de minimisation suivant

$$\min_{g(x,y)=0} f(x, y), \text{ où } f(x, y) = x^2 + y^2 \text{ et } g(x, y) = x + y - 1.$$

✓ On a les gradients suivants :

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \ \nabla g(x, y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

✓ Condition nécessaire du 1ier ordre : $\nabla f\,(x^*, y^*) + \lambda \nabla g\,(x^*, y^*) = 0$

$$\begin{pmatrix} 2x^* \\ 2y^* \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ i.e. } \begin{cases} 2x^* + \lambda = 0 \\ 2y^* + \lambda = 0 \end{cases} \Rightarrow x^* = y^*$$

Et comme $g\,(x^*, y^*) = 0$ alors $(x^*, y^*) = (1/2,\ 1/2)$ et $\lambda = -2x^* = -1$.

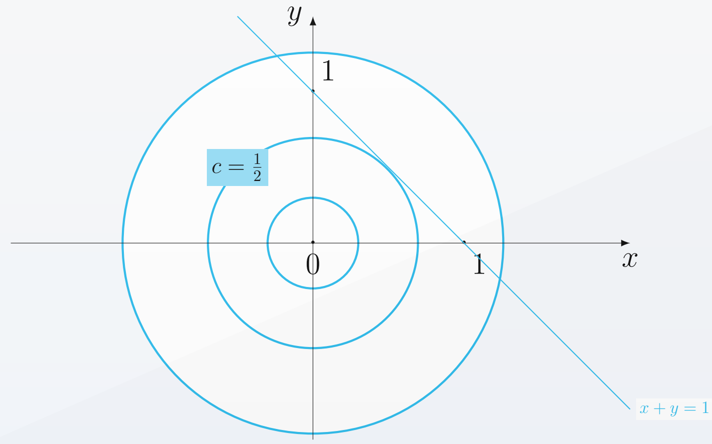
✓ Condition nécessaire du 2ième ordre : $(L\,(x^*, y^*)\,\xi, \xi) \geqslant 0,\ \forall\, \xi \in M\,(x^*, y^*)$

$$F\,(x^*, y^*) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ et } G\,(x^*, y^*) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \text{ alors } M\,(x^*, y^*) = \mathbb{R}^2.$$

Et donc, $L\,(x^*, y^*) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ est définie positive.

✓ Donc, le problème admet une solution unique qui est $(1/2, 1/2)$ et le multiplicateur de Lagrange est $\lambda = -1$.

✓ Encore une fois, on peut résoudre ce problème géométriquement. Les courbes de niveau $f(x, y) = c$ sont les cercles d'équations $x^2 + y^2 = c$ (bien sûr $c \geqslant 0$).



Le minimum est atteint au point où la courbe de niveau $x^2+y^2 = c$ est tangente à la droite $x+y = 1$. Ce point est $(1/2, 1/2)$ et la valeur minimale est $c = \frac{1}{4}+\frac{1}{4} = 1/2$.

2.3 Formule du Min-Max

Soit $\mathscr{L}(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$, une fonction.

Définition 2.3

Point-selle

On dit qu'un point $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ est un point-selle de $\mathscr{L}(\cdot,\cdot)$ si le point x^* minimise la fonction

$$\mathscr{L}\,(\cdot, \lambda^*) : x \in \mathbb{R}^n \longmapsto \mathscr{L}\,(x, \lambda^*)$$

et si le point λ^* maximise la fonction

$$\mathscr{L}\,(x^*, \cdot) : \lambda \in \mathbb{R}^m \longmapsto \mathscr{L}^*\,(x^*, \lambda)$$

c'est-à-dire, si on a

$$\max_{\lambda\in\mathbb{R}^m} \mathscr{L}\,(x^*, \lambda) = \mathscr{L}\,(x^*, \lambda^*) = \min_{x\in\mathbb{R}^n} \mathscr{L}\,(x, \lambda^*)\,.$$

Proposition 2.1

min-max

Si $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ est un point-selle de $\mathscr{L}(\cdot,\cdot)$, alors

$$\max_{\lambda\in\mathbb{R}^m} \min_{x\in\mathbb{R}^n} \mathscr{L}(x, \lambda) = \mathscr{L}\,(x^*, \lambda^*) = \min_{x\in\mathbb{R}^n} \max_{\lambda\in\mathbb{R}^m} \mathscr{L}(x, \lambda).$$

Preuve :

- Tout d'abord, par définition du point-selle (x^*, λ^*), nous avons les inéquations

$$\mathscr{L}(x^*, \lambda) \leqslant \mathscr{L}(x^*, \lambda^*) \leqslant \mathscr{L}(x, \lambda^*), \ \forall\, x \in \mathbb{R}^n, \ \forall\, \lambda \in \mathbb{R}^m.$$

- Posons $F(x) = \max\limits_{\lambda \in \mathbb{R}^m} \mathscr{L}(x, \lambda)$, $\forall\, x \in \mathbb{R}^n$. On veut montrer que $\min\limits_{x \in \mathbb{R}^n} F(x) = \mathscr{L}(x^*, \lambda^*)$:
  - ✓ On a
$$\mathscr{L}(x^*, \lambda^*) \leqslant \mathscr{L}(x, \lambda^*) \leq \max_{\lambda \in \mathbb{R}^m} \mathscr{L}(x, \lambda) = F(x), \ \forall\, x \in \mathbb{R}^n.$$
    Donc, on a $\mathscr{L}(x^*, \lambda^*) \leq \min\limits_{x \in \mathbb{R}^n} F(x)$.
  - ✓ D'autre part, par définition du point-selle, on a
$$F(x^*) = \max_{\lambda \in \mathbb{R}^m} \mathscr{L}(x^*, \lambda) = \mathscr{L}(x^*, \lambda^*)$$
  - ✓ Alors,
$$\min_{x \in \mathbb{R}^n} F(x) \leqslant F(x^*) = \mathscr{L}(x^*, \lambda^*)$$
    Donc, on a $\min\limits_{x \in \mathbb{R}^n} F(x) \leqslant \mathscr{L}(x^*, \lambda^*)$.
    On obtient ainsi le résultat $\min\limits_{x \in \mathbb{R}^n} \max\limits_{\lambda \in \mathbb{R}^m} \mathscr{L}(x, \lambda) = \mathscr{L}(x^*, \lambda^*)$.
- Maintenant, posons $G(\lambda) = \min\limits_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda)$. On veut montrer que $\max\limits_{\lambda \in \mathbb{R}^m} G(\lambda) = \mathscr{L}(x^*, \lambda^*)$:
  - ✓ On a
$$G(\lambda) = \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda) \leqslant \mathscr{L}(x^*, \lambda) \leqslant \mathscr{L}(x^*, \lambda^*), \ \forall\, \lambda \in \mathbb{R}^m.$$
    Donc, on obtient $\max\limits_{\lambda \in \mathbb{R}^m} G(\lambda) \leqslant \mathscr{L}(x^*, \lambda^*)$.
  - ✓ D'autre part, d'après la définition de point-selle, on a
$$\mathscr{L}(x^*, \lambda^*) = \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda^*) = G(\lambda^*)$$
  - ✓ Alors
$$\mathscr{L}(x^*, \lambda^*) = G(\lambda^*) \leq \max_{\lambda \in \mathbb{R}^m} G(\lambda)$$
    Donc, on obtient $\mathscr{L}(x^*, \lambda^*) \leqslant \max\limits_{\lambda \in \mathbb{R}^m} G(\lambda)$.
    Ainsi, on obtient donc le résultat $\max\limits_{\lambda \in \mathbb{R}^m} \min\limits_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda) = \mathscr{L}(x^*, \lambda^*)$. □

Maintenant, nous allons montrer le lien entre le point-selle et la solution d'un problème d'optimisation avec contraintes.

Définition 2.4

Considérons toujours le même problème de minimisation (P) suivant

$$(P)\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x) \\ g_i(x)=0,\ i=1,\dots,m. \end{cases}$$

La fonction de Lagrange associée au problème (P), est par définition la fonction

$$\mathscr{L}(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i g_i(x).$$

Théorème 2.5

(x^*,λ^*) est un point-selle pour $\mathscr{L}(x,\lambda)$ si et seulement si

1. $\mathscr{L}\left(x^*,\lambda^*\right)=\min\limits_{x\in\mathbb{R}^n}\mathscr{L}\left(x,\lambda^*\right)$.
2. $g_i\left(x^*\right)=0,\ \forall\, i\in\{1,\dots,m\}$.

Preuve :

- Condition nécessaire
  - ✓ D'après la définition du point-selle, on a automatiquement 1.
  - ✓ D'autre part

$$\mathscr{L}\left(x^*,\lambda^*\right)\geqslant\mathscr{L}\left(x^*,\lambda\right),\ \forall\,\lambda\in\mathbb{R}^m.$$

Donc,

$$\begin{aligned} f\left(x^*\right)+\sum_{i=1}^{m}\lambda_i^* g_i\left(x^*\right) &\geqslant f\left(x^*\right)+\sum_{i=1}^{m}\lambda_i g_i\left(x^*\right) \\ \sum_{i=1}^{m}\left(\lambda_i-\lambda_i^*\right)g_i\left(x^*\right) &\leq 0,\ \forall\ \left(\lambda_i\right)_{1\leqslant i\leqslant m}\in\mathbb{R}^m. \end{aligned} \tag{7}$$

  - ✓ Alors forcément $g_i\left(x^*\right)=0,\ \forall\, i\in\{1,\dots,m\}$.
  - ✓ En effet, supposons qu'il existe $k\in\{1,\dots,m\}$ tel que $g_k\left(x^*\right)\neq 0$. Alors, en choisissant $\left(\lambda_1^*,\dots,\lambda_{k-1}^*,\lambda_k,\lambda_{k+1}^*,\dots,\lambda_m^*\right)$, où λ_k est quelconqne dans $\mathbb{R}$, on obtient à partir de (7) ceci

$$\begin{aligned} \sum_{i=1}^{m}\left(\lambda_i-\lambda_i^*\right)g_i\left(x^*\right) &= \left(\lambda_k-\lambda_k^*\right)g_k\left(x^*\right) \\ &\leq 0,\ \forall\ \lambda_k\in\mathbb{R}. \end{aligned}$$

Or suivant le signe de $g_k\left(x^*\right)$, on peut trouver λ_k tel que $\left(\lambda_k-\lambda_k^*\right)g_k\left(x^*\right)>0$. Ce qui sera contradictoire.

Condition suffisante

✓ D'après 2, on a

$$\begin{aligned}
&\mathscr{L}\left(x^*, \lambda^*\right)=f\left(x^*\right) \\
\text{et } &\mathscr{L}\left(x^*, \lambda\right)=f\left(x^*\right)+\sum_{i=1}^{m} \lambda_i g_i\left(x^*\right)=f\left(x^*\right)=\mathscr{L}\left(x^*, \lambda^*\right) \\
\text{i.e.} \quad &\max_{\lambda \in \mathbb{R}^m} \mathscr{L}^*\left(x^*, \lambda\right)=\mathscr{L}\left(x^*, \lambda^*\right).
\end{aligned}$$

✓ Cette dernière relation et celle en 1 entraînent que le point (x^*, λ^*) est un point-selle pour $\mathscr{L}(x, \lambda)$. □

Corollaire 2.1

Si (x^*, λ^*) est un point-selle de $\mathscr{L}(x, \lambda)$ alors x^* est un optimum global de (P).

