Optimization and Algorithms
November 17, 2022

Write your name: ______________________

Write your student number: ______________________

Write your answers (A, B, C, D, E, or F) to problems 1 to 4 in this box
Your answer to problem 1: ________
Your answer to problem 2: ________
Your answer to problem 3: ________
Your answer to problem 4: ________

Exam: Part A

1. *Positioning without covering a set.* Consider the set

$$S = \{x = (x_1, \ldots, x_n) \in \mathbf{R}^n \colon \; -1 \leq x_i \leq 1, \text{ for } 1 \leq i \leq n\}.$$

(In dimension $n = 2$, the set S looks like a filled square.)

We want to position a ball

$$B(c) = \{x \in \mathbf{R}^n \colon \; \|x - c\|_2 \leq r\}$$

such that the center c of the ball is as close as possible to a given point $d \in \mathbf{R}^n$ and such that the ball does not cover the set S (that is, we do not want to have $S \subseteq B(c)$). The radius $r > 0$ of the ball is given and cannot be changed.

For future reference, let $V \subset \mathbf{R}^n$ be the set of vectors of size n whose components are either 1 or -1. For example, in dimension $n = 2$, we have

$$V = \left\{ \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Note that, for $n = 2$, the set V is the set of vertices of the square S. (For general n, the set V has 2^n elements.)

Consider the following optimization problems.

(A)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c - d\|_2 \\ \text{subject to} & \max\{\|v - c\|_2 : v \in V\} > r \end{array} \tag{1}$$

(B)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c-d\|_2 \\ \text{subject to} & \max\{\|v-c\|_2 : v \in V\} = r \end{array} \tag{2}$$

(C)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c-d\|_2 \\ \text{subject to} & \min\{\|v-c\|_2 : v \in V\} > r \end{array} \tag{3}$$

(D)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c-d\|_2 \\ \text{subject to} & \max\{\|v-c\|_2 : v \in V\} \le r \end{array} \tag{4}$$

(E)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c-d\|_2 \\ \text{subject to} & \min\{\|v-c\|_2 : v \in V\} \le r \end{array} \tag{5}$$

(F)

$$\begin{array}{ll} \underset{c}{\text{minimize}} & \|c-d\|_2 \\ \text{subject to} & \min\{\|v-c\|_2 : v \in V\} = r \end{array} \tag{6}$$

One of these optimization is suitable for the given context.

Which one?

Write your answer (A, B, C, D, E, or F) in the box at the top of page 1

Hint: think about this problem in dimension $n = 2$.

2. *Unconstrained optimization.* Consider the optimization problem

$$\underset{x \in \mathbf{R}^2}{\text{minimize}} \quad (a^T x - b)^2 + (\|x\|_2 - c)_+^2 + \|x-d\|_2^2\,, \tag{7}$$

where

$$a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad b = 2, \quad \text{and} \quad c = 10.$$

The point

$$x^\star = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

is a global minimizer of (7) for one of the following choices of d:

(A) $d = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$

(B) $d = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

(C) $d = \begin{bmatrix} -4 \\ -3 \end{bmatrix}$

(D) $d = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

(E) $d = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

(F) $d = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$

Which one?

Write your answer (A, B, C, D, E, or F) in the box at the top of page 1

3. *Least-squares.* Let $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ be given, with n being an even number, and consider the following optimization problems:

(A)

$$\underset{x}{\text{minimize}} \quad \|A\, \texttt{rev}(x) - b\|_2^2$$

where $\texttt{rev}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that returns the input written in reverse order (that is, written from the last component to the first component). Examples (for $n = 4$):

$$\texttt{rev}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 0 \\ -1 \\ 4 \end{bmatrix}, \quad \texttt{rev}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \quad \texttt{rev}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ 4 \\ 1 \\ -1 \end{bmatrix}.$$

(B)

$$\underset{x}{\text{minimize}} \quad \|A\, \texttt{sort}(x) - b\|_2^2$$

where $\texttt{sort}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that returns the input sorted in ascending order. Examples (for $n = 4$):

$$\texttt{sort}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 0 \\ 3 \\ 4 \end{bmatrix}, \quad \texttt{sort}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 2 \end{bmatrix}, \quad \texttt{sort}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}.$$

