CHAPITRE 3

MÉTHODE D'UZAWA

3.1 Algorithme d'Uzawa : contraintes d'égalités affines

Soit le problème d'optimisation avec contraintes d'égalités affines suivant :

$$\begin{cases} \min f(x) \\ Ax = b \end{cases} \tag{3.1}$$

où $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction de classe $\mathcal{C}^1$, A est une matrice rectangulaire de taille $m \times n$ avec $m < n$ et $rang(A) = m$, $b \in \mathbb{R}^m$.Dans ce cas, on note $h(x) = Ax - b$.

La fonction de Lagrange est donnée par :

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x). \tag{3.2}$$

La condition de Lagrange est donnée par :

$$\begin{cases} Df(x) + \lambda^T Dh(x) = 0, \\ Ax = b. \end{cases} \Leftrightarrow \begin{cases} \nabla f(x) + A^T \lambda = 0, \\ Ax = b. \end{cases} \tag{3.3}$$

On réécrit ces équations, pour $\rho > 0$, sous la forme :

$$\nabla f(x) + A^T \lambda = 0, \tag{3.4}$$

$$\lambda = \lambda + \rho (Ax - b). \tag{3.5}$$

Cette dernière forme suggère alors l'algorithme suivant :

Algorithme 1 (Algorithme d'Uzawa : contraintes d'égalités affines)

1: **Initialisation :** On prend un multiplicateur de départ $\lambda^0 \in \mathbb{R}^m$. On fixe un pas $\rho > 0$.
2: Puis on itère sur $k \in \mathbb{N}$:
   1. Calculer x^k tel que $\nabla f(x^k) + A^T \lambda^k = 0$.
   2. Calculer $\lambda^{k+1} = \lambda^k + \rho (Ax^k - b)$.

Pour que l'algorithme soit bien défini il faudra montrer l'existence d'un vecteur x^k solution de 1.

Théorème 3.1 (Cas de contraintes d'égalité affines). *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable et α-convexe sur l'ensemble non vide des contraintes $\mathcal{C} = \{x \in \mathbb{R}^n : Ax = b\}$, où A est une matrice rectangulaire de taille $m \times n$ avec $m < n$ et $rang(A) = m$. Alors,*

1. *Il existe un unique minimiseur x^* de f sur $\mathcal{C}$.*
2. *Pour tout $0 < \rho < 2\alpha / \|A\|^2$, pour tout λ^0 de départ, l'algorithme d'Uzawa converge c'est à dire*

$$\lim_{k \to +\infty} x^k = x^*.$$

3. *De plus, si ∇f est continue, alors $\left(\lambda^k\right)$ est convergente vers un unique λ^* solution de* (3.5).

Démonstration. 1. est classique.

2. Commençons par vérifier l'existence de x^k. On introduit la fonction de Lagrange $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ du problème (3.1) définie par :

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T (Ax - b).$$

Supposons λ^k connu. On considère l'application suivant :

$$v \in \mathbb{R}^n \longmapsto \mathcal{L}\left(v, \lambda^k\right) = f(v) + \left\langle \lambda^k, Av - b \right\rangle.$$

C'est une fonction strictement convexe de v ; en effet,
– f est strictement convexe (f est α-convexe).
– Le terme $\left\langle \lambda^k, Av - b \right\rangle$ est également une fonction affine de v et en particulier c'est une fonction convexe de v.
– De plus, λ^k étant fixé, la fonction $v \longmapsto \mathcal{L}\left(v, \lambda^k\right)$ est coercive car f est différentiable et α- convexe donc coercive.

