OPT201 – Optimisation Différentiable I

Contrôle des connaissances

- Durée : 1 h 30 (on ne peut probablement pas tout faire...).
- Dans chaque sujet, on peut admettre des résultats énoncés dans des questions ou sous-questions *précédentes* pour traiter la question ou sous-question courante.
- Ne pas écrire en rouge.
- On peut consulter tous les documents *distribués au cours*, ainsi que ses propres notes.
- *Éteindre tout appareil connecté à un réseau.*

Trois sujets indépendants

1. Calcul de solution. On considère le problème d'optimisation à deux variables $x = (x_1, x_2) \in \mathbb{R}^2$ suivant :

$$\begin{cases} \min_x \; x_1 + x_2 \\ x_1 \geqslant x_2 \\ x_1^2 + x_2^2 \leqslant 1. \end{cases} \tag{1}$$

1. Montrez théoriquement que ce problème a au moins une solution (*théoriquement* veut dire ici *sans calculer une solution*, ce qui est le sujet des points suivants, mais par des arguments de nature topologique).
2. Montrez que les contraintes du problème sont qualifiées en tout point admissible.
3. Calculez tous les points stationnaires du problème, composantes primales et duales (c'est-à-dire, les solutions des conditions d'optimalité du premier ordre).
4. Déterminez la ou les solutions (globales) du problème.

2. Face exposée d'un convexe. Soit C un convexe de $\mathbb{R}^n$.

- On dit qu'une partie E de C est *exposée* s'il existe $\xi \in \mathbb{R}^n$ tel que $E = \arg\min \{\xi^\mathsf{T} x : x \in C\}$ (ce dernier ensemble est celui des minimiseurs du problème d'optimisation $\min \{\xi^\mathsf{T} x : x \in C\}$).
- On dit qu'une partie F de C est une *face* de C si F est convexe et si, lorsque x_0 et $x_1 \in C$ vérifient $(x_0 + x_1)/2 \in F$, on a nécessairement x_0 et $x_1 \in F$.

1. Montrez que toute partie exposée d'un convexe est une face.
2. La réciproque n'est pas nécessairement vraie : trouver un exemple d'ensemble convexe et de face de celui-ci qui n'est pas exposée (il faut que $n \geqslant 2$).

   [*Indication* : Tout point de la courbe $\{x \in \mathbb{R}^2 : x_2 = x_1^2\}$ est une face exposée du convexe $\{x \in \mathbb{R}^2 : x_2 \geqslant x_1^2\}$.]

3. Dualisation wolfienne d'un problème convexe différentiable. On considère les problèmes d'optimisation (P) et (D) suivants :

$$(P) \quad \begin{cases} \inf_{x \in \mathbb{R}^n} \; f(x) \\ c(x) \leqslant 0 \end{cases} \qquad \text{et} \qquad (D) \quad \begin{cases} \sup_{(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m} \; \ell(x, \lambda) \\ \lambda \geqslant 0 \\ \nabla_x \ell(x, \lambda) = 0, \end{cases}$$

dans lesquels $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction *convexe* différentiable, les composantes de $c : \mathbb{R}^n \to \mathbb{R}^m$ sont *convexes* différentiables et

$$\ell : \mathbb{R}^n \times \mathbb{R}^m : (x, \lambda) \mapsto \ell(x, \lambda) = f(x) + \lambda^\mathsf{T} c(x)$$

est le lagrangien du problème (P). On note $\text{val}(P)$ et $\text{val}(D)$ les valeurs optimales de (P) et de (D), respectivement.

1. *Dualité faible.* Montrez que $\text{val}(D) \leqslant \text{val}(P)$.

   [*Indication* : Ce n'est pas une dualisation min-max. Il faut trouver une démonstration directe en utilisant un point admissible arbitraire x de (P), un point admissible arbitraire $(\tilde{x}, \tilde{\lambda})$ de (D) et la convexité de f et des c_i.]

2. *Dualité forte I.* On suppose dans ce numéro que (P) a une solution x_* et que la contrainte de (P) y est qualifiée au sens de la définition 4.25 du syllabus du cours (c'est-à-dire que les cônes tangent et linéarisant sont égaux). Montrez que

   (a) il existe $\lambda_* \in \mathbb{R}^m$ tel que (x_*, λ_*) est solution de (D),

   (b) $\text{val}(D) = \text{val}(P)$.

