

## C11, Optimisation avec contrainte

**Objectif :**  $f \in C(\mathbb{R}^n, \mathbb{R})$ ,  $K$  un sous ensemble fermé de  $\mathbb{R}^n$ . On s'intéresse aux 3 questions suivantes :

1. Existence de  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .  
(On dit alors que  $x \in \operatorname{argmin}_K f$ .)
2. Unicité de  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .  
(Si il y a existence et unicité, on dit alors que  $x = \operatorname{argmin}_K f$ .)
3. Comment calculer ce point  $\bar{x}$  de  $K$  minimisant  $f$  sur  $K$ .

C'est un problème de minimisation avec contrainte car la variable est contrainte à être dans  $K$ , sous ensemble de  $\mathbb{R}^n$ .

On peut remplacer  $f \in C(\mathbb{R}^n, \mathbb{R})$  par  $f \in C(K, \mathbb{R})$

# Existence de $\bar{x}$

## Théorème

$f \in C(\mathbb{R}^n, \mathbb{R})$ .

1. *L'ensemble  $K$  est fermé, borné, non vide.*

*Alors, il existe  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .*

2. *L'ensemble  $K$  est fermé non vide et  $f(x) \rightarrow +\infty$  quand  $\|x\| \rightarrow +\infty$ .*

*Alors, il existe  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .*

## Existence de $\bar{x}$ , démonstration, premier cas

$f \in C(\mathbb{R}^n, \mathbb{R})$ . L'ensemble  $K$  est fermé, borné, non vide.

Dans ce cas, l'ensemble  $K$  est compact et, comme  $f$  est continue,  $f$  est bornée sur  $K$  et atteint ses bornes. Donc, il existe  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

## Existence de $\bar{x}$ , démonstration, deuxième cas

L'ensemble  $K$  est fermé non vide et  $f(x) \rightarrow +\infty$  quand  $\|x\| \rightarrow +\infty$ .

Soit  $x_0 \in K$ . Il existe  $R > 0$  tel que

$$\|x\| > R \implies f(x) > f(x_0).$$

On pose  $B_R = \{x \in \mathbb{R}^n, \|x\| \leq R\}$ . **Noter que  $x_0 \in B_R \cap K$ .**

Comme  $B_R \cap K$  est compact et  $f$  est continue, il existe  $\bar{x} \in B_R \cap K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in B_R \cap K$ .

En particulier,  $f(\bar{x}) \leq f(x_0)$  et donc

$$f(\bar{x}) \leq f(x) \text{ pour tout } x \in K.$$

On a bien trouvé  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

# Unicité de $\bar{x}$

## Définition

*L'ensemble  $K$  est convexe si*

$$x, y \in K, t \in [0, 1] \implies tx + (1 - t)y \in K.$$

## Théorème

*$f \in C(\mathbb{R}^n, \mathbb{R})$ . On suppose  $f$  strictement convexe et  $K$  convexe. Alors, il existe au plus un point  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .*

**Démonstration :** Soient  $\bar{x} \in K$  et  $\bar{\bar{x}} \in K$  tels que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$  et  $f(\bar{\bar{x}}) \leq f(x)$  pour tout  $x \in K$ .

On a donc  $f(\bar{x}) = f(\bar{\bar{x}})$ . Si  $\bar{x} \neq \bar{\bar{x}}$ , par strict convexité de  $f$ ,

$$f\left(\frac{\bar{x} + \bar{\bar{x}}}{2}\right) < \frac{1}{2}(f(\bar{x}) + f(\bar{\bar{x}})) = f(\bar{x}),$$

et, par convexité de  $K$ ,  $\frac{\bar{x} + \bar{\bar{x}}}{2} \in K$ , ce qui est impossible. Donc  $\bar{x} = \bar{\bar{x}}$ .

# Existence et unicité de $\bar{x}$

On regroupe le théorème d'existence et le théorème d'unicité :

## Théorème

$f \in C(\mathbb{R}^n, \mathbb{R})$ . On suppose  $f$  strictement convexe et  $K$  convexe fermé non vide. On suppose  $K$  borné ou  $f(x) \rightarrow +\infty$  quand  $\|x\| \rightarrow +\infty$ .

Alors, il existe un unique  $\bar{x} \in K$  tel que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

# Condition nécessaire d'optimalité

On suppose que  $f \in C^1(\mathbb{R}^n, \mathbb{R})$ .

