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

EE 4530 APPLIED CONVEX OPTIMIZATION

9 April 2021, 9:00–12:20
Block 2 (10:50-12:20)
Open book, strictly timed take-home exam. (Electronic) copies of the book and the course slides allowed. No other tools except a basic pocket calculator permitted.
Upload answers during 12:15–12:25

This block consists of three questions (25 points); more than usual, and this will be taken into account during grading. 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 4 (7 points)

Consider the function $f(x, y) = x^2 + x \cos y$ with $x, y \in \mathbb{R}$.

(a) Compute the gradient of $f(x, y)$.

(b) Compute the Hessian of $f(x, y)$.

(c) Is the function convex? Why or why not? Try to find the easiest way to prove this.

(d) In the point $x = 0, y = -\pi/2$, mathematically describe the set of all possible descent directions. Illustrate this also in a 2D plot. Which is the best descent direction?

(e) In the point $x = 0, y = -\pi/2$, is the Hessian positive definite, negative definite or neither? Can you relate this outcome to your plot (d)? How is such a point called?

Solution

(a) The gradient is

$$\nabla f(x, y) = \begin{bmatrix} 2x + \cos y \\ -x \sin y \end{bmatrix}$$

(b) The Hessian is

$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & -\sin y \\ -\sin y & -x \cos y \end{bmatrix}$$

(c) The function is not convex. The easiest way to see this is to derive the determinant of the Hessian which is given by $-2x \cos y - \sin^2 y$. If this determinant is negative, the matrix cannot be positive definite and hence the function is not convex. It is clear that $-2x \cos y - \sin^2 y$ can be negative. Take for instance $x = 1$ and $y = 0$.

(d) The gradient in this point is $\nabla f(0, -\pi/2) = [0, 0]^T$. All directions $[x, y]^T$ for which $[0, 0][x, y]^T = 0$ is negative is a descent direction. So no direction is a descent direction (if you say all directions are descent directions, that is also fine). So there is also no best direction. The point that was picked was a bit unfortunate and that is why the answer was kind of trivial. A plot would not reveal much in this case.

(e) The Hessian in this point is

$$\nabla^2 f(0, -\pi/2) = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}.$$

The determinant is negative and hence there is one positive and negative eigenvalue. So the matrix is neither positive nor negative definite. This means the point is a saddle point. This can also be observed from (d), since the gradient is zero.

Question 5 (9 Points)

Consider an imaging problem where we want to retrieve a series of images from a series of measurements. Every time instant k, we observe an m-dimensional measurement vector $y(k) \in \mathbb{R}^m$, which depends on the n-dimensional image vector $x(k) \in \mathbb{R}^n$ as

$$y(k) = Ax(k) + n(k),$$

where $A \in \mathbb{R}^{m \times n}$ is the $m \times n$ measurement matrix and $n(k) \in \mathbb{R}^m$ is an m-dimensional noise vector. Given A, the goal is to recover $x(k)$ from $y(k)$ and this for all time instants k.

The standard approach to tackle this would be to solve the following least squares problem for every time instant k:

$$\underset{x(k)}{\text{minimize}} \quad \|Ax(k) - y(k)\|_2^2$$

(a) In most imaging problems, we have $m < n$ and A has rank m. For this scenario, explain why this problem has multiple solutions. Mathematically describe the set of solutions.

(b) Which solution in the set described in (a) would you prefer? And why?

From the above, it clearly would be advantageous to exploit some constraints to narrow down the possible solutions. Therefore, it is often assumed that the image $x(k)$ is sparse and that the sparsity pattern does not change over time k (note that the values of $x(k)$ can still change over time k). In order to exploit this knowledge, it is beneficial to write the data model jointly for all time indices $k = 1, 2, \ldots, K$ as

$$Y = AX + N,$$

where $Y = [y(1), \ldots, y(K)]$, $X = [x(1), \ldots x(K)]$ and $N = [n(1), \ldots, n(K)]$. So the knowledge we have about the images is that the matrix X is sparse in the row dimension, i.e., only a few rows are non-zero. In other words, denoting the rows of X by the K-dimensional vectors $\tilde{x}(i) \in \mathbb{R}^K$, with $i = 1, 2, \ldots, n$, only a few vectors $\tilde{x}(i)$ will be non-zero. Assuming the upper bound on the number of non-zero rows is S, the imaging problem can thus be formulated as

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \|Y - AX\|_F^2 \\ \text{subject to} \quad & \#\ \text{nonzero}\ \tilde{x}(i) \leq S. \end{aligned}$$

(c) Mathematically formulate the constraint described in the above problem. To do this, make use of the n-dimensional vector $\chi = [\|\tilde{x}(1)\|_2, \ldots, \|\tilde{x}(n)\|_2]^T$.

