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

EE 4530 APPLIED CONVEX OPTIMIZATION

20 January 2021, 13:30–16:50
Block 1 (13:30-15:00)
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 14:55–15:05

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.

1 Question 1 (9 points)

For each of the following 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) $f(x) = \|x\|_{1/2}$ with $\mathrm{dom} f = \mathbf{R}^n$;

(b) $f(x) = \sqrt{\|x\|_2}$ with $\mathrm{dom} f = \mathbf{R}^n$;

(c) $f(x) = -\max_j \sqrt{x_j}$ with $\mathrm{dom} f = \mathbf{R}^n$;

(d) $f(x) = -\min_i a_i^T x$ with $\mathrm{dom} f = \mathbf{R}^n$;

(e) $f(X) = \lambda_{\min}(X)$ with $\lambda_{\min}(\cdot)$ indicating the minimum eigenvalue and $\mathrm{dom} f = \mathbf{S}^n$.

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

(g) Consider the set $\mathcal{C} = \{x \in \mathbb{R}^2 | x^\top P x \leq 1\}$ for some $P \in \mathbb{S}^n_{++}$.

1) Define $\mathcal{B} = \{x \in \mathbb{R}^2 | \|x\|_1 \leq 1\}$. Sketch the set $\mathcal{D} = \mathcal{C} \setminus \mathcal{B}$, with $P = I$ identity matrix. Is the set $\mathcal{D}$ convex in $\mathbb{R}^2$? Is it convex on the positive part of $\mathbb{R}^2$, i.e. $x \succeq 0$? *(Don't prove it, just draw your conclusion by inspection of your sketch)*

2) Define $\mathcal{C}_1 = \mathcal{C}$ with $P = I$, and $\mathcal{C}_2 = \mathcal{C}$ with $P = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$. Sketch the set $\mathcal{C}_1 \bigcap \mathcal{C}_2$: is it convex in $\mathbb{R}^2$? *(Don't prove it, just draw your conclusion by inspection of your sketch)*

Solution

(a) Neither convex nor concave. For $0 < p < 1$ it is not a norm, does not satisfy the triangle inequality.

(b) Neither convex nor concave. Notice here that the domain is all $\mathbb{R}^N$. As analogy consider $f(x) = \sqrt{|x|}$, for $x \in \mathbb{R}$.

(c) Neither convex nor concave. It is the pointwise maximum of concave functions.

(d) It is convex. It is the pointwise minimum of concave functions (linear functions are either convex and concave), with switched sign.

(e) It is concave. The infimum of any family of linear functions is concave.

(f) For $\gamma \geq 0$ and $p \geq 1$.

(g.1) No it is not convex. It is the set difference between a unit norm ℓ_2 ball and the unit ℓ_1 ball. It is convex in $\mathbb{R}^2_+$.

(g.2) Yes, it is the intersection of the unit Euclidean norm ball and an ellipsoid with semi axis 1 and $\frac{1}{2}$.

2 Question 2 (8 points)

Consider the function

$$f(x_1, x_2) = 2x_1^2 + 6x_2^2 + 2x_1x_2 + 2x_1 + 3x_2 + 3.$$

(a) Write this function in the form

$$f(x) = \frac{1}{2}x^\top Px + bx + c,$$

Is this function convex or not? Why?

(b) Derive the steepest descent direction related to the norm $\|x\|_P$. *To solve this you need to minimize $v^T\nabla f(x)$ subject to $\|v\|_P = 1$ using the Lagrangian. Give all steps of the derivation.*

(c) Consider a descent algorithm based on the direction derived in (b). To which type of descent algorithm does this correspond?

Consider now a regularized version of the above function with an ℓ_1 term, i.e., consider:

$$\min_x \; \frac{1}{2}x^\top Px + bx + c + \lambda\|x\|_1$$

(d) What is the regularization term used for? Write the optimality condition(s).

(e) Compute and plot the subdifferential of the second term of the function, for a fixed λ. *It is enough to compute the subgradient with respect to only a single scalar variable x_j*

(f) Perform one iteration of subgradient descent starting from the point (1,1). Does the function value decrease or not by taking such a step? Do you expect it always to decrease?

Solution

(a) $P = \begin{bmatrix} 4 & 2 \\ 2 & 12 \end{bmatrix}$. The function is convex because P is positive definite.

