Chapitre3. Optimisation avec contraintes - méthodes globales

III.1 Méthode du gradient projeté

a- Principe

La méthode du gradient projeté consiste à projeter sur le domaine *C* (convexe, ferme, non vide) les points obtenus à chaque itération par la méthode du gradient a pas fixe. C'est-à-dire, soit $\rho > 0$ un pas de descente tel que :

$$x_{k+1} = P_C(x_k - \rho\nabla f(x_k))$$

Sa convergence est assurée sous les mêmes hypothèses que pour la méthode du gradient à pas fixe par le théorème suivant :

Soient *C* un sous-ensemble convexe fermé non vide de $\mathbb{R}^n$, $f: C \to \mathbb{R}$ une application α-elliptique
On suppose de plus que $\nabla f: \mathbb{R}^n \to \mathbb{R}^n$ *est M-lipschitzienne (où* α *et* M *sont respectivement les constantes d'ellipticité, c'est-à-dire,* $\exists M > 0, \forall (x, y) \in : \mathbb{R}^n, \|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$*)*
Si le pas de descente ρ est choisi dans un intervalle $[\beta_1, \beta_2]$ tel que :

$$0 < \rho < \frac{2\alpha}{M^2} \quad \text{Où} \quad 0 < \beta_1 < \beta_2 < \frac{2\alpha}{M^2}$$

Alors la méthode du gradient projeté converge géométriquement vers le minimum de la fonction *f* sur le domaine *C*
Soit (x_k) la suite de la méthode de gradient projeté définie par :

$$\begin{cases} x_0 \in C \quad donné, \\ x_k = P_C\big(x_k - \rho_k \nabla f(x_k)\big) \; \forall k \in \mathbb{N} \\ P_C(x_k) = arg \min\limits_{y \in C} \|x - y\|^2 \end{cases}$$

De plus,
1. Si $\rho_k = \alpha > 0$ pour tout $k \in \mathbb{N}$, on dit que le système est gradient projeté à pas fixe.
2. Si ρ_k est le point de minimum global de la fonction *f* définie par : $f(\rho_k) = (x_k - \rho_k\nabla f(x_k))$, on dit que le système est gradient projeté à pas optimale.

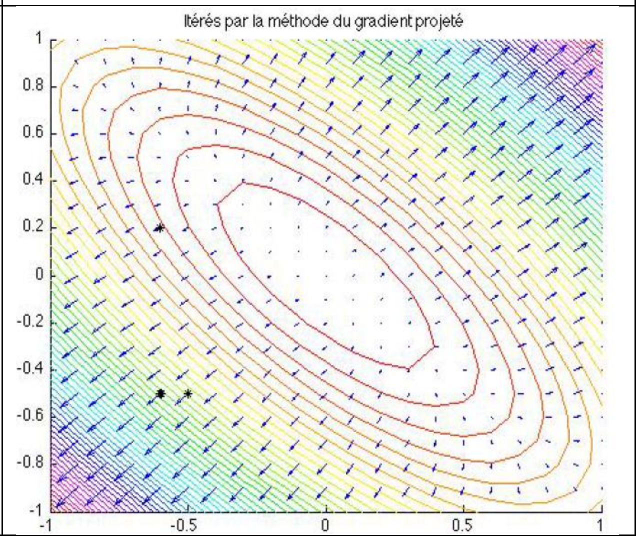
b- Algorithme du gradient projeté

Recherché du gradient projeté
1- **Initialisation** : $x_0 \in \mathbb{R}^n$, $k = 0$ $\tilde{x}_1 = (x_0 - \rho_0\nabla f(x_0)$ Où ρ_0 éventuellement calculé par une méthode de recherche linéaire $x_1 = P_C\,(\tilde{x}_1)$ 2- **Tant que** $\|x_{k+1} - x_k\| < \varepsilon$ (tolérance) $\tilde{x}_{k+1} = x_k - \rho_k\nabla f(k)$ Où ρ_k éventuellement calculé par une méthode de recherche linéaire $x_{k+1} = P_C\,(\tilde{x}_{k+1})$ $k = k + 1$ Fin

Exemple
On cherche à résoudre numériquement le problème : $f: C \to \mathbb{R}$ différentiable, à l'aide d'une méthode de gradient projeté à pas constant.

