Exercises : Markov Decision Process

June 12, 2023

Exercise 1 (A simple MDP)**.** *Let* $\mathcal{X} = \{0, 1, 2, 3\}$, $\mathcal{A} = \{0, 1\}$*. Let* $(\boldsymbol{X}_t)_{t \in [\![1,5]\!]}$ *be a controlled Markov chain, such that, if* $a = 0$*, it stays in its state, and if* $a = 1$ *it has a probability* 0.5 *of going* 1 *up (if possible, otherwise stay in place), and* 0.5 *of going* 1 *down (if possible, otherwise stay in place).*
Solve by Dynamic Programming the following optimization problem.

$$\text{Max} \qquad \mathbb{E}\Big[\sum_{t=0}^{4} \boldsymbol{X}_t^2 \mid \boldsymbol{X}_0 = 0\Big]$$

You can represent the cost-to-go and the optimal policy as matrices, each column representing one time-step.

Exercise 2. *Consider a unit that have* 3 *possible states : New, Working, Broken. When the unit is New at the beginning of one year, it will be Working at the beginning with probability* 0.75 *and Broken with probability* 0.25*. If it is in Working condition it can be either maintained or not. If maintained, for a cost of* 2*, it stay in Working condition with probability* 1*. If not maintainted, there is a probability of* 0.5 *of staying in the same condition, and of* 0.5 *of being Broken. If broken you can either stay this way, for a cost of* 5*, or repair it for a cost of* 10*, making it new for the next step.*
We want to manage the unit over an horizon of $T = 5$ *steps, starting with a new unit. Find the policy with minimal expected cost.*

Exercise 3 (Optimal stopping time)**.** *Consider the following "push your luck" game. At turn t the player gain* 1 *point with probability* $0 < p < 1$*, and loose everything with probability* $1 - p$*. At the end of the turn she chooses to stop, earning her current points or continue - with the risk of loosing all. Solve the problem of maximizing expected earned points directly and by dynamic programming.*

Answers: We see that if it is optimal to stop when you reach x point, for any $y > x$ it is also optimal to stop. So the problem consists in choosing the number of points (or equivalently the number of turn) after which to stop. After x turn you win x points, with probability p^x. Thus the expected value is $f(x) = xp^x$, and we have $f'(x) = (1-p)^x(1 + x\ln(1-p))$ which is a concave function maximized for $x = -1/\ln(1-p)$. The optimal number of turn is thus $\lfloor -1/\ln(1-p) \rfloor$ or $\lceil -1/\ln(1-p) \rceil$.
The state x is the current number of point. The Bellman equation reads $V(x) = \max\{x, pV(x+1)\}$. For x large enough we can guess that it is optimal to stop, thus $V(x) = x$.

Exercises: Convex analysis

June 12, 2023

Convex sets

Exercise 1 (Perspective function)**.** *Let* $P : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ *be the* perspective function *defined as* $P(x,t) = x/t$, *with* $\mathrm{dom}(P) = \mathbb{R}^n \times \mathbb{R}_+^*$.

1. *Show that the image by* P *of the segment* $[\binom{x}{s}, \binom{y}{t}]$ *is the segment* $[P(\binom{x}{s}), P(\binom{y}{t})]$, *i.e.* $P([\binom{x}{s}, \binom{y}{t}]) = [P(\binom{x}{s}), P(\binom{y}{t})]$.
2. *Show that, if* $C \subset \mathbb{R}^n \times \mathbb{R}_+^*$ *is convex, then* $P(C)$ *is convex.*
3. *Show that, if* $D \subset \mathbb{R}^n$, *then* $P^{-1}(D)$ *is convex.*

Answers:

1. Let x and y be element of $\mathbb{R}^n \times \mathbb{R}_+^*$.
$$\begin{aligned} &P(\theta\tbinom{x}{s} + (1-\theta)\tbinom{y}{t}) \\ &= \frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} \\ &= \mu P(\tbinom{x}{s}) + (1-\mu)P(\tbinom{y}{t}) \end{aligned}$$
with $\mu(\theta) = \frac{\theta s}{\theta s + (1-\theta)t}$. Note that $\theta \mapsto \mu$ is monotonous, and $\mu([0,1]) = [0,1]$. Thus, $P([x,y]) = [P(x), P(y)]$.

2. Consider two elements of $P(C)$, $P(x)$ and $P(y)$. To show convexity we need to show that $[P(x), P(y)] \subset P(C)$. By 1. we have $[P(x), P(y)] = P([x,y])$ and $[x,y] \subset C$ by convexity of C.

3. Now assume that $\binom{x}{s} \in P^{-1}(C)$ and $\binom{y}{t} \in P^{-1}(C)$. We need to show that $\frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} \in C$. This comes from $\frac{\theta x + (1-\theta)y}{\theta s + (1-\theta)t} = \mu(x/t) + (1-\mu)(y/s)$ with $\mu = \frac{\theta t}{\theta t + (1-\theta)s}$.

Exercise 2 (Dual cones)**.** *Recall that, for any set* $K \subset \mathbb{R}^n$, $K^\oplus := \{y \in \mathbb{R}^n \mid \forall x \in K,\ \langle y, x\rangle \geq 0\}$. *We say that* K *is self dual if* $K^\oplus = K$.

1. *Show that* $K = \mathbb{R}_+^n$ *is self dual.*
2. *We consider the set of symmetric matrices* S_n *with the scalar product* $\langle A, B\rangle = \mathrm{tr}(AB)$. *Show that* $K = S_n^+(\mathbb{R})$ *is self dual.*
3. *Let* $\|\cdot\|$ *be a norm, show that* $K = \{(x,t) \mid \|x\| \leq t\}$ *has for dual* $K^\oplus = \{(z,\lambda) \mid \|z\|_\star \leq \lambda\}$, *where* $\|z\|_\star := \sup_{x:\|x\|\leq 1} z^\top x$.

Answers:

1. obvious

2. Let $Y \in S_n \setminus S_n^+$. Then there exists $v \in \mathbb{R}^n$, $v^\top Y v < 0$. Moreover, $v^\top Y v = tr(v^\top X v) = tr(v^\top v X) < 0$. Hence we have $X = v^\top v \in S_n^+$ such that $\langle Y, X\rangle < 0$, i.e. $Y \notin (S_n^+)^\oplus$.

   On the other hand, consider $Y \in S_n^+$. We have the following decomposition $Y = \sum_{i=1}^n \lambda_i q_i^\top q_i$, where $\lambda_i \geq 0$ are the eigenvalues, and q_i the associated eigenvectors. Thus, for any $X \in S_n^+$, we have
$$\langle Y, X\rangle = \mathrm{tr}(X\sum_{i=1}^n \lambda_i q_i^\top q_i) = \mathrm{tr}(\sum_{i=1}^n \lambda_i q_i^\top X q_i) \geq 0$$
   hence $Y \in (S_n^+)^\oplus$

Exercise 3. *We consider the set of* $n \times n$ *symmetric real matrices* $S_n(\mathbb{R})$.

1. *Show that* $\langle A, B\rangle = \mathrm{tr}(AB)$ *is a scalar product on* S_n.
2. *Show that the set of semi-definite positive matrices* $K = S_n^+(\mathbb{R})$ *is a cone.*

3. *Show that $K = S_n^+(\mathbb{R})$ is self dual (i.e. $K = K^\oplus$ for this scalar product).*

Answers:

1. It is symetric and bilinear. $\mathrm{tr}(AA) = \sum_{ij} a_{ij}^2 = 0$ implie $A = 0$.
2. Let A and B be in S_{++}^n, and $t > 0$ and $t' \geq 0$. Then we have, for all $x \in \mathbb{R}^n$, $x \neq 0$,
$$x^\top (tA + t'B)x = tx^\top Ax + t'x^\top Bx > 0.$$
3. Let $Y \in S_n \setminus S_n^+$. Then there exists $v \in \mathbb{R}^n$, $v^\top Y v < 0$. Moreover, $v^\top Y v = tr(v^\top X v) = tr(v^\top v X) < 0$. Hence we have $X = v^\top v \in S_n^+$ such that $\langle Y, X\rangle < 0$, i.e. $Y \notin (S_n^+)^\oplus$.

