Elements d’Optimisation

Olivier Brun

brun@laas.fr

LAAS-CNRS
7 Av. Colonel Roche, 31077 Toulouse, France.

INSA, 2010.

Outline

Introduction

- **Objectifs**
  - Rappeler quelques éléments de base de la théorie de l'optimisation.
  - La présentation évite de rentrer dans les détails mathématiques et se contente d'introduire quelques notions nécéssaires à la bonne compréhension du cours.
  - Optimisation convexe, programmation linéaire, programmation linéaire mixte, recherche locale.

Optimisation Convexe

Ensembles convexes

- Un ensemble $\mathcal{K} \subset \mathbb{R}^n$ est dit convexe si le segment joignant deux points $\mathbf{x}, \mathbf{y} \in \mathcal{K}$ appartient à $\mathcal{K}$

$$\alpha \mathbf{x} + (1 - \alpha) \mathbf{y} \in \mathcal{K} \quad \forall \alpha \in [0, 1], \ \forall \mathbf{x}, \mathbf{y} \in \mathcal{K}$$

  - Exemples : boule unité $\mathcal{K} = \{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\| \leq 1\}$, ellipsoides, polyhédres,...
  - Les ensembles convexes de $\mathbb{R}$ sont les intervalles.
  - L'intersection d'ensembles convexes est un ensemble convexe.

Fonctions convexes

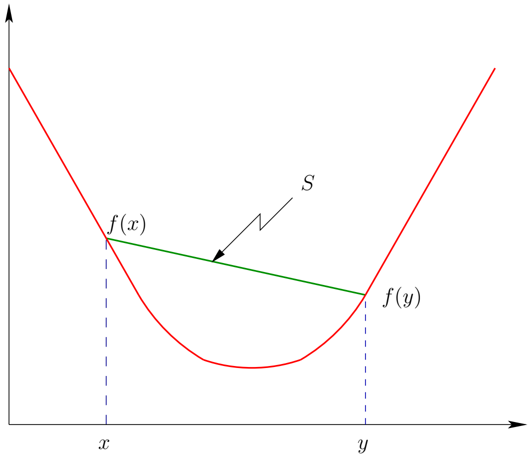
- Etant donné un ensemble convexe $\mathcal{K} \subset \mathbb{R}^n$, une fonction $f(\mathbf{x}) : \mathcal{K} \rightarrow \mathbb{R}$ est dite convexe sur $\mathcal{K}$ si

$$f\left(\alpha\,\mathbf{x} + (1-\alpha)\,\mathbf{y}\right) \leq \alpha\,f(\mathbf{x}) + (1-\alpha)\,f(\mathbf{y}) \quad \forall \alpha \in (0,1),\ \forall \mathbf{x}, \mathbf{y} \in \mathcal{K}$$

  - On dit que f est strictement convexe si l'inégalité est stricte.
  - Exemple : les fonctions linéaires $f(\mathbf{x}) = \sum_i c_i x_i$ sont convexes, $e^{\mathbf{x}}$ est strictement convexe...
  - L'addition, la multiplication par un scalaire et l'opérateur max préservent la convexité.

Fonctions convexes

- **Fonctions convexes** : tout segment S joignant les points $(\mathbf{x}, f(\mathbf{x}))$ et $(\mathbf{y}, f(\mathbf{y}))$ se trouve “au dessus” de la courbe de f.



Optimisation Convexe

- **Problème d'optimisation :**

$$\begin{aligned} \text{Minimiser} \quad & f(\mathbf{x}) \\ \text{s.t} \quad & \mathbf{x} \in \mathcal{K} \end{aligned}$$

- **Optimalité :** un point admissible $\mathbf{x}^* \in \mathcal{K}$ est dit optimal si $f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in \mathcal{K}$.
  - On parle d'optimum local si l'inégalité n'est vraie que dans un voisinage de $\mathbf{x}^*$.
  - Unicité du minimum global si f strictement convexe.

