Convex Optimization Exam

03/06/2022
3 hours – documents allowed
Answers in English or French

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. "Classifying" an optimization problem consists in precising in which of the category presented in chapter 5 it falls (LP, QP, QCQP, SOCP, SDP, unconstrained or not, differentiable or not, continuous or not, convex or not).

Some usefull recalls

i) Recall that a step τ is deemed admissible in the backtracking step rules if $f(x_k + \tau d_k) \leq f(x_k) + \alpha\tau g_k^\top d_k$ for some $\alpha \in]0, 1/2[$.

ii) For $p, q \in]1, +\infty[$, $1/p + 1/q = 1$, we also have $q = \frac{p}{p-1}$ and $\frac{q}{p} + 1 = q$

iii) Let S_n be the set of symmetric real valued matrices. Then all $A \in S_n$ is diagonalizable. We denote S_n^+ (resp. S_n^{++}) the set of semidefinite (resp. definite) symmetric matrix, i.e. all eigenvalues are non-negative (resp. strictly positive). For $A, B \in S_n$, $A \preceq B$ iff $B - A$ is semidefinite positive (denoted $B - A \succeq 0$.

iv) S_n is an euclidean space, whose canonical scalar product is $\langle A, B\rangle = tr(AB)$.

Exercice 1: Warm-up **??** *points*

(a) (1 point) In Figure **??** we represent level set of some function. Are there some cases where the function cannot be convex ? Briefly justify.

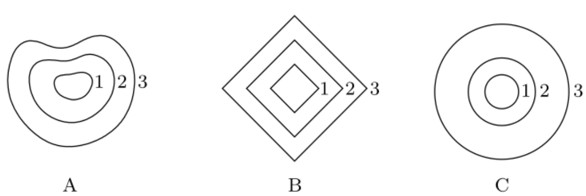

Figure 1: level set of potentially convex function ?

(b) (1 point) Consider $f : \mathbb{R}^2 \to \mathbb{R}$, $f : x \mapsto x_1^2 + x_2^2$, and $C = \{x_1 + x_2 \geq 1\}$. Gives, for every $x \in \mathbb{R}_2$ the normal cone $N_C(x)$. Use this to solve $\min_{x \in C} f(x)$ through the convex optimality condition.

Solution:

(a) A cannot be convex as level set are not convex (0.25 pts), B can be convex (0.25 pts) and C cannot be convex because of the spacing between the level line (0.5 pts).

(b) $N_C(x) = 0$ if $x_1 + x_2 > 1$, $N_C(x) = \emptyset$ if $x_1 + x_2 < 1$ and $N_C(x) = \{-\lambda(1,1) | \lambda \in \mathbb{R}^+\}$ if $x_1 + x_2 > 1$ (1 pts). We need to solve $-\nabla f(x) \in N_C(x)$ (0.5 pts), yielding $x^\sharp = (1/2, 1/2)$ (0.5 pts).

Exercice 2: Projection over the L_1 ball ?? *points*

Let $a \in \mathbb{R}^n$. We consider the following optimization problem.

$$\begin{array}{cl} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2}\|x - a\|^2 \\ \text{s.t.} & \|x\|_1 \le 1 \end{array}$$

(a) (1 point) Classify this problem, and justify that we have strong duality. Justify existence and unicity of optimal solution.

(b) (1 point) Write the (Lagrangian) dual problem as

$$\max_{\lambda \ge 0} g(\lambda) := \sum_{k=1}^{n} g_k(\lambda) - \lambda$$

where g_k should be given as analytical formula (i.e. without "min").

(c) (1 point) Show that $g'(\lambda) = \sum_{k=1}^{n}(|a_k| - \lambda)^+ - 1$.

(d) (1 point) Suggest an efficient method to find the optimal multiplier $\lambda^\sharp$.

(e) (1 point) Explain how to obtain the optimal primal solution $x^\sharp$ from $\lambda^\sharp$.

Solution:

(a) This problem is a convex (0.25 pts) QP (0.25 pts) (quadratic objective and linear constraints). Constraints are qualified as 0 is a Slater's point. (0.5 pts).

(b) The Lagrangian reads

$$L(x, \lambda) = \sum_{k=1}^{n} \left\{ \frac{1}{2}(x_k - a_k)^2 + \lambda |x_k| \right\} - \lambda$$

thus the dual problem reads

$$\max_{\lambda \ge 0} \sum_{k=1}^{n} g_k(\lambda) - \lambda$$

where

$$g_k(\lambda) = \min_{x_k} \frac{1}{2}(x_k - a_k)^2 + \lambda |x_k| = \begin{cases} -\lambda^2/2 + \lambda |a_k| & \text{if } \lambda < |a_k| \\ a_k^2/2 & \text{if } \lambda > |a_k| \end{cases}$$

(see last question for computational details)

(c) $g_k'(\lambda) = (|a_k| - \lambda)\mathbb{1}_{\lambda < |a_k|}$

(d) $g'(0) = \|a\|_1 - 1$, and $g'(\lambda) \to -1$. Thus 0 is optimal if $\|a\|_1 \le 1$, otherwise sorting $|a_k|$ and dichotomy yields the linear part of g' where it change sign, and then a simple linear equation provide $\lambda^\sharp$ such that $g'(\lambda^\sharp) = 0$.

