

Cours 12 : Optimisation non linéaire sous contraintes

Christophe Gonzales

LIP6 – Université Paris 6, France

Problème à résoudre

Optimisation sous contraintes

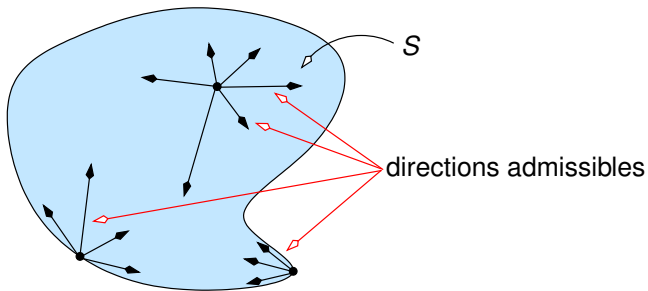
- $f : \mathbb{R}^n \mapsto \mathbb{R}$ fonction à minimiser
- $S \subseteq \mathbb{R}^n$
- Problème d'optimisation : $\min_{x \in S} f(x)$

	f	S
prog linéaire	linéaire	polyèdre convexe
prog quadratique	forme quadratique semi définie positive	polyèdre convexe
prog convexe	convexe	ensemble convexe
prog non linéaire	quelconque	ensemble quelconque

Directions admissibles

Définition d'une direction admissible

- Soit $f : \mathbb{R}^n \mapsto \mathbb{R}$, $S \subseteq \mathbb{R}^n$, $x \in S$
- $d =$ direction admissible si $\exists \bar{\lambda} > 0$ t.q. :
 $x + \lambda d \in S \quad \forall 0 \leq \lambda \leq \bar{\lambda}$



Proposition 1

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
 - \hat{x} minimum local de f sur S
- $\implies \forall$ direction admissible d en \hat{x} , on a $\vec{\nabla}f(\hat{x})^T d \geq 0$

Corollaire

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
- une condition nécessaire pour que $\hat{x} \in S$ soit minimum local de f est que $\vec{\nabla}f(\hat{x})^T d \geq 0, \forall$ direction admissible d en \hat{x}

Démonstration de la proposition :

d direction admissible en $\hat{x} \implies \exists \bar{\lambda} > 0$ t.q. $\hat{x} + \lambda d \in S, \forall 0 \leq \lambda \leq \bar{\lambda}$

Soit $g(\lambda) = f(\hat{x} + \lambda d)$

hypothèses $\implies g$: minimum local en $\lambda = 0$

développement limité : $g(\lambda) = g(0) + \lambda g'(0) + o(\lambda)$

si $g'(0) < 0$ alors, pour λ petit, $\lambda g'(0) + o(\lambda) < 0$

$\implies g(\lambda) < g(0) \implies$ impossible : $g(0)$ minimum local

$\implies g'(0) \geq 0$. Or $g'(\lambda) = \vec{\nabla} f(\hat{x} + \lambda d)^T d$

et donc $\vec{\nabla} f(\hat{x})^T d \geq 0$

Cas particulier d'ensemble S

- $S = \{x : g_i(x) \leq 0, i = 1, \dots, m\}$

- $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1

- x^* : point quelconque de S

- contrainte serrée = contrainte telle que $g_i(x^*) = 0$

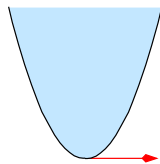
- $I(x^*) = \{\text{indices } i \text{ t.q. } g_i(x^*) = 0\}$
= \{\text{indices des contraintes serrées}\}

- $D(x^*) = \{d : \vec{\nabla} g_i(x^*)^T d \leq 0, i \in I(x^*)\}$
 \implies direction admissible en $x^* \in D(x^*)$



l'inverse n'est pas forcément vrai :

$$S = \{x : x_1^2 - x_2 \leq 0\}$$



$$D(0) = \{(d_1, d_2) : d_2 \geq 0\}$$

$$\implies d = (1, 0) \in D(0)$$

Conditions de Kuhn et Tucker (2/4)