## Proposition

Soit  $\bar{x} \in K$  tel. que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$

1. Si  $\bar{x} \in \overset{\circ}{K}$ , alors  $\nabla f(\bar{x}) = 0$ .
2. Si  $K$  convexe,  $\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq 0$  pour tout  $x \in K$ .

Démonstration du premier cas Si  $\bar{x} \in \overset{\circ}{K}$ , il existe  $\varepsilon > 0$  tel que

$$\|\bar{x} - z\| \leq \varepsilon \implies z \in K.$$

Soit  $y \in \mathbb{R}^n$ ,  $y \neq 0$ , on a  $\bar{x} + ty \in K$  si  $-\frac{\varepsilon}{\|y\|} \leq t \leq \frac{\varepsilon}{\|y\|}$  et donc, en posant  $\varphi(t) = f(\bar{x} + ty)$ ,

$$\varphi(0) = f(\bar{x}) \leq \varphi(t) \text{ si } -\frac{\varepsilon}{\|y\|} \leq t \leq \frac{\varepsilon}{\|y\|}.$$

On en déduit  $\varphi'(0) = 0$ , c'est-à-dire

$\varphi'(0) = \nabla f(\bar{x}) \cdot y = 0$  pour tout  $y \in \mathbb{R}^n$  et donc  $\nabla f(\bar{x}) = 0$ .

## Condition nécessaire d'optimalité, deuxième cas

$f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $K$  convexe,  $\bar{x} \in K$  tel. que  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$

Soit  $x \in K$ . par convexité de  $K$ ,  $\bar{x} + t(x - \bar{x}) \in K$  si  $t \in ]0, 1[$ . Donc,

$$\frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t} \geq 0 \text{ pour tout } 0 < t < 1.$$

quand  $t \rightarrow 0$  ceci donne

$$\nabla f(\bar{x}) \cdot (x - \bar{x}) \geq 0.$$

$K$  défini par de contraintes "égalités"

$$f \in C^1(\mathbb{R}^n, \mathbb{R}), n \geq 1$$

$$g_i \in C^1(\mathbb{R}^n, \mathbb{R}), i \in \{1, \dots, p\} (p \geq 1).$$

$$K = \{x \in \mathbb{R}^n; g_i(x) = 0, i = 1, \dots, p\}.$$

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_p \end{bmatrix} \in C^1(\mathbb{R}^n, \mathbb{R}^p)$$

$$K = \{x \in \mathbb{R}^n; g(x) = 0\}.$$

# Multiplicateurs de lagrange

## Théorème

Soit  $f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ .

$K = \{x \in \mathbb{R}^n; g(x) = 0\}$ .

$\bar{x} \in K$ ,  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

On suppose que  $\text{rang}(J_g(\bar{x})) = p$  (et donc  $p \leq n$ )

Alors, il existe  $\lambda_1, \dots, \lambda_p$  dans  $\mathbb{R}$  tels que

$$\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0.$$

Rappel :  $J_g(x) \in \mathcal{M}_{p,n}$  et  $L_i(J_g(x)) = (\nabla g_i(x))^t$

Le cas  $p = n$  est facile car  $\{\nabla g_1(\bar{x}), \dots, \nabla g_n(\bar{x})\}$  est alors une base de  $\mathbb{R}^n$

## Courbes de niveau d'une fonction, $n = 2$

$h \in C^1(\mathbb{R}^2, \mathbb{R})$ ,  $c \in \mathbb{R}$ ,  $\Gamma_{h,c} = \{x \in \mathbb{R}^2; h(x) = c\}$ .

Paramétrisation locale de  $\Gamma_{h,c}$

$x \in C^1(]-1, 1[, \mathbb{R}^2)$ ,  $\{x(t); t \in ]-1, 1[ \} \subset \Gamma_{h,c}$ . Alors

$$h(x(t)) = c \text{ pour tout } t \in ]-1, 1[,$$

et donc  $\nabla h(x(t)) \cdot x'(t) = 0$  pour tout  $t \in ]-1, 1[$ .

Le vecteur  $\nabla h(x(0))$  est orthogonal à la courbe  $\Gamma_{h,c}$  au point  $x(0)$ .

