

Continuous Optimization Exam

04/06/2021

3 hours – documents allowed

The exam is made of 4 independent exercises, in roughly increasing difficulty. If necessary, you can admit the results of previous questions. When using the recalls, cite them.

Some useful recalls

- i) If $X \sim \mathcal{N}(\mu, \Sigma)$ is a Gaussian vector, then, for any vector u , we have $u^\top X \sim \mathcal{N}(u^\top \mu, u^\top \Sigma u)$.
- ii) The Fenchel transform of a function $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is given by $f^*(x) = \sup_{y \in \mathbb{R}^n} x^\top y - f(y)$.
- iii) Assume that f is a convex proper lsc function. f is μ -strongly convex, iff f^* is differentiable with $\frac{1}{\mu}$ -Lipschitz gradient.
- iv) Assume that f is a convex proper lsc function. Then $\lambda \in \partial f(x)$ iff $x \in \partial f^*(\lambda)$.
- v) Assume that f is a convex proper lsc function. Then $\lambda \in \partial f(x)$ iff $x \in \arg \max_y \lambda^\top y - f(y)$

Exercise 1: Warm-up

0 points

- (a) (1 point) On what conditions on the set C is \mathbb{I}_C a proper lower semicontinuous, convex function ?
- (b) (2 points) Write the KKT conditions for the following problem.

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \ln \left(\sum_{i=1}^n e^{x_i} \right) \\ \text{s.t.} \quad & \sum_{i=1}^n x_i = 0 \\ & \sum_{i=1}^n x_i^2 \leq 1 \end{aligned}$$

Are they necessary and/or sufficient conditions of optimality for this problem ?

- (c) (2 points) We consider the following problem

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

with value v and the following penalized versions

$$\begin{aligned} (P_t^{in}) \quad & \min_{x \in \mathbb{R}^n} \quad f(x) - t \sum_{i=1}^n \ln(-x_i) \quad \text{and} \quad (P_t^{out}) \quad \min_{x \in \mathbb{R}^n} \quad f(x) + t \sum_{i=1}^n (x_i)^+ \\ \text{s.t.} \quad & Ax = b, \quad x < 0 \quad \quad \quad \text{s.t.} \quad Ax = b \end{aligned}$$

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

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 ? What can you say about x_t^{in} ? Can you compare v_t^{in} and v ? Same questions for (P_t^{out})

Solution:

1. If C is closed convex, then so is $\text{epi}(\mathbb{I}_C) = C \times \mathbb{R}_+$, implying that \mathbb{I}_C is convex lsc. (0.75 pts) It is proper if C is non-empty. (0.25 pts)
2. The KKT conditions reads, there exists $\lambda \in \mathbb{R}^n$ and $\mu \in \mathbb{R}^+$ such that (1 pts)

$$\begin{cases} \frac{e^{x_i}}{\sum_{i=1}^n e^{x_i}} + \lambda_i + 2\mu x_i = 0 \\ \mu_i \geq 0 \\ \sum_i x_i = 0, \quad \sum_i x_i^2 \leq 1 \\ \mu = 0 \text{ or } \sum_i x_i^2 = 1 \end{cases}$$

The problem is convex, (0.25 pts) and qualified (0 is a Slater's point) (0.25 pts), thus conditions are necessary and sufficient. (0.5 pts)

3. For t going to 0 (0.25 pts) we have that (P_t^{in}) tends toward (P) : in the sense that $v_t^{(in)} \rightarrow v$ and x_t goes toward an optimal solution (0.25 pts). For t small enough we have $v_t^{in} \geq v$. (0.25 pts) In any case x_t^{in} is admissible. (0.25 pts)

For t going to $+\infty$ (0.25 pts), we have that (P_t^{out}) tends toward (P) in the sense that $v_t^{(out)} \rightarrow v$ and x_t^{out} goes toward an optimal solution (0.25 pts). For t large enough, x_t^{out} is optimal for (P) . (0.25 pts) We always have $v_t^{(out)} \leq v$. (0.25 pts)

