

# Chapitre 3 : algorithmes d'optimisation locale, avec contraintes

Problème à étudier : minimisation

de  $J : \Omega \rightarrow \mathbb{R}$  où

$$\Omega = \{x \in \mathbb{R}^n \mid C_E(x) = 0 \text{ et } C_I(x) \leq 0\}$$

avec  $\begin{cases} C_E : \mathbb{R}^n \rightarrow \mathbb{R}^p \text{ (et } J, C_E, C_I) \\ C_I : \mathbb{R}^n \rightarrow \mathbb{R}^q \text{ régulières} \end{cases}$

Condition d'existence d'une solution : KKT (Chap 1.)

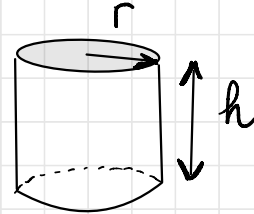
utilisent  $\mathcal{L}(x, \lambda, \mu) = J(x) + \langle \lambda, C_E(x) \rangle + \langle \mu, C_I(x) \rangle$

variables primales

variables duales

## Exemples "concrets"

1) Problème de la canette (TD1, Ex 3)



$J(r, h)$

On veut minimiser  $(2\pi r h + 2\pi r^2)$   
 $(r, h)$

sous la contrainte  $\pi r^2 h \geq V_0$   
(volume fixé)  
(ou supérieur à  $V_0$ )

→ KKT

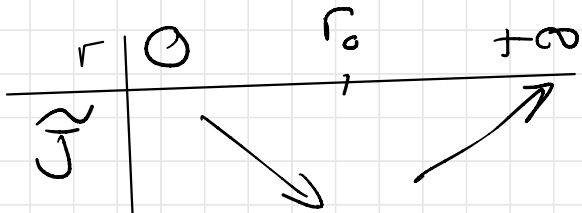
→ Elimination de  $h$  dans  $J$ :

$$h = \frac{V_0}{\pi r^2} \rightsquigarrow \hat{J}(r) = 2\pi r^2 + \frac{2\pi r V_0}{\pi r^2} = 2\pi r^2 + \frac{2V_0}{r}$$

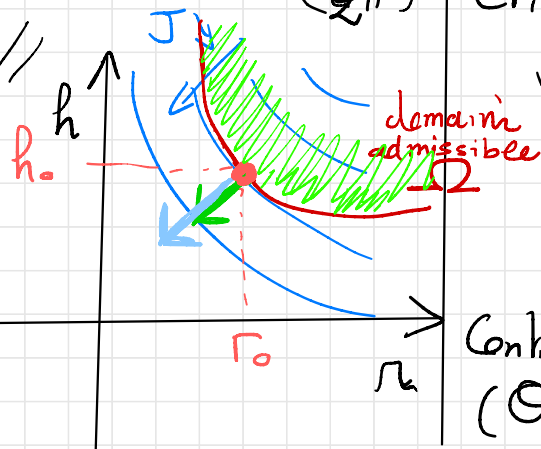
On minimise  $\tilde{J}$  :

$$\tilde{J}'(r) = 4\pi r - \frac{2\sqrt{v_0}}{r^2} \text{ soit}$$

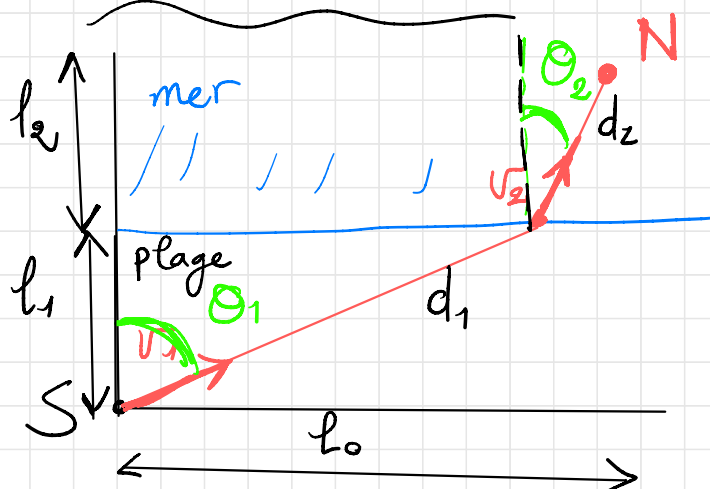
$$\tilde{J}'(r_0) = 0 \text{ si } r_0 = \left(\frac{\sqrt{v_0}}{2\pi}\right)^{\frac{1}{3}}$$