Preuve :

- (x^*, λ^*) point-selle de $\mathscr{L}(x, \lambda)$ entraine que les conditions 1 et 2 du **Théorème 2.5** sont réalisées.
- Alors, ces conditions nous donnent

$$\begin{aligned}
&f\left(x^*\right)+\underbrace{\sum_{i=1}^{m} \lambda_i^* g_i\left(x^*\right)}_{=0} \leq f(x)+\sum_{i=1}^{m} \lambda_i^* g_i(x), \ \forall\, x \in \mathbb{R}^n \\
\text{i.e.} \quad &f\left(x^*\right) \leqslant f(x)+\sum_{i=1}^{m} \lambda_i^* g_i(x), \ \forall\, x \in \mathbb{R}^n.
\end{aligned}$$

- En particulier

$$f\left(x^*\right) \leqslant f(x), \ \forall\, x \in \mathbb{R}^n \text{ tel que } g_i(x)=0, \ \forall\, i \in\{1, \ldots, m\}.$$

□

Note :

- Connaissant λ^*, le problème de minimisation avec contraintes se ramène donc au problème de minimisation sans contraintes

$$\left(P^*\right) \quad \mathscr{L}\left(x^*, \lambda^*\right)=\min_{x \in \mathbb{R}^n} \mathscr{L}\left(x, \lambda^*\right).$$

- Comment trouver un tel $\lambda^* \in \mathbb{R}^m$? Si l'on se rappelle que

$$\begin{aligned}
\mathscr{L}\left(x^*, \lambda^*\right)=\min_{x \in \mathbb{R}^n} \mathscr{L}\left(x, \lambda^*\right) &=\min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} \mathscr{L}(x, \lambda) \\
&=\max_{\lambda \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda)
\end{aligned}$$

alors on est naturellement conduit à chercher λ^* comme solution du problème de maximisation

$$(D)\ \begin{cases} G(\lambda^*) = \max\limits_{\lambda \in \mathbb{R}^m} G(\lambda) \\ \text{avec } G(\lambda) = \min\limits_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda) \end{cases}$$

- On appelle (D) le problème dual du problème (P), qui de ce point de vue, devient le problème primal.
- Le problème dual apparait donc à nouveau comme un problème d'optimisation sans contraintes.

Exemple 2.3 : Cas où f est une fonction quadratique et $g_i, i = 1, \ldots, m$ des fonctions affines. Soit

$$f(x) = \frac{1}{2}\langle Ax, x\rangle_n + \langle b, x\rangle_n + c$$

où A est une matrice d'ordre n, symétrique et définie positive. Soit g_i définie par

$$g_i(x) = \sum_{j=1}^{n} b_{ij}x_j - d_i,\ i = 1, \ldots, m,$$

Vectoriellement, les fonctions affines g_i peuvent s'écrire

$$g(x) = Bx - d, \quad \text{où} \quad B = [b_{ij}]_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant m}} \quad \text{et} \quad d = \begin{pmatrix} d_1 \\ \vdots \\ d_m \end{pmatrix}.$$

On cherche à résoudre le problème

$$(P)\ \begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ Bx = d \end{cases}$$

- On définit d'abord la fonction $\mathscr{L}(x, \lambda)$ par

$$\mathscr{L}(x, \lambda) = \frac{1}{2}\langle Ax, x\rangle_n + \langle b, x\rangle_n + c + \langle Bx - d, \lambda\rangle_m, \quad \text{où} \quad x \in \mathbb{R}^n \text{ et } \lambda \in \mathbb{R}^m.$$

$\mathscr{L}(x, \lambda)$ s'écrit aussi sous la forme suivante :

$$\begin{aligned} \mathscr{L}(x, \lambda) &= \frac{1}{2}\langle Ax, x\rangle_n + \langle b, x\rangle_n + \langle Bx, \lambda\rangle_m - \langle d, \lambda\rangle_m + c \\ &= \frac{1}{2}\langle Ax, x\rangle_n + \langle b, x\rangle_n + \left\langle B^\top \lambda, x\right\rangle_n - \langle d, \lambda\rangle_m + c \end{aligned}$$

- Calculons $G(\lambda) = \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda)$. Soit x_λ le minimum, alors

$$\frac{\partial \mathscr{L}}{\partial x}(x_\lambda,\lambda) = 0, \text{ c'est-à-dire } Ax_\lambda + b + B^\top\lambda = 0. \tag{8}$$

Et c'est bien le minimum uniqne pour chaque λ car le Hessien $\frac{\partial^2 \mathscr{L}}{\partial x^2}(x,\lambda) = A$ qui est définie positive.

- D'où, la fonction duale s'écrit

$$\begin{aligned}
G(\lambda) &= \min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda) = \mathscr{L}(x_\lambda,\lambda) \\
&= \frac{1}{2}\langle Ax_\lambda, x_\lambda\rangle_n + \left\langle b + B^\top\lambda, x_\lambda\right\rangle_n - \langle d,\lambda\rangle_m + c \\
&= \frac{1}{2}\left\langle -b - B^\top\lambda, -A^{-1}\left(b + B^\top\lambda\right)\right\rangle_n + \left\langle b + B^\top\lambda, -A^{-1}\left(b + B^\top\lambda\right)\right\rangle_n \\
&- \langle d,\lambda\rangle_m + c \\
&= -\frac{1}{2}\left\langle b + B^\top\lambda, +A^{-1}\left(b + B^\top\lambda\right)\right\rangle_n - \langle d,\lambda\rangle_m + c \\
&= -\frac{1}{2}\left\langle b, A^{-1}b\right\rangle_n - \frac{1}{2}\left\langle b, A^{-1}B^\top\lambda\right\rangle_n \\
&- \frac{1}{2}\left\langle B^\top\lambda, A^{-1}b\right\rangle_n - \frac{1}{2}\left\langle B^\top\lambda, A^{-1}B^\top\lambda\right\rangle_n - \langle d,\lambda\rangle_m + c \\
&= -\frac{1}{2}\left\langle BA^{-1}B^\top\lambda, \lambda\right\rangle_m - \left\langle BA^{-1}b + d, \lambda\right\rangle_m - \frac{1}{2}\left\langle b, A^{-1}b\right\rangle_n + c
\end{aligned}$$

- Le point λ^* qui réalise le maximum de $G(\lambda)$ sur $\mathbb{R}^n$ vérifie nécessairement

$$\nabla G(\lambda^*) = 0, \text{ i.e, } -BA^{-1}B^\top\lambda^* - BA^{-1}b - d = 0.$$

Si $M = BA^{-1}B^\top$ et si B est de rang m, alors

$$\lambda^* = -M^{-1}\left(BA^{-1}b + d\right) \tag{9}$$

Et c'est bien l'unique solution car le Hessien de G est

$$\nabla^2 G(\lambda) = -BA^{-1}B^\top = -M$$

qui est définie négative car A^{-1} est définie positive et B est de rang m.

- Connaissant λ^*, on peut alors résoudre le problème (P^*) : $\min_{x\in\mathbb{R}^n} \mathscr{L}(x,\lambda^*)$ qui aura comme solution la solution x^* du problème (P). Cette solution est donnée par (8) pour $\lambda = \lambda^*$, i.e,

$$x^* = x_{\lambda^*} = -A^{-1}\left(b + B^\top\lambda^*\right).$$

Exemple 2.4 : Considérons l'exemple suivant : $\min_{x\in S} \frac{1}{2}(Ax,x)+(b,x)$,

avec $A=\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, $b=\begin{pmatrix} -2 \\ 0 \end{pmatrix}$ et $S=\left\{x=(x_1,x_2)\in\mathbb{R}^2 : x_1+x_2=0\right\}$.

- Par rapport à l'exemple précédent, on a $B=(1,1)$, $d=0$ et la fonctionnelle

$$\mathscr{L}(x,\lambda)=x_1^2+x_1x_2+x_2^2-2x_1+\lambda\,(x_1+x_2)$$

- D'après (9), on a

$$\begin{aligned}
\lambda^* &= -M^{-1}\left(BA^{-1}b+d\right) \\
M = BA^{-1}B^\top &= (1,1)\left[\frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\right]\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \frac{1}{3}(1,1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{2}{3} \\
BA^{-1}b &= (1,1)\left[\frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\right]\begin{pmatrix} -2 \\ 0 \end{pmatrix} \\
&= \frac{1}{3}(1,1)\begin{pmatrix} -4 \\ 2 \end{pmatrix} = -\frac{2}{3}
\end{aligned}$$

Alors $\lambda^*=(-3/2)(-2/3)=1$ et la fonctionnelle $\mathscr{L}\,(x_1,\,\lambda^*)$ s'écrit

$$\begin{aligned}
\mathscr{L}\,(x_1,\,\lambda^*) &= x_1^2+x_1x_2+x_2^2-2x_1+x_1+x_2 \\
&= x_1^2+x_1x_2+x_2^2-x_1+x_2
\end{aligned}$$

- A partir de (8), on a

$$\begin{aligned}
x^* = x_{\lambda^*} &= -A^{-1}\left(b+B^\top\lambda^*\right) \\
&= -\frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} -2+1 \\ 0+1 \end{pmatrix} \\
&= -\frac{1}{3}\begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
\end{aligned}$$

Note : A partir de (8) et (9), on peut montrer que

$$P_S\,(x_0)=\left(I-B^\top\left(BB^\top\right)^{-1}B\right)x_0,$$

où $S=\{x\in\mathbb{R}^n;\ Bx=0\}$ avec B une matrice (m,n) de rang m.

Preuve :
En effet ;

$$P_S(x_0) = \underset{\substack{x\in\mathbb{R}^n \\ Bx=0}}{\operatorname{argmin}} \|x - x_0\|^2$$

- On introduit la fonctionnelle

$$\begin{aligned} f(x) &= \|x - x_0\|^2 = (x - x_0, x - x_0) \\ &= (x, x) - 2(x, x_0) + \|x_0\|^2 \end{aligned}$$

  où $A = 2I, b = -2x_0, c = \|x_0\|^2$ et $d = 0$.

- Alors, à partir de (9), on a

$$\begin{aligned} \lambda^* &= -M^{-1}\left(BA^{-1}b + d\right) \\ &= -2\left(BB^\top\right)^{-1}\left(\frac{1}{2}B(-2x_0)\right) = 2\left(BB^\top\right)^{-1} Bx_0 \end{aligned}$$

- Et à partir de (8), nous avons

$$\begin{aligned} Ax^* &= Ax_{\lambda^*} = -b - B^\top\lambda^* \\ 2x^* &= 2x_0 - B^\top\left(2\left(BB^\top\right)^{-1} Bx_0\right) \\ &= 2\left(x_0 - B^\top\left(BB^\top\right)^{-1} Bx_0\right) \end{aligned}$$

  c'est-à-dire

$$P_S(x_0) = x^* = \left(I - B^T\left(BB^T\right)^{-1} B\right)x_0. \qquad \square$$

2.4 Méthodes numériques

• 2.4.1 - Méthode de linéarisation

✓ Dans cette méthode, la fonction objectif et les contraintes sont linéarisées à chaque itération. Rappelons tout d'abord la méthode du gradient dans le cas d'optimisation sans contraintes $\min_{x\in\mathbb{R}^n} F(x)$. Les itérations de la méthode du gradient s'écrivent

$$x^{k+1} = x^k - \gamma\nabla F\left(x^k\right), \ \gamma > 0.$$

Cette méthode peut être écrite sous la forme

$$x^{k+1} = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\left(\frac{1}{2\gamma}\left\|x - x^k\right\|^2 + \left(\nabla F\left(x^k\right), x - x^k\right)\right).$$

✓ Alors, pour le problème avec contraintes

$$(P)\begin{cases} \min_{x\in\mathbb{R}^n} f(x) \\ g_i(x) = 0, \ i = 1, \dots, m, \end{cases}$$

on utilise la même idée après avoir linéarisée les contraintes, c'est-à-dire, en les écrivant comme suit

$$g_i(x) \simeq g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right), \left(x - x^k\right)\right),\ i = 1, \ldots, m.$$