- **Optimisation convexe :**
  - $\mathcal{K}$ est un ensemble fermé et convexe,
  - f est convexe et continuellement différentiable sur $\mathcal{K}$.
  - Tout minimum local est un minimum global.

Conditions d'optimalité

- Gradient de f au point $\mathbf{x}$: $\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_i}(\mathbf{x})\right)^T_{i=1,\ldots,n}$.
- Produit scalaire des vecteurs $\mathbf{d}$ et $\nabla f(\mathbf{x})$:

$$\mathbf{d}^T \nabla f(\mathbf{x}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x})\, d_i$$

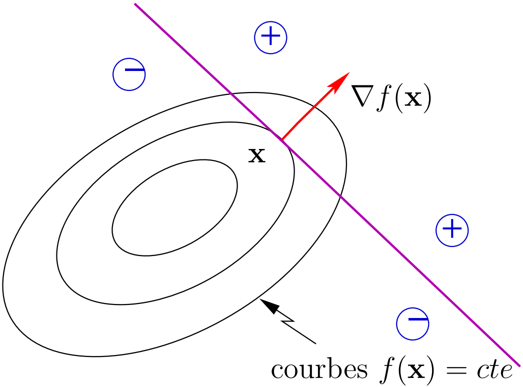
- Développement en série de Taylor (premier ordre)

$$\mathbf{y} = \mathbf{x} + \alpha \mathbf{d} \Longrightarrow f(\mathbf{y}) \approx f(\mathbf{x}) + \alpha \mathbf{d}^T \nabla f(\mathbf{x}) + o(\alpha \mathbf{d})$$

pour toute direction $d \in \mathbb{R}^n$ et pour $\alpha > 0$ suffisamment petit.

Conditions d'optimalité (2)

- La direction **d** est une direction de
  - maximisation si **d** forme un angle aigüe avec $\nabla f(\mathbf{x})$: $\mathbf{d}^T \nabla f(\mathbf{x}) > 0$.
  - minimisation si **d** forme un angle obtu avec $\nabla f(\mathbf{x})$: $\mathbf{d}^T \nabla f(\mathbf{x}) < 0$.
- Le gradient est normal aux courbes iso-coûts $f(\mathbf{x}) = cte$.



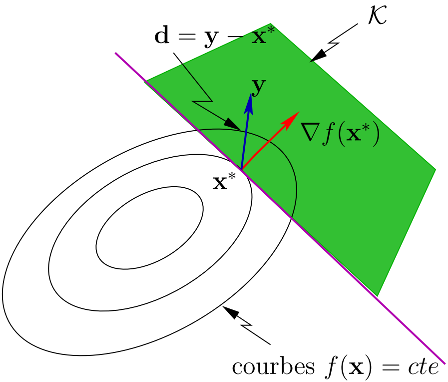
Conditions d'optimalité (3)

- **Direction admissible** : la direction $\mathbf{d}$ est admissible en $\mathbf{x} \in \mathcal{K}$ ssi il existe $\alpha > 0$ tel que $\mathbf{y} = \mathbf{x} + \alpha \mathbf{d} \in \mathcal{K}$.
- **Principe du minimum** : $\mathbf{x}^* \in \mathcal{K}$ est une solution optimale ssi

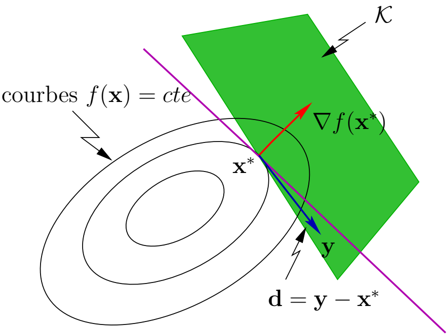
$$(\mathbf{y} - \mathbf{x}^*)^T \nabla f(\mathbf{x}^*) \geq 0 \quad \forall y \in \mathcal{K}$$

  - Toute direction admissible $\mathbf{d} = \mathbf{y} - \mathbf{x}$ est une direction de maximisation, i.e. elle forme un angle aigüe avec le gradient.
  - Cette condition est suffisante grâce à la convexité.
  - Si $\mathcal{K} = \mathbb{R}^n$, on retrouve $\nabla f(\mathbf{x}^*) = 0$.