$$\inf_{x\in C} f(x,y) = \inf_{x\in C}(2x^2 + 3xy + 2y^2)$$

Et $\in \mathbb{R}^n$, un ensemble de contraintes. $C = \left\{x \le -\frac{1}{2},\ y \le -\frac{1}{2}\right\}$
Le pas $\rho = 10^{-4}$. Le gradient converge et $l'erreur\ < 10^{-8}$ en 8 itérations
Suite des itérations (sur la figure) obtenus par Matlab :

x	y	
-0.6000	0.2000	
-0.5998	-0.5000	
-0.5994	-0.5000	
-0.5989	-0.5000	
-0.5985	-0.5000	
-0.5980	-0.5000	
-0.5974	-0.5000	
-0.5000	-0.5000	
-0.5000	-0.5000	

Ainsi la méthode du gradient conjugué permet en théorie de déterminer le minimum d'une fonction elliptique _a dérivée lipchitzienne sur une convexe ferme quelconque. C'est cependant illusoire : on ne sait pas en général construire l'operateur de projection sur un convexe. Les seuls exemples notables étant lorsque $C = \prod_{i=1}^{n}[a_i, b_i]$ est un produit d'intervalles, ou lorsque C'est une boule fermée $C = \overline{B(x_0, r)}$. Aussi emploie-t-on plutôt, du moins lorsque les contraintes sont affinées, la méthode d'Uzawa, que nous allons voir, qui met à profit la notion de dualité et résout le problème dual, où l'operateur de projection est alors ne peut plus simple, les contraintes n'étant plus alors que des contraintes de signe.

c- Méthode des directions réalisables du gradient projeté

Considérons le même problème $\mathcal{P}$ que dans la méthode des directions réalisables où la fonction $f \in C^1 \subset \mathbb{R}$ et où est défini à l'aide d'un ensemble de contraintes linéaires:

$$(\mathcal{P}) \begin{cases} Minf(x) \\ sujet \text{ à } a_i^T x_k < b_i \ \text{i} = 1, \dots, \text{m} \end{cases}$$

Etant donné une solution réalisable x_k de $(\mathcal{P})$ définissons l'ensemble des indices des contraintes actives

$$E_{ck} = \{i : a_i^T x_k = b_i\}$$

La direction de descente est obtenue en projetant le vecteur $[-\nabla f(x_k)]$ sur le sous espace

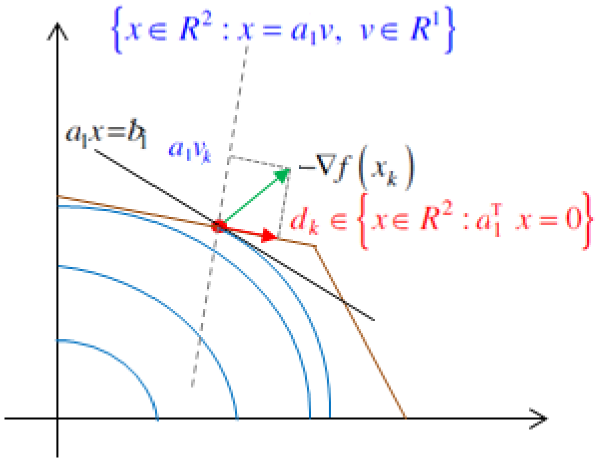
$$\{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\}$$

Pour dériver une formulation explicite de d_k, on note A_k la matrice $|E_{ck}| \times n$ dont les lignes sont les vecteurs a_i^T dont les indices $i \in E_{ck}$. De plus on suppose que les vecteurs a_i^T dont les indices $i \in E_{ck}$ sont linéairement indépendants de sorte que la matrice A_k soit de rang $|E_{ck}| < n$.
Ainsi, le sous espace $\{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\}$ est orthogonal au sous espace $\left\{x \in \mathbb{R} : x = A_k^T v, i \in \mathbb{R}^{|E_{ck}|}\right\}$ et de plus

$$\mathbb{R}^n = \{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\} + \left\{x \in \mathbb{R} : x = A_k^T v, i \in \mathbb{R}^{|E_{ck}|}\right\}$$

Puisque d_k est la projection de $[-\nabla f(x_k)]$ sur $\{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\}$ alors $d_k \in \{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\}$ et de plus il existe $v_k \in \mathbb{R}^{|E_{ck}|}$ tel que

$$[-\nabla f(x_k)] = d_k + A_k^T v_k$$



Ainsi
$$-A_k \nabla f(x_k) = A_k d_k + A_k A_k^T v_k$$
$$-A_k \nabla f(x_k) = A_k A_k^T v_k \quad \text{puisque} \quad A_k d_k = 0$$

