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

EE 4530 APPLIED CONVEX OPTIMIZATION

19 January 2022, 13:30–16:30

Open book. Copies of the book and the course slides allowed. 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.

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) $f(x) = \exp(|x^3|)$ with $\mathrm{dom} f = \mathbf{R}$;

(b) $f(X) = \det(X)$, with $X \in \mathbb{S}^n$. *(Hint: use the definition of convexity.)*

(c) The set $\mathcal{S} = \{s \in \mathbf{R}^n | f(y) \geq f(x) + s^\top(y - x)\}$. Which set is it? *(Hint: use the definition of convexity.)*

(d) $f(x) = -\log(\min\{-(x^2) + 1, -(x-1)^2 + 1\})$ with $\mathrm{dom} f = [0, 1]$;

(e) $f(x) = \inf_z(\sum_{i=1}^n x_i z_i^2)$ with $x, z \in \mathbf{R}^n$.

(f) Consider $f(x) = \frac{1}{2}\|b - Ax\|_2^2 - \gamma\|x\|_2^2$, with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$. For which conditions on A and γ is the function $f(x)$ convex in $\mathbb{R}^n$?

(g) Sketch the set $\mathcal{C} = \{x \in \mathbf{R}^2 |\, \mathrm{trace}(A(x-1)(x-1)^\top) \leq 1 \ \bigcap \ x_1 = x_2\}$ with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Is it convex (by inspection of your drawing)? *(Hint: recall the cyclic property* $\mathrm{trace}(ABC) = \mathrm{trace}(BCA)$ *and* $\mathrm{trace}(a) = a$ *if* a *scalar.)*

Solution

(a) It is convex. A way to show it is through the composition rule.

(b) Neither convex nor concave. To see that it is not convex, pick two elements X and Y of the domain, such that $X = \text{Diag}([\mathbf{1}_l \ \mathbf{0}_{n-l}])$ and the matrix $Y = I - X$; i.e., X is the diagonal matrix with the first l diagonal entries equal to 1, and Y is the diagonal matrix with the last $n - l$ diagonal entries equal to 1. For any $\theta \in (0, 1)$, it holds:

$$\det(\theta X + (1-\theta)Y) \geq \theta \det(X) + (1-\theta)\det(Y) = 0. \tag{1}$$

Hence the function is not convex.

To see that is not concave, consider now the diagonal matrix $X = \text{Diag}([-1 \ \mathbf{0}_{n-1}])$ and the diagonal matrix $Y = X + I$. For any $\theta \in (0, 1)$, it holds:

$$\det(\theta X + (1-\theta)Y) \leq \theta \det(X) + (1-\theta)\det(Y) = 0. \tag{2}$$

Hence the function is not concave.

(c) It is the subdifferential set, which is always convex. To prove it, we need to show that $\theta s_1 + (1-\theta)s_2 \in \mathcal{S}$, for any $s_1, s_2 \in \mathcal{S}$ and $\theta \in (0, 1)$.

$$f(x) + (\theta s_1 + (1-\theta)s_2)^\top (y - x) \tag{3}$$
$$= \theta(f(x) + s_1^\top (y - x)) + (1-\theta)(f(x) + s_2^\top (y - x)) \tag{4}$$
$$\leq \theta f(y) + (1-\theta) f(y) \tag{5}$$
$$= f(y) \tag{6}$$

(d) It is convex. It is the composition of the pointwise minimum of concave functions (i.e. concave) with the negative logarithm (convex and non-increasing). According to the composition rule, this is convex.

(e) It is concave. It is the infimum of linear (hence concave) functions, indexed by the variable z, over the variable x.

(f) For $\gamma \leq 0$ and for any A.

(g) It is convex. It is the intersection between an ellipse centered at the point $(1, 1)$ and the line bisecting the first and the third quadrant; both sets are convex and hence their intersection.

Question 2 (10 points)

Consider the optimization problem

$$\begin{aligned} &\underset{x}{\text{minimize}}\ f_0(x) \\ &\text{subject to } x \leq 0, \end{aligned} \tag{7}$$

with $x \in \mathbb{R}^n$ and $f_0(x)$ a convex function.

(a) Approximate the optimization problem by an unconstrained problem using the logarithmic barrier function. You can set the approximation parameter to $t = 1$ for simplicity.

(b) Compute the gradient and Hessian of the new objective function as a function of the gradient and Hessian of $f_0(x)$.

Assume now that $f_0(x) = x^T x + 2 \cdot 1^T x + n$, where 1 is the n-dimensional vector of all ones.