La surface est minimale par  $r_0 = \left(\frac{\sqrt{v_0}}{2\pi}\right)^{\frac{1}{3}}$   
 et  $h_0 = \frac{\sqrt{v_0}}{\pi r_0^2} \parallel h$



2) Problème du sauveur (TD2, Ex4) 2



Critère à minimiser

$$J(\theta_1, \theta_2) = \frac{d_1}{\sqrt{v_1}} + \frac{d_2}{\sqrt{v_2}}$$

$$= \frac{l_1}{\cos\theta_1 \sqrt{v_1}} + \frac{l_2}{\cos\theta_2 \sqrt{v_2}}$$

Contrainte :  $l_0 = l_1 \tan\theta_1 + l_2 \tan\theta_2$   
 $(\theta_1, \theta_2 \in [0, \frac{\pi}{2}])$

\* Méthode de substitution : compliquée.

\* KKT :

$$\mathcal{L}(\theta_1, \theta_2, \lambda) = \frac{l_1}{\cos \theta_1 v_1} + \frac{l_2}{\cos \theta_2 v_2} + \lambda (l_1 \tan \theta_1 + l_2 \tan \theta_2 - l_0)$$

Sous réserve de qualification, on a

(après simplifications)

$$\begin{cases} \sin \theta_1 + \lambda v_1 = 0 \\ \sin \theta_2 + \lambda v_2 = 0 \\ l_1 \tan \theta_1 + l_2 \tan \theta_2 = l_0 \end{cases}$$

$$\Rightarrow \frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2} \quad (\text{loi de Descartes})$$

On remplace  $\theta_1$  et  $\theta_2$  en fonction de  $\lambda$  dans la condition d'admissibilité :

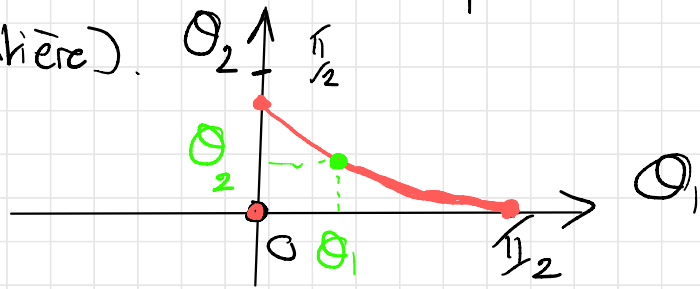
$$\lambda l_1 v_1 \sqrt{1 - \lambda^2 v_2^2} + \lambda l_2 v_2 \sqrt{1 - \lambda^2 v_1^2} = -l_0 \sqrt{1 - \lambda^2 v_1^2} \sqrt{1 - \lambda^2 v_2^2}$$

En élevant au carré, on obtient une équation de degré 4 (en  $\lambda^2$ ) qu'on peut résoudre et en déduire la valeur de  $\lambda < 0$ .

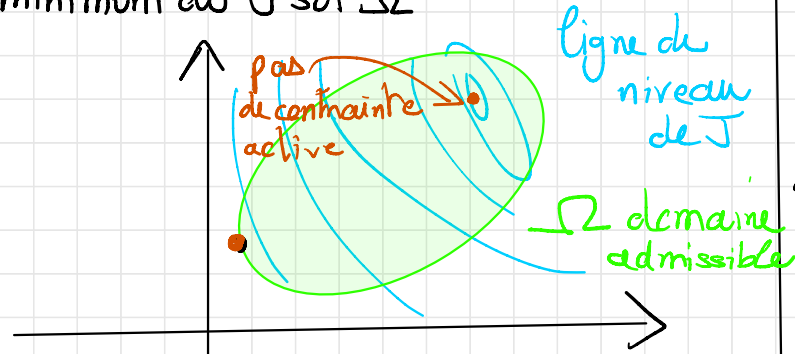
(\*) exactement ou numérique.