- Idée de base des méthodes de direction admissible (cf. cours optimisation non-linéaire du routage).



$\mathbf{x}^* =$ minimum global.



$\mathbf{x}^* \neq$ minimum global.

Conditions d'optimalité de Karush-Kuhn-Tucker

- L'ensemble $\mathcal{K}$ des solutions admissibles est défini par :

$$\mathbf{x} \in \mathcal{K} \iff \begin{cases} h_i(\mathbf{x}) = 0 & i = 1, \ldots, k \\ g_j(\mathbf{x}) \leq 0 & j = 1, \ldots, m \end{cases}$$

Theorem (Conditions KKT)

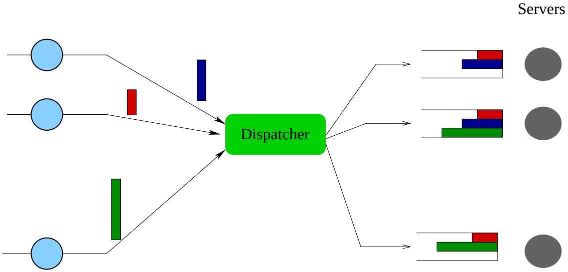
Si $\mathbf{x}^*$ est un minimum local, il existe des multiplicateurs $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$ et $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m_+$ tels que

$$\begin{aligned} \lambda_j g_j(\mathbf{x}^*) &= 0 \quad j = 1, \ldots, m \\ \nabla f(\mathbf{x}^*) &= -\sum_i \mu_i \nabla h(\mathbf{x}^*) - \sum_j \lambda_j \nabla g(\mathbf{x}^*) \end{aligned}$$

- La projection de $-\nabla f$ sur $\mathcal{K}$ est le vecteur nul.

Exemple : routage dans une ferme de serveurs

- **Un dispatcher centralisé route des jobs vers des serveurs**
  - Objectif : minimiser la durée moyenne de traitement.
  - Taux d'arrivée des jobs $\bar{\lambda}$ (Poisson).
  - Serveur j : capacité r_j, coût c_j/job et trafic x_j.
  - $\frac{c_1}{r_1} \leq \frac{c_2}{r_2} \ldots \leq \frac{c_S}{r_S}$
  - Modèle M/G/1/PS pour les serveurs.



$$\text{Min}_x \quad \sum_{j\in\mathcal{S}} c_j \frac{x_j}{r_j - x_j}$$

s.t.

$$\sum_{j\in\mathcal{S}} x_j = \bar{\lambda}$$
$$x_j \geq 0 \quad j \in \mathcal{S}$$

Exemple : routage dans une ferme de serveurs (2)

- **Conditions KKT** : $\exists \mu \in \mathbb{R}_+$ t.q.

$$\frac{\partial f}{\partial x_j}(\mathbf{x}^*) = \frac{c_j}{r_j} \frac{1}{\left(1 - x_j/r_j\right)^2} \geq \mu$$

avec égalité ssi $x_j^* > 0$.

- **Conclusion :** le routage optimal ne route des jobs que vers les serveurs ayant un coût marginal $\frac{\partial f}{\partial x_j}(\mathbf{x}^*)$ minimal.