(b) By forming the Lagrangian and using the KKT conditions, we find the normalized steepest descent direction $v^\star = -\frac{P^{-1}\nabla f(x)}{\|\nabla f(x)\|_{P^{-1}}}$. Then, the steepest descent direction (i.e. the un-normalized) is $\Delta x_{sd} = -P^{-1}\nabla f(x)$.

(c) Because P is also the Hessian of the function $f(x)$, this descent algorithm corresponds to the Newton method.

(d) It is used to promote sparsity of the solution. The condition is $0 \in \nabla f(x) + \lambda\partial\|x\|_1$.

(e) [figure] The subdifferential is the set $\partial|x_i| = \begin{cases} \lambda & x_i > 0 \\ -\lambda & x_i < 0 \\ [-\lambda, \lambda] & x_i = 0 \end{cases}$

(f) Subgradient descent is not a descent method, thus the cost function value could increase. To perform one iteration of subgradient descent, you need to compute a subgradient of the ℓ_1 norm (at this point you should already have computed the gradient of the differentiable part of the function). Remember that $h(x) = \|x\|_1 = \sum_i \underbrace{|x_i|}_{h_i(x)}$. Then exploiting the summation rule $\partial h(x) = \partial \sum_i h_i(x) = \sum_i \partial h_i(x)$ we have:

$$\sum_{i:x_i \neq 0} \operatorname{sgn}(x_i)\, \boldsymbol{e}_i \in \partial h(\boldsymbol{x}).$$

Remember that the variable is a vector, and as such is the (sub)gradient, reason why the basis vectors $\boldsymbol{e}_i$ are in the above expression.

A subgradient iteration is then:

$$x^1 = x^0 - \alpha g^0,$$

with $g^0 \in \nabla f(x^0) + \partial h(x^0)$.

3 Question 3 (8 Points)

Consider the following problem:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{m}(a_i^\top x - b_i)^2 \tag{1}$$

where $a_i \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$.

(a) Is the problem convex? Argue why and find the optimal solution $x^\star$.

Consider now a regularized version of the above problem:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{m}(a_i^\top x - b_i)^2 + \lambda \sum_{i=1}^{n} x_i^2 \tag{2}$$

(b) Is the problem convex? Argue why and find the optimal solution $x^\star$.

(c) What can you say about the difference or similarity between the solutions of problems (1) and (2)? When can the second formulation be of help?

Another variant of the problem is by considering:

$$\underset{x}{\text{minimize}} \quad \frac{1}{2}\sum_{i=1}^{m}(a_i^\top x - b_i)^2 \quad \text{subject to} \quad \sum_{i=1}^{n} x_i^2 \le 1 \tag{3}$$

(d) Which category of canonical optimization problems does the one above belong to (e.g. linear, etc.)? Sketch some level curves and the constraint set of this general category of problems.

(e) Write the update formula related to problem (3) in case of projected gradient descent.

(f) Write the KKT conditions and the dual problem for (3). Can problems (2) and (3) have the same solution? If yes, when? If not, why?

Solution

(a) Yes. The cost function can be written in vector notation as $\frac{1}{2}\|Ax - b\|^2$, i.e., in a least square formulation. Expanding the function, we have $\frac{1}{2}[x^\top A^\top Ax - 2b^\top Ax + b^\top b]$. Because $A^\top A \succeq 0$, the function is convex. The optimal solution is found by setting $\nabla f(x) = 0$, that is $x^\star = A^\dagger b$, where $A^\dagger = (A^\top A)^{-1}A^\top$.

(b) Yes, as long as $\lambda \ge 0$. In this case, the optimal solution is given by $x^\star = (A^\top A + 2\lambda I)^{-1}A^\top b$.

(c) When the matrix to invert, i.e. $A^\top A$ is ill-conditioned, i.e., it has a high condition number or it is nearly singular. By adding the positive definite matrix $2\lambda I$, with $\lambda \ge 0$, the overall inverse benefits.

(d) QCQP. Either the level curves of the function and the constraint set are ellipsoids.

(e) Is it a descent step and a projection in the Euclidean unit norm ball. So if the new point is inside the ball, nothing changes, while if it is outside of the ball, the projection is found by scaling the vector for its ℓ_2 norm.

(f) Problem (2) is a particular case of (3). They can be equivalent for a particular choice of λ. In particular, by obtaining the optimal $\lambda^\star$ from the dual problem of (3) and plugging it into (2), the two problems are equivalent, i.e. they have the same solution $x^\star$.