On en déduit les valeurs de  $\theta_1$  et  $\theta_2$ .

Réciproquement, il s'agit bien d'un minimum car la fonction  $J$  est  $C^0$  sur un ensemble compact (et le minimum n'est pas atteint sur sa frontière).



On va chercher à construire une suite d'approximations  $(x_k)_{k \in \mathbb{N}}$  d'un minimum de  $J$  sur  $\Omega$



Il est nécessaire d'adapter les méthodes du paragraphe 2, en particulier si les contraintes sont actives au point de minimum.

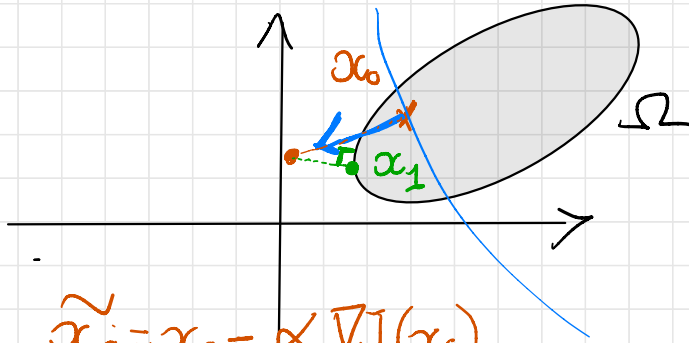
Trois approches possibles vont être étudiées 4

→ méthode du gradient projeté : à chaque étape d'une méthode de gradient, on rajoute une opération de projection afin de revenir dans le domaine admissible

→ méthode de pénalisation : on se ramène à minimiser une fonction sans contraintes en pénalisant (de plus en plus) les contraintes.

→ méthode duale : on cherche à approcher un point selle du Lagrangien correspondant à une solution de KKT ( $x, \lambda, \mu$  : variables primales et duales à rechercher simultanément)

### 3.1) Méthode du gradient projeté



$$\tilde{x}_0 = x_0 - \alpha \nabla J(x_0)$$

puis

$$x_1 = P_{\Omega}(\tilde{x}_0)$$

et plus généralement

$$x_{k+1} = P_{\Omega}(x_k - \alpha_k \nabla J(x_k))$$

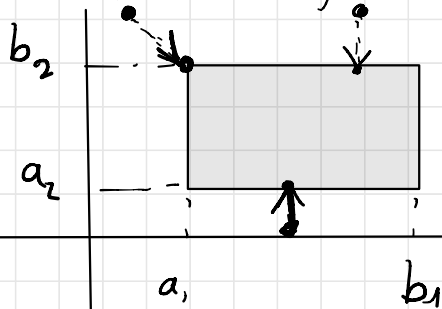
où  $P_{\Omega} : \begin{pmatrix} \mathbb{R}^n \rightarrow \Omega \\ \alpha \mapsto P_{\Omega}(x) \end{pmatrix}$  est

une fonction de projection sur  $\Omega$ .

Dans certains cas,  $P_{\Omega}$  est facile à définir et même à expliciter :

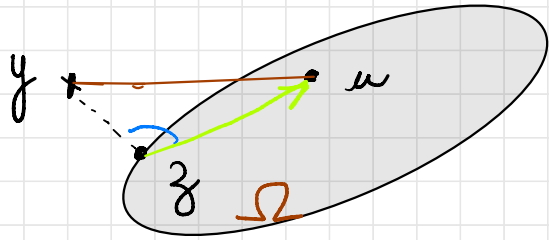
→ cas  $\Omega$  hypercube :

$$\Omega = \{x \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i\} \\ = [a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$$



$$P_{\Omega}((x_1, x_2)) = (\text{med}(x_1, a_1, b_1), \text{med}(x_2, a_2, b_2))$$

→ cas  $\Omega$  convexe fermé : dans ce cas,  
 il existe une unique projection orthogonale  
 sur  $\Omega$  : si  $y \in \mathbb{R}^n$ ,  $P_{\Omega}(y) = z$  est tel  
 que  $\forall u \in \Omega, \|y - z\| \leq \|y - u\|$



et vérifie :  
 $\forall u \in \Omega, \langle y - z, u - z \rangle \leq 0$   
 (expression non explicite)

