# Eléments de correction des Exercices et Problèmes complémentaires du chapitre 10

Dilshad Surroop, Delphine Bresch-Pietri

27 janvier 2023

## 1 Minimisation sous contraintes

Soit $J(y_1, y_2) := (\frac{y_1}{2} - a)^2 + (y_2 - b)^2$, $c_1(y_1, y_2) := y_1^2 + y_2^2 - 1$, $c_2(y_1, y_2) = -y_1$, $c_3(y_1, y_2) = -y_2$. Le gradient de $J$ s'écrit $\nabla J(y_1, y_2) = (\frac{1}{2}y_1 - a,\, 2(y_2 - b))^T$.

- Si $a < 0$, alors la première coordonnée de $\nabla J(y_1, y_2)$ est positive. Pour que $\nabla J(y_1, y_2)$ soit dans le cône des contraintes actives, il faut nécessairement $c_2$ active, *i.e.* $y_1 = 0$.
  - Si $b < 0$, alors $\frac{\partial J}{\partial y_2}(y_1, y_2) > 0$, et de même, il faut que $c_3$ soit active. Le seul point où $c_2$ et $c_3$ sont actives en même temps est $(0, 0)$.
  - Si $b > 1$, alors $\frac{\partial J}{\partial y_2}(y_1, y_2) < 0$, et il faut cette fois-ci $c_1$ active. Donc $(y_1, y_2) = (0, 1)$.
  - Si $b \in [0, 1]$, $(y_1, y_2) = (0, b)$ réalise le minimum de $J$.
- Si $a > 1/2$, alors $\frac{\partial J}{\partial y_1}(y_1, y_2) < 0$, donc $c_1$ est active.
  - Si $b < 0$, alors $\frac{\partial J}{\partial y_2}(y_1, y_2) > 0$, donc $c_3$ est ative. Le seul point où $c_1$ et $c_3$ sont actives en même temps est $(1, 0)$.
  - Si $b > 0$, les conditions de Karush-Kuhn-Tucker s'écrivent

$$\begin{aligned}
\frac{y_1}{2} - a + 2\lambda_1 y_1 - \lambda_2 &= 0\\
2y_2 - 2b + 2\lambda_1 y_2 - \lambda_3 &= 0\\
y_1\lambda_2 &= 0\\
y_2\lambda_3 &= 0\\
\lambda_{1,2,3} &\geq 0
\end{aligned}$$

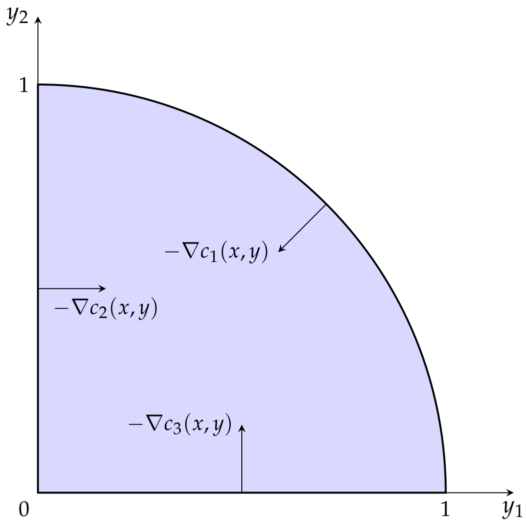
    Si $y_1 = 0$, alors d'après la première équation, $\lambda_2 = -a < 0$, ce qui n'est pas possible, donc $y_1 \neq 0$, et donc $\lambda_2 = 0$. De même, on obtient $\lambda_3 = 0$. $y_1$ et $y_2$ vérifient alors

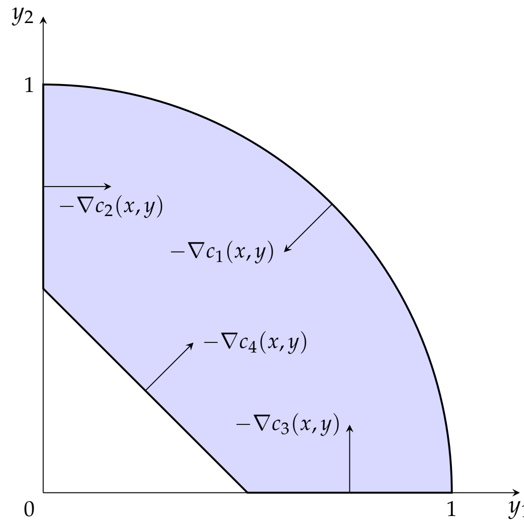
$$y_1 = \frac{2a}{1 + 4\lambda_1}, \qquad y_2 = \frac{b}{1 + \lambda_1}.$$

    De plus, comme $c_1$ est active, $y$ vérifie $y_1^2 + y_2^2 = 1$, soit donc pour $\lambda_1$ :

$$\frac{4a^2}{(1 + 4\lambda_1)^2} + \frac{b^2}{(1 + \lambda_1)^2} = 1$$

    On remarque que $\phi(\lambda_1) := \frac{4a^2}{(1+4\lambda_1)^2} + \frac{b^2}{(1+\lambda_1)^2}$ est continue, vérifie $\phi(0) = 4a^2 + b^2 > 1$, et est strictement décroissante telle que $\phi(\lambda_1) \to 0$ quand $\lambda_1 \to +\infty$. Il existe donc un unique $\lambda_1$ vérifiant $\phi(\lambda_1) = 1$. En notant $\lambda_1^* > 0$ cette solution, on obtient $y_1 = 2a/(1 + 4\lambda_1^*)$ et $y_2 = b/(1 + \lambda_1^*)$.
  - Si $b = 0$, alors en reprenant le point précédent, on a toujours $y_1 \neq 0$, $\lambda_2 = 0$. Cette fois-ci, ou $y_2 = 0$, ou $\lambda_3 = 0$, et alors $2y_2 + 2\lambda_1 y_2 = 0$, soit $y_2 = 0$ (car $\lambda_1 \geq 0$). Dans ce cas, $(1, 0)$ réalise le minimum de $J$.

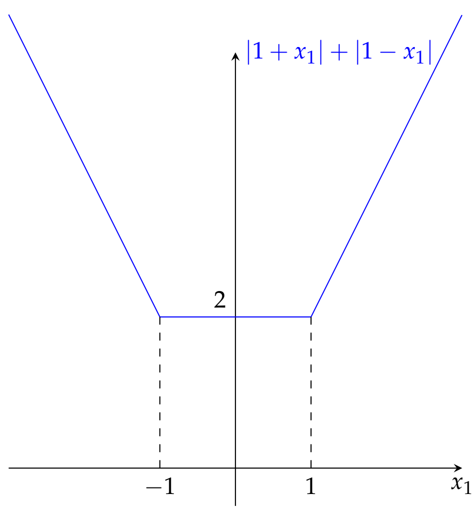



— Si $a \in [0, 1/2]$
 — Si $b < 0$, alors $c_3$ est active, donc $y_2 = 0$. Ainsi $(2a, 0)$ réalise le minimum de $J$.
 — Si $b \in [0, 1]$, alors $(2a, b)$ réalise le minimum de $J$ (puisque $J \geq 0$, et $J(2a, b) = 0$ avec $(2a, b)$ vérifiant les contraintes).
 — Si $b > 1$, alors $c_1$ est active, et on reprend l'étude faite dans le cas $a > 1/2$, $b > 0$.

## 2 Problème de Weber

1. La fonction $\varphi : x \in \mathbb{R}^3 \mapsto \sum_{i=1}^{m} m_i \|x - y_i\|$ est continue, et telle que $\lim_{\|x\| \to +\infty} \varphi(x) = +\infty$. Il existe donc un minimum global au problème de minimisation de $\varphi$ sur $\mathbb{R}^3$.
2. Ce minimum n'est pas nécessairement unique. On considère par exemple $y_1 = (-1, 0)$, $y_2 = (1, 0)$ et $m_1 = 1$, $m_2 = 1$. Alors on cherche à minimiser $\varphi(x) := \|y_1 - x\| + \|y_2 - x\| = 2|x_2| + |x_1 - 1| + |x_1 + 1|$. Or pour $x = (x_1, 0)$ avec $x_1 \in [-1, 1]$, $x$ réalise le minimum de $\varphi$ (cf. graphe), d'où la non-unicité du minimum.



# 3 Optimisation géométrique

1. L'aire du quadrilatère est égale à $\frac{1}{2}ad\sin\theta_1 + \frac{1}{2}bc\sin\theta_2$. Maximiser l'aire revient donc à maximiser $ad\sin\theta_1 + bc\sin\theta_2$. La contrainte est la longueur du côté $BD$ (en pointillé) qui doit être égale pour chacun des triangles $ABD$ et $BCD$. D'après le théorème d'Al Kashi, cette longueur est égale à $a^2 + d^2 - 2ad\cos\theta_1$ pour le triangle $ABD$, et à $b^2 + c^2 - 2bc\cos\theta_2$ pour le triangle $BCD$.
2. Le Lagrangien du problème d'optimisation avec contrainte égalité est
$$\mathcal{L}(\theta_1, \theta_2, \lambda) := -ad\sin\theta_1 - bc\sin\theta_2 + \lambda(a^2 + d^2 - 2ad\cos\theta_1 - b^2 - c^2 + 2bc\cos\theta_2)$$
3. Les points stationnaires du Lagrangien vérifient
$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\theta_1}(\theta_1, \theta_2, \lambda) &= 0 = -ad\cos\theta_1 + 2ad\lambda\sin\theta_1 \\ \frac{\partial\mathcal{L}}{\partial\theta_2}(\theta_1, \theta_2, \lambda) &= 0 = -bc\cos\theta_2 - 2bc\lambda\sin\theta_2 \\ \frac{\partial\mathcal{L}}{\partial\lambda}(\theta_1, \theta_2, \lambda) &= 0 = a^2 + d^2 - 2ad\cos\theta_1 - b^2 - c^2 + 2bc\cos\theta_2 \end{aligned}$$
Les deux premières équations donnent $\lambda = \frac{1}{2}\operatorname{cotan}\theta_2 = -\frac{1}{2}\operatorname{cotan}\theta_1$, donc $\theta_2 = \pi - \theta_1$. En remplaçant $\theta_2$ dans la troisième équation, avec $\cos\pi - \theta_1 = -\cos\theta_1$, on obtient
$$\cos\theta_1 = \frac{a^2 + d^2 - b^2 - c^2}{2(ad + bc)}.$$
Donc $\theta_1 = \arccos\frac{a^2+d^2-b^2-c^2}{2(ab+bc)}$, et $\theta_2 = \pi - \theta_1$.
4. Pour $a = b = c = 1$, $\theta_1 = \arccos 0 = \frac{\pi}{2}$, et $\theta_2 = \frac{\pi}{2}$. Le quadrilatère $ABCD$ est donc un carré. Pour $a = d = 1$, et $b = c = 2 + \sqrt{3}$, on trouve $\theta_1 = \frac{5}{6}\pi$ et $\theta_2 = \frac{\pi}{6}$.

# 4 Tarification de billets d'avion

1. Les hommes d'affaire éprouvent plus de plaisir que les touristes à passer le samedi soir chez eux d'où $z_b > z_n$. Quitte à changer l'échelle de l'agrément, on peut supposer $z_n = 0$. Les hommes d'affaire ayant a priori un revenu plus important que les touristes, il leur est moins pénible de payer le prix du billet d'où $\alpha_n > \alpha_b$. Ces deux pénibilités sont néanmoins positives.
2. Pour un touriste, un simple bilan des sources d'agrément permet d'obtenir $u = z_n - \alpha_n p + t$. Pour un homme d'affaire, son agrément de passer la nuit du samedi sur place étant nul, on a $u = -\alpha_b p + t$.
3. Pour donner envie aux touristes de passer la nuit du samedi sur place, il faut que l'agrément correspondant soit positif, soit $-\alpha_n p_n + t \geq 0$. Pour les hommes d'affaires soient encouragés à effectuer un séjour sans la nuit du samedi sur place, on impose que l'agrément correspondant soit positif soit $z_b - \alpha_b p_b + t \geq 0$.
4. Pour que le voyage avec séjour le samedi soit plus intéressant pour les touristes, on impose que son agrément est plus grand que celui sans, ce qui donne $-\alpha_n p_n + t \geq -\alpha_n p_b + t$. De même, pour les hommes d'affaires, on veut $z_b - \alpha_b p_b + t \geq -\alpha_b p_b + t$. En simplifiant $t$, on a les relations voulues.
5. On souhaite résoudre le problème
$$\max_{p_n, p_b \in \mathbb{R}} \frac{1}{2}(p_n + p_b) = -\min_{p_n, p_b \in \mathbb{R}_+} -\frac{1}{2}(p_n + p_b) \tag{1}$$

sous les contraintes

$$-p_n \leq 0\ , \quad -p_b \leq 0 \tag{2}$$
$$\alpha_n p_n - t \leq 0 \tag{3}$$
$$-z_b + \alpha_b p_b - t \leq 0 \tag{4}$$
$$p_n - p_b \leq 0 \tag{5}$$
$$-z_b + \alpha_b p_b - \alpha_b p_n \leq 0 \tag{6}$$

6. A l'aide des conditions de KKT, et puisqu'on a quatre contraintes, on a $2^4$ cas à traiter.
7. Considérons les relations $-\alpha_n p_n + t \geq 0$ et $z_b - \alpha_b p_b \geq -\alpha_b p_n$. En les additionnant, on obtient $-\alpha_n p_n + t + z_b - \alpha_b p_b \geq -\alpha_b p_n \geq -\alpha_n p_n$ car $\alpha_b < \alpha_n$. On obtient donc $t + z_b - \alpha_b p_b \geq -0$. Par ailleurs, on peut observer que la contrainte $p_n \leq p_b$ n'est alors pas limitante pour le problème (à l'aide par exemple des conditions de KKT).
8. On cherche à résoudre

$$\min_{p_n, p_b \in \mathbb{R}_+} -\frac{1}{2}(p_n + p_b) \tag{7}$$

sous les contraintes

$$-p_n \leq 0\ , \quad -p_b \leq 0 \tag{8}$$
$$\alpha_n p_n - t \leq 0 \tag{9}$$
$$-z_b + \alpha_b p_b - \alpha_b p_n \leq 0 \tag{10}$$

On suppose que les deux dernières contraintes sont actives, cad $p_n = t/\alpha_n$ et $p_b = p_n + z_b/\alpha_b$. Cherchons à résoudre les conditions de KKT correspondantes

$$\begin{cases} -\dfrac{1}{2} + \lambda_3 \alpha_n - \lambda_4 \alpha_b = 0 \\ -\dfrac{1}{2} + \lambda_4 \alpha_b = 0 \end{cases} \Leftrightarrow \begin{cases} \lambda_3 = \dfrac{1}{\alpha_n} > 0 \\ \lambda_4 = \dfrac{1}{2\alpha_b} > 0 \end{cases} \tag{11}$$

Les fonctions étant affines donc convexes, on en déduit qu'il s'agit bien du minimum recherché. A l'optimum, on a donc $p_b - p_n = z_b/\alpha_b$, cad que la ristourne accordée aux touristes correspond au ratio entre l'agrément de passer la nuit du samedi soir chez soi et la pénibilité de payer le prix du billet, pour les hommes d'affaires.
9. Dans ce cas, on s'intéresse au problème

$$\min_{p_n, p_b \in \mathbb{R}_+} -\frac{1}{f_n + f_b}(f_n p_n + f_b p_b) \tag{12}$$

sous les contraintes

$$-p_n \leq 0\ , \quad -p_b \leq 0 \tag{13}$$
$$\alpha_n p_n - t \leq 0 \tag{14}$$
$$-z_b + \alpha_b p_b - \alpha_b p_n \leq 0 \tag{15}$$

où $f_n$ et $f_b$ représentent les flux de touristes et hommes d'affaires respectivement. Le minimum est atteint au même point que précédemment.

