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

EE 4530 APPLIED CONVEX OPTIMIZATION

8 April 2025, 09:00 - 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.

Question 1 (10 points)

For each of the following sets or functions, prove its convexity, concavity, or both/neither, by using the definition or one of the properties of convex sets/functions.

Exercises

(a) The function $f(x) = \|x\|_\infty$, for $\textbf{dom}\ f = \mathbf{R}^n$.
(*Show a proof, simply stating that norms are convex is not enough to get the point.*) **(1 point)**

(b) The function $f(x) = \|x\|_0$, for $\textbf{dom}\ f = \mathbf{R}^n$. **(2 points)**

(c) The function $f(x) = -(\log(x))^2$, with $\textbf{dom}\ f = \mathbf{R}_{++}$. **(2 points)**

(d) The set $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{S}_2$, where $\mathcal{S}_1 = \{ax_1 + bx_2 \in \mathbf{R}^n \mid a, b \geq 0\}$ and $\mathcal{S}_2 = \{x \in \mathbf{R}^n \mid c^T x \leq b\}$ for some non-zero c. **(1 point)**

(e) The set $\mathcal{S} = \{g(c) \mid c \in \mathcal{C}\}$, where $g(x) = d^T x + b$ and $\mathcal{C}$ is some convex set. **(2 points)**

(f) Show that the *dual cone*, e.g., $K^* = \{y \mid x^T y \geq 0, \text{ for all } x \in K\}$ of the cone K is always convex. Do not forget to explain why the convexity of the dual cone does not depend on the convexity of the original cone K. **(2 points)**

Solution

(a) Note that $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$, which is a maximum over convex functions, and thus convex. It would be neat to also show that $|x_i|$ is a convex function in x_i.

(b) The function is neither convex nor concave. Two simple counter examples suffices. For example $x_1 = [1, 0, 1]^T$ and $x_2 = [1, 0, 0]^T$ to disprove convexity, and $x_1 = [-1, 0, 1]^T$ and $x_2 = [1, 0, 0]^T$ to disprove concavity.

(c) The second derivative is given by $f''(x) = \frac{2(\log(x)-1)}{x^2}$, for which $f''(x) \leq 0$ if $x \leq e$ and $f''(x) \geq 0$ if $x \geq e$. The function is concave for $x \in (0, e]$ and convex for $x \in [e, \rightarrow)$. Therefore, for **dom** f, the function is neither convex nor concave.

(d) The set is convex, since it is an *intersection* of two convex sets (a cone and a halfspace).

(e) Note that $g(x)$ is an affine function w.r.t x. The image of a convex set ($\mathcal{C}$) under an affine function ($g(x)$) is/remains convex.

(f) Note that the dual cone is the intersection of a (infinite) set of homogeneous halfspaces (i.e. of the form $a^T y$). Hence it is a closed convex cone.

$$K^* = \{y \mid x^T y \geq 0, \text{ for all } x \in K\} = \bigcap_{x \in K} \{y \mid x^T y \geq 0\} \tag{1}$$

Observe that a halfspace defined by $\{y \mid x^T y \geq 0\}$ is convex for any selection of x. Therefore, K^* is convex even if K itself is not.

Question 2 (9 points)

Let us consider the following quadratic minimization problem:

$$\min_x \frac{1}{2}x^T Ax + b^T x,$$

with $A = \begin{bmatrix} p & q \\ r & p \end{bmatrix}$ and $b = \begin{bmatrix} s \\ q \end{bmatrix}$.

(a) For which values of p, q, r, s is the above problem convex?

(1 points)

(b) For which values of p, q, r, s is the optimal point $x^* = 0$?
Hint: the inverse of a 2×2 *matrix* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *is* $A^{-1} = \frac{1}{det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

(2 points)

Let us use a steepest descent algorithm to find the minimum. Let us consider the Mahalanobis norm for steepest descent, i.e. $||v||_\Sigma = \sqrt{v^T \Sigma^{-1} v}$, where Σ is a positive definite covariance matrix.

(c) What is the steepest descent step for this norm? **(1 point)**

(d) Compute one iteration of the steepest descent step with exact line search, starting from $x^0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and assuming $\Delta x_{sd} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, $p = 2$, $q = -1$ $r = 1$ and $s = -3$ **(2 points)**

(e) Compare the convergence of steepest descent with normalized steepest descent (still considering exact line search). Which algorithm converges first? **(1 point)**

(f) Let us now consider Newton's method with exact line saerch to find the optimum. Show that for any A and b (given that the problem is convex and twice differentiable) Newton's method converges in one iteration. **(2 points)**

Solution

(a) It is convex if A is positive semi-definite. That is the case if $p \geq 0$ and $p^2 \geq rq$. s is arbitrary.