   On the other hand, consider $Y \in S_n^+$. We have the following decomposition $Y = \sum_{i=1}^n \lambda_i q_i^\top q_i$, where $\lambda_i \geq 0$ are the eigenvalues, and q_i the associated eigenvectors. Thus, for any $X \in S_n^+$, we have
$$\begin{aligned}\langle Y, X\rangle &= \mathrm{tr}(X \sum_{i=1}^n \lambda_i q_i^\top q_i) \\ &= \mathrm{tr}(\sum_{i=1}^n \lambda_i q_i^\top X q_i) \geq 0\end{aligned}$$
   hence $Y \in (S_n^+)^\star$

Convex functions

Exercise 4 (Moving average)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function.*

1. *Show that, $s \mapsto \int_0^1 f(st)dt$ is convex.*
2. *Show that, $\mathbb{R}_+^* \ni T \mapsto 1/T \int_0^T f(t)dt$ is convex.*

Answers:

1. Obvious from convexity of f and monotonicity of the integral.
2. Change of variable $u = t/T$.

Exercise 5 (Partial infimum)**.** *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ be a convex function and $C \subset \mathbb{R}^m$ a convex set. Show that the function*
$$g : x \mapsto \inf_{y \in C} f(x, y)$$
is convex.

Answers: Consider x_1 and x_2 in $\mathrm{dom}(g)$. For $\varepsilon > 0$, we have y_i such that $f(x_i, y_i) \leq g(x_i) + \varepsilon$. Thus,
$$\begin{aligned}&g(\theta x_1 + (1-\theta)x_2) \\ &\quad = \inf_{y\in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\quad \leq f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\quad \leq \theta f(x_1, y_1) + (1-\theta)f(x_2, y2) \\ &\quad \leq \theta g(x_1) + (1-\theta)g(x_2) + \varepsilon\end{aligned}$$
taking the limit in ε yields the result.

Exercise 6 (log determinant)**.** *Let, for any $X \in S_n$, $f(X) = \ln(\det(X))$ for $X \succ 0$, $-\infty$ otherwise. Consider, for $Z \succ 0$, and $V \in S_n$, the function $g : t \mapsto f(Z + tV)$.*

1. *Show that $g(t) = \sum_{i=1}^n \ln(1 + t\lambda_i) + f(Z)$, where the λ_i are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$.*
2. *Show that g is concave. Conclude that f is concave.*

Answers:

1. We have
$$\begin{aligned}g(t) &= f(Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}) \\ &= \ln\det(Z) + \ln\det(I + tZ^{-1/2}VZ^{-1/2}) \\ &= f(Z) + \sum_{i=1}^n \ln(1 + t\lambda_i)\end{aligned}$$
2. Concavity of g is obvious as sum of concave functions. We have $f(tX + (1-t)Y) = g(t)$, with $Z = X$ and $V = Y - X$. Hence f is concave.

Exercise 7 (Perspective function)**.** *Let $\phi : E \to \mathbb{R} \cup \{+\infty\}$. The perspective of ϕ is defined as $\tilde{\phi} : \mathbb{R}_+^* \times E \to \mathbb{R}$ by*
$$\tilde{\phi}(\eta, y) := \eta\phi(y/\eta).$$
Show that ϕ is convex iff $\tilde{\phi}$ is convex.

Answers:
$$\begin{aligned}(\eta, y, z) \in \mathrm{epi}\,\tilde{\phi} &\Leftrightarrow \eta\phi(y/\eta) \leq z \\ &\Leftrightarrow \phi(y/\eta) \leq z/\eta \\ &\Leftrightarrow (y/\eta, z/\eta) \in \mathrm{epi}\,\phi\end{aligned}$$

Thus epi ϕ is the image of epi $\tilde{\phi}$ through the perspective function which preserve convexity (see exercise 1).

Fenchel transform and subdifferential

Exercise 8 (Norm)**.** *Let* $\|\cdot\|$ *be a norm on* $\mathbb{R}^n$ *and* $\|y\|_\star := \sup_{x:\|x\|\leq 1} y\top x$ *be its dual norm. Let* $f : x \mapsto \|x\|$*. Compute* $f^\star$ *and* $\partial f(0)$*.*

Answers: Recall that $f^\star(y) = \sup_x y^\top x - \|x\|$. We have $y^\top x \leq \|x\|\|y\|_\star$. Thus, if $\|y\|_\star \leq 1$, we have $f^\star(y) \leq \sup_x \|x\|(\|y\|_\star - 1) \leq 0$ attained for $x = 0$.
Otherwise, if $\|y\|_\star > 1$, there exists x such that $y^\top x > 1$, and we have, for all $t > 0$, $f^\star(y) \geq t(y^\top x - \|x\|)$. Consequently $f^\star(y) = \mathbb{I}_{\|\cdot\|_\star \leq 1}$.
By Fenchel-Young, $\partial f(0) = \{y \in \mathbb{R}^n \mid \|y\|_\star \leq 1\}$.

Exercise 9 (Log sum exp)**.** *We consider* $f(x) := \ln(\sum_{i=1}^n e^{x_i})$*.*

1. *Show that* f *is convex. Hint : recall Holder's inequality* $x^\top y \leq \|x\|_p\|y\|_q$ *for* $1/p + 1/q = 1$*.*
2. *Show that* $f^\star(y) = \sum_{i=1}^n y_i \ln(y_i)$ *if* $y \geq 0$ *and* $\sum_i y_i = 1$*,* $+\infty$ *otherwise.*

Answers:

1. Let $x, y \in \mathbb{R}^n$ and set $u_i = e^{x_i}$ and $v_i = e^{y_i}$. For $\theta \in [0, 1]$, we have

$$
\begin{aligned}
f(\theta x + (1-\theta)y) &= \ln(\sum_{i=1}^n e^{\theta x_i + (1-\theta) y_i}) \\
&= \ln(\sum_{i=1}^n u_i^\theta v_i^{1-\theta})
\end{aligned}
$$

   We use $p = 1/\theta$ and $q = 1/(1-\theta)$ in Holder's inequality to get

$$
\begin{aligned}
f(\theta x + (1-\theta)y) &\leq \ln\left(\left(\sum_{i=1}^n u_i\right)^\theta \left(\sum_{i=1}^n u_i\right)^{1-\theta}\right) \\
&= \theta f(x) + (1-\theta) f(y)
\end{aligned}
$$

2.

Exercises: Optimality conditions

June 12, 2023

Exercise 1. *Solve the following optimization problem*

$$\begin{aligned} \underset{x,y\in\mathbb{R}^2}{\text{Min}} \quad & (x-1)^2+(y-2)^2 \\ & x \leq y \\ & x+2y \leq 2 \end{aligned}$$

Answers: The problem is convex and qualified through Slater's condition (e.g. $(-1,0)$). Lagrangian

$$\mathcal{L}(x,y,\mu) = (x-1)^2+(y-2)^2 + \mu_1(x-y)+\mu_2(x+2y-2)$$

KKT conditions

$$\begin{cases} 2(x-1)+\mu_1+\mu_2=0 \\ 2(y-2)-\mu_1+2\mu_2=0 \\ x\leq y, \quad x+2y\leq 2 \\ \mu_1\geq 0, \mu_2\geq 0 \\ \mu_1=0 \quad \text{or} \quad x=y \\ \mu_2=0 \quad \text{or} \quad x+2y=2 \end{cases}$$

If $\mu_1=\mu_2=0$ we get $x=1$, $y=2$ thus $x+2y=5>2$ not admissible.
If $\mu_1=0$ and $\mu_2>0$, we get $x=2-2y$ and $\mu_2=2(1-x)=4y-2$, leading to $2(y-2)+2(4y-2)=0$. Thus, $y=4/5$, $x=2/5$, $\mu_1=0$, $\mu_2=6/5>0$ satisfy KKT conditions, and thus is optimal by convexity.