10. Il faut alors inclure des pénibilités à payer le prix de la nuit d'hôtel $\beta_n$ et $\beta_p$. La fonction à maximiser est alors $\frac{1}{f_n+f_b}(f_n p_n + f_b p_b) + c$ sous les contraintes

$$-p_n \leq 0\,, \quad -p_b \leq 0 \quad -c \leq 0 \tag{16}$$
$$-\alpha_n p_n - \beta_n c + t \geq 0 \tag{17}$$
$$z_b - \alpha_b p_b - \beta_b c + t \geq 0 \tag{18}$$
$$-\alpha_n p_n - \beta_n c \geq -\alpha_n p_b \tag{19}$$
$$z_b - \alpha_b p_b \geq -\alpha_b p_n - \beta_b c \tag{20}$$

# 5 Inégalité de Kantorovich

1. Le produit scalaire entre $x$ et $x$ s'écrit $x^T x = (\sum_{i=1}^n x_i^2)$. Pour $Q = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, $Qx = (\lambda_1 x_1, \dots, \lambda_n x_n)^T$, d'où $x^T Q x = \sum_{i=1}^n \lambda_i x_i^2$. De même, $x^T Q^{-1} x = \sum \lambda_i^{-1} x_i^2$, d'où le résultat.
2. $q(x)$ vérifie

$$\begin{aligned} q(x) &= \frac{(\sum_{i=1}^n x_i^2)^2}{(\sum_{i=1}^n \lambda_i x_i^2)(\sum_{i=1}^n \lambda_i^{-1} x_i^2)} = \left[\frac{\sum_{i=1}^n \lambda_i x_i^2}{\sum_{i=1}^n x_i^2} \times \frac{\sum_{i=1}^n \lambda_i^{-1} x_i^2}{\sum_{i=1}^n x_i^2}\right]^{-1} \\ &= \left[f(\lambda_1, \dots, \lambda_n, x) \times g(\lambda_1, \dots, \lambda_n, x)\right]^{-1}, \end{aligned}$$

d'où le résultat.
3. On considère $\lambda_1, \dots, \lambda_n$ fixés. Pour $i = 1, \dots, n$, $\lambda_1 \leq \lambda_i \leq \lambda_n$ donc il existe $\alpha_i \in [0,1]$ tel que $\lambda_i = (1-\alpha_i)\lambda_1 + \alpha_i \lambda_n$. Ainsi, $f$ se réécrit

$$\begin{aligned} f(x) &= \lambda_1 \frac{\sum_{i=1}^n (1-\alpha_i) x_i^2}{\sum_{i=1}^n x_i^2} + \lambda_n \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum_{i=1}^n x_i^2} \\ &= \lambda_1 \left(1 - \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum x_i^2}\right) + \lambda_n \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum_{i=1}^n x_i^2}. \end{aligned}$$

On pose alors $\alpha(x) := \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum_{i=1}^n x_i^2}$. La fonction $f$ vérifie alors $f(x) = (1-\alpha(x))\lambda_1 + \alpha(x)\lambda_n$ et $\alpha(x) \in [0,1]$.
Par ailleurs, par la convexité de la fonction inverse, pour $i = 1, \dots, n$,

$$\frac{1}{\lambda_i} = \frac{1}{(1-\alpha_i)\lambda_1 + \alpha_i \lambda_n} \leq (1-\alpha_i)\frac{1}{\lambda_1} + \alpha_i \frac{1}{\lambda_n}$$

Ainsi, il vient

$$\begin{aligned} g(x) = \sum_{i=1}^n \frac{1}{\lambda_i} \frac{x_i^2}{\sum_{i=1}^n x_i^2} &\leq \left(1 - \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum_{i=1}^n x_i^2}\right) + \frac{\sum_{i=1}^n \alpha_i x_i^2}{\sum_{i=1}^n x_i^2} \lambda_n \\ &\leq \big(1-\alpha(x)\big)\lambda_1 + \alpha(x)\lambda_n. \end{aligned}$$

4. En multipliant les deux résultats obtenus en 3, on a

$$f(x)g(x) \leq (1-\alpha(x))^2 + \alpha(x)^2 + (1-\alpha(x))\big(\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}\big).$$

Or $(1-\alpha(x))^2 + \alpha(x)^2 \leq 1$, et pour $\alpha(x) \in [0,1]$, $\alpha(x)(1-\alpha(x)) \leq 1/4$, donc

$$\begin{aligned} f(x)g(x) &\leq 1 + \frac{1}{4}\frac{\lambda_1^2 + \lambda_n^2}{\lambda_1 \lambda_n} = \frac{\lambda_1^2 + \lambda_1\lambda_n + \lambda_n^2}{\lambda_1 \lambda_n} \\ &\leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n}. \end{aligned}$$

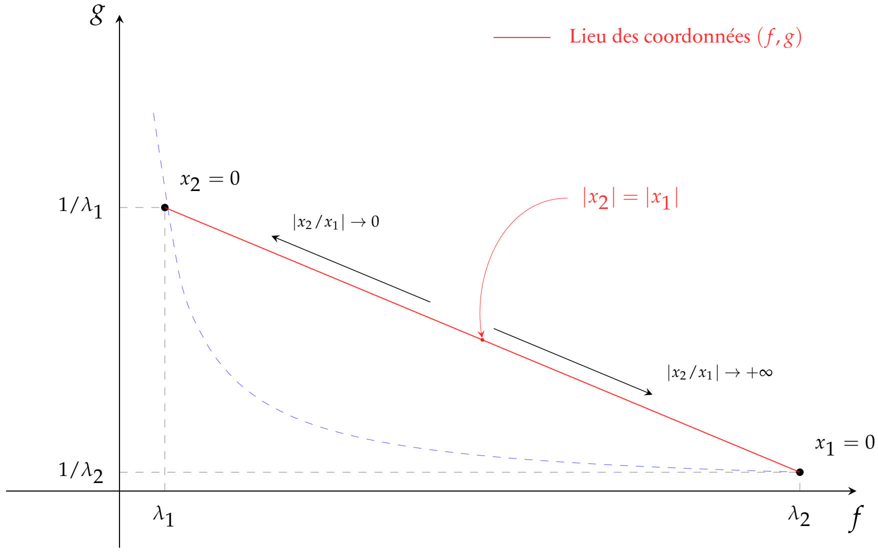



FIGURE 1 – Ensemble des points $(f, g)$ pour $x$ variable, $\lambda_1, \lambda_2$ fixés

En inversant l'inégalité, on obtient bien

$$q(x) = \frac{1}{f(x)g(x)} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}.$$

Pour $Q$ symétrique définie positive, d'après le théorème spectral, il existe une matrice $P$ orthogonale telle que $Q = P\Delta P^{-1}$ avec $\Delta = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ diagonale et $P^{-1} = P^T$. On se ramène au cas diagonal en considérant le changement de variable $\widetilde{x} = Px$. Dans ce cas, $\widetilde{x}^T\widetilde{x} = x^TP^TPx = x^Tx$, et $\widetilde{x}^TQ\widetilde{x} = x^TP^T(P\Delta P^{-1})Px = x^T\Delta x$. Ainsi,

$$q(\widetilde{x}) = \frac{(x^Tx)^2}{(x^T\Delta x)(x^T\Delta^{-1}x)} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}.$$

5. L'algorithme du gradient à pas optimal vérifie

$$\begin{aligned} x^{k+1} &= x^k - l^k\nabla h(x^k) \\ l^k &= \operatorname*{argmin}_{l\in\mathbb{R}} h(x^k + l\nabla h(x^k)). \end{aligned}$$

La matrice $Q$ est symétrique définie positive, donc $h$ est $\alpha$-convexe, tout comme $g : l \mapsto h(x^k - l\nabla h(x^k))$. Ainsi, $l^k$ est caractérisé par $g'(l) = 0$, soit $\nabla h(x^k)^T\nabla h(x^k - l\nabla h(x^k)) = 0$. Ici, $\nabla h(x) = Qx - b$, donc, avec $g^k = \nabla h(x^k) = Qx^k - b$, $(g^k)^T(g^k - lQg^k) = 0$, soit donc $l = \frac{(g^k)^Tg^k}{(g^k)^TQg^k}$, et donc

$$x^{k+1} = x^k - \frac{(g^k)^Tg^k}{(g^k)^TQg^k}g^k.$$

6. Pour le problème de minimisation quadratique $h$, la solution est $x^* = Q^{-1}b$, donc $Q(x^k - x^*) = Qx^k - b = g^k$. De plus,

$$\begin{aligned} E(x^k) = \frac{1}{2}(x^k - x^*)^TQ(x^k - x^*) &= \frac{1}{2}[Q^{-1}(Qx^k - b)]^T(Qx^k - b) \\ &= \frac{1}{2}(g^k)^TQ^{-1}g^k \end{aligned}$$

Ainsi, avec $l = \frac{(g^k)^T g^k}{(g^k)^T Q g^k}$ défini précédemment,

$$\begin{aligned}
E(x^{k+1}) &= \frac{1}{2}(x^k - x^* - lg^k)Q(x^k - x^* - lg^k) \\
&= E(x^k) - l(g^k)^T Q(x^k - x^*) + \frac{1}{2}l^2 (g^k)^T Q g^k \\
&= E(x^k) - \frac{[(g^k)^T g^k]^2}{(g^k)^T Q g^k} + \frac{1}{2}\frac{[(g^k)^T g^k]^2}{(g^k)^T Q g^k} \\
&= E(x^k) - \frac{1}{2}\frac{[(g^k)^T g^k]^2}{(g^k)^T Q g^k \times (g^k)^T Q^{-1} g^k} \times \underbrace{(g^k)^T Q^{-1} g^k}_{=2E(x^k)} \\
&= E(x^k)\big(1 - \frac{[(g^k)^T g^k]^2}{(g^k)^T Q g^k \times (g^k)^T Q^{-1} g^k}\big)
\end{aligned}$$

D'après l'inégalité de Kantorovitch,

$$1 - \frac{[(g^k)^T g^k]^2}{(g^k)^T Q g^k \times (g^k)^T Q^{-1} g^k} \leq 1 - \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2} = \frac{(\lambda_n - \lambda_1)^2}{(\lambda_n + \lambda_1)^2},$$

d'où finalement

$$E(x^{k+1}) \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 E(x^k)$$

7. Meilleur est le conditionnement de la matrice (le ratio de la plus grande et de la plus petite valeur propre de $Q$ se rapproche de 1), meilleure est la vitesse de convergence (le coefficient multiplicatif dans l'inégalité précédente se rapprochant de 0).

# 6 Dualité

1. La solution est $x = 1$, et $|1| = 1$. Le Lagrangien du problème est $\mathcal{L}(x, \lambda) = |x| + \lambda(1 - x)$. La fonction duale vérifie $f(x) = +\infty$ si $x < 1$ et $f(x) = 0$ si $x \geq 1$. D'après le théorème du point selle, la solution du problème vérifie $\mathcal{L}(x, \lambda) = \inf_{x\in\mathbb{R}} \sup_{\lambda\geq 0} \mathcal{L}(x, \lambda) = \inf_{x\in\mathbb{R}} |x| + f(x)$. On retrouve bien $x = 1$ comme solution du problème.
2. Le Lagrangien du problème est $\mathcal{L}(x, \lambda, \nu) = c^T x + \nu^T(Ax - b) - \lambda^T x$. Le problème dual est

$$\sup_{\substack{(\lambda,\nu)\\ \lambda\geq 0}} \inf_{x\in\mathbb{R}^n} \mathcal{L}(x, \lambda, \nu)$$

Pour $\lambda \in \mathbb{R}^n, \lambda \geq 0$ et $\nu \in \mathbb{R}^n$ fixé, le Lagrangien en $x^*$ (où son minimum est atteint) vérifie $\nabla_x \mathcal{L}(x^*, \lambda, \nu) = 0$, soit $c + A^T\nu - \lambda = 0$. En remplaçant donc $\lambda = A^T\nu - c$ dans $\mathcal{L}$, on obtient $\mathcal{L}(x^*, \lambda, \nu) = -\nu^T b = -b^T\nu$. Le problème dual est donc

$$\sup_{\substack{(\lambda,\nu)\\ \lambda\geq 0}} \mathcal{L}(x^*, \lambda, \nu) = \sup_{\substack{(\lambda,\nu)\\ c + A^T\nu - \lambda = 0\\ \lambda\geq 0}} -b^T\nu$$

3. Le Lagrangien du problème est cette fois-ci $\mathcal{L}(x, \lambda) = c^T x + \lambda^T(Ax - b)$, avec $\lambda \in \mathbb{R}^m$, $\lambda \geq 0$. Comme précédemment, à $\lambda$ fixé, le minimum atteint en $x^*$ vérifie $\nabla_x \mathcal{L}(x^*, \lambda) = 0$, soit $c + A^T\lambda = 0$. Donc $\mathcal{L}(x^*, \lambda) = -b^T\lambda$. Le problème dual est donc

$$\sup_{\lambda\geq 0} \mathcal{L}(x^*, \lambda, \nu) = \sup_{\substack{\lambda\geq 0\\ A^T\lambda + c = 0}} -b^T\lambda$$

# 7 Résultats sur la minimisation d'une fonction elliptique

1. Le développement de Taylor de $J$ s'écrit

$$J(v) = J(u) + \nabla J(u)^T(v-u) + \int_0^1 (\nabla J(u+t(v-u)) - \nabla J(u))^T(v-u)\,dt.$$

$\nabla J$ étant M-Lipschitzien (i.e. $\|\nabla J(y) - \nabla J(x)\| \leq M\|y-x\|$), il vient pour $t \in [0,1]$,

$$\|(\nabla J(u+t(v-u)) - \nabla J(u))^T(v-u)\ \| \leq tM\|v-u\|^2.$$

Ainsi, en injectant cette inégalité dans le développement de Taylor

$$\begin{aligned} J(v) - J(u) &\leq \nabla J(u)^T(v-u) + M\|v-u\|^2 \int_0^1 t\,dt \\ &\leq \nabla J(u)^T(v-u) + \frac{M}{2}\|v-u\|^2. \end{aligned}$$

Par ailleurs, par hypothèse, il existe $\alpha > 0$ tel que pour $t \in [0,1]$,

$$(\nabla J(u+t(v-u)) - \nabla J(u))^T(v-u)\ \geq t\alpha\|v-u\|^2.$$

Et de même, on a,

$$\begin{aligned} J(v) - J(u) &\geq \nabla J(u)^T(v-u) + \alpha\|v-u\|^2 \int_0^1 t\,dt \\ &\geq \nabla J(u)^T(v-u) + \frac{\alpha}{2}\|v-u\|^2. \end{aligned}$$

Les deux inégalités fournissent le résultat demandé.

2. La fonction $J$ est fortement convexe (puisque $(\nabla(v)-\nabla(u))^T(v-u) \geq \alpha\|u-v\|^2$ pour $u, v \in \mathcal{U}$), donc d'après le théorème 12 (avec $\mathcal{U}$ convexe) le problème d'optimisation $\min_{u\in\mathcal{U}} J(u)$ admet une unique solution.

3. Théorème du point selle.

4. On suppose qu'il existe $\rho > 0$ tel que $\lambda = P(\lambda + \rho\phi(u))$. Par définition de la projection, $\lambda \in (\mathbb{R}_+)^m$. Pour $i \in \{,1,\ldots,n\}$, si $\lambda_i > 0$, alors nécessairement $\phi_i(u) = 0$. En effet, comme $\phi_i(u) \leq 0$, si on avait $\phi_i(u) < 0$, alors $\lambda + \rho\phi_i(u) < \lambda$, donc $P(\lambda + \rho\phi_i(u)) < \lambda$, ce qui contredit l'hypothèse faite. Donc $\phi_i(u)\lambda_i = 0$ pour tout $i \in \{1,\ldots,n\}$. Ainsi, on a bien pour tout $\mu \in (\mathbb{R}_+)^m$, $\sum_i \phi_i(u)(\mu_i - \lambda_i) \leq 0$.

   Réciproquement, supposons $\lambda \in (\mathbb{R}_+)^m$ et, pour tout $\mu \in (\mathbb{R}_+)^m$, $\sum_i \phi_i(u)(\mu_i - \lambda_i) \leq 0$. En particulier, pour $\mu = 0$, on obtient $\sum_i \phi_i(u)\lambda_i \geq 0$. Comme $\phi_i(u) \leq 0$ et $\lambda_i \geq 0$, alors nécessairement $\phi_i(u)\lambda_i = 0$ pour $i = 1,\ldots,n$. Ainsi, si $\lambda_i = 0$, alors pour tout $\rho > 0$, $\lambda_i + \rho\phi_i(u) < \lambda_i = 0$, donc $P(\lambda_i) = \lambda_i = 0$. Et si $\lambda_i > 0$, alors pour tout $\rho > 0$, $\rho\phi_i(u) = 0$, donc $P(\lambda_i) = P(\lambda + \rho\phi_i(u))$.

5. En reprenant la preuve (réciproque) précédente, avec $\mu = 0$, on obtient immédiatement le résultat désiré.

6. Si $(u,\lambda)$ est un point selle du Lagrangien, alors d'après les conditions KKT, pour $i = 1,\ldots,n$, $\lambda_i\phi_i(u) = 0$ avec $\lambda_i \geq 0$, donc en particulier $\lambda^T\phi_i(u) = 0$. Donc d'après le résultat de la question 5, il existe $\rho > 0$ tel que $\lambda = P(\lambda + \rho\phi(u))$.

   Par définition du point selle, $\mathcal{L}(u,\lambda) = \inf_v \mathcal{L}(v,\lambda)$, donc pour $\theta \in [0,1]$ et $v \in \mathcal{U}$, $\mathcal{L}(u,\lambda) \leq \mathcal{L}(u + \theta(v-u), \lambda)$.

7. Le développement de Taylor de $\mathcal{L}(u+\theta(v-u),\lambda)$ donne

$$\mathcal{L}(u+\theta(v-u),\lambda) = \mathcal{L}(u,\lambda) + \theta\nabla_u\mathcal{L}(u,\lambda)^T(v-u) + o(\theta),$$

où $\nabla_u$ désigne le gradient par rapport à la première varible $u$. D'après l'inégalité de la question 6, pour $\theta \in [0,1]$,

$$\theta\nabla_u\mathcal{L}(u,\lambda)^T(v-u) + o(\theta) \geq 0;$$

soit, en divisant par $\theta > 0$ et en faisant tendre $\theta$ vers 0,

$$\nabla_u\mathcal{L}(u,\lambda)^T(v-u) \geq 0.$$

Le calcul du gradient du Lagrangien par rapport à $u$ donne

$$\nabla_u\mathcal{L}(u,\lambda)^T(v-u) = \nabla J(u)^T(v-u) + \lambda^T J_\phi(u)(v-u)$$

Par ailleurs, le développement de Taylor sur $\phi$ s'écrit

$$\phi(v) = \phi(u) + J_\phi(u)(v-u) + \int_0^1 (J_\phi(u+t(v-u)) - J_\phi(u))(v-u)dt.$$

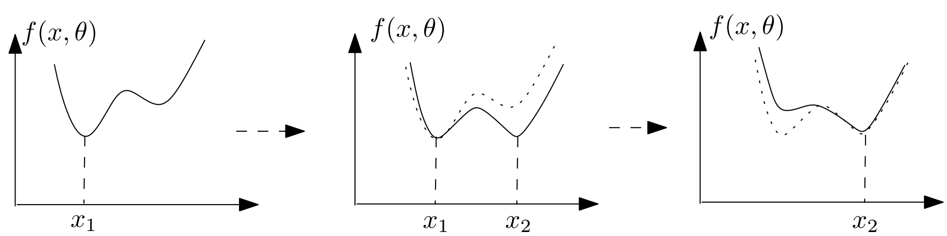
Les fonctions $\phi_i$ sont convexes, donc $(\nabla\phi_i(u+t(v-u)) - \nabla\phi_i(u))^T(v-u) \geq 0$. Comme $\lambda \geq 0$, on obtient donc $\lambda^T(J_\phi(u+t(v-u)) - J_\phi(u))(v-u) \geq 0$. Ainsi, $\lambda^T\phi(v) \geq \lambda^T\phi(u) + \lambda^T J_\phi(u)(v-u) = \lambda^T J_\phi(u)(v-u)$ puisque $\lambda^T\phi(u) = 0$. Ainsi, en reprenant l'inégalité sur le gradient du Lagrangien, on obtient

$$\begin{aligned} 0 &\leq \nabla_u\mathcal{L}(u,\lambda)^T(v-u) = \nabla J(u)^T(v-u) + \lambda^T J_\phi(u)(v-u) \\ &\leq \nabla J(u)^T + \lambda^T\phi(v). \end{aligned}$$

## 8 Sur un théorème de C. Berge

1. (a) Par continuité de f sur un compact.

