Delft University of Technology
Faculty of Electrical Engineering, Mathematics, and Computer Science
Signal Processing Systems Group

EE 4530 APPLIED CONVEX OPTIMIZATION

14 April 2023, 9:00–12:00

Open book: copies of the book and the course slides allowed, as well as a one page cheat sheet. No other tools except a basic pocket calculator permitted.

Answer in English. Make clear in your answer how you reach the final result; the road to the answer is very important. Write your name and student number on each sheet and use one sheet per question.

Hint: Avoid losing too much time on detailed calculations, write down the general approach first.

Question 1 (10 points)

For each of the following sets or functions, explain if it is convex, concave, or neither convex nor concave. Prove it (by using the definition or some of the basic rules we have encountered in the course).

(a) The set

$$D = \{(y, b) \mid (y - c)^T A^{-1}(y - c) \leq b\} \tag{1}$$

with fixed c and A and where $A \in \mathbf{S}^n_{++}$. Make a sketch of this set!

(b) The function

$$f(y) = \sqrt{\sum_{i=1}^{k} (\,|a + b_i y_i|)^{\,2}} \tag{2}$$

(c) A set of length N sequences defining a bounded discrete cosine function:

$$S = \{x \in \mathbf{R}^N \mid c(k) = \sum_{n=1}^{N} x_n \cos(\frac{\pi}{N}(n + \frac{1}{2})k) < M \text{ for } k = 1, ..., N\} \tag{3}$$

(d) The following set in $\mathbf{R}^2$:

$$C = \{y + Ax \mid ||x||_2 \leq 1\} \tag{4}$$

with $A = \begin{bmatrix} 1/t & 2/t \\ 2/t & 4/t \end{bmatrix}$ and an arbitrary fixed y and fixed $t > 0$.

(e) The function $f(x) = \min\{f_1(x), f_2(x)\}$ on $\textbf{dom}f = \mathbf{R}_{++}$, where

$$f_1(x) = -3x_1^2 + 4x_1x_2 - 5x_2^2$$
$$f_2(x) = \log(x_1) + \log(x_2)$$

(f) The set of Hankel matrices of dimension n. Recall that the Hankel matrix created from a sequence $a := [a_1, ...a_n]$ is defined as :

$$H(a) := \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ a_2 & a_3 & \cdots & a_{n+1} \\ \vdots & \vdots & \cdots & \vdots \\ a_n & a_{n+1} & \cdots & a_{2n-1} \end{bmatrix}$$

(g) The function

$$f(x) = ||x||_{1/k} + ||x||_k \text{ with } \textbf{dom}f = \mathbf{R}^n$$

for an arbitrary positive integer k.

Solution

(a) It is an elliptic cone (i.e. a cone with elliptic cross-section), therefore, convex.

(b) This is the l2-norm of an affine function of y, hence convex.

(c) For a fixed k, $\sum_{n=1}^{N} x_n \cos(\frac{\pi}{N}(n + \frac{1}{2})k)$ is an affine function, therefore, $c(k) < M$ is a halfspace. The set S, therefore, is the intersection of N halfspaces and hence convex.

(d) It is a degenerate ellipsoid (A is positive semi-definite, but singular), which is convex.

(e) The function $f_1(x)$ is a quadratic function with a negative definite Hessian, hence concave. $f_2(x)$ is concave as well. Therefore, it is a pointwise minimum of concave functions, hence concave.

(f) It is easy to prove using the definition. Given a Hankel matrix based on two sequences a and b. The convex combination of their entries, i.e. $c_i = \theta a_i + (1 - \theta)b_i$ will preserve the Hankel structure.

(g) Neither convex or concave. The p-norm is only convex for $p \geq 1$, so both functions in the sum cannot be convex at the same time.

Question 2 (10 points)

Consider the following optimization problem in the domain $\textbf{dom}f = \{x \mid x \in \mathbf{R}^2,\, x_2 > 0\}$:

$$\min_x f(x) = \frac{3x_1^2}{2x_2}$$

(a) What is the optimum value and where is it obtained?

(b) Sketch the contour plot (i.e sublevel sets) of the function! Indicate the objective function value at at least 2 contour lines.

Let us solve this optimization problem with a steepest descent method with the quadratic norm $||x||_P = (x^T P x)^{1/2}$!