Ainsi, on cherche x^{k+1} le point qui minimise la fonction

$$\begin{cases} x^{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left(\dfrac{1}{2\gamma}\left\|x - x^k\right\|^2 + \left(\nabla f\left(x^k\right), x - x^k\right)\right) \\ \text{sous les contraintes :} \\ g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right), x - x^k\right) = 0,\ \forall\, i = 1, \ldots, m. \end{cases}$$

✓ Soit la fonction de Lagrange associée à ce problème

$$\begin{aligned} \mathscr{L}(x, \lambda) = \frac{1}{2\gamma}\left\|x - x^k\right\|^2 + \left(\nabla f\left(x^k\right), x - x^k\right) \\ + \sum_{i=1}^{m} \lambda_i \left(g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right), x - x^k\right)\right). \end{aligned}$$

Les conditions d'optimalité du point-selle nous donnent

$$\mathscr{L}_x\left(x^{k+1}, \lambda^{k+1}\right) = 0 \tag{10}$$
$$\mathscr{L}_\lambda\left(x^{k+1}, \lambda^{k+1}\right) = 0 \tag{11}$$

Appliquant respectivement les conditions (10) et (11) à la fonctionnelle $\mathscr{L}(x, \lambda)$, on a

$$\begin{aligned} \frac{1}{\gamma}\left(x^{k+1} - x^k\right) + \nabla f\left(x^k\right) + \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i\left(x^k\right) = 0 \\ g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right),\ x^{k+1} - x^k\right) = 0,\ i = 1, \ldots, m,\ i.e., \end{aligned}$$

$$\begin{cases} \dfrac{1}{\gamma}\left(x^{k+1} - x^k\right) + \displaystyle\sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i\left(x^k\right) = -\nabla f\left(x^k\right),\ \gamma > 0, \\ \left(\nabla g_i\left(x^k\right), x^{k+1} - x^k\right) = -g_i\left(x^k\right),\ i = 1, \ldots, m. \end{cases} \tag{12}$$

C'est la méthode de linéarisation qui consiste à résoudre un système linéaire de $(n+m)$ équations à $(n+m)$ inconnues $\left(x^{k+1}, \lambda^{k+1}\right)$.

Exemple 2.5 : Reprenons l'exemple

$$\min_{x+y=1} \left(x^2 + y^2\right)$$

- En posant $f(x,y) = x^2 + y^2$ et $g(x,y) = x + y - 1$, on a les gradients

$$\nabla f(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}, \ \nabla g(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- Alors, on a à partir de la linéairisation (12) le système suivant

$$\begin{cases} \dfrac{1}{\gamma}\begin{pmatrix} x_{k+1} - x_k \\ y_{k+1} - y_k \end{pmatrix} + \lambda_{k+1}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = -\begin{pmatrix} 2x_k \\ 2y_k \end{pmatrix} \\ \qquad x_{k+1} - x_k + y_{k+1} - y_k = -(x_k + y_k - 1) \end{cases}$$

- C'est-à-dire

$$\begin{cases} \dfrac{1}{\gamma}x_{k+1} + \lambda_{k+1} = \left(\dfrac{1}{\gamma} - 2\right)x_k \\ \dfrac{1}{\gamma}y_{k+1} + \lambda_{k+1} = \left(\dfrac{1}{\gamma} - 2\right)y_k \\ \quad x_{k+1} + y_{k+1} = 1 \end{cases}$$

2.4.2 - Méthodes Duales

La méthode de linéarisation ne contient pas explicitement le lagrangien. Les méthodes que nous allons décrire ici sont des méthodes qui utilisent explicitement les propriétés du lagrangien. Reprenons le problème sous contraintes suivant

$$(P)\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x) \\ g_i(x) = 0, \ i = 1, \ldots, m, \end{cases}$$

où la fonction de Lagrange associée au problème (P) est donnée par

$$\mathscr{L}(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

Méthode d'Uzawa Rappelons que le point-selle (x^*, λ^*) de $\mathscr{L}(\cdot,\cdot)$ est défini par

$$\mathscr{L}(x^*, \lambda^*) = \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda^*)$$
$$\mathscr{L}(x^*, \lambda^*) = \max_{\lambda \in \mathbb{R}^m} \mathscr{L}(x^*, \lambda)$$

La méthode d'Uzawa consiste à faire les étapes suivantes :

✓ Étant donné λ^k, on cherche x^k tel que

$$\mathscr{L}\left(x^k, \lambda^k\right) = \min_{x \in \mathbb{R}^n} \mathscr{L}\left(x, \lambda^k\right).$$

✓ Après avoir déterminer x^k, on applique la méthode du gradient au problème

$$\max_{\lambda\in\mathbb{R}^m} \mathscr{L}\left(x^k,\lambda\right), \text{ c'est-à-dire } \lambda^{k+1} = \lambda^k + \gamma \mathscr{L}_\lambda\left(x^k,\lambda^k\right), \ \gamma > 0.$$

Comme

$$\mathscr{L}(x,\lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x), \text{ alors } \mathscr{L}_\lambda\left(x^k,\lambda^k\right) = \begin{pmatrix} g_1\left(x^k\right) \\ \vdots \\ g_m\left(x^k\right) \end{pmatrix} = g\left(x^k\right).$$

Algorithme : L'algorithme d'Uzawa s'écrit alors :

1. Étant donné λ^0.
2. Déterminer x^k par :
$$\mathscr{L}\left(x^k,\lambda^k\right) = \min_{x\in\mathbb{R}^n} \mathscr{L}\left(x,\lambda^k\right)$$
3. Calculer λ^{k+1} par : $\lambda^{k+1} = \lambda^k + \gamma g\left(x^k\right), \ \gamma > 0$.

Méthode d'Arrow-Hurwicz On repart toujours du point-selle (x^*,λ^*) de $\mathscr{L}(\cdot,\cdot)$

$$\mathscr{L}\left(x^*,\lambda^*\right) = \min_{x\in\mathbb{R}^n} \mathscr{L}\left(x,\lambda^*\right)$$
$$\mathscr{L}\left(x^*,\lambda^*\right) = \max_{\lambda\in\mathbb{R}^m} \mathscr{L}\left(x^*,\lambda\right)$$

Cette fois-ci, on applique la méthode du gradient aux deux problèmes suivants :

$$\min_{x\in\mathbb{R}^n} \mathscr{L}\left(x,\lambda^k\right) \text{ et } \max_{\lambda\in\mathbb{R}^m} \mathscr{L}\left(x^k,\lambda\right),$$

lorsque $\left(x^k,\lambda^k\right)$ est donné. Ceci nous amène à l'algorithme d'Arrow-Hurwicz.

Algorithme : L'algorithme d'Arrow-Hurwicz s'écrit comme suit :

1. Étant donné (x^0,λ^0).
2. Déterminer (x^{k+1},λ^{k+1}) par :
$$\begin{aligned} x^{k+1} &= x^k - \gamma\left(\nabla f\left(x^k\right) + \sum_{i=1}^m \lambda_i^k \nabla g_i\left(x^k\right)\right) \\ \lambda^{k+1} &= \lambda^k + \gamma g\left(x^k\right), \ \gamma > 0 \end{aligned}$$

Exemple 2.6 : Reprenons encore une fois l'exemple $\min\limits_{x+y=1} x^2+y^2$. On a,

$$\mathscr{L}(x,y,\lambda) = x^2+y^2+\lambda(x+y-1).$$

- **Méthode d'Uzawa**

✓ On a $\mathscr{L}\left(x^k,y^k,\lambda^k\right) = \min\limits_{(x,y)\in\mathbb{R}^2} \mathscr{L}\left(x,y,\lambda^k\right)$.

✓ Alors,

$$\begin{aligned}\mathscr{L}_x\left(x^k,y^k,\lambda^k\right) = 0 \;\Rightarrow\; 2x^k+\lambda^k = 0\\ \mathscr{L}_y\left(x^k,y^k,\lambda^k\right) = 0 \;\Rightarrow\; 2y^k+\lambda^k = 0\end{aligned}$$

✓ L'algorithme d'Uzawa appliqué au problème de minimisation s'écrit : λ^0 donné

$$\begin{cases} 2x^k+\lambda^k = 0\\ 2y^k+\lambda^k = 0\\ \lambda^{k+1} = \lambda^k+\gamma\left(x^k+y^k-1\right),\ \gamma>0\end{cases} \tag{13}$$

✓ Pour cet exemple, l'algorithme (13) se ramène à des itérations sur λ seulement. En effet, $x^k = -\frac{\lambda^k}{2}$ et $y^k = -\frac{\lambda^k}{2}$, alors

$$\lambda^{k+1} = \lambda^k+\gamma\left(-\lambda^k-1\right),\ \text{ i.e. } \lambda^{k+1} = (1-\gamma)\lambda^k-\gamma$$

- **Méthode d'Arrow - Hurwicz**

✓ Étant donné (x^0,λ^0)

✓ Calculer

$$\begin{aligned}\begin{pmatrix} x^{k+1}\\ y^{k+1}\end{pmatrix} = \begin{pmatrix} x^k\\ y^k\end{pmatrix} - \gamma\begin{pmatrix} 2x^k+\lambda^k\\ 2y^k+\lambda^k\end{pmatrix}\\ \lambda^{k+1} = \lambda^k+\gamma\left(x^k+y^k-1\right)\end{aligned}$$

✓ C'est-à-dire plus simplement

$$\begin{aligned} x^{k+1} &= (1-2\gamma)x^k-\gamma\lambda^k\\ y^{k+1} &= (1-2\gamma)y^k-\gamma\lambda^k\\ \lambda^{k+1} &= \lambda^k+\gamma\left(x^k+y^k-1\right)\end{aligned}$$

- ### 2.4.3 - Méthode de Newton

Les conditions d'optimalité du point-selle de la fonction de Lagrange $\mathscr{L}(x, \lambda)$ sont

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) = 0 \\ \qquad\qquad\qquad g_i(x) = 0,\ i = 1, \ldots, m. \end{cases} \tag{14}$$

Le système (14) est un système non-linéaire de $(n+m)$ inconnues (x, λ). La méthode de Newton consiste, à partir d'un point $\left(x^k, \lambda^k\right)$:

✓ À linéariser (14) au voisinage de $\left(x^k, \lambda^k\right)$;

✓ Et à définir $\left(x^{k+1}, \lambda^{k+1}\right)$ comme la solution du problème linéarisé :

$$\begin{cases} \nabla f\left(x^k\right) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i\left(x^k\right) + \left(\nabla^2 f\left(x^k\right) + \sum_{i=1}^{m} \lambda_i^k \nabla^2 g_i\left(x^k\right)\right)\left(x^{k+1} - x^k\right) \\ \qquad\qquad + \sum_{i=1}^{m} \left(\lambda_i^{k+1} - \lambda_i^k\right) \nabla g_i\left(x^k\right) = 0 \\ \qquad g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right), \left(x^{k+1} - x^k\right)\right) = 0,\ i = 1, \ldots, m. \end{cases} \tag{15}$$

✓ En remplaçant les termes de (15) par les dérivées correspondantes du Lagrangien

$$\mathscr{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x),$$

on peut réécrire le système (15) sous la forme suivante :

$$\left[\begin{array}{c|c} \frac{\partial^2 \mathscr{L}}{\partial x^2}\left(x^k, \lambda^k\right) & \nabla g_1\left(x^k\right) \ldots \nabla g_m\left(x^k\right) \\ \hline \nabla g_1^\top\left(x^k\right) & \\ \vdots & 0 \\ g_m^\top\left(x^k\right) & \end{array}\right] \begin{bmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{bmatrix} = \begin{bmatrix} -\frac{\partial \mathscr{L}}{\partial x}\left(x^k, \lambda^k\right) \\ -g\left(x^k\right) \end{bmatrix}$$

- ### 2.4.4 - Méthode de pénalisation

Pour pénaliser le problème sous contraintes suivant

$$(P) \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ g_i(x) = 0,\ i = 1, \ldots, m, \end{cases}$$

✓ On pénalise la fonction objectif $f(x)$ par le terme de «pénalisation» $\frac{r}{2}\|g(x)\|^2$.

