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

EE 4530 APPLIED CONVEX OPTIMIZATION

12 April 2024, 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 (8.5 points)

For each of the following sets or functions, prove its convexity, concavity, or both.
The following hint might turn useful:

1. *if* $a, b \geq 0$ *and* $\theta \in [0,1]$ *then* $a^{\theta} b^{(1-\theta)} \leq \theta a + (1-\theta) b$

Exercises

(a) The function $f(x) = e^x$. Show it only by showing the convexity or non convexity of its epigraph. **(2 points)**

(b) The set $\mathcal{S} = \{(x,t) \in \mathbb{R}^2 | t \leq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, t \geq 0\}$ for two constant parameters μ, σ.
Hint: you can graphically show the convexity or non-convexity of the set, by explicitly sketching the set and coloring its interior. **(2 points)**

(c) $f(x) = \inf_{y \in [1,2]} yax^2$ with $\mathrm{dom} f \in \mathbb{R}$ and $a < 0$. **(1 point)**

(d) The set $\mathcal{S} = \{e^{j\theta} || e^{j\theta}| = 1\}$ for $\theta \in [0, 2\pi]$, where j is the imaginary unit. Draw a sketch of the set. **(1 point)**

(e) The set $\mathcal{S} = \{\theta | \sin\theta \leq 1\}$ for $\theta \in [0, 2\pi]$. Draw a sketch of the set. **(1 point)**

(f) The function $f(x) = x^\top A x$ with $A = \sum_{i=1}^{N} a_i a_i^\top$ **(1.5 points)**

Solution

(a) 2p It is convex. We can show this by showing that the epigraph of f is a convex set:

$$\mathrm{epi} f = \{(x,t) \in \mathbb{R}^{n+1} | e^x \leq t\}$$

Consider two distinct points $\bar{x}_1 = [x_1, t_1]^\top$ and $\bar{x}_2 = [x_2, t_2]^\top$, both belonging to epi$f$. We need to show that $\forall\ \theta \in [0, 1]$ the point $\bar{x} := \theta\bar{x}_1 + (1-\theta)\bar{x}_2$ belongs to epi$f$. Expanding this expression we have:

$$\bar{x} = \theta\bar{x}_1 + (1-\theta)\bar{x}_2 = [\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2]$$

and:

$$e^{\theta x_1 + (1-\theta)x_2)} = e^{\theta x_1} e^{(1-\theta)x_2} \leq t_1^\theta t_2^{1-\theta} \leq \theta t_1 + (1-\theta)t_2$$

from which we conclude that $\bar{x} \in$ epi$f$ and that f is convex.

(b) 2p It is not convex, since it is the set of points in $\mathbb{R}^2$ lying below the graph of the Gaussian function, which can be graphically be shown to not be convex. Formally, $\mathcal{S}$ is the hypograph of the function $f(x) = \mathcal{N}(x; \mu, \sigma^2)$. Since function $f(x)$ is not concave (it is quasi concave), then its hypograph cannot be convex.

(c) 1p It is concave, since for any $y \in [1, 2]$ the function yax^2 is concave in x, due to the fact that $a < 0$.

(d) 1p It is not convex, since it is the circumference of a circle.

(e) 1p It is convex, since it is simply the line segment $\theta \in [0, 2\pi]$.

(f) 1.5p It is convex since the Hessian of the function is A, which is always positive semidefinite due to the fact that it is a matrix given by the sum of outer products of a vector with itself.

Question 2 (10.5 points)

Consider the following optimization problem:

$$\begin{aligned} \min_x \; & f(x) := \frac{1}{2}(x+2)^2 \\ \text{s. t.} \; & x \leq -3 \end{aligned} \tag{$\mathcal{P}$}$$

(a) Draw the objective function and the constraint set. Compute the KKT conditions and find $x^\star$. **(2.5 points)**

Suppose that now our constraint set $\mathcal{C}$ is: $x \in [-8, 2]$, while the objective function $f(x)$ does not change.

(b) How do you expect the lagrangian multiplier(s) associated to the inequality constraints to be (higher, smaller or equal to 0)? Why? **(0.5 points)**

Although this optimization problem is manageable in a low-dimensional setting, where solutions can be obtained using KKT conditions, projected gradient descent, or even by visual inspection, it becomes considerably more challenging in higher dimensions with a diverse set of constraints. Hence, we proceed by developing an iterative algorithm tailored specifically to solve constrained problems.
Specifically, assume that at the kth iteration of our algorithm, our estimate x_k is $x_k = -6$. Denote with $\hat{f}_k(x)$ the first-order approximation of $f(\cdot)$ at the point x_k.