Exercise 2 (First order optimality condition)**.** *Consider, for f differentiable,*

$$(P) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) \qquad s.t. \quad x\in X$$

Recall that

$$T_X(x_0)=\Big\{\ d\in\mathbb{R}^n \mid \exists d_k\to d, \exists t_k\searrow 0, \ s.t.\ x_0+t_kd_k\in X\Big\}$$

and $K^\oplus=\big\{\lambda \mid \lambda^\top x\geq 0,\ \forall x\in K\big\}$.
Show that

1. *If x_0 is an optimal solution to (P), then $\nabla f(x_0)\in[T_X(x_0)]^+$.*
2. *If f is convex, X is closed convex, and $\nabla f(x_0)\in[T_X(x_0)]^\oplus$, then x_0 is an optimal solution to (P).*

Answers:

1. Assume that $\nabla f(x_0)\notin[T_X(x_0)]^\oplus$. Then we have $d\in T_X(x_0)$ such that $d^\top\nabla f(x_0)<0$. By continuity of scalar product we have, for k large enough, $d_k^\top\nabla f(x_0)<0$. We have $x_0+t_kd_k\in X$, and $f(x_0+t_kd_k)=f(x_0)+t_kd_k^\top\nabla f(x_0)+o(t_kd_k)$. Thus, for k large enough, $f(x_0+t_kd_k)<f(x_0)$.
2. By convexity of X, we have, for $x\in X$, $(x-x_0)\in T_X(x_0)$. Further, by convexity of f, $f(x)\geq f(x_0)+\big\langle\nabla f(x_0)\,, x-x_0\big\rangle\geq f(x_0)$.

Exercise 3. *In the following cases, are the KKT conditions necessary / sufficient ?*

1.

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & 12x_1-5x_2+3x_3 \\ s.t. \quad & x_1+2x_2-x_3=5 \\ & x_1-x_2\geq -2 \\ & 2x_1-4x_2\leq 12 \end{aligned}$$

2.

$$\begin{aligned}\min_{x_1,x_2} \quad & 4x_1^2 - x_1x_2 + x_2^2 - 12x_1 \\ s.t. \quad & x_1 - 2x_2 + x_3 = 5 \\ & x_1^2 + 3x_2^2 \leq 10 \\ & x_1, x_2, x_3 \geq 0\end{aligned}$$

3.

$$\begin{aligned}\min_{x_1,x_2,x_3} \quad & e^{x_1} - x_1x_2 + x_3^3 \\ s.t. \quad & \ln(e^{x_1-4x_2} + e^{x_1+x_3}) \leq 2x_1 + 3 \\ & 2x_1^2 + x_2^2 \leq 2\end{aligned}$$

4.

$$\begin{aligned}\min_{x_1,x_2} \quad & -x_1 \\ s.t. \quad & -x_2 - (x_1-1)^3 \leq 0 \\ & x_1, x_2 \geq 0\end{aligned}$$

5.

$$\begin{aligned}\min_{x_1,x_2} \quad & -x_1 \\ s.t. \quad & x_2 - (x_1-1)^3 \leq 0 \\ & x_1, x_2 \geq 0\end{aligned}$$

Answers:

1. CNS as problem is linear, thus convex and qualified everywhere
2. CNS as problem is convex and qualified by Slater
3. CN as constraints are convex and qualified by Slater but objective is nonconvex
4. CNS, constraints are qualified due to "positive-independence" condition.
5. Neither. Indeed, no sufficient qualification conditions are satisfied and we can even check that the constraints are not qualified at $x_0 = (1,0)$. Indeed, we have ($x_1 \geq 0$ is not active at x_0)

$$T^\ell_{x_0}X = \{x \mid x_2 - 0 \leq 0,\ x_2 \leq 0\} = \mathbb{R} \times \{0\};$$

$$T_{x_0}X = \mathbb{R}^+ \times \{0\}.$$

Exercise 4. *Solve the following problem using first order optimality conditions*

$$\begin{aligned}\min_{x_1,x_2} \quad & -2(x_1-2)^2 - x_2^2 \\ s.t. \quad & x_1^2 + x_2^2 \leq 25 \\ & x_1 \geq 0\end{aligned}$$

Answers: First note that the constraint set is convex, and $(1,1)$ is a Slater's point, ensuring qualification everywhere.
The Lagrangian reads

$$\mathcal{L}(x_1,x_2,\mu_1,\mu_2) = -2(x_1-2)^2 - x_2^2 + \mu_1(x_1^2+x_2^2-25) - \mu_2 x_1$$

The KKT conditions thus read

$$\begin{cases} -4(x_1-2) + 2\mu_1 x_1 - \mu_2 = 0 \\ -2x_2 + 2\mu_1 x_2 = 0 \\ x_1^2 + x_2^2 \leq 25 \\ x_1 \geq 0 \\ \mu_1, \mu_2 \geq 0 \\ \mu_1 = 0 \quad \text{or} \quad x_1^2 + x_2^2 = 25 \\ \mu_2 = 0 \quad \text{or} \quad x_1 = 0 \end{cases}$$

If $\mu_1 = \mu_2 = 0$, we have $x_1 = 2$ and $x_2 = 0$ which satisfies the primal constraints. Thus $\begin{pmatrix}2\\0\end{pmatrix}, \begin{pmatrix}0\\0\end{pmatrix}$ is a primal-dual point satisfying KKT conditions with associated value 0.
If $\mu_1 = 0$ and $\mu_2 > 0$ we have $x_1 = x_2 = 0$ with $\mu_2 = 8 > 0$ which is a primal-dual point with value -8.
If $\mu_2 = 0$ and $\mu_1 > 0$ we have

$$\begin{cases} -4(x_1-2) + 2\mu_1 x_1 = 0 \\ -2x_2 + 2\mu_1 x_2 = 0 \\ x_1 \geq 0 \\ \mu_1 > 0 \\ x_1^2 + x_2^2 = 25 \end{cases}$$

Thus, either $x_2 = 0$ or $\mu_1 = 1$. In the first case we get $x_1 = 5$, $x_2 = 0$, thus $\mu_1 = 6/5 > 0$ and $\mu_2 = 0$ which is a KKT point with value -18. In the second case we get $x_1 = 4$ and $x_2 = \pm 3$, with $\mu_1 = 1$ and $\mu_2 = 0$ which are two KKT points with value -17.
Finally, if $\mu_2 > 0$ and $\mu_1 > 0$, we have $x_1 = 0$ and $x_2 = \pm 5$ with $\mu_1 = 1$ and $\mu_2 = 8$, which are two KKT points with value -33, and thus the global minima.

Exercises: Duality

June 12, 2023

Exercise 1 (Dual formulation)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}^m$. Show that*

1. $\mathbb{I}_{g(x)=0} = \sup_{\lambda\in\mathbb{R}^m} \lambda^\top g(x)$
2. $\mathbb{I}_{g(x)\leq 0} = \sup_{\lambda\in\mathbb{R}^m_+} \lambda^\top g(x)$
3. $\mathbb{I}_{g(x)\in C} = \sup_{\lambda\in -C^\oplus} \lambda^\top g(x)$ *where C is a closed convex cone, and $C^\oplus := \{\lambda \in \mathbb{R}^m \mid \lambda^\top c \geq 0, \forall c \in C\}$.*

Answers:

1. If $g(x) \neq 0$ there is $i \in [m]$ such that $g_i(x) \neq 0$, and we choose λ_i accordingly.
2. Same reasoning.
3. If $g(x) \in C$, $\lambda^\top g(x) \leq 0$, and $0 \in -C^\oplus$. If $g(x) \notin C$, by separation of the convex compact $\{g(x)\}$ from the closed convex set C there exists $\lambda \in \mathbb{R}^n$ such that $\lambda^\top g(x) > b > \lambda^\top c$ for all $c \in C$. As C is a cone, $tc \in C$ for all $t > 0$, and thus $\lambda \in -C^\oplus$. Further $b \geq 0$, thus $t\lambda^\top g(x) \to +\infty$ when $t \to \infty$.