Exemple : routage dans une ferme de serveurs (3)

- **Routage optimal** :
  - L'ensemble des serveurs utilisés par le partage de charge optimal est $\mathcal{S}^* = \{1, \ldots, j^*\}$, où

$$j^* = max \left\{ j \in \mathcal{S} \mid \sum_{k=1}^{j} \sqrt{c_k r_k} > \left(\sum_{k=1}^{j} r_k - \bar{\lambda} \right) \sqrt{\frac{c_j}{r_j}} \right\}$$

  - La charge du serveur $j \in \mathcal{S}^*$ en utilisant une politique optimale est donnée par,

$$x_j^* = r_j \left(1 - \sqrt{\frac{c_j}{r_j}} \frac{\sum_{k \in \mathcal{S}^*} r_k - \bar{\lambda}}{\sum_{k \in \mathcal{S}^*} \sqrt{c_k r_k}} \right)$$

tandis que $x_j^* = 0$ pour $j \notin \mathcal{S}^*$.

Programmation Linéaire

Programmation Linéaire

- Un programme linéaire est un problème d'optimisation convexe de la forme

$$\begin{array}{rl} \text{Minimiser} & z = \sum_i c_i x_i \\ \text{s.t} & \\ & \sum_i a_{ji} x_i \leq b_j \quad j = 1, \ldots, m \end{array}$$

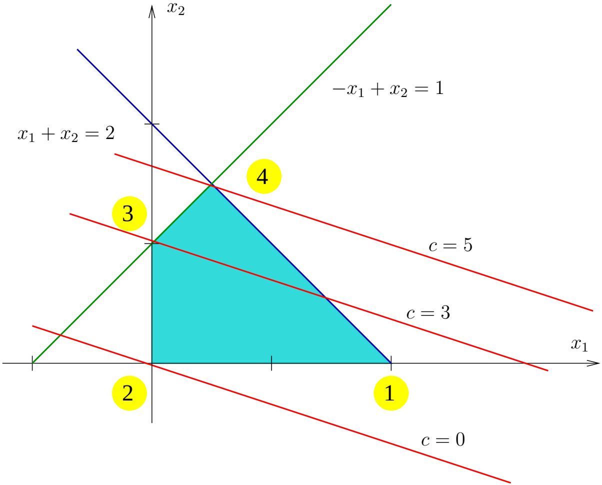
ou sous forme matricielle

$$\begin{array}{rl} \text{Minimiser} & z = \mathbf{c}^T \mathbf{x} \\ \text{s.t} & \mathbf{A}\, \mathbf{x} \leq \mathbf{b} \end{array}$$

- $\sum_i a_{ji} x_i = b_j$ est équivalent à $\sum_i a_{ji} x_i \leq b_j$ et $-\sum_i a_{ji} x_i \leq -b_j$.

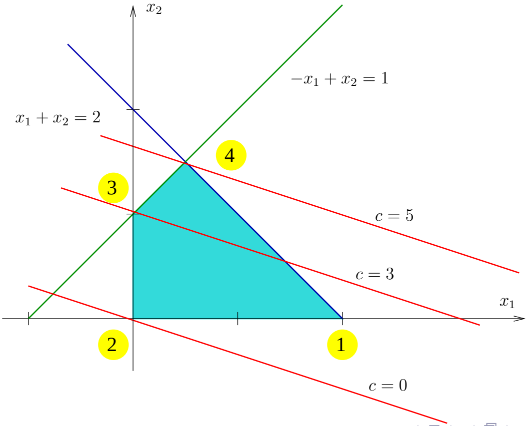
Propriétés fondamentales (1)

$$
\begin{aligned}
\text{Max} \quad & z = x_1 + 3x_2 \\
\text{s.t} \quad & \\
& -x_1 + x_2 \leq 1 \\
& x_1 + x_2 \leq 2 \\
& x_1 \geq 0, x_2 \geq 0
\end{aligned}
$$