(c) Calculate the steepest descent step for $P = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$

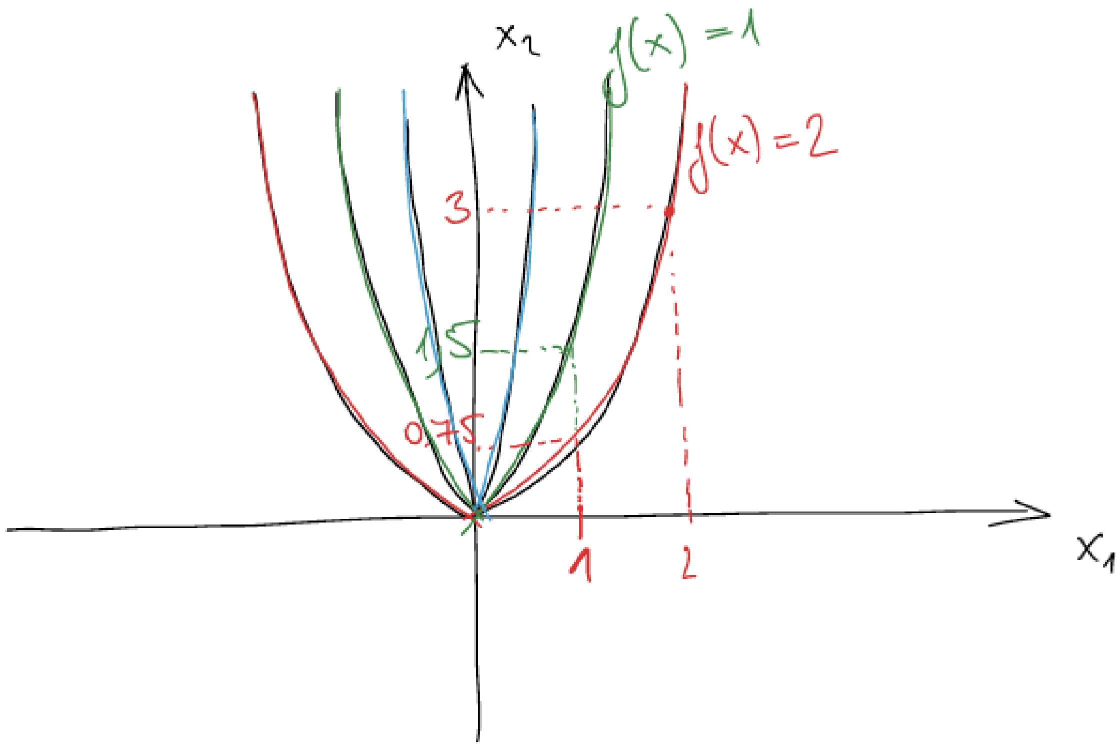
Let us assume that we start at $x^{(0)} = \begin{bmatrix} 1/4 & 3 \end{bmatrix}$.

(d) Is it a good idea to use this steepest descent step at this particular point? Why? If not, what would be a better choice? *Hint:* Inspect your sketch to answer.

(e) For which choice of P would the steepest descent step be equivalent to the Newton step? Calculate the Newton step at $x^{(0)}$!

Solution

(a) The optimum value is 0, obtained at any $x_1 = 0$.



(b)

(c)

$$
\begin{aligned}
\nabla_{sd} &= -P^{-1}\nabla f(x) \\
\frac{df}{dx_1} &= \frac{3x_1}{x2} \\
\frac{df}{dx_2} &= \frac{-3x_1^2}{2x_2^2} \\
P^{-1} &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \text{ therefore,} \\
\nabla_{sd} &= \begin{bmatrix} \frac{6x_1}{x_2} & \frac{-3x_1^2}{x_2^2} \end{bmatrix}
\end{aligned}
$$

(d) As a rule of thumb, if the shape of the norm ball approximates well the sublevel sets, we can expect a good convergence rate. The chosen P is a circle of radius $1/2$. Inspecting the contour plot, an ellipse with a long vertical axis would be a better choice at this point.

(e) The Newton step is the steepest descent direction for the local Hessian norm. Therefore,

$$
P = \nabla^2 f(x^{(0)}) = \begin{bmatrix} \frac{6}{x_2} & \frac{6}{-x_2^2} \\ \frac{-2x_1}{x_2} & \frac{x_1^2}{2x_2^2} \end{bmatrix}
$$

Question 3 (10 points)

Let us consider the problem of minimizing the difference between two quadratic functions under a lower and upper bound on the norm of the unknown variable. In other words, consider the problem

$$\min_{x} \ f_1(x) - f_2(x), \quad \text{s.t.} \ \ L \leq \|x\| \leq U, \tag{5}$$

where $f_i(x) = x^T A_i x - 2b_i^T x + c_i$, with $A_i \in \mathbb{R}^{n \times n}$ a symmetric positive semi-definite matrix, $b_i \in \mathbb{R}^n$ and $c_i \in \mathbb{R}$.