Exercise 2 (Linear Programming)**.** *Consider the following linear problem (LP)*

$$(P) \quad \min_{x\geq 0} \quad c^\top x \qquad s.t. \quad Ax = b$$

1. *Show that the dual of (P) is an LP.*
2. *Show that the dual of the dual of (P) is equivalent to (P).*

Answers:

1. The dual of (P) is

$$(D) \quad \max_{\lambda} \quad -b^\top\lambda \qquad \text{s.t.} \quad A^\top\lambda + c \leq 0$$

2. Direct by computing the dual of (D).

Exercise 3 (Quadratically Constrained Quadratic Programming)**.** *Consider the problem*

$$(QCQP) \quad \min_{x\in\mathbb{R}^n} \quad \frac{1}{2}x^\top P_0 x + q_0^\top x + r_0$$
$$\frac{1}{2}x^\top P_i x + q_i^\top x + r_i \leq 0 \quad \forall i \in [m]$$

where $P_0 \in S^n_{++}$ and $P_i \in S^n_+$.

1. *Show by duality that, for $\mu \in \mathbb{R}^m_+$, there exists P_μ, q_μ and r_μ,such that $g(\mu) = -\frac{1}{2}q_\mu P_\mu^{-1} q_\mu + r_\mu \leq val(P)$.*
2. *Give an easy condition under which $val(P) = \sup_{\mu\geq 0} g(\mu)$.*

Answers:

1. Simply write the dual function we get

$$P_\mu = P_0 + \sum_i \mu_i P_i, \quad q_\mu = q_0 + \sum_i \mu_i q_i$$

$$r_\mu = r_0 + \sum_i \mu_i r_i$$

2. The problem is convex, Slater's condition ensure constraint qualification, thus a condition would be the existence of a strictly satisfying all constraints.

Exercise 4 (Conic Programming)**.** *Let $K \subset \mathbb{R}^n$ be a closed convex pointed cone, and denote $x \preceq_K y$ iff $y \in x + K$. Consider the following program, with $A \in M_{m,n}$ and $b \in \mathbb{R}^m$.*

$$(P) \quad \min_{x\in\mathbb{R}^n} \quad c^\top x \qquad s.t. \quad Ax = b, \quad x \preceq_K 0$$

1. *Show that (P) is a convex optimization problem.*
2. *Denote $\mathcal{L}(x, \lambda, \mu) = c^\top x + \lambda^\top(Ax - b) + \mu^\top x$. Show that $\text{val}(P) = \text{Min}_{x\in\mathbb{R}^n} \sup_{\lambda\in\mathbb{R}^m, \mu\in K^\oplus} \mathcal{L}(x, \lambda, \mu)$.*
3. *Give a dual problem to (P).*

Answers:

1. $x \preceq_K 0$ iff $x \in -K$, and $-K$ is convex.
2. If $x \in -K$, for any $\mu \in K^\oplus$, $\mu^\top x \leq 0$, thus $\sup_{\mu\in K^\oplus} \mu^\top x = 0$. If $-x \notin K = K^{\oplus\oplus}$, there exists $\lambda \in K^\oplus$, such that $-x^\top\lambda < 0$, hence $\sup_{\mu\in K^\oplus} \mu^\top x = +\infty$. (or see ex 1).
3. By min-max duality we consider

$$\underset{\lambda\in\mathbb{R}^m, \mu\in K^\oplus}{\text{Max}} -b^\top\lambda + \inf_{x\in\mathbb{R}^n} (A^\top\lambda + c + \mu)^\top x +$$

wich yields

$$(D) \quad \text{Max} \quad -b^\top\lambda$$
$$A^\top\lambda + c + \mu = 0 \quad \mu \in K^\oplus$$

Exercise 5 (Duality gap)**.** *Consider the following problem*

$$\underset{x\in\mathbb{R}, y\in\mathbb{R}_*^+}{\text{Min}} \quad e^{-x}$$
$$s.t. \quad x^2/y \leq 0$$

1. *Find the optimal solution of this problem.*
2. *Write and solve the (Lagrangian) dual problem. Is there a duality gap ?*

Answers:

$$\mathcal{L}(x, y; \mu) = e^{-x} + \mu x^2/y$$
$$g(\mu) = \inf_{x\in\mathbb{R}, y>0} e^{-x} + \mu x^2/y = 0$$

as the term inside the inf is positive, and choosing $x = t$, $y = t^3$ goes to 0 for all μ.

Exercise 6 (Two-way partitionning)**.** *Let $W \in S_n$ be a symmetric matrix, consider the following problem.*

$$(P) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad x^\top W x$$
$$s.t. \quad x_i^2 = 1 \qquad \forall i \in [n]$$

1. *Consider a set of n element that you want to partition in 2 subsets, with a cost $c_{i,j}$ if i and j are in the same set, and a cost $-c_{i,j}$ if they are in a different set. Justify that it can be solved by solving (P).*
2. *Is (P) a convex problem ?*
3. *Show that, for any $\lambda \in \mathbb{R}^n$ such that $W + \text{diag}(\lambda) \succeq 0$, we have $\text{val}(P) \geq -\sum \lambda_i$. Deduce a lower bound on $\text{val}(P)$.*

Answers:

1. The constraint ensures that $x_i \in \{-1, 1\}$, each value representing one subset. We set $W_{i,j} = c_{i,j}$.
2. No, because the set of admissible points is $\{-1, 1\}^n$.
3. The Lagrangian of (P) is

$$\mathcal{L}(x, \lambda) = x^\top W x + \sum_{i=1}^n \lambda_i(x_i^2 - 1)$$
$$= x^\top(W + diag(\lambda))x - \sum_{i=1}^n \lambda_i$$

And we have,

$$g(\lambda) = \inf_x x^\top(W + diag(\lambda))x - \sum_{i=1}^n \lambda_i$$
$$= -\sum_{i=1}^n \lambda_i - \mathbb{I}_{W+diag(\lambda)\geq 0} \leq val(P)$$

Thus, if λ_{min} is the small eigenvalue of W we have $W + diag(\lambda) \geq 0$, and $val(P) \geq n\lambda_{min}$.

Exercise 7 (Linear SVM : duality)**.** *Consider the following problem (see : `https://www.youtube.com/watch?v=IOetFPgsMUc` for background)*

$$\min_{w\in\mathbb{R}^d, b\in\mathbb{R}} \frac{1}{2}\|w\|^2$$
$$s.t. \quad y_i(w^\top x_i + b) \geq 1 \qquad \forall i \in [n]$$
$$\eta_i \geq 0 \qquad \forall i \in [n]$$

1. *In which case can we guarantee strong duality ?*

2. *Write the dual of this optimization problem and express the optimal primal solution $(w^\sharp, b^\sharp)$ in terms of the optimal dual solution.*

Exercise 8. *We consider the following problem.*

$$\begin{aligned} \underset{x_1,x_2}{\text{Min}} \quad & x_1^2 + x_2^2 & (1)\\ s.t. \quad & (x_1-1)^2 + (x_2-1)^2 \leq 1 & (2)\\ & (x_1-1)^2 + (x_2+1)^2 \leq 1 & (3) \end{aligned}$$

1. *Classify this problem (After 5th course)*
2. *Find the optimal solution and value of this problem.*
3. *Write and solve the KKT equation for this problem.*
4. *Derive and solve the Lagrangian dual of this problem.*
5. *Do we have strong duality ? If yes, could we have known it from the start ? If not, can you comment on why ?*

Answers:

1. This is a convex QCQP
2. The only admissible point, and hence the optimal solution is $(1, 0)$, with value 1.
3. The Lagrangian is

$$\begin{aligned} \mathcal{L}(x,\lambda) =& x_1^2 + x_2^2 + \lambda_1((x_1-1)^2 + (x_2-1)^2 - 1)\\ &+ \lambda_2((x_1-1)^2 + (x_2+1)^2 - 1) \end{aligned}$$

KKT condition are

- Gradient of Lagrangian is null :

$$\begin{aligned} 2x_1 + 2\lambda_1(x_1-1) + 2\lambda_2(x_1-1) = 0\\ 2x_2 + 2\lambda_1(x_2-1) + 2\lambda_2(x_2+1) = 0 \end{aligned}$$

- x is primal feasible : $(x_1-1)^2 + (x_2-1)^2 \leq 1$ and $(x_1-1)^2 + (x_2+1)^2 \leq 1$
- λ is dual feasible $\lambda_1 \geq 0$, $\lambda_2 \geq 0$.
- Complementary slackness:

$$\begin{aligned} \lambda_1 = 0 \quad & \text{or} \quad (x_1-1)^2 + (x_2-1)^2 = 1\\ \lambda_2 = 0 \quad & \text{or} \quad (x_1-1)^2 + (x_2+1)^2 = 1 \end{aligned}$$

x feasible is $x = (1, 0)$, which imply $2 = 0$ which is impossible. Thus there is no pair (x, λ) satisfying the KKT equations. The KKT equations fails to give the optimal solution because the constraints are not qualified.

4. The Lagrange dual function is

$$\begin{aligned} g(\lambda_1,\lambda_2) &= \inf_{x_1,x_2} \mathcal{L}(x,\lambda)\\ &= \inf_{x_1,x_2} (1+\lambda_1+\lambda_2)(x_1^2 + x_2)^2\\ &\quad - 2(\lambda_1+\lambda_2)x_1 - 2(\lambda_1-\lambda_2)x_2 + \lambda_1 + \lambda_2\\ &= \lambda_1 + \lambda_2 - \frac{(\lambda_1+\lambda_2)^2 + (\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}\\ &\quad \text{if } 1+\lambda_1+\lambda_2 > 0 \end{aligned}$$

The dual problem reads

$$\begin{aligned} \underset{\lambda}{\text{Max}} \quad & \frac{\lambda_1+\lambda_2 - (\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2}\\ \text{s.t.} \quad & \lambda_1 \geq 0, \lambda_2 \geq 0 \end{aligned}$$

By symmetry the optimum is attained at $\lambda_1 = \lambda_2$, thus the dual reads

$$\underset{\lambda_1 \geq 0}{\text{Max}} \quad \frac{2\lambda_1}{2\lambda_1 + 1}$$

Which has value 1 and no solution.

5. The dual problem have the same value as the primal problem, thus we have strong duality.

However there does not exist a dual multiplier, which is why there is no solution to the KKT equations.

We could not guarantee the existence of a primal-dual optimal solution through KKT as the constraints were not qualified.

Exercises: optimization problem classes

Exercise 1 (Hyperbolic constraints as SOCP)**.**

1. *Show that, for all* $x \in \mathbb{R}^n$, $y \in \mathbb{R}$, $z \in \mathbb{R}$,

$$x^\top x \leq yz, \quad y \geq 0, \quad z \geq 0$$

iff

$$\left\| \begin{pmatrix} 2x \\ y - z \end{pmatrix} \right\| \leq y + z \quad y \geq 0, \quad z \geq 0$$

2. *Represent the following problem as an SOCP*

$$\begin{aligned} (P) \text{ Max} \quad & \left(\sum_{i=1}^n 1/(a_i^\top x - b) \right)^{-1} \\ s.t. \quad & Ax > b \end{aligned}$$

Answers:

1. Assume $y \geq 0, z \geq 0$, then $y + z \geq 0 \iff xy \geq 0$. We now assume $y + z \geq 0$, then we have

$$\begin{aligned} & \left\| \begin{pmatrix} 2x \\ y - z \end{pmatrix} \right\| \leq y + z \\ \iff & \left\| \begin{pmatrix} 2x \\ y - z \end{pmatrix} \right\|^2 \leq (y + z)^2 \\ \iff & 4x^\top x + (y - z)^2 \leq y^2 + 2yz + z^2 \\ \iff & 4x^\top x \leq 4yz \\ \iff & x^\top x \leq yz \end{aligned}$$

2. Since $t \mapsto 1/t$ is decreasing (P) is equivalent to

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n 1/(a_i^\top x - b) \\ s.t. \quad & Ax > b \end{aligned}$$

By adding the lift variables z_i, (P) is equivalent to the problem

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n z_i \\ s.t. \quad & Ax > b \\ & 1/(a_i^\top x - b) \geq z_i \geq 0, \forall i \in [n] \end{aligned}$$

which is equivalent to

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n z_i \\ s.t. \quad & Ax \geq b \\ & 1 \leq z(a_i^\top x - b), \forall i \in [n] \\ & z_i \geq 0, \forall i \in [n] \end{aligned}$$

By question 1. it is equivalent to the following SOCP:

$$\begin{aligned} \text{Min} \quad & \sum_{i=1}^n z_i \\ s.t. \quad & -Ax \leq -b \\ & \left\| \begin{pmatrix} 2 \\ a_i^\top x - b - z \end{pmatrix} \right\| \leq a_i^\top x - b + z, \forall i \in [n] \\ & z_i \geq 0, \forall i \in [n] \end{aligned}$$

Exercise 2. *We consider a physical function* Φ *that is approximated as the superposition of multiple simple phenomenon (e.g. waves). Each simple phenomenon* $p \in [P]$ *is represented by a function* $\Phi_p : \mathbb{R}^d \to \mathbb{R}$.
We have data points $(x^k, y^k)_{k \in [n]}$, *and want to find the* Φ *that match* at best *the data while being a linear combination of* Φ_p.
Propose a least-square regression that answer this question.

Answers: We define the matrix $M \in \mathbb{R}^{n\times P}$ with coefficients $M_{k,p} = \Phi_p(x_k)$. We propose the following last square regression problem:

$$\underset{\alpha\in\mathbb{R}^P}{\text{Min}} \qquad \|M\alpha - y\|^2 + \lambda\|\alpha\|_1$$

Exercise 3. *Consider a chocolate manufacturing company that produces only two types of chocolate – A and B. Both the chocolates require Milk and Choco only. To manufacture each unit of A and B, the following quantities are required:*

- *Each unit of A requires 1 unit of Milk and 3 units of Choco*
- *Each unit of B requires 1 unit of Milk and 2 units of Choco*

The company kitchen has a total of 5 units of Milk and 12 units of Choco. On each sale, the company makes a profit of

- *6 per unit A sold*
- *5 per unit B sold.*

Model this as an LP.

Exercise 4. *A classical extension of the least-square problem, which has strong theoretical and practical intereset is the LASSO problem*

$$\underset{x\in\mathbb{R}^p}{\text{Min}} \qquad \|Ax - b\|^2 + \lambda\|x\|_1$$

Show that this problem can be cast as a QP problem.

Answers: The LASSO problem is equivalent to the following QP problem

$$\begin{aligned}\underset{x\in\mathbb{R}^n, z\in\mathbb{R}^n}{\text{Min}} \qquad & x^\top A^\top Ax - 2b^\top Ax + \lambda\sum_{i=1}^n z_i \\ \text{s.t.} \qquad & x_i \le z_i \\ & -x_i \le z_i\end{aligned}$$

Exercise 5. *Consider the following optimization problem.*

$$\begin{aligned}\underset{x\in\mathbb{R}^n}{\text{Min}} \qquad & c^\top x \\ s.t. \qquad & Ax = b \\ & x_i \in \{0,1\} \qquad \forall i \in I\end{aligned}$$

Write this problem as a QCQP. Is it convex ?