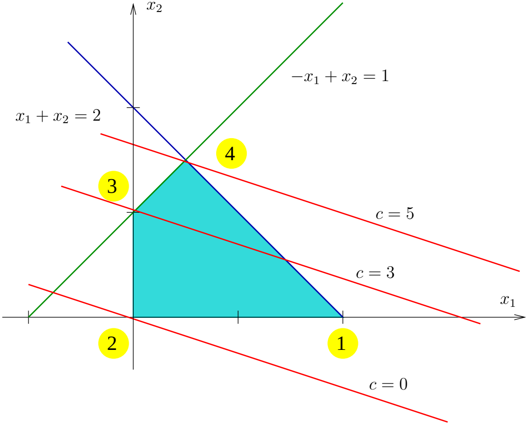


Propriétés fondamentales (2)

- Ensemble des solutions : intersection des demi-plans satisfaisant l'inégalité linéaire.
- Point numérotés : points extrêmaux.
- Les courbes iso-coûts $c = x_1 + 3x_2$ sont // les unes aux autres.



- La valeur max de z correspond à la plus grande valeur de c pour laquelle la ligne iso-coût a au moins un point commun avec l'ensemble des solutions : $c = 5$, $x_1 = \frac{1}{2}$, $x_2 = \frac{3}{2}$.



Propriété fondamentale
L'optimum d'un PL est toujours atteint en un point extrêmal.

Les différents cas de figure

- **Quatre possibilités**
  - **Pas de solution** : si on remplace $x_1 + x_2 \leq 2$ par $x_1 + x_2 \leq -1$ l'ensemble solution est vide.
  - **Solution non bornée** : si on enlève $x_1 + x_2 \leq 2$, alors $\max z = +\infty$.
  - **Solution unique** : elle correspond à un point extrêmal.
  - **Infinité de solution** : si on remplace $z = x_1 + 3x_2$ par $z = 2(x_1 + x_2)$ alors les lignes iso-coûts sont parallèles à la contrainte $x_1 + x_2 = 2$ et le maximum est atteint en tout point de ce segment.

La méthode du Simplex

- **Généralisation en dimension *n* de ce qu'on a vu dans $\mathbb{R}^2$**
  - Phase 1 : obtention d'une solution admissible de base (si existence)
  - Phase 2 : détermination de l'optimum global.

- **Méthode itérative :**
  - Visite consécutivement des solutions de base (points extrêmaux du polyhédre),
  - Diminue le coût à chaque itération,
  - Identifie le minimum global quand il est atteint.

- **Algorithme implémentés dans de nombreux solveurs linéaires** : CPLEX, Xpress MP, lp solve, matlab, etc.

- **Extrêmement efficace** : on peut résoudre des problèmes avec des milliers de variables et de contraintes en des temps calcul raisonnables.

Programmation Linéaire Mixte

Programmation Linéaire Mixte

- Exemple des PL mixtes avec variables 0-1 :

$$
\begin{array}{rl}
\text{Minimiser} & z = \sum_{i=1}^{n} c_i x_i \\
\text{s.t} & \\
 & \sum_i a_{ji} x_i \le b_j \quad j = 1, \ldots, m \\
 & x_i \in \{0, 1\} \quad i = 1, \ldots, k \\
 & x_i \ge 0 \quad i = k+1, \ldots, n
\end{array}
$$

- Solution partielle :
  - $N_0 \subset \{1, \ldots, k\}$: variables binaires valant 0,
  - $N_1 \subset \{1, \ldots, k\}$: variables binaires valant 1,
  - $N_U \subset \{1, \ldots, k\}$: variables binaires non fixées.