(c) What is the solution for x of the original problem (7) and why? Is the solution for x of the approximate (log-barrier) problem larger or smaller if you consider a single entry? Explain also why.

(d) Plug in the gradient and Hessian of $f_0(x)$ to update the results of (b). Write down the update equations for the gradient descent method and the Newton descent method of the approximate (log-barrier) problem. Assume no line search is considered.

(e) Assume $x = -2 \cdot 1$ is selected as initial point, with 1 again the n-dimensional vector of all ones. Is it then possible for the gradient descent method and the Newton descent method of (d) to obtain the same updates? If so, under what condition of the step sizes?

(f) Next to the log-barrier method, can you think of another iterative method to solve this optimization problem? Work out this method for the considered problem.

Solution

(a) The new problem is

$$\underset{x}{\text{minimize}}\ f_0(x) - \sum_{i=1}^{n} \log(-x_i).$$

(b) Note that in this case, we have

$$f_i(x) = x_i = e_i^T x,$$

where e_i is the i-th canonical vector. The gradient and Hessian are then given by

$$\nabla f(x) = \nabla f_0(x) - \sum_{i=1}^{n} \frac{1}{x_i} e_i = \nabla f_0(x) - [1/x_1, 1/x_2, \ldots, 1/x_n]^T,$$

$$\nabla^2 f(x) = \nabla^2 f_0(x) + \sum_{i=1}^{n} \frac{1}{x_i^2} e_i e_i^T = \nabla^2 f_0(x) + \text{diag}(1/x_1^2, 1/x_2^2, \ldots, 1/x_n^2).$$

(c) The objective can be written as $f_0(x) = (x+1)^T(x+1)$ and the solution of the original problem is thus given by $x = -1$. Clearly the constraint is satisfied. For the approximate problem, the solution per entry is smaller since the gradient of the new objective function in $x = -1$ is 1 for every entry.

(d) The gradient and Hessian of the new objective function are given by

$$\begin{aligned}\nabla f(x) &= 2(x+1) - [1/x_1, 1/x_2, \ldots, 1/x_n]^T, \\ \nabla^2 f(x) &= 2I + \mathrm{diag}[1/x_1^2, 1/x_2^2, \ldots, 1/x_n^2].\end{aligned}$$

The gradient descent method is given by

$$x^{(k+1)} = x^{(k)} - t_{\mathrm{GD},k}\nabla f(x^{(k)}) = x^{(k)} - t_{\mathrm{GD},k}\{2(x^{(k)}+1) - [1/x_1^{(k)}, 1/x_2^{(k)}, \ldots, 1/x_n^{(k)}]^T\},$$

with step size $t_{\mathrm{GD},k}$. The Newton descent method by

$$\begin{aligned}x^{(k+1)} &= x^{(k)} - t_{\mathrm{HD},k}[\nabla^2 f(x^{(k)})]^{-1}\nabla f(x^{(k)}) \\ &= x^{(k)} - t_{\mathrm{HD},k}\{2I + \mathrm{diag}[1/(x_1^{(k)})^2, 1/(x_2^{(k)})^2, \ldots, 1/(x_n^{(k)})^2]\}^{-1} \\ &\quad \cdot \{2(x^{(k)}+1) - [1/x_1^{(k)}, 1/x_2^{(k)}, \ldots, 1/x_n^{(k)}]^T\},\end{aligned}$$

with step size $t_{\mathrm{HD},k}$.

(e) Because the initial point has the same entries, after every update, both the gradient and Newton descent method will lead to a new solution with the same entries, which we can denote as $\alpha^{(k)}$, i.e., $\alpha^{(k)} = x_i^{(k)}$, $\forall i$. The Hessian in every step is thus given by $\nabla^2 f(x^{(k)}) = [2 + 1/(\alpha^{(k)})^2]I$. And as a consequence, the two methods are equal when

$$t_{\mathrm{GD},k} = t_{\mathrm{HD},k}[2 + 1/(\alpha^{(k)})^2]^{-1}.$$

(f) We can consider a projected gradient descent method. Starting from $x^{(0)}$, the algorithm computes for $k = 0, 1, \ldots$:

$$\begin{aligned}\tilde{x}^{(k+1)} &= x^{(k)} - t\nabla f_0(x^{(k)}) = x^{(k)} - t2(x^{(k)}+1), \\ x_i^{(k+1)} &= \min\{0, \tilde{x}_i^{(k+1)}\}, \ \forall i.\end{aligned}$$

Question 3 (10 points)

Consider a linear program of the form

$$\begin{aligned} &\underset{x}{\text{minimize}}\ c^T x \\ &\text{subject to } Ax \preceq b, \end{aligned} \tag{8}$$

where $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Let us further assume that b can be modeled as

$$b(u) = b_0 + Bu, \tag{9}$$

with $b_0 \in \mathbb{R}^m$, $B \in \mathbb{R}^{m\times p}$, and u some p-dimensional variable. You can think of Bu as the deviation of b around its nominal value b_0.