D'après la condition nécessaire et suffisante d'optimalité du premier ordre d'un problème d'optimisation sans contraintes, la fonction $\mathcal{L}\left(\cdot, \lambda^k\right)$ possède un unique minimiseur $x^k \in \mathbb{R}^n$ caractérisé par :

$$\nabla_v \mathcal{L}\left(x^k, \lambda^k\right) = 0.$$

c'est à dire

$$\nabla f\left(x^k\right) + A^T \lambda^k = 0.$$

Cela prouve l'existence et l'unicité de x^k.
Travaillons ensuite sur la convergence des λ^k : nous avons

$$\begin{aligned}
\left\|\lambda^{k+1} - \lambda\right\|^2 &= \left\|\lambda^k + \rho\left(Ax^k - b\right) - \lambda - \rho(Ax - b)\right\|^2 \\
&= \left\|\lambda^k - \lambda + \rho A\left(x^k - x\right)\right\|^2 \\
&= \left\|\lambda^k - \lambda\right\|^2 + 2\rho\left\langle \lambda^k - \lambda, A\left(x^k - x\right)\right\rangle + \rho^2 \left\|A\left(x^k - x\right)\right\|^2.
\end{aligned}$$

Nous avons $\left\|A\left(x^k - x\right)\right\|^2 \le \|A\|^2 \left\|x^k - x\right\|^2$. D'autre part, en utilisant les relations sur les gradients et α-convexe de f, on obtient :

$$\begin{aligned}
\left\langle \lambda^k - \lambda, A\left(x^k - x\right)\right\rangle &= \left\langle A^T\left(\lambda^k - \lambda\right), x^k - x\right\rangle \\
&= -\left\langle \nabla f\left(x^k\right) - \nabla f(x), x^k - x\right\rangle \le -\alpha \left\|x^k - x\right\|^2.
\end{aligned}$$

Ainsi,

$$\left\|\lambda^{k+1} - \lambda\right\|^2 \le \left\|\lambda^k - \lambda\right\|^2 - \gamma \left\|x^k - x\right\|^2, \tag{3.6}$$

avec $\gamma := \rho\left(2\alpha - \rho\|A\|^2\right)$. En particulier si $0 < \rho < 2\alpha / \|A\|^2$, alors $\gamma > 0$.

La suite $n \longmapsto \|\lambda^n - \lambda\|^2$ est décroissante, minorée (par 0), donc convergente vers une limite notée ℓ. Ensuite on renverse l'inégalité (3.6) pour écrire

$$\gamma \left\|x^k - x\right\|^2 \leq \left\|\lambda^k - \lambda\right\|^2 - \left\|\lambda^{k+1} - \lambda\right\|^2 \xrightarrow[n \to +\infty]{} \ell - \ell = 0.$$

Cela démontre la convergence de la suite $\left(x^k\right)$ vers x, mais pas nécessairement la convergence de la suite $\left(\lambda^k\right)$.

3. Nous avons $rang\,(A) = rang\left(A^T\right) = m$. Donc, d'après le théorème du rang, nous avons :

$$dim\left(ker\left(A^T\right)\right) + rang\left(A^T\right) = m, \text{ d'où } dim\left(ker\left(A^T\right)\right) = 0 \Rightarrow A^T \text{est injective.}$$

Par conséquent, AA^T est inversible. en effet, soit $x \in \mathbb{R}^n$ tel que $AA^T x = 0$, alors on a :

$$\left\|A^T x\right\|^2 = \left\langle A^T x, A^T x\right\rangle = \left\langle x, AA^T x\right\rangle = 0, \tag{3.7}$$

donc $A^T x = 0$ et $x = 0$ puisque A^T est injective. En utilisant la relation $A^T \lambda^k = -\nabla f\left(x^k\right)$, on obtient :

$$AA^T \lambda^k = -A \nabla f\left(x^k\right),$$

donc, on a :

$$\lambda^k = -\left(AA^T\right)^{-1} A \nabla f\left(x^k\right).$$

Comme $x^k \xrightarrow[k \to +\infty]{} x$, par continuité de ∇f on obtient la convergence des λ^k vers un vecteur $\lambda \in \mathbb{R}^m$. Enfin, par passage à la limite dans la relation suivante :

$$\nabla f\left(x^k\right) + A^T \lambda^k = 0,$$

on trouve que λ satisfait (3.4).
Pour l'unicité de λ, nous avons :

$$\lambda = -\left(AA^T\right)^{-1} A \nabla f\,(x),$$

ce qui définit λ de manière unique puisque x est également défini de manière unique.

