Optimization and Algorithms 2023/24

Instituto Superior Técnico

2^{nd} exam – January 29, 2024

Student ID: __________ **Name:** ____________________________________

- Exam duration 2h
- Mark the answers to multiple-choice questions in the table below. For a question worth C points with n options to choose from, you will lose $C/(n-1)$ points if you answer it incorrectly. There is no penalty for leaving it blank
- At the end of the exam you should turn in the table of answers to multiple-choice questions, as well as the answers to open-ended problems **in separate sheets**
- Make sure that you fill out your **name and student ID in all sheets**
- **Justify** your answers to all open-ended problems. Portuguese-speaking students should answer in **Portuguese**, others in **English**

Answers to multiple-choice questions

	a	b	c	d	e	f	g	h	i
Q1									
Q2									
Q3									
Q4									

Q 1. [1.5 val] For any given points $x, x_0 \in \mathbb{R}^n$ a twice differentiable convex function f satisfies

$$f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{3}{2}\|x - x_0\|^2$$

Choose the statement below that follows from this.

a) $\nabla f(x_0) \neq 0$

b) $f(x)$ is quadratic

c) Maximum eigenvalue: $\lambda_{\max}(\nabla^2 f(x_0)) \leq 2$

d) $f((1-\alpha)x + \alpha x_0) \geq (1-\alpha)f(x) + \alpha f(x_0) - 3\frac{\alpha(1-\alpha)}{2}\|x - x_0\|^2$ for all $0 \leq \alpha \leq 1$

e) Minimum eigenvalue: $\lambda_{\min}(\nabla^2 f(x_0)) \geq 3$

f) $f(x_0) + \nabla f(x_0)^T(x - x_0)$ is quadratic in x

Q 2. [1.5 val] A company has developed three new products and is considering bringing them to market under the following constraints:

- At most two of them will be commercialized.
- The company operates two factories that could produce any of these products. However, to keep costs low the chosen products should be produced at the same factory.

The main problem variables are x_1, x_2, x_3, which denote the weekly production volume for products 1, 2, and 3, respectively.

- The expected profit is $5x_1 + 7x_2 + 3x_3$.
- If produced in the first factory, the required manufacturing time would be $3x_1 + 4x_2 + 2x_3$ hours. A maximum of 30 hours per week could be allocated for this. In the second factory the same production volume would take $4x_1 + 6x_2 + 2x_3$ hours to complete, in no more than 40 hours.
- The market cannot absorb more than 7 units of product 1 per week, so it is pointless to produce beyond this volume. For the same reason, no more than 5 units of product 2 and 9 units of product 3 should be produced per week.

The following formulation is proposed to maximize profit while respecting the stated constraints:

$$
\begin{aligned}
\underset{\substack{x_i,\ i=1,\ldots,3,\\ y_i,\ i=1,\ldots,4}}{\text{maximize}} \quad & 5x_1 + 7x_2 + 3x_3 \\
\text{subject to} \quad & \boxed{\quad ① \quad}, \\
& y_1 + y_2 + y_3 \leq 2, \\
& x_i \geq 0,\ i = 1, \ldots, 3, \\
& y_i \in \{0,\ 1\},\ i = 1, \ldots, 4
\end{aligned}
\tag{1}
$$

Choose below the content for box ①.

a) $3x_1 + 4x_2 + 2x_3 \leq 30$
$4x_1 + 6x_2 + 2x_3 \leq 40$
$x_1 \leq 7, \quad x_2 \leq 5, \quad x_3 \leq 9$

b) $3x_1 + 4x_2 + 2x_3 \leq 30 + 1000y_4$
$4x_1 + 6x_2 + 2x_3 \leq 40 + 1000(1 - y4)$
$x_1 \leq 7y_1, \quad x_2 \leq 5y_2, \quad x_3 \leq 9y_3$

c) $3x_1 + 4x_2 + 2x_3 \leq 30 + 1000(1 - y_4)$
$4x_1 + 6x_2 + 2x_3 \leq 40 + 1000y4$
$x_1y_1 \geq 7, \quad x_2y_2 \geq 5, \quad x_3y_3 \geq 9$