\hat{x} = minimum local de f sur S

$$I(\hat{x}) = \{i_1, \dots, i_k\}$$

Supp que $D(\hat{x}) = \{\text{toutes les directions admissibles en } \hat{x}\}$

Soit \bar{A} = matrice des $-\vec{\nabla} g_i(\hat{x})$, $i \in I(\hat{x})$

$$\bar{A} = \begin{bmatrix} -\frac{\partial g_{i_1}}{\partial x_1}(\hat{x}) & -\frac{\partial g_{i_2}}{\partial x_1}(\hat{x}) & \dots & -\frac{\partial g_{i_k}}{\partial x_1}(\hat{x}) \\ \dots & \dots & \dots & \dots \\ -\frac{\partial g_{i_1}}{\partial x_n}(\hat{x}) & -\frac{\partial g_{i_2}}{\partial x_n}(\hat{x}) & \dots & -\frac{\partial g_{i_k}}{\partial x_n}(\hat{x}) \end{bmatrix}$$

Rappel Proposition 1 :

\hat{x} minimum local $\implies \forall$ direction admissible d en \hat{x} , $\vec{\nabla} f(\hat{x})^T d \geq 0$

\implies le système $d^T \bar{A} \geq 0$, $\vec{\nabla} f(\hat{x})^T d < 0$ n'a pas de solution

Rappel : Lemme de Farkas

- A une matrice $m \times n$
- c un vecteur de taille n
- les deux énoncés suivants sont équivalents :
 - 1 $Ax \geq 0 \implies c^T x \geq 0$
 - 2 il existe $v \geq 0$ tel que $v^T A = c^T$

le système $d^T \bar{A} \geq 0, \vec{\nabla} f(\hat{x})^T d < 0$ n'a pas de solution

$$\implies (d^T \bar{A} \geq 0 \implies \vec{\nabla} f(\hat{x})^T d \geq 0)$$

$$\implies \exists \text{ vecteur } \lambda \geq 0 \text{ tel que } \bar{A}\lambda = \vec{\nabla} f(\hat{x})$$

$$\implies \vec{\nabla} f(\hat{x}) = \sum_{i \in I(\hat{x})} -\lambda_i \vec{\nabla} g_i(\hat{x}) \implies \vec{\nabla} f(\hat{x}) + \sum_{i \in I(\hat{x})} \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$$

Conditions de Kuhn et Tucker (4/4)

⇒ conditions de Kuhn et Tucker (1951) :

Conditions nécessaires d'optimalité de Kuhn et Tucker

- f de classe C^1
- hypothèse de qualification des contraintes :
 \hat{x} = point de S tel que $D(\hat{x}) = \{\text{directions admissibles en } \hat{x}\}$
- alors condition nécessaire pour que \hat{x} = minimum local de f :
 $\exists \lambda_i \geq 0, i = 1, \dots, m$, tels que :

$$\textcircled{1} \quad \vec{\nabla} f(\hat{x}) + \sum_{i=1}^m \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$$

$$\textcircled{2} \quad \lambda_i g_i(\hat{x}) = 0, \quad \forall i \in \{1, \dots, m\}$$



lorsque $i \notin I(\hat{x})$, il suffit de fixer $\lambda_i = 0$



l'hypothèse de qualification des contraintes est primordiale pour le théorème de Kuhn et Tucker

Exemple :

$$S = \{x \in \mathbb{R}^2 : g_i(x) \leq 0, i = 1, 2, 3\}$$

$$g_1(x_1, x_2) = -x_1 \quad g_2(x_1, x_2) = -x_2 \quad g_3(x_1, x_2) = x_2 - (1 - x_1)^3$$