✓ Et on résout le problème (d'optimisation sans contraintes) suivant :

$$\min_{x \in \mathbb{R}^n} \varphi(x, r), \text{ où } \varphi(x, r) = f(x) + \frac{r}{2}\|g(x)\|^2 \text{ et } r > 0.$$

✓ Lorsque $r \longrightarrow +\infty$, on montre que les contraintes du problème (P) sont satisfaites.

Exemple 2.7 : Considérons le problème de minimisation suivant

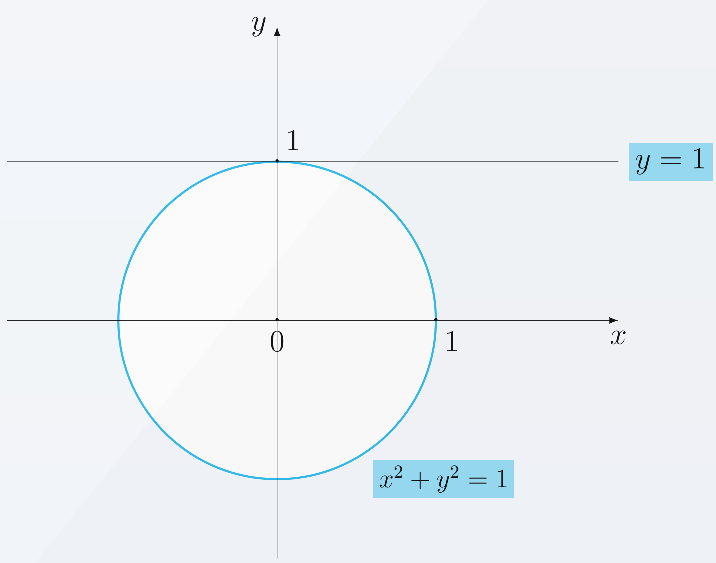
$$\min_{y=1} \frac{1}{2}\left(x^2 + y^2\right).$$

- Il est tout à fait évident que la solution de ce problème est $(0, 1)$.
- La fonctionnelle pénalisée est définie par

$$\varphi(x, y, r) = \frac{1}{2}\left(x^2 + y^2\right) + \frac{r}{2}(y - 1)^2$$

où en voulant déterminer le minimum, on annulle son gradient :

$$\begin{cases} \frac{\partial \varphi}{\partial x} = 0 \\ \frac{\partial \varphi}{\partial y} = 0 \end{cases} \Rightarrow \begin{cases} x = 0 \\ y + r(y - 1) = 0 \end{cases} \Rightarrow \begin{cases} x = 0 \\ y = \frac{r}{1+r} \end{cases}$$



r	y
1	0.5
2	0.67
10	0.91
20	0.95
50	0.98
100	0.99
∞	1

3 Contraintes d'inégalité

3.1 Conditions de Kuhn-Tucker

On s'intéresse maintenant au problème de minimisation suivant :

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} f(x) \\ g_i(x) = 0,\ i = 1, \ldots, m, \\ h_i(x) \leqslant 0,\ i = 1, \ldots, p, \end{cases} \tag{16}$$

où les fonctions f, $g_i, i = 1, \ldots, m$ et $h_i, i = 1, \ldots, p$ sont de classe C^1.

• 3.1.1 - Conditions nécessaires du 1^{er} ordre

Définition 3.1

Soit $x^* \in \mathbb{R}^n$ un point admissible, c'est-à-dire, $g_i(x^*) = 0,\ i = 1, \ldots,$ m et $h_i(x^*) \leq 0, i = 1, \ldots, p$. Soit I^* l'ensemble des contraintes saturées, c'est-à-dire,

$$I^* = \{i \mid h_i(x^*) = 0, i = 1, \ldots, p\}.$$

Alors x^* est dit point régulier si les vecteurs $\nabla g_i(x^*), i = 1, \ldots, m$ et $\nabla h_i(x^*), i \in I^*$, sont linéairement indépendants.

Théorème 3.1

Kuhn-Tucker

Soit x^* un minimum du problème (16). Si x^* est un point régulier, alors il existe $\lambda \in \mathbb{R}^m$ et $\mu \in \mathbb{R}^p$ où $\mu \geqslant 0$, tels que

$$\begin{cases} \nabla f(x^*) + \sum\limits_{i=1}^{m} \lambda_i \nabla g_i(x^*) + \sum\limits_{i=1}^{p} \mu_i \nabla h_i(x^*) = 0 \\ \mu_i h_i(x^*) = 0,\ i = 1, \ldots, p \end{cases}$$

Exemple 3.1 : Considérons le problème de minimisation suivant :

$$\begin{cases} \min\ (x-1)^2 + y - 2 \\ y - x = 1 \\ x + y \leqslant 2 \end{cases}$$

- Définissons et déterminons les gradients associés aux fonctions suivantes

$$f(x,y) = (x-1)^2 + y - 2, \qquad \nabla f(x,y) = \begin{pmatrix} 2(x-1) \\ 1 \end{pmatrix}$$

$$g(x,y) = y - x - 1, \qquad \nabla g(x,y) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$h(x,y) = x + y - 2, \qquad \nabla h(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

- Remarquons qu'ici, tout point de $\mathbb{R}^2$ est régulier. En effet :
  - ✓ Soit que la contrainte $h(x,y)$ n'est pas saturée et dans ce cas tout revient à calculer $\nabla g(x,y) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ qui est libre.
  - ✓ Soit que la contrainte $h(x,y)$ est saturée et dans ce cas I^* correspond à la contrainte $h(x,y)$ et on a $\nabla g(x,y) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ et $\nabla h(x,y) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ qui sont linéairement indépendants.
- Les conditions de Kuhn-Tucker disent qu'il existe $\lambda \in \mathbb{R}$ et $\mu \in \mathbb{R}$ où $\mu \geqslant 0$ t.q.

$$2(x-1) - \lambda + \mu = 0, \tag{17}$$
$$1 + \lambda + \mu = 0, \tag{18}$$
$$\mu(x + y - 2) = 0, \tag{19}$$

et aussi le point (x,y) doit être admissible, c'est-à-dire,

$$y - x = 1 \text{ et } x + y \leqslant 2. \tag{20}$$

- L'équation (18) nous donne $\lambda = -1 - \mu$, alors d'après (17) on obtient $x = \frac{1-2\mu}{2}$.
- D'autre part à partir de (20), nous avons $y = 1 + x = \frac{3-2\mu}{2}$.
- Et donc (19) devient en remplaçant x et y par leur valeur

$$\mu(u + y - 2) = -2\mu\mu = 0 \;\Rightarrow\; \mu = 0.$$

- Alors $x = 1/2$ et $y = 3/2$, et ces deux valeurs vérifient bien $x + y \leqslant 2$. Donc

$$(x^*, y^*) = \left(\frac{1}{2}, \frac{3}{2}\right), \; \lambda^* = -1 \text{ et } \mu^* = 0.$$

1$^{\text{er}}$ Cas particulier : Le premier cas particulier que nous allons étudier, est le problème

$$\begin{cases} \min f(x) \\ g_i(x) = 0, \; i = 1, \ldots, m, \\ x \geqslant 0. \end{cases} \tag{21}$$

✓ Ici les contraintes d'inégalités sont définies par les fonctions suivantes

$$h_i(x) = -x_i \; i = 1, \ldots, n.$$

✓ Soit x^* un minimum régulier, alors il existe $\lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^n, \mu \geqslant 0$ tels que

$$\begin{cases} \nabla f\left(x^*\right) + \sum\limits_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) + \sum\limits_{i=1}^{n} \mu_i \nabla h_i\left(x^*\right) = 0 \\ \hfill \mu_i h_i\left(x^*\right) = 0, \; i = 1, \ldots, n \end{cases}$$

✓ Sachant que $\nabla h_i\left(x^*\right) = \begin{pmatrix} 0 \\ \vdots \\ -1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i^{\text{ ième composante}}$, alors on obtient

$$\begin{aligned} \nabla f\left(x^*\right) + \sum_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) &= \mu \geqslant 0, \\ \mu_i h_i\left(x^*\right) &= 0, \;\; \text{c'est-à-dire} \;\; \mu_i x_i^* = 0. \end{aligned}$$

✓ Alors les conditions de Kuhn-Tucker pour le problème (21) s'écrivent :

$$\text{Il existe}\, \lambda \in \mathbb{R}^m \text{ tel que :}$$

$$\begin{cases} \nabla f\left(x^*\right) + \sum\limits_{i=1}^{m} \lambda_i \nabla g_i\left(x^*\right) \geqslant 0, \\ \left(\dfrac{\partial f}{\partial x_j}\left(x^*\right) + \sum\limits_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial x_j}\left(x^*\right)\right) x_j^* = 0, \; j = 1, \ldots, n, \\ g_i\left(x^*\right) = 0, \; i = 1, \ldots, m, \\ x^* \geqslant 0. \end{cases}$$

L'intérêt de cette écriture est que la variable μ est éliminée.

Exemple 3.2 : Considérons le problème de minimisation suivant :

$$\begin{cases} \min(x-2)^2 + (y-2)^2 \\ x + 2y = 4 \\ x \geqslant 0, y \geqslant 0 \end{cases}$$

- On introduit les fonctions f, g suivantes et leur gradient

$$f(x, y) = (x-2)^2 + (y-2)^2, \;\; \nabla f(x, y) = \begin{pmatrix} 2(x-2) \\ 2(y-2) \end{pmatrix}$$

$$g(x, y) = x + 2y - 4, \qquad \nabla g(x, y) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

- Les conditions de Kuhn-Tucker nous donnent : Il existe $\lambda \in \mathbb{R}$ tel que

$$2(x-2)+\lambda \geq 0 \tag{22}$$
$$2(y-2)+2\lambda \geqslant 0 \tag{23}$$
$$(2(x-2)+\lambda)x = 0 \tag{24}$$
$$(2(y-2)+2\lambda)y = 0 \tag{25}$$

- Et comme le point (x, y) doit être admissible, alors

$$x+2y-4=0,\ x \geqslant 0,\ y \geqslant 0 \tag{26}$$

- L'équation (24) implique que $x = 0$ ou $2(x-2)+\lambda = 0$.
- Alors que l'équation (25) donne $y = 0$ ou $y-2+\lambda = 0$.
- Nous avons alors à étudier les différents cas suivants :
  - ✓ 1^{er} cas : $x = 0$ et $y = 0$. Impossible d'après (26) !
  - ✓ $2^{\text{ième}}$ cas : $x = 0$ et $y-2+\lambda = 0$. Alors d'après (26), on a $y = 2$ et donc $\lambda = 0$. Ce qui est impossible d'après (22).
  - ✓ $3^{\text{ième}}$ cas : $2(x-2)+\lambda = 0$ et $y = 0$. Alors d'après (26), on a $x = 4$ et donc $\lambda = -4$. Ce qui est impossible d'après (23).
  - ✓ $4^{\text{ième}}$ cas :

$$\begin{cases} 2(x-2)+\lambda = 0 \\ \quad y-2+\lambda = 0 \end{cases}$$

On a aussi d'après (26), $x+2y-4 = 0$. Alors, on a à résoudre le système

$$\begin{cases} 2x-4+\lambda = 0 \\ y-2+\lambda = 0 \\ x+2y-4 = 0 \end{cases}$$

Les deux premières équations donne $2x-y-2 = 0$, donc on obtient finalement

$$\begin{cases} 2x-y = 2 \\ x+2y = 4 \end{cases}$$

Ce qui fournit la solution $x = 8/5$ et $y = 6/5$ et $\lambda = 2-y = 4/5$.
Le point $(x^*, y^*) = (8/5, 6/5)$ et $\lambda = 4/5$ vérifient bien toutes les équations.