d) $3x_1 + 4x_2 + 2x_3 \leq 30y_4$
$4x_1 + 6x_2 + 2x_3 \leq 40y4$
$x_1 \geq 7, \quad x_2 \geq 5, \quad x_3 \geq 9$

e) $3x_1 + 4x_2 + 2x_3 \leq 30$
$7x_1 + 5x_2 + 9x_3 \geq 1$
$y_1 + y_2 + y_3 \leq y_4$

f) $4x_1 + 6x_2 + 2x_3 \leq 40$
$7x_1 + 5x_2 + 9x_3 \geq 70$

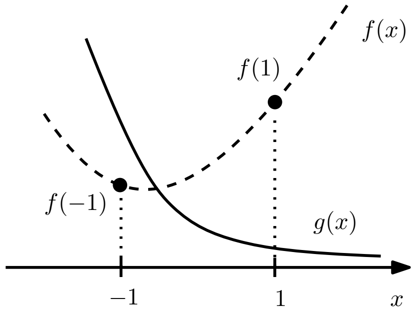
Q 3. [1.5 val] Consider the constrained optimization problem

$$
\begin{aligned}
\underset{x,\,y}{\text{minimize}} \quad & (x-1)^2 + 2(y-2)^2 \\
\text{subject to} \quad & x - 2y = 0
\end{aligned}
$$

whose optimal solution is $(x^*, y^*) = (2,\ 1)$. The augmented Lagrangian method, initialized with penalty factor $c_0 = 2$ and Lagrange factor $\lambda_0 = -1$, is used to iteratively solve the problem. Assuming that the method converges very close to the true solution of the original constrained problem within 20 iterations, what should be the final value λ_{20}?

a) -6 **b)** -4 **c)** -2 **d)** 2 **e)** 4 **f)** 6

Q 4. [1.5 val] You are given a non-negative function $f(x)$ whose values are unimportant except at $x = \pm 1$, as shown in the figure.



You wish to tune a non-negative function of the form $g(x) = \log(1+e^{-\gamma x})$, with adjustable parameter γ, so that it becomes as *dissimilar* as possible to f. Similarity between f and g is quantified by the inner product of $(f(-1), f(1))$ with $(g(-1), g(1))$, that is, $f(-1)g(x)|_{x=-1} + f(1)g(x)|_{x=1}$. Which of the following expressions reflects a first-order stationarity condition for γ to minimize the similarity?

a) $f(-1)\log(1+e^{\gamma})+f(1)\log(1+e^{-\gamma}) = 0$

b) $f(-1)e^{\gamma} = f(1)e^{-\gamma}$

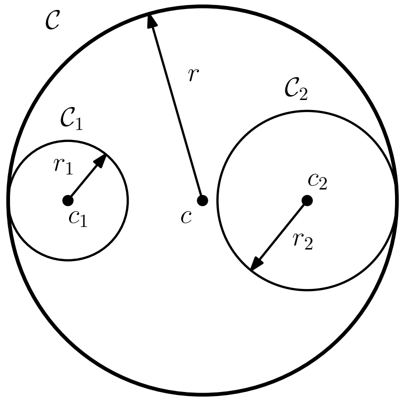
c) $\gamma = 1$ or $\gamma = -1$

d) $\left.\frac{dg}{d\gamma}\right|_{x=-1} = \left.\frac{dg}{d\gamma}\right|_{x=1} = 0$

e) $f(-1)\frac{e^{\gamma}}{1+e^{\gamma}} = f(1)\frac{e^{-\gamma}}{1+e^{-\gamma}}$

f) $f(-1)(1+e^{-\gamma}) = f(1)(1+e^{\gamma})$

P 1. [5.5 val] We wish to determine the tightest circle $\mathcal{C}$ in 2D that encloses a given collection of smaller circles $\mathcal{C}_1, \ldots, \mathcal{C}_N$. Each circle $\mathcal{C}_n$ is defined by its center $c_n \in \mathbb{R}^2$ and radius $r_n \geq 0$, such that $\mathcal{C}_n = \{x \in \mathbb{R}^2 : \|x - c_n\|_2 \leq r_n\}$. The figure below illustrates this.



a) **[1 val]** Given circles $\mathcal{C}$ and $\mathcal{C}_1$, argue that $\mathcal{C}$ encloses (i.e., contains) $\mathcal{C}_1$ if $\|c - c_1\|_2 + r_1 \leq r$.