   (b) On peut considérer une variation continue de $\theta$ induisant le comportement illustré sur la Figure 1b.




$x^\star(\theta) = \{x_1\}$ $\quad$ $x^\star(\theta) = \{x_1, x_2\}$ $\quad$ $x^\star(\theta) = \{x_2\}$

   La variation du nombre d'élements de $x^*(\theta)$ est discontinue.

2. (a) $(\theta_n)$ est convergente et $a$ et $b$ sont continues, aussi $C = \bigcup_{n\in\mathbb{N}} [a(\theta_n); b(\theta_n)]$ est un intervalle, qui contient $(x_n)$. On peut donc en extraire une sous-suite convergente $(x_{\phi(n)})$ de limite $x \in C$.

(b) Par l'absurde, supposone que $x \notin x^\star(\theta)$. Comme $x^\star(\theta)$ est non-vide, il existe $\hat{x} \in x^\star(\theta)$ qui vérifie, par définition,

$$f(\hat{x}, \theta) < f(x, \theta) \tag{21}$$

Soit $(\hat{x}_{\phi(n)})$ une suite convergente de limite $\hat{x}$. Comme $a$ et $b$ sont continues, la suite $([a(\theta_n); b(\theta_n)])$ converge vers $([a(\theta); b(\theta)])$ et que $\hat{x} \in [a(\theta); b(\theta)]$, on peut construire $(\hat{x}_{\phi(n)})$ telle que

$$\forall n \in \mathbb{N}, \quad \hat{x}_{\phi(n)} \in [a(\theta_{\phi n}); b(\theta_{\phi(n)})] \tag{22}$$

Par continuité de $f$ et selon (21), il existe $n_0 \in \mathbb{N}$ tel que, pour $\quad n \geq n_0$, $f(\hat{x}_{\phi(n)}, \theta_{\phi(n)}) < f(x_{\phi(n)}, \theta_{\phi(n)})$. Aussi, pour $n \geq n_0$, $\quad x_{\phi(n)} \notin x^\star(\theta_{\phi(n)})$, ce qui contredit la définition de $(x_n)$.

(c) On a, en supposant l'existence de la limite,

$$\begin{aligned} \lim_{n\to\infty} f^\star(\theta_n) &= \lim_{n\to\infty} f^\star(\theta_{\phi(n)}) \\ &= \lim_{n\to\infty} f(x_{\phi(n)}, \theta_{\phi(n)}) \\ &= f(x, \theta) \quad \text{par continuité de f} \\ &= f^\star(\theta) \quad \text{selon la question précédente.} \end{aligned} \tag{23}$$

Cette dernière égalité assure que la limite existe et est finie, ce qui permet de conclure sur la continuité de $f^\star$.

# 9 Meilleur antécédent par une matrice non-inversible

1. La fonction $u \in \mathbb{R}^n \mapsto \|Au - b\|$ est bien convexe, puisque pour $\lambda \in [0, 1]$, $u, v \in \mathbb{R}^n$, par inégalité triangulaire,

$$\begin{aligned} \|A((1-\lambda)u + \lambda v) - b\| &= \|(1-\lambda)(Au - b) + \lambda(Av - b)\| \\ &\leq (1-\lambda)\|(Au - b) + \lambda(Av - b). \end{aligned}$$

Les fonctions $x \mapsto \|Ax - b\|$ et $\varphi : x \mapsto \|Ax - b\|^2$ possèdent les mêmes points optimaux. Le gradient de $\varphi$ s'écrit $\nabla\varphi(u) = 2AA^T u - 2A^T b$. Le problème étant convexe, $u$ réalise le minimum de $\varphi$ si et seulement si $\nabla\varphi(u) = 0$, i.e. $AA^T u = A^T b$.

2. La matrice $A$ n'est pas inversible, donc il existe $x \in \mathbb{R}^n$, $x \neq 0$, tel que $Ax = 0$, donc $A^T Ax = 0$. Le noyau de $A^T A$ étant non nul, $A^T A$ n'est pas inversible non plus.

3. La matrice $A^T A$ vérifie $(A^T A)^T = A^T A$, donc est symétrique. D'après le théorème spectral, elle est donc diagonalisable en base orthonormée. Soit $(v_1, \ldots, v_n)$ une telle base, i.e. telle que $A^T A v_i = \lambda_i e_i$, avec $\lambda_i$ valeur propre associée à $v_i$. Alors $\lambda_i = \langle A^T A v_i, v_i \rangle = \langle Av_i, Av_i \rangle = \|Av_i\|^2 \geq 0$. La matrice $A^T A$ est donc symétrique positive.
   De plus, $A^T A$ vérifie $A^T A = P \Delta P^T$ avec $P = (v_1, \ldots, v_n)$. Donc $A^T A = \sum_{i=1}^n \lambda_i v_i v_i^T = \sum_{i=1}^m \lambda_i v_i v_i^T$ (puisque $\lambda_i = 0$ pour $i > m$).

4. La famille $(v_1, \ldots, v_n)$ forme une base de $\mathbb{R}^n$. Pour $j = 1, \ldots, m$, $Pv_j = A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T) v_j = A^T A \frac{1}{\lambda_j} v_j = v_j$, et pour $j > m$, $Pv_j = 0$. Ainsi, $\mathrm{Ker}(P) = \{v_{m+1}, \ldots, v_n\}$, donc $\dim(\mathrm{Ker}(P)) = n - m$, et donc rang $P = m$.
   On a $P^2 = A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T) A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T) = A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T)(\sum_{i=1}^m \lambda_i v_i v_i^T)(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T)$.
   Pour $i \neq j$, $\frac{1}{\lambda_i} v_i v_i^T \lambda_j v_j v_j^T = 0$, et pour $i = j$, $\frac{1}{\lambda_i} v_i v_i^T \lambda_i v_i v_i^T = v_i v_i^T$ (les vecteurs $v_i$ forment une base orthonormale). Donc $P^2 = A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T) \sum_{i=1}^m (v_i v_i^T) = A^T A(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T) = P$.

D'après la définition de $P$, on a $\operatorname{Im} P \subset \operatorname{Im} A^T$. De plus, $\dim \operatorname{Im}(P) = \operatorname{rang} P = m = \operatorname{rang} A^T$. Donc $\operatorname{Im} P = \operatorname{Im} A^T$.

L'opérateur $P$ est donc bien un projecteur. Montrons maintenant que $\operatorname{Ker} P = (\operatorname{Im} P)^T$. Le noyau de $P$ est $\operatorname{Ker} P = \{v_{m+1}, \dots, v_n\}$, et son image est $\operatorname{Im} A^T$. Donc $\dim \operatorname{Ker} P = n - m = \dim(\operatorname{Im} A^T)^T$. Soit maintenant $y \in \operatorname{Im} A^T$. Il existe $x \in \mathbb{R}^n$ tel que $y = A^T x$. Pour $i = m+1, \dots, n$, $\langle y, v_i \rangle = \left\langle A^T x, v_i \right\rangle = \langle x, Av_i \rangle = 0$. Donc $\operatorname{Ker} P \subset (\operatorname{Im} A^T)^T$. Par égalité des dimensions, on a donc bien $\operatorname{Ker} P = (\operatorname{Im} A^T)^T$ : $P$ est donc la projection orthogonale sur $\operatorname{Im} A^T$.

5. Soit $u^* := \left(\sum_{i=1}^m \frac{1}{\lambda_i} v_i v_i^T\right) A^T b$. Alors $A^T A u^* = P A^T b$. Or d'après ce qui précède, $P$ est la projection orthogonale sur $\operatorname{Im} A^T$, et comme $A^T b \in \operatorname{Im} A^T$, on a $PA^T b = A^T b$. Ainsi, $A^T A u^* = A^T b$. D'après la question 1, $u^*$ réalise bien l'optimum recherché. Si $m = n$, alors $A^T A$ est inversible, et $(A^T A)^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i} v_i v_i^T$, donc $Au^* = A(A^T A)^{-1} A^T b = AA^{-1}(A^T)^{-1} A^T b = b$.

# 10 Introduction aux méthodes de points intérieurs

1. Il s'agit de la minimisation d'une fonction fortement convexe (son hessien est $P$, définie positive) sur un ensemble convexe. Par théorème du cours, ce problème admet donc une unique solution.
2. Par identification, on définit $N$ matrice $n \times 2n$ telle que $Ny = z$. Comme $z_i = y_i - y_{i+n}$, on obtient $N = (I_n \quad - I_n)$. On a alors

$$\frac{1}{2} x^T Q x + c^T x = \frac{1}{2} y^T N^T P N y + f^T N y = \frac{1}{2} z^T P z + f^T z \tag{24}$$

$$\begin{cases} Ax = b \\ -x \leq 0 \end{cases} \Leftrightarrow \begin{cases} MNy + s = Mz + s = b \\ s \geq 0 \\ y \geq 0 \end{cases} \Leftrightarrow \begin{cases} -s + b = Mz \leq b \\ y \geq 0 \end{cases} \tag{25}$$

3. On définit le lagrangien

$$\mathcal{L} = \frac{1}{2} x^T Q x + c^T x + \lambda^T (Ax - b) - \mu^T x \tag{26}$$

Le problème dual est alors

$$\max_{\mu \geq 0, \lambda} \min_x \mathcal{L}(x, \lambda, \mu) \tag{27}$$

On résout le problème de minimisation en $x$ qui a pour solution

$$x^\star(\lambda, \mu) = Q^{-1} \left(\mu - c - A^T \lambda\right) \tag{28}$$

et on remplace dans le problème dual pour obtenir

$$\max_{\mu \geq 0, \lambda} \frac{1}{2} x^\star(\lambda, \mu)^T Q x^\star(\lambda, \mu) + (c + A^T \lambda - \mu)^T x^\star(\lambda, \mu) - \lambda^T b \tag{29}$$

$$= \max_{\mu \geq 0, \lambda} -\frac{1}{2} x^\star(\lambda, \mu)^T Q x^\star(\lambda, \mu) - \lambda^T b \tag{30}$$

On peut alors résoudre le problème de maximisation en $\lambda$, dont la solution est donnée par

$$\nabla_\lambda \left(-\frac{1}{2} x^\star(\lambda, \mu)^T Q x^\star(\lambda, \mu) - \lambda^T b\right) = -AQ^{-1} A^T \lambda - b = 0 \tag{31}$$

soit $\lambda^\star(\mu) = -(AQ^{-1}A^T)^{-1}b$, où l'inverse existe en supposant que $A$ est de rang plein. On obtient ainsi pour expression finale du problème dual

$$\max_{\mu\geq 0} -\frac{1}{2}x^\star(\lambda^*(\mu),\mu)^T Q x^\star(\lambda^*(\mu),\mu) - \lambda^*(\mu)b \tag{32}$$

$$= \max_{\mu\geq 0} -\frac{1}{2}\left(\mu - c + A^T(AQ^{-1}A^T)^{-1}b\right)^T Q^{-1}\left(\mu - c + A^T(AQ^{-1}A^T)^{-1}b\right) + b^T(AQ^{-1}A^T)^{-T}b \tag{33}$$

$$= \max_{\mu\geq 0} -\frac{1}{2}\mu^T Q^{-1}\mu + d^T\mu + h \tag{34}$$

avec $g = Q^{-T}\left(-c + A^T(AQ^{-1}A^T)^{-1}b\right)$ et $h = -\frac{1}{2}\left(c - A^T(AQ^{-1}A^T)^{-1}b\right)^T Q^{-1}\left(c - A^T(AQ^{-1}A^T)^{-1}b\right) + b^T(AQ^{-1}A^T)^{-T}b$.

4. S'ensuit par définition des conditions de KKT et définition des contraintes égalités.
5. Les conditions de KKT s'écrivent alors

$$\begin{cases} Qx^\star + c + A^T\lambda^* - \mu^* = 0 \\ Ax^\star = b \\ -x^\star \leq 0 \\ \mu^\star \geq 0 \\ \mu_i^\star x_i^\star = 0 \end{cases} \tag{35}$$

6. On définit $f(x) = \frac{1}{2}x^TQx + c^Tx - \varepsilon\sum\ln(x_i)$. Le hessien de $f$ s'écrit $\nabla^2 f(x) = Q + \varepsilon diag(1/x_i^2) \geq Q > 0$. On conclut que $f$ est une fonction fortement convexe, que l'on cherche à minimiser sur l'ensemble $\left\{x \in {\mathbb{R}_+}^{2n+m} \mid Ax = b\right\}$ qui est convexe. Ce problème admet bien une unique solution.
7. Le lagrangien associé au problème pénalisé est $\mathcal{L} = \frac{1}{2}x^TQx + c^Tx - \varepsilon\sum\ln(x_i) + \lambda^T(b - Ax)$ pour $\lambda \in \mathbb{R}^m$. Les conditions de KKT sont

$$\begin{cases} Qx^\star + c - \varepsilon(1/x_i)_{1\leq i\leq n} - A^T\lambda^* = 0 \\ Ax^\star = b \\ x^\star \geq 0 \end{cases} \tag{36}$$

On définit $s \in \mathbb{R}^n$ tel que $s_i = \varepsilon/x_i$ pour obtenir l'équivalence avec les équations voulues.
8. On cherche à annuler la fonction

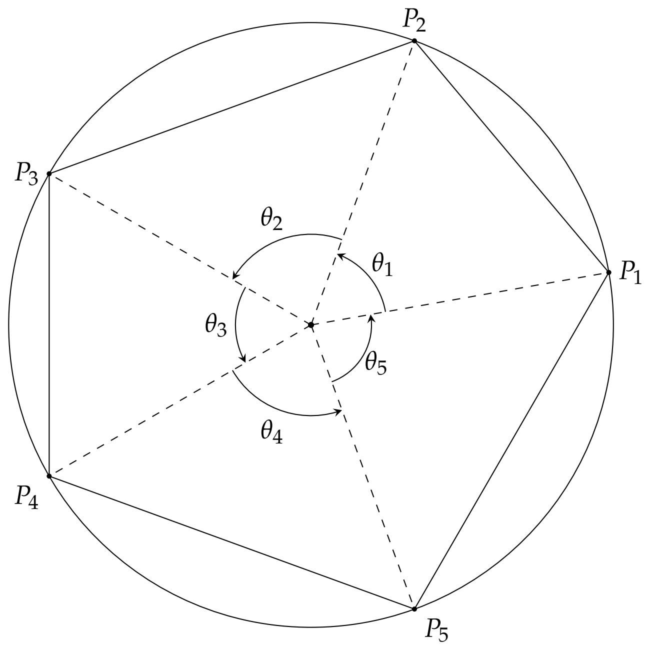
$$g : \mathbb{R}_+^{4n+2m} \times \mathbb{R}^m \to \mathbb{R}^{4n+3m} \tag{37}$$

$$(x, s, \lambda) \mapsto (Ax - b, -Qx + A^T\lambda + s - c, x \bullet s - \varepsilon\mathbf{1}) \tag{38}$$

où $\bullet$ représente le produit vectoriel composante par composante (produit d'Hadamard) et $\mathbf{1}$ le vecteur unité. Son jacobien s'écrit

$$J(x, s, \lambda) = \begin{pmatrix} A & 0 & 0 \\ -Q & I & A^T \\ diag(s_i) & diag(x_i) & 0 \end{pmatrix} \tag{39}$$

Un algorithme de Newton avec projection correspondrait à itérer le calcul $(x_{k+1}, s_{k+1}, \lambda_{k+1}) = P\left((x_k, s_k, \lambda_k) - J(x_k, s_k, \lambda_k)^{-1}g(x_k, s_k, \lambda_k)\right)$ où $P$ projecteur sur $\mathbb{R}_+^{4n+2m} \times \mathbb{R}^m$.
9. La solution du problème pénalisé serait a priori (intuitivement du moins) atteinte en faisant tendre $\varepsilon$ vers 0, cad numériquement en résolvant consécutivement une séquence de problème à $\varepsilon_k$ fixés pour une certaine suite $(\epsilon_k)$ tendant vers 0.
10. Les conditions de stationnarité imposent $x_is_i = \varepsilon$ d'où $x^Ts = \sum_{i=1}^{2n+m} x_is_i = (2n+m)\varepsilon$. A l'optimum, on a donc $\varepsilon = \frac{x^Ts}{2n+m}$. On choisit ainsi $\epsilon$ comme une fraction de cette valeur d'équilibre, pour considérer une décroissance "géométrique" de $\epsilon$.