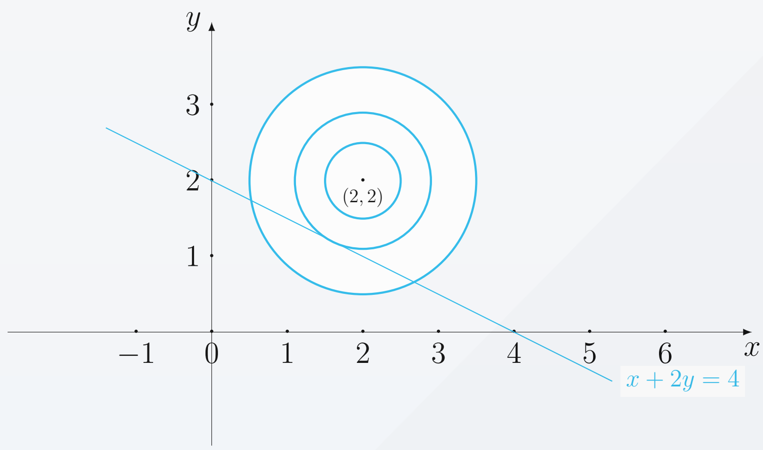
- Vérifions la régularité du point (x^*, y^*). On a $\nabla g(x^*, y^*) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ qui est libre, donc (x^*, y^*) est un point régulier.

L'exemple qu'on vient de traiter peut être résolu géométriquement.

- Les courbes de niveau de la fonction à minimiser $f(x,y)$ données par

$$(x-2)^2 + (y-2)^2 = c,$$

qui sont des cercles centrés au point $(2,2)$.



- La solution est le point tel que la droite $x+2y=4$ est tangente au cercle.
- Un vecteur directeur à la droite est $\vec{t} = \begin{pmatrix} 1 \\ -1/2 \end{pmatrix}$ et une normale au cercle est $\vec{n} = \begin{pmatrix} 2(x-2) \\ 2(y-2) \end{pmatrix}$, alors on obtient la condition d'orthogonalité suivante

$$\begin{aligned} \vec{n}\cdot\vec{t} = 0 \ &\Leftrightarrow\ 2(x-2)-(y-2) = 0 \\ &\Leftrightarrow\ 2x - y = 2 \end{aligned}$$

- Aussi la solution appartient à la droite d'équation $x+2y=4$. D'où, on a à résoudre

$$\begin{cases} 2x - y = 2 \\ x + 2y = 4 \end{cases}$$

- Ce qui donne $x = 8/5$ et $y = 6/5$ et c'est l'unique solution.

2$^{\text{ième}}$ Cas particulier : Le deuxième cas particulier est le problème suivant

$$\begin{cases} \min f(x) \\ g_i(x) \leq 0, \ \ i = 1,\ldots,m \\ x \geqslant 0, \end{cases} \tag{27}$$

✓ Si x^* est un minimum régulier, alors il existe $\lambda \in \mathbb{R}^m, \lambda \geqslant 0, \mu \in \mathbb{R}^n, \mu \geqslant 0$ tels que

$$\begin{cases} \nabla f\left(x^*\right)+\sum_{i=1}^m \lambda_i \nabla g_i\left(x^*\right)+\sum_{i=1}^n \mu_i \nabla h_i\left(x^*\right)=0, \\ \lambda_i g_i\left(x^*\right)=0,\ i=1,\ldots,m, \\ \mu_i h_i\left(x^*\right)=0,\ i=1,\ldots,n, \end{cases}$$

où $h_i(x)=-x_i,\ i=1,\ldots,n$.

✓ Là aussi on peut éliminer μ comme dans le premier cas (au problème (21)).

✓ On obtient ainsi les conditions suivantes :

$$\begin{cases} \lambda \in \mathbb{R}^m,\ \lambda \geqslant 0,\ g_i\left(x^*\right) \leqslant 0,\ i=1,\ldots,m,\ x^* \geqslant 0, \\ \nabla f\left(x^*\right)+\sum_{i=1}^m \lambda_i \nabla g_i\left(x^*\right) \geqslant 0, \\ \left(\frac{\partial f}{\partial x_j}\left(x^*\right)+\sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_j}\left(x^*\right)\right) x_j^*=0,\ j=1,\ldots,n, \\ \lambda_i g_i\left(x^*\right)=0,\ i=1,\ldots,m. \end{cases}$$

Exemple 3.3 : Considérons le problème de minimisation suivant :

$$\begin{cases} \min\left(x^2+3y^2-4x-6y\right) \\ x+2y \leq 4 \\ x \geqslant 0, y \geqslant 0 \end{cases}$$

- On définit les fonctions et leur gradient comme suit

$$f(x,y)=x^2+3y^2-4x-6y,\quad \nabla f(x,y)=\begin{pmatrix} 2x-4 \\ 6y-6 \end{pmatrix}$$

$$g(x,y)=x+2y-4,\qquad \nabla g(x,y)=\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

- On obtient les conditions suivantes : $\lambda \in \mathbb{R}, \lambda \geqslant 0$

$$2x-4+\lambda \geqslant 0, \tag{28}$$
$$6y-6+2\lambda \geqslant 0, \tag{29}$$
$$(2x-4+\lambda)x=0, \tag{30}$$
$$(6y-6+2\lambda)y=0, \tag{31}$$
$$\lambda(x+2y-4)=0, \tag{32}$$
$$x+2y \leqslant 4,\ x \geqslant 0,\ y \geqslant 0. \tag{33}$$

- L'équation (32) donne $\lambda=0$ ou $x+2y-4=0$.

- Si $\lambda = 0$, alors (30) donne $x = 0$ ou $x = 2$:
  - ✓ Si $x = 0$, alors (28) n'est pas vérifiée donc $x = 2$.
  - ✓ Dans ce cas (31) génère $(y-1)y = 0 \Rightarrow y = 0$ ou $y = 1$.
  - ✓ Si $y = 0$, alors (29) n'est pas vérifiée ; donc $y = 1$.

  Le point $(x^*, y^*) = (2, 1)$ et $\lambda = 0$ vérifient toutes les équations.
- Si $\lambda \neq 0$, des calculs similaires montrent qu'il n'y a pas de solution dans ce cas.
- Là encore pour vérifier que le point $(x^*, y^*) = (2, 1)$ est régulier, ça revient à calculer $\nabla g(x^*, y^*) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ qui est libre (puisque la seule contrainte saturée est $g(x, y)$).

3.1.2 - Conditions suffisantes du $2^{\text{ième}}$ ordre

Retournons au problème de départ, c'est-à-dire au problème d'optimisation suivant :

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} f(x), \\ g_i(x) = 0, \ i = 1, \ldots, m, \\ h_i(x) \leqslant 0, \ i = 1, \ldots, p. \end{cases} \tag{34}$$

Introduisons la fonction de Lagrange suivante

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \mu_i h_i(x).$$

Théorème 3.2

Soit x^* un point admissible et régulier. On suppose que les fonctions f, g_i, $1 \leq i \leq m$ et $h_i, 1 \leqslant i \leqslant p$ sont de classe C^2 dans un voisinage de x^*. S'il existe $\lambda^* \in \mathbb{R}^m$, $\mu^* \in \mathbb{R}^p$, $\mu^* \geq 0$ tels que

$$\begin{cases} L_x(x^*, \lambda^*, \mu^*) = 0 \\ \quad \mu_i^* h_i(x^*) = 0, \ 1 \leq i \leq p \end{cases}$$

(où L_x est le gradient de L en x) et si pour tout $\xi \in \mathbb{R}^n$ et $\xi \neq 0$ tel que

$$\begin{aligned} &(\nabla g_i(x^*), \xi) = 0, \quad 1 \leqslant i \leq m, \\ &(\nabla h_i(x^*), \xi) = 0, \quad i \in I^*, \ \mu_i^* > 0, \\ &(\nabla h_i(x^*), \xi) \geqslant 0, \quad i \in I^*, \ \mu_i^* = 0, \end{aligned}$$

on a

$$(L_{xx}(x^*, \lambda^*, \mu^*)\, \xi, \ \xi) > 0,$$

alors x^* est un minimum local pour le problème (34).

Note :

- Sous les conditions du **Théorème 3.2**, x^* est localement unique.
- Sous les conditions du **Théorème 3.1**, les réels $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}^p$, $\mu \geq 0$ sont déterminés de manière unique.

Exemple 3.4 : Reprenons les exemples précédents.

1. Considérons le problème de minimisation suivant :

$$\begin{cases} \min\ (x-1)^2 + (y-2) \\ y - x = 1 \\ x + y \leq 2 \end{cases} \tag{35}$$

✓ On a $(x^*, y^*) = (1/2,\ 3/2)$, $\lambda^* = -1$, $\mu^* = 0$. Soit les fonctions g et h suivantes

$$g(x,y) = y - x - 1,\ \ h(x,y) = x + y - 2.$$

✓ Soit $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \in \mathbb{R}^2$ avec $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, tel que

$$\begin{aligned} \left(\nabla g\,(x^*, y^*)\,, \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right) &= 0 \iff -\xi + \eta = 0 \\ \left(\nabla h\,(x^*, y^*)\,, \begin{pmatrix} \xi \\ \eta \end{pmatrix} \right) &\geqslant 0 \iff \xi + \eta \geqslant 0 \end{aligned}$$

donc $\eta = \xi$ et $\xi \geqslant 0$.

✓ Et comme $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ alors il faut considérer les points $\begin{pmatrix} \xi \\ \xi \end{pmatrix}$ avec $\xi > 0$.

✓ Considérons la fonction de Lagrange suivante

$$\begin{aligned} L(x,y,\lambda,\mu) &= (x-1)^2 + y - 2 + \lambda(y - x - 1) + \mu(x + y - 2) \\ \nabla_{(x,y)} L &= \begin{pmatrix} 2(x-1) - \lambda + \mu \\ 1 + \lambda + \mu \end{pmatrix} \\ \nabla^2_{(x,y)} L &= \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

✓ Alors $(\xi, \xi) \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \xi \end{pmatrix} = 2\xi^2 > 0$, d'où $(1/2, 3/2)$ est une solution, de plus elle est unique.

2. Considérons maintenant le problème de minimisation suivant :

$$\begin{cases} \min\ (x-2)^2+(y-2)^2 \\ x+2y=4 \\ x\geq 0 \\ y\geq 0 \end{cases}$$

- ✓ On a $(x^*,y^*)=(8/5,\ 6/5),\ \lambda^*=4/5$ et automatiquement $\mu_1^*=\mu_2^*=0$ (car dans les conditions de Kuhn-Tucker, on doit avoir $\mu_1^* x^*=0$ et $\mu_2^* y^*=0$).
- ✓ On a la fonction de Lagrange suivante

$$L\,(x,y,\lambda,\mu_1,\mu_2)=(x-2)^2+(y-2)^2+\lambda(x+2y-4)-\mu_1 x-\mu_2 y$$
$$\nabla(x,y)L=\begin{pmatrix} 2(x-2)+\lambda-\mu_1 \\ 2(y-2)+2\lambda-\mu_2 \end{pmatrix}$$
$$\nabla^2_{(x,y)}L=\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

- ✓ La matrice $\nabla^2_{(x,y)}L$ est définie positive sur $\mathbb{R}^2$.
- ✓ Donc, en particulier la condition suffisante du $2^{\text{ième}}$ ordre est vérifiée.
- ✓ D'où $(8/5,\ 6/5)$ est l'unique solution.