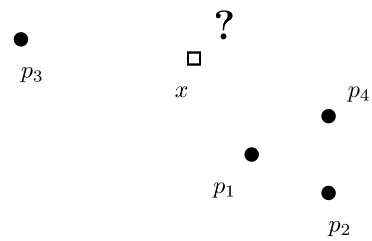
b) **[1 val]** In 3D solution space ($c \in \mathbb{R}^2$, $r \in \mathbb{R}$) sketch the region (set) where $\|c\|_2 \leq r$.

c) **[1.5 val]** Based on the first result, formulate the problem of finding the smallest circle $\mathcal{C}$ that fully contains a given collection $\mathcal{C}_1, \ldots, \mathcal{C}_N$.

d) **[1 val]** Is the above problem convex? If so, prove it, otherwise explain why not.

e) **[1 val]** From the properties of the cost function and feasible set Ω, explain why this problem is always solvable.

P 2. [5 val] We wish to choose the location $x = (x_1,\, x_2) \in \mathbb{R}^2$ for a broadcast station that will serve N communities located at $p_1, \ldots, p_N$. Each $p_n = (p_{n1},\, p_{n2}) \in \mathbb{R}^2$ is a 2D location on the map (see the figure below).



To optimally place the station we consider the following cost functions, which penalize in different ways the selection of a given location x

$$\phi_1(x) = \sum_{n=1}^{N} w_n \|x - p_n\|_2^2, \qquad \phi_2(x) = \sum_{n=1}^{N} w_n e^{\|x-p_n\|_2}, \qquad \phi_3(x) = \sum_{n=1}^{N} w_n \|x - p_n\|_1$$

where w_n is a weighting factor that reflects the importance of community n (e.g., it may be proportional to its population). The goal is to select the location which minimizes the chosen penalization.

a) [2 val] Express the unconstrained minimization problem using cost function ϕ_1 as a least-squares problem. Determine the optimal solution and interpret it.

Note: If you don't know how to handle the weighting factors solve the problem with $w_n = 1$ for a partial grade.

b) [1 val] If all communities are relatively close, except for a single one in a more "remote" location (see the figure), and all weights w_n are similar, which penalization ϕ_1, ϕ_2, ϕ_3 should yield a location for the broadcast station that is closer to the remote community? Why?

c) [2 val] Suppose that the station must be located along a highway which, in the relevant portion of the 2D map, is described by the line $x_2 = 10 - 4x_1$. Write the KKT conditions for the optimization problem using cost ϕ_3 (you don't have to solve the KKT system).

P 3. [3.5 val] The Kullback-Leibler divergence provides a measure of disparity, from the perspective of information theory, between two vectors with positive entries, $u, v \in (\mathbb{R}^+)^n$. It is defined as

$$D_{\mathrm{KL}}(u, v) = \sum_{i=1}^{n} u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i$$

a) [2 val] To analyze D_{KL} we will resort to the relative entropy $f(u) = \sum_{i=1}^{n} u_i \log(u_i)$, which is known to be a convex function with gradient

$$\nabla f(u) = \begin{bmatrix} \log(u_1) \\ \vdots \\ \log(u_n) \end{bmatrix} + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Use it to write D_{KL} as $D_{\mathrm{KL}}(u, v) = f(u) - \big(f(v) + \nabla f(v)^T (u - v)\big)$

b) [1.5 val] Based on the convexity of f, writing D_{KL} as above proves that it is non-negative, as you would expect from a disparity that is akin to a distance metric. Why can you assert that $D_{\mathrm{KL}} \geq 0$?