Let us first assume that u is a deterministic unknown value constrained to a ball of radius 1, i.e., $\|u\|_2^2 \le 1$. In other words, we have two unknown variables now, x and u.

(a) Based on (8) and (9) as well as the fact that $\|u\|_2^2 \le 1$, write down the new problem in both x and u. Is this problem convex?

(b) Write down the Lagragian function and the KKT conditions of the problem in (a).

(c) Derive the dual problem of the problem in (a).

(d) Is there strong duality? Explain clearly why or why not.

Solution

(a) The new problem is given by

$$\begin{aligned} &\underset{x,u}{\text{minimize}}\ c^T x \\ &\text{subject to } Ax - Bu - b_0 \preceq 0 \\ &\text{subject to } u^T u - 1 \le 0. \end{aligned}$$

This problem is convex since first of all the objective function is linear and hence convex. Second, the inequality constraint functions are either linear or quadratic with a positive semidefinite Hessian (I in this case), which all are convex functions.

(b) The Lagrangian function is given by

$$\begin{aligned} L(x, u, \nu, \lambda) &= c^T x + \nu^T(Ax - Bu - b_0) + \lambda(u^T u - 1) \\ &= \lambda u^T u - \nu^T Bu + (c^T + \nu^T A)x - \nu^T b_0 - \lambda, \end{aligned}$$

where $\nu \succeq 0$ and $\lambda \ge 0$. The KKT conditions can be written as

1. $Ax - Bu - b_0 \preceq 0$, $u^T u - 1 \le 0$
2. $\nu \succeq 0$, $\lambda \ge 0$
3. $\nu_i(a_i^T x - b_i^T u - b_{0,i}) = 0$, $\forall i$, $\lambda(u^T u - 1) = 0$

4. $c + A^T\nu = 0$, $2\lambda u - B^T\nu = 0$

(c) The dual function is given by

$$g(\nu,\lambda) = \min_{x,u} L(x,u,\nu,\lambda) = \min_{x,u}\{\lambda u^T u - \nu^T B u + (c^T + \nu^T A)x - \nu^T b_0 - \lambda\}.$$

The Lagrangian is linear in x and quadratic in u so the dual function is given by

$$g(\nu,\lambda) = \begin{cases} -\frac{1}{4\lambda}\nu^T B B^T \nu - \nu^T b_0 - \lambda & \text{if } c^T + \nu^T A = 0 \\ -\infty & \text{if } c^T + \nu^T A \neq 0 \end{cases}.$$

The dual problem becomes

$$\begin{array}{ll} \underset{\nu,\lambda}{\text{maximize}} & -\frac{1}{4\lambda}\nu^T B B^T \nu - \nu^T b_0 - \lambda \\ \text{subject to} & c^T + \nu^T A = 0, \quad \nu \succeq 0, \quad \lambda \geq 0. \end{array}$$

(d) It is clear that it is always possible to find a u for which $\|u\|_2 < 1$. Take for instance $u = 0$. Then the question is whether we can find an x for which $Ax - b_0 \prec 0$. Every inequality $a_i^T x - b_{0,i} < 0$ forms an open half plane, so all constraints together form an intersection of open half planes. Only in some very exceptional cases this intersection can be empty without the constraint set being empty (for instance if $a_j = -a_i$ and $b_{0,j} = -b_{0,i}$ for some i and j). But in general the intersection is not empty. So Slater's condition is generally satisfied and we have strong duality.

Question 4 (10 points)

As in Question 3, let us consider the linear program (8) as well as the model (9).