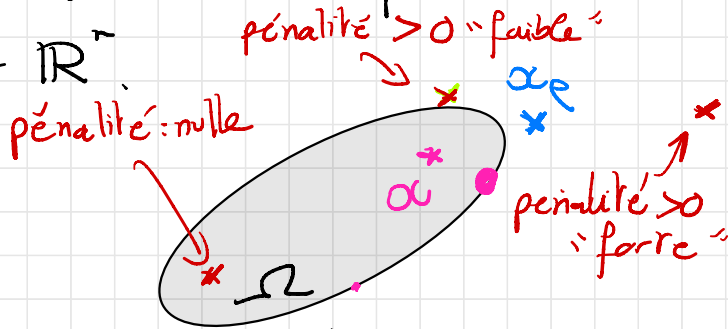
\* Le pas  $\alpha_k$  peut être adapté (après l'étape de projection) en essayant de maintenir une condition de descente entre  $\alpha_k$  et  $\alpha_{k+1}$ .

\* Lorsque la projection n'est pas explicite, il est possible d'utiliser une étape de "restauration".

\* Exemples (à venir) : cas de fonctionnelles quadratiques avec contraintes affines (IDT) et cas de la canette (implémentation).

### 3.2) Méthode de pénalisation

\* L'idée consiste à "pénaliser" la contrainte afin de se ramener à un problème sans contraintes sur  $\mathbb{R}^n$ .



On définit une fonction de pénalisation :

$$p : \mathbb{R}^n \rightarrow \mathbb{R} \text{ tq } (p \in C^1)$$

$$\bullet \forall x \in \Omega, p(x) = 0$$

$$\bullet \forall x \notin \Omega, p(x) > 0$$

et on cherche à minimiser la nouvelle fonction

$$\tilde{J}_\rho(x) = J(x) + \rho p(x)$$

où  $\rho > 0$  est un facteur d'amplification de la pénalisation.

Par une méthode de minimisation de type descente appliquée à  $\tilde{J}_\rho$  (cas sans contrainte) on obtient une solution  $\alpha_\rho$  minimisant  $\tilde{J}_\rho$ .

On répète cette minimisation en faisant tendre  $\rho$  vers l'infini. Sous certaines conditions, on peut montrer que

$$\lim_{\rho \rightarrow +\infty} \alpha_\rho = \alpha^* \text{ minimum de } J \text{ sur } \Omega.$$

(voir TD 4)

Le choix de la fonction de pénalisation n'est pas unique. Il est guidé par la définition de  $\Omega$ :

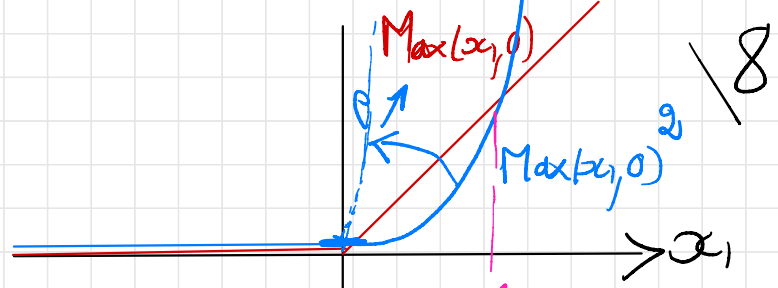
$$\Omega = \{x \in \mathbb{R}^n / C_E(x) = 0 \text{ et } C_I(x) \leq 0\}$$

\* Exemple

$$\Omega = \{x = (x_1, x_2) \in \mathbb{R}^2 / x_1 \leq 0\}$$



$$p(x_1, x_2) = \text{Max}(x_1, 0) \quad \text{ou} \quad \text{Max}(x_1, 0)^2$$



Plus généralement, on peut utiliser l'une des deux fonctions de pénalisation suivante :

\* norme 1:

$$p_1(x) = \sum_{i=1}^p |C_{E_i}(x)| + \sum_{j=1}^q \text{Max}(0, C_{I_j}(x))$$

\* norme 2:

$$p_2(x) = \sum_{i=1}^p (C_{E_i}(x))^2 + \sum_{j=1}^q \text{Max}(0, C_{I_j}(x))^2$$



Ici, la pénalisation en norme 2 est plus adaptée à un algorithme d'optimisation de type gradient ( $p_2$  est  $C^1$  mais  $p_1$  ne l'est pas).