Optimization and Algorithms 2023/24

Instituto Superior Técnico

Answers to the 2nd exam – January 29, 2024

- Answers are only schematically presented, as they are meant to complement the students' own solutions to the problems

---

Answers to multiple-choice questions

	a	b	c	d	e	f	g	h	i
Q1					✓				
Q2		✓							
Q3			✓						
Q4					✓				

Q 1. [1.5 val] For any given points $x, x_0 \in \mathbb{R}^n$ a twice differentiable convex function f satisfies

$$f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0) + \frac{3}{2}\|x - x_0\|^2$$

Choose the statement below that follows from this.

a) $\nabla f(x_0) \neq 0$

b) $f(x)$ is quadratic

c) Maximum eigenvalue: $\lambda_{\max}(\nabla^2 f(x_0)) \leq 2$

d) $f((1-\alpha)x + \alpha x_0) \geq (1-\alpha)f(x) + \alpha f(x_0) - 3\frac{\alpha(1-\alpha)}{2}\|x - x_0\|^2$ for all $0 \leq \alpha \leq 1$

e) Minimum eigenvalue: $\lambda_{\min}(\nabla^2 f(x_0)) \geq 3$

f) $f(x_0) + \nabla f(x_0)^T(x - x_0)$ is quadratic in x

Note: The expression shows that f is strongly convex with factor $\nu = 3$. Its curvature is thus no smaller that that of a quadratic $x^T Ax$ with $A = 3I$. Equivalently, $\nabla^2 f(x_0) - 3I \succeq 0$ is positive semidefinite, and all its eigenvalues are non-negative. In particular, for the smallest one we have $\lambda_{\min}(\nabla^2 f(x_0) - 3I) = \lambda_{\min}(\nabla^2 f(x_0)) - 3 \geq 0$.

Q 2. [1.5 val] A company has developed three new products and is considering bringing them to market under the following constraints:

- At most two of them will be commercialized.
- The company operates two factories that could produce any of these products. However, to keep costs low the chosen products should be produced at the same factory.

The main problem variables are x_1, x_2, x_3, which denote the weekly production volume for products 1, 2, and 3, respectively.

- The expected profit is $5x_1 + 7x_2 + 3x_3$.
- If produced in the first factory, the required manufacturing time would be $3x_1 + 4x_2 + 2x_3$ hours. A maximum of 30 hours per week could be allocated for this. In the second factory the same production volume would take $4x_1 + 6x_2 + 2x_3$ hours to complete, in no more than 40 hours.
- The market cannot absorb more than 7 units of product 1 per week, so it is pointless to produce beyond this volume. For the same reason, no more than 5 units of product 2 and 9 units of product 3 should be produced per week.

The following formulation is proposed to maximize profit while respecting the stated constraints:

$$\begin{aligned} \underset{\substack{x_i,\, i=1,\ldots,3,\\ y_i,\, i=1,\ldots,4}}{\text{maximize}} \quad & 5x_1 + 7x_2 + 3x_3 \\ \text{subject to} \quad & \boxed{\quad ① \quad}, \\ & y_1 + y_2 + y_3 \leq 2, \\ & x_i \geq 0,\ i = 1, \ldots, 3, \\ & y_i \in \{0,\, 1\},\ i = 1, \ldots, 4 \end{aligned} \tag{1}$$

Choose below the content for box ①.

a) $3x_1 + 4x_2 + 2x_3 \leq 30$
$4x_1 + 6x_2 + 2x_3 \leq 40$
$x_1 \leq 7, \quad x_2 \leq 5, \quad x_3 \leq 9$

b) $3x_1 + 4x_2 + 2x_3 \leq 30 + 1000y_4$
$4x_1 + 6x_2 + 2x_3 \leq 40 + 1000(1 - y4)$
$x_1 \leq 7y_1, \quad x_2 \leq 5y_2, \quad x_3 \leq 9y_3$

c) $3x_1 + 4x_2 + 2x_3 \leq 30 + 1000(1 - y_4)$
$4x_1 + 6x_2 + 2x_3 \leq 40 + 1000y4$
$x_1y_1 \geq 7, \quad x_2y_2 \geq 5, \quad x_3y_3 \geq 9$

d) $3x_1 + 4x_2 + 2x_3 \leq 30y_4$
$4x_1 + 6x_2 + 2x_3 \leq 40y4$
$x_1 \geq 7, \quad x_2 \geq 5, \quad x_3 \geq 9$

e) $3x_1 + 4x_2 + 2x_3 \leq 30$
$7x_1 + 5x_2 + 9x_3 \geq 1$
$y_1 + y_2 + y_3 \leq y_4$

f) $4x_1 + 6x_2 + 2x_3 \leq 40$
$7x_1 + 5x_2 + 9x_3 \geq 70$

Note: Binary variables y_1, y_2, y_3 define which products will be manufactured (0 means no, 1 means yes). Up to two products out of three equates to $y_1 + y_2 + y_3 \leq 2$, whereas conditions such as $x_1 \leq 7y_1$ simultaneously (1) impose a limit on the quantity x_i of a product if it is selected for manufacturing (the corresponding $y_i = 1$), and (2) force x_i to be zero if the product is not selected ($y_i = 0$).