(C)

$$\underset{x}{\text{minimize}} \quad \|A\, \texttt{circ}(x) - b\|_2^2$$

where $\texttt{circ}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that returns the input circulated by one component (that is, the first component of the input x becomes the second component of the output, the second component of x becomes the third component of the output, and so on, and the last component of x becomes the first component of the output). Examples (for $n = 4$):

$$\texttt{circ}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 4 \\ -1 \\ 0 \end{bmatrix}, \quad \texttt{circ}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 2 \\ -1 \\ 1 \end{bmatrix}, \quad \texttt{circ}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ -1 \\ 1 \\ 4 \end{bmatrix}.$$

(D)

$$\underset{x}{\text{minimize}} \quad \|A\,\texttt{cent}(x) - b\|_2^2$$

where $\texttt{cent}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that shifts the input so that it becomes centered at the origin; more precisely, given the input x, the map subtracts from x the average of the components of x. Examples (for $n = 4$):

$$\texttt{cent}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix} - \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \\ 3/2 \end{bmatrix}, \quad \texttt{cent}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

$$\texttt{cent}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}.$$

(E)

$$\underset{x}{\text{minimize}} \quad \|A\,\texttt{trim}(x) - b\|_2^2$$

where $\texttt{trim}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that trims one component at the beginning and at the end of the input; that is, given the input x, the map zeroes the first and last components of x, while keeping the other components of x intact. Examples (for $n = 4$):

$$\texttt{trim}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \texttt{trim}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \quad \texttt{trim}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 0 \end{bmatrix}.$$

(F)

$$\underset{x}{\text{minimize}} \quad \|A\,\texttt{swap}(x) - b\|_2^2$$

where $\texttt{swap}\colon \mathbf{R}^n \to \mathbf{R}^n$ is the map that swaps each consecutive pair of components of x (that is, given $x = (x_1, x_2, \ldots, x_n)$, it swaps x_1 with x_2, then x_3 with x_4, and so on, until x_{n-1} with x_n). Examples (for $n = 4$):

$$\texttt{swap}\left(\begin{bmatrix} 4 \\ -1 \\ 0 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 4 \\ 3 \\ 0 \end{bmatrix}, \quad \texttt{swap}\left(\begin{bmatrix} 2 \\ -1 \\ 1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \quad \texttt{swap}\left(\begin{bmatrix} -1 \\ 1 \\ 4 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \\ 4 \\ 4 \end{bmatrix}.$$

One of the above problems is <u>not</u> a least-squares problem. Which one?

Write your answer (A, B, C, D, E, or F) in the box at the top of page 1

4. *Newton algorithm.* Consider the Newton algorithm with the generic update equation given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{8}$$

where α_k denotes the stepsize.

Suppose that the Newton algorithm is applied to the function $f\colon \mathbf{R}^2 \to \mathbf{R}$,

$$f(a,b) = (2a-b)^2 + b^2,$$

and suppose that the current iterate is

$$x_k = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

One of the following vectors is then the vector d_k in (8).

(A)

$$\begin{bmatrix} 0 \\ 4 \end{bmatrix}$$

(B)

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

(C)

$$\begin{bmatrix} 0 \\ -4 \end{bmatrix}$$

(D)

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

(E)

$$\begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

(F)

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Which one?

Write your answer (A, B, C, D, E, or F) in the box at the top of page 1

Optimization and Algorithms
November 17, 2022

Exam: Part B

1. *Convex problem.* Show that the following problem is convex,

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \sum_{k=1}^{K} \phi\left(a_k^T x - b_k\right) + \lambda \psi\left(\|x - c\|_2 - r\right) \\ \text{subject to} & \|x\|_1 \leq \rho, \end{array} \tag{1}$$

where the vectors $a_k \in \mathbf{R}^n$ (for $1 \leq k \leq K$) and $c \in \mathbf{R}^n$. The numbers b_k (for $1 \leq k \leq K$), $\lambda > 0$, $r > 0$, and $\rho > 0$ are also given.