3. Finalement, considérons le problème de minimisation

$$\begin{cases} \min\ (x^2+3y^2-4x-6y) \\ x+2y\leqslant 4 \\ x\geqslant 0 \\ y\geqslant 0 \end{cases}$$

- ✓ On définit la fonctionnelle L et on détermine ses dérivées comme suit :

$$L\,(x,y,\mu_1,\mu_2,\mu_3)=x^2+3y^2-4x-6y+\mu_1(x+2y-4)-\mu_2 x-\mu_3 y$$
$$\nabla(x,y)L=\begin{pmatrix} 2x-4+\mu_1-\mu_2 \\ 6y-6+2\mu_1-\mu_3 \end{pmatrix}$$
$$\nabla^2(x,y)L=\begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$$

- ✓ La matrice $\nabla^2(x,y)L$ est définie positive sur $\mathbb{R}^2$.
- ✓ Et on a la condition suffisante qui est vérifiée.

3.2 Dualité, point-selle

Nous nous intéressons ici au problème

$$(P)\begin{cases}\min\ f(x)\\ g_i(x)\leq 0,\ i=1,\ldots,m,\end{cases}$$

où $x\in\mathbb{R}^n$, les fonctions f et g_i, $1\leqslant i\leqslant m$, sont supposées de classe C^1.
Rappelons que les conditions de Kuhn-Tucker pour le problème (P) s'écrivent :

Si x^* est un mimimum régulier de (P), alors il existe $\lambda^*\in\mathbb{R}^m$, $\lambda^*\geqslant 0$ tels que

$$\begin{cases}\nabla f\left(x^*\right)+\sum_{i=1}^{m}\lambda_i^*\nabla g_i\left(x^*\right)=0,\\ \qquad\qquad\lambda_i^* g_i\left(x^*\right)=0,\ i=1,\ldots,m,\end{cases}$$

de plus x^* est admissible, i.e., $g_i\left(x^*\right)\leq 0$, $1\leq i\leq m$.

La condition pour que x^* soit un point régulier dans le cas du problème (P) est que x^* soit admissible et que $\nabla g_i\left(x^*\right)$, $i\in I^*$, soient linéairement indépendants, où

$$I^*=\left\{i\mid g_i\left(x^*\right)=0,\ i=1,\ldots,m\right\}.$$

Définition 3.2

1. La fonction de Lagrange (ou Lagrangien) associée au problème (P) est définie

$$\mathscr{L}(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i g_i(x).$$

2. Le problème (P) est appelé problème primal.

Théorème 3.3

Conditions du point-selle :
(x^*,λ^*) avec $\lambda^*\geqslant 0$ est un point-selle du Lagrangien $\mathscr{L}(x,\lambda)$ associé au problème primal (P) si et seulement si les conditions suivantes sont satisfaites :

1. x^* minimise $\mathscr{L}\left(x,\lambda^*\right)$ sur $\mathbb{R}^n$,
2. $g_i\left(x^*\right)\leq 0$, $\quad i=1,\ldots,m$,
3. $\lambda_i^* g_i\left(x^*\right)=0$, $\quad i=1,\ldots,m$.

Preuve :
Le point (x^*,λ^*) est un point-selle si et seulement si

$$\mathscr{L}\left(x^*,\lambda\right)\leqslant\mathscr{L}\left(x^*,\lambda^*\right)\leqslant\mathscr{L}\left(x,\lambda^*\right),\ \forall\, x\in\mathbb{R}^n,\forall\,\lambda\geqslant 0.$$

Condition nécessaire

- ✓ On a

$$\mathscr{L}\left(x^*,\lambda^*\right) \leq \mathscr{L}\left(x,\lambda^*\right),\ \forall\, x \in \mathbb{R}^n \ \text{ alors, on obtient } \boxed{1}.$$

- ✓ Maintenant

$$\begin{aligned}\mathscr{L}\left(x^*,\lambda\right) &\leqslant \mathscr{L}^*\left(x^*,\lambda^*\right)\\ f\left(x^*\right)+\sum_{i=1}^m \lambda_i g_i\left(x^*\right) &\leq f\left(x^*\right)+\sum_{i=1}^m \lambda_i^* g_i\left(x^*\right)\\ \sum_{i=1}^m\left(\lambda_i-\lambda_i^*\right) g_i\left(x^*\right) &\leq 0,\ \forall\ \lambda \geqslant 0\end{aligned}$$

- ✓ Ceci extraînce, forcément, $g_i\left(x^*\right) \leq 0, \forall\ i \in \{1,\ldots,m\}$ car s'il existe $k \in \{1,\ldots,m\}$ tel que $g_k\left(x^*\right)>0$ alors le choix de

$$\lambda=\left(\lambda_1^*,\lambda_2^*,\ldots,\lambda_{k-1}^*,\lambda_k,\lambda_{k+1}^*,\ldots,\lambda_n^*\right) \geqslant 0,$$

  nous donne $\left(\lambda_k-\lambda_k^*\right) g_k\left(x^*\right)$, pour tout $\lambda_k \geqslant 0$.
- ✓ Et il suffit de prendre $\lambda_k>\lambda_k^*$ ce qui nous donne $\left(\lambda_k-\lambda_k^*\right) g_k\left(x^*\right)>0$, ceci est une contradiction. Alors, la condition $\boxed{2}$ est vérifiée.
- ✓ Reste à montrer la condition $\boxed{3}$. A partir de

$$\sum_{i=1}^m\left(\lambda_i-\lambda_i^*\right) g_i\left(x^*\right) \leq 0,\ \forall\ \lambda \geqslant 0$$

  on a en particulier (en prenant $\lambda=0$),

$$-\sum_{i=1}^m \lambda_i^* g_i\left(x^*\right) \leq 0 \ \Rightarrow\ \sum_{i=1}^m \lambda_i^* g_i\left(x^*\right) \geqslant 0.$$

  Or $g_i\left(x^*\right) \leq 0$ d'après la condition $\boxed{2}$, donc on obtient

$$\begin{aligned}&\sum_{i=1}^m \lambda_i g_i\left(x^*\right) \leq 0, \text{ car } \ \lambda^* \geqslant 0, \ \text{ d'où}\\ &\sum_{i=1}^m \lambda_i^* g_i\left(x^*\right)=0 \Rightarrow\ \lambda_i^* g_i\left(x^*\right)=0, \forall\, i \in \{1,\ldots,m\},\end{aligned}$$

  car tous les termes $\lambda_i^* g_i\left(x^*\right)$ sont de même signe.

Condition suffisante

- ✓ La condition $\boxed{3}$ implique que $\mathscr{L}\left(x^*,\lambda^*\right) \leq \mathscr{L}\left(x,\lambda^*\right),\ \forall\, x \in \mathbb{R}^n$.
- ✓ Reste à montrer que

$$\mathscr{L}\left(x^*,\lambda\right) \leqslant \mathscr{L}\left(x^*,\lambda^*\right),\ \forall\ \lambda \geqslant 0.$$

✓ On a

$$\mathscr{L}(x^*, \lambda) = f(x^*) + \sum_{i=1}^{m} \lambda_i g_i(x^*) \leq f(x^*)$$

car d'après 2, on a $g_i(x^*) \leqslant 0$ et donc $\sum_{i=1}^{m} \lambda_i g_i(x^*) \leqslant 0$ pour tout $\lambda \geqslant 0$.

✓ D'autre part

$$\mathscr{L}(x^*, \lambda^*) = f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) = f(x^*),$$

d'après 3.

✓ D'où

$$\mathscr{L}(x^*, \lambda) \leqslant \mathscr{L}(x^*, \lambda^*), \; \forall \, \lambda \geqslant 0. \qquad \square$$

Théorème 3.4

Si (x^*, λ^*) est un point-selle avec $\lambda^* \geqslant 0$, alors x^* est un minimum global du problème (P).

Preuve :

- D'après le Théorème 3.3, on a $g_i(x^*) \leq 0, \; \forall \, i, 1 \leq i \leq m$.
- Il reste à montrer que

$$f(x^*) \leqslant f(x), \; \forall \, x, \; g_i(x) \leqslant 0, \; i = 1, \ldots, m.$$

- On a $\mathscr{L}(x^*, \lambda^*) \leqslant \mathscr{L}(x, \lambda^*), \; \forall \, x \in \mathbb{R}^n$ alors

$$f(x^*) + \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \leqslant f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x), \; \forall \, x \in \mathbb{R}^n.$$

- Mais d'après le Théorème 3.3, on a

$$\lambda_i^* g_i(x^*) = 0, \; \forall \, i, 1 \leqslant i \leqslant m,$$

alors

$$f(x^*) \leq f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x), \; \forall \, x \in \mathbb{R}^n.$$

Et en particulier

$$f(x^*) \leq f(x), \; \forall \, x, \; g_i(x) \leq 0, i = 1, \ldots, m, \; \text{car} \; \lambda^* \geq 0. \qquad \square$$

Note : Si les fonctions f et $g_i, i = 1, \ldots, m$, sont convexes et s'il existe un point $\xi^0 \in \mathbb{R}^n$ tel que $g_i(\xi^0) < 0,\ i = 1, \ldots, m$, alors les conditions du point-selle (Théorème 3.4) sont des conditions nécessaires et suffisantes.

Résumons la situation :

- Problème primal :

$$\begin{cases} \min\ f(x) \\ g_i(x) \leq 0, i = 1, \ldots, m. \end{cases}$$

- Lagrangien :

$$\mathscr{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

- Fonction duale :

$$G(\lambda) = \min_{x \in \mathbb{R}^m} \mathscr{L}(x, \lambda)$$

- Problème dual : $\max\limits_{\lambda \in D} G(\lambda)$, où $D = \{\lambda \mid G(\lambda) \text{ existe}, \lambda \geqslant 0\}$.
- Comme dans le cas d'optimisation avec contraintes égalités, on détermine λ^* t.q.

$$G(\lambda^*) = \max_{\lambda \in D} G(\lambda)$$

puis on détermine x^* par

$$\mathscr{L}(x^*, \lambda^*) = \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda^*).$$

Exemple 3.5 : Considérons le problème de minimisation suivant :

$$\min_{x+y \leq -2} (x^2 + 2x + y^2 + y - 1)$$

- On définit et détermine les dérivées du lagrangien

$$\begin{aligned} \mathscr{L}(x, y, \lambda) &= x^2 + 2x + y^2 + y - 1 + \lambda(x + y + 2) \\ \frac{\partial \mathscr{L}(x, y, \lambda)}{\partial x} &= 2x + 2 + \lambda = 0 \ \Rightarrow\ x = -\frac{\lambda + 2}{2} \\ \frac{\partial \mathscr{L}(x, y, \lambda)}{\partial y} &= 2y + 1 + \lambda = 0 \ \Rightarrow\ y = -\frac{\lambda + 1}{2}. \end{aligned}$$

- De plus $\nabla^2_{(x,y)} \mathscr{L} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ qui est définie positive.
- Donc $\left(-\frac{\lambda+2}{2}, -\frac{\lambda+1}{2}\right)$ est bien le minimum de $\mathscr{L}(x, y, \lambda)$ pour chaque λ.