Binary variable y_4 defines where the production will happen (0 means first factory, 1 means second). If $y_4 = 0$ the first inequality reads $3x_1 + 4x_2 + 2x_3 \leq 30$, as expected to define limits on production for the first factory, but the second one will be $4x_1+6x_2+2x_3 \leq 1040$. As long as the x_i stay within the stated production bounds (7, 5, 9 units per week), the latter inequality will always be true — by a large margin —, and therefore conditions for the second (unused) factory will be effectively dropped from the set of constraints, as intended.

Q 3. [1.5 val] Consider the constrained optimization problem

$$\begin{aligned} \underset{x,\,y}{\text{minimize}} \quad & (x-1)^2 + 2(y-2)^2 \\ \text{subject to} \quad & x - 2y = 0 \end{aligned}$$

whose optimal solution is $(x^*, y^*) = (2,\ 1)$. The augmented Lagrangian method, initialized with penalty factor $c_0 = 2$ and Lagrange factor $\lambda_0 = -1$, is used to iteratively solve the problem. Assuming that the method converges very close to the true solution of the original constrained problem within 20 iterations, what should be the final value λ_{20}?

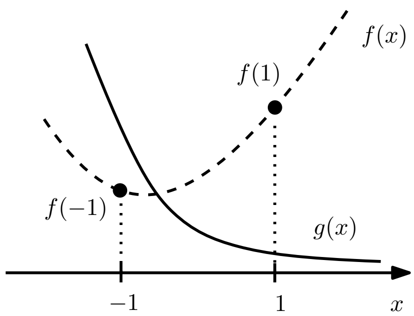
a) -6 **b)** -4 **c)** -2 **d)** 2 **e)** 4 **f)** 6

Note: For cost and constraint functions $f(x, y) = (x-1)^2 + 2(y-2)^2$ and $h(x, y) = x - 2y$, respectively, the KKT conditons assert that at an optimal point $(x^*,\ y^*)$ we must have

$$\nabla f(x^*,\ y^*) + \lambda \nabla h(x^*,\ y^*) = \begin{bmatrix} 2(x^*-1) \\ 4(y^*-2) \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix} + \lambda \begin{bmatrix} 1 \\ -2 \end{bmatrix} 0$$

so the Lagrange multiplier is $\lambda^* = -2$. Upon convergence, the Lagrange factor in the augmented Lagrangian method will approach this, $\lambda_{20} \approx \lambda^*$. The values of c_0 and λ_0 are not important here, as it is simply assumed here that they are such that convergence of the augmented Lagrangian occurs to the close vicinity of the solution to the original constrained problem.

Q 4. [1.5 val] You are given a non-negative function $f(x)$ whose values are unimportant except at $x = \pm 1$, as shown in the figure.



You wish to tune a non-negative function of the form $g(x) = \log(1 + e^{-\gamma x})$, with adjustable parameter γ, so that it becomes as *dissimilar* as possible to f. Similarity between f and g is quantified by the inner product of $(f(-1),\ f(1))$ with $(g(-1),\ g(1))$, that is, $f(-1)g(x)|_{x=-1} + f(1)g(x)|_{x=1}$. Which of the following expressions reflects a first-order stationarity condition for γ to minimize the similarity?

a) $f(-1)\log(1+e^{\gamma}) + f(1)\log(1+e^{-\gamma}) = 0$ **b)** $f(-1)e^{\gamma} = f(1)e^{-\gamma}$

c) $\gamma = 1$ or $\gamma = -1$ **d)** $\left.\frac{dg}{d\gamma}\right|_{x=-1} = \left.\frac{dg}{d\gamma}\right|_{x=1} = 0$

e) $f(-1)\frac{e^{\gamma}}{1+e^{\gamma}} = f(1)\frac{e^{-\gamma}}{1+e^{-\gamma}}$ **f)** $f(-1)(1+e^{-\gamma}) = f(1)(1+e^{\gamma})$