The function $\phi \colon \mathbf{R} \to \mathbf{R}$ is defined as

$$\phi(t) = \begin{cases} -t, & \text{if } t < 0 \\ 2t, & \text{if } t \geq 0. \end{cases}$$

The function $\psi \colon \mathbf{R} \to \mathbf{R}$ is defined as

$$\psi(t) = \begin{cases} 0, & \text{if } t < 0 \\ t, & \text{if } 0 \leq t < 1 \\ 2t - 1, & \text{if } t \geq 1. \end{cases}$$

2. *Equivalent problems?* Bob wants to compute a point in the hyperplane $H(s, r) = \left\{x \in \mathbf{R}^n \colon s^T x = r\right\}$ that is closest to given points $p_k \in \mathbf{R}^n$, for $1 \leq k \leq K$, in the mean squared-distance sense. That is, Bob wants to find a global minimizer of the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{K} \sum_{k=1}^{K} \|x - p_k\|_2^2 \\ \text{subject to} & x \in H(s, r). \end{array} \tag{2}$$

Alice claims that this problem is equivalent to compute a point in the hyperplane $H(s, r)$ that is closest to the center-of-mass of the points p_k, for $1 \leq k \leq K$. That is, Alice claims that the global minimizers of (2) are the same as the global minimizers of

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \left\| x - \left(\sum_{k=1}^{K} p_k\right) / K \right\|_2^2 \\ \text{subject to} & x \in H(s, r). \end{array} \tag{3}$$

Is Alice correct or wrong? If you think Alice is wrong, then provide a counter-example: that is, provide an hyperplane $H(s, r)$ and points $p_1, \ldots, p_K$ such that the global minimizers of (2) and (3) are not the same. If you think Alice is correct, then prove that the global minimizers of (2) and (3) are the same (for any hyperplane $H(s, r)$ and points $p_1, \ldots, p_K$).

3. *Distance between two parallel hyperplanes.* Consider two parallel hyperplanes in $\mathbf{R}^n$ given by

$$H_1 = \left\{x \in \mathbf{R}^n \colon s^T x = r_1\right\}, \text{ and } H_2 = \left\{x \in \mathbf{R}^n \colon s^T x = r_2\right\},$$

where s is a nonzero vector in $\mathbf{R}^n$, and r_1 and r_2 are two distinct numbers ($r_1 \neq r_2$).

We wish to find the distance between these two hyperplanes. That is, denoting this distance by $d\left(H_1, H_2\right)$, we wish to compute

$$d\left(H_1, H_2\right) = \inf\left\{\|x_1 - x_2\|_2 : x_1 \in H_1, x_2 \in H_2\right\}. \tag{4}$$

Give a closed-form expression for the distance $d\left(H_1, H_2\right)$ in terms of s, r_1, and r_2.

Hint: use KKT conditions on the problem (4).

4. *Dead-zone quadratic penalty.* The dead-zone quadratic penalty function with bandwidth $B > 0$ is the function $\phi \colon \mathbf{R}^n \to \mathbf{R}$ defined as

$$\phi_B(x) = \left(\|x\|_2 - B\right)_+^2.$$

Let $H(s, r)$ be a given hyperplane

$$H(s, r) = \left\{x \in \mathbf{R}^n \colon s^T x = r\right\},$$

with $s \in \mathbf{R}^n$ and $r \in \mathbf{R}$.

Give a closed-form expression for a global minimizer of the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \phi_B\left(x - c\right) \\ \text{subject to} & x \in H(s, r), \end{array}$$

where $c \in \mathbf{R}^n$ and $B > 0$ are given. Assume that c is not in the hyperplane ($c \notin H(s, r)$) and that s is a nonzero vector ($s \neq 0$).

Solution of Exam 1 – Part A

Problem 1 Answer: A

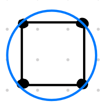
Details

1. Note that $S \not\subset B(c)$ if and only if $V \not\subset B(c)$. That is, the ball $B(c)$ does not cover S if and only if there exists a vector $v \in V$ such that $v \notin B(c)$.

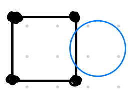
Thus:

$$
\begin{aligned}
S \not\subset B(c) &\Leftrightarrow \underset{v \in V}{\exists} : v \notin B(c) \\
&\Leftrightarrow \underset{v \in V}{\exists} : \|v - c\|_2 > r \\
&\Leftrightarrow \underbrace{\max \{ \|v - c\|_2 : v \in V \} > r.}_{\text{option A}}
\end{aligned}
$$

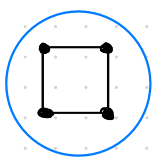
2. Option B is incorrect because it allows



3. Option C is incorrect because it forbids



4. Option D is incorrect because it allows



⑤ Option E is incorrect because it allows the configuration shown in ②

⑥ Option F is incorrect because it allows the configuration shown in ②

Problem 2 Answer: E

Details

① The problem is convex, and the function

$$f(x) = (a^T x - b)^2 + (\|x\|_2 - c)_+^2 + \|x - d\|_2^2$$

is differentiable at $x^* = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$.