Answers: The constraint $x_i \in \{0,1\}$ is equivalent to $x_i(1-x_i) = 0$. We define $q_i := (0,\cdots,0,1,0,\cdots,0)$ as the vector with all coordinates equal to 0 except the ith which equals 1. We set $Q_i = 2diag(q_i) = 2q_iq_i^\top$. Then, this problem is equivalent to the following QCQP problem

$$\begin{aligned}\underset{x\in\mathbb{R}^n}{\text{Min}} \qquad & c^\top x \\ \text{s.t.} \qquad & Ax = b \\ & \frac{1}{2}x^\top Q_i x + q_i^\top x \le 0 \qquad \forall i \in I \\ & -\frac{1}{2}x^\top Q_i x - q_i^\top x \le 0 \qquad \forall i \in I\end{aligned}$$

It is not convex in the general case (i.e. if the admissible set is neither empty or reduced to a singleton) since the set $\{0,1\}^n$ is not convex. Remark that $-Q_i$ is not positive.

Exercise 6. *Consider a facility that plan to deliver product to clients by drone (thus in direct line). Assume that you have N clients, each with position (in $\mathbb{R}^2$) x_n. The drone make each time a direct travel from the facility location to the client. Assume that the drone have a maximum range of R, and that you want to minimize the average travel distance while being able to serve all of your clients.*
Model the problem of choosing the facility location as an SOCP.

Answers: We want to minimize the average travel distance $\frac{1}{N}\sum_{n=1}^N \|x_n - y\|$ from a center y to the clients (x_n) while being able to serve all of your clients. We modelize this by the problem

$$\underset{y\in\mathbb{R}^2}{\text{Min}} \quad \frac{1}{N}\sum_{n=1}^N \|x_n - y\| \tag{1}$$

$$\text{s.t.} \quad \|x_n - y\| \le R, \forall n \in [N] \tag{2}$$

By adding lift variables z_n, this equivalent to the following SOCP:

$$\underset{y\in\mathbb{R}^2, z\in\mathbb{R}^N}{\text{Min}} \quad \frac{1}{n}\sum_{n=1}^N z_n \tag{3}$$

$$\text{s.t.} \quad \|x_n - y\| \le R, \forall n \in [N] \tag{4}$$

$$\|x_n - y\| \le z_n, \forall n \in [N] \tag{5}$$

Exercise 7. *Consider the following robust linear program*

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & c^\top x \\ s.t. \quad & (a_i + R_i\delta_i)^\top x \le b_i \quad \forall \|\delta_i\|_2 \le 1, \quad \forall i \in [m] \end{aligned}$$

where R_i are positive real numbers. Write this problem as an SOCP.

Answers: The constraint $(a_i + R_i\delta_i)^\top x \le b_i, \quad \forall \|\delta_i\|_2 \le 1$ is equivalent to

$$\max_{\delta_i \,|\, \|\delta_i\|_2 \le 1} (a_i + R_i\delta_i)^\top x \le b_i$$

which is equivalent to

$$R_i \max_{\delta_i \,|\, \|\delta_i\|_2 \le 1} \delta_i^\top x \le b_i - a_i^\top x.$$

However, $\max_{\delta_i \,|\, \|\delta_i\|_2 \le 1} \delta_i^\top x = \|x\|$. Indeed, this result is trivial for $x = 0$, for $x \neq 0$, we have that $\|x\|$ is an upper bound by Cauchy-Swhartz inequality which is attained for $\delta_i = \frac{x}{\|x\|}$. Thus, our problem is equivalent to the following SOCP:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & c^\top x \\ \text{s.t.} \quad & R_i\|x\| \le -a_i^\top x + b_i, \qquad \forall i \in [m] \end{aligned}$$

Exercise 8. *Let $F(\theta)$ be a symmetric matrix parametrized by $\theta \in \mathbb{R}^d$ whose coefficients are linear in θ. Model the problem of finding the parameter $\theta \in \Theta$, where Θ is a polyhedron, minimizing $\kappa(\theta)$ as an SDP.*
What happen if the coefficient of $F(\theta)$ are affine in θ ? Suggest a solution method ? (hard)

Exercise 9. *Consider a finite set $X = \{x_i\}_{i\in[n]}$, and $\mathcal{P}^+$ the set of probabilities on X. For $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$, with* supp$(\mathbb{Q}) = X$*, we define the* Kullback-Leibler *divergence as*

$$d_{KL}(\mathbb{P}|\mathbb{Q}) = \sum_{i=1}^n p_i \ln(p_i/q_i)$$

where $p_i = \mathbb{P}(X = x_i)$ and $q_i = \mathbb{Q}(X = x_i)$.
Let X be 100 equidistant points spanning in $[-1, 1]$. Let $\mathbb{Q}$ be uniform on X.
We are looking for the probability $\mathbb{P}$ on X such that

- $\mathbb{E}_{\mathbb{P}}[\boldsymbol{X}] \in [-0.1, 0.1]$
- $\mathbb{E}_{\mathbb{P}}[\boldsymbol{X}^2] \in [0.5, 0.6]$
- $\mathbb{E}_{\mathbb{P}}[3\boldsymbol{X}^2 - 2\boldsymbol{X}] \in [-0.3, -0.2]$
- $\mathbb{P}(\boldsymbol{X} < 0) \in [0.3, 0.4]$

that minimize the Kullback-Leibler divergence from $\mathbb{Q}$.
Model this problem as an optimization problem. In which class does it belongs ?

Exercise 10. *Consider that you sell a given product over T days. The demand for each day is d_t. Having a quantity x_t of items in stock have a cost (per day) of cx_t. You can order, each day, a quantity q_t, and have to satisfy the demand.*
For each of the following variation : model the problem, explicit the class to which it belongs, and give the optimal solution if easily found.

1. *Without any further constraint / specifications.*
2. *There is an "ordering cost": each time you order, you have to pay a fix cost κ.*
3. *Instead of an "ordering cost" there is a maximum number of days at which you can order a replenishment.*

Exercises: Gradient algorithms

Exercise 1 (A quadratic example in $\mathbb{R}^2$)**.** *Consider, for $\gamma > 0$, $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$. We apply the gradient descent method with optimal step, starting at $x^{(0)} = (\gamma, 1)$.*

1. *Show that f is m-convex with M-Lipschitz gradient. Find the tightest m and M constants.*

2. *Show that*

$$x^{(k)} = \left(\gamma\Big(\frac{\gamma-1}{\gamma+1}\Big)^k, \Big(-\frac{\gamma-1}{\gamma+1}\Big)^k\right)$$

and

$$f(x^{(k)}) = \frac{\gamma(\gamma+1)}{2}\Big(\frac{\gamma-1}{\gamma+1}\Big)^{2k} f(x^{(0)})$$

3. *Show that, on this example, the convergence is exactly linear, that is $f(x^{(k)}) - v^\sharp$ is a geometric series. Give its reason. Compare with the theoretical bound.*

4. *When is this algorithm fast and slow ?*

Exercise 2 (Strongly convex - optimal step)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a m-convex $\mathcal{C}^2$ function. Define, for given $x^{(0)}$,*

$$\begin{aligned}
&\tilde{f}_k : t \mapsto f(x^{(k)} - t\nabla f(x^{(k)})) \\
&t^{(k)} = \arg\min_{t\in\mathbb{R}} \tilde{f}_k(t) \\
&x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f(x^{(k)})
\end{aligned}$$

1. *Show that there exists $M \geq m$ such that $mI \preceq \nabla^2 f(x^{(k)}) \preceq MI$*

2. *Show that, for any interesting t (to be defined) we have*

$$\tilde{f}_k(t) \leq f(x^{(k)}) - t\|\nabla f(x^{(k)})\|_2^2 + \frac{Mt^2}{2}\|\nabla f(x^{(k)})\|_2^2$$

3. *Show that,*

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{1}{2M}\|\nabla f(x^{(k)})\|_2^2$$

4. *Show that $f(x^\sharp) \geq f(x) - \frac{1}{2m}\|\nabla f(x)\|_2^2$*

5. *Show that*

$$f(x^{(k+1)}) - f(x^\sharp) \leq (1 - \frac{m}{M})[f(x^{(k)}) - f(x^\sharp)]$$

6. *Show that the algorithm converges, and give its convergence speed.*

Answers:

1. Let $S = \mathrm{lev}_{f(x^{(0)})} f$. By strong convexity it is a bounded set. f being C^2, its Hessian is continuous and thus bounded on S, where all $x^{(k)}$ lives.

