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

EE 4530 APPLIED CONVEX OPTIMIZATION

25 January 2023, 13:30–16:30

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 function

$$f(A) = \inf_{\|x\|_2=1} \operatorname{tr}(Axx^\top), \tag{1}$$

with $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ not positive semidefinite (PSD)

(b) The function $g(x) = x^\top Qx$, with $Q = [\nabla^2 f(a,b,c)]^2$, where $f(a,b,c) : \mathbf{R}^3 \to \mathbf{R}$ is the determinant of the matrix:

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \tag{2}$$

for some fixed values $a, b, c \in \mathbf{R}$.

(c) The function $f(x) = \log(|x|)$ with $\operatorname{dom} f = \mathbf{R}$

(d) The function $f(x) = \sup_{\|y\|_2=1}\{|y^\top Ax|, \|Ax-y\|_2^2\}$, for $x, y \in \mathbf{R}^n$ and $A \in \mathbf{R}^{n\times n}$

(e) The function $f(x) = \frac{x(x-1)\cdots(x-r+1)}{r!}$ for $r-1 < x < +\infty$

(f) The set of Vandermonde matrices of dimension n: $\mathcal{V} = \{V(x) \in \mathbb{R}^{n\times n} | \; x \in \mathbb{R}^n\}$. A Vandermonde matrix V with parameter $x := [x_0, \ldots, x_{n-1}]^\top \in \mathbf{R}^n$ is defined as:

$$V(x) := \begin{bmatrix} 1 & x_0 & \cdots & x_0^{n-1} \\ 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n-1} & \cdots & x_{n-1}^{n-1} \end{bmatrix}$$

(g) The set $\mathcal{E} = \{(x,\epsilon) \mid x \in \mathbf{R}^n, \|Ax-b\|_\infty > \epsilon\}$. Which set is it?

Solution

(a) It is the infimum over the variable x of a linear function in A, so it is both convex and concave.

(b) It is convex due to the positive semidefiniteness of Q. Since $Q = (\nabla^2 f(a,b,c))(\nabla^2 f(a,b,c))$, it suffices to note that $\forall x \in \mathbf{R}^3$ it holds $x^\top Q x = \|\nabla^2 f(a,b,c)x\|_2^2 \geq 0$. Another way to show its convexity is given by computing the Hessian matrix $\nabla^2 f(a,b,c)$, which is an antidiagonal matrix. When squared, $\nabla^2 f(a,b,c)$ becomes diagonal and it reveals in its diagonal the eigenvalues of matrix Q, which are all non-negative.

(c) It is neither convex nor concave. It is actually quasi-convex. This can be quickly seen by inspection.

(d) It is convex, since it is the pointwise supremum of convex functions in x, indexed by y.

(e) It is convex. To show this, we can use the second order condition $f''(x) \geq 0$ for all $x \in]r-1, +\infty]$. The function $f(x)$ is the product of r linear terms in x. Then, by the product rule, its derivative f' is the sum of r terms, each one consisting of $r-1$ products. For instance, one term is $(x-1)\cdots(x-r+1)$, corresponding to the factor x missing. The second derivative f'' is then the sum of an even greater number of terms, each one consisting of $r-2$ products, with factors linear in x. This can be for instance $(x-2)\cdots(x-r+1)$, with the factors x and $(x-1)$ removed. Since $x > r-1$, all the products are positive and their sum is also positive. Thus $f''(x) > 0$ for all the x of interest.

(f) It is not a convex set. It is easy to show that $\theta V(x) + (1-\theta)V(y)$ is in general not Vandermonde.