Note: We have the following similarity metric, written as a function of the adjustable parameter γ

$$s(\gamma) \triangleq f(-1)g(x)|_{x=-1} + f(1)g(x)|_{x=1} = f(-1)\log(1+e^{\gamma}) + f(1)\log(1+e^{-\gamma}) \geq 0$$

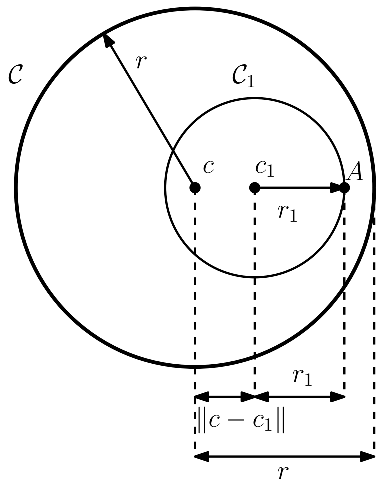
To make this as close to zero as possible we differentiate with respect to γ and set the derivative to zero

$$\frac{ds}{d\gamma} = f(-1)\frac{e^{\gamma}}{1+e^{\gamma}} - f(1)\frac{e^{-\gamma}}{1+e^{-\gamma}} = 0$$

In general, a point that satisfies a first-order stationarity condition like this one is not necessarily a minimum. While the question does not require you to verify this, convince yourself that here it is indeed a minimum.

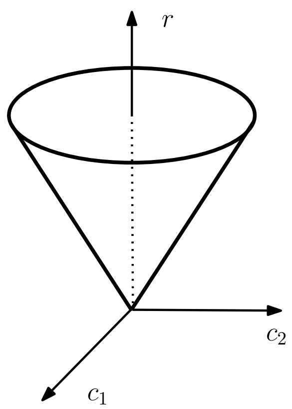
P 1 (Smallest covering circle)**.** We wish to find the smallest circle $\mathcal{C}$ that will cover a given collection of circles $\mathcal{C}_1, \ldots, \mathcal{C}_N$.

a) $\mathcal{C}_1$ is inside $\mathcal{C}$ if its furthest point from the center c, marked A in the figure below, is at a distance that does not exceed the radius r.



The distance from c to A is $\|c - c_1\| + r_1$.

b) The equation $\|c\| \leq r$ defines a cone in 3D solution space ($c \in \mathbb{R}^2$, $r \in \mathbb{R}$). This is commonly known as the Lorentz cone.



c)

$$\begin{aligned} \underset{c,\,r}{\text{minimize}} \quad & r \\ \text{subject to} \quad & \|c - c_n\| + r_n \le r, \quad n = 1, \dots, N, \\ & (r \ge 0) \quad \text{(redundant, given the other constraints)} \end{aligned}$$

d) For the problem above to be convex its cost and inequality constraint functions should be convex. Concatenate optimization variables c and r into a single column vector $x = [c^T\, r]^T$. Then the cost is an affine function $f(x) = \begin{bmatrix} 0 & 1 \end{bmatrix} x$, hence convex. Each constraint function is $g_n(x) = \| \begin{bmatrix} I & 0 \end{bmatrix} x - c_n\| + \begin{bmatrix} 0 & -1 \end{bmatrix} x + r_n$, which is the sum of an affine function $\begin{bmatrix} 0 & -1 \end{bmatrix} x + r_n$ with the composition of a norm with an affine function $\begin{bmatrix} I & 0 \end{bmatrix} x - c_n$, hence convex.

e) An inequality $\|c\| \le r$ defines the Lorentz cone discussed previously. Each inequality constraint $\|c - c_n\| \le r - r_n$ is obviously closely related to this, and defines a similar (closed) cone, except that it is translated so that its vertex is located at $(c_n,\, r_n)$ instead of (0, 0). The feasible set Ω is therefore an intersection of a finite number of translated Lorentz cones. The shape of that intersection is not important here, but along the r dimension Ω will extend from some $r_{\min} \ge 0$ to $+\infty$. Also, Ω is closed as an intersection of N closed sets (cones).