Ceci implique que

$$v_k = -(A_k A_k^T)^{-1} A_k \nabla f(x_k)$$

Et par conséquent

$$d_k = -\nabla f(x_k) - A_k^T v_k = -[I - A_k^T (A_k A_k^T)^{-1} A_k] \nabla f(x_k)$$

(On note que la matrice $A_k A_k^T$ de dimension $|E_{ck}| \times |E_{ck}|$ est non singulier puisque la matrice A_k de dimension $|E_{ck}| \times n$ est plein de rang par hypothèse).

$$\mathbb{R}^n = \{x \in \mathbb{R} : a_i^T x_k = 0, i \in E_{ck}\} + \left\{x \in \mathbb{R} : x = A_k^T v, i \in \mathbb{R}^{|E_{ck}|}\right\}$$

$$d_k = -\nabla f(x_k) - A_k^T v_k = -[I - A_k^T (A_k A_k^T)^{-1} A_k] \nabla f(x_k)$$

Si $d_k = 0$ et $v_k \geq 0$ alors :

$$A_k x_k \leq 0$$
$$v_k^T A_k x_k = 0 \text{ puisque } A_k x_k \leq 0$$
$$0^T A_k x_k = 0$$
$$v_k \geq 0$$

Ainsi les conditions KKT sont vérifiées à x_k
Si $d_k \neq 0$

$$\nabla f(x_k)^T d_k < 0 \ \text{(Direction de descente)}$$
$$A_i^T d_k < 0 \ \text{(Direction réalisable)}$$

Puisque $a_i^T x_k < b_i$ lorsque $i \notin E_{ck}$, ils'ensuit que
Si $A_i^T d_k \leq 0$, alors $a_i^T x_k + \rho_k a_i^T d_k < b_i$ pour toute valeu $\rho_k \geq 0$;
Si $A_i^T d_k > 0$, alors $a_i^T x_k + \rho_k a_i^T d_k \leq b_i \Leftrightarrow \rho_k a_i^T d_k \leq b_i - a_i^T x_k$

$$\Leftrightarrow \rho_k \leq \frac{b_i - a_i^T x_k}{a_i^T d_k}$$

Par conséquent :

$$\bar{\rho}_k = \min_{i \notin E_{ck}} \left\{ \frac{b_i - a_i^T x_k}{a_i^T d_k} : a_i^T d_k > 0 \right\}$$

Le critère d'arrêt est satisfait lorsque $\nabla f(x_k)^T d_k \geq 0$ pour un certain indice k, sinon la méthode s'arrête à la limite quand $k \to \infty$.

d- Algorithme du gradient projeté, Direction réalisable

Recherché du gradient projeté, Direction réalisable

1- Initialisation : $x_0 \in \mathbb{R}^n$,
$k = 0$
Le point courant : $x_k \in C$
Soit A_0 la matrice dont les lignes correspondent aux contraintes saturées
Calculer la matrice de saturation
$P_0 = I - A_0^T (A_0 A_0^T)^{-1} A_0$
$d_k = -P_0 \nabla f(x_k)$
$v_0 = -(A_0 A_0^T)^{-1} A_0 \nabla f(x_k)$
$v_i = \min_{j=1,\dots,n} \{v_j\}$
Si ($d_k = 0$ et $v_i < 0$) Alors
Calculer A'_0, P'_0, d'_k
$d'_k = d_k$
    Sinon ($d_k = 0$ et $v_i \geq 0$) Alors
     Le point x_k satisfait les conditions necessires de KKT
     Fin des itérations
   Fin Si
Fin Si
Determiner $\alpha_{max} = Max\{\alpha / x_k + \alpha d_k \in C\}$ puis x_{k+1} tel que :
$f(x_{k+1}) = \min_{0<\alpha<\alpha_{max}} \{f(x_k + \alpha d_k)\}$
k=k+1
Tant que (condition d'arret non vérifiée)
Fin

III.2 Méthode de Lagrange-Newton pour des contraintes égalité / inégalité

a- Principe

La méthode de Lagrange-Newton pour des contraintes inégalité consiste à introduire dans la méthode de Newton un multiplicateur de Lagrange associé à la contrainte $Ax - b = 0$.
On va faire appel au cas particulier d'un problème quadratique avec contraintes en égalité affines.