# 11 Polygones inscrits du cercle de longueur maximale

1. Il s'agit d'une inégalité de Jensen. Par convexité, elle est vérifiée pour $n = 2$. Supposons qu'elle est vraie pour $n \geq 2$, *i.e.*

$$f\big(\frac{1}{n}\sum_{i=1}^{n} x_i\big) \leq \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

On applique d'abord l'inégalité de convexité à $y_1 := \big(\frac{1}{n}\sum_{i=1}^{n} x_i\big)$ et $y_2 := x_{n+1}$ :

$$\begin{aligned} f\big((1-\frac{1}{n+1})\big(\frac{1}{n}\sum_{i=1}^{n} x_i\big) + \frac{1}{n+1}x_{n+1}\big) &\leq (1-\frac{1}{n+1})f\big(\frac{1}{n}\sum_{i=1}^{n} x_i\big) + \frac{1}{n+1}f(x_{n+1}) \\ &\leq (1-\frac{1}{n+1})\frac{1}{n}\sum_{i=1}^{n} f(x_i) + \frac{1}{n+1}f(x_{n+1}) \end{aligned}$$

en utilisant l'hypothèse de récurrence. En remarquant que $(1-\frac{1}{n+1}) \times \frac{1}{n} = \frac{1}{n+1}$, on obtient bien

$$f\big(\frac{1}{n+1}\sum_{i=1}^{n+1} x_i\big) \leq \frac{1}{n+1}\sum_{i=1}^{n+1} f(x_i).$$

2. On considère un polygone régulier $P_1P_2\ldots P_n$ inscrit sur le cercle unité. On note $\theta_i \in [0, 2\pi]$ l'angle $\widehat{P_iOP_{i+1}}$ ($\widehat{P_nOP_1}$ si $i = n$). La longueur du côté $P_iP_{i+1}$ est égale à $2\sin(\theta_i/2)$. On cherche donc à maximiser $2\sum_{i=1}^{n}\sin(\theta_i/2)$. Or d'après l'inégalité précédente, avec $-\sin$ convexe sur $[0, \pi]$,

$$\sum_{i=1}^{n}\sin\big(\frac{\theta_i}{2}\big) \leq n\sin\big(\frac{\theta_1 + \ldots + \theta_n}{2n}\big)$$

avec égalité si et seulement si les $\theta_i$ sont égaux. On maximise donc le périmètre du polygone $P_1P_2\ldots P_n$ lorsque tous les angles $\theta_i$ sont égaux, *i.e.* lorsque le polygone est régulier.

## 12 Parallélépipède maximal

1. On s'intéresse au problème d'optimisation suivant

$$\min_{c(x,y,z)=0} -f(x,y,z)$$

Le Lagrangien est donc

$$\mathcal{L}(x,y,z,\lambda) = -f(x,y,z) + \lambda c(x,y,z).$$

Les conditions de stationnarité sont

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= 0 = -8yz + \frac{2x}{a^2}\lambda \\
\frac{\partial \mathcal{L}}{\partial y} &= 0 = -8xz + \frac{2y}{b^2}\lambda \\
\frac{\partial \mathcal{L}}{\partial z} &= 0 = -8xy + \frac{2z}{c^2}\lambda \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= 0 = c(x,y,z)
\end{aligned}$$

2. Le volume solution est non nul, donc $x, y, z \neq 0$. En utilisant les conditions de stationnarité, on obtient bien $\lambda \neq 0$.
3. Multiplier les trois premières équations entre elles donne $8^3x^2y^2z^2 = 8\frac{xyz}{a^2b^2c^2}\lambda^3$, donc $V = \sqrt{\frac{\lambda^3}{8abc}}$.
4. Multiplier la première équation par $x$ donne $-8xyz + \frac{2x^2}{a^2}\lambda = 0$, soit $-V = \frac{2x^2}{a^2}\lambda$. De même, en multipliant la deuxième équation par $y$ (resp. la troisième par $z$), on obtient $-V = \frac{2y^2}{b^2}$ (resp. $-V = \frac{2z^2}{c^2}$). Comme $\lambda \neq 0$, on a donc $\frac{x^2}{a^2} = \frac{y^2}{b^2} = \frac{c^2}{z^2}$. En remplaçant ces égalités dans la contrainte $c$ (dernière équation), on obtient $3\frac{x^2}{a^2} = 1$, soit $x = a/\sqrt{3}$. On obtient de même $y = b/\sqrt{3}$ et $z = c/\sqrt{3}$, et donc $V = \frac{8abc}{3\sqrt{3}}$.

## 13 Convexité

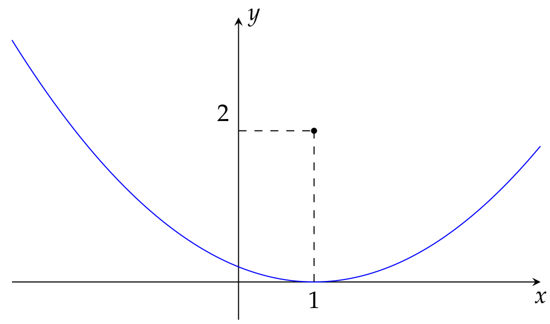
1. Soient $(\lambda^1, \lambda^2) \in (\mathbb{R}^n)^2$, et $t \in [0,1]$. Alors

$$\begin{aligned}
f(x + (1-t)\lambda^1 + t\lambda^2) &= f\big((1-t)(x + \sum_{i=1}^n \lambda_i^1 y_i) + t(x + \sum_{i=1}^n \lambda_i^2 y_i)\\
&\leq (1-t)f\big(x + \sum_{i=1}^n \lambda_i^1 y_i\big) + tf\big(x + \sum_{i=1}^n \lambda_i^2 y_i\big)
\end{aligned}$$

par convexité de $f$. Donc $\varphi := \lambda \in \mathbb{R}^n \mapsto f(x + \sum_{i=1}^n \lambda_i y_i)$ est convexe. De plus, pour $x \in \mathbb{R}^n$, $f(x)$ est de la forme $f(x) = \frac{1}{2}x^TAx + b^Tx$. Donc

$$f(x + \sum_{i=1}^n \lambda_i y_i) = \frac{1}{2}x^TAx + b^T + \frac{1}{2}\big(\sum_{i=1}^n \lambda_i x^TAy_i + \lambda_i y_i^TAx\big) + \sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j y_i^TAy_j.$$

La fonction $\varphi$ est donc elle aussi quadratique.




2. Soit $\lambda \in \mathbb{R}^n$. Alors les dérivées premières et secondes de $\varphi$ s'écrivent

$$\frac{\partial \varphi}{\partial \lambda_i}(\lambda) = \nabla f(x + \sum_{i=1}^{n} \lambda_i y_i)^T y_i,$$
$$\frac{\partial^2 \varphi}{\partial \lambda_j \partial \lambda_i}(\lambda) = y_j^T H_f(x + \sum_{i=1}^{n} \lambda_i y_i) y_i.$$

Une condition simple pour que $\varphi$ soit strictement convexe est de choisir $y_i = e_i$ où $(e_i)_i$ est la base canonique de $\mathbb{R}^n$. Dans ce cas, $\frac{\partial^2 \varphi}{\partial \lambda_j \partial \lambda_i}(\lambda) = y_j^T H_f(x + \sum_{i=1}^{n} \lambda_i y_i) y_i. = [H_f(x + \sum_{i=1}^{n})]_{i,j}$, i.e. $H_\varphi(\lambda) = H_f(x + \sum_{i=1}^{n} \lambda_i y_i$. Comme $f$ est strictement convexe, pour tout $x \in \mathbb{R}^n$ et $u \in \mathbb{R}^n$, $u^T H_f(x) u > 0$. Donc pour tout $\lambda \in \mathbb{R}^n$, $u^T H_\varphi(\lambda) u = u^T H_f(x + \sum_{i=1}^{n} \lambda_i y_i) u > 0$, donc $\varphi$ est strictement convexe.

Le résultat reste vrai si $(y_i)_i$ forme une base orthogonale de $\mathbb{R}^n$.

# 14 Projeté sur une parabole

1. On cherche à minimiser la distance $f(x,y) := (x-1)^2 + (y-2)^2$ sous la contrainte $c(x,y) := 5y - (x-1)^2 = 0$.
2. Le Lagrangien du problème est $\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda c(x,y)$. Les conditions de Lagrange sont donc

$$\frac{\partial \mathcal{L}}{\partial x}(x,y,\lambda) = 0 = 2(1-\lambda)(x-1)$$
$$\frac{\partial \mathcal{L}}{\partial y}(x,y,\lambda) = 0 = 2(y-2) + 5\lambda$$
$$\frac{\partial \mathcal{L}}{\partial \lambda}(x,y,\lambda) = 0 = c(x,y)$$

3. D'après la première équation, $\lambda = 1$ ou $x = 1$. Si $\lambda = 1$, alors $y = -1/2$ d'après la deuxième équation, ce qui est incompatiable avec la contrainte puisque $y = \frac{1}{5}(x-1)^2 > 0$. Donc $x = 1$, et avec la contrainte, on obtient $y = 0$. Donc si $(x,y)$ est solution, alors $(x,y) = (1,0)$. Par ailleurs, $f(x,y)$ tend vers l'infini lorsque $\|(x,y)\|$ tend vers l'infini sur l'espace $c^{-1}\{0\}$. Donc le problème admet bien une solution, qui est donc $(1,0)$.
4. Comme $y = (x-1)^2/5$, on a $f(x,y) = (x-1)^2 + \left(\frac{(x-1)^2}{5} - 2\right)^2$. Cette fonction est convexe en $x$, et sa dérivée $\frac{2}{25}(x-1)(2x^2 - 4x + 7)$, qui est nulle seulement en $x = 1$. Le minimum est donc atteint en $x = 1$. La contrainte donne alors $y = 0$.

# 15 Existence de minima

Les ensembles $\{(x_1,x_2)\in\mathbb{R}^2; x_1^2+x_2^2=1, 0\le x_1\le 1/2, 0\le x_2\}$ et $\{(x_1,x_2)\in\mathbb{R}^2; x_1^2+2x_2^2\le 1, x_1+x_2=10\}$ sont compacts. Les fonctions $(x_1,x_2)\in\mathbb{R}^2\mapsto x_1+x_2$ et $(x_1,x_2)\in\mathbb{R}^2\mapsto x_1+2x_2$ sont continues, donc atteignent leur minimum sur un compact : (61) et (62) possèdent donc des solutions. En revanche, (63) n'admet pas de solution, puisque $\inf_{x_1+x_2=1} x_1=-\infty$.

# 16 Pénalité intérieure

1. Pour $x\in\partial S$ sur la frontière de $S$, il existe $i$ tel que $g_i(x)=0$, donc $J_k(x)=+\infty$. Nécessairement, le minimum $x_k$ de $J_k$ est dans l'intérieur de $S$.
2. Pour $k\in\mathbb{N}$, $J_k$ admet un minimum en $x_k$, donc pour tout $y\in\mathbb{R}^n$, $J_k(x_k)\le J_k(y)$ et $J_{k+1}(x_{k+1})\le J_{k+1}(y)$. En particulier, pour $y=x_{k+1}$ pour la première inégalité, $y=x_k$ dans la seconde, on obtient bien
$$\begin{aligned} J_k(x_k)&\le J_k(x_{k+1})\\ J_{k+1}(x_{k+1})&\le J_{k+1}(x_{k+1})\end{aligned}$$
3. En développant les deux inégalités précédentes, on a
$$\begin{aligned} J(x_k)+\varepsilon_k p(x_k)&\le J(x_{k+1})+\varepsilon_k p(x_{k+1})\\ J(x_{k+1})+\varepsilon_{k+1}p(x_{k+1})&\le J(x_k)+\varepsilon_{k+1}p(x_k)\end{aligned}$$
En multipliant la première inégalité par $\varepsilon_{k+1}$, la seconde par $\varepsilon_k$ et en sommant, on obtient
$$\varepsilon_{k+1}J(x_k)+\varepsilon_k J(x_{k+1})+\varepsilon_k\varepsilon_{k+1}(p(x_k)+p(x_{k+1}))\le\varepsilon_{k+1}J(x_{k+1})+\varepsilon_k J(x_k)+\varepsilon_k\varepsilon_{k+1}(p(x_k)+p(x_{k+1})),$$
soit donc
$$(\varepsilon_k-\varepsilon_{k+1})J(x_{k+1})\le(\varepsilon_k-\varepsilon_{k+1})J(x_k).$$
La suite $\varepsilon_k$ est strictement positive et strictement décroissante, donc $\varepsilon_k-\varepsilon_{k+1}>0$, et donc $J(x_{k+1})\le J(x_k)$.
4. Par définition, $J(x_k)$ vérifie
$$\begin{aligned} J(x_k)&=J_k(x_k)-\varepsilon_k p(x_k)\\ &\le J_k(x_{k+1})-\varepsilon_k p(x_k)\\ &=J(x_{k+1})+\varepsilon_k(p(x_{k+1})-p(x_k))\end{aligned}$$
Donc $\varepsilon_k(p(x_{k+1})-p(x_k))\ge J(x_k)-J(x_{k+1})\ge 0$ d'après la question précédente. On a donc bien $p(x_k)\le p(x_{k+1})$.
5. D'après l'inégalité de la question 2,
$$\begin{aligned} J_{k+1}(x_{k+1})&\le J_{k+1}(x_k)=J(x_k)+\varepsilon_{k+1}p(x_k)\\ &\le J(x_k)+\varepsilon_k p(x_k)=J_k(x_k),\end{aligned}$$
puisque $0<\varepsilon_{k+1}<\varepsilon_k$.
6. La fonction $J$ est continue en $x^*$, donc pour tout $\delta>0$, il existe $\eta>0$ tel que pour $x$ tel que $\|x-x^*\|<\eta$, on a $|J(x)-J(x^*)|<\delta/2$. En prenant un $x^\delta$ dans la boule ouverte de rayon $x^*$ et de rayon $\eta>0$, on a donc bien $J(x^\delta)<J(x^*)+\delta/2$.
7. Pour $k>l$, d'après l'inégalité de la question 5 que l'on itère, on a $J_k(x_k)\le J_l(x_l)$. Et comme le minimum de $J_l$ est atteint en $x_l$, on a $J_l(x_l)\le J_l(x^\delta)$. On a donc bien $J_k(x_k)\le J_l(x^\delta)$.

8. La première inégalité $J(x^*) \leq J(x_k)$ tient du fait que $x^*$ minimise $J$. Par ailleurs, pour $k > l$, $J(x_k) = J_k(x_k) - \varepsilon_k p(x_k) \leq J_k(x_k)$ car $\varepsilon_k p(x_k) \geq 0$. De plus, $J_k(x_k) \leq J_l(x^\delta) = J(x^\delta) + \varepsilon_l p(x^\delta) < J(x^\delta) + \delta$ en utilisant les inégalités des questions 6 et 7.

   On a donc montré que pour tout $\delta > 0$, il existe $l \in \mathbb{N}$ tel que pour $k > l$, $|J(x_k) - J(x^*)| < \delta$. Par définition de la limite, on a bien

$$\lim_{k \to +\infty} J(x_k) = J(x^*)$$

9. D'après la question 1, $x_n$ est une suite de $S$ compact, donc il existe une sous-suite $x_{n_k}$ convergeante dans $S$, disons vers $a$. D'après le résultat précédent, $\lim_{k\to+\infty} J(x_{n_k}) = J(a) = J(x^*)$. Par unicité du mininum $x^*$, on a donc $a = x^*$. Et par unicité du point d'adhérence sur un compact, la suite complète $x_n$ converge vers $a = x^*$.

# 17 Convexité 2

1. Soient $x_1, x_2 \in L_t$ et $\lambda \in [0,1]$. L'ensemble $E$ est convexe, donc $(1-\lambda)x_1 + \lambda x_2 \in L_t$. La fonction $f$ étant convexe, on a

$$\begin{aligned} f((1-\lambda)x_1 + \lambda x_2) &\leq (1-\lambda)f(x_1) + \lambda f(x_2) \leq (1-\lambda)t + \lambda t \\ &\leq t, \end{aligned}$$

   i.e. $(1-\lambda)x_1 + \lambda x_2 \in L_t$, donc $L_t$ est convexe.

2. Soient $x_1, x_2 \in L$ et $\lambda \in [0,1]$. Pour tout $i \in \{1, \ldots, n\}$, par convexité de $g_i$, on a

$$g_i((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)g_i(x_1) + \lambda g_i(x_2) \leq 0.$$

   i.e., $(1-\lambda)x_1 + \lambda x_2 \in L$, donc $L$ est convexe.

3. Soient $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ et $\lambda \in [0,1]$. Alors par convexité de $f$ et $g$

$$\begin{aligned} h\big((1-\lambda)(x_1,y_1) + \lambda(x_2,y_2)\big) &= f((1-\lambda)x_1 + \lambda x_2) + g((1-\lambda)y_1 + \lambda y_2) \\ &\leq (1-\lambda)(f(x_1) + g(y_1)) + \lambda(f(x_2) + g(y_2)) \\ &\leq (1-\lambda)h(x_1,y_1) + \lambda h(x_2,y_2); \end{aligned}$$

   $h$ est donc convexe.

# 18 Réservoir cylindrique

1. Le volume du cylindre en fonction de son diamètre et de sa hauteur est donné par

$$V(d,h) = \pi\big(\frac{d^2}{4}\big)h$$

   Sa surface est

$$S(d,h) = \pi d h + \pi \frac{d^2}{2}$$

2. Le lagrangien du système est

$$\mathcal{L}(d,h,\lambda) = V(d,h) + \lambda(S(d,h) - S_0)$$

3. Conditions d'optimalité

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial d}(d,h,\lambda) = 0 = \frac{1}{2}\pi dh + \lambda\pi h + \lambda\pi d \\
\frac{\partial \mathcal{L}}{\partial h}(d,h,\lambda) = 0 = \frac{1}{4}\pi d^2 + \lambda\pi d \\
\frac{\partial \mathcal{L}}{\partial \lambda}(d,h,\lambda) = 0 = \pi dh + \pi\frac{d^2}{2} - S_0.
\end{aligned}$$

La deuxième équation donne $\lambda = -d/4$. En réinjectant $\lambda$ dans la première équation, il vient $d = h$. Ainsi, la dernière équation fournit $\frac{3}{2}\pi d^2 = S_0$, soit $d = \sqrt{\frac{2S_0}{3\pi}}$ (attention : conditions nécessaires mais pas suffisantes, cf exercice minima liés) .