Exercise 2: Support function (4 points)

0 points

For any set $C \subset \mathbb{R}^n$, we define its support function

$$\sigma_C : x \mapsto \sup_{c \in C} c^\top x$$

- (2 points) Assume that C and D are closed convex sets. Using a separation theorem, show that $C = D$ if and only if their support functions are equal.
- (2 points) For any set C , recall that the indicator function \mathbb{I}_C take value 0 on C and $+\infty$ outside. Show that, for any non empty set C , the Fenchel transform of its indicator function of set C is its support function, i.e. $\mathbb{I}_C^* = \sigma_C$. Deduce a second proof for the previous question.

Solution:

- (a) If $C = D$ their support function are equals. (0.5 pts) Now assume that $\sigma_C = \sigma_D$ and $C \neq D$. Without loss of generality we assume that there exists $x_0 \in D \setminus C$. As C is closed, it can be strictly separated from $\{x_0\}$, meaning that there is a vector a such that $\sup_{x \in C} a^\top x \leq b < a^\top x_0$. Thus $\sigma_C(a) \leq b < a^\top x_0 \leq \sigma_D(a)$. (1.5 pts)
- (b) $\mathbb{I}_C^*(x) = \sup_{y \in \mathbb{R}^n} x^\top y - \mathbb{I}_C(x) = \sup_{y \in C} x^\top y = \sigma_C(x)$. (0.5 pts) If C is non-empty closed convex, then \mathbb{I}_C is proper convex lsc (0.5 pts), and $\mathbb{I}_C = \mathbb{I}_C^{**} = \sigma_C^*$ (0.5 pts). If $\sigma_C = \sigma_D$, then $\sigma_C^* = \sigma_D^*$, and as C and D are closed convex, $\mathbb{I}_C = \mathbb{I}_D$, hence $C = D$. (0.5 pts).

Exercise 3: A linear problem with Gaussian cost

0 points

In the following we assume that $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^m$ are given matrices ; c is a gaussian random variable with mean $\bar{c} \in \mathbb{R}^n$ and variance $\Sigma \in \mathbb{R}^{n \times n}$.

(a) (2 points) We consider the following optimization problem

$$(P_\gamma) \quad \min_{x \in \mathbb{R}^n} \quad \mathbb{E}[c^\top x] + \gamma \text{Var}(c^\top x)$$

Show that P_γ is a quadratic program. Comment on the complexity of solving P_γ . (Hint : answer should depend on the value of the parameter $\gamma \in \mathbb{R}$).

(b) (2 points) We now consider the following problem

$$(P'_\alpha) \quad \min_{x \in \mathbb{R}^n, z \in \mathbb{R}} \quad z$$

$$\text{s.t.} \quad Ax \leq b$$

$$\mathbb{P}[c^\top x \geq z] \leq \alpha$$

Show that, for $\alpha \in]0, 0.5]$, (P'_α) is equivalent to an SOCP, using $\phi(t) = \frac{1}{\sqrt{2\pi}} \int_t^{+\infty} e^{-u^2/2} du$ (which is 1 minus the cdf of a centered gaussian), or its inverse ϕ^{-1} .

What happen if $\alpha \in]0.5, 1]$?

Solution:

(a) $\text{Var}(c^\top x) = \mathbb{E}[(c^\top x - \bar{c}^\top x)^2] = x^\top \mathbb{E}[(c - \bar{c})^\top (x - \bar{c})] x = x^\top \Sigma x$ (0.25 pts). Thus, (P_γ) reads

$$\min_{x \in \mathbb{R}^n} \quad \bar{c}^\top x + \gamma x^\top \Sigma x \quad (1\text{pts})$$

$$\text{s.t.} \quad Ax \leq b$$

If $\gamma = 0$, the problem is linear, which is simplest (0.25 pts). If $\gamma > 0$, the problem is quadratic convex, which is simple (0.25 pts). If $\gamma < 0$, the problem is quadratic non-convex, which is hard (0.25 pts).