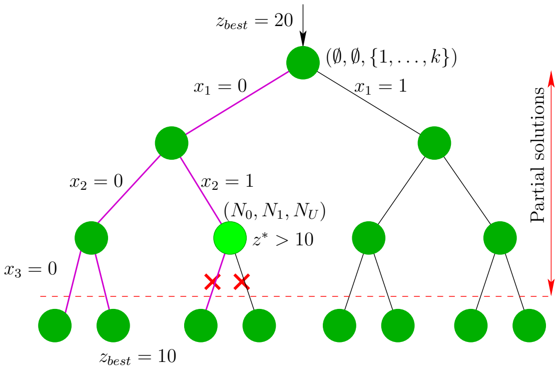
Borne inférieure

- Borne inférieure z^* sur le coût d'une solution partielle obtenue en résolvant le problème relaxé :

$$
\begin{aligned}
\text{Minimiser} \quad & z = \sum_{i=1}^{n} c_i \, x_i \\
\text{s.t} \quad & \\
& \sum_i a_{ji} x_i \leq b_j \quad j = 1, \ldots, m \\
& x_i \geq 0 \quad i = k+1, \ldots, n \\
& 0 \leq x_i \leq 1 \quad i \in N_U \\
& x_i = 0 \quad i \in N_0 \\
& x_i = 1 \quad i \in N_1
\end{aligned}
$$

- Toute solution complète **x** obtenue à partir de la solution partielle (N_0, N_1, N_U) en fixant les valeurs des variables $j \in N_U$ à 0 ou à 1 aura un coût supérieur ou égal à z^*.

- **Principe :**
  - Explorer l'arbre représentant l'ensemble des solutions,
  - Mettre à jour la borne sup quand une solution complète améliorante est découverte,
  - Calculer une borne inférieure en chaque noeud et élaguer l'arbre si $z^* \geq z^{best}$.



Branch and Bound (2)

- **Algorithme du Branch-and-Bound :**

procedure BRANCH-AND-BOUND((N_U, N_0, N_1))
    LinearProgramming($(N_U, N_0, N_1, \mathbf{x}, z^*)$)
    if $N_U = \emptyset$ or $x_i \in \{0, 1\}$ $\forall i \in N_U$ **then**
        if $z^* < z^{best}$ **then**
            $z^{best} = z^*$ et $\mathbf{x}^{best} = \mathbf{x}$
        end if
    else ▷ au moins un x_i non binaire
        if $z^* \geq z^{best}$ **then**
            return ▷ bounding
        else ▷ branching
            Choisir $i \in N_U$ tel que $x_i \neq 0, 1$
            Branch-and-Bound($N_U - i, N_0 \cup \{i\}, N_1$)
            Branch-and-Bound($N_U - i, N_0, N_1 \cup \{i\}$)
        end if
    end if
end procedure

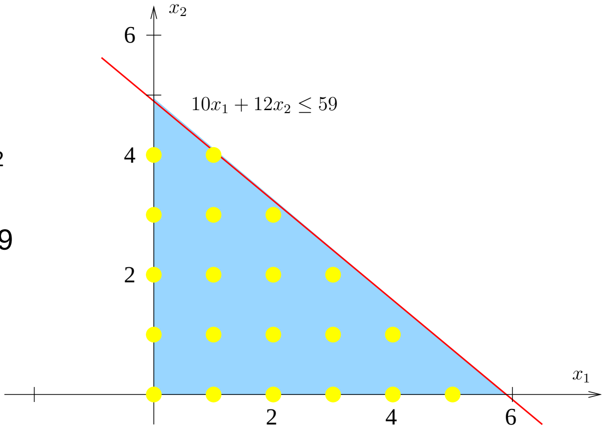
Branch and Bound (3)

- Généralisation à des variables entières.
- Implémenté par de nombreux solveurs : CPLEX, Xpress MP, lp solve, etc.
- Performances limitées à des problèmes de "petites" tailles :
  - Beaucoup de temps passé pour prouver l' optimalité,
  - Incontournable pour évaluer des approximations.
- La borne inférieure obtenue par relaxation linéaire est souvent grossière :
  - L'idée de base du Branch-and-Cut est d'améliorer la borne inférieure en ajoutant des contraintes vérifiées par toute solution entière mais excluant une partie des solutions continues (enveloppe convexe).