- La fonction duale et sa dérivée deviennent

$$
\begin{aligned}
G(\lambda) &= \frac{(\lambda+2)^2}{4} - \lambda - 2 + \frac{(\lambda+1)^2}{4} - \frac{\lambda+1}{2} - 1 + \lambda(-\lambda - 3/2 + 2) \\
&= \frac{\lambda^2}{4} + \lambda + 1 - \lambda - 2 + \frac{\lambda^2}{4} + \frac{\lambda}{2} + \frac{1}{4} - \frac{\lambda}{2} - \frac{1}{2} - 1 - \lambda^2 + \frac{\lambda}{2} \\
&= -\frac{\lambda^2}{2} + \frac{\lambda}{2} - \frac{9}{4} \\
G'(\lambda) &= -\lambda + 1/2 = 0 \ \Rightarrow\ \lambda = 1/2.
\end{aligned}
$$

- Comme $\lambda^* = 1/2 > 0$, et que $G''(\lambda) = -1 < 0$, alors c'est l'unique solution du problème $\max\limits_{\lambda \geq 0} G(\lambda)$. D'où

$$
\begin{aligned}
x^* &= -\lambda^*/2 - 1 = -1/4 - 1 = -5/4 \\
y^* &= -\lambda^*/2 - 1/2 = -1/4 - 1/2 = -3/4
\end{aligned}
$$

- D'autre part le problème considéré est convexe car $\nabla^2 f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ est définie positive (en particulier elle est semi-définie positive) et $g(x,y)$ est linéaire (donc g est convexe et il existe $(x^0, y^0) \in \mathbb{R}^2$ tel que $g\left(x^0, y^0\right) < 0$).
- Donc $(x^*, y^*) = (-5/4,\ -3/4)$ est une solution du problème en plus c'est l'unique solution globale.

Exemple 3.6 : Programmation quadratique Soit le problème suivant

$$
(\mathrm{P}) \begin{cases} \min\ f(x) \\ Bx \leqslant c \end{cases}
$$

où $f(x) = \frac{1}{2}(Ax,\ x) + (b,\ x)$ avec $x \in \mathbb{R}^n$, A une matrice d'ordre n symétrique, B est une matrice de type (m,n) avec $m \leqslant n$ et $c \in \mathbb{R}^m$.

- Si A est semi-définie positive alors f est convexe.
- Et comme $g(x) = Bx - c$ est linéaire, alors g est convexe.
- Il suffit alors qu'il existe un point $\xi^0 \in \mathbb{R}^n$ tel que $B\xi^0 < c$, pour que les conditions du point-selle soient des conditions nécessaires et suffisantes d'optimalité.
- Soit $\mathscr{L}(x,\lambda)$ le Lagrangien associé au problème (P), i.e,

$$
\begin{aligned}
\mathscr{L}(x,\lambda) &= \frac{1}{2}(Ax,x)_n + (b,x)_n + (Bx - c, \lambda)_m \\
&= \frac{1}{2}(Ax,x)_n + (b,x)_n + \left(B^\top \lambda, x\right)_n - (c,\lambda)_m
\end{aligned}
$$

- Alors les conditions du point-selle sont :

$$\begin{cases} \nabla_x \mathscr{L}(x^*, \lambda^*) = 0 \\ \qquad\qquad Bx^* \leqslant c \\ \qquad \lambda_i^* g_i(x^*) = 0 \;\Leftrightarrow\; \lambda_i^* \left(\sum_{j=1}^{n} b_{ij} x_j^* - c_i \right) = 0, \; 1 \leqslant i \leqslant m \\ \qquad\quad \text{et } \lambda^* \geqslant 0, \quad \text{où on a posé } \; B = (b_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \end{cases}$$

- Donc

$$\begin{cases} Ax^* + b + B^\top \lambda^* = 0 \\ Bx^* \leqslant c \\ \lambda^* \geqslant 0 \\ (Bx^* - c, \lambda^*)_m = 0 \end{cases}$$

- La dernière équation, $(Bx^* - c, \lambda^*)_m = 0$, est équivalente à $\lambda_i^* g_i(x^*) = 0, \; 1 \leq i \leq m$, car

$$(Bx^* - c, \lambda^*)_m = \sum_{i=1}^{m} \lambda_i^* g_i(x^*), \; \text{où } \; g_i(x^*) \leq 0 \; \text{ et } \; \lambda_i^* \geq 0.$$

Conclusion : Si A est semi-définie positive et si le point x^* est régulier, alors x^* est un minimum de (P) si et seulement si $\exists\, \lambda^* \in \mathbb{R}^m$, $\lambda^* \geqslant 0$, t.q.

$$\begin{cases} Ax^* + b + B^\top \lambda^* = 0 \\ Bx^* \leq c \\ (Bx^* - c, \lambda^*)_m = 0 \end{cases}$$

- Ecrivons maintenant le problème dual :

$$G(\lambda) = \min_{x \in \mathbb{R}^n} \mathscr{L}(x, \lambda).$$

- Comme A est semi-définie positive alors une conclusion nécessaire et suffisante est

$$\nabla_x \mathscr{L}(u, \lambda) = 0, \; i.e. \; Ax + b + B^\top \lambda = 0.$$

- De plus si A est définie positive, alors il y'a une solution unique pour chaque λ donné et qui est

$$\boxed{x_\lambda = -A^{-1}\left(B^\top \lambda + b\right)}$$

- Alors, la fonction duale est définie comme suit

$$\begin{aligned} G(\lambda) &= \frac{1}{2}(Ax_\lambda, x_\lambda)_n + (b, x_\lambda)_n + \left(B^\top\lambda, x_\lambda\right)_n - (c,\lambda)_m \\ &= -\frac{1}{2}\left(BA^{-1}B^\top\lambda,\lambda\right)_m - \left(BA^{-1}b + c,\lambda\right)_m - \frac{1}{2}\left(b, A^{-1}b\right)_n, \end{aligned}$$

(voir des calculs similaires effectués précédemment). $G(\lambda)$ existe pour tout $\lambda \in \mathbb{R}^m$, par conséquent

$$D = \{\lambda \mid G(\lambda) \text{ existe}, \lambda \geqslant 0\} = \{\lambda;\ \lambda \geqslant 0\}.$$

- Donc, on a à résoudre le problème dual suivant pour déterminer λ^* :

$$\boxed{\max_{\lambda\geqslant 0} G(\lambda)}$$

- D'autre part

$$\begin{aligned} \max_{\lambda\geqslant 0}(G(\lambda)) &= -\min_{\lambda\geqslant 0}(-G(\lambda)) \\ &= -\min\left(\frac{1}{2}\left(BA^{-1}B^\top\lambda,\lambda\right)_m + \left(BA^{-1}b + c,\lambda\right)_m\right) \\ &\quad - \frac{1}{2}\left(b, A^{-1}b\right)_m \end{aligned}$$

- Alors, tout revient à résoudre le problème (D)

$$\text{(D)} \quad \min_{\lambda\geqslant 0}\left(\frac{1}{2}\left(BA^{-1}B^\top\lambda,\lambda\right)_m + \left(BA^{-1}b + c,\lambda\right)_m\right)$$

- Une fois λ^* obtenu, alors on a

$$\boxed{x^* = -A^{-1}\left(B^T\lambda^* + b\right)}$$

Le problime (D) est plus simple à résoudre que le problème initial (P) car la contrainte pour le problème (D) est seulement $\lambda \geqslant 0$.

Exemple 3.7 : On considère le problème suivant :

$$\begin{cases} \min\ (x^2 + 2y^2) \\ 1 - x - y \leq 0 \end{cases}$$

- La fonction $f(x,y)$, la matrice B, le vecteur b et la constante c sont donnés par

$$f(x,y) = x^2 + 2y^2 \Rightarrow A = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

$$b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, B = (-1,-1) \text{ et } c = -1$$

- Alors, on a à résoudre le problème

$$\min_{\lambda \geqslant 0} \left(1/2\left(BA^{-1}B^{\top}\lambda, \lambda\right) + \left(BA^{-1}b + c, \lambda\right)\right)$$

$$A^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix}$$

$$\begin{aligned} BA^{-1}B^{\top} &= (-1,-1)\begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \end{pmatrix} \\ &= (-1,-1)\begin{pmatrix} -1/2 \\ -1/4 \end{pmatrix} = 1/2 + 1/4 = 3/4 \end{aligned}$$

- Plus simplement, il nous faut résoudre

$$\min_{\lambda \geqslant 0}\left(\frac{3}{8}\lambda^2 - \lambda\right) = \min_{\lambda \geqslant 0} G(\lambda)$$

$$G'(\lambda) = 3/4\lambda - 1 = 0 \;\Rightarrow\; \lambda = 4/3$$

- Comme $\lambda = \frac{4}{3} > 0$ et $G''(\lambda) = \frac{3}{4} > 0$, alors $\lambda^* = \frac{4}{3}$ est l'unique solution de

$$\min_{\lambda \geqslant 0} G(\lambda)$$

- Donc

$$\begin{aligned} \begin{pmatrix} x^* \\ y^* \end{pmatrix} &= -A^{-1}\left(B^{\top}\lambda^* + b\right) \\ &= -\begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} 4/3 \begin{pmatrix} -1 \\ -1 \end{pmatrix} = 4/3\begin{pmatrix} 1/2 \\ 1/4 \end{pmatrix}, \text{ i.e. } x^* = 2/3 \text{ et } y^* = 1/3 \end{aligned}$$

3.3 Méthodes numériques

Nous nous intéressons, encore une fois, au problème

$$(P) \begin{cases} \min_{x \in \mathbb{R}^n} f(x) \\ g_i(x) \leq 0, \; i = 1, \ldots, m. \end{cases}$$

- ### 3.3.1 - Méthode de linéarisation

✓ Comme dans le cas des contraintes égalités, on linéarise les contraintes au voisinage du point x^k et on minimise la fonction

$$\left(\nabla f\left(x^k\right), x - x^k\right) + \frac{1}{2\gamma}\left\|x - x^k\right\|^2,$$

sous les contraintes linéarisées, pour obtenir le point x^{k+1}.

La linéarisation du problème (P) revient alors à résoudre le problème quadratique :

$$\begin{cases} \min\left[\left(\nabla f\left(x^k\right), x - x^k\right) + \frac{1}{2\gamma}\left\|x - x^k\right\|^2\right] \\ \text{sous contraintes :} \\ g_i\left(x^k\right) + \left(\nabla g_i\left(x^k\right), x - x^k\right) \leq 0,\ i = 1, \ldots, m, \end{cases}$$

où $\gamma > 0$. Ainsi x^{k+1} est déterminé comme la solution de ce dernier problème.

- ### 3.3.2 - Méthodes Duales

Algorithme d'Uzawa L'algorithme d'Uzawa utilise une méthode de gradient pour résoudre le problème dual.

Algorithme : L'algorithme d'Uzawa de résolution du problème (P) s'écrit ici :

1. Étant donné $\lambda^0 \geq 0$.
2. Déterminer x^k tel que :

$$\mathscr{L}\left(x^k, \lambda^k\right) = \min_{x \in \mathbb{R}^n} \mathscr{L}\left(x, \lambda^k\right),\ \text{où } \mathscr{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

3. Calculer λ^{k+1} par :

$$\lambda^{k+1} = \operatorname{Proj}_{\mathbb{R}_+^m}\left(\lambda^k + \gamma g\left(x^k\right)\right),\ \gamma > 0.$$

$$\left(\operatorname{Proj}_{\mathbb{R}_+^m}(\lambda)\right)_i = \max\left\{\lambda_i, 0\right\},\ 1 \leq i \leq m.$$

Exemple 3.8 :

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} \dfrac{1}{2}(Ax,x)+(b,x) \\ Bx \leqslant c \end{cases}$$

où $b \in \mathbb{R}^n$, A est une matrice (n,n) définie positive, B une matrice (m,n) et $c \in \mathbb{R}^m$. Le calcul des différents points de l'algorithme donne :

1. λ^0 donné, $\lambda^0 \geqslant 0$,
2. $Ax^k + b + B^T\lambda^k = 0$,
3. $\lambda^{k+1} = \max\left\{\lambda^k + \gamma\left(Bx^k - c\right), 0\right\}$.

