Cours OPT 201

Optimisation Différentiable – Théorie et Algorithmes

Exercices de la séance 3
(conditions d'optimalité)

Analyse convexe

1. Propriétés de convexité
2. Cônes duaux
3. Autour du lemme de Farkas

Calcul de solution par les conditions d'optimalité (contraintes d'égalité)

4. Quotient de Rayleigh
5. Éliminations hasardeuses

Conditions d'optimalité

6. Interprétation géométrique des conditions d'optimalité de KKT
7. Signification des multiplicateurs

Calcul de solution par les conditions d'optimalité (contraintes d'inégalité)

8. Minimisation d'une fonction linéaire sur une boule
9. Bifurcation de solutions

1 Propriétés de convexité

1. On suppose que f est convexe et on se donne $x_0,\ x_1 \in \operatorname{dom} f$ et $t > 1$ tel que $(1-t)x_0 + tx_1 \in \operatorname{dom} f$. Alors $f((1-t)x_0 + tx_1) \geqslant (1-t)f(x_0) + tf(x_1)$.

2. *Inégalité de Jensen.* Montrez que f est convexe si, et seulement si, pour tout m-uple $(t_1, \ldots, t_m) \in \Delta_m$ et pour tout $(x_1, \ldots, x_m) \in (\operatorname{dom} f)^m$, on a
$$f\left(\sum_{i=1}^{m} t_i x_i\right) \leqslant \sum_{i=1}^{m} t_i f(x_i). \tag{3.1}$$

3. *Inégalité des moyennes.* La moyenne géométrique est inférieure à la moyenne arithmétique : si $\{a_i\}_{i=1}^{n}$ sont n nombres positifs, on a
$$\left(\prod_{i=1}^{n} a_i\right)^{1/n} \leqslant \frac{1}{n}\sum_{i=1}^{n} a_i.$$
<u>Remarque</u>. Pour retenir le sens de cette inégalité, il suffit de constater que l'inégalité inverse n'a pas lieu si un des a_i est nul et un autre ne l'est pas.

2 Cônes duaux

1. Soit
$$P := \{x \in \mathbb{R}^n : x_1 \leqslant \cdots \leqslant x_n\}.$$
Montrez que le dual de cet ensemble pour le produit scalaire euclidien de $\mathbb{R}^n$ s'écrit
$$P^+ = \left\{d \in \mathbb{R}^n : \sum_{i=1}^{j} d_i \leqslant 0,\ \text{pour } j = 1, \ldots, n-1,\ \text{et } \sum_{i=1}^{n} d_i = 0\right\}.$$

On dit qu'un cône K d'un espace vectoriel muni d'un produit scalaire est un *auto-dual* si $K^+ = K$. Montrez que les cônes ci-dessous sont auto-duaux.

2. $\mathbb{R}^n_+$ (orthant positif de $\mathbb{R}^n$) pour le produit scalaire euclidien de $\mathbb{R}^n$.

3. $\{x \in \mathbb{R}^n : x_1^2 + \cdots + x_{n-1}^2 \leqslant x_n^2,\ x_n \geqslant 0\}$ pour le produit scalaire euclidien de $\mathbb{R}^n$.

4. $\mathcal{S}^n_+$ (ensemble des matrices d'ordre n symétriques semi-définies positives) pour le produit scalaire $\langle A, B\rangle = \operatorname{tr} AB$.

3 Autour du lemme de Farkas

1. Le cône convexe $\{\sum_i \alpha_i v_i : \text{les } \alpha_i \text{ sont positifs}\}$ engendré par (i.e., l'enveloppe conique de) *un nombre fini* de vecteurs v_i est fermé.

2. On considère le cône convexe fermé

$$K := \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 \leqslant x_3^2,\ x_3 \geqslant 0\}$$

et l'application linéaire

$$A : x \in \mathbb{R}^3 \to \mathbb{R}^2 : x \mapsto (x_1, x_2 + x_3).$$

Montrez que le cône $\{Ax : x \in K\}$ n'est pas fermé dans $\mathbb{R}^2$.