(a) Is problem (5) convex? Explain also why or why not.

Using the eigenvalue decomposition of $\bar{A} = A_1 - A_2$, i.e., $\bar{A} = U \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n) U^T$, where we assume $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, and making the change of variables $y = U^T x$, we can transform the original problem (5) into

$$\min_{y} \sum_{i=1}^{n} (\sigma_i y_i^2 - 2\bar{b}_i y_i) + \bar{c}, \quad \text{s.t.} \ \ L \leq \|y\| \leq U, \tag{6}$$

where $\bar{b} = U^T(b_1 - b_2)$ and $\bar{c} = c_1 - c_2$.

(b) Prove that in the optimal point y^* of (6), we have that $\bar{b}_i y_i^* \geq 0$ for $i = 1, 2, \ldots, n$. *Hint: Since $y^* = [y_1^*, y_2^*, \ldots, y_n^*]^T$ is feasible, also $y_{\text{flip}}^* = [y_1^*, \ldots, y_{k-1}^*, -y_k^*, y_{k+1}^*, \ldots, y_n^*]^T$ is feasible, but it is not necessarily optimal and hence its cost function is larger. Use this fact to construct the proof.*

Using (b), we can make another change of variables

$$y_i = \text{sign}(\bar{b}_i)\sqrt{z_i}, \ \ i = 1, 2, \ldots, n,$$

where $z_i \geq 0$, $i = 1, 2, \ldots, n$. This allows us to transform problem (6) into

$$\min_{z \succeq 0} \ \sum_{i=1}^{n} (\sigma_i z_i - 2|\bar{b}_i|\sqrt{z_i}) + \bar{c} \ \ \text{s.t.} \ \ L^2 \leq \sum_{i=1}^{n} z_i \leq U^2. \tag{7}$$

Note that $z \succeq 0$ represents the domain of the cost function.

(c) Is problem (7) convex? Explain also why or why not.

(d) Write down the Lagrangian of (7).

(e) Derive the dual function and formulate the dual problem.

Solution

(a) It is not convex since it is a difference of two convex functions. Also the constraint set is not a convex set.

(b) Plugging y^* and y^*_{flip} into the cost, we obtain

$$\sum_{i=1}^{n}(\sigma_i (y_i^*)^2 - 2\bar{b}_i y_i^*) \leq \sum_{i=1, i\neq k}^{n}(\sigma_i (y_i^*)^2 - 2\bar{b}_i y_i^*) + \sigma_k(-y_k^*)^2 + 2\bar{b}_k y_k^*,$$

which results into

$$-\bar{b}_k y_k^* \leq \bar{b}_k y_k^* \;\Rightarrow \bar{b}_k y_k^* \geq 0.$$

Since we can repeat this for every k, the result holds.

(c) This problem is convex. The domain of the cost function is a convex set. The cost function is convex in its domain, since it consists of a linear part, which is always convex, and a weighted sum of negative square roots, which is also convex in the domain of positive variables. The constraint functions are all linear, so that makes the constraints also convex.

(d) Using λ_1 and λ_2 as the dual variables for the lower and upper bound, the Lagrangian can be written as

$$\begin{aligned} L(z,\lambda) &= \sum_{i=1}^{n}(\sigma_i z_i - 2|\bar{b}_i|\sqrt{z_i}) + \bar{c} + \lambda_1\left(L^2 - \sum_{i=1}^{n} z_i\right) + \lambda_2\left(\sum_{i=1}^{n} z_i - U^2\right) \\ &= \sum_{i=1}^{n}\left((\sigma_i - \lambda_1 + \lambda_2)z_i - 2|\bar{b}_i|\sqrt{z_i}\right) + \lambda_1 L^2 - \lambda_2 U^2 + \bar{c} \end{aligned}$$

(e) Computing the derivative of the Lagrangian with respect to z_i and equating to zero, we obtain

$$z_i = \frac{\bar{b}_i^2}{(\sigma_i - \lambda_1 + \lambda_2)^2},$$

subject to the condition $\sigma_i - \lambda_1 + \lambda_2 \geq 0$. Note that all these conditions for $i = 1, 2, \ldots, n$ together can equivalently be written as the single condition $\sigma_n - \lambda_1 + \lambda_2 \geq 0$. The dual function is then given by

$$g(\lambda) = \inf_{z \succeq 0} L(z,\lambda) = \begin{cases} -\displaystyle\sum_{i=1}^{n}\frac{\bar{b}_i^2}{\sigma_i - \lambda_1 + \lambda_2} + \lambda_1 L^2 - \lambda_2 U^2 + \bar{c} & \text{if } \sigma_n - \lambda_1 + \lambda_2 \geq 0 \\ -\infty & \text{otherwise} \end{cases}.$$