Cette méthode est simple à mettre en œuvre. Dans certains cas, il existe des résultats de convergence de  $x_p$  vers  $x^*$ . (voir TD 4). Cependant, cette méthode se heurte à des difficultés pratiques, en particulier le fait que  $J_p$  est mal conditionnée vis à vis de la recherche d'un minimum lorsque  $p$  est grand.

D'autre part, les approximations  $x_p$  obtenues ne sont pas dans  $\Omega$ . Ce dernier point peut être corrigé, en pénalisation non pas les points extérieurs mais les points intérieurs près de la frontière de  $\Omega$ .

### 3.3) Méthode duale

On rappelle l'expression du Lagrangien associé:  
 $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p \times (\mathbb{R}_+)^q \rightarrow \mathbb{R}$  où

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

$\lambda_i \in \mathbb{R}$        $\mu_j \in \mathbb{R}_+$

↑                      ↑  
variables      variable  
primales      duales

et  $\Omega = \left\{ x \in \mathbb{R}^n \mid c_{E_i}(x) = 0 \text{ et } c_{I_j}(x) \leq 0 \right\}$

<sup>à vérifier</sup>  
On remarque que le problème de minimisation de  $f$  sur  $\Omega$  peut se récrire avec  $Z$ :

$$\inf_{x \in \Omega} f(x) = \inf_{x \in \mathbb{R}^n} \left( \sup_{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q} Z(x, \lambda, \mu) \right) \quad (P)$$

(problème primal)

Le problème dual consiste à échanger Inf et Sup:

$$\sup_{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q} \left( \inf_{x \in \mathbb{R}^n} Z(x, \lambda, \mu) \right) = \sup_{(\lambda, \mu) \in X^*} f^*(\lambda, \mu) \quad (D)$$

cà  $X^* = \{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q \mid f^*(\lambda, \mu) > -\infty\}$   
(domaine admissible de la fonction duale  $f^*$ )

Le problème dual (D) 10  
n'est pas toujours forcément équivalent au problème primal.  
Il est par contre plus simple à résoudre (moyennant le calcul de  $f^*$ )  
car on peut montrer que  $f^*$  est toujours concave et  $X^*$  est toujours convexe (exercice)

Il existe un lien entre problèmes dual et primal:

$$\sup_{(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q} \left( \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu) \right) \leq \inf_{x \in \mathbb{R}^n} \left( \sup_{(\lambda, \mu)} \mathcal{L}(x, \lambda, \mu) \right)$$

(principe de dualité faible)

Lorsque l'inégalité est stricte, on parle de saut de dualité qu'on mesure avec l'écart

$$\inf_{(\lambda, \mu)} \left( \sup_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu) \right) - \sup_{x \in \mathbb{R}^n} \left( \inf_{(\lambda, \mu)} \mathcal{L}(x, \lambda, \mu) \right)$$

On s'intéresse au cas où ce saut est nul. On rappelle la notion de point selle:

Def: soit  $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_+^q$  11

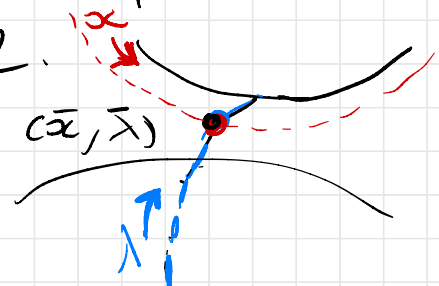
On dit que  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  est un point selle du Lagrangien si

$$\left\{ \begin{array}{l} \forall (\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q, \mathcal{L}(\bar{x}, \lambda, \mu) \leq \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) \\ \forall x \in \mathbb{R}^n, \mathcal{L}(x, \bar{\lambda}, \bar{\mu}) \leq \mathcal{L}(x, \bar{\lambda}, \bar{\mu}) \end{array} \right.$$

On a donc

$$\begin{aligned} f(\bar{x}) &= \inf_{x \in \mathbb{R}^n} \left( \sup_{(\lambda, \mu)} \mathcal{L}(x, \lambda, \mu) \right) = \sup_{(\lambda, \mu)} \left( \inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu) \right) \\ &= \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) \end{aligned}$$

et  $\bar{x}$  est bien un point de minimum de  $f$  sur  $\Omega$ .