(b) We have,

$$\mathbb{P}(c^\top x \geq z) = \phi\left(\frac{\beta - \bar{c}^\top x}{\|\Sigma^{1/2}x\|}\right)$$

thus (0.5 pts)

$$\mathbb{P}(c^\top x \geq z) \leq \alpha \Leftrightarrow \frac{\beta - \bar{c}^\top x}{\|\Sigma^{1/2}x\|} \geq \phi^{-1}(\alpha)$$

$$\phi^{-1}(\alpha)\|\Sigma^{1/2}x\|_2 + \bar{c}^\top x \leq z$$

Thus (P'_α) reads (0.5 pts)

$$\min_{x \in \mathbb{R}^n} \quad \phi^{-1}(\alpha)\|\Sigma^{1/2}x\|_2 + \bar{c}^\top x$$

$$\text{s.t.} \quad Ax \leq b$$

which is an SOCP (0.5 pts) if $\phi^{-1}(\alpha) \geq 0$, that is if $]0, 0.5]$. If $\alpha \in]0.5, 1]$, the problem is non-convex. (0.5 pts)

Exercice 4: Prox operator and Moreau-regularization

0 points

For any $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ proper convex function we define the proximal operator

$$\text{prox}_f : x \mapsto \arg \min_{y \in \mathbb{R}^n} f(y) + \frac{1}{2}\|y - x\|^2$$

and Moreau regularization of parameter $\mu > 0$

$$f_\mu : x \mapsto \inf_{y \in \mathbb{R}^n} f(y) + \frac{1}{2\mu} \|y - x\|^2$$

We want to study the *proximal point algorithm* given by the following sequence

$$x^{(k+1)} = \text{prox}_{\mu f}(x^{(k)}).$$

- (a) (1 point) Show that prox_f and f_μ are well defined. For C closed convex non empty, and $f = \mathbb{I}_C$ recognize prox_f and f_μ .
- (b) (1 point) Show that x^\sharp is a minimizer of f if and only if it minimizes f_μ , if and only if $x^\sharp = \text{prox}_f(x^\sharp)$.
- (c) (1 point) Show that $f_\mu(x) = \frac{1}{2\mu} \|x\|^2 - \frac{1}{\mu} (\mu f + \frac{1}{2} \|\cdot\|^2)^*(x)$.
- (d) (1 point) Show that $\text{prox}_{\mu f}(x) = \arg \max_y x^\top y - \mu f(y) - \frac{1}{2} \|y\|^2$.
- (e) (1 point) Show that $\nabla f_\mu(x) = \frac{1}{\mu} (x - \text{prox}_{\mu f}(x))$.
- (f) (1 point) Interpret the proximal point algorithm as a gradient algorithm.
- (g) (2 points) Writing

$$f_\mu(x) = \min_{y,z} f(y) + \frac{1}{2\mu} \|z\|^2$$

$$\text{s.t. } x - y = z$$

and using duality show that $f_\mu(x) = (f^* + \frac{\mu}{2} \|\cdot\|^2)^*(x)$. Deduce that f_μ has $\frac{1}{\mu}$ -Lipschitz gradient.

- (h) (1 point) Show that, if f is a proper convex lowersemicontinuous function, admitting a minimizer, the proximal point algorithm converges toward a minimizer of f .
- (i) (1 point) For the following problem

$$\min_{x \in \mathbb{R}^n} g(x) + h(x)$$

with $g \in \mathcal{C}^1$, we introduce the proximal gradient algorithm given as

$$x^{(k+1)} = \text{prox}_{\mu h} \left(x^{(k)} - \mu \nabla g(x^{(k)}) \right)$$

Recognize the proximal gradient algorithm for $h = 0$ first and then for $h = \mathbb{I}_C$, with g proper convex lowersemicontinuous, and $C \subset \text{dom}(g)$ closed convex non-empty.