□

3.2 Algorithme d'Uzawa : contraintes d'inégalité

3.2.1 Contraintes d'inégalité affines

Soit le problème d'optimisation avec contraintes d'inégalité affines défini par :

$$\begin{cases} \min f\,(x) \\ x \in \mathcal{C} \end{cases} \tag{3.8}$$

où $\mathcal{C} = \{x \in \mathbb{R}^n : Ax \leq b\}$ avec A une matrice rectangulaire de taille $m \times n$, $m < n$, $rang(A) = m$ et $b \in \mathbb{R}^m$.

Rappelons que si x est un point de minimum local de f sur $\mathcal{C}$, et si f est différentiable en x, alors on peut écrire les conditions de KKT sous la forme suivante :

$$\nabla f\,(x) + A^T \mu = 0, \tag{3.9}$$

$$\mu \geq 0,\ Ax - b \leq 0,\ \langle \mu, Ax - b\rangle = 0. \tag{3.10}$$

Le lemme suivant permet de réécrire le deuxième jeu d'équations sur μ de manière plus compacte :

Lemme 3.1. *Soient $\Omega = \mathbb{R}_+^m$ et $\rho > 0$. Pour tout $\mu \in \Omega$, $A \in \mathcal{M}_{m,n}$ et $b \in \mathbb{R}^m$, nous avons :*

$$(\mu \geq 0,\ Ax - b \leq 0,\ \langle \mu, Ax - b \rangle = 0) \Leftrightarrow \mu = \mathcal{P}_\Omega \left[\mu + \rho \left(Ax - b\right)\right].$$

Démonstration. On procède par double implication.

1. Supposons que $\mu \geq 0$, $Ax - b \leq 0$ et $\langle \mu, Ax - b \rangle = 0$ et montrons que $\mu = \mathcal{P}_\Omega \left[\mu + \rho \left(Ax - b\right)\right]$. Comme $\mu \geq 0$, $\mu \in \Omega$; il s'agit donc de montrer que pour tout $\lambda \in \Omega$, nous avons :

$$\langle (\mu + \rho \left(Ax - b\right)) - \mu, \lambda - \mu \rangle \leq 0.$$

Or,

$$\begin{aligned} \langle (\mu + \rho \left(Ax - b\right)) - \mu, \lambda - \mu \rangle &= \langle \rho \left(Ax - b\right), \lambda - \mu \rangle \\ &= \rho \left\langle Ax - b, \lambda \right\rangle - \rho \left\langle Ax - b, \mu \right\rangle \leq 0 \end{aligned}$$

puisque $\langle Ax - b, \mu \rangle = 0$, $\rho > 0$, $Ax - b \leq 0$ et $\lambda \geq 0$. D'après Théorème 1.2, nous avons :

$$\mu = \mathcal{P}_\Omega \left[\mu + \rho \left(Ax - b\right)\right].$$

2. Réciproquement, supposons que $\mu = \mathcal{P}_\Omega \left[\mu + \rho \left(Ax - b\right)\right]$, en utilisant Théorème 1.2, on obtient $\mu \geq 0$ et

$$\langle \rho \left(Ax - b\right), \lambda - \mu \rangle \leq 0, \ \ \text{pour tout } \lambda \geq 0.$$

En prenant $\lambda = 0$, on obtient $\langle \rho \left(Ax - b\right), \mu \rangle \geq 0$ et en prenant $\lambda = 2\mu$, on obtient $\langle \rho \left(Ax - b\right), \mu \rangle \leq 0$. Alors, nous avons $\langle Ax - b, \mu \rangle = 0$.
D'après la formule de projection sur $\Omega = \mathbb{R}_+^m$, on peut écrire pour chaque $i \in \{1, 2, \ldots, m\}$,

$$\mu_i = \max \left(\mu_i + \rho \left(Ax - b\right)_i, 0\right).$$

En particulier, $\mu_i \geq \mu_i + \rho \left(Ax - b\right)_i$ donc $\left(Ax - b\right)_i \leq 0$. Cela montre que $Ax - b \leq 0$

□

Ainsi on peut réécrire les conditions d'optimalité sous la forme suivante, pour tout $\rho > 0$:

$$\mu \geq 0,\ \nabla f\left(x\right) + A^T \mu = 0, \tag{3.11}$$
$$\mu = \mathcal{P}_\Omega \left[\mu + \rho \left(Ax - b\right)\right]. \tag{3.12}$$