(c) Compute $\hat{f}_k(x)$ and solve:

$$\begin{aligned} s_k := \operatorname*{argmin}_x \; & \hat{f}_k(x) \\ \text{s. t.} \; & x \in \mathcal{C} \end{aligned}$$

Is the value of the original function $f(\cdot)$ attained at s_k lower or higher than the value of $\hat{f}_k(\cdot)$ at s_k? Do you always expect such property to hold? Draw a sketch of this optimization problem. **(3 points)**

The point s_k is the minimizer of $\hat{f}_k(x)$ and not of $f(x)$. Thus in general $f(s_k)$ might be even higher than the value attained by the previous iterate x_k, i.e., $f(x_k)$. Thus, we usually compute the iterate at iteration $k+1$, i.e. x_{k+1} with the following update:

$$x_{k+1} = \alpha^\star x_k + (1 - \alpha^\star) s_k$$

where $\alpha^\star$ is the scalar value minimizing the following line-search problem:

$$\begin{aligned} \alpha^\star = \operatorname*{argmin}_\alpha \; & f(\alpha x_k + (1-\alpha) s_k) \\ \text{s. t.} \; & \alpha \in [0, 1] \end{aligned}$$

(d) Is x_{k+1} always feasible? Why? Compute $\alpha^\star$ and x_{k+1} **(3 points)**

(e) Solve the original problem $\mathcal{P}$ with a projected gradient descent, starting at $x_0 = -6$ and a step size $t = 0.5$. How many steps do you need to reach the optimal point? **(1.5 points)**

Solution

In this question, in particular (c) and (d), we will develop the so called Frank-Wolfe method, also called conditional gradient method, used to solve constrained optimization problem.

(a) The Lagrangian is $L(x, \lambda) = \frac{1}{2}(x+2)^2 + \lambda(x+3)$ and the KKT conditions:

- $x^\star + 3 \leq 0$
- $\lambda^\star \geq 0$
- $\lambda^\star(x^\star + 3) = 0$
- $\frac{\partial}{\partial x} L(x, \lambda) = 0$

From the last equation we get $x^\star + 2 + \lambda^\star = 0 \Rightarrow x^\star = -\lambda^\star - 2$. Substituting into the complementary slackness condition we find $-\lambda^{\star 2} + \lambda^\star = -\lambda^\star(\lambda^\star - 1) = 0$. From this equality $\lambda^\star$ can either be 0 or +1. If $\lambda^\star = 0$ we have $x^\star = -2$, which is not feasible because the primal feasibility is not satisfied. Thus we conclude $\lambda^\star = 1$ (the constraint is active) and $x^\star = -3$.

(b) Since the objective function is a parabola centered at $x = -2$, and this point is strictly feasible, then the Lagrange multipliers are equal to 0.

(c) We have:

$$\hat{f}_k(x) = f(x_k) + f'(x_k)(x - x_k) = 8 - 4(x+6) = -4x - 16$$

Thus the optimization problem we need to solve is:

$$\begin{aligned} s_k := \underset{x}{\operatorname{argmin}} \ & -4x - 16 \\ \text{s. t.} \quad & -8 \leq x \leq 2 \end{aligned}$$

Since this is a linear program, its optimum lies in the border of the feasible set, specifically at $x = 2$. Thus $s_k = 2$. It holds $f(s_k) = 8 > -24 = \hat{f}_k(s_k)$. This was expected since the first-order Taylor approximation of a function is a global underestimator of the function it approximates.

(d) The point x_{k+1} is always feasible because it is the convex combination of x_k and s_k, which are both feasible points. To find this point, we need look at the line segment $\alpha x_k + (1-\alpha)s_k$ and find the $\alpha^\star$ such that the point $\alpha^\star x_k + (1-\alpha^\star)s_k$ minimizes $f(\cdot)$.