(d) Is this problem convex? Why or why not? If not, can you relax the constraint such that it becomes a convex problem?

(e) Derive a first-order method to solve the convex problem in (d). To do this, first write the problem as an unconstrained problem $\|Y - AX\|_F^2 + \lambda f(X)$, where $f(X)$ is related to the constraint in (d) and λ is considered fixed.

Solution

(a) The matrix A has a right null space of dimension $n-m$ which can be characterized by a set of $n-m$ orthonormal basis vectors stacked in $V_n \in \mathbb{R}^{n\times(n-m)}$. In other words, we have that $AV_n = 0$. These can be computed from the SVD of A for instance. So if $x^*(k)$ is a solution to the problem, then it is clear that also $x^*(k) + V_n a$ is a solution where $a \in \mathbb{R}^{n-m}$ is an arbitrary vector. As a result, there is an infinite amount of solutions. A possible expression for $x^*(k)$ is given by the pseudo inverse of A times $y(k)$, i.e.,

$$x^*(k) = A^\dagger y(k) = A^T(AA^T)^{-1}y(k).$$

(b) From all the solutions, usually the solution with the minimal norm is preferred. From (a), we know that the set of solutions is given by $\{A^T(AA^T)^{-1}y(k) + V_n a | a \in \mathbb{R}^{n-m}\}$. Since V_n is orthogonal to A, the two terms describing this set are always orthogonal. So it is not possible to further minimize the norm of the first term. So the minimal norm solution is given by $x^*(k) = A^T(AA^T)^{-1}y(k)$. Note that this is the standard least squares solution.

(c) The constraint can easily be formulated using the zero norm of χ. More specifically, the constraint can mathematically be expressed as $\|\chi\|_0 \leq S$.

(d) This constraint is not convex because of the composition rule and the fact that the zero norm is not convex. We can relax this zero norm to a 1-norm, leading to $\|\chi\|_1 = \sum_{i=1}^n \|\tilde{x}(i)\|_2 \leq S$. This is sometimes called the 1-2 norm of the matrix X. Note in this context that we might have to scale the upper bound S depending on the energy of the row vectors.

(e) We can first write the problem as a LASSO type of problem where we bring the constraint into the cost function. This leads to

$$\underset{X}{\text{minimize}} \quad \|Y - AX\|_F^2 + \lambda \sum_{i=1}^n \|\tilde{x}(i)\|_2.$$

This becomes an unconstrained nondifferentiable convex optimization problem. So a simple subgradient descent algorithm can be considered. The gradient of the first term is well-known and this is given by $2A^TAX - 2A^TY$. A subgradient for the second term w.r.t. $\tilde{x}(i)$ is given by $g_i = \tilde{x}(i)/\|\tilde{x}(i)\|_2$ when $\tilde{x}(i) \neq 0$ and given by $g_i = 0$ when $\tilde{x}(i) = 0$. So for the full matrix X this subgradient is $G = [g_1, \ldots, g_n]^T = \text{diag}^\dagger(\chi)X$ where the pseudo-inverse $(\cdot)^\dagger$ only inverts the non-zero values and keeps the zero values. So the update equation becomes

$$X^{(k+1)} = X^{(k)} - \alpha_k[2A^TAX^{(k)} - 2A^TY + \lambda\text{diag}^\dagger(\chi^{(k)})X^{(k)}].$$

The step size α_k can follow some of the well-known step size rules.

Question 6 (9 points)

Consider the problem

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \le 0 \text{ holds for at least } k \text{ values of } i,
\end{array}
$$

with variable $x \in \mathbb{R}^n$, where the objective f_0 and the constraint functions f_i, $i = 1, \ldots, m$ (with $m \ge k$), are convex. We require that only k of the m constraints hold, instead of all m of them. In general this is a hard combinatorial problem, which requires solving all $\binom{m}{k}$ convex problems obtained by choosing subsets of k constraints, and selecting the one with the smallest objective value. In this question, we explore some ways to relax or restrict this problem into a convex problem.