# 19 Convexité de l'image de petites boules par une application régulière

1. Soit $Q_n(x) = (1-\gamma)(x - A^{-1}b)$ avec $0 < \gamma < 1$. Ici, $P(x) = Ax - b$ ; le Jacobien de $P$ est $P'(x) = A$, qui est bien Lipschitz de constante $L = \|A\|_2$. De plus, $\|P(x) - P'(x)Q_n(x)\| = \|\gamma(Ax - b)\| = \gamma\|P(x)\|$. Ensuite, $\|Q_n(x)\| = \|(1-\gamma)x - A^{-1}b\| = \|(1-\gamma)A^{-1}(Ax-b)\| \leq (1-\gamma)\|A^{-1}\|_2\|P(x)\| = \lambda\|P(x)\|$ avec $\lambda = (1-\gamma)\|A^{-1}\|_2$. Enfin, $L\lambda^2/2 \times \|P(x^0)\| = L(1-\gamma)^2/2 \times \|P(x^0)\|$. Or,

$$\gamma + \frac{L(1-\gamma)^2}{2}\|P(x^0)\| \underset{\gamma\to 1^-}{\longrightarrow} 1^-,$$

donc il existe $\gamma_0 < 1$ tel que $\gamma_0 + \frac{L(1-\gamma)^2}{2} < 1$, ce qui valide les quatre hypothèses (i)-(iv).
2. Par définition, $P(x^{n+1}) = P(x^n - Q_n(x^n))$. En utilisant le développement de Taylor, il vient

$$\begin{aligned}
P(x^{n+1}) &= P(x^n) - \int_0^1 P'(x^n - tQ_n(x^n))Q_n(x^n)\,dt \\
&= \int_0^1 (P(x^n) - P'(x^n - tQ_n(x^n))Q_n(x^n))\,dt
\end{aligned}$$

En écrivant $P'(x^n - tQ_n(x^n)) = P'(x^n) + (P'(x^n - tQ_n(x^n)) - P'(x^n))$, on obtient

$$P(x^{n+1}) = \int_0^1 (P(x^n) - P'(x^n)Q_n(x^n))\,dt + \int_0^1 (P'(x^n) - P'(x^n + tQ_n(x^n))Q_n(x^n)\,dt$$

Soit, en passant à la norme et par inégalité triangulaire,

$$\|P(x^{n+1})\| \leq \int_0^1 \|(P(x^n) - P'(x^n)Q_n(x^n))\|\,dt + \int_0^1 \|(P'(x^n) - P'(x^n + tQ_n(x^n))Q_n(x^n)\|\,dt.$$

En utilisant (ii) sur la première intégrale, et (i) sur le second intégrande, on obtient

$$\|P(x^{n+1})\| \leq \gamma\|P(x)\| + L\int_0^1 t\|Q_n(x^n)\|^2\,dt.$$

Enfin, la majoration donnée par (iii) donne

$$\|P(x^{n+1})\| \leq \gamma\|P(x^n)\| + \frac{L\gamma^2}{2}\|P(x^n)\|^2.$$

3. Par récurrence, on a bien $\delta_0 \leq \delta_0$, et pour $n \in \mathbb{N}$,

$$\begin{aligned}\delta_{n+1} &= \gamma + \frac{L\lambda^2}{2}\|P(x^{n+1})\| \\ &\leq \gamma + \frac{L\lambda^2}{2}(\gamma\|P(x)\| + \frac{L\lambda^2}{2}\|P(x^n)\|^2)\end{aligned}$$

d'après le résultat de la question 2. Or par définition, $\delta_n = \gamma + \frac{L\lambda^2}{2}\|P(x^n)\|$, donc

$$\begin{aligned}\delta_{n+1} &\leq \gamma + \frac{L\lambda^2}{2}\delta_n\|P(x^n)\| \\ &\leq \gamma + \frac{L\lambda^2}{2}\|P(x^n)\| = \delta_n,\end{aligned}$$

puisque, par hypothèse de récurrence, $\delta_n \leq \delta_0 < 1$.

4. On a vu que

$$\begin{aligned}\|P(x^{n+1})\| &\leq \gamma\|P(x^n)\| + \frac{L\lambda^2}{2}\|P(x^n)\|^2 = (\gamma + \frac{L\lambda^2}{2}\|P(x^n)\|)\|P(x^n)\| \\ &\leq \delta_n\|P(x^n)\| \leq \delta_0\|P(x^n)\|\end{aligned}$$

puisque $\delta_n \leq \delta_0$ d'après la question précédente. De cette inégalité, on déduit par itération

$$\|P(x^n)\| \leq \delta_0^n\|P(x^0)\|.$$

Ainsi, en reprenant la définition de $\delta_n$,

$$\delta_n = \gamma + \frac{L\lambda^2}{2}\|P(x^n)\| \leq \gamma + \frac{L\lambda^2}{2}\delta_0^n\|P(x^n)\| \leq \gamma(1 + c_1\delta_0^n).$$

Or pour tout $x$, $1 + x \leq \exp(x)$. En utilisant cette inégalité avec $x = c_1\delta_0^n$, on a $\delta_n \leq \gamma\exp(c_1\delta_0^n)$.

5. Comme $\|P(x^{n+1})\| \leq \delta_n\|P(x^n)\|$, on a

$$\begin{aligned}\|P(x^{n+1})\| &\leq \delta_n\|P(x^n)\| \leq \delta_n\delta_{n-1}\|P(x^{n-1})\| \leq \dots \\ &\leq \|P(x^0)\|\prod_{i=0}^{n}\delta_i.\end{aligned}$$

Or d'après l'inégalité précédente,

$$\prod_{i=0}^{n}\delta_i \leq \gamma^{n+1}\exp\big(c_1\sum_{i=0}^{n}\delta_0^i\big)$$

Par ailleurs, la série géométrique de terme général $\delta_0$ vérifie

$$\sum_{i=0}^{n}\delta_0^i = \delta_0\frac{1-\delta_0^{n+1}}{1-\delta_0} \leq \frac{1}{1-\delta_0},$$

puisque $\delta_0 < 1$. Ainsi, on a bien

$$\|P(x^{n+1})\| \leq \|P(x^0)\|\gamma^{n+1}\exp\big(\frac{c_1}{1-\delta_0}\big).$$

6. Soit $\varepsilon > 0$. Pour $p \geq 0$, d'après la définition de $x^{n+p}$,

$$\begin{aligned} x^{n+p} &= x^{n+p-1} - Q_{n+p-1}(x^{n+p-1}) = x^{n+p-2} - Q_{n+p-2}(x^{n+p-2}) - Q_{n+p-1}(x^{n+p-1}) = \dots \\ &= x^n - \sum_{i=0}^{p-1} Q_{n+i}(x^{n+i}) \end{aligned}$$

Donc, en utilisant l'inégalité (iii),

$$\begin{aligned} \|x^{n+p} - x^n\| &\leq \sum_{i=0}^{p-1} \|Q_{n+i}(x^{n+i})\| \\ &\leq \lambda \sum_{i=0}^{p-1} \|P(x^{n+i})\| \end{aligned}$$

Et en utilisant le résultat de la question 5,

$$\begin{aligned} \|x^{n+p} - x^n\| &\leq \lambda \|P(x^0)\| \exp\big(\frac{c_1}{1-\lambda_0}\big) \gamma^n \sum_{i=0}^{p-1} \gamma^i \\ &\leq \lambda \|P(x^0)\| \exp\big(\frac{c_1}{1-\lambda_0}\big) \gamma^n \times \frac{1}{1-\delta_0} \end{aligned}$$

Le second membre de l'inégalité précédente tend vers 0 (indépendamment de $p$) lorsque $n$ tend vers $+\infty$ puisque $\gamma < 1$. Donc il existe $n_0 \in \mathbb{N}$ tel que pour $n \geq n_0$, et pour $p > 0$, $\|x^{n+p} - x^n\| \leq \varepsilon$. La suite $(x^n)$ est donc de Cauchy dans $\mathbb{R}$ qui est complet, donc converge. Par passage à la limite dans l'inégalité de la question 5, on obtient $\lim_{n\to+\infty} \|P(x^n)\| = 0$, soit donc $\|P(x^*)\| = 0$.

7. Soit $\alpha \in ]0, 2[$. Avec $P'(x)z = P(x)$, on a

$$\|P(x) - P'(x)Q_n(x)\| = \|P(x) - \alpha P'(x)z\| = (1-\alpha)\|P(x)\|.$$

L'hypothèse (ii) est bien vérifiée. Et d'après l'hypothèse d'uniforme inversibilité sur $P'$, $m\|\alpha z\| \leq \|P'(x)(\alpha z)\| = \alpha \|P(x)\|$. Donc $\|Q_n(x)\| = \|\alpha z\| \leq \frac{\alpha}{m}\|P(x)\|$. L'hypothèse (iii) est bien vérifiée avec $\lambda = \alpha/m \geq 0$.

8. Avec $Q_n = \alpha_n z^n$, la suite $x^n$ est les opérateurs $P, Q_n$, vérifient les hypothèses (i), (ii) et (iii) d'après la question précédente. Pour $\|P(x^0)\|$ suffisamment petit, *i.e* tel que $\gamma + \frac{L\lambda^2}{2}\|P(x^0)\| < 1$, l'hypothèse (iv) est aussi vérifiée. D'où, d'après la question 6, la convergence de la séquence $(x^n)$.

9. La suite $(x^n)$ converge vers $x^*$ solution de $P(x) = 0$. De plus, par définition de $x^n$, $x^{n+1} - x^n = -\alpha^n z^n$, donc $\|x^* - x^0\| \leq \sum_{k=0}^{+\infty} \|\alpha^k z^k\| \leq \frac{2}{m} \sum_{k=0}^{+\infty} \|P'(z^k) z^k\|$ (en utilisant l'hypothèse supplémentaire). Donc $\|x^* - x^0\| \leq 2 \sum_{k=0}^{+\infty} \|P(x^k)\|$. Et d'après l'inégalité de la question 5, on obtient

$$\|x^* - x^0\| \leq \frac{2}{m(1-\gamma)} \exp\big(\frac{c_1}{1-\delta_0}\big) \|P(x^0)\|$$

10. Le développement de Taylor de $f$ en $x^0$ s'écrit

$$y^i = f(x^i) = f(x^0) + f'(x^0)(x^i - x^0) + \int_0^1 (f'(x^0 + t(x^i - x^0)) - f'(x^0))(x^i - x^0)\, dt.$$

On note $\varepsilon_i := \int_0^1 (f'(x^0+t(x^i-x^0))-f'(x^0))(x^i-x^0)\,dt$. D'une part, $\|x^i-x^0\| = \frac{1}{2}\|x^1-x^2\|$, et d'autre part,

$$\|f'(x^0+t(x^i-x^0))-f'(x^0))(x^i-x^0)\| \leq tL\|x^i-x^0\|^2$$

On obtient donc

$$\begin{aligned}\|\int_0^1 f'(x^0+t(x^1-x^0))-f'(x^0))(x^1-x^0)\,dt\| &\leq \frac{1}{2}L \times \frac{1}{4}\|x^1-x^2\|^2 \\ &\leq \frac{1}{8}L\|x^1-x^2\|^2.\end{aligned}$$

Ainsi, $y^i = f(x^0)+f'(x^0)(x^i-x^0)+\varepsilon_i$ avec $\|\varepsilon_i\| \leq \frac{L}{8}\|x^1-x^2\|^2$.

11. En utilisant les expressions de $y^1$ et de $y^2$ précédentes,

$$y^0 = \frac{y^1+y^2}{2} = f(x^0)+\frac{1}{2}f'(x^0)(x^1-x^0+x^2-x^0)+\frac{\varepsilon_1+\varepsilon_2}{2}$$

Or, par définition de $x^0$, $x^1+x^2-2x^0 = 0$, et d'après la question précédente, $\varepsilon_0 := \frac{1}{2}(\varepsilon_1+\varepsilon_2)$ vérifie bien $\|\varepsilon_0\| \leq \frac{L}{8}\|x^1-x^2\|^2$

12. On définit $P(x) = f(x)-y^0$. Alors $P(x^0) = f(x^0)-y^0 = -\varepsilon^0$. Or d'après la question 8, si $\|P(x_0)\|$ est suffisament petit (qu'on peut choisir si on prend aussi $\varepsilon$ suffisamment petit), la séquence $x^n$ converge (avec les fonctions $Q_n$ définies à cette question) vers $x^*$ vérifiant $P(x^*) = 0$, *i.e.* $f(x^*) = y^0$. Et avec l'hypothèse d'inversibilité avec une constante $m$, d'après la question 9, $x^*$ vérifie $\|x^*-x^0\| \leq m\|P(x^0)\| = m\|f(x^0)-y^0\|$.

13. La distance entre $x^*$ et $a$ vérifie

$$\|x^*-a\| \leq \|x^*-x^0\|+\|x^0-a\|$$

Comme $x^0 \in B(a,\varepsilon)$, il existe $0<\eta<1$ tel que $\|x^0-a\| = \eta\varepsilon$. De plus, d'après la question précédente, $\|x^*-x^0\| \leq m\|f(x^0)-y^0\| = m\|\varepsilon_0\| = \frac{mL}{8}\|x^1-x^2\|^2 \leq \frac{mL}{8}\varepsilon^2$. Donc pour $\varepsilon$ suffisamment petit, $\frac{mL}{8}\varepsilon^2 < (1-\eta)\varepsilon$, de sorte que $\|x^*-a\| \leq \varepsilon$ : on a alors bien $x^* \in B(a,\varepsilon)$.

14. D'après la question précédente, $y^0 = f(x^*)$ avec $x^* \in B(a,\varepsilon)$. Donc pour $\varepsilon$ suffisamment petit, l'image $F$ de la boule $B(a,\varepsilon)$ est convexe. On a donc montré que pour tout $a$, il existe $\varepsilon > 0$ tel que $f(B(a,\varepsilon)$ est convexe.

15. Soit $f(x) = \begin{pmatrix} x_1x_2-x_1 \\ x_1x_2+x_2 \end{pmatrix}$. La fonction $f$ est de classe $C^1$, son Jacobien s'écrit $f'(x) = \begin{pmatrix} x_2-1 & x_1 \\ x_2 & x_1+1 \end{pmatrix}$. Le Jacobien est linéaire en $x$, donc Lipschitz. De plus, pour $x_2-x_1-1 \neq 0$, son inverse est

$$f'(x)^{-1} = \frac{1}{x_2-x_1-1}\begin{pmatrix} x_1+1 & -x_1 \\ -x_2 & x_2-1 \end{pmatrix} \tag{40}$$

Localement autour de 0, on a pour tout $y$, $\|f'(x)y\| \geq \frac{1}{2}\|y\|$. $f$ vérifie donc les hypothèses du théorème en 0 : il existe donc $\varepsilon > 0$ tel que l'image par $f$ de $B(0,\varepsilon)$ est convexe.

# 20 Programmation linéaire robuste

1. Soit $(x,y) \in E_i^2$ et soit $\lambda \in [0,1]$. $D_i(\lambda x+(1-\lambda)y) = \lambda D_i x+(1-\lambda)D_i y \leq \lambda d_i+(1-\lambda)d_i = d_i$. Donc $\lambda x+(1-\lambda)y \in E_i$ cad $E_i$ convexe.

2. Pour tout $a_i \in E_i$, on doit avoir $a_i^T x \leq b_i$ pour tout $b_i \in [\gamma_i, \delta_i]$, ce qui est équivalent à $a_i^T x \leq \gamma_i$.
3. On obtient directement

$$\max_{D_i a \leq d_i} a^T x \leq \gamma_i\,, \quad i = 1, \ldots, p$$

4. (a) En réécrivant le problème de maximisation en un problème de minimisation, on obtient $\mathcal{L}(a, \lambda) = -a^T x + \lambda^T (D_i a - d_i)$

   (b) Si $D_i a - d_i \leq 0$, alors le sup est atteint pour $\lambda = 0$ et vaut $-a^T x$. Autrement, si l'une des coordonnées est positive, on peut toujours choisir la coordonnée correspondante de $\lambda$ aussi grande que l'on veut et donc $\sup_{\lambda \geq 0} \mathcal{L}(a, \lambda) = +\infty$.

   (c) $\mathcal{L}(a, \lambda) = a^T (D_i^T \lambda - x) - \lambda^T d_i$. Si $x = D_i^T \lambda$, alors la fonction est constante et $\inf_a \mathcal{L}(a, \lambda) = -\lambda^T d_i$. Sinon, on peut toujours choisir $a$ aussi grand que l'on veut en valeur absolue et de signe opposé à celui de la coordonnée non-nulle de $x - D_i^T$, d'où $\inf_a \mathcal{L}(a, \lambda) = -\infty$.

   (d) Le problème de minimisation étudié est linéaire. Selon le théorème 29, il existe donc un point selle qui est solution du problème dual

$$\sup_{\lambda \geq 0} \inf_a \mathcal{L}(a, \lambda) = \sup_{\lambda \geq 0\,,\, x = D_i^T \lambda} \left(-\lambda^T d_i\right) = -\inf_{\lambda \geq 0\,,\, x = D_i^T \lambda} \lambda^T d_i$$

5. Si l'on reporte le problème ci-dessus, on obtient

$$\begin{aligned} &\min \frac{1}{2} x^T Q x + r^T x \\ &\text{t.q.} \quad \min_{\lambda_i \geq 0\,,\quad D_i^T \lambda_i = x} d_i^T \lambda_i \leq \gamma_i\,, \quad i = 1, \ldots, p \end{aligned}$$

   ce qui permet de conclure.

# 21 Dualité de Fenchel-Rockafellar

1. Par définition de la transformée de Fenchel, on a

$$\begin{aligned} f^\star(\varphi) + f(x) &\geq x^T \varphi \\ g^\star(-\varphi) + g(x) &\geq -x^T \varphi \end{aligned}$$

   On obtient le résultat voulu en sommant. En prenant les inf et sup, on obtient $p \geq d$.

