# Optimisation sous contraintes
# Examen

**Le partiel est prévu pour une durée de 2h30.**

**La rigueur de la rédaction est un élément d'évaluation important.**

# Un problème sous contraintes

Pour $c \in \mathbb{R}^n$, on considère le problème

$$\inf_{x \in C} \langle c, x \rangle , \tag{P}$$

où $C = \{x \in \mathbb{R}^n / \|x\|_2 \leq 1, x \succeq 0\}$.

*Notation :* pour $z \in \mathbb{R}^n$, on écrira $z^- \in \mathbb{R}^n$ défini pour $i \in \{1, \dots n\}$ par $(z^-)_i = \begin{cases} z_i & \text{si } z_i < 0, \\ 0 & \text{sinon} \end{cases}$.

1. Montrer que le problème (P) est bien posé, et en donner un minimiseur global évident si $c \geq 0$.

On suppose **dans toute la suite** qu'il existe $i_0 \in \{1, \dots n\}$ tel que $c_{i_0} < 0$.

2. Écrire l'ensemble $C$ grâce à $n+1$ contraintes d'inégalité (les plus simples possibles) $\mathcal{C}^1$ sur $\mathbb{R}^n$, notées $f_0, f_1, \dots, f_n$, où $f_0$ est associée à la contrainte $\|x\|_2 \leq 1$.
3. Montrer que tous les points de $C$ sont réguliers par rapport aux contraintes le défissant. *On pourra différencier les cas en fonctions de si $x \in C$ satisfait $\|x\|_2 < 1$ ou $\|x\|_2 = 1$.*
4. Soit $x$ un minimiseur global du problème.
   (a) Déduire de la précédente question qu'il existe $\lambda_0 \geq 0$, $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n_+$ tels que

   $$c + 2\lambda_0 x - \lambda = 0_{\mathbb{R}^n}$$

   avec $\lambda_0 = 0$ si $\|x\|_2 < 1$, et pour tout $i \in \{1, \dots, n\}$, $\lambda_i = 0$ si $x_i > 0$.
   (b) Justifier que $\lambda_0 > 0$.
   (c) Démontrer que

   $$x = -\frac{c^-}{\|c^-\|_2}.$$

---

# Algorithme de Frank-Wolfe

On se donne $C$ un ensemble non vide, convexe et fermé borné de $\mathbb{R}^n$. $f$ étant une fonction de classe $C^1$ sur $\mathbb{R}^n$, convexe sur $C$ et de gradient $L$-Lipschitz sur $C$. On considère le problème (bien posé)

$$\begin{aligned} \text{min.} \quad & f(x) \\ \text{s.c.} \quad & x \in C, \end{aligned}$$

dont on note $p^\star$ l'optimum. Pour le résoudre, on étudie l'*algorithme de Frank-Wolfe*. Cet algorithme initialisé en $x_0 \in C$ quelconque est défini pour tout $k \in \mathbb{N}$ par

$$\begin{cases} s_k \in \arg\min_{s \in C} \langle \nabla f(x_k), s \rangle , \\ x_{k+1} = (1 - \lambda_k) x_k + \lambda_k s_k, \end{cases}$$

où, pour $k \in \mathbb{N}$, $\lambda_k = \frac{2}{k+2}$.

1. (a) Sous les hypothèses satisfaites par $f$, montrer que l'algorithme est bien défini.
   (b) Que présuppose cet algorithme pour pouvoir être appliqué en pratique ?
   (c) À l'aide des résultats de l'exercice 1, expliciter la partie conduisant au calcul de $s_k$ sous forme de pseudo-code dans le cas où l'ensemble $C$ est donné par les contraintes $x \geq 0,\ \|x\|_2 \leq 1$.

   On s'intéresse désormais à la convergence de l'algorithme.
2. Dans cette question uniquement, on se place dans le cas particulier où $n = 1$,
$$f : x \mapsto \tfrac{1}{2}x^2, \quad C = [-1, 1], \quad x_0 = 1.$$
   (a) Montrer que
$$\forall j \in \mathbb{N}, \qquad s_{2j} = -1,\ s_{2j+1} = 1, \quad x_{2j} = \frac{1}{2j+1},\ x_{2j+1} = -\frac{1}{2j+1}.$$
   (b) Montrer alors que la suite $(f(x_k) - p^\star)$ converge vers 0 en donnant la vitesse de convergence.
3. Définissant le diamètre de $C$ via
$$d_C := \sup_{(x,y)\in C^2} \|x - y\|,$$
   on veut montrer que
$$\forall k \in \mathbb{N}^*, \quad f(x_k) - p^\star \leq 2L\frac{d_C^2}{k+2}.$$
   On donne le lemme suivant valable pour une suite positive $(\alpha_k)$, et $M > 0$ :
$$\forall k \in \mathbb{N},\ \alpha_{k+1} \leq (1 - \lambda_k)\alpha_k + M\lambda_k^2 \quad \Longrightarrow \quad \forall k \in \mathbb{N}^*,\ \alpha_k \leq \frac{4M}{k+2}.$$
   (a) Montrer que
$$\forall x, y \in C, \quad f(y) \leq f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|x - y\|^2.$$
   (b) En déduire que
$$\forall k \in \mathbb{N}, \quad f(x_{k+1}) \leq f(x_k) + \lambda_k \langle \nabla f(x_k), x^\star - x_k\rangle + \frac{L}{2}d_C^2\lambda_k^2,$$
   où $x^\star$ est un minimiseur du problème.
   (c) Conclure à l'aide de la suite définie pour $k \in \mathbb{N}$ par $\alpha_k = f(x_k) - p^\star$.
4. À la lumière de l'ensemble des questions précédentes, quels avantages et inconvénients voyez-vous à cet algorithme par rapport à l'algorithme du gradient projeté ?