<u>Conséquence</u>. L'image par une application linéaire d'un cône convexe fermé n'est pas nécessairement fermée.

3. *Théorème de l'alternative de Farkas* (1902). Soient A une matrice $m \times n$ et $b \in \mathbb{R}^m$. Alors, des deux affirmations suivantes, une et une seule est vraie :

   (i) $\exists\, x \geqslant 0 : Ax = b$,
   (ii) $\exists\, y : A^\mathsf{T} y \geqslant 0$ et $b^\mathsf{T} y < 0$.

4. *Théorème de l'alternative de Motzkin* (1936). Soient $A \in \mathbb{R}^{m_A \times n}$, $B \in \mathbb{R}^{m_B \times n}$ et $C \in \mathbb{R}^{m_C \times n}$ des matrices ayant un même nombre de colonnes. Alors, des deux affirmations suivantes, une et une seule est vraie :

   (i) $\exists\, x \in \mathbb{R}^n : Ax = 0$, $Bx \leqslant 0$ et $Cx < 0$,
   (ii) $\exists\, (\alpha, \beta, \gamma) \in \mathbb{R}^{m_A} \times \mathbb{R}_+^{m_B} \times \mathbb{R}_+^{m_C} : A^\mathsf{T}\alpha + B^\mathsf{T}\beta + C^\mathsf{T}\gamma = 0$ et $\gamma \neq 0$.

   <u>Remarques</u>. Cette alternative peut s'utiliser sans les matrices A et B (on obtient alors l'*alternative de Gordan*, 1873), mais pas sans la matrice C (dans ce dernier cas, (i) et (ii) sont toutes les deux trivialement vraies). Elle permet d'avoir des conditions duales exprimant la *compatibilité d'égalités et d'inégalités linéaires* (strictes et non strictes) homogènes. La version non homogène de l'alternative de Motzkin est énoncée en exercice dans le syllabus du cours.

4 Quotient de Rayleigh

Soient $\mathbb{E}$ un espace euclidien de dimension finie, muni du produit scalaire $\langle\cdot,\cdot\rangle$ et de la norme associée $\|\cdot\| := \langle\cdot,\cdot\rangle^{1/2}$ et $A : \mathbb{E} \to \mathbb{E}$ une application linéaire auto-adjointe (pour ce produit scalaire), c'est-à-dire qui vérifie

$$\forall\, (x, y) \in \mathbb{E}^2, \quad \langle Ax, y\rangle = \langle x, Ay\rangle.$$

Le *quotient de Rayleigh* de A dans la direction non nulle $x \in \mathbb{E}$ est défini par

$$q(x) := \frac{\langle Ax, x\rangle}{\|x\|^2}.$$

1. Montrez que les problèmes

$$(P_1) \quad \inf_{x \neq 0}\, q(x) \qquad \text{et} \qquad (P_2) \quad \sup_{x \neq 0}\, q(x),$$

   dans lequel on minimise et maximise le quotient de Rayleigh sur $\Omega = \mathbb{R}^n \backslash \{0\}$, ont une solution.

2. On considère les problèmes

$$(P_1') \quad \begin{cases} \inf \langle Ax, x\rangle \\ \|x\| = 1 \end{cases} \qquad \text{et} \qquad (P_2') \quad \begin{cases} \sup \langle Ax, x\rangle \\ \|x\| = 1. \end{cases}$$

   (*i*) Montrez que la contrainte $c(x) \equiv \|x\|^2 - 1$ est qualifiée en tout point admissible [3]

   (*ii*) Montrez que les vecteurs propres unitaires (i.e., de norme 1) de A sont les points stationnaires de (P_1') et que les valeurs propres associées en sont les multiplicateurs associés et les valeurs critiques.

   (*iii*) Montrez que la valeur optimale de (P_1') [resp. de (P_2')] est la plus petite valeur propre $\lambda_{\min}(A)$ [resp. la plus grande valeur propre $\lambda_{\max}(A)$] de A et que les solutions sont les vecteurs propres unitaires correspondants.