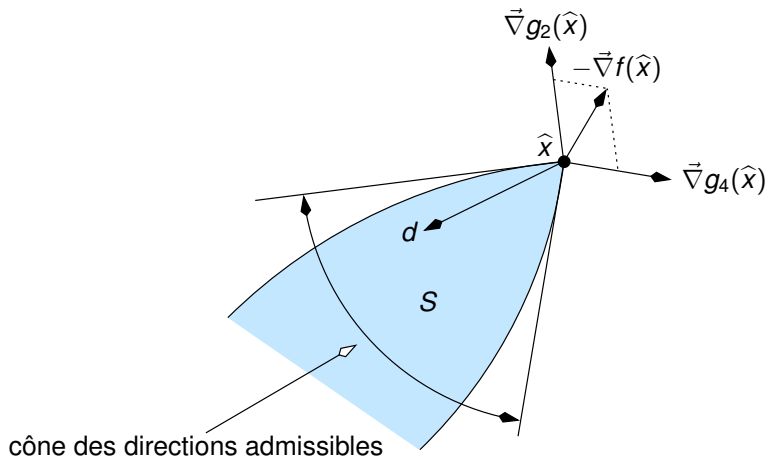
$$\text{si } f(x_1, x_2) = -x_1 \text{ alors } \hat{x} = (1, 0), I(\hat{x}) = \{2, 3\}$$

$$\vec{\nabla} f(\hat{x}) = (-1, 0)$$

$$\lambda_2 \vec{\nabla} g_2(\hat{x}) + \lambda_3 \vec{\nabla} g_3(\hat{x}) = \lambda_2(0, 1) + \lambda_3(0, -1) = (0, \lambda_2 - \lambda_3)$$

$$\implies \vec{\nabla} f(\hat{x}) \neq \lambda_2 \vec{\nabla} g_2(\hat{x}) + \lambda_3 \vec{\nabla} g_3(\hat{x})$$

Interprétation géométrique de Kuhn et Tucker



Kuhn et Tucker : $-\vec{\nabla}f(\hat{x}) =$ combinaison linéaire (à coeffs positifs) des $\vec{\nabla}g_j(\hat{x})$

$\implies -\vec{\nabla}f(\hat{x})$ forme un angle obtus avec toute direction admissible

transparents précédents :

conditions de Kuhn et Tucker = conditions nécessaires

sous certaines hypothèses de convexité : conditions suffisantes

⇒ on va s'intéresser maintenant à des propriétés de convexité

Proposition 2

- f de classe C^1
- $S \subseteq \mathbb{R}^n$: ensemble **convexe**
- f convexe sur $S \iff \forall x, y \in S :$

$$f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x)$$

Démonstration :

Supposons f convexe sur S

$$\implies f(\lambda y + (1 - \lambda)x) \leq \lambda f(y) + (1 - \lambda)f(x) \quad \forall 0 \leq \lambda \leq 1$$

$$\implies \forall 0 < \lambda \leq 1, \quad \frac{f(\lambda y + (1 - \lambda)x) - f(x)}{\lambda} \leq f(y) - f(x)$$

$$\begin{aligned} \text{Or } \lim_{\lambda \downarrow 0} \frac{f(\lambda y + (1 - \lambda)x) - f(x)}{\lambda} &= \lim_{\lambda \downarrow 0} \frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \\ &= \vec{\nabla} f(x)^T (y - x) \end{aligned}$$

$$\implies \vec{\nabla} f(x)^T (y - x) \leq f(y) - f(x)$$

$$\implies f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x)$$

Réciproque :

$$\text{supp } f(y) \geq f(x) + \vec{\nabla} f(x)^T (y - x) \quad \forall x, y \in S$$

Soit 2 points $x_1, x_2 \in S$ et $x = \lambda x_1 + (1 - \lambda)x_2$, $\lambda \in [0, 1]$

$$\implies \begin{cases} f(x_1) \geq f(x) + \vec{\nabla} f(x)^T (x_1 - x) \\ f(x_2) \geq f(x) + \vec{\nabla} f(x)^T (x_2 - x) \end{cases}$$

$$\begin{aligned} \implies \lambda f(x_1) + (1 - \lambda)f(x_2) &\geq f(x) + \vec{\nabla} f(x)^T [\lambda(x_1 - x) + (1 - \lambda)(x_2 - x)] \\ &\geq f(x) + \vec{\nabla} f(x)^T [\lambda x_1 + (1 - \lambda)x_2 - x] \\ &\geq f(x) = f(\lambda x_1 + (1 - \lambda)x_2) \end{aligned}$$