On montre que cette méthode converge lorsque

$$0 < \gamma < 2\,\frac{\lambda_{\min}}{\|B\|^2}.$$

Exemple 3.9 : On considère le problème suivant :

$$\begin{cases} \min\limits_{(x,y)\in\mathbb{R}^2} 5(x+5)^2 + y^2 \\ x + y \leqslant 1 \end{cases} \tag{36}$$

- On définit la fonction $f(x,y) = \frac{1}{2}(AX,X) + (b,X) + C$ où $X = \begin{pmatrix} x \\ y \end{pmatrix}$ et

$$A = \begin{pmatrix} 10 & 0 \\ 0 & 2 \end{pmatrix},\ b = \begin{pmatrix} 50 \\ 0 \end{pmatrix},\ C = 125,$$
$$B = (1,1),\ c = 1 \text{ et } \lambda_{\min} = 2$$

- On effectue les calculs suivants :

$$\|B\| = \sup_{\xi\in\mathbb{R}^2, \xi\neq 0} \frac{\|B\xi\|}{\|\xi\|}$$
$$B\xi = (1,1)\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \xi_1 + \xi_2 \ \Rightarrow\ \|B\xi\| = |\xi_1 + \xi_2|$$
$$\frac{\|B\xi\|}{\|\xi\|} = \frac{|\xi_1+\xi_2|}{\sqrt{\xi_1^2+\xi_2^2}} \leqslant \sqrt{2} \ \Rightarrow\ \sup_{\zeta\in\mathbb{R}^2, \xi\neq 0} \frac{\|B\xi\|}{\|\xi\|} \leqslant \sqrt{2}$$

- Mais pour le point $(1,1) \neq (0,0)$, on constate que

$$\frac{\left\|B\begin{pmatrix}1\\1\end{pmatrix}\right\|}{\left\|\begin{pmatrix}1\\1\end{pmatrix}\right\|}=\sqrt{2},\quad \text{alors, on obtient } \|B\|=\sqrt{2}.$$

- Alors, on doit prendre $0<\gamma<2\cdot\frac{2}{2}=2$ pour s'assurer que l'algorithme converge. Prenons alors $\gamma=1$ et démarrons pour la $1^{\text{ère}}$ itération avec la valeur $\lambda^0=1$:

$$\begin{aligned}&AX^0+b+B^\top\lambda^0=0\\&\begin{pmatrix}10&0\\0&2\end{pmatrix}\begin{pmatrix}x_0\\y_0\end{pmatrix}=\begin{pmatrix}-50\\0\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}-49\\1\end{pmatrix}\\&\begin{cases}10x_0=-49\\2y_0=1\end{cases}\Rightarrow\begin{cases}x_0=-4.9\\y_0=0.5\end{cases}\end{aligned}$$

- Or $\lambda^1=\max\left\{\lambda^0+\gamma\left(Bx^0-c\right),0\right\}$, où

$$\lambda^0+\gamma\left(Bx^0-c\right)=1+(-4.9+0.5-1)=1-5.4<0.$$

- Alors $\lambda^1=0$ et donc pour la $2^{\text{ième}}$ itération, on effectue :

$$\begin{aligned}&AX^1=-b-B^\top\lambda^1=-b\\&\begin{pmatrix}10&0\\0&2\end{pmatrix}\begin{pmatrix}x_1\\y_1\end{pmatrix}=\begin{pmatrix}-50\\0\end{pmatrix}\Rightarrow\begin{cases}x_1=-5\\y_1=0\end{cases}\end{aligned}$$

qui est bien la solution recherchée du problème (36).

Méthode d'Arrow-Hurwicz Il s'agit d'utiliser la méthode du gradient pour les deux problèmes

$$\min_{x\in\mathbb{R}^n}\mathscr{L}\left(x,\lambda^k\right)\quad\text{et}\quad\max_{\lambda\geq 0}\mathscr{L}\left(x^k,\lambda\right)$$

Algorithme : L'algorithme d'Arrow-Hurwicz de résolution du problème (P) s'écrit :

1. Étant donné (x^0,λ^0), $\lambda^0\geq 0$.
2. Déterminer (x^{k+1},λ^{k+1}) par :

$$\begin{aligned}x^{k+1}&=x^k-\gamma\left(\nabla f\left(x^k\right)+\sum_{i=1}^m\lambda_i^k\nabla g_i\left(x^k\right)\right)\\\lambda^{k+1}&=\underset{\mathbb{R}^m_+}{\mathsf{Proj}}\left(\lambda^k+\gamma g\left(x^k\right)\right),\ \gamma>0\end{aligned}$$

- ### 3.3.3 - Extension de la méthode de Newton : Méthode de Wilson

✓ Reprenons la méthode de Newton dans le cas des contraintes égalités, c'est-à-dire,

$$\begin{cases} \nabla f\left(x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^k\nabla g_i\left(x^k\right)+\left(\nabla^2 f\left(x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^k\nabla^2 g_i\left(x^k\right)\right)\left(x^{k+1}-x^k\right) \\ \qquad\qquad +\sum\limits_{i=1}^{m}\left(\lambda_i^{k+1}-\lambda_i^k\right)\nabla g_i\left(x^k\right)=0 \\ g_i\left(x^k\right)+\left(\nabla g_i\left(x^k\right),\left(x^{k+1}-x^k\right)\right)=0,\ i=1,\cdots,m \end{cases}$$

✓ Ce dermier système s'écrit aussi

$$\begin{cases} \nabla f\left(x^k\right)+\left(\nabla^2 f\left(x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^k\nabla^2 g_i\left(x^k\right)\right)\left(x^{k+1}-x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^{k+1}\nabla g_i\left(x^k\right)=0 \\ g_i\left(x^k\right)+\left(\nabla g_i\left(x^k\right),\left(x^{k+1}-x^k\right)\right)=0 \end{cases}$$

✓ On observe alors que x^{k+1} est solution du problème d'optimisation quadratique :

$$(Q)\begin{cases} \min\left\{\frac{1}{2}\left(\left(\nabla^2 f\left(x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^k\nabla^2 g_i\left(x^k\right)\right)\left(x-x^k\right),\ x-x^k\right)+\left(\nabla f\left(x^k\right),\ x-x^k\right)\right\} \\ \text{Sous les contraintes :} \\ \left(\nabla g_i\left(x^k\right),\ x-x^k\right)+g_i\left(x^k\right)=0,\ i=1,...,m, \end{cases}$$

et que λ^{k+1} n'est autre que le vecteur dual optimal de ce programme quadratique. Au lieu d'utiliser la formule de Newton classique, on peut donc résoudre le problème (Q) à chaque itération pour obtenir x^{k+1} et λ^{k+1} à partir de x^k et λ^k.

✓ Cette extension de la méthode de Newton est due à Wilson. La méthode de Wilson présente l'avantage de se généraliser facilement au cas des contraintes d'inégalités.

✓ Si au lieu de contraintes d'égalités, on a des contraintes d'inigalités du type $g_i(x)\leq 0$, il est facile de montrer qu'il suffit de considérer dans (Q) des contraintes du type :

$$\left(\nabla g_i\left(x^k\right),\ x-x^k\right)+g_i\left(x^k\right)\leq 0.$$

On a alors à résoudre à chaque étape, le programme quadratique : **Méthode de Wilson**

$$\begin{cases} \min\left\{\frac{1}{2}\left(\left(\nabla^2 f\left(x^k\right)+\sum\limits_{i=1}^{m}\lambda_i^k\nabla^2 g_i\left(x^k\right)\right)\left(x-x^k\right),\ x-x^k\right)+\left(\nabla f\left(x^k\right),\ x-x^k\right)\right\} \\ \text{Sous les contraintes} \\ \left(\nabla g_i\left(x^k\right),\ x-x^k\right)+g_i\left(x^k\right)\leq 0,\ i=1,...,m. \end{cases}$$

- ### 3.3.4 - Méthode de pénalisation

✓ La fonction de pénalisation $\varphi_r(x)$ associée au problème (P) est définie par

$$\varphi_r(x) = f(x) + \frac{r}{2}\sum_{i=1}^{m} (g_i(x))^2 u_i(g_i), \quad \text{où} \quad u_i(g_i) = \begin{cases} 0 & \text{si } g_i(x) \leqslant 0 \\ 1 & \text{si } g_i(x) > 0 \end{cases}$$

✓ On cherche ici à résoudre le problème d'optimisation sans contraintes suivant :

$$\min_{x\in\mathbb{R}^n} \varphi_r(x)$$

✓ Pour voir le lien avec le Lagrangien associé au problème (P)

$$\mathscr{L}(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x),$$

on pose $\lambda_i = \frac{r}{2} g_i(x) u_i(g_i)$. Alors, on a $\varphi_r(x) = \mathscr{L}(x,\lambda)$ et on montre (sous certaines conditions) que

$$\lambda_i^* = \lim_{r\to+\infty} r g_i(x) u_i(g_i).$$

Exemple 3.10 : Reprenons l'exemple (déjà traité) suivant :

$$\begin{cases} \min (x^2 + 2y^2) \\ 1 - x - y \leq 0 \end{cases}$$

- Ici, on définit $\varphi_r(x,y)$ comme suit

$$\varphi_r(x,y) = x^2 + 2y^2 + \frac{r}{2}(1-x-y)^2\delta, \quad \text{où} \quad \delta = \begin{cases} 0 & \text{si } 1-x-y \leq 0 \\ 1 & \text{si } 1-x-y > 0 \end{cases}$$

- Si $\delta = 0$, i.e, $1 - x - y \leq 0$, alors on obtient

$$\begin{cases} \frac{\partial\varphi_r}{\partial x} = 2x = 0 \\ \frac{\partial\varphi_r}{\partial y} = 4y = 0 \end{cases} \Rightarrow x = y = 0, \quad \text{ce qui est impossible}$$

- Donc $\delta = 1$, et on obtient alors

$$\begin{cases} \frac{\partial\varphi_r}{\partial x} = 2x - r(1-x-y) = 0 \\ \frac{\partial\varphi_r}{\partial y} = 4y - r(1-x-y) = 0 \end{cases} \Rightarrow x = 2y$$

  Par suite, on a

$$\begin{cases} x = \frac{2r}{4+3r} \\ y = \frac{r}{4+3r} \end{cases}$$

  et lorsque $r \to +\infty$ on obtient $x^* = 2/3$ et $y^* = 1/3$.

- D'autre part, on a

$$\begin{aligned}\lambda^* &= \lim_{r\to+\infty} r(1-x-y)\delta \\ &= \lim_{r\to+\infty} r\left(1-\frac{3r}{4+3r}\right) \\ &= \lim_{r\to\infty} \frac{4r}{4+3r} = 4/3\end{aligned}$$

Bibliographie :

[1] **V. Karmanov. Programmation mathématique. Editions Mir, Moscou, 1977.**

[2] **M. Minoux. Programmation mathématique. Théorie et algorithmes. Lavoisier, Paris, 2008.**

[3] **B. T. Polyak. Introduction to Optimization. Optimization Software, New York, 1987.**