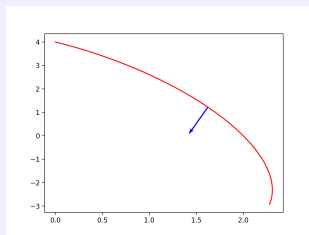


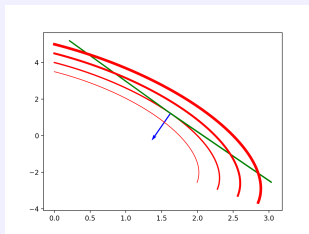
Figure: Courbe de niveau de  $h$  en rouge,  $\nabla h(x(0))$  en bleu

# Multiplicateur de Lagrange, $n = 2$ , $p = 1$

Dans cette figures les courbes de niveau de  $f$  sont en rouge. L'épaisseur de la courbe est liée à la valeur de  $f$ .

La courbe de niveau de  $g$ , c'est-à-dire  $\Gamma_{g,0}$ , est en vert.

Lorsque que l'on parcourt la courbe de niveau de  $g$ , la fonction  $f$  atteint son minimum (en  $\bar{x}$ ) lorsque les courbes de niveau de  $f$  et  $g$  sont tangentes, c'est-à-dire lorsque leurs vecteurs normaux  $\nabla g(\bar{x})$  ou  $\nabla f(\bar{x})$  sont colinéaires, et il existe donc  $\lambda$  tel que  $\nabla f(\bar{x}) = -\lambda \nabla g(\bar{x})$  (si  $\nabla g(\bar{x}) \neq 0$ ).



Rappel  $K = \Gamma_{g,0}$ . On cherche l'argmin de  $f$  sur  $K$

## Multiplicateur de Lagrange, $n = 2$ , $p = 1$ , autre méthode

On peut aussi paramétriser  $\Gamma_{g,0}$  au voisinage de  $\bar{x}$

Soit  $x \in C^1(]-1, 1[, \mathbb{R}^2)$ ,  $\{x(t); t \in ]-1, 1[ \} \subset \Gamma_{g,0}$  et  $x(0) = \bar{x}$ .

La fonction  $t \mapsto f(x(t))$  est minimale en  $t = 0$ . Sa dérivée est nulle en 0:

$$\nabla f(\bar{x}) \cdot x'(0) = 0.$$

Mais on sait aussi que  $\nabla g(x(0)) \cdot x'(0) = 0$ .

Ce qui prouve encore, si  $x'(0) \neq 0$ , que  $\nabla g(\bar{x})$  et  $\nabla f(\bar{x})$  sont colinéaires,

# Multiplicateur de Lagrange, démonstration (1)

Hypothèses :  $f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ .  $K = \{x \in \mathbb{R}^n; g(x) = 0\}$ .  
 $\bar{x} \in K$ ,  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .  $\text{rang}(J_g(\bar{x})) = p < n$ .

**Etape 1 :** Il existe  $G$  s.e.v de  $\mathbb{R}^n$  tel que  $\mathbb{R}^n = G \oplus \ker(J_g(\bar{x}))$  et  $J_g(\bar{x})$  bijection de  $G$  sur  $\text{Im}(J_g(\bar{x}))$  (exercice 1 du td1, démonstration du théorème du rang).

Comme  $\text{rang}(J_g(\bar{x})) = p$ ,  $\text{Im}(J_g(\bar{x})) = \mathbb{R}^p$ . Donc  $J_g(\bar{x})$  est une bijection de  $G$  sur  $\mathbb{R}^p$ .

on pose  $F = \ker(J_g(\bar{x}))$ .

Pour  $(y, z) \in G \times F$ , on pose  $\bar{f}(y, z) = f(y + z)$  et  $\bar{g}(y, z) = g(y + z)$   
 $\bar{x} = \bar{y} + \bar{z}$ .

Intérêt :  $D_1\bar{g}(y, z)(h) = Dg(y + z)(h) = J_g(y + z)h$  pour tout  $h \in G$  et donc  $D_1\bar{g}(\bar{y}, \bar{z})$  est inversible de  $G$  dans  $\mathbb{R}^p$ .

# Multiplicateur de Lagrange, démonstration (2)

Etape 2 :

$\bar{K} = \{(y, z) \in G \times F; y + z \in K\} = \{(y, z) \in G \times F; \bar{g}(y, z) = 0\}$ .