② So, x^* is a global minimizer for f if and only if

$$\nabla f(x^*) = 0 \iff 2(a^T x^* - b)a + 2(\|x^*\|_2 - c)_+ \frac{x^*}{\|x^*\|_2} + 2(x^* - d) = 0$$

$$\iff d = (a^T x^* - b)a + (\|x^*\|_2 - c)_+ \frac{x^*}{\|x^*\|_2} + x^*$$

$$\iff d = \underbrace{\left([1 \ 1] \begin{bmatrix} -3 \\ 4 \end{bmatrix} - 2 \right)}_{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underbrace{\left(\left\| \begin{bmatrix} -3 \\ 4 \end{bmatrix} \right\|_2 - 10 \right)_+}_{0} \frac{1}{\left\| \begin{bmatrix} -3 \\ 4 \end{bmatrix} \right\|_2} \begin{bmatrix} -3 \\ 4 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

$$\iff d = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} -3 \\ 4 \end{bmatrix}$$

$$\iff d = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

Problem 3 Answer: B

Details

(1) Recall that an LS problem is an optimization problem of the form

$$\min_{x} \quad \| \mathcal{A} x - \beta \|_2^2,$$

for some matrix $\mathcal{A}$ and vector β

(2) In option A, the map $\text{rev}: \mathbb{R}^n \to \mathbb{R}^n$ is a linear map, that is, it can be written in the form $\text{rev}(x) = Mx$, for some matrix M. Indeed, taking the special case $n=4$, we have

$$\text{rev}\Bigg(\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}\Bigg) = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}.$$

This means that

$$\min_{x} \quad \| A\,\text{rev}(x) - b \|_2^2 \iff \min_{x} \quad \| \underbrace{AM}_{\mathcal{A}}\, x - \underbrace{b}_{\beta} \|_2^2$$

that's a LS problem

(3) The reasoning above applies also to options C, D, E, and F; only the matrix will change:

- option C

$$\text{circ}\Bigg(\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}\Bigg) = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}$$

- option D

$$\text{cent}\Bigg(\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}\Bigg) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} - \begin{bmatrix} (x_1+x_2+x_3+x_4)/4 \\ (x_1+x_2+x_3+x_4)/4 \\ (x_1+x_2+x_3+x_4)/4 \\ (x_1+x_2+x_3+x_4)/4 \end{bmatrix} = \underbrace{\begin{bmatrix} 3/4 & -1/4 & -1/4 & -1/4 \\ -1/4 & 3/4 & -1/4 & -1/4 \\ -1/4 & -1/4 & 3/4 & -1/4 \\ -1/4 & -1/4 & -1/4 & 3/4 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}$$

- <u>option E</u>

$$\mathrm{trim}\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}$$

with $\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}$

- <u>option F</u>

$$\mathrm{swap}\left(\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}\right) = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{x}$$

Problem 4

Answer: E

<u>Details</u>

① In the Newton algorithm $x_{k+1} = x_k + \alpha_k d_k$, the direction d_k is given by

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

② For $f(a,b) = (2a-b)^2 + b^2$, we have

$$\nabla f(a,b) = \begin{bmatrix} 8a - 4b \\ -4a + 4b \end{bmatrix} \qquad \nabla^2 f(a,b) = \begin{bmatrix} 8 & -4 \\ -4 & 4 \end{bmatrix}$$

③ Evaluating ∇f and $\nabla^2 f$ at $x_k = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ gives

$$\nabla f(x_k) = \begin{bmatrix} 0 \\ 4 \end{bmatrix} \qquad \nabla^2 f(x_k) = \begin{bmatrix} 8 & -4 \\ -4 & 4 \end{bmatrix}$$

④ Thus,

$$d_k = -\begin{bmatrix} 8 & -4 \\ -4 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