   (*iv*) Montrez que les conditions suffisantes du second ordre de (P_1') sont vérifiées en une solution si, et seulement si, la plus petite valeur propre de A est simple (i.e., l'espace vectoriel associé est de dimension 1).

3. On note λ_i les valeurs propres de A. Montrez que

$$\max_i |\lambda_i| = \|A\| \tag{3.2}$$

$$\lambda_{\max}(A) = \|A\| \quad (\text{si } A \succcurlyeq 0) \tag{3.3}$$

$$\lambda_{\min}(A) = \|A^{-1}\|^{-1} \quad (\text{si } A \succ 0). \tag{3.4}$$

5 Éliminations hasardeuses de variable ou contrainte

Les exemples suivant montrent qu'il faut être très prudent lorsqu'on élimine des contraintes. Dans les deux cas ci-dessous, l'élimination de contrainte conduit à une erreur, à cause d'une perte d'information.

Exemple 1

On considère le problème

$$\begin{cases} \min_{(x_1,x_2)} \; x_1^2 + (x_2-1)^2 \\ \alpha\, x_1^2 = x_2, \end{cases} \tag{3.5}$$

où $\alpha > 0$.

1. *Sans élimination.*

   (a) Montrez que la contrainte est qualifiée en tout point.

   (b) Déterminez la ou les solutions du problème (3.5) en fonction de $\alpha > 0$.

   On utilisera les conditions du deuxième ordre pour déterminer si les points stationnaires donnés par les conditions de Lagrange sont des minima ou des maxima.

[3] Il n'est pas indispensable d'élever la norme au carré, car $x \mapsto \|x\|$ est différentiable sur la sphère unité, mais cela permet d'avoir la même degré d'homogénéité que le critère.

2. *Avec élimination.* On remplace dans le critère de (3.5), x_1^2 par sa valeur donnée par la contrainte, ce qui conduit alors au problème

$$\min_{x_2} \frac{1}{\alpha} x_2 + (x_2 - 1)^2. \tag{3.6}$$

Déterminez l'unique solution du problème (3.6). Celle-ci n'est pas nécessairement solution de (3.5). Pourquoi ?

Leçon : *il est donc préférable d'utiliser les conditions d'optimalité des problèmes avec contraintes.*

Exemple 2

On considère le problème

$$\begin{cases} \inf_{(x,y)\in\mathbb{R}^2\times\mathbb{R}} e_1^\mathsf{T} Ay + \frac{1}{2}\|x\|_2^2 \\ x = Ay, \end{cases}$$

où

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{et} \qquad A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

1. *Sans élimination.* Calculez la solution du problème.
2. *Avec élimination de x.* On élimine x en remplaçant x par Ay dans l'objectif et en éliminant la contrainte. On obtient

$$\inf_{y\in\mathbb{R}} e_1^\mathsf{T} Ay + \frac{1}{2}\|Ay\|_2^2.$$

3. *Avec élimination de y.* On élimine y en remplaçant Ay par x dans l'objectif et en éliminant la contrainte. On obtient

$$\inf_{x\in\mathbb{R}^2} e_1^\mathsf{T} x + \frac{1}{2}\|x\|_2^2.$$

Expliquez pourquoi on retrouve parfois la solution de la formulation avant élimination et parfois pas.

6 Interprétation géométrique des conditions d'optimalité de KKT

Les dessins de la figure 2 représentent un ensemble admissible $X = \{x \in \mathbb{R}^2 : c_i(x) \leqslant 0,\ i = 1, 2, 3\}$ d'un problème d'optimisation dont la solution est supposée être le point $x_* \in X$ (intersection des courbes $c_2(\cdot) = 0$ et $c_3(\cdot) = 0$). Les conditions d'optimalité du premier ordre affirment que le gradient de f en x_* (pour le produit scalaire euclidien) est dans un certain cône. La figure donne quatre possibilités (les cônes ont été translatés en x_*). Quel est le cône dans lequel doit se trouver le gradient $\nabla f(x_*)$? Justifiez votre réponse.