Solution:

- (a) For any $x \in \mathbb{R}^n$, $y \mapsto f(y) + \frac{1}{2} \|y - x\|^2$ is proper lsc and strongly convex, thus admits a unique minimum. (0.5 pts)

$\text{prox}_f(x) = \arg \min_{y \in C} \|x - y\|^2$ is the projection on set C . (0.25 pts) $f_\mu(x) = \frac{1}{2\mu} \min_{y \in C} \|x - y\|^2 = \frac{d(x,C)^2}{2\mu}$

- (b) If x^\sharp minimizes f , we have,

$$f(x) + 1/2 \|x - x^\sharp\|^2 \geq f(x^\sharp) = f(x^\sharp) + 1/(2\mu) \|x^\sharp - x^\sharp\|^2$$

thus x^\sharp minimizes f_μ and $x^\sharp = \text{prox}_f(x^\sharp)$ (for $\mu = 1$). (0.5 pts)

On the other hand, $\tilde{x} = \text{prox}_f(v)$ minimizes f_1 iff

$$0 \in \partial f(\tilde{x}) + (\tilde{x} - v)$$

taking $v = \tilde{x} = x^\sharp$ we get $0 \in \partial f(x^\sharp)$. (0.5 pts)

(c) We have

$$\begin{aligned} f_\mu(x) &= \inf_y f(y) + \frac{1}{2\mu} \|x - y\|^2 \\ &= \inf_y \frac{\|x\|^2}{2\mu} - \frac{1}{\mu} x^\top y + \frac{1}{2\mu} \|y\|^2 \\ &= \frac{\|x\|^2}{2\mu} - \frac{1}{\mu} \sup_y \{x^\top y - \frac{1}{2} \|y\|^2\} \end{aligned}$$

(d) We have

$$\begin{aligned} \text{prox}_{\mu f}(x) &= \arg \min_y f(y) + \frac{1}{2} \|x - y\|^2 \\ &= \arg \min_y f(y) + \frac{1}{2} \|x\|^2 - x^\top y + \frac{1}{2} \|y\|^2 \\ &= \arg \max_y x^\top y - f(y) - \frac{1}{2} \|y\|^2 \end{aligned}$$

(e) Using the the recalls we have that $g = \nabla(\mu f + \frac{1}{2} \|\cdot\|^2)^*(x)$ iff $g \in \arg \max_y x^\top y - (\mu f + \frac{1}{2} \|y\|^2)$. Hence the previous questions yields the result.

(f) From the previous question we have $\text{prox}_{\mu f}(x) = x - \mu \nabla f_\mu(x)$ (0.5 pts). Thus $x^{(k+1)} = x - \mu \nabla f_\mu(x^{(k)})$ and the proximal point algorithm is the gradient with fixed step size μ applied to f_μ . (0.5 pts)

(g) We have the Lagrangian (0.5 pts) $\mathcal{L}(y, z, \lambda) = f(y) + \frac{1}{2\mu} \|z\|^2 + \lambda^\top (x - y - z)$ the dual function is (0.5 pts)

$$\begin{aligned} g(\lambda) &= \inf_y \{(f(y) - \lambda^\top y)\} + \inf_z \frac{1}{2\mu} \|z\|^2 - \lambda^\top z + \lambda^\top x \\ &= -f^*(\lambda) - \frac{\mu}{2} \|\lambda\|^2 + \lambda^\top x \end{aligned}$$

Strong duality (0.5 pts) yields

$$f_\mu(x) = \sup_\lambda g(\lambda) = (f^* + \frac{\mu}{2} \|\cdot\|^2)(x).$$

As f^* is convex, $f^* + \frac{\mu}{2} \|\cdot\|^2$ is μ -convex and thus $(f^* + \frac{\mu}{2} \|\cdot\|^2)$ is $1/\mu$ -smooth.

(h) As f_μ is $1/\mu$ -smooth (its gradient is $1/\mu$ -Lipschitz), the fixed step gradient algorithm (equivalent to the proximal point algorithm) is converging toward the minimum of f_μ (0.5 pts), which is the minimum of f (0.5 pts).