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 2 (15:20-16:50)
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 16:45–16:55

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 4 (7 points)

Consider the function $f(x) = (2x_1 + x_2^2)^2$.

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

(b) At the point $s = [0,1]^T$, we consider the search direction $d = [1,-2]^T$. Show that d is a descent direction.

(c) Assume we do a line search from the point s in the direction of d. Find the stepsize α that minimizes $f(s + \alpha d)$. For the optimal α, report the value $f(s + \alpha d)$.

(d) Compute the Hessian of $f(x)$.

(e) Run one Newton's step with fixed stepsize $t = 1$ starting from $s = [0,1]^T$. What is the resulting $f(x)$? Does this procedure outperform the method considered in (c)?

Solution

(a) The gradient is

$$\nabla f(x) = \begin{bmatrix} 8x_1 + 4x_2^2 \\ 8x_1x_2 + 4x_2^3 \end{bmatrix}$$

(b) In the point $s = [0,1]^T$, we have $\nabla f(s) = [4,4]^T$. Hence, $\nabla f(s)^T d = 4 - 8 = -4 < 0$ and thus d is a descent direction.

(c) We need to find the α that minimizes

$$f(s+\alpha d) = f([\alpha, 1-2\alpha]^T) = [2\alpha + (1-2\alpha)^2]^2 = (2\alpha + 1 - 4\alpha + 4\alpha^2)^2$$
$$= (1 - 2\alpha + 4\alpha^2)^2 = [(2\alpha - 1/2)^2 + 3/4]^2.$$

Minimizing the square is like mimimizing $(2\alpha - 1/2)^2 + 3/4$ since it is always positive. Clearly, $\alpha = 1/4$ is the minimizer. The final function value is $(3/4)^2 = 9/16$. Note that the initial function value was $f(s) = 1$ se we clearly decreased the function.

(d) The Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} 8 & 8x_2 \\ 8x_2 & 8x_1 + 12x_2^2 \end{bmatrix}.$$

(e) First of all, we compute the Hessian in the point s:

$$\nabla^2 f(s) = \begin{bmatrix} 8 & 8 \\ 8 & 12 \end{bmatrix}.$$

The Hessian-based descent direction is then given by

$$d_H = -(\nabla^2 f(s))^{-1} \nabla f(s) = -\frac{1}{8} \begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

So for step size $t = 1$, the function value is

$$f(s + d_H) = f([0,1]^T + [-0.5, 0]^T) = f([-0.5, 1]^T) = (-1+1)^2 = 0.$$

This is clearly smaller than the one obtained by the first-order method we used before, eventhough the optimal step size was used over there. This is because the Hessian-based descent direction is better.

2 Question 5 (9 Points)

Suppose you want to solve a set of noisy linear equations with binary unknowns. Such a problem can be solved by Boolean least squares, which can be formulated as

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i \in \{-1, 1\} \quad i = 1, 2, \ldots, n. \end{array}$$

Here, the unknown variable is $x \in \mathbb{R}^n$, whereas $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given. This is a basic problem arising, for instance, in digital communications. A brute force solution is to check all 2^n possible values of x, which is usually impractical.

(a) Show that the problem is equivalent to

$$\begin{array}{ll} \text{minimize} & \operatorname{trace}(A^T A X) - 2b^T A x + b^T b \\ \text{subject to} & X = xx^T \\ & X_{i,i} = 1 \quad i = 1, 2, \ldots, n \end{array}$$

in the variables X and x. Explain why this is not a convex problem.

(b) Approximate the problem of (a) by a semi-definite program. Do this by relaxing $X = xx^T$ to $X \succeq xx^T$ and using the following property:

$$\text{If} \quad C \succ 0 \quad \text{then} \quad \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad A - BC^{-1}B^T \succeq 0$$

Show the resulting problem and explain why it is a semi-definite program.

(c) Describe a first-order method to solve the convex problem in (b). Also present the related equations.

(d) Another way to convexify the Boolean least squares problem is by relaxing the Boolean constraints $x_i \in \{-1, 1\}$ by their convex hull. Show how this convex problem looks like.

(e) Describe a first-order method to solve the convex problem in (d). Also present the related equations.

Solution

(a) Plugging in the first constraint into the cost function as well as the third constraint, and thereby eliminating the matrix variable X, we obtain the original problem. This is not convex because of the first equality constraint which is quadratic and not affine.

(b) From the shown property, $X \succeq xx^T$ is equivalent to

$$\begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0.$$

This is a linear matrix inequality. The other constraints as well as the cost are affine. That is what we need to obtain a semi-definite program.

(c) We can run a projected gradient descent algorithm, where to project on the intersection of the positive semi-definite cone and the affine constraint we run an alternating projections onto convex sets algorithm. Starting from the kth iteration with iterates $\tilde{X}^{(k)}$ and $\tilde{x}^{(k)}$, we start with the gradient descent step:

$$\begin{aligned} X^{(k+1)} &= \tilde{X}^{(k)} - \alpha_k A^T A, \\ x^{(k+1)} &= \tilde{x}^{(k)} + \alpha_k A^T b. \end{aligned}$$

Then we carry out the projection on the intersection of the two constraint sets:

$$(\tilde{X}^{(k+1)}, \tilde{x}^{(k+1)}) = \Pi(X^{(k+1)}, x^{(k+1)}).$$

This latter is computed by an alternating projection algorithm using the following two projection functions:

$$\Pi_{PSD}\left(\begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix}\right) \overset{SVD}{=} \Pi_{PSD}\left(\sum_{i=1}^{n+1} \lambda_i u_i u_i^T\right) = \sum_{i=1}^{n+1} \max(0, \lambda_i) u_i u_i^T = \begin{bmatrix} \tilde{X} & \tilde{x} \\ \tilde{x}^T & 1 \end{bmatrix},$$

$$\Pi_{diag}(X) = \tilde{X} \quad \text{with} \quad \tilde{X}_{i,j} = \begin{cases} X_{i,j} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