2. (a) Soit $(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^n$ et soit $\lambda \in [0, 1]$. Par définition, on obtient

$$V(\lambda x_1 + (1-\lambda) x_2) \leq f(x_0) + g(x_0 + \lambda x_1 + (1-\lambda) x_2)\,, \quad x_0 \in \mathbb{R}^n$$

   Soit $x_0 = \lambda x + (1-\lambda) y$ pour un certain couple $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, alors, les fonction $f, g$ étant convexes, il s'ensuit

$$\begin{aligned} V(\lambda x_1 + (1-\lambda) x_2) &\leq f(\lambda x + (1-\lambda) y) + g(\lambda (x + x_1) + (1-\lambda)(y + x_2)) \\ &\leq \lambda\, [g(x) + f(x + x_1)] + (1-\lambda)\, [f(y) + g(y + x_2)] \end{aligned}$$

   En prenant l'inf sur $(x, y)$ dans le membre de droite, on obtient le résultat voulu.

   (b) $p = V(0)$ est direct. Et, pour tout $x \in \mathbb{R}^n$, $V(x) \geq V(0) + \varphi^T x$ car $\varphi \in \partial V(0)$. Aussi, $V^\star(\varphi) = \sup_{x \in \mathbb{R}^n} \left(\varphi^T x - V(x)\right) = -V(0)$.

(c) On a, par définition,

$$\begin{aligned}
V^\star(\varphi) &= \sup_{x\in\mathbb{R}^n} \left[\varphi^T x - V(x)\right] \\
&= \sup_{x\in\mathbb{R}^n} \left[\varphi^T x - \inf_{s\in\mathbb{R}^n} \left[f(s) + g(s+x)\right]\right] \\
&= \sup_{x\in\mathbb{R}^n} \sup_{s\in\mathbb{R}^n} \left[\varphi^T x - f(s) - g(s+x)\right]
\end{aligned}$$

En poursuivant le calcul précédent, on a

$$\begin{aligned}
V^\star(\varphi) &= \sup_{x\in\mathbb{R}^n} \sup_{s\in\mathbb{R}^n} \left[\varphi^T (x+s-s) - f(s) - g(s+x)\right] \\
&= \sup_{x\in\mathbb{R}^n} \sup_{s\in\mathbb{R}^n} \left[(-\varphi)^T s - f(s) + \varphi^T(s+x) - g(s+x)\right] \\
&= \sup_{(s+x)\in\mathbb{R}^n} \sup_{s\in\mathbb{R}^n} \left[(-\varphi)^T s - f(s) + \varphi^T(s+x) - g(s+x)\right] \\
&= \sup_{u\in\mathbb{R}^n} \sup_{s\in\mathbb{R}^n} \left[(-\varphi)^T s - f(s) + \varphi^T u - g(u)\right] \\
&= \sup_{s\in\mathbb{R}^n} \left[(-\varphi)^T s - f(s)\right] + \sup_{u\in\mathbb{R}^n} \left[\varphi^T u - g(u)\right] \\
&= f^\star(-\varphi) + g^\star(\varphi)
\end{aligned}$$

(d) Des deux question précédentes, on obtient $p = -f^\star(-\varphi) - g^\star(\varphi)$ où $\varphi \in \partial V(0)$. Donc $p \leq \sup_{s\in\mathbb{R}^n} \left[-f^\star(-s) - g^\star(s)\right] = \sup_{s\in\mathbb{R}^n} \left[-f^\star(s) - g^\star(-s)\right] = d$.

3. On a d'une part $p \geq d$ et d'autre part $p \leq d$, d'où $p = d$ qui est l'égalité recherchée. Les inf et sup sont atteints, par exemple, dans le cas où l'une des fonctions $f$ et $g$ est strictement convexe.

4. (a) Comme $g(x) = +\infty$ si $\|x\|_2 > 1$, on en déduit que $\inf_{x\in\mathbb{R}^n}(f(x)+g(x)) = \inf_{\|x\|_2\leq 1}(f(x)+g(x))$.

(b) On a

$$g^*(\varphi) = \sup_{x\in\mathbb{R}^n} \left(\varphi^T x - g(x)\right) = \sup_{\|x\|_2\leq 1} \varphi^T x = \|\varphi\|_2$$

où la dernière inégalité provient de Cauchy-Schwarz.

(c) Il suffit de calculer

$$f^*(\varphi) = \sup_{x\in\mathbb{R}^n} \left(\varphi^T x - \frac{1}{2}\|Ax-b\|_2^2\right)$$

On trouve que ce maximum est atteint pour $x = (A^TA)^{-1}(\varphi + A^Tb)$, ce qui correspond à

$$f^*(\varphi) = \frac{1}{2}\varphi^T Q\varphi + c^T\varphi + d$$

avec $Q = (A^TA)^{-1}$, $c = (A^TA)^{-1}A^Tb$ et $d = -\frac{1}{2}b^Tb$. En appliquant le théorème de Fenchel-Rockafellar, on conclut que

$$p = \max(-f^\star(\varphi) - \|\varphi\|_2) = -\min(\frac{1}{2}\varphi^T Q\varphi + c^T\varphi + d + \|\varphi\|_2)$$

qui est un problème convexe sans contrainte.

## 22 Méthode de faisceaux proximale

**Partie 1 : Instabilité de la méthode des faisceaux**

1. **Application.**
   (a) On a $f(x_0) = 1/2$, $g_0 = x_0 = 1$, $f(x_1) = \varepsilon^2/2$ et $g_1 = x_1 = -\varepsilon$ d'où $\varphi_1(x) = \max\left\{\frac{1}{2} + x - 1, \frac{\varepsilon^2}{2} - \varepsilon(x+\epsilon)\right\}$
   (b) $x_2$ est l'intersection des deux droites déterminées ci-dessus. Soit
$$\frac{1}{2} + x_2 - 1 = -\frac{\varepsilon^2}{2} - \varepsilon x_2 \Leftrightarrow x_2 = \frac{1-\varepsilon}{2}$$
   (c) Le minimum de $f$ est zéro et atteint en l'origine. On observe que $f(x_1) = \epsilon^2/2$ est plus proche de ce minimum que $f(x_2) = (1-\epsilon)^2/8$ : la méthode des faisceaux n'est pas une méthode de descente. Le fait de faire tendre $\varepsilon$ vers 0 empire ce phénomène : plus $x_1$ est proche de $x^*$ et plus $x_2$ en sera éloigné. Algorithmiquement, on parle de méthode instable.
2. Le problème considéré n'est rien d'autre que l'enveloppe de Moreau de paramètre $\mu$ de la fonction $\varphi_k$. Le terme régularisant a pour effet de forcer à ce que les $x_k$ restent relativement proches et donc à ce que $\varphi_k$ soit une approximation raisonnable de $f$ au point de travail. On espère ainsi obtenir une méthode de descente.
3. **Application.**
   (a) $x_2$ est tel que $0 \in \partial\varphi_1(x_2) + \frac{1}{\mu}(x_2 - x_1)$ avec
$$\partial\varphi_1(x_2) = \begin{cases} -\varepsilon & \text{si } x_2 < \dfrac{1-\varepsilon}{2} \\ [-\varepsilon, 1] & \text{si } x_2 = \dfrac{1-\varepsilon}{2} \\ 1 & \text{si } x_2 > \dfrac{1-\varepsilon}{2} \end{cases}$$
   Supposons $x_2 \geq \frac{1-\varepsilon}{2}$, alors, il existe $g \in [-\epsilon, 1]$ tel que
$$0 = g + 3(x_2 + \varepsilon) \Leftrightarrow x_2 = -\varepsilon - \frac{g}{3} \leq -\varepsilon - \frac{\varepsilon}{3} < \frac{1-\varepsilon}{2}$$
   pour $\varepsilon$ suffisamment faible, ce qui est absurde. Aussi, $x_2 < \frac{1-\varepsilon}{2}$ et est tel que
$$0 = -\varepsilon + 3(x_2 + \varepsilon) \Leftrightarrow x_2 = -\varepsilon + \frac{\varepsilon}{3} = -\frac{2}{3}\varepsilon$$
   (b) $x_2$ est maintenant plus proche de l'origine que $x_1$ et ce quelle que soit la valeur de $\varepsilon$. La méthode proximale a été stabilisée.

**Partie 2 : Analyse de convergence**

4. On a $f(x) \geq f(x_i) + g_i^T(x - x_i)$ pour tout $i = 0, \dots, k$ par convexité de $f$ et définition de $g_i$ ce qui permet de conclure.
5. L'inégalité provient directement de l'étape (b) de l'algorithme et du fait que $f(x_k) = f(x_{k+1})$ si $k \notin K$. $f$ étant bornée inférieurement, on a donc $\sum_{k\in K} \delta_k < \infty$. Par ailleurs, par définition de $y_{k+1}$,
$$\varphi_k(y_{k+1}) + \frac{1}{2\mu_k}\|y_{k+1} - x_k\|^2 \leq \varphi_k(x_k) + \frac{1}{2\mu_k}\|x_k - x_k\|^2 = \varphi_k(x_k) \leq f(x_k)$$
   donc $\delta_k \geq 0$ et le résultat voulu s'ensuit.

6. Par définition de $y_{k+1}$, on a $0 \in \partial\varphi_k(y_{k+1}) + \frac{1}{\mu_k}(y_{k+1} - x_k)$. On en déduit

$$\begin{aligned}\varphi_k(y) &\geq \varphi_k(y_{k+1}) + \frac{1}{\mu_k}(x_k - y_{k+1})^T(y - y_{k+1}) \\ &= \varphi_k(y_{k+1}) + \frac{1}{\mu_k}(x_k - y_{k+1})^T(y - x_k) + \frac{1}{\mu_k}\|x_k - y_{k+1}\|^2\end{aligned}$$

7. On a

$$\|y - x_{k+1}\|^2 = \|y - y_{k+1}\|^2 \leq \|y - x_k\|^2 + 2(y - x_k)^T(x_k - y_{k+1}) + \|x_k - y_{k+1}\|^2$$

Et, en utilisant la question précédente,

$$\begin{aligned}\|y - x_{k+1}\|^2 &\leq \|y - x_k\|^2 + 2\mu_k\left(\varphi_k(y) - \varphi_k(y_{k+1}) - \frac{1}{2\mu_k}\|x_k - y_{k+1}\|^2\right) \\ &\leq \|y - x_k\|^2 + 2\mu_k\left(f(y) - f(x_k) + \underbrace{f(x_k) - \varphi_k(y_{k+1}) - \frac{1}{2\mu_k}\|x_k - y_{k+1}\|^2}_{=\delta_k}\right)\end{aligned}$$

8. En remplaçant dans l'inégalité précédente, on a

$$\|z - x_{k+1}\|^2 \leq \|z - x_k\|^2 + 2\mu_k(\delta_k - \eta)$$

On a $\lim \delta_k = 0$ pour $k \to \infty$ et $k \in K$. Ainsi, il existe $k_0 \in \mathbb{N}$ tel que, pour $k \geq k_0$,

$$\|z - x_{k+1}\|^2 \leq \|z - x_k\|^2 - \mu_k\eta$$

En sommant, on obtient

$$\|z - x_N\|^2 \leq \|z - x_{k_0}\|^2 - \eta\sum_{k=k_0}^{N}\mu_k \underset{k\to\infty}{\to} -\infty$$

D'où la contradiction. Ainsi, sous les hypothèses considérées, la suite $x_k$ converge vers un minimum de $f$.

# 23 Programmation sur un cône

**Partie 1 : Applications**

1. (a) Direct en introduisant $t \in \mathbb{R}$ et en notant que l'ajout de la constant $r^TQ^{-1}r$ ne modifie pas le problème.

   (b) On a

   $$z^TQz + 2r^Tz = z^TR^TRz + 2r^TR^{-1}Rz = \|Rz + R^{-1}r\|^2 - r^TQ^{-1}r \tag{41}$$

   ce qui permet de conclure.

   (c) On observe que $t+r^TQ^{-1}r \geq 0$ sur l'ensemble des contraintes. Ainsi, minimiser $t+r^TQ^{-1}r$ sur cet ensemble est équivalent à minimiser $\sqrt{t + r^TQ^{-1}r}$. On conclut avec $\alpha = e_{n+1}$, $A_1 = (R \quad 0)$, $b_1 = R^{-1}r$, $c_1 = e_{n+1}$, $d_1 = 0$, $A_i = 0$, $b_i = 0$, $c_i^T = -m_{i-1}$ la $i-1$ième ligne de $M$ et $d_i = -p_{i-1}$ pour tout $i = 2, \ldots, m$.

2. (b) Direct par définition du problème.

   (c) On a $\max\left\{a_i^Tx \mid a_i \in \mathcal{E}_i\right\} = \max\left\{(\bar{a}_i + P_iu)^Tx \mid \|u\| \leq 1\right\} = \bar{a}_i^Tx + \|P_ix\|$ par Cauchy-Schwarz. Le problème est donc bien de la forme voulue avec $A_i = P_i$, $b_i = 0$, $c_i = \bar{a}_i$ et $d_i = -b_i$.

**Partie 2 : Résolution**

3. Direct par linéarité et par la convexité de la norme.
4. On applique la règle de la chaîne à $g_0 = g^2$. On a $\nabla g_0(x) = 2g(x)\nabla g(x) = 2A_i^T(A_ix + b_i)$, ce qui permet de conclure.
5. Le problème dual est

$$\max_{\lambda\geq 0}\min_x \alpha^T x + \sum_{i=1}^N \lambda_i(\|A_ix + b_i\| - c_i^T x - d_i)$$

La solution du problème de minimisation se traduit par l'équation

$$0 = \alpha + \sum_{i=1}^N \lambda_i \left(\frac{A_i^T(A_ix + b_i)}{\|A_ix + b_i\|} - c_i\right)$$

Avec cette contrainte, la fonction coût se reformule alors comme

$$\begin{aligned}
&\alpha^T x + \sum_{i=1}^N \lambda_i(\|A_ix + b_i\| - c_i^T x - d_i)\\
&= -\sum_{i=1}^N \lambda_i \left(\frac{A_i^T(A_ix + b_i)}{\|A_ix + b_i\|} - c_i\right)^T x + \sum_{i=1}^N \lambda_i(\|A_ix + b_i\| - c_i^T x - d_i)\\
&= \sum_{i=1}^N \lambda_i \frac{b_i^T(A_ix + b_i)}{\|A_ix + b_i\|} - \sum_{i=1}^N \lambda_i d_i
\end{aligned}$$

En définissant $z_i = -\lambda_i \frac{A_ix+b_i}{\|A_ix+b_i\|}$, on obtient donc que la fonction objectif s'écrit

$$-\sum_{i=1}^N \left(b_i^T z_i + d_i\lambda_i\right) \tag{42}$$

sous les contraintes

$$0 = \alpha - \sum_{i=1}^N \left(A_i^T z_i + c_i\lambda_i\right) \tag{43}$$

$$\lambda_i \geq 0\,, \quad \|z_i\| \leq \lambda_i \tag{44}$$

6. (b)

$$\begin{aligned}
\eta(x, z, \lambda) &= x^T \sum_{i=1}^N (A_i^T z_i + c_i\lambda_i) + \sum_{i=1}^N (b_i^T z_i + d_i\lambda_i)\\
&= \sum_{i=1}^N \left[(A_ix + b_i)^T z_i + (c_i^T x + d_i)\lambda_i\right] \geq \sum_{i=1}^N \left[-\|A_ix + b_i\|\|z_i\| + (c_i^T x + d_i)\lambda_i\right] \geq 0
\end{aligned}$$

   (c) Le problème est convexe. Sous réserve que les contraintes soient qualifiées en un minimum, il existe donc un point selle correspondant et il vérifie $\eta(x, z, \lambda) = 0$. On peut donc chercher un tel point pour résoudre le problème de minimisation.
7. On a par la borne supérieure de $\eta$ que $\eta$ converge vers 0 si $\psi$ converge vers $-\infty$. Ainsi, minimiser $\psi$ revient à chercher les points selles du problème et donc un point minimum candidat selon la question précédente. Par ailleurs, par définition de la barrière logarithmique, en partant d'un point strictement faisable, on va itérer et obtenir un autre points strictement faisable. D'où le nom de points intérieurs donné à cette méthode. Le caractère primal-dual est lui lié au fait que cette méthode considère la différence entre les deux problèmes, le saut de dualité.