3. *Dualité forte II.* On suppose maintenant que la contrainte de (P) est affine, de la forme

   $$Ax \leqslant b, \qquad \text{avec } A \in \mathbb{R}^{m \times n} \text{ et } b \in \mathbb{R}^m,$$

   et qu'*elle n'est pas réalisable* (il n'existe pas de $x \in \mathbb{R}^n$ tel que $Ax \leqslant b$, donc $\text{val}(P) = +\infty$; mais a priori cela n'implique pas que $\text{val}(D) = +\infty$; c'est ce que l'on va montrer).
   Montrer que

   (a) on peut trouver un $\mu \in \mathbb{R}^m$ tel que $\mu \geqslant 0$, $A^\mathsf{T}\mu = 0$ et $b^\mathsf{T}\mu = -1$,

   (b) si les contraintes de (D) sont réalisables, alors $\text{val}(D) = +\infty$.

OPT201 – Optimisation Différentiable I

Contrôle des connaissances (solutions succinctes)

Barème : 24 points, qui seront ramenés à 20 (on peut évaluer les réponses avec des points non entiers).

1. Barème : 8 points

1. **Barème** : 1 point
   L'ensemble admissible est non vide (contient $(0,0)$) et compact (c'est l'intersection de la boule unité, qui est compacte, et de l'ensemble fermé $\{x \in \mathbb{R}^2 : x_1 \geqslant x_2\}$, donc un compact). Le critère est continu (linéaire en réalité). Donc le problème a au moins une solution (théorème de Weierstrass).
2. **Barème** : 1 point
   L'ensemble admissible est décrit par des contraintes convexes et (QC-S) est vérifiée : le point $(0, -1/2)$ vérifie les contraintes strictement.
3. **Barème** : 5 points (1 point pour les conditions de KKT, 4 points pour le reste, dont 1 point pour des calculs corrects)
   Le problème et sa solution sont représentés à la figure 1.

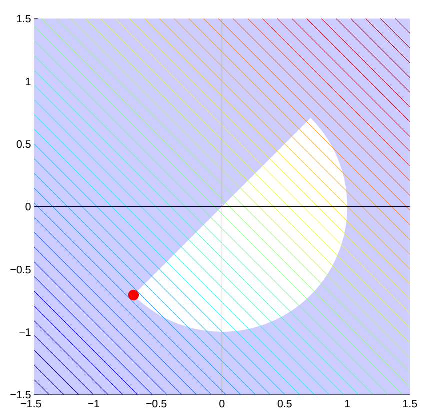

Figure 1: L'ensemble admissible du problème (en clair), sa solution et unique point stationnaire (point rouge) et les courbes de niveau du critère

Le lagrangien est la fonction $\ell : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ définie en $(x, \lambda) \in \mathbb{R}^2 \times \mathbb{R}^2$ par

$$\ell(x,\lambda) = x_1 + x_2 + \lambda_1(-x_1 + x_2) + \lambda_2(x_1^2 + x_2^2 - 1).$$

Les conditions de KKT s'écrivent

$$\begin{cases} 1 - \lambda_1 + 2\lambda_2 x_1 = 0 \\ 1 + \lambda_1 + 2\lambda_2 x_2 = 0 \\ 0 \leqslant \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \perp \begin{pmatrix} x_1 - x_2 \\ 1 - x_1^2 - x_2^2 \end{pmatrix} \geqslant 0. \end{cases}$$

- **Cas 1 :** $\lambda_1 = \lambda_2 = 0$.
  Non compatible avec les deux premières conditions.

- **Cas 2 :** $\lambda_1 = 0$ et $x_1^2 + x_2^2 = 1$.

  Il faut alors résoudre
  $$\begin{cases} 1 + 2\lambda_2 x_1 = 0 \\ 1 + 2\lambda_2 x_2 = 0 \\ x_1^2 + x_2^2 = 1 \\ x_1 \geqslant x_2 \\ \lambda_2 \geqslant 0. \end{cases}$$
  Les 3 premières conditions donnent
  $$x_1 = x_2 = \pm\frac{\sqrt{2}}{2} \qquad \text{et} \qquad \lambda_2 = \mp\frac{\sqrt{2}}{2}.$$
  La dernière condition impose finalement
  $$x_1 = x_2 = -\frac{\sqrt{2}}{2} \qquad \text{et} \qquad \lambda_2 = \frac{\sqrt{2}}{2}.$$

- **Cas 3 :** $x_1 = x_2$ et $\lambda_2 = 0$.

  Non compatible avec les deux premières conditions.

- **Cas 4 :** $x_1 = x_2$ et $x_1^2 + x_2^2 = 1$.