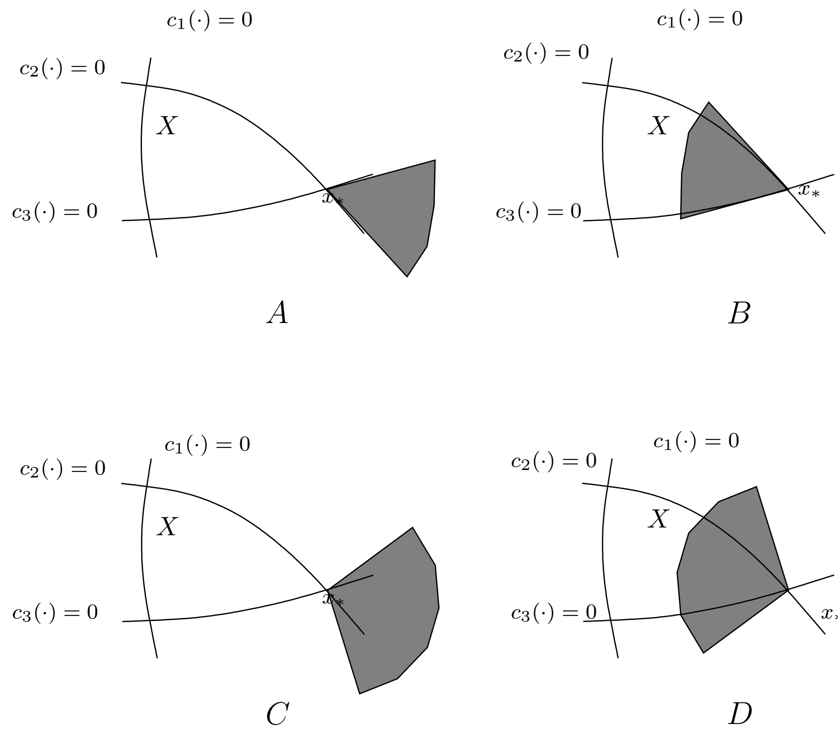

Fig. 2: Interprétation géométrique des conditions d'optimalité de KKT

7 Signification des multiplicateurs

On considère le problème d'optimisation perturbé sur $\mathbb{R}^n$:

$$(P_p) \quad \begin{cases} \min f(x) \\ c_i(x) + p_i = 0, \quad i \in E \\ c_i(x) + p_i \leqslant 0, \quad i \in I, \end{cases}$$

dont on suppose les données f et c différentiables. Soit $(\bar{x}, \bar{\lambda})$ une solution primale-duale de (P_0). On suppose que pour p voisin de 0 dans $\mathbb{R}^{|E|+|I|}$ le problème (P_p) admet une solution primale-duale $(\bar{x}(p), \bar{\lambda}(p))$ unique dans un voisinage de $(\bar{x}, \bar{\lambda})$. On suppose également que

$$\bar{x}(0) = \bar{x}, \quad \bar{\lambda}(0) = \bar{\lambda} \quad \text{et} \quad p \to (\bar{x}(p), \bar{\lambda}(p)) \text{ est différentiable.}$$

Ce sont des hypothèses très fortes, mais on cherche seulement ici à trouver la signification des multiplicateurs optimaux, sans que les hypothèses soient les plus faibles possibles.