$$\begin{cases} min\left(\frac{1}{2}x^TQx - c^Tx\right) \\ Ax = b \end{cases}$$

Où Q est une matrice carrée de $\mathcal{M}_{n\times n}$ ($\mathbb{R}$) et c un vecteur de $\mathbb{R}^n$, $A_{n\times n}$ $\in$ ($\mathbb{R}$) , $b \in \mathbb{R}^p$. Si on écrit les relations de ***Kuhn-Tucker*** (conditions d'optimalité au 1[ier] ordre), on a:

$$\begin{cases} \nabla_x\left(\frac{1}{2}x^TQx - c^Tx + \mu^T(Ax-b)\right) = 0 \\ Ax = b \end{cases}$$

Où $\mu \in \mathbb{R}^p$ est le multiplicateur de Lagrange associé à la contrainte $Ax - b = 0$. Par conséquent, le couple optimal $(x^*,\ \mu^*)$ est la solution du système

$$\begin{cases} Qx^* - c + A^T\mu^* = 0 \\ Ax^* = b \end{cases}$$

Que l'on peut aussi écrire:

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ \mu^* \end{bmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}$$

Si la matrice $\mathcal{M} = \begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}$ est inversible, ce système admet une solution que l'on peut calculer par n'importe quelle m´méthode pour la résolution des systèmes linéaires.
Supposons maintenant que le problème ne soit plus quadratique. On va résoudre le système d'optimalité par la méthode de Newton. Considérons le problème en égalité suivant :

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

Où $f: \mathbb{R}^n \to \mathbb{R}$ $\quad h: \mathbb{R}^n \to \mathbb{R}^p$ sont suffisamment régulières. Les conditions du premier ordre s'écrivent (au moins formellement)

$$\begin{cases} \nabla_x \mathcal{L}L(x,y) = 0 \\ h(x) = 0 \end{cases}$$

Où $\mu \in \mathbb{R}^p$ et $\mathcal{L}(x,\mu) = f(x) + \mu h(x)$ est le Lagrangien du problème. On peut alors résoudre ce système d'équations non linéaires par la méthode de *Newton*.
On a alors ce que l'on appelle la méthode de *Lagrange-Newton*.

On remarque que si l'on ajoute $\nabla h(x_k)^T\mu_k$ à la première ligne de l'équation précédente, on aura alors la forme équivalente :

$$\begin{bmatrix} D^2_{xx}\mathcal{L}(x_k,\mu_k) & \nabla h(x_k)^T \\ \nabla h(x_k) & 0 \end{bmatrix}\begin{bmatrix} d_k \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} \nabla h(x_k)^T \\ h(x_k) \end{bmatrix}$$

Où $D^2_{xx}\mathcal{L}(x_k,\mu_k)$ est le Hessien par rapport à *x* de $\mathscr{L}$.

b- Algorithme du Lagrange-Newton

Recherché de Lagrange-Newton
1- **Initialisation** : $(x_{0,}\mu_0) \in \mathbb{R}^n \times \mathbb{R}^p$ 2- Itération $k = 0$ **2. Tant que** $\|x_{k+1} - (x_k + d_k)\| < \varepsilon$ (tolérance) alors Résoudre le système linéaire suivant $$\begin{bmatrix} D^2_{xx}\mathcal{L}(x_k,\mu_k) & \nabla_x h(x_k)^T \\ \nabla_x h(x_k) & 0 \end{bmatrix}\begin{bmatrix} d_k \\ y_k \end{bmatrix} = -\begin{bmatrix} \nabla_x h(x_k,\mu_k)^T \\ h(x_k) \end{bmatrix}$$ Poser $x_{k+1} = x_k + d_k$ $\mu_{k+1} = \mu_k + y_k$ $k = k + 1$ Fin

c- Lagrangien :

Considérons le problème d'optimisation suivant

$$\begin{cases} Min f(x) \\ sujet \text{ à } f_i(x) \leq 0 \quad i = 1,\dots,m \\ x \in C \end{cases}$$

Suppose que le lagrangien associe au problème d'optimisation possède un minimum global x^* sur C lorsque le vecteur de multiplicateur

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

$\lambda = \lambda^*$. Si $f_i(x^*) \leq 0$, $\lambda_i^* \geq 0$ et $\lambda_i^* f_i(x^*) = 0$ pour tout $i = 1,\dots,m$, alors x^* est une solution optimale globale du problème de minimisation.