Optimization and algorithms, IST Dec 2022, João Xavier

Solution of Exam 1 - Part B

Problem 1

① The problem

$$\min_x \quad \underbrace{\sum_{k=1}^{K} \phi(a_k^T x - b_k) + \lambda \Psi(\|x-c\|_2 - r)}_{f(x)}$$

$$\text{s.t.} \quad \underbrace{\|x\|_1 - r}_{g(x)} \leq 0$$

is convex if f and g are convex, which we now show

② g is convex

- g is the sum of a norm (a well-known convex function) with a constant (which does not affect convexity)
- so, g is convex

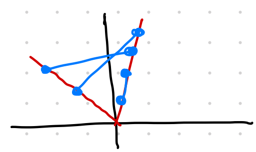
③ f is convex

- write f as nonnegative linear combination of $f_1, \ldots, f_K$, and h:

$$f = f_1 + f_2 + \cdots + f_K + \lambda h,$$

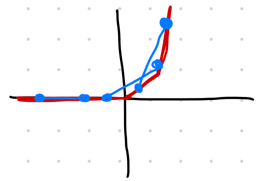
where $f_k(x) = \phi(a_k^T x - b_k)$ and $h(x) = \Psi(\|x-c\|_2 - r)$

- if we show $f_1, \ldots, f_K, h$ are convex, the function f will be convex
- f_k is convex: f_k is the composition $\phi \circ p_k$, where $p_k(x) = a_k^T x - b_k$ is affine and ϕ is convex (it suffices to look at the graph of ϕ:



)

- <u>h is convex</u> h is the composition $\psi \circ q$, where ψ is convex and non-decreasing (as obvious from its graph



),

and q is convex: q is the composition of a convex function, $\|\cdot\|_2$, with the affine map $x \mapsto x - c$ (the additive constant $-r$ is irrelevant for the convexity of q)

Problem 2 Alice is right:

(1) Bob's problem is

$$\min_x \underbrace{\frac{1}{K}\sum_{k=1}^{K} \|x - p_k\|_2^2}_{f_{BOB}(x)}$$

$$\text{s.t. } s^T x = r,$$

whose KKT conditions are

$$\begin{cases} \nabla f_{BOB}(x) = s\lambda \\ s^T x = r \end{cases} \rightarrow \begin{cases} 2(x - \bar{p}) = s\lambda \\ s^T x = r, \end{cases}$$

where $\bar{p} = \frac{1}{K}\sum_{k=1}^{K} p_k$

(2) Alice's problem is

$$\min_x \underbrace{\|x - \bar{p}\|_2^2}_{f_{ALICE}(x)}$$

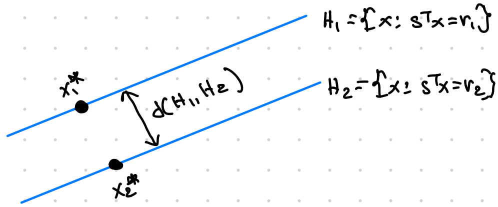
$$\text{s.t. } s^T x = r,$$

whose KKT conditions are

$$\begin{cases} \nabla f_{ALICE}(x) = s\lambda \\ s^T x = r \end{cases} \rightarrow \begin{cases} 2(x - \bar{p}) = s\lambda \\ s^T x = r \end{cases}$$

(3) The KKT conditions of Bob and Alice are the same; so, the minimizers are the same

Problem 3



① The goal is to find $d(H_1, H_2)$:

② The quantity $d(H_1,H_2)$ is equal to $\|x_1^* - x_2^*\|_2$ whenever x_1^*, x_2^* are solutions of the problem

$$\min_{x_1, x_2} \quad \underbrace{\tfrac{1}{2}\|x_1 - x_2\|^2}_{f(x_1,x_2)}$$

$$\text{s.t.} \quad \underbrace{s^T x_1 - r_1}_{h_1(x_1,x_2)} = 0$$

$$\underbrace{s^T x_2 - r_2}_{h_2(x_1,x_2)} = 0,$$

whose KKT conditions are

$$\begin{cases} \nabla f(x_1,x_2) = \nabla h_1(x_1,x_2)\lambda_1 + \nabla h_2(x_1,x_2)\lambda_2 \\ h_1(x_1,x_2) = 0 \\ h_2(x_1,x_2) = 0 \end{cases} \rightarrow \begin{cases} \begin{bmatrix} x_1 - x_2 \\ x_2 - x_1 \end{bmatrix} = \begin{bmatrix} s \\ 0 \end{bmatrix}\lambda_1 + \begin{bmatrix} 0 \\ s \end{bmatrix}\lambda_2 \\ s^T x_1 = r_1 \\ s^T x_2 = r_2 \end{cases} \rightarrow \begin{cases} x_1 - x_2 = s\lambda_1 & (i) \\ x_2 - x_1 = s\lambda_2 & (ii) \\ s^T x_1 = r_1 & (iii) \\ s^T x_2 = r_2 & (iv) \end{cases}$$