$\implies f$ convexe sur S

Rappel : Proposition 1 : minimum local

- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1
 - \hat{x} minimum local de f sur S
- $\implies \forall$ direction admissible d en \hat{x} , on a $\vec{\nabla}f(\hat{x})^T d \geq 0$

Proposition 3 : minimum global

- $S \subseteq \mathbb{R}^n$: ensemble **convexe**
- $f : \mathbb{R}^n \mapsto \mathbb{R}$ de classe C^1 **convexe sur S**
- \hat{x} minimum **global** de f sur $S \iff \forall y \in S, \vec{\nabla}f(\hat{x})^T (y - \hat{x}) \geq 0$

Démonstration :

\hat{x} minimum global de $f \implies \hat{x}$ minimum local de f

Proposition 1 $\implies \vec{\nabla} f(\hat{x})^T d \geq 0 \quad \forall$ direction d admissible en \hat{x}

S convexe $\implies (y \in S \implies (y - \hat{x})$ admissible)

$\implies \forall y \in S, \vec{\nabla} f(\hat{x})^T (y - \hat{x}) \geq 0$

Réciproque : supp que $\vec{\nabla} f(\hat{x})^T (y - \hat{x}) \geq 0 \quad \forall y \in S$

f convexe $\implies f(y) \geq f(\hat{x}) + \vec{\nabla} f(\hat{x})^T (y - \hat{x})$ d'après proposition 2

$\implies f(y) \geq f(\hat{x})$

$\implies \hat{x} =$ optimum global de f sur S

Rappel : Conditions de Kuhn et Tucker

- f de classe C^1
- $I(x^*) = \{\text{indices } i \text{ t.q. } g_i(x^*) = 0\}$
- $D(x^*) = \{d : \vec{\nabla} g_i(x^*)^T d \leq 0, i \in I(x^*)\}$
- \hat{x} = point de S tel que $D(\hat{x}) = \{\text{directions admissibles en } \hat{x}\}$
- alors condition nécessaire pour que \hat{x} = minimum local de f :
 $\exists \lambda_i \geq 0, i = 1, \dots, m$, tels que :

$$\textcircled{1} \quad \vec{\nabla} f(\hat{x}) + \sum_{i=1}^m \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$$

$$\textcircled{2} \quad \lambda_i g_i(\hat{x}) = 0, \quad \forall i \in \{1, \dots, m\}$$

Conditions suffisantes d'optimalité de Kuhn et Tucker

- g_i **convexes** de classe C^1 , $i = 1, \dots, m$
- $S = \{x : g_i(x) \leq 0, i = 1, \dots, m\}$
- f de classe C^1 **convexe** sur S
- \hat{x} = point de S tel que $\exists \lambda_i \geq 0, i = 1, \dots, m$, tels que :

$$\textcircled{1} \quad \vec{\nabla} f(\hat{x}) + \sum_{i=1}^m \lambda_i \vec{\nabla} g_i(\hat{x}) = 0$$

$$\textcircled{2} \quad \lambda_i g_i(\hat{x}) = 0, \quad \forall i \in \{1, \dots, m\}$$

$\implies \hat{x}$ = minimum **global** de f sur S

Démonstration :

Soit $y \in S$, $y \neq \hat{x}$. On veut montrer que $f(y) \geq f(\hat{x})$

les g_i convexes $\implies S$ convexe

$\implies \hat{d} = y - \hat{x}$ direction admissible en \hat{x}

$\implies \vec{\nabla} g_i(\hat{x})^T \hat{d} \leq 0 \quad \forall i \in I(\hat{x})$

Or conditions ① et ② de Kuhn et Tucker équivalentes à :

$\vec{\nabla} f(\hat{x})^T d \geq 0 \quad \forall$ direction d telle que $\vec{\nabla} g_i(\hat{x})^T d \leq 0 \quad \forall i \in I(\hat{x})$