Exemple

Min $f(x,y) = x^2 + 2y^2$

$sujet$ à $g_i(x,y) \leq 0 \quad i = 1,\dots,m$

$\mathcal{L}(\lambda,x) = x^2 + \lambda(2x + 5)$

$\mathcal{L}$ est convexe en (x,y). Donc le minimum de la fonction *f* atteint lorsque $\nabla_x \mathcal{L}(\lambda,x) = 0$

$\nabla_{x,y}\mathcal{L}(\lambda,x,y) = 2x + \lambda = 0 \iff x = -\lambda$

$2x + 5 = -2\lambda + 5 = 0 \Leftrightarrow \lambda = \dfrac{5}{2} \geq 0$

$2x + 5 = 0 \Rightarrow (2x + 5)\lambda = 0$

Donc $x = -\lambda = -\frac{5}{2}$

Conditions d'optimalité de *Karush-Kuhn-Tucker* (KKT) de premier ordre

Pour obtenir des conditions plus facilement vérifiables, il faut poser des hypothèses sur C et sur les fonctions f et f_i.
Si C est convexe, si f et f_i sont différentiables et convexes dans le d'optimisation problème

$$\begin{cases} Min f(x) \\ sujet \text{ à } f_i(x) \leq 0 \quad i = 1, \dots, m \\ x \in C \end{cases}$$

Alors le lagrangien

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

est aussi une fonction convexe sur C, $\lambda \geq 0$ puisque $\lambda_i \geq 0$ et $f_i(x)$ convexe $\Rightarrow \lambda_i f_i(x)$ convexe et par conséquent $f(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$ est convexe

Alors les conditions d'optimalité de Karush-Kuhn-Tucker (*KKT*) de premier ordre est la suivante :

S'il existe un λ^* tel que

*Karush-Kuhn-Tucker (**KKT**)*
$\nabla_x \mathcal{L}(\lambda^*, x^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0$ $\lambda_i^* f_i(x^*) = 0 \quad i = 1, \dots, n$ $f_i(x^*) \leq 0 \quad i = 1, \dots, n$ $\lambda_i^* \geq 0 \quad i = 1, \dots, n$ $x^* \in C$

Exemple

$$Min\ 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2$$

$$\text{Sujet à} \quad x_1^2 + x_2^2 \leq 5$$
$$3x_1 + x_2 \leq 6$$

$$\nabla \mathcal{L}(\lambda, x) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) = 0 \iff \begin{bmatrix} 4x_1 + 2x_2 - 10 \\ 2x_1 + 2x_2 - 10 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 0$$

$x_1^2 + x_2^2 - 5 \leq 0$
$3x_1^2 + x_2^2 - 6 \leq 0$
$\lambda_1(x_1^2 + x_2^2 - 5) = 0$
$\lambda_2(3x_1^2 + x_2^2 - 6) = 0$
$\lambda_1, \lambda_2 \geq 0$

Nous pouvons faire différentes hypothèses sur quelle contrainte est active dans le but d'identifier des valeurs de x_1, x_2, λ_1, λ_2 satisfaisant les conditions *KKT*.
Supposons que la première contrainte est active et que la seconde ne l'est pas:

$x_1^2 + x_2^2 - 5 = 0$
$3x_1^2 + x_2^2 - 6 < 0 \Rightarrow \lambda_2 = 0$
Nous retrouvons alors le système avec 3 équations et 3 inconnus suivant:

$4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 = 0$
$2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 = 0$
$x_1^2 + x_2^2 = 5$
Donc : $x_1 = 1, x_2 = 2, \lambda_1 = 1$ et $\lambda_2 = 0$

Nous pouvons vérifier que ce résultat satisfait le système précédent et par conséquent il satisfait aussi les conditions KKT.

Conditions K-K-T suffisantes

Le lagrangien étant convexe si $\lambda^* \geq 0$ (démontrer précédemment), alors par l'inégalité du gradient il s'ensuit que pour tout $x \in C$

$$\mathrm{f(x)} + \sum_{i=1}^{m} \lambda_i^* f_i(x) \geq \mathrm{f(x^*)} + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(x^*)}_{=0} + \underbrace{\left[\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*)\right]^T}_{=0} (x - x^*)$$