The inequality $\|c\| \le r$ implies that, if $\|c\| \to +\infty$, then $r \to +\infty$ as well. As an intersection of cones, any point $x \in \Omega$ such that $\|x\| \to +\infty$ must also satisfy $r \to +\infty$. This means that $f(x) = \begin{bmatrix} 0 & 1 \end{bmatrix} x$ is coercive in Ω. A constrained optimization problem is solvable if it has (1) a continuous and coercive cost function, and (2) a closed feasible set Ω.

P 2 (Choosing the location of a broadcast station)**.** Choose the location x for a broadcast station that will serve N communities, located at $p_1, \dots, p_N$, under different optimality criteria.

a) Minimization of $\phi_1(x)$ is equivalent to

$$\underset{x}{\text{minimize}} \quad \|Ax - b\|_2^2$$

where

$$A = \begin{bmatrix} \sqrt{w_1} I_2 \\ \vdots \\ \sqrt{w_N} I_2 \end{bmatrix}, \qquad b = \begin{bmatrix} \sqrt{w_1} p_1 \\ \vdots \\ \sqrt{w_N} p_N \end{bmatrix}$$

The $2N \times 2$ matrix A has clearly full column rank, hence the optimal solution is given by the usual pseudoinverse expression

$$x = (A^T A)^{-1} A^T b = \Big(I_2 \sum_{n=1}^{N} w_n \Big)^{-1} \sum_{n=1}^{N} w_n p_n = \frac{\sum_{n=1}^{N} w_n p_n}{\sum_{n=1}^{N} w_n}$$

which is just the weighted centroid of the communities.

b) If x were far from a given "remote" community, say, p_r, then $\|x - p_r\|$ would be significantly larger than the remaining $\|x - p_n\|$. Due to its exponential mapping of distances to penalizations, out of the three given cost functions ϕ_2 should be the one where the discrepancy in norms translates into the largest discrepancy in terms of penalizations. Then, that choice for x might be suboptimal, because a lower penalization could be obtained by incrementally moving the broadcast station closer to the remote community; this would strongly decrease the large penalization with only a moderate increase in the remaining ones. For ϕ_1, ϕ_3 the reduction in overall penalization from such a move in x would be smaller, at best.

c)

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \phi_3(x) \\ \text{subject to} & h(x) = 0 \end{array}$$

with

$$\phi_3(x) = \sum_{n=1}^{N} w_n(|x_1 - p_{n1}| + |x_2 - p_{n2}|), \qquad\qquad h(x) = 4x_1 + x_2 - 10$$

KKT:

$$\begin{cases} \nabla\phi_3(x) + \lambda\nabla h(x) = 0 \\ h(x) = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{n=1}^{N} w_n \frac{d}{dx_1}|x_1 - p_{n1}| + 4\lambda = 0 \\ \sum_{n=1}^{N} w_n \frac{d}{dx_2}|x_2 - p_{n2}| + \lambda = 0 \\ 4x_1 + x_2 = 10 \end{cases}$$

where

$$\frac{d}{dx_1}|x_1 - p_{n1}| = \frac{x_1 - p_{n1}}{|x_1 - p_{n1}|} = \operatorname{sign}(x_1 - p_{n1})$$

and similarly for x_2.

P 3 (Non-negativity of Kullback-Leibler divergence)**.**

$$D_{\mathrm{KL}}(u, v) = \sum_{i=1}^{n} u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i$$

a) Start with the given alternative expression

$$\begin{aligned} D_{\mathrm{KL}}(u, v) &= f(u) - \big(f(v) + \nabla f(v)^T(u - v)\big) \\ &= \sum_i u_i \log(u_i) - v_i \log(v_i) - (\log(v_i) + 1)(u_i - v_i) \\ &= \sum_i u_i \log(u_i) - u_i \log(v_i) - u_i + v_i \\ &= \sum_i u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i \end{aligned}$$

b) Since f is convex (given) its graph lies above the linearization around any point

$$f(u) \geq f(u_0) + \nabla f(u_0)^T(u - u_0) \quad \Leftrightarrow \quad f(u) - \big(f(u_0) + \nabla f(u_0)^T(u - u_0)\big) \geq 0$$

The result $D_{\mathrm{KL}}(u, v) \geq 0$ follows by taking $u_0 = v$.