1. Montrez que l'on a pour p voisin de zéro :

$$f(\bar{x}(p)) = \ell(\bar{x}(p), \bar{\lambda}(p)) + \bar{\lambda}(p)^{\mathsf{T}} p,$$

   où ℓ est le lagrangien du problème (P_0) :

$$\ell(x, \lambda) = f(x) + \sum_{i \in E \cup I} \lambda_i c_i(x).$$

2. Montrez que
$$\nabla_p(f \circ \bar{x})(0) = \bar{\lambda}$$
et interprétez.
3. On considère le problème qui consiste à minimiser $\|x\|_2^2$ (le carré de la norme ℓ_2) sur une partie X de $\mathbb{R}^2$ délimitée par des contraintes affines. Chaque dessin A, B et C de la figure 3 représente un domaine admissible et la courbe de niveau 1 du critère (en pointillés). Dans le premier dessin (A), l'ensemble admissible est
$$X := \{x \in \mathbb{R}^2 : x_1 \geqslant 0\},$$
dans le second (dessin B)
$$X := \{x \in \mathbb{R}^2 : x_1 \geqslant 1\}$$
et dans le troisième (dessin C)
$$X := \{x \in \mathbb{R}^2 : x_1 \geqslant 1,\ x_2 \leqslant 0,\ x_2 \geqslant x_1 - 2\}.$$
Sans écrire les conditions d'optimalité, mais **en examinant les figures** et en utilisant la **technique de perturbation des contraintes**, donner la solution de ces problèmes **et la valeur des multiplicateurs optimaux**.

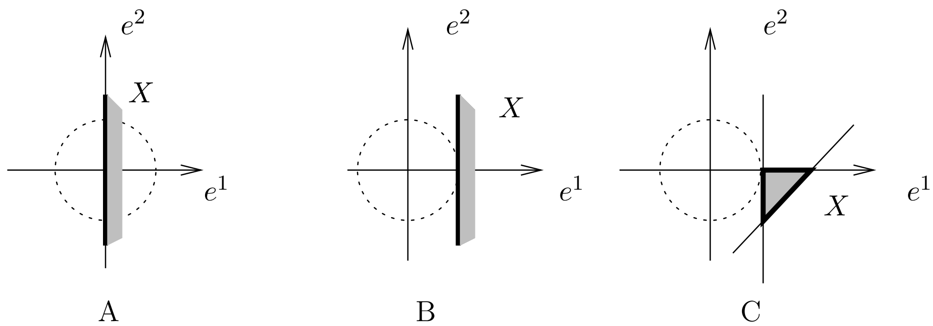

Fig. 3: L'ensemble des multiplicateurs optimaux est un singleton

8 Minimisation d'une fonction linéaire sur une boule

Soit $c \in \mathbb{R}^n$ *non nul* (le cas où $c = 0$ est trivial). Calculez les solutions du problème
$$\begin{cases} \min c^\mathsf{T} x \\ x \in X, \end{cases}$$
lorsque

1. $X = \{x \in \mathbb{R}^n : \|x\|_1 \leqslant 1\}$,
2. $X = \{x \in \mathbb{R}^n : \|x\|_p \leqslant 1\}$, avec $1 < p < \infty$,
3. $X = \{x \in \mathbb{R}^n : \|x\|_\infty \leqslant 1\}$.

En déduire l'**inégalité de Hölder** : pour tout x et $y \in \mathbb{R}^n$, et pour $1 \leqslant p \leqslant +\infty$, on a
$$|x^\mathsf{T} y| \leqslant \|x\|_p \|y\|_{p'}, \quad \text{avec } \tfrac{1}{p} + \tfrac{1}{p'} = 1.$$
Lorsque $1 < p < +\infty$, on a égalité dans l'inégalité de Hölder si, et seulement si, x est parallèle au vecteur de composante $\operatorname{sgn}(y_i)|y_i|^{p'/p}$ (qui est y si $p = 2$).

9 Bifurcation de solutions

On considère le problème

$$\begin{cases} \min\,(f(x) = x_2^2 - x_1^2) \\ x_2 \geqslant 2|x_1| - \varepsilon, \end{cases}$$

où ε est un paramètre réel.

1. Formulez le problème de manière à n'avoir que des contraintes linéaires (et donc différentiables).
2. Montrez l'existence d'une solution quelle que soit la valeur de $\varepsilon \in \mathbb{R}$.
3. En utilisant les conditions d'optimalité, décrivez les points stationnaires (en particulier les solutions) en fonction de $\varepsilon \in \mathbb{R}$.