Alors, pour tout $x^* \in C$

$$f(x^*) - f(x) \leq \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) \leq 0$$

et

$$f(x^*) \leq f(x)$$

III.3 Méthode de Newton projeté (pour des contraintes de borne)

a- Principe

La méthode de Newton projeté relève d’une idée analogue à celle d´enveloppée lors du gradient projeté : puisque les itérés successifs ne satisfont pas les contraintes, on les projette sur l’ensemble des contraintes. On s’intéresse ici qu’au cas où on aura des contraintes de bornes de la forme :

$$(\mathcal{P}) \quad \begin{cases} min f(x) \\ a \leq x \leq b \end{cases}$$

Remarquons que beaucoup de problèmes duaux (où μ est alors l’inconnue) sont également de cette forme. On commence par un problème plus simple :

$$(\mathcal{P}) \quad \begin{cases} min f(x) \\ x \geq 0 \end{cases}$$

Où f est C^2 sur $\mathbb{R}^n$. Si H_k désigne la matrice Hessienne $[D^2 f(x)]$ une itération de la méthode de Newton est de la forme :

$$x_{k+1} = x_k - [H_k]^{-1} \nabla f(x_k)$$

On peut affiner un peu en introduisant un pas $\alpha_k > 0$ qui donne la nouvelle formulation :

$$x_{k+1} = x_k - \alpha_k [H_k]^{-1} \nabla f(x_k)$$

Si on projette sur l'ensemble des contraintes, on obtient alors :

$$x_{k+1} = (x_k - \alpha_k [H_k]^{-1}(x_k))^+$$

Où :

$$(f(x))^+ = max(f(x), 0)$$

Ceci donne alors la méthode de *Newton projetée*.

b- Algorithme de Newton projeté.

Recherché de Newton projetée
1- **Initialisation** : Poser n° Choisir $x_0 \in \mathbb{R}^n$ $x_0 \geq 0$ Choisir une tolérance $\varepsilon > 0$ 2- Itération k : $k = 0$ **2. Tant que** $\|x_{k+1} - x_k\| < \varepsilon$ alors Application de règle anti zig-zig $w_k = \|x_k - (x_k - \nabla f(x_k))^+\|$ $\varepsilon_k = \min(\varepsilon, w_k)$ $I_k^+ = \left\{ i / \ 0 \leq x_{ki} \leq \varepsilon_k \ et \ \frac{\partial f}{\partial x_i}(x_k) > 0 \right\}$ $D_k = [H_k]^{-1}$ où $H_{kij} = \begin{cases} 0 \ \ si \ i \neq j \ \ et \ (i \in I_k^+ \ \ ou \ \ j \in I_k^+) \\ \frac{\partial^2}{\partial x_i \partial x_j}(x_k) \end{cases}$ sinon $P_k = D_k \nabla f(x_k)$ (choix de direction de descente) $x_k(\alpha) = (x_k - \alpha P_k)^+$ (choix du pas par une recherche linéaire) α_k est choisi avec une règle de type d'Armijo $x_{k+1} = x_k(\alpha_k)$ Fin pour i $k = k + 1$ Fin

III.4 Méthode de pénalisation

a- Principe

Les méthodes de pénalisation sont très utilisées en pratique car elles sont très simples. Elles partent du principe suivant : on remplace le problème avec contraintes

$$(\mathcal{P}) \ \begin{cases} minf(x) \\ \ \ x \geq 0 \end{cases}$$

par un problème sans contraintes en intégrant les contraintes x ∈ C⊂$\mathbb{R}^n$ dans une fonctionnelle f
La fonction f devient d'autant plus grande que les contraintes ne sont pas respectées,

On minimise f sur Rn par les méthodes de minimisation de problèmes sans contraintes.

$$(\mathcal{P}_\varepsilon)\ \begin{cases} \min f(x) + \frac{1}{\varepsilon}\alpha(x) \\ \mathrm{x} \in \mathbb{R}^n \end{cases}$$

Où $\alpha: \mathbb{R}^{\mathrm{n}} \to \mathbb{R}$ est une fonction de pénalisation des contraintes et $\varepsilon > 0$. Le but est de trouver des fonctions
α telles que les problèmes $(\mathcal{P})$ et $\mathcal{P}_\varepsilon)$ soient équivalents, c'est-`a-dire, tels qu'ils aient les mêmes solutions.
Dans ce cas, on dit que la pénalisation est exacte. On peut, par exemple, choisir :

$$\alpha(x) = \begin{cases} 0 & si\ \ \mathrm{x} \in \mathrm{C} \\ +\infty & si\ \ \mathrm{x} \in \mathrm{C} \end{cases}$$

Si α n'est même pas continue, les algorithmes vus précédemment pour les problèmes sans contraintes ne sont pas applicables de résoudre le problème, donc on doit alors modifier la fonction.