2. For any t such that $y := x^{(k)} - t\nabla f(x^{(k)}) \in S$, there exists $z \in [x^{(k)}, y]$, such that

$$\begin{aligned}
f(y) =& f(x^{(k)}) - \nabla f(x^{(k)})^\top (y - x^{(k)}) \\
&+ \frac{1}{2}(y - x^{(k)})^\top \nabla^2 f(z)(y - x^{(k)})
\end{aligned}$$

replacing y by its value, and using the upper bound on $\nabla^2 f(z)$ yields the result.

3. Use $t = 1/M$ in the upperbound of the previous question. Note that this choice of t minimizes said upper bound.

4. We have, by strong convexity,

$$\begin{aligned}
f(y) \geq& f(x) + \nabla f(x)^\top (y - x) + \frac{m}{2}\|y - x\|^2 \\
\geq& f(x) + \nabla f(x)^\top (\tilde{y} - x) + \frac{m}{2}\|\tilde{y} - x\|^2 \\
=& f(x) - \frac{1}{2m}\|\nabla f(x)\|^2
\end{aligned}$$

where $\tilde{y}$ minizes the right hand side. We then apply to $y = x^\sharp$.

5. By the previous question we have $\|\nabla f(x^{(k)})\|^2 \geq 2m(f(x^{(k)}) - v^\sharp)$. Question 3 then yields

$$f(x^{(k+1)}) \leq f(x^{(k)}) - m/M(f(x^{(k)}) - v^\sharp)$$

substracting $v^\sharp$ on each sides yields the result.

6. Recursively we get $f(x^{(k)}) - v^\sharp \leq c^k(f(x^{(0)}) - v^\sharp)$, with $c = 1 - m/M$. In particular, for any $\varepsilon > 0$, we have $f(x^{(k)}) - v^\sharp \leq \varepsilon$ after at most $\frac{\ln(\varepsilon) - \ln(f(x^{(0)} - v^\sharp))}{\ln(c)}$ iterations.

Exercise 3 (Strictly convex case)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2, convex function with M-Lipschitz gradient such that $f(x^\sharp) = \inf f$. We define, for given $x^{(0)}$.*

$$x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$$

with $t \leq 1/M$.

1. *Show that, for all x and y $(y-x)^\top \nabla^2 f(z)(y-x) \leq M\|y-x\|_2^2$*
2. *Show that*

$$f(y) \leq f(x) + \nabla f(x)^\top (y-x) + \frac{M}{2}\|y-x\|^2$$

3. *Show that $f(x^{(k+1)}) \leq f(x^{(k)}) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2$*
4. *Show that*

$$f(x^{(k+1)}) \leq v^\sharp + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2$$

5. *Deduce that*

$$f(x^{(k+1)}) \leq v^\sharp + \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2)$$

6. *Show that*

$$\sum_{i=1}^{k} f(x^{(i)}) - v^\sharp \leq \frac{1}{2t}\|x^{(0)} - x^\sharp\|^2$$

7. *Conclude that*

$$f(x^{(k)}) - v^\sharp \leq \frac{1}{2kt}\|x^{(0)} - x^\sharp\|^2.$$

Answers:

1. We have $\nabla^2 f(x) \preceq MI$, thus $(y-x)^\top (MI - \nabla^2 f(x))(y-x) \geq 0$.
2. Using Taylor remainder theorem we have the existence of $z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(z)(y-x)$$

3. We obtain

$$f(x^{(k+1)}) \leq f(x^{(k)}) - (1 - \frac{Mt}{2})\|\nabla f(x^{(k)})\|_2^2$$

and with the condition on t we have $1 - Mt/2 \geq 1/2$.

4. We use the convexity inequality to get $f(x^{(k)}) \leq f(x^\sharp) + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp)$ and the result of the previous question.
5. We have

$$\begin{aligned}
&\frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2) \\
&= \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k)} - x^\sharp - t\nabla f(x^{(k)})\|^2) \\
&= \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2
\end{aligned}$$

Thus,

$$\begin{aligned}
f(x^{(k+1)}) &\leq v^\sharp + \nabla f(x^{(k)})^\top (x^{(k)} - x^\sharp) - \frac{t}{2}\|\nabla f(x^{(k)})\|^2 \\
&= v^\sharp + \frac{1}{2t}(\|x^{(k)} - x^\sharp\|^2 - \|x^{(k+1)} - x^\sharp\|^2)
\end{aligned}$$

6. Sum the previous inequality
7. As $f(x^{(i)}) - v^\sharp$ is non-increasing we have that the last term of the sum is lower than the mean of the sum.

Exercise 4 (Kelley's convergence)**.** *We are going to prove that, if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, and X a non-empty polytope (bounded polyhedron) then Kelley's cutting plane algorithm is converging. Consider $x_1 \in X$. We consider a sequence of points $(x^{(k)})_{k\in\mathbb{N}}$ such that $x^{(k+1)}$ is an optimal solution to*

$$(\mathcal{P}^{(k)}) \qquad \underline{v}^{(k+1)} = \underset{x\in X}{\operatorname{Min}} \quad z \qquad s.t. \quad f(x^{(\kappa)}) + \langle g^{(\kappa)}, x - x^{(\kappa)} \rangle \leq z \quad \forall \kappa \in [k]$$

where $g^{(k)} \in \partial f(x^{(k)})$.
Denote $v = \min_{x\in X} f(x)$.

1. *Show that v exists and is finite, and that there exists a sequence $x^{(k)}$.*

2. *Show that there exists L such that, for all k_1 and k_2, we have $\|f(x^{(k_1)}) - f(x^{(k_2)})\| \leq L\|x^{(k_1)} - x^{(k_2)}\|$, and $\|g^{(k)}\| \leq L$.*

3. *Let $K_\varepsilon = \{k \in \mathbb{N} \mid f(x^{(k)}) > v + \varepsilon\}$ be the set of index such that $x^{(k)}$ is not an ε-optimal solution. Show that $f(x_k) \to v$ if and only if K_ε is finite for all $\varepsilon > 0$*

4. *Consider $k_1, k_2 \in K_\varepsilon$, such that $k_2 > k_1$. Show that*

$$f(x^{(k_1)}) + \langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \rangle \leq \underline{v}^{(k_2)} \leq v$$

5. *Show that $\varepsilon + f(x^{(k_1)}) + \langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \rangle < f(x^{(k_2)})$*

6. *Show that $\varepsilon < 2L\|x^{(k_2)} - x^{(k_1)}\|$.*

7. *Prove that $f(x^{(k)}) \to v$.*

8. *(Optional - hard) Find a complexity bound for the method (that is a number of iteration N_ε after which you are sure to have obtained a ε-optimal solution).*

Answers:

1. f is finite convex and thus continuous on X which is compact, yielding the existence and finiteness of v.

   f is subdifferentiable, thus we have the existence of $g^{(k)}$, and an optimal solution to $\mathcal{P}^{(k)}$ exists as the solution of a bounded linear programm.

2. We have seen that on any compact Kincluded in the domain of a convex function f, f is L-Lipschitz. Here $\text{dom}(f) = \mathbb{R}^n$, so on the compact $K = X + B(0, \varepsilon)$ f is L-Lipschitz, and on X any subgradient g is of norm lower than L.

3. $f(x_k) \to v$ iff $\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N}, \quad k \geq K \implies f(x_k) \leq v + \varepsilon$. Hence $K_\varepsilon \subset [N_\varepsilon]$.