(g) It is not convex. It is the complementary set of the epigraph of the function $f(x) = \|Ax - b\|_\infty$, which is convex and hence also its epigraph $\mathrm{epi}(f) = \{(x,\epsilon) \mid x \in \mathbf{R}^n, \epsilon \in \mathbf{R}, \|Ax-b\|_\infty \leq \epsilon\}$

Question 2 (10 points)

Functional neuroimaging (such as functional ultrasound, fUS) data analysis often aims to identify neural processes that occur during the execution of a certain task, e.g. processing of visual stimuli shown to the experimental subject. However, the cerebral blood volume (CVB) signal measured by fUS is only a proxy for neural activity. The measured time series $\mathbf{y}$ at a given voxel is often assumed to be equal to the underlying neural activity $\mathbf{s}$ convolved with a so-called hemodynamic response function (HRF) that acts as a low-pass filter. Let us denote the Toeplitz matrix (with columns populated by a delayed version of the HRF filter coefficients) with $\mathbf{H}$. Further, we will assume that the measurement $\mathbf{y}$ is corrupted by additive Gaussian noise $\mathbf{e}$.

(a) Write down the minimization problem to recover the neural activity $\mathbf{s}$ based on the measurements $\mathbf{y}$, additive noise $\mathbf{e}$ and the known matrix $\mathbf{H}$ in the maximum likelihood sense.

Due to the fact that $\mathbf{H}$ in general has many highly correlated columns, the problem is ill-posed. Therefore, a pragmatic solution is to assume that the neural activity is sparse, i.e. the vector $\mathbf{s}$ has at most k nonzero elements, i.e., the stimulus was shown no more than k times to the subject during the measurements.

(b) Write down the optimization problem considering the sparsity assumption!

(c) Is this problem convex? If yes, prove convexity! If not, propose an appropriate unconstrained convex relaxation to the problem!

(d) Now, let us take a gradient descent step at $\mathbf{s}^{(0)} = \mathbf{1}$ as a starting point. Calculate the gradient at this point!

(e) Next, how would you determine the step size t using exact line search? No need to make calculations, but clearly explain the 3 steps you need to take to determine t.

(f) Is gradient descent always applicable in this problem?

Solution

(a)

$$\underset{\mathbf{s}}{\text{minimize}} ||\mathbf{y} - \mathbf{H}\mathbf{s}||_{\mathbf{2}}^{\mathbf{2}}$$

(b)

$$\begin{aligned} &\underset{\mathbf{s}}{\text{minimize}} ||\mathbf{y} - \mathbf{H}\mathbf{s}||_{\mathbf{2}}^{\mathbf{2}} \\ &\text{subject to } \mathbf{card}(\mathbf{s}) \leq \mathbf{k} \end{aligned}$$

(c) It is not convex, as the inequality constraint is not convex. A possible convex relaxation is achieved by the l1-norm heuristic as a regularization term:

$$\underset{\mathbf{s}}{\text{minimize}} ||\mathbf{y} - \mathbf{H}\mathbf{s}||_{\mathbf{2}}^{\mathbf{2}} + \lambda ||\mathbf{s}||_{\mathbf{1}}$$

(d)

$$\nabla_{\mathbf{s}}(||\mathbf{y}-\mathbf{Hs}||_{\mathbf{2}}^{\mathbf{2}}+\lambda||\mathbf{s}||_{\mathbf{1}})=\mathbf{2H^{T}Hs}-\mathbf{2H^{T}y}+\lambda\mathbf{1}$$

(e) Let us denote the step size with t. The first step is to write the update in terms of t.

$$\mathbf{s^{(1)}}=\mathbf{s^{(0)}}-\mathbf{t}\nabla\mathbf{f}(\mathbf{s^{(0)}})$$

Then, we have to evaluate the objective function at $\mathbf{s^{(1)}}$ (we get an expression which is paramterized with t):

$$f(\mathbf{s^{(1)}})=\|\mathbf{y}-\mathbf{Hs^{(1)}}\|_{\mathbf{2}}^{\mathbf{2}}+\lambda\|\mathbf{s^{(1)}}||_{\mathbf{1}}$$

Finally, setting the derivative of the objective function with respect to t to 0 and solving for t gives us the step size:

$$\frac{\partial f(\mathbf{s^{(1)}})}{\partial t}=0$$

(f) No, because it is not differentiable.