- **Pénalisation extérieure**

Modifier la fonction à minimiser *f* pour qu'elle prenne de « grandes » valeurs en dehors du domaine admissible.
On suppose que α vérifié les propriétés suivantes :

1. La fonction α est continue sur $\mathbb{R}^n$
2. $\forall \mathrm{x} \in \mathbb{R}^n$, $\alpha(x) \geq 0$
3. $\alpha(x) = 0 \Leftrightarrow \mathrm{x} \in \mathrm{C}$

Nous donnons quelques exemples de pénalisation pour différentes contraintes :

1. Pour une contrainte $x \geq 0$, on a la fonction $\alpha(x) = \|x^-\|^2$
2. Pour une contrainte $h(x) = 0$, on a la fonction $\alpha(x) = \|h(x\|^2$
3. Pour une contrainte $g(x) \geq 0$, on a la fonction $\alpha(x) = \|g(x)^-\|^2$

Où $\|-\|$ est la norme euclidienne de $\mathbb{R}^n$ et $x^- = \max(-x, 0)$.

- **Pénalisation intérieure**

Même principe que la pénalité extérieure, mais pour obtenir un minimum approché par l'intérieur du domaine (donc toujours admissible). On définit :

Contraintes $g(x) \leq 0$:

$$\alpha(x,\varepsilon) = f(x) - \varepsilon \sum_{j=1}^{\ell} \log(-g_j(x))$$

Introduire le terme $\frac{1}{\varepsilon}\alpha(x)$ qui tend vers l'infini lorsque *x* s'approche de la frontière δC de l'ensemble des contraintes *C*.
Créer une "*barrière*" au bord de l'ensemble admissible et on s'approche du minimum du problème $(\mathcal{P})$ de "*l'intérieur*".
Le choix de la fonction de pénalisation α se fait si les deux conditions sont vérifiées

1. La fonction α est continue sur $C \subset \mathbb{R}^n$
2. $\mathrm{x} \notin \mathrm{C} \Rightarrow \alpha(x) = \infty$

a- La méthode de pénalisation dans le cas de contraintes égalités

Dans le cas de problème d'optimisation avec contraintes de type égalité, la fonction pénalité utilisée est la fonction quadratique. Le problème à résoudre est le suivant :

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

La fonction quadratique de pénalisation $q(x_k, \mu_k)$ pour $(\mathcal{P})$ est

$$\alpha(x_k, \mu_k) = f(x_k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2\,(x_k)$$

Où $\mu_k > 0$ est le paramètre de pénalisation.
En tendant μ_k vers zéro, nous pénalisons les violations des contraintes.
Le principe est donc : on considère une séquence de valeurs de $\{\mu_k\}$ avec $\mu_k \to 0$, $k \to 0$ et on calcule le minimum x_k de $\alpha(x_k, \mu_k)$ pour chaque itération k.

Exemple. On considère le problème suivant :

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

La fonction de pénalisation quadratique associée est :

$$\alpha(x, \mu) = x + y + \frac{1}{2\mu}(x^2 + y^2 - 2)^2$$

Pour $\mu = 1$, on trouve le minimum de q est de $(-1.1\,, -1.1)$
Pour $\mu = 0.1$, on trouve le minimum de q est de $(-1\,, -1)$
La solution du problème est donc $(-1\,, -1)$

b- La méthode de pénalisation dans le cas de contraintes mixtes

On considère le problème d'optimisation avec contraintes mixtes suivant :

$$(\mathcal{P})\ \begin{cases} \min f(x) \\ g_i(x) = 0,\ \ 1 \le i \le m \\ h_j(x) \ge 0,\ \ 1 \le j \le \ell \end{cases}$$

On définit la fonction q de pénalisation comme suit :

$$\alpha(x_k, \mu_k) = f(x_k) + \frac{1}{2\mu_k}\sum_{i=1}^{m} g_i^2\,(x_k) + \frac{1}{2\mu_k}\sum_{j=1}^{\ell}(h_j^-\,(x_k))^2$$

Où $h^- = \max(-h, 0)$

c- Algorithme de pénalisation.