The line segment is given by $-6\alpha + 2(1-\alpha) = -8\alpha + 2$. Thus:

$$\begin{aligned} \alpha^\star &= \underset{\alpha \in [0,1]}{\operatorname{argmin}} f(-8\alpha + 2) \\ &= \underset{\alpha \in [0,1]}{\operatorname{argmin}} \frac{1}{2}(-8\alpha + 4)^2 \end{aligned}$$

from which $\alpha^\star = 0.5$. Thus $x_{k+1} = -2$.

(e) The gradient step is given by:

$$x_1 = -6 - 0.5(-6 + 2) = -4$$

and since the point is already in the feasible set, the projection does not affect the point. Then, the next iterate is:

$$x_2 = -4 - 0.5(-4 + 2) = -3$$

which is the optimum of the problem. So in total 2 iterations were necessary to reach convergence.

Question 3 (10 points)

Consider the following unconstrained optimization problem:

$$f(x) = 3|x_1| + 2|x_2|$$

(a) Is the above problem convex? Is it differentiable? **(1 point)**

(b) Sketch the function using at least 2 of its sublevel sets. Indicate the objective function value at one of the sublevel sets, and give the coordinates of the points where the sublevel set crosses the axes. (**1 point**)

(c) What is the subdifferential of the function at the point $x = (0, 2)$? (**3 points**)

(d) Calculate $x^{(1)}$, i.e. the point obtained after one iteration of a (sub)gradient descent algorithm for minimizing $f(x)$, starting from $x^{(0)} = (2, 2)$. Use exact line search for the step size. Explain your steps and all details of your calculation. (**3 points**)

(e) Indicate the search line (i.e. along which you performed the exact line search) on your sketch from question (b). Discuss whether your result for $x^{(1)}$ matches your expectations based on this sketch. (**2 points**)

Solution

(a) It is convex, but non-differentiable.

(b) See Figure 1 below.

(c) Taking $f_1(x) = 3|x_1|$ and $f_2(x) = 2|x_2|$ and using the addition rule of subgradient calculus:

$$\partial f(x) = \partial f_1(x) + \partial f_2(x)$$

At $x = (0, 2)$ $f_2(x)$ is differentiable, therefore, $\partial f_2(x) = \nabla f_2(x) = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$. However, at $x = (0, 2)$ $f_1(x)$ is not differentiable. It can be written as the pointwise maximum of 2 functions:

$$f_1(x) = \max(g_1(x), g_2(x)) = \max(3x_1, -3x_1)$$

According to the finite pointwise maximum rule

$$\partial f_1(x) = \mathbf{Co} \bigcup \{\partial g_i(x) | g_i(x) = f_1(x)\} \qquad (1)$$

As both functions (i.e. $g_1(x)$ and $g_2(x)$) are active at $x = (0, 2)$, we need to take the convex hull of the union of their subdifferentials. $\partial g_1(x) = \nabla g_1(x) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, while $\partial g_2(x) = \nabla g_2(x) = \begin{bmatrix} -3 \\ 0 \end{bmatrix}$. The convex hull of their union gives a line segment between $(3, 0)$ and $(-3, 0)$ for $\partial f_1(x)$. Finally, $\partial f(x) = \partial f_1(x) + \partial f_2(x)$ will be a line segment between $(3, 2)$ and $(-3, 2)$.

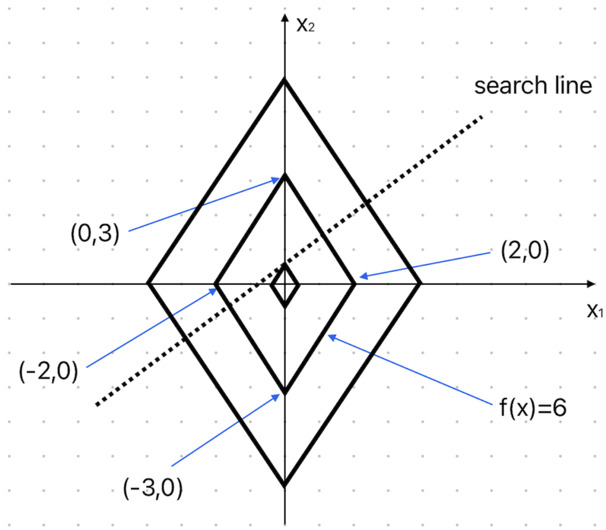

Figure 1: Countour lines of $f(x)$ and search line of gradient descent from point (2,2)

(d) At $(2,2)$ the function is differentiable, so we can use gradient descent. First, find an expression for $x^{(1)}$:

$$x^{(1)} = x^{(0)} - t\nabla f(x^{(0)}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - t \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

To determine the step size, we have to find

$$\underset{t}{\operatorname{argmin}}\, f(\begin{bmatrix} 2 \\ 2 \end{bmatrix} - t \begin{bmatrix} 3 \\ 2 \end{bmatrix}) = 3|2 - 3t| + 2|2 - 2t|.$$

This is a piece-wise linear function. Simple inspection reveals that the minimum is attained at $t = \frac{2}{3}$.