  Il faut alors résoudre
  $$\begin{cases} 1 - \lambda_1 + 2\lambda_2 x_1 = 0 \\ 1 + \lambda_1 + 2\lambda_2 x_2 = 0 \\ x_1 = x_2 \\ x_1^2 + x_2^2 = 1 \\ \lambda_1 \geqslant 0 \\ \lambda_2 \geqslant 0. \end{cases}$$
  Les 4 premières conditions donnent
  $$x_1 = x_2 = \pm\frac{\sqrt{2}}{2}, \qquad \lambda_1 = 0 \qquad \text{et} \qquad \lambda_2 = \mp\frac{\sqrt{2}}{2}.$$
  La dernière condition impose finalement
  $$x_1 = x_2 = -\frac{\sqrt{2}}{2} \qquad \lambda_1 = 0 \qquad \text{et} \qquad \lambda_2 = \frac{\sqrt{2}}{2}.$$

4. **Barème** : 1 point

   Il y a un unique point stationnaire $x_1 = x_2 = -\sqrt{2}/2$. C'est donc la solution du problème, qui n'en a pas d'autres. On notera qu'il n'y a pas complémentarité stricte pour la première contrainte (elle est active et son multiplicateur est nul) et que cette contrainte est superflue.

2. Barème : 4 points

1. **Barème** : 2 points (1 point pour la convexité de E et 1 point pour son autre propriété)

   Soit $E = \arg\min\{\xi^\mathsf{T} x : x \in C\}$.

   Alors E est convexe comme ensemble de solutions d'un problème convexe.

   On note $\alpha = \min\{\xi^\mathsf{T} x : x \in C\}$. Soient x_0 et $x_1 \in C$ tels que $x = (x_0 + x_1)/2 \in E$. Alors $\xi^\mathsf{T} x_1 \geqslant \alpha$, $\xi^\mathsf{T} x_2 \geqslant \alpha$ et $\xi^\mathsf{T} x_0 + \xi^\mathsf{T} x_1 = 2\alpha$. Nécessairement, $\xi^\mathsf{T} x_1 = \alpha$ et $\xi^\mathsf{T} x_2 = \alpha$, si bien que x_1 et $x_2 \in E$.

2. **Barème** : 2 points

   Le singleton $\{0\}$ de l'ensemble $C := \mathbb{R}^2_+ \cup \{x \in \mathbb{R}^2 : x_2 \geqslant x_1^2\}$ est une face de C, mais n'est pas une partie exposée.

3. Barème : 12 points

1. **Barème** : 3 points

   Si (P) et/ou (D) n'est pas réalisable, l'inégalité est claire. Dans le cas contraire, on peut trouver $(\tilde{x},\tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ tel que $\nabla_x \ell(\tilde{x},\tilde{\lambda}) = 0$ et $x \in \mathbb{R}^n$ tel que $c(x) \leqslant 0$. Voici deux manières de poursuivre.

   - On a

   $$\begin{array}{rcl} \ell(\tilde{x},\tilde{\lambda}) & = & f(\tilde{x}) + \tilde{\lambda}^\mathsf{T} c(\tilde{x}) \\ & \leqslant & f(x) - \nabla f(\tilde{x})^\mathsf{T}(x-\tilde{x}) + \tilde{\lambda}^\mathsf{T} c(\tilde{x}) \qquad [\text{convexité de } f] \\ & = & f(x) + \tilde{\lambda}^\mathsf{T} c'(\tilde{x})(x-\tilde{x}) + \tilde{\lambda}^\mathsf{T} c(\tilde{x}) \qquad [\nabla_x \ell(\tilde{x},\tilde{\lambda}) = 0] \\ & \leqslant & f(x) + \tilde{\lambda}^\mathsf{T} c(x) \qquad [\text{convexité des } c_i \text{ et } \tilde{\lambda} \geqslant 0] \\ & \leqslant & f(x) \qquad [c(x) \leqslant 0 \text{ et } \tilde{\lambda} \geqslant 0]. \end{array}$$

     On en déduit que $\operatorname{val}(D) \leqslant \operatorname{val}(P)$.

   - Comme f et les c_i sont convexes et que $\tilde{\lambda} \geqslant 0$, $x' \mapsto \ell(x',\tilde{\lambda})$ est convexe. Cette dernière fonction est alors minimisée en $\tilde{x}$ car $\nabla_x \ell(\tilde{x},\tilde{\lambda}) = 0$. Ceci conduit à la première inégalité ci-dessous et, pour la seconde, on utilise le fait que $c(x) \leqslant 0$ et $\tilde{\lambda} \geqslant 0$:

   $$\ell(\tilde{x},\tilde{\lambda}) \leqslant \ell(x,\tilde{\lambda}) \leqslant f(x).$$

     On en déduit que $\operatorname{val}(D) \leqslant \operatorname{val}(P)$.

