Optimization and Algorithms
November 25, 2021

Write your name: ______________________________

Write your student number: ______________________________

Exam

1. *Nonconvex function.* (3 points) One of the following functions $f\colon \mathbf{R} \to \mathbf{R}$ is **not** convex:

(A) $f(x) = (x^2 - x)_+ - x$
(B) $f(x) = -\left((x_+)\right)^2 + x^2 + x$
(C) $f(x) = (x - x_+)^2 - x$
(D) $f(x) = \left((x_+)\right)^2 - x^2 + x$
(E) $f(x) = x_+ + x^2 - x$
(F) $f(x) = (x + x_+)^2 - (x_+)^2$

Which one?

Write your answer (A, B, C, D, E, or F) here: __________

2. *Least-squares.* (2 points) Consider the following six optimization problems:

(A)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \|A(c+x) - b\|_2^2 + \rho \|x\|_2^2$$

(B)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \|Ax - (Bx + b)\|_2^2 + \rho \|x - c\|_2^2$$

(C)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \|Ax\|_2^2 + \rho \|(B(x-c) + b)\|_2^2$$

(D)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \|(Ax - b) + \rho(Bx)\|_2^2 + \|x - c\|_2^2$$

(E)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \left(\|Ax - b\|_2 + \rho \|Bx\|_2\right)^2 + \|x - c\|_2^2$$

(F)
$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad (Ax + Bx - b)^T (Ax + Bx - b) + \rho x^T x$$

In each of the six problems above, the variable to optimize is $x \in \mathbf{R}^n$. The matrices A and B, and the vector c are given. The scalar ρ is also given and is positive: $\rho > 0$.

One of the optimization problems above is **not** a least-squares problem.

Which one?

Write your answer (A, B, C, D, E, or F) here: __________

3. *Optimal value of a constrained problem.* (3 points) Consider the constrained problem

$$\begin{array}{ll} \underset{x_1,\ldots,x_N}{\text{minimize}} & \underbrace{\frac{1}{2}\sum_{n=1}^{N} x_n^T R_n x_n}_{f(x_1,\ldots,x_N)} \\ \text{subject to} & x_1 + \cdots + x_N = s, \end{array}$$

where the variable to optimize is $(x_1, \ldots, x_N)$, with $x_n \in \mathbf{R}^d$ for $1 \le n \le N$. The matrices $R_n \in \mathbf{R}^{d \times d}$ are given for $1 \le n \le N$. Assume that each R_n is a symmetric, positive-definite matrix. The vector $s \in \mathbf{R}^d$ is also given.

One of the following expressions is the minimum value that f attains over the feasible set, that is, one of the following expressions is the number $\min\{f(x_1, \ldots, x_N) \colon x_1 + \cdots + x_N = s\}$:

(A) $\frac{1}{2} s^T (R_1 + \cdots + R_N)\, s$

(B) $\frac{1}{2} s^T \left(R_1^{-1} + \cdots + R_N^{-1}\right)^{-1} s$

(C) $\frac{1}{2} s^T \left(R_1^{-2} + \cdots + R_N^{-2}\right)^{-1} s$

(D) $\frac{1}{2} s^T \left(R_1^{-1} + \cdots + R_N^{-1}\right) s$

(E) $\frac{1}{2} s^T (R_1 + \cdots + R_N)^{-1}\, s$

(F) $\frac{1}{2} s^T \left(R_1^2 + \cdots + R_N^2\right) s$

Which one?

Write your answer (A, B, C, D, E, or F) here: __________

4. *Sparse linear regression with asymmetric loss.* (4 points) Consider the optimization problem

$$\underset{s\in\mathbf{R}^n, r\in\mathbf{R}}{\text{minimize}} \quad \underbrace{\sum_{k=1}^{K} \alpha\left((s^T x_k + r - y_k)_-\right)^2 + \beta\left((s^T x_k + r - y_k)_+\right)^2 + \rho\left\|s\right\|_1}_{f(s,r)},$$

where the variable to optimize is $(s,r) \in \mathbf{R}^n \times \mathbf{R}$. The vectors $x_k \in \mathbf{R}^n$ and the scalars $y_k \in \mathbf{R}$ are given for $1 \le k \le K$. The scalars α, β, and ρ are given and denote positive constants: $\alpha > 0$, $\beta > 0$, and $\rho > 0$. The functions $(\cdot)_-$ and $(\cdot)_+$ are defined as $(z)_- = \max\{-z, 0\}$ and $(z)_+ = \max\{z, 0\}$ for $z \in \mathbf{R}$.

Show that the function f is convex.