4. By subgradient inequality $f(y) \geq f(x^{(k)}) + \langle g^{(k)}, y - x^{(k)} \rangle$. Thus, for all k, $v \geq v^{(k)}$. Further, note that $v^{(k)} = f(x^{(k)})$, hence using $k = k_1$, and $y = x^{(k_2)}$ we get

$$f(x^{(k_1)}) + \langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \rangle \leq \underline{v}^{(k_2)} \leq v.$$

5. As $k_2 \in K_\varepsilon$, we have $f(x^{(k_2)}) = v^{(k_2)} > v + \varepsilon \geq f(x^{(k_1)}) + \langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \rangle + \varepsilon$ by the previous question.

6. We have

$$\varepsilon < |f(x^{(k_2)}) - f(x^{(k_1)})| + |\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \rangle| \leq 2L\|x^{(k_2)} - x^{(k_1)}\|$$

   by Cauchy-Schwartz and question 2.

7. If $f(x^{(k)}) \not\to v$, then there exists $\varepsilon > 0$ such that $(x^{(k)})_{k \in K_\varepsilon})$ is not finite. As X is compact we can exctract a converging subsequence, that is $x^{(\sigma(k)}$ such that $x^{(\sigma(k))} \to x^\star$ and $\sigma(k) \in K_\varepsilon$, which is in contradiction with the result of 6.

Exercises: Constrained Optimization

Exercise 1 (Penalization). *We consider the following problem*

$$(P) \quad \min_{x\in\mathbb{R}^n} \quad f(x) \qquad s.t. \quad Ax = b, \quad x \leq 0$$

with value v and the following penalized versions

$$(P_t^{in}) \quad \min_{x\in\mathbb{R}^n} f(x) - t\sum_{i=1}^n \ln(-x_i) \qquad s.t. Ax = b, \quad x < 0$$

and

$$(P_t^{out}) \quad \min_{x\in\mathbb{R}^n} f(x) + t\sum_{i=1}^n (x_i)^+ \qquad s.t. \quad Ax = b$$

with associated value v_t^{in} and v_t^{out}, and an optimal solution x_t^{in} and x_t^{out}.

1. *Intuitively, assuming that f is "well behaved", for t going to which value does (P_t^{in}) tends to the original problem (P) ? In which sense ?*
2. *What can you say about x_t^{in} ?*
3. *Can you compare v_t^{in} and v ?*
4. *Same questions for (P_t^{out}).*

Answers: For t going to 0 we have that (P_t^{in}) tends toward (P) : in the sense that $v_t^{(in)} \to v$ and x_t goes toward an optimal solution . For t small enough we have $v_t^{in} \geq v$. In any case x_t^{in} is admissible.
For t going to $+\infty$, we have that (P_t^{out}) tends toward (P) in the sense that $v_t^{(out)} \to v$ and x_t^{out} goes toward an optimal solution. For t large enough, x_t^{out} is optimal for (P). We always have $v_t^{(out)} \leq v$.

Exercise 2 (Decomposition by prices). *We consider the following energy problem:*

- *you are an energy producer with N production units*
- *you have to satisfy a given demand planning for the next 24h (i.e. the total output at time t should be equal to d_t)*
- *the time step is the hour, and each unit have a production cost for each planning given as a convex quadratic function of the planning*
- *For each unit i, the production planning $u^i = (u_t^i)_{t\in[24]}$ has to satisfy polyhedral constraints $u^i \in U^i$.*

1. *Model this problem as an optimization problem. In which class does it belongs ? How many variables ?*
2. *Apply Uzawa's algorithm to this problem. Why could this be an interesting idea ?*
3. *Give an economic interpretation to this method.*
4. *What would happen if each unit had production constraints ?*

Exercise 3 (Kelley's convergence). *We are going to prove that, if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, and X a non-empty polytope (bounded polyhedron) then Kelley's cutting plane algorithm is converging.*
Consider $x_1 \in X$. We consider a sequence of points $(x^{(k)})_{k\in\mathbb{N}}$ such that $x^{(k+1)}$ is an optimal solution to

$$(\mathcal{P}^{(k)}) \quad \underline{v}^{(k+1)} = \underset{x\in X}{\mathrm{Min}}\; z \qquad s.t.\; f(x^{(\kappa)}) + \left\langle g^{(\kappa)}, x - x^{(\kappa)} \right\rangle \leq z \quad \forall \kappa \in [k]$$

where $g^{(k)} \in \partial f(x^{(k)})$.
Denote $v = \min_{x\in X} f(x)$.

1. *Show that v exists and is finite, and that there exists a sequence $x^{(k)}$.*

2. *Show that there exists L such that, for all k_1 and k_2, we have $\|f(x^{(k_1)}) - f(x^{(k_2)})\| \leq L\|x^{(k_1)} - x^{(k_2)}\|$, and $\|g^{(k)}\| \leq L$.*

3. *Let $K_\varepsilon = \{k \in \mathbb{N} \mid f(x^{(k)}) > v + \varepsilon\}$ be the set of index such that $x^{(k)}$ is not an ε-optimal solution. Show that $f(x_k) \to v$ if and only if K_ε is finite for all $\varepsilon > 0$*

4. *Consider $k_1, k_2 \in K_\varepsilon$, such that $k_2 > k_1$. Show that*
$$f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle \leq \underline{v}^{(k_2)} \leq v$$

5. *Show that $\varepsilon + f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle < f(x^{(k_2)})$*

6. *Show that $\varepsilon < 2L\|x^{(k_2)} - x^{(k_1)}\|$.*

7. *Prove that $f(x^{(k)}) \to v$.*

8. *(Optional - hard) Find a complexity bound for the method (that is a number of iteration N_ε after which you are sure to have obtained a ε-optimal solution).*

Answers:

1. f is finite convex and thus continuous on X which is compact, yielding the existence and finiteness of v.

   f is subdifferentiable, thus we have the existence of $g^{(k)}$, and an optimal solution to $\mathcal{P}^{(k)}$ exists as the solution of a bounded linear programm.

2. We have seen that on any compact Kincluded in the domain of a convex function f, f is L-Lipschitz. Here $\text{dom}(f) = \mathbb{R}^n$, so on the compact $K = X + B(0, \varepsilon)$ f is L-Lipschitz, and on X any subgradient g is of norm lower than L.

3. $f(x_k) \to v$ iff $\forall \varepsilon > 0, \exists N_\varepsilon \in \mathbb{N}, \quad k \geq K \implies f(x_k) \leq v + \varepsilon$. Hence $K_\varepsilon \subset [N_\varepsilon]$.

4. By subgradient inequality $f(y) \geq f(x^{(k)}) + \left\langle g^{(k)}, y - x^{(k)} \right\rangle$. Thus, for all k, $v \geq v^{(k)}$. Further, note that $v^{(k)} = f(x^{(k)})$, hence using $k = k_1$, and $y = x^{(k_2)}$ we get
$$f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle \leq \underline{v}^{(k_2)} \leq v.$$

5. As $k_2 \in K_\varepsilon$, we have $f(x^{(k_2)}) = v^{(k_2)} > v + \varepsilon \geq f(x^{(k_1)}) + \left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle + \varepsilon$ by the previous question.

6. We have
$$\begin{aligned}\varepsilon &< |f(x^{(k_2)}) - f(x^{(k_1)})| + |\left\langle g^{(k_1)}, x^{(k_2)} - x^{(k_1)} \right\rangle| \\ &\leq 2L\|x^{(k_2)} - x^{(k_1)}\|\end{aligned}$$

   by Cauchy-Schwartz and question 2.

7. If $f(x^{(k)}) \not\to v$, then there exists $\varepsilon > 0$ such that $(x^{(k)})_{k \in K_\varepsilon})$ is not finite. As X is compact we can exctract a converging subsequence, that is $x^{(\sigma(k)}$ such that $x^{(\sigma(k))} \to x^\star$ and $\sigma(k) \in K_\varepsilon$, which is in contradiction with the result of 6.