(d) This problem looks like

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|^2 \\ \text{subject to} & x_i \in [-1, 1] \quad i = 1, 2, \ldots, n. \end{array}$$

(e) Here also, we can run a projected gradient descent agorithm:

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

where the gradient is

$$g^{(k)} = 2A^T A x^{(k)} - 2A^T b,$$

and the projection function $\Pi(\cdot)$ is defined as follows

$$\Pi(x) = \begin{cases} x & \text{if } x \in [-1, 1] \\ \text{sign}(x) & \text{if } x \notin [-1, 1] \end{cases}$$

3 Question 6 (9 points)

Consider the convex optimization problem

$$\begin{array}{ll} \text{minimize} & -c^T x + \sum_{i=1}^{m} y_i \log y_i \\ \text{subject to} & x \succeq 0 \\ & Px = y \\ & 1^T x = 1 \end{array}$$

with variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. The matrix $P \in \mathbb{R}^{m \times n}$ is such that $P^T 1 = 1$, i.e., the columns of P sum up to 1.

(a) Give the KKT conditions of this problem.

(b) Derive the dual function of this problem.

(c) What is the dual problem? Try to limit the number of dual variables involved in the problem.

(d) Now simplify the inequality constraint in this dual problem by using the fact that $P^T 1 = 1$ and introducing a new dual variable w to replace the dual variable related to $Px = y$. Furthermore, eliminate the dual variable related to $1^T x = 1$ by optimizing for it. Show that all these steps lead to the following simplified dual problem in the variable w:

$$\begin{array}{ll} \text{maximize} & -\log\left(\sum_{i=1}^{m} \exp w_i\right) \\ \text{subject to} & P^T w \succeq c \end{array}$$

Solution

(a) Assuming functions are taken element wise, the Lagrangian of this problem is

$$L(x, y, \lambda, \nu, \mu) = -c^T x + y^T \log y - \lambda^T x + \nu^T (Px - y) + \mu(1^T x - 1),$$

where $\lambda \succeq 0$. The first set of KKT conditions corresponds to the primal constraints. The second one is $\lambda \succeq 0$. The third set of KKT conditions corresponds to the complementary slackness conditions given by $\lambda_i x_i = 0$ for $i = 1, 2, \ldots, n$. And finally, the fourth set of KKT conditions equates the gradient of the Lagrangian with respect to x and y to zero:

$$\frac{\partial L}{\partial x} = -c - \lambda + P^T \nu + \mu 1 = 0 \;\Rightarrow\; \lambda = P^T \nu + \mu 1 - c, \tag{1}$$

$$\frac{\partial L}{\partial y} = \log y + 1 - \nu = 0 \;\Rightarrow\; \log y = \nu - 1 \Rightarrow y = \exp(\nu - 1). \tag{2}$$

(b) It is clear that the minimum of the Lagrangian is $-\infty$ when (1) is not satisfied. If (1) is satisfied, the linear term in x of the Lagrangian is always 0 and the minimum for y is reached at (2). This gives the following dual function:

$$g(\lambda, \nu, \mu) = \begin{cases} -1^T \exp(\nu - 1) - \mu & \text{if } \lambda = P^T \nu + \mu 1 - c \\ -\infty & \text{otherwise} \end{cases}$$

(c) For the dual problem we have to maximize this dual function under the constraint that $\lambda \succeq 0$. So we can easily eliminate λ in the dual problem, leading to

$$\underset{\nu,\mu}{\text{maximize}} \quad g(\nu, \mu),$$

where

$$g(\nu, \mu) = \begin{cases} -\mathbf{1}^T \exp(\nu - 1) - \mu & \text{if } P^T \nu + \mu \mathbf{1} \succeq c \\ -\infty & \text{otherwise} \end{cases}$$

So we obtain the dual problem

$$\begin{array}{ll} \underset{\nu,\mu}{\text{maximize}} & -\mathbf{1}^T \exp(\nu - 1) - \mu \\ \text{subject to} & P^T \nu + \mu \mathbf{1} \succeq c \end{array}$$

(d) Using the fact that $P^T \mathbf{1} = \mathbf{1}$, we can replace $P^T \nu + \mu \mathbf{1} \succeq c$ by $P^T(\nu + \mu \mathbf{1}) \succeq c$. Now replacing ν with the variable $w = \nu + \mu \mathbf{1}$, the dual problem becomes

$$\begin{array}{ll} \underset{w,\mu}{\text{maximize}} & -\mathbf{1}^T \exp(w - \mu \mathbf{1} - 1) - \mu \\ \text{subject to} & P^T w \succeq c \end{array}$$

Or equivalently

$$\begin{array}{ll} \underset{w,\mu}{\text{maximize}} & -\exp(-\mu - 1)\mathbf{1}^T \exp w - \mu \\ \text{subject to} & P^T w \succeq c \end{array}$$

Now we can elminate μ by setting the derivative of the cost w.r.t. μ to zero, since μ is not involved in a constraint. This gives us

$$\exp(-\mu - 1)\mathbf{1}^T \exp w - 1 = 0 \ \Rightarrow \ \mu = \log(\mathbf{1}^T \exp w) - 1.$$

So finally, the dual problem becomes

$$\begin{array}{ll} \underset{w}{\text{maximize}} & -\log(\mathbf{1}^T \exp w) \\ \text{subject to} & P^T w \succeq c \end{array}$$