5. *A simple control problem.* (4 points) Consider the optimization problem

$$\begin{array}{ll} \underset{\{x_t\colon 1\le t\le T\},\{u_t\colon 1\le t\le T-1\}}{\text{minimize}} & \frac{1}{2}\left\|x_T\right\|_2^2 + \frac{\rho}{2}\sum_{t=1}^{T-1}\left\|u_t\right\|_2^2 \\ \text{subject to} & x_1 = x_{\text{initial}} \\ & x_{t+1} = x_t + D_t u_t \quad \text{for } 1 \le t \le T-1, \end{array}$$

where the variables to optimize are $x_t \in \mathbf{R}^n$ for $1 \le t \le T$ and $u_t \in \mathbf{R}^p$ for $1 \le t \le T-1$. The vector $x_{\text{initial}} \in \mathbf{R}^n$ and the matrices $D_t \in \mathbf{R}^{n\times p}$ are given for $1 \le t \le T-1$. The scalar ρ is also given and denotes a positive constant: $\rho > 0$.

Give a closed-form solution for the optimal $\{u_t\colon 1 \le t \le T-1\}$.

6. *Moureau envelope.* (4 points) Let $f\colon \mathbf{R} \to \mathbf{R}$ be a convex function. For $\lambda > 0$, we define a function $e_\lambda[f]\colon \mathbf{R} \to \mathbf{R}$ as follows: for $x \in \mathbf{R}$, the image of x under the function $e_\lambda[f]$ is the number $\min\{f(u) + \frac{1}{2\lambda}(u-x)^2\colon u \in \mathbf{R}\}$.

That is, the function $e_\lambda[f]$ maps each number x to the number $e_\lambda[f](x)$, where $e_\lambda[f](x)$ is the minimum value attained by $f(u) + \frac{1}{2\lambda}(u-x)^2$ as u varies in $\mathbf{R}$.

Let $\lambda_1 > 0$ and $\lambda_2 > 0$. Show that

$$e_{\lambda_1}[e_{\lambda_2}[f]](x) = e_{\lambda_1+\lambda_2}[f](x),$$

for each $x \in \mathbf{R}$.

João Xavier

OPTIMIZATION AND ALGORITHMS

(Instituto Superior Técnico)

– Solution of the exam –

Problem 1

D

Problem 2

E

Problem 3

B

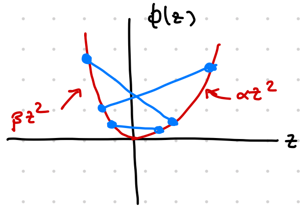
Problem 4

- Note that $f(s,r) = \sum_{k=1}^{K} \underbrace{\phi(s^T x_k + r - y_k)}_{f_k(s,r)} + \rho \underbrace{\|s\|_1}_{f_0(s,r)}$,

  where $\phi : \mathbb{R} \to \mathbb{R}$, $\phi(z) = \alpha (z_-)^2 + \beta (z_+)^2$.

- The function ϕ is convex. This can be seen directly from its graph:

$$\phi(z) = \begin{cases} \alpha z^2 & ; \text{ if } z \geq 0 \\ \beta z^2 & ; \text{ if } z < 0 \end{cases} \longrightarrow$$



- The function $f_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $f_k(s,r) = \phi(s^T x_k + r - y_k)$, can be decomposed as

$$f_k = \phi \circ g_k,$$

  where $g_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $g_k(s,r) = s^T x_k + r - y_k$.

  Because g_k is affine and ϕ is convex, we conclude that f_k is convex (for $1 \leq k \leq K$)

- The function $f_0 : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $f_0(s,r) = \|s\|_1$, can be decomposed as

$$f_0 = q \circ p,$$

where $p : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $p(s,r) = s$, and $q : \mathbb{R}^n \to \mathbb{R}$, $q(z) = \|z\|_1$.

Because p is an affine map and q is a (well-known) convex function, we conclude that f_0 is convex.

- Finally, $f = f_1 + \dots + f_K + \rho f_0$, being a linear combination with nonnegative weights of convex functions, is itself convex.

Problem 5

The constraints $\begin{cases} x_1 = x_{\text{initial}} \\ x_{t+1} = x_t + D_t u_t, \text{ for } 1 \leq t \leq T-1 \end{cases}$ imply $x_T = x_{\text{initial}} + D_1 u_1 + \dots + D_{T-1} u_{T-1}$.

Thus, our problem can be written as

$$\underset{u_1, \dots, u_{T-1}}{\text{minimize}} \quad {}^1\!/_2 \| x_{\text{initial}} + D_1 u_1 + \dots + D_T u_T \|^2 + \frac{\rho}{2} \sum_{t=1}^{T-1} \|u_t\|^2,$$

or, in a more compact form, as

$$\underset{u}{\text{minimize}} \quad \underbrace{{}^1\!/_2 \| x_{\text{initial}} + D u \|^2 + \frac{\rho}{2} \|u\|^2}_{f(u)},$$

where $D = [D_1 \ D_2 \ \cdots \ D_{T-1}]$ and $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{T-1} \end{bmatrix}$.