③ Subtract (iii) and (iv) to get

$$s^T(x_1 - x_2) = r_1 - r_2,$$

and plug $x_1 - x_2 = s\lambda_1$ (from (i)) to arrive at

$$s^T s \lambda_1 = r_1 - r_2 \quad \Rightarrow \quad \lambda_1 = \frac{r_1 - r_2}{\|s\|_2^2}$$

④ Using $\lambda_1 = \dfrac{r_1 - r_2}{\|s\|_2^2}$ in (i) gives $\quad x_1 - x_2 = \dfrac{(r_1 - r_2)}{\|s\|_2^2} s$

⑤ This means $\quad d(H_1, H_2) = \|x_1 - x_2\| = \dfrac{|r_1 - r_2|}{\|s\|_2}$

Problem 4

① The problem is

$$\min_{x} \underbrace{\left(\|x-c\|_2 - B \right)_+^2}_{f(x)}$$

$$\text{s.t.} \quad \underbrace{s^T x - r}_{h(x)} = 0,$$

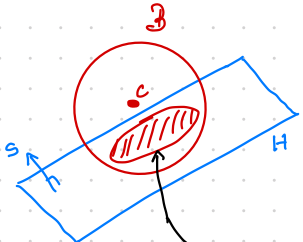
whose KKT conditions are

$$\begin{cases} \nabla f(x) = \nabla h(x) \lambda \\ h(x) = 0 \end{cases} \longrightarrow \begin{cases} 2\left(\|x-c\| - B\right)_+ \dfrac{x-c}{\|x-c\|_2} = s\lambda \\ s^T x = r \end{cases}$$

② Two cases can arise: $\left(\|x-c\| - B\right)_+ = 0$ or $\left(\|x-c\| - B\right)_+ > 0$

③ <u>case $\left(\|x-c\| - B\right)_+ = 0$</u>

- this means that we are investigating the existence of a minimizer x that satisfies $\|x-c\|_2 \le B$, that is, x is in the ball $\mathcal{B} = \{u : \|u-c\|_2 \le B\}$



- in this case, KKT system becomes

$$\begin{cases} 0 = s\lambda \ (\Rightarrow \lambda = 0 \text{ because } s \neq 0) \\ s^T x = r, \ (\Rightarrow x \in H) \end{cases}$$

with solutions being all pairs (x, λ) satisfying

$$x \in \underbrace{H \cap \mathcal{B}} \quad \text{and} \quad \lambda = 0$$

④ <u>case $\left(\|x-c\|_2 - B\right)_+ > 0$</u>

- in this case, KKT system becomes

$$\begin{cases} 2\left(\|x-c\|_2 - B\right) \dfrac{x-c}{\|x-c\|_2} = s\lambda & \text{(i)} \\ s^T x = r & \text{(ii)} \end{cases}$$

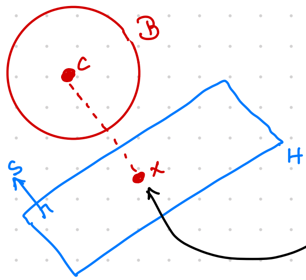
- (i) implies

$$x = c + \underbrace{\frac{\|x-c\|_2}{2(\|x-c\|_2 - B)}\lambda}_{\gamma} s,$$

that is, x is of the form $x = c + \gamma s$, for some $\gamma \in \mathbb{R}$

- plug $x = c + \gamma s$ into (ii) to get

$$s^T(c+\gamma s) = r \Rightarrow \gamma = \frac{r - s^T c}{\|s\|_2^2},$$



which means

$$x = c + \frac{r - s^T c}{\|s\|_2^2} s$$

- such x above is a solution in case it also satisfies $x \notin B$