On sait que \hat{d} vérifie $\vec{\nabla} g_i(\hat{x})^T \hat{d} \leq 0 \quad \forall i \in I(\hat{x})$

Hypothèse de la proposition $\implies \vec{\nabla} f(\hat{x})^T \hat{d} = f(\hat{x})^T (y - \hat{x}) \geq 0$

Or Prop. 3 : \hat{x} minimum global de $f \iff \forall y \in S, \vec{\nabla} f(\hat{x})^T (y - \hat{x}) \geq 0$

$\implies \hat{x}$ minimum global de f

Définition de la fonction de Lagrange

- *Problème d'origine :*

$$\min f(x)$$

$$\text{s.c. } g_i(x) \leq 0, \quad i = 1, \dots, m$$

- *Fonction de Lagrange :*

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

- $\lambda_i \geq 0$: *multiplicateur de Lagrange*

$$\text{Fonction de Lagrange : } L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$$

Fonction primale / duale

- Fonction primale : $L^*(x) = \max_{\lambda \geq 0} L(x, \lambda)$
- Fonction duale : $L_*(\lambda) = \min_x L(x, \lambda)$

Problème primal / dual

- Problème primal : $L^*(\hat{x}) = \min_x L^*(x) = \min_x \max_{\lambda \geq 0} L(x, \lambda)$
- Problème dual : $L_*(\hat{\lambda}) = \max_{\lambda \geq 0} L_*(\lambda) = \max_{\lambda \geq 0} \min_x L(x, \lambda)$

$$\begin{aligned} L^*(x) &= \max_{\lambda \geq 0} L(x, \lambda) \\ &= \max_{\lambda \geq 0} \left(f(x) + \sum_{i=1}^m \lambda_i g_i(x) \right) \\ &= \begin{cases} f(x) & \text{si } g_i(x) \leq 0 \forall i \\ +\infty & \text{si } g_i(x) > 0 \text{ pour un } i \end{cases} \end{aligned}$$

Problème primal : $L^*(\hat{x}) = \min_x L^*(x) \implies$ ne s'intéresser qu'aux x tels que $g_i(x) \leq 0$

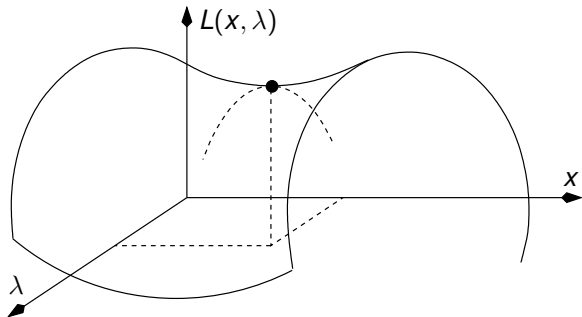
\implies le problème devient alors : $\begin{cases} \min f(x) \\ \text{s.c. } g_i(x) \leq 0, \quad i = 1, \dots, m \end{cases}$

Définition d'un point selle

- Soit $\bar{x} \in S$, $\bar{\lambda} \geq 0$
- $(\bar{x}, \bar{\lambda})$ est un point selle de $L(x, \lambda)$ si :

$$L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x \in S$$

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \quad \forall \lambda \geq 0$$



Proposition 4 : Caractérisation d'un point Selle

- Soit $\bar{x} \in S$, $\bar{\lambda} \geq 0$
- $(\bar{x}, \bar{\lambda})$ est un point selle de $L(x, \lambda)$ si et seulement si :
 - 1 $L(\bar{x}, \bar{\lambda}) = \min_{x \in S} L(x, \bar{\lambda})$
 - 2 $g_i(\bar{x}) \leq 0 \forall i \in \{1, \dots, m\}$
 - 3 $\bar{\lambda}_i g_i(\bar{x}) = 0 \forall i \in \{1, \dots, m\}$

Démonstration :

$\text{supp } (\bar{x}, \bar{\lambda}) \text{ point selle} \implies L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x, \forall \lambda \geq 0$