Nous avons l'algorithme suivant :

Algorithme 2 (Algorithme d'Uzawa : contraintes d'inégalité affines)

1 : On prend un multiplicateur de départ $\mu^0 \in \mathbb{R}_+^m$. On fixe un pas $\rho > 0$.
2 : Puis on itère sur $n \geq 0$:
 – Calculer x^n te que $\nabla f\left(x^n\right) + A^T \mu^n = 0$.
 – calculer $\mu^{n+1} = \mathcal{P}_\Omega \left[\mu^n + \rho \left(Ax^n - b\right)\right]$.

Pour la convergence d'Algorithme 2, nous avons le théorème suivant :

Théorème 3.2. *(Cas de contraintes d'inégalité affines) Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable et α-convexe c'est à dire*

$$\forall x, y \in \mathbb{R}^n,\ \langle \nabla f\left(x\right), \nabla f\left(x\right), x - y \rangle \geq \alpha \left\|x - y\right\|^2,$$

et soit un ensemble de contraintes non vide $\mathcal{C} = \{x \in \mathbb{R}^n : Ax - b \leq 0\}$, où A est une matrice rectangulaire de taille $m \times n$ avec $m < n$. Nous avons :

1. *Il existe un unique minimiseur x de f sur $\mathcal{C}$.*
2. *Pour tout $\rho \in \left]0, 2\alpha/\|A\|^2\right[$, pour tout μ^0 point de départ, Algorithme 2 est bien défini et converge c'est à dire*
$$\lim_{n\to+\infty} x^n = x.$$
3. *Si, de plus, A est surjective et ∇f est continue, alors la suite (μ^n) est convergente vers unique point μ solution de* (3.12).

Démonstration. La preuve est pratiquement identique à celle du Théorème 3.1. la seule différence provient de la projection sur Ω. Nous avons :

$$\left\|\mu^{n+1}-\mu\right\|^2 = \left\|\mathcal{P}_\Omega\left[\mu^n+\rho\left(Ax^n-b\right)\right]-\mathcal{P}_\Omega\left[\mu+\rho\left(Ax-b\right)\right]\right\|^2$$

En utilisant (1.6), on obtient :

$$\begin{aligned}\left\|\mu^{n+1}-\mu\right\|^2 &\le \left\|\left(\mu^n+\rho\left(Ax^n-b\right)\right)-\left(\mu+\rho\left(Ax-b\right)\right)\right\|^2\\ &\le \left\|\mu^n-\mu+\rho A\left(x^n-x\right)\right\|^2\\ &\le \left\|\mu^n-\mu\right\|^2+2\rho\left\langle\mu^n-\mu, A\left(x^n-x\right)\right\rangle+\rho^2\left\|A\left(x^n-x\right)\right\|^2.\end{aligned}$$

Nous avons $\|A(x^n-x)\|^2 \le \|A\|^2\|x^n-x\|^2$. D'autre part, f est α-convexe, on obtient :

$$\begin{aligned}\left\langle\mu^n-\mu, A\left(x^n-x\right)\right\rangle &= \left\langle A^T\left(\mu^n-\mu\right), x^n-x\right\rangle\\ &= -\left\langle\nabla f\left(x^n\right)-\nabla f\left(x\right), x^n-x\right\rangle\\ &\le -\alpha\left\|x^n-x\right\|^2\end{aligned}$$

Ainsi,

$$\left\|\mu^{n+1}-\mu\right\|^2 \le \left\|\mu^n-\mu\right\|^2-\gamma\left\|x^n-x\right\|^2, \tag{3.13}$$

avec $\gamma := \rho\left(2\alpha-\rho\|A\|^2\right)$. En particulier, si $0<\rho<2\alpha/\|A\|^2$, alors $\gamma>0$.