Remarque : si  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  est un point selle de  $\mathcal{L}$ , on a bien

$$\begin{cases} \nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0 \\ C_E(\bar{x}) = 0 \\ C_I(\bar{x}) \leq 0 \\ \bar{\mu} \geq 0 \end{cases}$$

c'est à dire que  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  vérifie bien KKT.

On a la CNS suivante d'existence d'un point selle par  $\mathcal{L}$  :

## Théorème (dualité)

19

Le point  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  est un point selle de  $\mathcal{L}$  si et seulement si :

- (i)  $\bar{x}$  est solution du problème primal
- (ii)  $(\bar{\lambda}, \bar{\mu})$  est solution du problème dual
- (iii) le saut de dualité est nul

Dans ce cas, on

$$f(\bar{x}) = \inf_x (\sup_{(\lambda, \mu)} \mathcal{L}) = \sup_{(\lambda, \mu)} (\inf_x \mathcal{L}) = \inf_{(\lambda, \mu)} f \quad \text{(exercice)}$$

Le problème de la recherche de  $\bar{x}$ , solution du problème initial, en la recherche d'un point selle de  $\mathcal{L}$  sur  $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}_+^q$ .

Le cas particulier convexe assure un saut de dualité nulle:

Théorème (dualité forte). On suppose  $f$  convexe,  $C_E$  affine,  $C_I$  convexe.

Si  $\bar{x}$  est solution du problème primal, et qu'il existe  $x_0 \in \Omega$  tq  $C_E(x_0) = 0$  et  $C_I(x_0) < 0$ , alors le saut de dualité est nul et il existe  $(\bar{\lambda}, \bar{\mu})$  tq  $(\bar{x}, \bar{\lambda}, \bar{\mu})$  est un point selle de  $\mathcal{L}$ .

Les algorithmes associés à ces deux théorèmes vont consister à rechercher un point selle de  $\mathcal{L}$  de la manière suivante

Alternativement, par  $(x_k, \lambda_k, \mu_k)$  donné

\* on minimise

$x \mapsto \mathcal{L}(x, \lambda_k, \mu_k)$  avec une itération de gradient par exemple

$$\rightarrow x_{k+1} = x_k - \rho_1 \nabla_x \mathcal{L}(x_k, \lambda_k, \mu_k)$$

\* on maximise

$(\lambda, \mu) \mapsto \mathcal{L}(x_{k+1}, \lambda, \mu)$  avec une itération de gradient prédictif:

$$\lambda_{k+1} = \lambda_k + \rho_2 \nabla_{\lambda} \mathcal{L}(x_{k+1}, \lambda_k, \mu_k)$$

et enfin projection

$$\mu_{k+1} = \text{Max}(0, \mu_k + \rho_2 \nabla_{\mu} \mathcal{L}(x_{k+1}, \lambda_{k+1}, \mu_k))$$

On parle ici de la méthode d'Uzawa

$\rho_1$  et  $\rho_2$  désignent les pas (constants) de gradient.

Sous certaines conditions, il est possible de montrer la convergence de  $(\alpha_k, \lambda_k, \mu_k)$  vers  $(\bar{\alpha}, \bar{\lambda}, \bar{\mu})$  point selle de  $\mathcal{L}$ .

(voir exemple en TD4)

Cet algorithme n'a pas les défauts des deux algorithmes précédents. Cependant, il ne converge pas forcément. Il est possible aussi d'améliorer l'étape de gradient à pas constant avec une stratégie de recherche linéaire ou une direction de descente améliorée (Newton).

Exemple d'implémentation sur le cas de la canette (voir code Salab).