(b) The minimum is reached when $Ax^* = -b$, that is, $x^* = -A^{-1}b$.
$A^{-1} = \begin{bmatrix} p & q \\ r & p \end{bmatrix}^{-1} = \frac{1}{p^2 - rq} \begin{bmatrix} p & -q \\ -r & p \end{bmatrix}$. Then,
$x^* = -A^{-1}b = \frac{1}{p^2 - rq} \begin{bmatrix} p & -q \\ -r & p \end{bmatrix} \begin{bmatrix} s \\ q \end{bmatrix} = \frac{1}{p^2 - rq} \begin{bmatrix} ps - q^2 \\ pq - rs \end{bmatrix}$.
This vector equals $\mathbf{0}$ if:

- $ps = q^2$ and $pq = rs$, or
- $s = 0$ and $q = 0$.

(c) This is a special type of quadratic norm, with $P = \Sigma^{-1}$. We have seen that for a quadratic norm $||x||_P = \sqrt{x^T P^{-1} x}$ the steepest descent step equals $-P^{-1}\nabla f(x)$. Therefore, in this case, $\Delta x_{sd} = -\Sigma \nabla f(x)$.

(d)

$$
\begin{aligned}
x^{(1)} &= x^{(0)} + t\Delta x_{sd} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\
f(x^{(1)}) &= \frac{1}{2} \begin{bmatrix} 1-t & 2t-1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1-t \\ 2t-1 \end{bmatrix} + \begin{bmatrix} -3 & -1 \end{bmatrix} \begin{bmatrix} 1-t \\ 2t-1 \end{bmatrix} = \\
&= 5t^2 - 5t \\
\frac{df}{dt} &= 10t - 5 = 0 \\
t &= 0.5 \\
x^{(1)} &= \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}
\end{aligned}
$$

(e) Normalized steepest descent step is just a scaled version of steepest descent. Therefore, exact line search will find the same optimal point along the same vector in each iteration. There is no difference in convergence.

(f) The Newton step $\nabla^2 f(x)^{-1} \nabla f(x) = -A^{-1}(Ax + b) = -x - A^{-1}b$.
The first iteration is $x^{(1)} = x^{(0)} + t(-x^0 - A^{-1}b)$.
We have seen that the optimal point $x^* = -A^{-1}b$.
So, for $t = 1$ the result of the first iteration is indeed $x^1 = x^0 - x^0 + x^* = x^*$.

Question 3 (11 points)

Consider the following quadratically constrained quadratic program (QCQP)

$$\begin{aligned} \min_{x_1,x_2} \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & (x_1-1)^2 + (x_2-1)^2 \leq 1 \\ & (x_1-1)^2 + (x_2+1)^2 \leq 1 \end{aligned}$$

with variable $x = [x_1, x_2]^T \in \mathbf{R}^2$.

(a) Sketch the feasible set of the problem and level sets of the objective. From this sketch, find the optimal point x^* and the optimal value p^*. **(2 points)**

(b) Give the Lagrangian function as well as the KKT conditions. **(3 points)**

(c) From these KKT conditions, do there exist Lagrange multipliers that prove the x^* from (a) is optimal? Explain why or why not. **(2 points)**

(d) Derive the Lagrange dual problem. Is this dual problem convex? Explain why or why not? **(2 points)**

(e) Solve the Lagrange dual problem. Is the dual optimum d^* attained? Does strong duality hold? Explain why or why not. **(2 points)**

Solution

(a) The level sets are circles around zero, and the two constraint sets are discs that touch at the point $[1,0]^T$. So there is only one feasible point which is then also the solution of the problem. As a result $x^* = [1,0]^T$ and $p^* = 1$.

(b) The Lagrangian function is given by

$$\begin{aligned} & L(x_1, x_2, \lambda_1, \lambda_2) \\ & = x_1^2 + x_2^2 + \lambda_1[(x_1-1)^2 + (x_2-1)^2 - 1] + \lambda_2[(x_1-1)^2 + (x_2+1)^2 - 1] \\ & = (1+\lambda_1+\lambda_2)x_1^2 + (1+\lambda_1+\lambda_2)x_2^2 - 2(\lambda_1+\lambda_2)x_1 - 2(\lambda_1-\lambda_2)x_2 + \lambda_1 + \lambda_2. \end{aligned}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. The KKT conditions are

$$\begin{aligned} & (x_1-1)^2 + (x_2-1)^2 \leq 1, (x_1-1)^2 + (x_2+1)^2 \leq 1, \\ & \lambda_1 \geq 0, \lambda_2 \geq 0, \\ & 2x_1 + 2\lambda_1(x_1-1) + 2\lambda_2(x_1-1) = 0, \\ & 2x_2 + 2\lambda_1(x_2-1) + 2\lambda_2(x_2+1) = 0, \\ & \lambda_1[(x_1-1)^2 + (x_2-1)^2 - 1] = 0, \\ & \lambda_2[(x_1-1)^2 + (x_2+1)^2 - 1] = 0. \end{aligned} \tag{2}$$

(c) At $x^* = [1,0]^T$, (2) does not have a solution. So there do not exist Lagrange multipliers to show that $x^* = [1,0]^T$ is optimal. This is due to the fact that there is no strictly feasible point in the constraint set and hence Slater's condition does not hold.