Recherché de pénalisation
1- **Initialisation** : Choisir $x_0 \in \mathbb{R}^n$ Choisir une tolérance $\varepsilon > 0$ $\mu_0 > 0$ 2- Itération k : $k = 0$ **2. Tant que** $\lVert \nabla f(x_{k,}\mu_k)) \rVert > \varepsilon$ alors Trouver le minimum approximatif x_k de $\alpha(x_k, \mu_k)$ $$\alpha(x_k, \mu_k) = f(x_k) + \frac{1}{2\mu_k} \sum_{i=1}^{m} g_i^2\,(x_k) + \frac{1}{2\mu_k} \sum_{j=1}^{\ell} (h_j^-(x_k))^2$$ Choisir le nouveau paramètre de pénalisation μ_{k+1} Choisir le nouveau x_{k+} $k = k + 1$ Fin

III.4 Méthode de dualité : méthode d'Uzawa

a- Principe

La méthode d'Uzawa est basée sur la résolution du problème de la dualité convexe.
L'idée générale est de considérer le Lagrangien $\mathscr{L}$ au lieu de la fonction f ; ce choix est motivé (au moins) par deux raisons :
_ La fonction Lagrangienne englobe à la fois la fonction f et les contraintes g et h et représente bien le problème.
_ Ensuite, On sait qu'un minimum global x^* est associe à un multiplicateur de Lagrange $\mu^* \geq 0$ tel que le couple (x^*, μ^*) est un point-selle de la fonction de Lagrange $\mathcal{L}(x, \mu) = f(x) + \mu^T\, g(x)$: Ceci légitime le fait d'envisager une approche duale du problème.
On peut également considérer le cas de contraintes d'égalités $h(x) = 0$ avec h affine de façon à conserver la convexité de *C* et la convexité de la fonction de Lagrange par rapport à la variable *x*:

Notre problème à considérer est :

$$(\mathcal{P}) \begin{cases} \min f(x) \\ g_i(x) = 0, \;\; 1 \leq i \leq m \\ h_j(x) \geq 0, \;\; 1 \leq j \leq \ell \end{cases}$$

Le lagrangien du problème est :

$$\mathcal{L}(x,\mu,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{\ell} \mu_j h_j(x)$$

Pour mieux comprendre la théorie de la dualité, on introduit la notion de point-selle (ou point col) du Lagrangien.

Définition : point-selle du Lagrangien
On appelle point-selle de $\mathscr{L}$ sur $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ tout triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_+^m$ vérifiant :

$$\forall (x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_+^m, \mathcal{L}(x^*, \mu, \lambda) \leq \mathcal{L}(x^*, \mu^*, \lambda^*) \leq \mathcal{L}(x, \mu^*, \lambda^*)$$

L'algorithme d'Uzawa se base sur ces théorèmes. Il consiste à chercher un triplet $(x^*, \mu^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}_+^m$ vérifiant les conditions de ***Karush-Kuhn-Tucker*** *KKT* de la façon suivante :

1. Pour (μ^*, λ^*) fixé dans $\mathbb{R}^l \times \mathbb{R}_+^m$, on va chercher le minimum sur $\mathbb{R}^n$ de la fonction définie ainsi

$$x \to \mathcal{L}(x, \mu^*, \lambda^*)$$

2. Pour x^* sur $\mathbb{R}^n$, on cherche le maximum sur $\mathbb{R}^l \times \mathbb{R}_+^m$ de la fonction définie ainsi :

$$(\mu, \lambda) \to \mathcal{L}(x^*, \mu, \lambda)$$

En faisant ces deux calculs simultanément, on obtient l'algorithme d'Uzawa suivant :

b- Algorithme de dualité : Algorithme d'Uzawa

Algorithme de dualité : Algorithme d'Uzawa
1- **Initialisation** : $k = 0$, Choisir $\mu_0 \in \mathbb{R}^l$ et Choisir $\lambda_0 \in \mathbb{R}_+^m$ Itération. **2. Tant que** le critère d'arrêt n'est pas satisfait : a- Calculer $x_k \in \mathbb{R}^n$ solution de $(\mathcal{P}_k) \begin{cases} \min \mathcal{L}(\mathrm{x}, \mu_k, \lambda_k) \\ x \in \mathbb{R}^n \end{cases}$ b- Calculer μ_{k+1} et λ_{k+1} avec $\begin{cases} \mu_{(k+1)i} = \mu_{ki} + \rho h_i(x_k), \ i = 1, \dots, l \\ \lambda_{(k+1)j} = \max(0, \lambda_{kj} + \rho g_i(x_k)), \ j = 1, \dots, m \end{cases}$ Où $\rho > 0$, est un réel fixe (par l'utilisateur).