Question 3 (10 points)

A certain disease is treated with a combination of two drugs, D_1 and D_2. Unfortunately, these drugs cause liver damage proportional to $x_1^2 + 2x_1 + 4x_2^2$, where x_i represents the number of doses of D_i administered. In order for the treatment to be effective, you need at least 10 mg of active ingredient A and another 10 mg of active ingredient B. Each dose of D_1 has 2 mg of A and 1 mg of B, whereas each dose of D_2 has 1 mg of A and 2 mg of B. Hence, you find yourself trying to solve the following problem

$$\begin{aligned} \underset{x}{\text{minimize}} \ & x_1^2 + 2x_1 + 4x_2^2 \\ \text{subject to } & 2x_1 + x_2 \geq 10 \\ & x_1 + 2x_2 \geq 10 \end{aligned}$$

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

(b) Derive the dual function and write down the dual problem.

(c) Write the KKT conditions for this problem.

(d) Is $x_1 = 2,5$, $x_2 = 5$, $\lambda_1 = 36$ and $\lambda_2 = 0$ an optimal solution to the problem? Prove your answer. Let us say that $\lambda_2 = 0$ is optimal. What can you conclude from that?

(e) So far we have implicitly assumed that $x \in \mathbb{R}^2$. Suppose now we constrain x to be an integer vector, i.e., $x \in \mathbb{Z}^2$? How does this affect the optimal objective value p^* of the problem?

Solution

(a) Yes it is convex since the objective function is convex (Hessian is diagonal and both diagonal elements are always positive), and the constraints are affine.

(b) The Lagrangian function is given by

$$L(x, \lambda) = x_1^2 + 2x_1 + 4x_2^2 + \lambda_1(10 - 2x_1 - x_2) + \lambda_2(10 - x_1 - 2x_2).$$

To find the minimum over x we set the derivative of the Lagrangian w.r.t x equal to zero. This gives

$$\begin{aligned} 2x_1 + 2 - 2\lambda_1 - \lambda_2 &= 0 \\ 8x_2 - \lambda_1 - 2\lambda_2 &= 0 \end{aligned}$$

This means the solution for x is

$$\begin{aligned} x_1 &= \lambda_1 + \lambda_2/2 - 1 \\ x_2 &= \lambda_1/8 + \lambda_2/4 \end{aligned}$$

The dual function is hence given by

$$\begin{aligned} g(\lambda) = \min_x L(x, \lambda) =& (\lambda_1 + \lambda_2/2 - 1)^2 + 2(\lambda_1 + \lambda_2/2 - 1) + 4(\lambda_1/8 + \lambda_2/4)^2 \\ &+ \lambda_1[10 - 2(\lambda_1 + \lambda_2/2 - 1) - (\lambda_1/8 + \lambda_2/4)] \\ &+ \lambda_2[10 - (\lambda_1 + \lambda_2/2 - 1) - 2(\lambda_1/8 + \lambda_2/4)]. \end{aligned}$$

The dual problem is then

$$\begin{aligned} \underset{x}{\text{maximize}}\ & g(\lambda) \\ \text{subject to}\ & \lambda_1 \geq 0 \\ & \lambda_2 \geq 0 \end{aligned}$$

(c) The KKT conditions are given by

$$\begin{aligned} \text{primal constraints:}\ & 2x_1 + x_2 \geq 10, \quad x_1 + 2x_2 \geq 10 \\ \text{dual constraints:}\ & \lambda_1 \geq 0, \quad \lambda_2 \geq 0 \\ \text{complementary slackness:}\ & \lambda_1(10 - 2x_1 - x_2) = 0, \quad \lambda_2(10 - x_1 - 2x_2) = 0 \\ \text{null gradient of Lagrangian:}\ & 2x_1 + 2 - 2\lambda_1 - \lambda_2 = 0, \quad 8x_2 - \lambda_1 - 2\lambda_2 = 0 \end{aligned}$$

(d) No, since it does not satisfy the fourth KKT condition. If $\lambda_2 = 0$ is optimal, then this means that the second constraint is not active, i.e., $x_1 + 2x_2 > 10$. In other words, even without this constraint, the optimal solution is the same.