$L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x \implies L(\bar{x}, \bar{\lambda}) = \min_x L(x, \bar{\lambda}) \implies \textcircled{1}$

$$\begin{aligned} L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) &\iff f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x}) \leq f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i g_i(\bar{x}) \\ &\iff \sum_{i=1}^m (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) \leq 0 \quad \forall \lambda \geq 0 \end{aligned}$$

Si $\exists i$ t.q. $g_i(\bar{x}) > 0$, alors λ_i suffisamment grand

$\implies \sum_{i=1}^m (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) > 0 \implies \text{impossible} \implies \textcircled{2} \text{ est vrai}$

Caractérisation d'un point Selle (3/4)

$$L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \iff \sum_{i=1}^m (\lambda_i - \bar{\lambda}_i) g_i(\bar{x}) \leq 0 \quad \forall \lambda \geq 0$$

$$\lambda = 0 \implies \sum_{i=1}^m \bar{\lambda}_i g_i(\bar{x}) \geq 0$$

$$\text{Or } \bar{\lambda} \geq 0 \text{ et } g_i(\bar{x}) \leq 0 \implies \sum_{i=1}^m \bar{\lambda}_i g_i(\bar{x}) \leq 0$$

$$\implies \sum_{i=1}^m \bar{\lambda}_i g_i(\bar{x}) = 0 \implies \bar{\lambda}_i g_i(\bar{x}) = 0 \quad \forall i \implies \textcircled{3}$$

Réciproque :

Supposons ①, ② et ③

$$\textcircled{1} \implies L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \quad \forall x \in S$$

$$\textcircled{3} \implies L(\bar{x}, \bar{\lambda}) = f(\bar{x})$$

$$L(\bar{x}, \lambda) = f(\bar{x}) + \sum_{i=1}^m \lambda_i g_i(\bar{x}) \leq f(\bar{x}) = L(\bar{x}, \bar{\lambda}) \quad \forall \lambda \geq 0$$

$$\implies L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda}) \implies (\bar{x}, \bar{\lambda}) \text{ point selle}$$

Condition d'optimalité du point selle

Si $(\hat{x}, \hat{\lambda})$ point selle de la fonction de Lagrange, alors :

$$\bullet \min_x L(x, \hat{\lambda}) = L(\hat{x}, \hat{\lambda}) = \max_{\lambda \geq 0} L(\hat{x}, \lambda)$$

$$\bullet \hat{x} = \text{solution de } \begin{cases} \min f(x) \\ \text{s.c. } g_i(x) \leq 0, \quad i = 1, \dots, m \end{cases}$$

Démonstration :

$$(\hat{x}, \hat{\lambda}) \text{ point selle} \implies L(\hat{x}, \lambda) \leq L(\hat{x}, \hat{\lambda}) \leq L(x, \hat{\lambda}) \quad \forall x, \forall \lambda \geq 0$$

$$L(\hat{x}, \lambda) \leq L(\hat{x}, \hat{\lambda}) \quad \forall \lambda \geq 0 \implies L(\hat{x}, \hat{\lambda}) = \max_{\lambda \geq 0} L(\hat{x}, \lambda)$$

$$L(\hat{x}, \hat{\lambda}) \leq L(x, \hat{\lambda}) \quad \forall x \implies L(\hat{x}, \hat{\lambda}) = L(\hat{x}, \hat{\lambda}) = \min_x L(x, \hat{\lambda})$$

$$L(\hat{x}, \hat{\lambda}) \leq L(x, \hat{\lambda}) \quad \forall x \iff f(\hat{x}) + \sum_{i=1}^m \hat{\lambda}_i g_i(\hat{x}) \leq f(x) + \sum_{i=1}^m \hat{\lambda}_i g_i(x)$$

Or ③ de la proposition 4 : $\hat{\lambda}_i g_i(\hat{x}) = 0 \quad \forall i \in \{1, \dots, m\}$

$$\implies f(\hat{x}) \leq f(x) + \sum_{i=1}^m \hat{\lambda}_i g_i(x) \leq f(x) \quad \forall x \in S$$

$\implies \hat{x} = \text{optimum global de } f \text{ sur } S$