Finally, substituting this back to equation (1):

$$x^{(1)} = x^{(0)} - t\nabla f(x^{(0)}) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2}{3} \end{bmatrix}$$

(d) See Figure 1 below. $x^{(1)}$, i.e. the minimum point along the line is the point where the search line crosses the x_2 axis. This is logical, considering that this point is the upper peak of a small diamond, representing a sublevel set with a small objective function value. Points around $x^{(1)}$ are outside the small diamond, hence, they have a larger objective function value.

Question 4 (11 points)

Suppose a user wants to take a flight from Amsterdam to Madrid and can select either our airline a or an alternative airline b. Our airline has two flights: flight $a1$ with price $c_{a1} > 0$ and flight $a2$ with price $c_{a2} > c_{a1} > 0$. The price for the alternative airline is $c_b > 0$. Our goal is to maximize the probability that a user will take our airline.

One would expect that a user always selects the cheapest airline but this does not account for irrational user behavior. Researchers in transportation tackle this issue by modeling the probability of taking airline a as $p_a = \frac{e^{-c_a}}{e^{-c_a}+e^{-c_b}}$, where c_a is the cost for airline a which is thus either c_{a1} or c_{a2}. As a result, the following problem is solved

$$
\begin{aligned}
(p_a^*, z_{a1}^*, z_{a2}^*) = \arg \max_{p_a, z_{a1}, z_{a2}} \quad & p_a \qquad (2)\\
\text{subject to} \quad & p_a \leq \frac{e^{-c_{a1}z_{a1}-c_{a2}z_{a2}}}{e^{-c_{a1}z_{a1}-c_{a2}z_{a2}} + e^{-c_b}},\\
& z_{a1}, z_{a2} \in \{0,1\},\\
& z_{a1} + z_{a2} = 1.
\end{aligned}
$$

Here the variables z_{a1} and z_{a2} are binary and indicate which flight from airline a the user will take, either flight $a1$ or flight $a2$.

(a) Show that problem (2) is non-convex even if z_{z1} and z_{a2} are not binary. To do this, make a rough plot of the function $f(x) = \frac{e^{-x}}{e^{-x}+c}$, where c is a constant, and conclude from the plot that $f(x)$ cannot be convex (check for instance the limits of $f(x)$ at $-\infty$ and $+\infty$). **(1 point)**

(b) Derive the solution of (2) just by observing the problem and using the shape of the function $f(x)$ in (a). **(2 points)**

Practical transportation problems are way more complex and an intuitive solution cannot easily be found. Therefore, as an alternative consider the following convex problem

$$
\begin{aligned}
(p_a^*, p_b^*, z_{a1}^*, z_{a2}^*) = \arg \min_{p_a, p_b, z_{a1}, z_{a2}} \quad & p_a[\log(p_a) - 1] + p_b[\log(p_b) - 1] + c_{a1}z_{a1} + c_{a2}z_{a2} + c_b p_b \qquad (3)\\
\text{subject to} \quad & p_a = z_{a1} + z_{a2},\\
& 0 \leq z_{a1} \leq 1,\ 0 \leq z_{a2} \leq 1,\\
& p_a + p_b = 1.
\end{aligned}
$$

Note that the variables z_{a1} and z_{a2} are continuous between 0 and 1. Also note that the domain of the objective function is implicitly restricted due to the logarithm.

(c) Prove that problem (3) is convex. **(2 points)**

(d) Give the Lagrangian and the KKT conditions of problem (3). **(3 points)**

(e) From the above KKT conditions, derive the solution to (3) and show that the solution for p_a is the same as the solution to (2). Pay particular attention to the complementary slackness conditions to set certain parameters to zero. **(3 points)**

Solution

(a) First of all, to show that $f(x)$ is non-convex, observe that the limit at ∞ is 1 and at $+\infty$ is 0. So the function looks like an inverse sigmoid function and is thus non-convex. Replacing the argument with an affine function does not change the non-convexity.