(d) The Lagrange dual function is given by

$$g(\lambda_1, \lambda_2) = \inf_{x_1, x_2} L(x_1, x_2, \lambda_1, \lambda_2),$$

where L is derived in (b). If $1 + \lambda_1 + \lambda_2 < 0$, then L is clearly unbounded below in x_1 and x_2. When, $1 + \lambda_1 + \lambda_2 \geq 0$, we can find the solutions for x_1 and x_2 by setting the derivative of L with respect to x_1 and x_2 to zero. This leads to

$$\begin{aligned} x_1 &= \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}, \\ x_2 &= \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2}. \end{aligned}$$

Consequently, we have

$$g(\lambda_1, \lambda_2) = \begin{cases} -\dfrac{(\lambda_1 + \lambda_2)^2 + (\lambda_1 - \lambda_2)^2}{1 + \lambda_1 + \lambda_2} + \lambda_1 + \lambda_2, & \text{if } 1 + \lambda_1 + \lambda_2 \geq 0, \\ -\infty, & \text{otherwise}, \end{cases}$$

The Lagrange dual problem is finally given by

$$\begin{aligned} \max_{\lambda_1, \lambda_2} \quad & \frac{\lambda_1 + \lambda_2 - (\lambda_1 - \lambda_2)^2}{1 + \lambda_1 + \lambda_2} \\ \text{s.t.} \quad & \lambda_1 \geq 0, \; \lambda_2 \geq 0. \end{aligned}$$

This problem is convex since the dual problem is always convex.

(e) Since g is symmetric, the optimum (if it exists) occurs at $\lambda_1 = \lambda_2$. Hence the maximum is the same as the maximum of the function

$$g(\lambda) = \frac{2\lambda}{1 + 2\lambda}.$$

Hence, the maximum d^* tends to 1 but will never be attained. So there is no strong duality.

Question 4 (10 points)

Consider a linear regression problem where you might have some outliers. In that case, instead of a least squares problem, a robust least squares problem is useful. Such a problem is given by

$$\min_x \sum_{i=1}^{m} \phi(a_i^T x - b_i), \tag{3}$$

where the loss function $\phi : \mathbf{R} \to \mathbf{R}$ is defined as

$$\phi(x) = \min_{y \in \mathbf{R}} |y| + \frac{1}{2}(x - y)^2. \tag{4}$$

This loss function is also known as the Huber penalty. The known parameters are the matrix $A \in \mathbf{R}^{m \times n}$ (with rows a_i^T) and the vector $b \in \mathbf{R}^m$ (with entries b_i).

(a) Prove that the above problem is convex. **(2 points)**

(b) Rewrite the above problem as an ℓ_1-norm regularized least squares problem by plugging (4) into (3). For this, you will need to introduce one new auxiliary variable y_i (replacing y in (4)) for every term in the sum of (3), leading to an additional optimization variable $y = [y_1, \ldots, y_m]^T \in \mathbf{R}^m$. **(3 points)**

(c) Can you now explain why this problem is robust to outliers compared to the classical least squares problem? Which variable represents the outliers? **(1 point)**

(d) Now introduce an auxiliary variable $t \in \mathbf{R}^m$ to rewrite the ℓ_1-norm term in the objective function obtained in (b) as a linear function in t using two additional vector inequalities based on y and t. What type of convex problem is this? **(2 points)**

(e) Can you think of another way to robustify the linear regression problem to outliers? Please write down the suggested optimization problem. Note that there is no single answer to this question. **(2 points)**

Solution

(a) First of all we need to prove that $\phi(x)$ is a convex function. For that we only need to prove that $f(x, y) = |y| + \frac{1}{2}(x - y)^2$ is convex in x and y since the minimization is over $\mathbf{R}$ which is convex. We know that $|y|$ is convex and we can further derive that the Hessian of $\frac{1}{2}(x - y)^2$ is

$$\mathbf{H} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \succeq \mathbf{0}.$$

Hence, $\phi(x)$ is convex. Further, the convex function of an affine function remains convex and a sum of convex functions also remains convex. Hence, the problem is convex.

(b) Introducing the variable $y \in \mathbf{R}^m$, we can rewrite the optimization problem as

$$\min_x \sum_{i=1}^{m} \min_{y_i} \left(|y_i| + \frac{1}{2}((a_i^T x - b_i) - y_i)^2\right),$$

or as

$$\min_{x,y} \quad \sum_{i=1}^{m} |y_i| + \frac{1}{2}((a_i^T x - b_i) - y_i)^2 = \|y\|_1 + \frac{1}{2}\|(Ax - b) - y\|^2.$$

(c) It is clear that this problem tries to fit the data b to the model Ax in a least squares sense up to a residual y which is forced to be sparse through the ℓ_1 norm term.

(d) The problem in (b) is equivalent to the following problem

$$\begin{aligned} \min_{x,t,y} \quad & \mathbf{1}^T t + \frac{1}{2}\|(Ax - b) - y\|^2 \\ \text{s.t.} \quad & -t \leq y \leq t \end{aligned}$$

Here t basically represents the modulus of the elements of y, i.e., $|y|$. This is a quadratic program (QP).

(e) A simple way to do this is to force the error of the regression problem to be sparse. This leads to

$$\min_{x} \quad \sum_{i=1}^{m} |a_i^T x - b_i| = \|Ax - b\|_1.$$