$(\bar{y}, \bar{z}) \in \bar{K}$ ,  $D_1 \bar{g}(\bar{y}, \bar{z})$  est inversible. Le théorème des fonctions implicites donne alors qu'il existe  $\varepsilon > 0$ ,  $\delta > 0$  tels que

$$\forall z \in B(\bar{z}, \varepsilon), \exists ! y \in B(\bar{y}, \delta) \text{ tel que } \bar{g}(y, z) = 0.$$

On note  $\phi(z)$  cette valeur de  $y$ . Le théorème donne aussi  $\phi \in C^1(F, G)$ .

$B(\bar{z}, \varepsilon)$  est la boule ouverte (de  $F$ ) de centre  $\bar{z}$  et de rayon  $\varepsilon$ .

$B(\bar{y}, \delta)$  est la boule ouverte (de  $G$ ) de centre  $\bar{y}$  et de rayon  $\delta$ .

Comme  $\bar{g}(\phi(z), z) = 0$  :

$$D_1 \bar{g}(\bar{y}, \bar{z}) \circ D\phi(\bar{z}) + D_2 \bar{g}(\bar{y}, \bar{z}) = 0$$

c'est-à-dire  $D\phi(\bar{z}) = -(D_1 \bar{g}(\bar{y}, \bar{z}))^{-1} \circ D_2 \bar{g}(\bar{y}, \bar{z})$ .

Noter  $D_2 \bar{g}(\bar{y}, \bar{z}) \in \mathcal{L}(F, \mathbb{R}^p)$  et  $D_1 \bar{g}(\bar{y}, \bar{z}) \in \mathcal{L}(G, \mathbb{R}^p)$

## Multiplicateur de Lagrange, démonstration (3)

$\bar{f}(\phi(\bar{z}), \bar{z}) \leq \bar{f}(\phi(z), z)$  pour tout  $z \in B(\bar{z}, \varepsilon)$

car  $\phi(z) + z \in K$  et  $\phi(\bar{z}) + \bar{z} = \bar{y} + \bar{z} = \bar{x}$ . Donc (min. sans contrainte)

$$D_1 \bar{f}(\bar{y}, \bar{z}) \circ D\phi(\bar{z}) + D_2 \bar{f}(\bar{y}, \bar{z}) = 0,$$

et donc, comme  $D\phi(\bar{z}) = -(D_1 \bar{g}(\bar{y}, \bar{z}))^{-1} \circ D_2 \bar{g}(\bar{y}, \bar{z})$ ,

$$D_2 \bar{f}(\bar{y}, \bar{z}) = D_1 \bar{f}(\bar{y}, \bar{z}) \circ (D_1 \bar{g}(\bar{y}, \bar{z}))^{-1} \circ D_2 \bar{g}(\bar{y}, \bar{z}).$$

Mais on a aussi

$$D_1 \bar{f}(\bar{y}, \bar{z}) = D_1 \bar{f}(\bar{y}, \bar{z}) \circ (D_1 \bar{g}(\bar{y}, \bar{z}))^{-1} \circ D_1 \bar{g}(\bar{y}, \bar{z}).$$

Ceci donne

$$Df(\bar{x}) = -\Lambda \circ Dg(\bar{x}),$$

avec  $\Lambda = -D_1 \bar{f}(\bar{y}, \bar{z}) \circ (D_1 \bar{g}(\bar{y}, \bar{z}))^{-1} \in \mathcal{L}(\mathbb{R}^p, \mathbb{R})$ ,  $Dg(\bar{x}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ .

Ceci donne bien  $\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0$ , où les  $\lambda_i$  sont les composantes de la matrice représentant  $\Lambda$ .

## Rappel de calcul différentiel

$$Df(\bar{x}) = -\Lambda \circ Dg(\bar{x}),$$

$$Df(\bar{x}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}), \Lambda \in \mathcal{L}(\mathbb{R}^p, \mathbb{R}), Dg(\bar{x}) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p).$$

Soit  $h \in \mathbb{R}^n$ ,  $u = Dg(\bar{x})(h) = J_g(\bar{x})h \in \mathbb{R}^p$ .

pour  $i \in \{1, \dots, p\}$ , on note  $u_i$  la  $i$ -eme composante de  $u$

$$u_i = L_i(J_g(\bar{x}))^t \cdot h = \nabla g_i(\bar{x}) \cdot h.$$

$$\Lambda \circ Dg(\bar{x})(h) = \Lambda(u) = \sum_{i=1}^p \lambda_i u_i = \left( \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) \right) \cdot h.$$

$$Df(\bar{x})(h) = \nabla f(\bar{x}) \cdot h.$$

$Df(\bar{x}) = -\Lambda \circ Dg(\bar{x})$  signifie

$$\nabla f(\bar{x}) + \sum_{i=1}^p \lambda_i \nabla g_i(\bar{x}) = 0.$$

# Utilisation pratique des multiplicateurs de Lagrange

$f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ .  $K = \{x \in \mathbb{R}^n; g(x) = 0\}$ .