La suite $\left(\|\mu^n-\mu\|^2\right)$ est décroissante et minorée par 0, donc convergente vers une limite notée ℓ. Ensuite on renverse l'inégalité (3.13) pour écrire

$$\gamma\left\|x^n-x\right\|^2 \le \left\|\mu^n-\mu\right\|^2-\left\|\mu^{n+1}-\mu\right\|^2 \xrightarrow[n\to+\infty]{} \ell-\ell=0.$$

Alors, la suite (x^n) converge vers x. □

3.2.2 Contraintes d'inégalité convexes

Soit le problème d'optimisation avec contraintes d'inégalité convexes défini par :

$$\begin{cases}\min f(x)\\ x\in\mathcal{C}\end{cases} \tag{3.14}$$

où $\mathcal{C}=\{x\in\mathbb{R}^n : g_i(x)\le 0,\ i=1,2,\dots,m\}$ et chaque contrainte $g_i:\mathbb{R}^n\to\mathbb{R}^m$ est supposée convexe.

Rappelons que si x est un point de minimum local de f sur $\mathcal{C}$, avec f et g_i différentiables en x et si x est un point régulier, alors on peut écrire les conditions de KKT sous la forme suivante :

$$\nabla f(x)+\sum_{i=1}^m \mu_i\nabla g_i(x)=0, \tag{3.15}$$

$$\mu\ge 0,\ g(x)\le 0,\ \langle\mu, g(x)\rangle=0. \tag{3.16}$$

Le lemme suivant permet de réécrire le deuxième jeu d'équations sur μ de manière plus compacte :

Lemme 3.2. *Soient $\Omega = \mathbb{R}_+^n$ et $\rho > 0$. Pour tout $\mu \in \mathbb{R}^m$ et $g(x) \in \mathbb{R}^m$, nous avons :*

$$(\mu \geq 0,\ g(x) \leq 0,\ \langle \mu, g(x) \rangle = 0) \Leftrightarrow \mu = \mathcal{P}_\Omega [\mu + \rho g(x)].$$

Démonstration. La preuve est identique à celle du Lemme 3.1, où $Ax - b$ est remplacé par $g(x)$. □

Ainsi, on peut réécrire les conditions d'optimalité de KKT sous la forme suivante, pour tout $\rho > 0$,

$$\nabla f(x) + \sum_{i=1}^{m} \mu_i g_i(x) = 0, \tag{3.17}$$

$$\mu = \mathcal{P}[\mu + \rho g(x)]. \tag{3.18}$$

Alors, nous avons l'algorithme suivant :

Algorithme 3 (Algorithme d'Uzawa : contraintes d'inégalité convexes)

1: On prend un multiplicateur de départ $\mu^0 \in \mathbb{R}_+^m$. On fixe un pas $\rho > 0$.
2: Puis on itère sur $n \geq 0$:
   i) Calculer x^n tel que $\nabla f(x^n) + \sum_{i=1}^m \mu_i^n g_i(x^n) = 0$.
   ii) Calculer $\mu^{n+1} = \mathcal{P}_\Omega [\mu^n + \rho g(x^n)]$.

Pour que l'algorithme soit bien défini il faudra montrer l'existence d'un vecteur x^n solution de i).

Théorème 3.3. *(Cas de contraintes d'inégalité convexes) Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable et α-convexe c'est à dire*

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

et soit un ensemble de contraintes non vide $\mathcal{C} = \{x \in \mathbb{R}^n : g_i(x) \leq 0,\ i = 1, 2, \ldots, m\}$avec g_i est différentiable et convexe. On suppose que g est lipschitzienne c'est à dire

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

Soit $x \in \mathcal{C}$ est un point régulier, alors nous avons :

1. *Il existe un unique minimiseur x de f sur $\mathcal{C}$.*
2. *Pour tout $\rho \in \,]0, 2\alpha/M^2[$, pour tout $\mu^0 \geq 0$ point de départ, Algorithme 3 est bien défini et converge c'est à dire*

$$\lim_{n \to +\infty} x^n = x.$$

3. *Si, de plus, la matrice $A = [\nabla g_1(x), \ldots, \nabla g_m(x)]^T$ est surjective, et si f et g sont de classe $\mathcal{C}^1$, alors la suite (μ^n) est convergente.*

Démonstration. 1. est classique.