The function f is strongly convex, which implies that it has a unique minimizer. Because f is differentiable, its global minimizer, say u^*, is the point satisfying

$$\nabla f(u^*) = 0 \quad \Rightarrow \quad D^T (x_{\text{initial}} + D u^*) + \rho u^* = 0$$

$$\Rightarrow \quad (D^T D + \rho I) u^* = -D^T x_{\text{initial}}$$

$$\Rightarrow \quad u^* = -(D^T D + \rho I)^{-1} D^T x_{initial}.$$

The last step is valid because $D^T D + \rho I$ is invertible; indeed, it is a positive-definite matrix: for $v \neq 0$, we have

$$v^T(D^T D + \rho I)v = \underbrace{\|Dv\|^2}_{\geq 0} + \underbrace{\rho}_{>0} \underbrace{\|v\|^2}_{>0} > 0.$$

Problem 6 Let x be fixed.

We have

$$e_{\lambda_1}[e_{\lambda_2}[f]](x) = \min\left\{ e_{\lambda_2}[f](u) + \frac{1}{2\lambda_1}(u-x)^2 : u \in \mathbb{R}\right\}$$

$$= \min\left\{ \min\left\{ f(v) + \frac{1}{2\lambda_2}(v-u)^2 : v \in \mathbb{R}\right\} + \frac{1}{2\lambda_1}(u-x)^2 : u \in \mathbb{R}\right\}$$

$$= \min\left\{ f(v) + \frac{1}{2\lambda_1}(u-x)^2 + \frac{1}{2\lambda_2}(v-u)^2 : u, v \in \mathbb{R}\right\}$$

$$= \min\left\{ f(v) + \boxed{\min\left\{ \frac{1}{2\lambda_1}(u-x)^2 + \frac{1}{2\lambda_2}(v-u)^2 : u \in \mathbb{R}\right\}} : v \in \mathbb{R}\right\}$$

$$= \min\left\{ f(v) + \boxed{\frac{1}{2(\lambda_1+\lambda_2)}(v-x)^2} : v \in \mathbb{R}\right\}$$

$$= e_{\lambda_1+\lambda_2}[f](x)$$

⊛ proof that $\min\left\{ \frac{1}{2\lambda_1}(u-x)^2 + \frac{1}{2\lambda_2}(v-u)^2 : u \in \mathbb{R}\right\} = \frac{1}{2(\lambda_1+\lambda_2)}(v-x)^2$

- Let $\phi(u) = \frac{1}{2\lambda_1}(u-x)^2 + \frac{1}{2\lambda_2}(v-u)^2$, which is a strongly convex function

- Its global minimizer, say u^*, can be found by solving $\dot{\phi}(u^*) = 0$:

$$\dot{\phi}(u^*) = 0 \Rightarrow \frac{u^*-x}{\lambda_1} + \frac{u^*-v}{\lambda_2} = 0$$

$$\Rightarrow u^* = \frac{\lambda_1^{-1}x + \lambda_2^{-1}v}{\lambda_1^{-1} + \lambda_2^{-1}}$$

• Finally,

$$\min\{\phi(u) : u \in \mathbb{R}\} = \phi(u^*)$$

$$= \tfrac{1}{2}\lambda_1^{-1}(u^* - x)^2 + \tfrac{1}{2}\lambda_2^{-1}(u^* - v)^2$$

$$= \tfrac{1}{2}\lambda_1^{-1}\left(\frac{\lambda_2^{-1}(v-x)}{\lambda_1^{-1}+\lambda_2^{-1}}\right)^2 + \tfrac{1}{2}\lambda_2^{-1}\left(\frac{\lambda_1^{-1}(x-v)}{\lambda_1^{-1}+\lambda_2^{-1}}\right)^2$$

$$= \tfrac{1}{2}\,\frac{\lambda_1^{-1}\lambda_2^{-2} + \lambda_2^{-1}\lambda_1^{-2}}{(\lambda_1^{-1}+\lambda_2^{-1})^2}(v-x)^2$$

$$= \tfrac{1}{2}\,\frac{\lambda_1^{-1}\lambda_2^{-1}\cancel{(\lambda_1^{-1}+\lambda_2^{-1})}}{(\lambda_1^{-1}+\lambda_2^{-1})^{\cancel{2}}}(v-x)^2$$

$$= \tfrac{1}{2}\,\frac{1}{\lambda_1+\lambda_2}(v-x)^2$$