Max $\quad z = 10x_1 + 11x_2$

s.t

$$10x_1 + 12x_2 \leq 59$$
$$x_1, x_2 \in \mathbb{N}$$



- Problème relaxé (PL) : $x^* = (5.9, 0)$ et $z^* = 59$.
  - En arrondissant : $x^* = (5, 0)$ et $z^* = 50$.
- Problème discret (PLNE) : $x^* = (1, 4)$ et $z^* = 54$.

Branch and Cut (2)

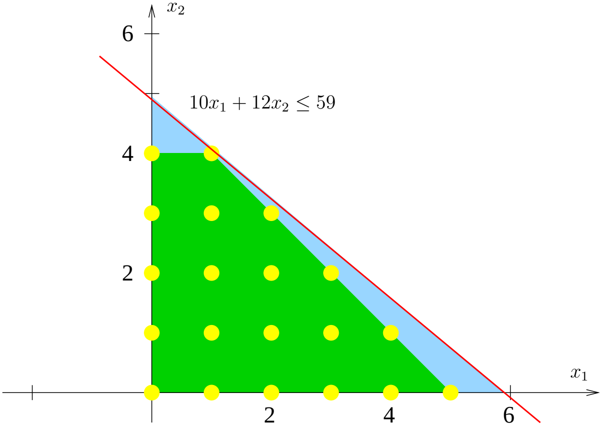
- Polyhédre convexe contenant toutes les solutions entières :

$$x_1 + x_2 \leq 5 \quad 0 \leq x_2 \leq 4 \quad x_1 \geq 0$$

$$\text{Max} \quad z = 10x_1 + 11x_2$$

s.t

$$x_1 + x_2 \leq 5$$
$$0 \leq x_2 \leq 4$$
$$x_1 \geq 0$$



- Borne inférieure : $x^* = (1, 4)$ et $z^* = 54$.

Dualité

Dualité

- On considére le problème suivant :

$$\begin{array}{rl} \text{Minimiser} & f(\mathbf{x}) \\ \text{s.t} & \\ & h_i(\mathbf{x}) = 0 \quad i = 1, \ldots, k \\ & g_j(\mathbf{x}) \leq 0 \quad j = 1, \ldots, m \\ & \mathbf{x} \in \mathcal{K} \end{array}$$

- **Lagrangien** :

$$L(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_i \mu_i h_i(\mathbf{x}) + \sum_j \lambda_j g_j(\mathbf{x}) \quad \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\lambda} \in \mathbb{R}_+^m$$

Dualité (2)

- **Fonction duale** :

$$W(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{K}} L(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\lambda}) \quad \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\lambda} \in \mathbb{R}_+^m$$

$$Dom(W) = \left\{ (\boldsymbol{\mu}, \boldsymbol{\lambda}) : \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\lambda} \in \mathbb{R}_+^m, \min_{\mathbf{x} \in \mathcal{K}} L(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\lambda}) > -\infty \right\}$$

- **Problème dual** :

$$\begin{array}{rc} \text{Maximiser} & W(\boldsymbol{\mu}, \boldsymbol{\lambda}) \\ \text{s.t} & \boldsymbol{\mu} \in \mathbb{R}^k, \boldsymbol{\lambda} \in \mathbb{R}_+^m \end{array}$$

Dualité (3)

- **Pour un problème d'optimisation convexe** :
  - La fonction duale $W(\boldsymbol{\mu}, \boldsymbol{\lambda})$ est concave et son domaine convexe,
  - Si le primal et le dual sont faisables, $f(\mathbf{x}^*) = W(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$,
  - Si le primal n'est pas borné (par en dessous), le dual est infaisable. Réciproquement, si le dual n'est pas borné (par en dessus), le primal est infaisable
  - A chaque solution optimale $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ du dual correspond une solution optimale du primal $\mathbf{x}^*$, et réciproquement.
  - La solution duale $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ correspond aux multiplicateurs de Lagrange des conditions KKT.