# 24 Optimisation robuste

1. Il s'agit d'une fonction fortement convexe, définie sur un ensemble convexe.
2. Il s'agit directement des conditions de KKT et de la définition des contraintes actives.
3. On a

$$\begin{aligned} x^* &= -Q^{-1}\left(r + A_{\mathcal{I}}^T \lambda\right) \\ -A_{\mathcal{I}} Q^{-1}\left(r + A_{\mathcal{I}}^T \lambda\right) &= b_{\mathcal{I}} \end{aligned}$$

   soit

$$\lambda = -\left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1}\left(A_{\mathcal{I}} Q^{-1} r + b_{\mathcal{I}}\right)$$

   et donc

$$x^* = Q^{-1}\left(-r + A_{\mathcal{I}}^T\left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1}\left(A_{\mathcal{I}} Q^{-1} r + b_{\mathcal{I}}\right)\right)$$

4. En multipliant (2) par $A_{\mathcal{I}}$, on a la relation $\lambda = -(A_{\mathcal{I}} A_{\mathcal{I}}^T)^{-1} A_{\mathcal{I}}\,(Qx^* + r)$.
5. On note $x^*(\delta)$ la solution du problème (**??**), qui est continue par rapport à $\delta$. Les contraintes $Ax - b$ étant continues par rapport à $x$, on en déduit que toute contrainte non-active en $x^*$ le reste pour $\delta$ suffisamment petit. Enfin, pour les contraintes actives pour la solution nominale, $\lambda$ est également une fonction continue de $\delta$, selon la question précédente. Aussi, pour $\delta$ suffisamment faible, $\lambda_i > 0$ car cela est le cas pour $\delta = 0$. Les contraintes actives du problème nominal sont donc toutes actives pour le problème infinitésimalement perturbé.
6. On débute avec $\lambda$ et $x^*$ donnés par (4) et (5) respectivement et on itère
   $\delta \to \delta + \epsilon$
   $\lambda \to \lambda - \left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1} \epsilon$
   if $any(\lambda) \leq 0$
   break
   else
   $x^* \to x^* + Q^{-1} A_{\mathcal{I}}^T \left(A_{\mathcal{I}} Q^{-1} A_{\mathcal{I}}^T\right)^{-1} \epsilon$
   end

**Partie 2 : Programmation robuste**

1. Pour tout $a_i \in E_i$, on doit avoir $a_i^T x \leq b_i$ pour tout $b_i \in [\gamma_i, \delta_i]$, ce qui est équivalent à $a_i^T x \leq \gamma_i$.
2. On obtient directement

$$\max_{D_i a \leq d_i} a^T x \leq \gamma_i\,, \quad i = 1, \ldots, p$$

3. (a) En réécrivant le problème de maximisation en un problème de minimisation, on obtient $\mathcal{L}(a, \lambda) = -a^T x + \lambda^T (D_i a - d_i)$
   (b) Si $D_i a - d_i \leq 0$, alors le sup est atteint pour $\lambda = 0$ et vaut $-a^T x$. Autrement, si l'une des coordonnées est positive, on peut toujours choisir la coordonnée correspondante de $\lambda$ aussi grande que l'on veut et donc $\sup_{\lambda \geq 0} \mathcal{L}(a, \lambda) = +\infty$.
   (c) $\mathcal{L}(a, \lambda) = a^T (D_i^T \lambda - x) - \lambda^T d_i$. Si $x = D_i^T \lambda$, alors la fonction est constante et $\inf_a \mathcal{L}(a, \lambda) = -\lambda^T d_i$. Sinon, on peut toujours choisir $a$ aussi grand que l'on veut en valeur absolue et de signe opposé à celui de la coordonnée non-nulle de $x - D_i^T$, d'où $\inf_a \mathcal{L}(a, \lambda) = -\infty$.

(d) Le problème est linéaire. Selon le théorème 29, il existe donc un point selle qui est solution du problème dual

$$\sup_{\lambda\geq 0}\inf_a \mathcal{L}(a,\lambda) = \sup_{\lambda\geq 0\,,\, x=D_i^T\lambda}\left(-\lambda^T d_i\right) = -\inf_{\lambda\geq 0\,,\, x=D_i^T\lambda}\lambda^T d_i$$

4. Si l'on reporte le problème auxiliaire étudié aux deux questions précédentes dans le problème de la question 1, on obtient

$$\begin{aligned}&\min \frac{1}{2}x^T Qx + r^T x\\ &\text{t.q.} \quad \min_{\lambda_i\geq 0\,,\quad D_i^T\lambda_i = x} d_i^T\lambda_i \leq \gamma_i\,, \quad i=1,\dots,p\end{aligned}$$

ce qui permet de conclure.

# 25 Méthode de gradient proximale accélérée

1. (a) $\text{Prox}_{lh}(x) = argmin_s \frac{1}{2}s^2 + \frac{1}{2l}(s-x)^2$ qui est atteint pour $ls+s-x=0$ soit $s=\frac{1}{1+l}x = \text{Prox}_{lh}(x)$. On a alors avec l'algorithme du gradient proximal :
$x_1 = \text{Prox}_{lh}(x_0 - lx_0) = \frac{1-l}{1+l}x_0$
$x_2 = \text{Prox}_{lh}(x_1 - lx_1) = \frac{(1-l)^2}{(1+l)^2}x_0$
$x_3 = \frac{(1-l)^3}{(1+l)^3}x_0$

(b) $x_1 = \text{Prox}_{lh}(x_0 - lx_0) = \frac{1-l}{1+l}x_0$
$t_2 = 3/2$
$\beta_2 = 0$
$y_2 = x_1 = \frac{1-l}{1+l}x_0$
$x_2 = \text{Prox}_{lh}(x_1 - lx_1) = \frac{(1-l)^2}{(1+l)^2}x_0$
$t_3 = 2$
$\beta_3 = (3/2-1)/2 = 1/4$
$y_3 = x_2 + \frac{1}{4}(x_2 - x_1) = \frac{(1-l)^2}{(1+l)^2}x_0 + \frac{1}{4}\frac{1-l}{1+l}x_0\left(\frac{1-l}{1+l}-1\right) = \frac{(1-l)^2}{(1+l)^2}x_0 + \frac{1}{4}\frac{1-l}{1+l}\frac{-2l}{1+l}x_0 = \frac{(1-l)}{(1+l)^2}\left(1-\frac{3}{2}l\right)x_0$

(c) On voit que, pour $l < 2/3$, la deuxième itération du gradient proximal accéléré est plus proche de l'origine (le minimiseur recherché) que pour la version standard. En réalité, il suffit que $l \leq 1$ car $g$ est 1-Lipschitzien comme on le verra dans la suite, mais cela n'est pas visible sur les premières itérations.

2. (a)

$$\begin{aligned}g(r_l(x)) &= g(x) + \int_0^1 \nabla g(x+s(r_l(x)-x))^T(r_l(x)-x)ds\\ &= g(x) + \nabla g(x)^T(r_l(x)-x) + \int_0^1 \left[\nabla g(x+s(r_l(x)-x)) - \nabla g(x)\right]^T(r_l(x)-x)ds\\ &\leq g(x) + \nabla g(x)^T(r_l(x)-x) + \frac{L}{2}\|r_l(x)-x\|^2\\ &\leq g(z) + \nabla g(x)^T(r_l(x)-z) + \frac{1}{2l}\|r_l(x)-x\|^2\end{aligned}$$

où la dernière inégalité provient de la convexité de $g$ et du fait que $L \leq 1/l$.

(b) Par définition de l'opérateur proximal, on a

$$0 \in \partial h(r_l(x)) + \frac{1}{l}\left(r_l(x) - x + l\nabla g(x)\right) \tag{45}$$

dont on conclut par convexité de $g$.

(c) En sommant les résultats des deux questions précédentes, on obtient

$$\begin{aligned}
f(z) &\geq f(r_l(x)) - \nabla g(x)^T(r_l(x) - z) - \frac{1}{2l}\|r_l(x) - x\|^2 + \frac{1}{l}(x - r_l(x) - l\nabla g(x))^T(z - r_l(x)) \\
&\geq f(r_l(x)) - \frac{1}{2l}\|r_l(x) - x\|^2 + \frac{1}{l}(x - r_l(x))^T(z - r_l(x)) \\
&\geq f(r_l(x)) + \frac{1}{2l}\|r_l(x) - x\|^2 + \frac{1}{l}(x - r_l(x))^T(z - x)
\end{aligned}$$

3. (a) On obtient directement les formules ci-dessus en utilisant le fait que $r_l(y_{k+1}) = x_{k+1}$.

   (b) S'obtient des deux inégalités précédentes en multipliant la première par $(t_{k+1} - 1)$ et en sommant.

   (c) $t_{k+1}(t_{k+1} - 1) = \frac{k+2}{2}\frac{k}{2} \leq \frac{(k+1)^2}{4} = t_k^2$. Le résultat voulu s'ensuit en multipliant le résultat de la question précédente par $t_{k+1}$ et en utilisant cette inégalité.

   (d) En complétant les carrés (ou par inégalité de Pythagore) et en utilisant la définition de $y_{k+1}$, on a

$$\begin{aligned}
&2l(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \\
&\qquad \|t_{k+1}(x_{k+1} - y_{k+1}) + t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^*\|^2 - \|t_{k+1}y_{k+1} - (t_{k+1} - 1)x_k - x^*\|^2 \\
&\qquad = \|t_{k+1}x_{k+1} - (t_{k+1} - 1)x_k - x^*\|^2 - \|t_{k+1}x_k - (t_k - 1)x_{k-1} - x^*\|^2
\end{aligned}$$

   ce qui est bien l'inégalité recherchée.

4. De la relation de la question précédente, on déduit que $\left(2lt_k^2 v_k + \|u_k\|^2\right)$ est une suite décroissante et donc que

$$v_k \leq \frac{1}{2lt_k^2}\|u_0\|^2$$

# 26 Autour de méthodes de pénalité

**Partie 1 : Méthode de points intérieurs**

1. $R$ est un fermé borné et $f$ est continue. Par théorème de Weierstrass, il existe donc au moins une solution à ce problème de minimisation.
2. $\gamma(c_i(x)) = \infty$ pour $c_i(x) \geq 0$. Or, si $x \notin R^0$, il existe $i \in \{1, \ldots, n\}$ tel que $c_i(x) \geq 0$. Aussi, $f_k(x) = \infty$ si $x \notin R^0$. Le minimum de $f_k$ est donc atteint sur $R^0$.
3. Par définition de $x_{k+1}$, pour tout $x \in \mathbb{R}^n$, $f_{k+1}(x) \geq f_{k+1}(x_{k+1})$ donc en particulier pour $x = x_k$. La suite $(\varepsilon_k)$ étant décroissante, on conclut que

$$f_{k+1}(x_{k+1}) \leq f_{k+1}(x_k) = f(x_k) + \varepsilon_{k+1}p(x_k) \leq f(x_k) + \varepsilon_k p(x_k) = f_k(x_k)$$

4. Par continuité de $f$, il existe $\sigma > 0$ tel que, si $\|x - x^*\| < \sigma$ alors $|f(x) - f(x^*)| = f(x) - f(x^*) < \delta/2$. Par ailleurs, comme $x^* \in R$, que $R = \overline{R^0}$, $B(x^*, \sigma) \cap R \neq \emptyset$. On choisit $x^\delta$ un élément de cet ensemble.
5. En itérant la relation de la question 3, on obtient $f_k(x_k) \leq f_l(x_l)$ pour tout $l \leq k$. Aussi, pour $k \geq K$, $f_k(x_k) \leq f_K(x_K) \leq f_K(x^\delta)$ par définition de $x_K$.

6. L'inégalité de gauche provient de la définition de $x^*$. Celle de droite du fait que

$$f(x_k) \leq f(x_k) + \epsilon_k p(x_k) = f_k(x_k) \leq f_K(x^\delta) < f(x^\delta) + \delta/2 < f(x^*) + \delta/2 + \delta/2$$

pour $k \geq K$ selon les deux question précédentes et par définition de $K$. On obtient le résultat voulu car $\delta$ est quelconque.

7. $x_k \in R$, fermé borné. Ainsi, il existe une sous-suite $x_{\varphi(k)}$ convergente de limite $l$. Par continuité de $f$ et la question précédente, $f(l) = f(x^*)$ dont on déduit $l = x^*$ par unicité du minimiseur de $f$ sur $R$. La suite $x_k$ admettant une unique valeur d'adhérence, elle est convergente, de limite $x^*$.

8. $C \cap R$ est un fermé borné de $R^n$, égal à l'adhérence de son intérieur. On peut donc reprendre l'étude précédente en considérant $C \cap R^0$ et $C \cap R$ en lieu et place de $R^0$ et $R$.

**Partie 2 : Slack variables pour contraintes égalités**

9. On reformule $c_i^{eq}(x) = 0$ sous la forme de deux contraintes inégalités $c_i^{eq}(x) \leq 0$ et $-c_i^{eq}(x) \leq 0$. Cela aboutit à la forme voulue en définissant $c_i = c_i^{in}$ pour $i = 1, \ldots, m_1$ et $c_{m_1+2j-1} = -c_{m_1+2j} = c_i^{eq}$ pour $j = 1, \ldots, m_2$. On ne peut pas appliquer directement la méthode de pénalités intérieures car l'ensemble admissible correspondant est d'intérieur vide (il n'existe pas $x_0$ tel que $c_i^{eq}(x_0) < 0$ et $-c_i^{eq}(x_0) < 0$ simultanément).

10. Par définition, $y \geq 0$. Avec le choix de $d$, on a par ailleurs $-s = c^{in}(x) \leq 0$. On a ainsi $-z \leq 0$. Il suffit donc de définir $g(z) = f(My)$.

11. Le lagrangien du problème est $\mathcal{L} = g(z) - \varepsilon \sum_{i=1}^{2n+m_1} \ln(z_i) + \lambda^T d(z)$. En définissant $w \in \mathbb{R}^{2n+m_1}$ tel que $w_i = \varepsilon / z_i$, on obtient le résultat voulu.

12. On définit $h(z, w, \lambda) = (\nabla g(z) + \sum_{i=1}^{m_1+m_2} \lambda_i \nabla d_i(z) - w, w \bullet z - \varepsilon \mathbf{1}, d)$ en notant $\bullet$ le produit composante par composante de deux vecteurs et $\mathbf{1}$ le vecteur unité.

13. Les espaces d'arrivée et de départ de la fonction $h$ étant de même dimension, on peut employer une méthode de Newton-Raphson avec projection pour résoudre cette équation. L'algorithme correspondant correspond à itérer la relation

$$(z_{l+1}, w_{l+1}, \lambda_{l+1}) = P\Big((z_l, w_l, \lambda_l) - (J_h(z_l, w_l, \lambda_l))^{-1} h(z_l, w_l, \lambda_l)\Big)$$

où $J_h$ jacobien de $h$ et $P$ projecteur sur $\mathbb{R}_+^{4n+2m_1} \times \mathbb{R}^{m_1+m_2}$.

14. Il suffit que le problème soit convexe.

15. Selon la partie précédente, $z_k$ converge vers $z^\star$. Donc, $MNz^*$ est une solution selon la reformulation effectuée.

# 27 Minimisation alternée

1. (a) Par définition de l'inf avec le choix particulier de $\xi = \lambda\xi_1 + (1-\lambda)\xi_2$.

   (b) Par convexité de $f_1$ et $f_2$, on obtient donc

$$f_1 \oplus f_2(\lambda x + (1-\lambda)y) \leq \lambda \left[f_1(\xi_1) + f_2(x - \xi_1)\right] + (1-\lambda)\left[f_1(\xi_2) + f_2(y - \xi_2)\right]$$

   En prenant l'inf en $\xi_1$ et $\xi_2$, on conclut donc.

2. (a) En utilisant la question précédente et le changement de variable $\varphi_2 = \varphi - \varphi_1$, on obtient

$$\begin{aligned}
(f_1^\star \oplus f_2^\star)^\star(x) &= \sup_{\varphi} \left( x^T \varphi - f_1^\star \oplus f_2^\star(\varphi) \right) \\
&= \sup_{\varphi} \left( x^T \varphi + \sup_{\varphi_1} \left[ -f_1^\star(\varphi_1) - f_2^\star(\varphi - \varphi_1) \right] \right) \\
&= \sup_{\varphi} \sup_{\varphi_1} \left( x^T \varphi - f_1^\star(\varphi_1) - f_2^\star(\varphi - \varphi_1) \right) \\
&= \sup_{\varphi_2} \sup_{\varphi_1} \left( x^T (\varphi_1 + \varphi_2) - f_1^\star(\varphi_1) - f_2^\star(\varphi_2) \right)
\end{aligned}$$

   (b) On sépare les sup et on reconnaît la définition de la transformée de Fenchel.

   (c) On a $f_1^{\star\star} + f_2^{\star\star} = f_1 + f_2$ selon le théorème de Moreau-Fenchel, car $f_1$ et $f_2$ sont convexes. En appliquant la transformée de Fenchel, on conclut que $(f_1^\star \oplus f_2^\star)^{\star\star} = (f_1 + f_2)^\star$ et on obtient le résultat voulu car $f_1^\star \oplus f_2^\star$ est convexe selon la question précédente.

3. Par identité de Moreau, on a $x = \mathrm{Prox}_{f_1+f_2}(x) + \mathrm{Prox}_{(f_1+f_2)^\star}(x) = \mathrm{Prox}_{f_1+f_2}(x) + \mathrm{Prox}_{f_1^\star \oplus f_2^\star}(x)$ selon le résultat de la question précédente.

4. 
$$\begin{aligned}
\mathrm{Prox}_{f_1^\star \oplus f_2^\star}(x) &= \arg\min_{s \in \mathbb{R}^n} \left( \min_{\xi} \left( f_1^\star(\xi) + f_2^\star(s - \xi) \right) + \frac{1}{2} \|s - x\|^2 \right) \\
&= \arg\min_{s, \xi \in \mathbb{R}^n} \left( f_1^\star(\xi) + f_2^\star(s - \xi) + \frac{1}{2} \|s - x\|^2 \right)
\end{aligned}$$

   Et on conclut avec le changement de variable $y = \xi$ et $z = s - \xi$.

5. Ce problème est équivalent à

$$\arg \min_{y \in \mathbb{R}^n} \left( f_1^\star(y) + \frac{1}{2} \|y + z - x\|^2 \right) = \mathrm{Prox}_{f_1^\star}(x - z)$$

   En utilisant l'identité de Moreau, on obtient donc $y_z = x - z - \mathrm{Prox}_{f_1}(x - z)$. De même, $z_y = x - y - \mathrm{Prox}_{f_1}(x - y)$.

6. Cet algorithme vise à résoudre le problème de la question 4. On obtiendrait alors $s = x - y - z = \mathrm{Prox}_{f_1^\star \oplus f_2^\star}(x)$ selon les questions 3 et 4. On pourrait alors obtenir cette valeur à l'aide uniquement de la connaissance de $\mathrm{Prox}_{f_1}$ et $\mathrm{Prox}_{f_2}$ selon la question 5.

# 28 Remplissage

On considère le problème

$$\begin{array}{ll}
\displaystyle\min_{x \in \mathbb{R}^n} & \displaystyle -\sum_{i=1}^{n} \ln(\alpha_i + x_i) \\
\text{tel que} & x \geq 0, \\
& \mathbb{1}^T x = 1
\end{array} \tag{46}$$

où $\alpha_i > 0$ constante ($1 \leq i \leq n$) et $\mathbb{1}$ est le vecteur de $\mathbb{R}^n$ dont toutes les composantes sont égales à 1. Il s'agit d'un problème d'allocation de ressources $x_i$ à $n$ entités, de capacités respectives $\ln(\alpha_i + x_i)$. On cherche ainsi à maximiser la capacité totale du réseau, avec des ressources totales fixes ($\mathbb{1}^T x = 1$).