2. Commençons par vérifier l'existence de x^n. On note que $\mathcal{L}(x, \mu^n) = f(x) + \sum_{i=1}^m \mu_i^n g_i(x)$ est une fonction strictement convexe de x, car f est strictement convexe, les contraintes g_i sont convexes et μ_i^n sont positifs. Alors, $\mathcal{L}(x, \mu^n)$ est coercive, car f est coercive et g_i est de Lipschitz. Ainsi, $\mathcal{L}(x, \mu^n)$ possède un unique minimiseur x^n sur $\mathbb{R}^n$, caractérisé par :

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

d'où

$$\nabla f\left(x^n\right)+\sum_{i=1}^m \mu_i^n g_i\left(x^n\right)=0.$$

Ce qui prouve l'existence et l'unicité de x^n.
Remarquons que par définition du minimiseur x^n,

$$\forall y \in \mathbb{R}^n,\ \mathcal{L}\left(x^n, \mu^n\right) \leq \mathcal{L}\left(y, \mu^n\right),$$

c'est à dire

$$\forall y \in \mathbb{R}^n,\ f\left(x^n\right)+\left\langle\mu^n, g\left(x^n\right)\right\rangle \leq f(y)+\left\langle\mu^n, g(y)\right\rangle . \tag{3.19}$$

Soit $z \in \mathbb{R}^n$ et $t \in\ ]0,1[$. En appliquant (3.19) avec $y=x^n+t\left(z-x^n\right)$, on obtient :

$$f\left(x^n+t\left(z-x^n\right)\right)-f\left(x^n\right)+\sum_{i=1}^m \mu_i^n\left[g_i\left(x^n+t\left(z-x^n\right)\right)-g_i\left(x^n\right)\right] \geq 0. \tag{3.20}$$

Mais par convexité des g_i, nous avons :

$$g_i\left(x^n+t\left(z-x^n\right)\right)-g_i\left(x^n\right) \leq t\left[g_i(z)-g_i\left(x^n\right)\right],\ i=1,2, \ldots, m.$$

D'après (3.20), on obtient :

$$f\left(x^n+t\left(z-x^n\right)\right)-f\left(x^n\right)+t \sum_{i=1}^m \mu_i^n\left[g_i(z)-g_i\left(x^n\right)\right] \geq 0.$$

En divisant par t et en passant à la limite quand $t \rightarrow 0$, on obtient :

$$\forall z \in \mathbb{R}^n,\ \left\langle\nabla f\left(x^n\right), z-x^n\right\rangle+\sum_{i=1}^m \mu_i^n\left[g_i(z)-g_i\left(x^n\right)\right] \geq 0. \tag{3.21}$$

Considérons à présent le point x et le vecteur μ. D'après les conditions de KKT, ils vérifient la relation suivante :

$$\nabla f(x)+\sum_{i=1}^m \mu_i \nabla g_i(x)=0. \tag{3.22}$$

En définissant comme plus haut le lagrangien $\mathcal{L}(x, \mu)=f(x)+\sum_{i=1}^m \mu_i g_i(x)$, on constate que la relation (3.22) s'écrit $\nabla_x \mathcal{L}(x, \mu)=0$, ce qui montre que x est le minimiseur unique de l'application $y \mapsto \mathcal{L}(y, \mu)$ sur $\mathbb{R}^n$ (l'existence et l'unicité d'un tel minimiseur s'obtiennent par les mêmes arguments que pour l'application $y \mapsto \mathcal{L}\left(y, \mu^n\right)$). Donc, on a :

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

En appliquant le même raisonnement que précédemment, on obtient :

$$\forall z \in \mathbb{R}^n,\ \langle\nabla f(x), z-x\rangle+\sum_{i=1}^m \mu_i\left[g_i(z)-g_i(x)\right] \geq 0. \tag{3.23}$$

En prenant $z = x$ dans (3.21) et $z = x^n$ dans (3.23), on obtient :

$$\langle \nabla f(x^n), x - x^n \rangle + \sum_{i=1}^{m} \mu_i^n [g_i(x) - g_i(x^n)] \geq 0,$$
$$\langle \nabla f(x), x^n - x \rangle + \sum_{i=1}^{m} \mu_i [g_i(x^n) - g_i(x)] \geq 0.$$