2. **Barème** : 4 points (1 point si on trouve (b) à partir de (a))

   Comme x_* est une solution de (P) et que la contrainte c est qualifiée en x_*, il existe un multiplicateur $\lambda_* \in \mathbb{R}^m$ tel que l'on ait les conditioins de KKT, à savoir

   $$\nabla_x \ell(x_*,\lambda_*) = 0 \qquad \text{et} \qquad 0 \leqslant \lambda_* \perp c(x_*) \leqslant 0.$$

   Comme x_* est admissible pour (P) et (x_*,λ_*) est admissible pour (D), le point 1 implique que

   $$\ell(x_*,\lambda_*) \leqslant \operatorname{val}(D) \leqslant \operatorname{val}(P) \leqslant f(x_*).$$

   Mais $\ell(x_*,\lambda_*) = f(x_*)$ par la relation de complémentarité $\lambda_* \perp c(x_*)$, si bien que l'on a égalité partout dans le systeme d'inégalités ci-dessus. On en déduit que (x_*,λ_*) est solution de (D) et que $\operatorname{val}(D) = \operatorname{val}(P)$.

3. **Barème** : 5 points

   (a) **Barème** : 3 points

   Il s'agit d'une expression duale de la non-réalisabilité d'un système linéaire, que l'on obtient en utilisant le lemme de Farkas. Voici deux manières de s'y prendre.

   - On part de la non-réalisabilité de $Ax \leqslant b$: dans ce cas

   $$b \notin \begin{pmatrix} A & I \end{pmatrix} (\mathbb{R}^n \times \mathbb{R}^m_+) = \left\{ y : \begin{pmatrix} A & I \end{pmatrix}^\mathsf{T} y \in \{0_{\mathbb{R}^m}\} \times \mathbb{R}^m_+ \right\}^+,$$

     où la dernière égalité provient du lemme de Farkas (le cône convexe $\mathbb{R}^n \times \mathbb{R}^m_+$ étant polyédrique, $(A\ \ I)(\mathbb{R}^n \times \mathbb{R}^m_+)$ est aussi polyédrique et donc son adhérence lui est égale). Il existe donc un y tel que $A^\mathsf{T} y = 0$, $y \geqslant 0$ et $b^\mathsf{T} y < 0$. Le résultat est obtenu en prenant $\mu = -y/(b^\mathsf{T} y)$.

   - On part du résultat à obtenir : il revient à montrer que

   $$\begin{pmatrix} 0 \\ -1 \end{pmatrix} \in \begin{pmatrix} A^\mathsf{T} \\ b^\mathsf{T} \end{pmatrix} (\mathbb{R}^m_+) = \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} : Ax + \alpha b \geqslant 0\}^+,$$

où la dernière égalité provient du lemme de Farkas (l'orthant positif $\mathbb{R}^m_+$ étant polyédrique, $(A \;\; I)^\mathsf{T}(\mathbb{R}^m_+)$ est aussi polyédrique et donc son adhérence lui est égale). Si cette appartenance n'était pas vérifiée, il existerait $(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ tel que $Ax + \alpha b \geqslant 0$ et $\alpha > 0$. Mais alors on aurait $A(-x/\alpha) - b \leqslant 0$ et le problème (P) serait réalisable.

(b) **Barème** : 2 points
Si les contraintes de (D) sont réalisables, il existe $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m_+$ tel que $\nabla_x \ell(x, \lambda) = 0$. Avec le μ trouvé ci-dessus et $t \geqslant 0$, $(x, \lambda + t\mu)$ réalise encore les contraintes de (D) et $\ell(x, \lambda + t\mu) = \ell(x, \lambda) + t$ tend vers $+\infty$ lorsque $t \to +\infty$. Donc $\operatorname{val}(D) = +\infty$.

Remarque. Le dual de Wolfe (D) se distingue du dual de Lagrange par un critère à minimiser explicite (alors que la fonction duale δ peut n'être connue que numériquement, pas analytiquement), mais avec une contrainte d'égalité non linéaire « $\nabla_x \ell(x, \lambda) = 0$ », qui peut être compliquée à prendre en compte numériquement (elle est non linéaire, non convexe, sa dérivée fait intervenir les dérivées secondes de f et c, elle est couplante donc ne permet pas de bénéficier d'un lagrangien décomposable). Il se peut d'ailleurs que (P) soit plus simple à résoudre numériquement que (D). Cependant, on peut parfois s'en servir pour obtenir des bornes inférieures de la valeur optimale de (P), grâce à l'inégalité de dualité faible, ce qui peut être utile dans la recherche d'un minimum global par les techniques de branchement. Retenons que

> *En présence de convexité, la valeur $\ell(x, \lambda)$ du lagrangien est une borne inférieure de $\operatorname{val}(P)$ en un couple (x, λ) vérifiant $\nabla_x \ell(x, \lambda) = 0$ et $\lambda \geqslant 0$.*

Quelques références sur le sujet : [2, 1].

Bibliographie

[1] C.J. Goh, X.Q. Yang (2002). *Duality in Optimization and Variational Inequality*. Optimization Theory and Applications 2. Taylor and Francis. 6

[2] P. Wolfe (1961). A duality theorem for non-linear programming. *Quaterly of Applied Mathematics*, 19, 239–244. [doi]. 6