This time we assume that u is known but random. More specifically, we assume that u has a specific distribution over $\mathcal{U} = [-1, 1]^p = \{u \mid \|u\|_\infty \leq 1\}$. We can then solve (8) for many different realizations of u, all leading to different solutions x. The resulting mapping from u into $x(u)$ is called the optimal policy, since it gives the optimal variable x for each deviation u. When there is not enough time to find this optimal policy, we seek a suboptimal policy. Here we will focus on finding a suboptimal affine policy

$$x_{\text{aff}}(u) = x_0 + Ku,$$

with $x_0 \in \mathbb{R}^n$ and $K \in \mathbb{R}^{n \times p}$. If we can estimate x_0 and K in advance, evaluating the new policy is very simple (matrix multiplication and addition). We will choose x_0 and K in order to minimize the expected value of the objective, while insisting that for any value of u, feasibility is maintained:

$$\begin{aligned} \underset{x_0, K}{\text{minimize}} \;& E\{c^T x_{\text{aff}}(u)\} \\ \text{subject to } & A x_{\text{aff}}(u) \preceq b(u), \; \forall u \in \mathcal{U}, \end{aligned} \tag{10}$$

with $b(u)$ as in (9). The expectation in the objective is over u, and the set of constraints guarantees feasibility.

(a) Work out the expected value in the objective value of (10).

(b) Note that because of $u \in \mathcal{U}$, problem (10) has an infinite number of m-dimensional constraints. Turn this infinite number of constraints into a finite number of constraints. To do that, look at the m dimensions separately and make use of the property

$$a^T u \leq b, \; \forall u \in \mathcal{U} \;\Leftrightarrow\; \max_{u \in \mathcal{U}}\{a^T u\} \leq b.$$

(Hint: observe the relation of the latter maximum with the dual norm.)

(c) Write down the new problem that is equivalent to (10) but has no expectations and no infinite number of constraints.

(d) Turn the problem of (c) into a linear program.

Solution

(a) The expected value can easily be written as

$$E\{c^T x_{\text{aff}}(u)\} = E\{c^T (x_0 + Ku)\} = c^T x_0.$$

(b) Writing one of the dimensions separately, we obtain the constraints

$$a_i^T x_0 + a_i^T K u \leq b_{0,i} + b_i^T u, \; \forall u \in \mathcal{U},$$

where a_i^T is the ith row of A and b_i^T is the ith row of B. This can be rewritten as

$$a_i^T x_0 - b_{0,i} + (K^T a_i - b_i)^T u \leq 0, \; \forall u \in \mathcal{U}.$$

Applying the provided trick, this infinite number of constraints can also be replaced by

$$a_i^T x_0 - b_{0,i} + \max_{u \in \mathcal{U}} \{(K^T a_i - b_i)^T u\} \leq 0.$$

Since $u \in [-1, 1]^p$ it is intuitively clear that this maximum is given by the sum of the absolute values of the entries of $K^T a_i - b_i$, i.e., the 1-norm of $K^T a_i - b_i$. Similarly, we see that the maximum is equal to the dual norm of the ∞-norm, which is the 1-norm. So the infinite number of constraints can be replaced by the single constraint

$$a_i^T x_0 - b_{0,i} + \|K^T a_i - b_i\|_1 \leq 0.$$

(c) So the new problem becomes

$$\begin{aligned} \underset{x_0, K}{\text{minimize}} \ & c^T x_0 \\ \text{subject to} \ & a_i^T x_0 - b_{0,i} + \|K^T a_i - b_i\|_1 \leq 0, \ \forall i. \end{aligned}$$

(d) To turn this into an LP, we introduce the new variables $t_{i,j} = k_j^T a_i - b_{i,j}$, where k_j is the jth column of the matrix K and $b_{i,j}$ is the jth entry of the vector b_i. Then the problem can be written as

$$\begin{aligned} \underset{x_0, K, \{t_{i,j}\}}{\text{minimize}} \ & c^T x_0 \\ \text{subject to} \ & a_i^T x_0 - b_{0,i} + \sum_{j=1}^{p} t_{i,j} \leq 0, \ \forall i \\ & t_{i,j} = |k_j^T a_i - b_{i,j}|, \ \forall i, j. \end{aligned}$$

This last constraint can be replaced by forcing $t_{i,j}$ to be larger than both $k_j^T a_i - b_{i,j}$ and $-k_j^T a_i + b_{i,j}$ so that it will always be larger than the positive value. The first constraint will then push $t_{i,j}$ to be as small as possible and hence $t_{i,j}$ has to be equal to $|k_j^T a_i - b_{i,j}|$. So the final equivalent LP is given by

$$\begin{aligned} \underset{x_0, K, \{t_{i,j}\}}{\text{minimize}} \ & c^T x_0 \\ \text{subject to} \ & a_i^T x_0 - b_{0,i} + \sum_{j=1}^{p} t_{i,j} \leq 0, \ \forall i \\ & t_{i,j} \geq k_j^T a_i - b_{i,j}, \ \forall i, j \\ & t_{i,j} \geq -k_j^T a_i + b_{i,j}, \ \forall i, j. \end{aligned}$$