En sommant les deux inégalités, on obtient :

$$\langle \nabla f(x) - \nabla f(x^n), x^n - x \rangle + \sum_{i=1}^{m} (\mu_i - \mu_i^n) [g_i(x^n) - g_i(x)] \geq 0$$

En utilisant la fortement convexe de f, on trouve :

$$\langle \mu^n - \mu, g(x^n) - g(x) \rangle \leq -\langle \nabla f(x^n) - \nabla f(x), x^n - x \rangle \leq -\alpha \|x^n - x\|^2 \tag{3.24}$$

Nous allons utiliser l'estimation (3.24) pour démontrer la convergence de la suite $(\|\mu^n - \mu\|)$. Pour cela, on écrit :

$$\begin{aligned}
\left\|\mu^{n+1} - \mu\right\|^2 &= \|\mathcal{P}_\Omega[\mu^n + \rho g(x^n)] - \mathcal{P}_\Omega[\mu + \rho g(x)]\|^2 \\
&\leq \|[\mu^n + \rho g(x^n)] - [\mu + \rho g(x)]\|^2 \\
&\leq \|(\mu^n - \mu) + \rho(g(x^n) - g(x))\|^2 \\
&\leq \|\mu^n - \mu\|^2 + 2\rho \langle \mu^n - \mu, g(x^n) - g(x) \rangle + \rho^2 M^2 \|x^n - x\|^2 \\
&\leq \|\mu^n - \mu\|^2 - \gamma \|x^n - x\|^2.
\end{aligned}$$

en utilisant le caractère de Lipschitz de g et l'inégalité (3.24) avec $\gamma := \rho\left(2\alpha - \rho M^2\right)$.
On conclut que la suite (x^n) est convergente vers x, mais pas nécessairement celle de la suite (μ^n).

3. Si les fonctions g_i sont de classe $\mathcal{C}^1$, alors l'application $y \mapsto A(y) = [\nabla g_1(y), \nabla g_2(y), \ldots, \nabla g_m(y)]^T$ est continue. De plus, (voir la preuve de la convergence de l'algorithme d'Usawa pour le cas des contraintes d'égalité affines), $A(y)$ étant surjective, la matrice $A^T(y)$ est injective, et dans ce cas, la matrice $A(y)A^T(y)$ est inversible. Par conséquent, son déterminant est non nul. Par continuité du déterminant et de l'application $y \mapsto A(y)A^T(y)$, puisque $det\left[A(y)A^T(y)\right] \neq 0$ c'est à dire que la matrice $A(x^n)A^T(x^n)$ est inversible.
Or, la relation

$$\nabla f(x^n) + \sum_{i=1}^{m} \mu_i^n \nabla g_i(x^n) = 0$$

s'écrit

$$\nabla f(x^n) + A^T(x^n)\mu^n = 0,$$

d'où

$$A(x^n)\nabla f(x^n) + A(x^n)A^T(x^n)\mu^n = 0.$$

Pour n assez grand, la matrice $A(x^n)A^T(x^n)$ est inversible, ce qui permet d'exprimer μ^n sous la forme suivante :

$$\mu^n = -\left[A(x^n)A^T(x^n)\right]^{-1} A(x^n)\nabla f(x^n).$$

Comme $x^n \xrightarrow[n\to+\infty]{} x$, on en déduit par continuité la convergence des μ^n vers une limite μ^* qui s'écrit

$$\mu^* = -\left[A(x)\,A^T(x)\right]^{-1} A(x)\,\nabla f(x)\,.$$

En passant à la limite dans la relation

$$\nabla f(x^n) + A^T(x^n)\,\mu^n = 0,$$

on obtient (par continuité des dérivées partielles de f)

$$\nabla f(x) + A^T(x)\,\mu^* = 0.$$

Ainsi, μ^* est solution de (3.17) ; c'est même l'unique solution du système, puisque par injectivité de $A^T(x)$, si deux vecteurs μ_1^*, μ_2^* satisfait

$$\nabla f(x) + A^T(x)\,\mu_1^* = 0 \text{ et } \nabla f(x) + A^T(x)\,\mu_2^* = 0$$

alors

$$A^T(x)\left[\mu_1^* - \mu_2^*\right] = 0 \Rightarrow \mu_1^* = \mu_2^*.$$

$\square$