1. La fonction à minimiser est continue et l'ensemble $K = \left\{ x \in \mathbb{R}^n \mid x \geq 0 \text{ et } \mathbb{1}^T x = 1 \right\}$ est un compact. On conclut donc par le théorème de Weierstrass.

2. Les conditions de KKT associées au problème avec contraintes inégalités sont

$$\begin{cases} -\dfrac{1}{\alpha_i + x_i^\star} - \lambda_i + \nu_1 - \nu_2 \ = 0\,,\ i = 1,\dots,n \\ \lambda_i x_i^\star = 0\,,\ i = 1,\dots,n \\ \nu_1(\mathbb{1}^T x^\star - 1) = 0 \\ \nu_2(\mathbb{1}^T x^\star - 1) = 0 \\ x^\star \geq 0\,,\ \mathbb{1}^T x^\star \leq 1\,,\ \mathbb{1}^T x^\star \geq 1 \\ \lambda, \nu_1, \nu_2 \geq 0 \end{cases}$$

On en déduit la forme désirée en définissant $\nu = \nu_1 - \nu_2$.

3. Faisons une distinction de cas sur le fait que $x_i^\star = 0$ ou non. Si $x_i^\star > 0$, alors on obtient $x_i^\star = \frac{1}{\nu} - \alpha_i$, ce qui implique que le membre de droite est strictement positif. Autrement, nécessairement, $x_i^\star = 0$, et alors, $\lambda_i = \nu - \frac{1}{\alpha_i} > 0$ qui est bien positif. En résumé, on obtient donc $x_i^\star = \max\left\{0, \frac{1}{\nu} - \alpha_i\right\}$.

   En remplaçant cette expression dans la contrainte $\mathbb{1}^T x^\star = 1$, on obtient la seconde égalité. Le membre de droite définit une fonction continue et strictement décroissante de $\nu$ sur $]0, 1/\min_i \alpha_i[$, qui tend vers $+\infty$ en 0 et 0 en $1/\min_i \alpha_i$, ce qui implique l'existence et l'unicité de $\nu$ voulues, par théorème des valeurs intermédiaires.

4. Imaginons un terrain partitionné en plusieurs intervalles et dont $\alpha_i$ est la hauteur sur un certain intervalle $i$. Supposons que l'on inonde ce terrain avec une quantité 1 d'eau. $x_i^\star$ correspond alors à la hauteur d'eau qui s'accumule sur l'intervalle $i$. La surface de l'eau est quant à elle à une hauteur $1/\nu$.

# 29 Gradient Projeté

1. (a) $x^\star + t(y - x^\star) = ty + (1-t)x^\star \in C$ par convexité de $C$.

   (b) Un développement de Taylor au premier order implique alors que

   $$f(x^\star + t(y - x^\star)) = f(x^\star) + \nabla f(x^\star)^T t(y - x^\star) + o(t) \geq f(x^\star)$$

   On en déduit le résultat désiré en simplifiant $f(x^\star)$ de part et d'autre, considérant $t > 0$, divisant par $t$ et faisant tendre $t$ vers 0.

   (c) $f$ étant convexe, on sait que $f(y) \geq f(x^\star) + \nabla f(x^\star)^T(y - x^\star)$ pour tout $y \in C$, ce qui permet de conclure.

2. (a) La fonction $h : y \in C \mapsto \|x - y\|^2$ est continue. De plus, pour tout $z \in C$, minimiser $h$ sur $C$ est équivalent à la minimiser sur $C \cap B(z, \|x - z\|)$, qui est un compact. Le théorème de Weierstrass assure donc l'existence d'un minimum.

   (b) Supposons qu'il existe deux minimiseurs $y_1^\star$ et $y_2^\star$ de $h$. Alors, par 2-convexité de la fonction $h$,

   $$h(y_1^\star) \geq h(y_2^\star) + 2(x - y_2^\star)^T(y_1^\star - y_2^\star) + \|y_1^\star - y_2^\star\|^2$$

   ce qui implique $0 \geq 2(x - y_2^\star)^T(y_1^\star - y_2^\star) + \|y_1^\star - y_2^\star\|^2$. Or $2(x - y_2^\star)^T(y_1^\star - y_2^\star) \geq 0$ selon la question 1). Par conséquent, $\|y_1^\star - y_2^\star\| = 0$.

   (c) Direct avec le fait que $\nabla h(y) = 2(y - x)$.

3. (a) Le projecteur s'écrit ici $P_C(x) = \min\{1, \max\{1/2, x\}\}$ et les itérations s'écrivent donc $x_{k+1} = P_C((1 - 2l)x_k)$. On obtient $x_1 = x_2 = 1/2$, qui est le minimum recherché. A contrario, pour un gradient usuel, $x_1 = x_2 = 0$, qui est le minimum de la fonction sur $\mathbb{R}$ entier.

(b) Dans ce cas, $P_C(x) = (\max(0, x_i))_{1\leq i\leq n}$. Les techniques de minimisation fondées sur la dualité (Uzawa, Arrow-Hurwicz) utilisent un gradient projeté pour tenir compte du fait que $\lambda \geq 0$.

(c) Cet algorithme nécessite de disposer d'une évaluation de la projection sur $C$. Selon l'ensemble $C$, il n'est pas toujours possible de disposer d'une formule analytique. En toute généralité, on est donc amené à résoudre un nouveau problème d'optimisation, qui a les mêmes contraintes que le problème originel mais porte sur une fonction quadratique. Ce nouveau problème n'a pas de raison d'être plus simple à résoudre que le problème originel.

4. (a) Selon la formule de Taylor avec reste intégral, on a

$$f(z) = f(x) + \int_0^1 \nabla f(x + t(z-x))^T (z-x)dt$$

En additionnant et soustrayant $\nabla f(x))^T(z-x)$, on obtient

$$\begin{aligned} f(z) &= f(x) + \nabla f(x)^T(z-x) + \int_0^1 (\nabla f(x+t(z-x)) - \nabla f(x))^T (z-x)dt \\ &\leq f(x) + \nabla f(x)^T(z-x) + \int_0^1 \|\nabla f(x+t(z-x)) - \nabla f(x)\| \, \|z-x\| \, dt \end{aligned}$$

selon l'inégalité de Cauchy-Schwarz. La lipschitziannité du gradient de $f$ entraîne par ailleurs l'inégalité suivante qui permet de conclure

$$\int_0^1 \|\nabla f(x+t(z-x)) - \nabla f(x)\| \, \|z-x\| \, dt \leq K \int_0^1 t \|z-x\|^2 \, dt = \frac{K}{2} \|z-x\|^2$$

(b) C'est une conséquence directe de la question 2c) avec $x = x_k - l\nabla f(x_k)$ et $z = x_{k+1}$. En appliquant cette inégalité à $y = x_k$, on en déduit $\|x_k - x_{k+1}\|^2 - l\nabla f(x_k)^T(x_k - x_{k+1}) \leq 0$ cad $\nabla f(x_k)^T(x_{k+1} - x_k) \leq -\frac{1}{l}\|x_k - x_{k+1}\|^2$ et on utilise la question précédente avec $x = x_k$ et $z = x_{k+1}$ pour obtenir le résultat désiré.

5. $f(x_k)$ décroissante et minorée donc convergente et, ainsi, $(x_{k+1} - x_k)$ converge vers zéro.

6. (a) $f$ étant fortement convexe, elle admet un minimum sur $\mathbb{R}^n$. Notons $\bar{x}$ ce minimum. Par $\alpha$-convexité, on a $f(x_k) \geq f(\bar{x}) + \nabla f(\bar{x})^T(x_k - \bar{x}) + \frac{\alpha}{2}\|x_k - \bar{x}\|^2 = f(\bar{x}) + \frac{\alpha}{2}\|x_k - \bar{x}\|^2$. La suite $f(x_k)$ étant convergente selon la question précédente, elle est bornée et on conclut que $(x_k)$ est elle-même bornée.

(b) La suite $(x_k)$ appartient à $C$ qui est fermé et est bornée. Elle admet donc une valeur d'adhérence.

(c) On considère une suite extraite $x_{\varphi(k)}$ qui converge vers $x^\star$. Alors, par continuité, $x_{\varphi(k)+1} = P_C(x_{\varphi(k)} - l\nabla f(x_{\varphi(k)}))$ converge vers $P_C(x^\star - l\nabla f(x^\star))$. Or, selon la question 5, $\left\|x_{\varphi(k)+1} - x_{\varphi(k)}\right\|$ tend vers zéro et on obtient donc le résultat voulu.

(d) Par caractérisation du projecteur (question 2)c)), pour tout $y \in C$,

$$0 \geq (x^\star - l\nabla f(x^\star) - P_C(x^\star - l\nabla f(x^\star)))^T (y - P_C(x^\star - l\nabla f(x^\star))) = -l\nabla f(x^\star)^T(y - x^\star)$$

Et l'on conclut donc à l'aide de la question 1)c) que $x^\star$ est un minimiseur global de $f$ sur $C$.

# 30 Dualité 2

1. Il s'agit de $\max_{\lambda\geq 0} \min_x \mathcal{L}(x, \lambda)$ avec $\mathcal{L}(x, \lambda) = c^T x + \lambda\left(\frac{1}{2}x^T Ax - 1\right)$.

2. $\nabla_x \mathcal{L}(x,\lambda) = c + \lambda A x$ qui s'annule pour $x^\star(\lambda) = -\frac{1}{\lambda} A^{-1} c$, $A$ étant inversible car définie positive. Ainsi,

$$\mathcal{L}(x^\star(\lambda),\lambda) = -\frac{1}{\lambda} c^T A^{-1} c + \lambda \left( \frac{1}{2\lambda^2} c^T A^{-1} A A^{-1} c - 1 \right) = -\frac{1}{2\lambda} c^T A^{-1} c - \lambda$$

Le maximum de cette fonction est atteint pour $\lambda^\star$ tel que $\frac{1}{2(\lambda^\star)^2} c^T A^{-1} c = 1$, cad $\lambda^\star = \sqrt{\frac{1}{2} c^T A^{-1} c}$. Le minimiseur correspondant est $x^\star = x^\star(\lambda^\star) = -\sqrt{\frac{2}{c^T A^{-1} c}} A^{-1} c$.

3. On pose le changement de variable $z = \frac{1}{\sqrt{2}} \sqrt{A} x$ où $\sqrt{A}$ racine carrée de la matrice $A$ (qui existe car $A$ définie-positive, et est elle-même définie positive et donc inversible). On obtient alors le problème voulu avec $d = \sqrt{2}\sqrt{A}^{-1} c$. La solution de ce problème est $z^\star = -\frac{d}{\|d\|}$. On retrouve ainsi

$$x^\star = -\sqrt{2}\sqrt{A}^{-1} \frac{d}{\|d\|} = -\sqrt{2}\sqrt{A}^{-1} \frac{\sqrt{2}\sqrt{A}^{-1} c}{\sqrt{2}\sqrt{c^T A^{-1} c}} = -\sqrt{\frac{2}{c^T A^{-1} c}} A^{-1} c$$

# 31 Complétion de matrice – Problème de Netflix

1. (a) Ces matrices sont de la forme $X = \begin{pmatrix} 1 & 2 \\ \star & 1 \end{pmatrix}$ puisque $\|Y - A \bullet X\|_F^2 = (1 - x_{11})^2 + (2 - x_{12})^2 + (1 - x_{22})^2 = 0$.

   (b) Il s'agit de $X_1 = \begin{pmatrix} 1 & 2 \\ 1/2 & 1 \end{pmatrix}$.

   (c) Pour toute matrice $X \in \mathbb{R}^{2\times 2}$, $L(X) \geq \gamma \operatorname{rg}(X)$. Or $L(X_1) = \gamma$, d'où $L(X) > L(X_1)$ pour toute matrice $X$ de rang 2.

   (d) $X_1$ est l'unique minimiseur de $L$ parmi les matrices de rang 1 et 2. Par ailleurs, la seule matrice de rang zéro est la matrice nulle, pour laquelle $L(0) = \|Y\|_F^2 = 1 + 2^2 + 1 = 6$. Ainsi, si $\gamma > 6$, $L(0) = 6 < L(X_1)$ et donc la solution du problème est la matrice nulle. Dans le cas contraire, la solution du problème est $X_1$. Le paramètre $\gamma$ sert d'arbitrage entre une priorité donnée au faible rang (qui privilégie la solution triviale nulle), et celle accordée au respect des données d'entrée (qui aboutit à la solution $X_1$).

2. Non. Par exemple, dans le cas $n = 1$, on a $rg(X) = 1$ si $X \neq 0$ et 0 sinon. Cette fonction est discontinue et non-convexe.

3. (a) $(\operatorname{rg} + \chi_S)^\star(\varphi) = \sup_{x\in\mathbb{R}} \{\varphi x - \operatorname{rg}(x) - \chi_S(x)\} = \sup_{x\in S} \{\varphi x - \operatorname{rg}(x)\}$. Pour $x \neq 0$, $\varphi x - \operatorname{rg}(x) = \varphi x - 1$ est maximisée sur $S$ pour $x$ du signe de $\varphi$ et admet pour valeur maximale $|\varphi| - 1$. Pour $x = 0$, $\varphi x - \operatorname{rg}(x) = 0$. En regroupant les deux cas, $\sup_{x\in S} \{\varphi x - \operatorname{rg}(x)\} = \max\{|\varphi\,| - 1, 0\}$.

   (b) A partir de la question précédente, on trouve

$$\begin{aligned} \sup_\varphi \{\varphi x - (\operatorname{rg} + \chi_S)^\star(\varphi)\} &= \sup_\varphi \{\varphi x - \max\{|\varphi| - 1, 0\}\} \\ &= \sup \left\{ \sup_{|\varphi|\geq 1} \{\varphi x - |\varphi| + 1\}, \sup_{|\varphi|\leq 1} \{\varphi x\} \right\} \end{aligned}$$

   Si $|x| \geq 1$, le premier terme du sup vaut $+\infty$. Si $|x| \leq 1$, les deux termes valent $|x|$.

4. (a) Par propriété de la biconjuguée, $\operatorname{Epi}(\|\cdot\|_\star + \chi_S)$ est l'enveloppe convexe de $\operatorname{Epi}(rg + \chi_S)$. Or,

$$\begin{aligned} \operatorname{Epi}(rg + \chi_S) &= \left\{ (X,y) \in \mathbb{R}^{n\times n} \times \mathbb{R} \mid y \geq \operatorname{rg}(X) + \chi_S(X) \right\} \\ &= \{(X,y) \in S \times \mathbb{R} \mid y \geq \operatorname{rg}(X)\} = \operatorname{Epi}(rg) \cap (S \times \mathbb{R}) \end{aligned}$$

(b) C'est une conséquence directe de la définition de l'enveloppe convexe sur $S$.

5. (a) $X \mapsto A \bullet X$ est une application linéaire et $\|\cdot\|_F$ est convexe puisque c'est une norme, donc $g$ est convexe. La convexité de $h$ peut être obtenue de façon similaire ou de la question précédente.

   (b) Dans le cas $n = 1$, on a $\|\cdot\|_\star = |\cdot|$ qui est non-différentiable en l'origine.

   (c) On note $g_1(X) = \|X\|_F^2$ qui est telle que

$$\begin{aligned} g_1(X+H) &= \mathrm{Tr}((X+H)^T(X+H)) = \mathrm{Tr}(X^TX) + 2\mathrm{Tr}(X^TH) + \mathrm{Tr}(H^TH) \\ &= g_1(X) + 2\mathrm{Tr}(X^TH) + o(\|H\|_F) \end{aligned}$$

   On en déduit que $g_1$ est différentiable, avec $dg_1(X) \cdot H = 2\mathrm{Tr}(X^TH) = 2\langle X, H\rangle_F$ (le produit scalaire associé à la norme de Frobenius). $g$ étant la composition de la fonction affine $X \mapsto Y - A \bullet X$ et de $g$, elle est différentiable. Par ailleurs, par règle de la chaîne, on a $dg(X) \cdot H = -\langle A \bullet (Y - A \bullet X), H\rangle_F$. Le gradient de $g$ est donc $-A \bullet (Y - A \bullet X) = -A \bullet (Y - X)$. (Plus exactement, $-\langle A \bullet (Y - A \bullet X), H\rangle_F = -(A \bullet (Y - A \bullet X)))_f^T H_f$ en notant $M_f$ le vecteur de $\mathbb{R}^{n^2}$ correspondant à la concaténation des coefficients de $M$ (pour un ordre donné) et le gradient est alors $\nabla g = -(A \bullet (Y - A \bullet X)))_f$, que l'on identifie alors à $-A \bullet (Y - A \bullet X)$.

6. Le cadre étudié correspond en tout point à celui du gradient proximal. Pour l'implémenter, il est nécessaire de disposer d'une évaluation de l'opérateur proximal de $h$.

7. A partir de $X_0$, on itérerait pour un pas $l > 0$ fixe supposé suffisamment petit (le gradient de $g$ est 1-Lipschitzien, comme nécessaire pour l'application de la méthode de gradient proximal) :
   - ☐ $G_k = -A \bullet (Y - X_k)$
   - ☐ $U, V, \sigma = SVD(X_K - lG_k)$
   - ☐ $X_{k+1} = U\mathrm{diag}(\mathrm{Prox}_{l\gamma\|\cdot\|_1}(\sigma))V$

   où une expression explicite de $\mathrm{Prox}_{\mu\|\cdot\|_1}$ est

$$(\mathrm{Prox}_{\mu\|\cdot\|_1}(x))_i = \begin{cases} x_i - \mu & \text{si } x_i \geq \mu \\ 0 & \text{si } |x_i| \leq l \\ x_i + \mu & \text{si } x_i \leq -\mu \end{cases}$$

   comme démontré dans le corrigé du TD5, exercice 5.1.