(b) Obviously, at the optimal point, the constraint is reached with equality. Otherwise, there would exist a better solution. Furthermore, as we observed in (a), the function $f(x)$ is monotonically decreasing. So the left-hand side of the first constraint is maximal when the $z_{a1} = 1$ and $z_{a2} = 0$ since $c_{a1} < c_{a2}$. The solution for p_a then follows naturally.

(c) The constraints are all linear. The cost function is further linear in z_{a1} and z_{a2}. Finally, the second-order derivative of the cost towards p_a is

$$\frac{\partial^2 f(p_a, z_{a1}, z_{a2})}{\partial^2 p_a} = 1/p_a,$$

which is positive in the domain of the logarithm. The same holds for p_b.

(d) The Lagrangian can be expressed as

$$\begin{aligned} L = & p_a[\log(p_a) - 1] + p_b[\log(p_b) - 1] + c_{a1}z_{a1} + c_{a2}cz_{a2} + c_b p_b \\ & + \lambda(p_a - z_{a1} - z_{a2}) \\ & - \mu_1 z_{a1} - \mu_2 z_{a2} \\ & + \tau_1(z_{a1} - 1) + \tau_2(z_{a2} - 1) \\ & + \nu(p_a + p_b - 1). \end{aligned}$$

The KKT conditions are given by

- Primal constraints: $p_a = z_{a1} + z_{a2}$, $0 \le z_{a1} \le 1$, $0 \le z_{a2} \le 1$, $p_a + p_b = 1$.
- Dual constraints: $\mu_1, \mu_2, \tau_1, \tau_2 \ge 0$.
- Complementary slackness: $\mu_1 z_{a1} = \mu_2 z_{a2} = \tau_1(z_{a1} - 1) = \tau_2(z_{a2} - 1) = 0$.
- Vanishing gradients:

$$\begin{aligned} \frac{\partial L}{\partial p_b} &= \log(p_b) + c_b + \nu = 0 \quad \Rightarrow \quad p_b^* = \frac{e^{-c_b}}{e^{\nu}} \\ \frac{\partial L}{\partial p_a} &= \log(p_a) + \lambda + \nu = 0 \quad \Rightarrow \quad p_a^* = \frac{e^{-\lambda}}{e^{\nu}} \\ \frac{\partial L}{\partial z_{a1}} &= c_{a1} - \lambda - \mu_1 + \tau_1 = 0 \quad \Rightarrow \quad \lambda = c_{a1} - \mu_1 + \tau_1 \\ \frac{\partial L}{\partial z_{a2}} &= c_{a2} - \lambda - \mu_2 + \tau_2 = 0 \quad \Rightarrow \quad \lambda = c_{a2} - \mu_2 + \tau_2 \end{aligned}$$

(e) From $p_a + p_b = 1$, we obtain $e^{\nu} = e^{-\lambda} + e^{-c_b}$ and thus we have

$$p_a^* = \frac{e^{-\lambda}}{e^{-\lambda} + e^{-c_b}}, \qquad p_b^* = \frac{e^{-c_b}}{e^{-\lambda} + e^{-c_b}}.$$

As e^{-c_b} and $e^{-\lambda}$ are positive, neither $p_a^*, p_b^*, z_{a1}^*, z_{a2}^*$ can be 1, so $\tau_1, \tau_2 = 0$ from complementary slackness. With this information, we update

$$\lambda = c_{a1} - \mu_1$$
$$\lambda = c_{a2} - \mu_2$$

As p_a is positive, μ_1 and μ_2 can not be both different from zero. Furthermore, we know that μ_1 and μ_2 should be positive. Then the only option for $\lambda = c_{a1} - \mu_1 = c_{a2} - \mu_2$ to be true is that λ is equal to the smallest cost, i.e., $\lambda = c_{a1}$, $\mu_1 = 0$ and $\mu_2 = c_{a2} - c_{a1}$. So, the final solution for p_a is

$$p_a^* = \frac{e^{-c_{a1}}}{e^{-c_{a1}} + e^{-c_b}}.$$

This is also the solution of (2).