On cherche à calculer  $\bar{x}$  tel que

$\bar{x} \in K$ ,  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

Une possibilité est donc de résoudre le système de  $(n + p)$  équations à  $(n + p)$  inconnues ( $\bar{x}$  dans  $\mathbb{R}^n$  et  $\lambda_1, \dots, \lambda_p$ ) :

$$\begin{aligned}\nabla f(x) + \sum_{i=1}^p \lambda_i \nabla g_i(x) &= 0 \text{ (} n \text{ équations),} \\ g(x) &= 0 \text{ (} p \text{ équations).}\end{aligned}$$

Il reste alors deux difficultés :

1.  $x$  solution de ce système n'est pas forcément un point où  $f$  réalise son minimum sur  $K$ .
2. Si, pour le point recherché,  $\text{rang}(J_g(\bar{x})) < p$ , le point  $\bar{x}$  peut ne pas être solution de ce système.

Toutefois, cette méthode est souvent très performante.

$K$  défini par de contraintes “inégalités”

$$f \in C^1(\mathbb{R}^n, \mathbb{R}), n \geq 1$$

$$g_i \in C^1(\mathbb{R}^n, \mathbb{R}), i \in \{1, \dots, p\} (p \geq 1).$$

$$K = \{x \in \mathbb{R}^n; g_i(x) \leq 0, i = 1, \dots, p\}.$$

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_p \end{bmatrix} \in C^1(\mathbb{R}^n, \mathbb{R}^p)$$

$$K = \{x \in \mathbb{R}^n; g(x) \leq 0\}.$$

# Théorème de Kuhn-Tucker

## Théorème

Soit  $f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ .

$K = \{x \in \mathbb{R}^n; g(x) \leq 0\}$ .

$\bar{x} \in K$ ,  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

On note  $I(\bar{x}) = \{i \in \{1, \dots, p\}; g_i(\bar{x}) = 0\}$  et on suppose que

$\{\nabla g_i(\bar{x}), i \in I(\bar{x})\}$  est une famille libre.

Alors, il existe  $\{\lambda_i, i \in I(\bar{x})\}$  inclus dans  $\mathbb{R}_+$  telle que

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Remarque : Si  $I(\bar{x}) = \emptyset$ , c'est-à-dire  $g_i(\bar{x}) < 0$  pour tout  $i \in \{1, \dots, p\}$ , on est alors ramené à un problème de minimisation sans contrainte et donc  $\nabla f(\bar{x}) = 0$ .

# Utilisation pratique du théorème de Kuhn-Tucker

$f \in C^1(\mathbb{R}^n, \mathbb{R})$ ,  $g \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ ,  $K = \{x \in \mathbb{R}^n; g(x) \leq 0\}$ .

On cherche à calculer  $\bar{x}$  tel que

$\bar{x} \in K$ ,  $f(\bar{x}) \leq f(x)$  pour tout  $x \in K$ .

Une possibilité est de résoudre le système de  $(n + p)$  équations à  $(n + p)$  inconnues ( $\bar{x}$  dans  $\mathbb{R}^n$  et  $\lambda_1, \dots, \lambda_p$ ) avec  $2p$  inégalités :

$$\begin{aligned}\nabla f(x) + \sum_{i=1}^p \lambda_i \nabla g_i(x) &= 0 \quad (n \text{ équations}), \\ \lambda_i g_i(x) &= 0, \quad i \in \{1, \dots, p\} \quad (p \text{ équations}), \\ g_i(x) &\leq 0, \quad i \in \{1, \dots, p\}, \\ \lambda_i &\geq 0, \quad i \in \{1, \dots, p\}.\end{aligned}$$

Ici encore, cette méthode est souvent très performante.