(e) If we include an integer constraint, then the problem becomes more constrained and hence the optimal function value will increase or remain the same.

Question 4 (10 points)

Suppose we have to solve a linear regression problem of the form

$$y_k \approx x_k^T w, \quad k = 1, 2, \ldots, K,$$

where for the kth data point, y_k is a given scalar and x_k is a given $N \times 1$ vector. The unknown regression coefficients are stacked in the $N \times 1$ vector w.

When $K \ll N$ we have too few equations to solve the problem and we need additional constraints. One option is to assume a sparse w. Let X be the wide $K \times N$ matrix stacking the vectors x_k^T as rows, y be the $K \times 1$ vector stacking the scalars y_k as entries, and $\epsilon > 0$ be a supplied constant. The popular Dantzig selector is then formulated through the following convex optimization problem:

$$\begin{aligned} \min_{w} \ & \|X^T(y - Xw)\|_\infty \\ \text{subject to } \ & \|w\|_1 \leq \epsilon \end{aligned}$$

The L∞-norm objective imposes that no feature has a high inner product with the residual, whereas the L1-norm constraint attempts to keep the weight vector w sparse.

Let us try to solve this problem using a first-order algorithm.

(a) First, derive a subgradient of $\|x\|_\infty$. To do this, assume the infinity norm is attained at index l, i.e., $\|x\|_\infty = |x_l|$. Also derive a subgradient of $\|x\|_1$.

(b) Denote $g_1(x)$ and $g_2(x)$ respectively as the subgradient of $\|x\|_\infty$ and $\|x\|_1$ at x derived in (a). Based on these two subgradients, derive a first-order algorithm to find the Dantzig selector.

Alternatively, we can rewrite the Dantzig selector problem as a linear program for which many algorithms have been developed. This is what we do next in different steps.

(c) Rewrite the L∞ objective as an auxiliary variable c by introducing two additional vector inequalities.

(d) Rewrite the L1-norm constraint as $1^T t \leq \epsilon$ with auxiliary $N \times 1$ vector t by introducing two additional vector inequalities.

(e) Based on (a) and (b), reformulate the original optimization problem as a linear program in w, c and t.

Solution

(a) Based on $\|x\|_\infty = |x_l|$, it is easy to see that a subgradient of $\|x\|_\infty$ is given by $[g_1(x)]_k = \text{sign}(x_k)\delta_{k,l}$, where $\delta_{k,l} = 1$ if $k = l$ and 0 otherwise. It is a subgradient and not a gradient, because there could be cases where the infinity norm is attained not only at index l but also at other indices. A subgradient of $\|x\|_1$ is given by $g_2(x) = \text{sign}(x)$.

(b) You can use a subgradient algorithm as

$$w^{(k+1)} = w^{(k)} - \alpha_k g^{(k)},$$

where

$$g^{(k)} = \begin{cases} X^T X g_1(X^T(y - Xw^{(k)})) & \text{if } \|w^{(k)}\|_1 \leq \epsilon \\ g_2(w^{(k)}) & \text{if } \|w^{(k)}\|_1 > \epsilon \end{cases}$$

(c) If we use c as cost, then we need the following two vector inequalities:

$$X^T(y - Xw) \le c1,$$
$$-X^T(y - Xw) \le c1,$$

where 1 is an all-one $N \times 1$ vector.

(d) We can rewrite the constraint as $1^T t \le \epsilon$ if we include the following two vector inequalities

$$w \le t,$$
$$-w \le t.$$

(e) The overall linear program can now be written as

$$\begin{aligned} & \min_{w,t,c} \; c \\ \text{subject to} \quad & X^T(y - Xw) \le c1, \\ & -X^T(y - Xw) \le c1, \\ & 1^T t \le \epsilon, \\ & w \le t, \\ & -w \le t. \end{aligned}$$