(e) We need to minimize

$$h_k(x_k) := \frac{1}{2}(x_k - a_k)^2 + \lambda |x_k|$$

For $x_k \ge 0$, the optimum is either $h_k(0) = a_k^2/2$ or $h_k(a_k - \lambda) = \frac{1}{2}\lambda^2 - \lambda^2 + \lambda a_k$, with $0 \le \lambda < a_k$.
For $x_k \le 0$, the optimum is either $h_k(0) = a_k^2/2$ or $h_k(a_k + \lambda) = \frac{1}{2}\lambda^2 - \lambda^2 - \lambda a_k$, with $0 \ge \lambda > a_k$.
Collecting, we have

$$x_k^\sharp(\lambda) = [a_k - sgn(a_k)\lambda]\mathbb{1}_{|a_k| \le \lambda}$$

Exercice 3: Unit step in Quasi Newton's Method ?? *points*

We consider a C^2 strongly-convex function $f : \mathbb{R}^n \to \mathbb{R}$, and the following algorithm, for given x_0, $x_{k+1} = x_k + t_k d_k$ where, for all $k \in \mathbb{N}$,$d_k = -M_k^{-1} g_k$, $g_k := \nabla f(x_k)$ and M_k is a symmetric definite positive matrix such that

$$d^\top M_k d \geq d^\top \nabla^2 f(x_k) d + o(\|d\|^2)$$

(a) (1 point) Show that this algorithm is a descent algorithm.

(b) (2 points) Assume that x_k converges toward the minimizer of f. Show that there exists K such that, for all $k \geq K$, $t_k = 1$ is admissible for backtracking step rule.

Solution:

(a) $d_k^\top g_k = -g_k^\top M_k^{-1} g_k \leq 0$(0.5 pts) as M_k and thus M_k^{-1} is definite positive (0.5 pts).

(b) We have

$$\begin{aligned} f(x_k + d_k) &= f(x_k) + d_k^\top g_k + \frac{1}{2} d_k^\top \nabla^2 f(x_k) d_k + o(\|d_k\|^2) \\ &\leq f(x_k) + d_k^\top g_k + \frac{1}{2} d_k^\top M_k d_k + o(\|d_k\|^2) \end{aligned}$$

Thus,(1 pts)

$$\begin{aligned} f(x_k + d_k) - f(x_k) - \alpha d_k^\top g_k &\leq (1-\alpha) d_k^\top g_k + \frac{1}{2} d_k^\top M_k d_k + o(\|d_k\|^2) \\ &= -(1-\alpha) d_k^\top M_k d_k + \frac{1}{2} d_k^\top M_k d_k + o(\|d_k\|^2) \quad (g_k = -M_k d_k) \\ &= -(\frac{1}{2} - \alpha) d_k^\top M_k d_k + o(\|d_k\|^2) \end{aligned}$$

We end by noting that $d_k^\top M_k d_k \geq \lambda_1 \|d_k\|^2$(0.5 pts) by strong convexity, and $d_k \to 0$(0.5 pts) by convergence assumption.

Exercice 4: Minimizing linear functions on a ball ?? *points*

For $c \in \mathbb{R}^n$ we are interested in finding the solution of

$$\min_{x \in \mathbb{R}^n} \left\{ c^\top x \quad | \quad \|x\|_p \leq 1 \right\}$$

to prove Hölder inequality, *i.e.*,, $|x^\top y| \leq \|x\|_p \|y\|_q$ for $p, q \in]1, +\infty[$ such that $1/p + 1/q = 1$.

(a) (1 point) Reformulate the problem as a differentiable problem, and write the KKT conditions.

(b) (1 point) Find the optimal solution for $1 < p < +\infty$.

(c) (1 point) Show that $\min_{\|x\|_p \leq 1} c^\top x = -\|x\|_q$.

(d) (1 point) Deduce Hölder inequality from the previous question.