The dual problem then becomes

$$\max_{\lambda \succeq 0} g(\lambda),$$

leading to

$$\max_{\lambda} \; -\sum_{i=1}^{n}\frac{\bar{b}_i^2}{\sigma_i - \lambda_1 + \lambda_2} + \lambda_1 L^2 - \lambda_2 U^2 + \bar{c} \;\; \text{s.t.} \;\; \sigma_n - \lambda_1 + \lambda_2 \geq 0, \;\lambda \succeq 0.$$

Question 4 (10 points)

Consider the following optimization problem:

$$\max_{x \prec b} \sum_{i=1}^{n} c_i^2 x_i - \sum_{i=1}^{n} \frac{a_i^2}{b_i - x_i}. \tag{8}$$

(a) Is this problem convex, concave, or neither one of the two? Use the second-order derivative to prove this. Make a sketch of the cost function for $n = 1$ to corroborate your answer.

(b) Is the maximum obtained at an x for which x_i lies *strictly* between $-\infty$ and b_i? To check this, derive the limits of the cost for $x_i \to -\infty$ and $x_i \to b_i$ and draw your conclusion from these limits.

(c) Write down the update rule of a first-order method to solve (8) where the step size is determined using line search. What measure do you take to handle the constraint $x \prec b$?

Let us now also add the additional constraint $l \prec x$ to problem (8), where we implicitly assume that $l \prec b$.

(d) Use the logarithmic barrier function to handle this constraint. How can the new cost function then be written? Make a sketch of the new cost for $n = 1$.

(e) For this new problem, write again the update rule of a first-order method where the step size is determined using line search. Make sure $l \prec x \prec b$ in every step of the algorithm.

Solution

(a) This is a concave problem since the domain is convex and the Hessian is negative definite. More specifically, if the cost is $f(x)$, the first-order partial derivative is given by

$$\frac{\partial f(x)}{\partial x_i} = c_i^2 - \frac{a_i^2}{(b_i - x_i)^2},$$

and the second-order partial derivative by

$$\frac{\partial^2 f(x)}{\partial x_i^2} = -\frac{2a_i^2}{(b_i - x_i)^3}.$$

Note that the other second-order partial derivatives are zero, and the Hessian is diagonal. Since all diagonal elements are negative in the domain $x \prec b$, the Hessian is negative definite.

(b) It is easy to show that both limits are $-\infty$ and thus the maximum must be obtained at an x for which x_i lies *strictly* between $-\infty$ and b_i.

(c) Using the expression of the derivative derived in (a), we can write the update rule per entry as

$$x_i^{(k+1)} = x_i^{(k)} + s_k \left(c_i^2 - \frac{a_i^2}{(b_i - x_i^{(k)})^2} \right),$$

where s_k is assumed positive since we need to take a step in the direction of the gradient to maximize the cost. The step size s_k is determined using

$$\underset{s_k > 0}{\text{argmax}} \; f(x^{(k+1)}).$$

Because of the fact that the cost becomes $-\infty$ for $x_i \to b_i$ and the fact that we do a line search for s_k, we don't need to take any specific action to guarantee that $x_i^{(k+1)} < b_i$.

(d) Since we are considering a maximization problem, the new cost becomes

$$\max_{x \prec b} \sum_{i=1}^{n} c_i^2 x_i - \sum_{i=1}^{n} \frac{a_i^2}{b_i - x_i} + \frac{1}{t} \sum_{i=1}^{n} \log(x_i - l_i).$$

(e) The update rule per entry now becomes

$$x_i^{(k+1)} = x_i^{(k)} + s_k \left(c_i^2 - \frac{a_i^2}{(b_i - x_i^{(k)})^2} + \frac{1}{t(x_i^{(k)} - l_i)} \right).$$

The positive step size s_k is again determined using

$$\underset{s_k > 0}{\text{argmax}} \; f(x^{(k+1)}).$$

As before, due to the shape of the cost, and the line search for the step size, we don't need to take any special action to guarantee that $l \prec x^{(k+1)} \prec b$.