(a) Can you think of a way to mathematically formulate the constraint as a single inequality by making use of the sign function, where $\operatorname{sign}(u) = 1$ if $u > 0$ and $\operatorname{sign}(u) = -1$ if $u \le 0$?

(b) Is this a convex constraint? If not, can you relax the constraint to a convex one?

(c) The relaxation in (b) is often too harsh. A more commonly used constraint is

$$
\sum_{i=1}^{m} (1 + y f_i(x))_+ \le m - k
$$

with $(u)_+ = \max\{0, u\}$ and the new variable $y > 0$. Show that this constraint guarantees that $f_i(x) \le 0$ holds for at least k values of i.
Hint: For each $u \in \mathbb{R}$, you can use that $(1 + yu)_+ \ge 1$ when $u > 0$ and $(1 + yu)_+ \ge 0$ when $u \le 0$.

Consider now the problem

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & \displaystyle\sum_{i=1}^{m} (1 + y f_i(x))_+ \le m - k \\
& y > 0,
\end{array}
$$

with variables x and y. This is a restriction of the original problem. In other words, if (x, y) is feasible for it, then x is feasible for the original problem. The goal is now to formulate this problem as a convex optimization problem

(d) As a first step, consider the epigraph form of the functions $(1 + y f_i(x))_+$ by introducing the new variables t_i. This allows you to rewrite the constraint $\sum_{i=1}^{m} (1 + y f_i(x))_+ \le m - k$ as a set of new constraints. Write down the resulting problem.

(e) Now, replace the variable y by $z = 1/y$ and the variables t_i by $u_i = (t_i - 1)z$. Show that the overal problem in the variables x, z and u_i then is convex.

Solution

(a) Based on the definition of the sign function, it is clear that the constraint is equivalent to saying that $\mathrm{sign}(f_i(x)) = -1$ for at least k values of i or equivalently $\mathrm{sign}(f_i(x)) = 1$ for at most $m - k$ values of i. This can be mathematically formulated as

$$\sum_{i=1}^{m} 1 + \mathrm{sign}(f_i(x)) \leq 2(m-k).$$

(b) Because of the sign function, this clearly is not a convex constraint. There are multiple ways to relax the function to a convex one. One option is to simply relax $1 + \mathrm{sign}(u)$ by $(u)_+$ [see (c) for a definition]. This will then lead to

$$\sum_{i=1}^{m} (f_i(x))_+ \leq 2(m-k).$$

Or alternatively, $1 + \mathrm{sign}(u)$ can be relaxed as $(1+u)_+$, leading to

$$\sum_{i=1}^{m} (1 + f_i(x))_+ \leq 2(m-k).$$

(c) To show that $\sum_{i=1}^{m} (1 + y f_i(x))_+ \leq m - k$ implies that $f_i(x) \leq 0$ holds for at least k values of i, we will use a proof by contradiction. In other words, we will show that if $f_i(x) \leq 0$ holds for less than k values of i (or $f_i(x) > 0$ holds for more than $m-k$ values of i), then $\sum_{i=1}^{m} (1 + y f_i(x))_+ > m - k$.

To prove this, suppose for simplicity that $f_i(x) > 0$ for $i = 1, \ldots, l$ and $f_i(x) \leq 0$ for $i = l+1, \ldots, m$, with $l > m - k$. Then using the hint it is clear that

$$\sum_{i=1}^{m} (1 + y f_i(x))_+ = \sum_{i=1}^{l} (1 + y f_i(x))_+ + \sum_{i=l+1}^{m} (1 + y f_i(x))_+ \geq l + 0 > m - k.$$

(d) We first transform this problem into its epigraph form by introducing the additional variables t_i for $i = 1, \ldots, m$. This leads to

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \displaystyle\sum_{i=1}^{m} t_i \leq m - k \\ & t_i \geq 0 \\ & t_i \geq 1 + y f_i(x) \\ & y > 0, \end{array}$$

(e) Introducing the change of variables $z = 1/y$ and $u_i = (t_i - 1)z$ we obtain the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & kz + \displaystyle\sum_{i=1}^{m} u_i \leq 0 \\ & u_i + z \geq 0 \\ & u_i - f_i(x) \geq 0 \\ & z > 0. \end{array}$$

Note that the last constraint can be replaced by $z \geq \epsilon$ with ϵ a small postive error in order to obtain all regular inequalities. All constraints are now clearly convex in the variables and hence the overall problem is convex.