Solution:

(a) The problem Lagrangian is

$$L(x, \lambda) = c^\top x + \lambda \sum_{i=1}^{n} (|x_i|^p - 1)$$

First order conditions reads

$$\begin{aligned} c_i + p\lambda x_i |x_i|^{p-2} &= 0 \\ \lambda = 0 \text{ OR } \|x\|^p &= 1 \\ \lambda &\geq 0 \\ \|x\|_p &\leq 1 \end{aligned}$$

(b) We can assume that $c \neq 0$, thus $\lambda \neq 0$, and $\|x\|^p = 1$. We have

$$\begin{aligned} p\lambda |x_i|^{p-1} &= |c_i| \\ (p\lambda)^{p/(p-1)} |x_i|^p &= |c_i|^{p/(p-1)} \\ (p\lambda)^q &= \sum_{i=1}^n |c_i|^q \\ p\lambda &= \left(\sum_{i=1}^n |c_i|^q \right)^{1/q} = \|c\|_q \end{aligned}$$

And

$$x_i^\sharp = -sgn(c_i) \left(\frac{c_i}{\|c\|_q} \right)^{q/p}$$

(c)

$$v^\sharp = -\sum_{i=1}^n c_i sgn(c_i) \left(\frac{c_i}{\|c\|_q} \right)^{q/p} = -\frac{\sum_{i=1}^n |c_i|^{\frac{q}{p}+1}}{\|c\|_q^{q/p}} = -\frac{\|c\|_q^q}{\|c\|_q^{q/p}} = -\|c\|_q$$

(d) We have $c^\top \frac{x}{\|x\|_p} \geq -\|c\|_q$. Multiplying by $\|x\|_q$ and replacing c by $-c$ we get the result.

Exercice 5: Sum of largest eigenvalues **??** *points*

We consider the function $f : S_n \to \mathbb{R}$ given as the sum of the $r \leq n$ largest eigenvalues, that is

$$f(A) = \sum_{k=1}^r \lambda_r(A)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n$ are the eigenvalues of A.

(a) (2 points) Show that

$$\begin{aligned} f(A) = \max_{X \in S_n} \quad & \langle A, X \rangle \\ \text{s.t.} \quad & tr(X) = r \\ & 0 \preceq X \preceq I \end{aligned}$$

Classify this problem.

(b) (1 point) Show that f is convex

(c) (3 points) Consider $A(x) := \sum_{i=1}^K x_i A_i$ where $A_i \in S_n$, and the problem

$$\min_{x \in \mathbb{R}^K} \quad f(A(x))$$

Using duality, reformulate this problem as an SDP.

Solution:

(a) We reformulate the given problem with $A = P^\top DP$ where P is an orthonormal matrix and D a diagonal matrix. We set $Y = PXP^\top$ to obtain

$$\begin{aligned} \max_{Y \in S_n} \quad & \langle D, Y \rangle \\ \text{s.t.} \quad & tr(Y) = r \\ & 0 \preceq Y \preceq I \end{aligned}$$

Without loss of generality Y can be chosen diagonal. The semidefinite constraint ensures that its (diagonal) coefficient are between 0 and 1. The trace constraint ensures that the sum is one, thus the optimal solution charges to 1 the highest coefficient of D, that is, the highest eigenvalues.

(b) Maximum of affine functions.

(c) To derive the dual of the problem in part (a), we first write it as a minimization

$$\begin{array}{ll} \text{minimize} & -\mathbf{tr}(AX) \\ \text{subject to} & \mathbf{tr}\, X = r \\ & 0 \preceq X \preceq I. \end{array}$$

The Lagrangian is

$$\begin{aligned} L(X, \nu, U, V) &= -\mathbf{tr}(AX) + \nu(\mathbf{tr}\, X - r) - \mathbf{tr}(UX) + \mathbf{tr}(V(X - I)) \\ &= \mathbf{tr}((-A + \nu I - U + V)X) - r\nu - \mathbf{tr}\, V. \end{aligned}$$

By minimizing over X we obtain the dual function

$$g(\nu, U, V) = \begin{cases} -r\nu - \mathbf{tr}\, V & -A + \nu I - U + V = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem is

$$\begin{array}{ll} \text{maximize} & -r\nu - \mathbf{tr}\, V \\ \text{subject to} & A - \nu I = V - U \\ & U \succeq 0, \quad V \succeq 0. \end{array}$$

If we change the dual problem to a minimization and eliminate the variable U, we obtain a dual problem for the SDP in part (a) of the assignment:

$$\begin{array}{ll} \text{minimize} & r\nu + \mathbf{tr}\, V \\ \text{subject to} & A - \nu I \preceq V \\ & V \succeq 0. \end{array}$$

By strong duality, the optimal value of this problem is equal to $f(A)$. We can therefore minimize $f(A(x))$ over x by solving the

$$\begin{array}{ll} \text{minimize} & r\nu + \mathbf{tr}\, V \\ \text{subject to} & A(x) - \nu I \preceq V \\ & V \succeq 0, \end{array}$$

which is an SDP in the variables $\nu \in \mathbf{R}$, $V \in \mathbf{S}^n$, $x \in \mathbf{R}^m$.