HOMEWORK
David F. Gleich
February 20, 2023

Remember that this is a graduate class. There may be elements of the problem statements that require you to fill in appropriate assumptions. You are also responsible for determining what evidence to include. An answer alone is rarely sufficient, but neither is an overly verbose description required. Use your judgement to focus your discussion on the most interesting pieces. The answer to "should I include 'something' in my solution?" will almost always be: Yes, if you think it helps support your answer.

Problem 0: Homework checklist

- Please identify anyone, whether or not they are in the class, with whom you discussed your homework. This problem is worth 1 point, but on a multiplicative scale.
- Make sure you have included your source-code and prepared your solution according to the most recent Piazza note on homework submissions.

Problem 1: Steepest descent

(Nocedal and Wright, Exercise 3.6) Let's conclude with a quick problem to show that steepest descent can converge very rapidly! Consider the steepest descent method with exact line search for the function $f(\mathbf{x}) = (1/2)\mathbf{x}^T\boldsymbol{Q}\mathbf{x} - \mathbf{x}^T\mathbf{b}$. Suppose that we know $\mathbf{x}_0 - \mathbf{x}^*$ is parallel to an eigenvector of $\boldsymbol{Q}$. Show that the method will converge in a single iteration.

First, we can take the derivative of the function and derive its gradient: $\nabla f(\mathbf{x}) = \boldsymbol{Q}\mathbf{x} - \mathbf{b}$. Since this is a convex optimization problem, $\mathbf{x}^*$ is the globally minimal point, we can denote:

$$\nabla f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}^*} = \boldsymbol{Q}\mathbf{x}^* - \mathbf{b} = 0 \tag{1}$$

Let $\mathbf{v}$ be one eigenvector of $\boldsymbol{Q}$, then we have $\mathbf{x}_0 - \mathbf{x}^* = k\mathbf{v}$. Starting from $\mathbf{x}_0$, we can denote the steepest descent as:

$$min_\alpha f(\mathbf{x}_0 - \alpha(\boldsymbol{Q}\mathbf{x}_0 - b) = min_\alpha (1/2)\alpha^2(\boldsymbol{Q}\mathbf{x}_0 - b)^T\boldsymbol{Q}(\boldsymbol{Q}\mathbf{x}_0 - b) - \alpha(\boldsymbol{Q}\mathbf{x}_0 - b)^T(\boldsymbol{Q}\mathbf{x}_0 - b) \tag{2}$$

Obviously the step length α is minimized at$\alpha = \frac{(\boldsymbol{Q}\mathbf{x}_0 - b)^T(\boldsymbol{Q}\mathbf{x}_0 - b)}{(\boldsymbol{Q}\mathbf{x}_0 - b)^T\boldsymbol{Q}(\boldsymbol{Q}\mathbf{x}_0 - b)}$

Then let's substitute α into $\mathbf{x}_1 = \mathbf{x}_0 - \nabla f(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}$, we have

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{x}_0 - \alpha(\boldsymbol{Q}\mathbf{x}_0 - b) \\
&= x^* + k\mathbf{v} - \frac{(\boldsymbol{Q}\mathbf{x}^* + k\boldsymbol{Q}\mathbf{v} - b)^T(\boldsymbol{Q}\mathbf{x}^* + k\boldsymbol{Q}\mathbf{v} - b)}{(\boldsymbol{Q}\mathbf{x}^* + k\boldsymbol{Q}\mathbf{v} - b)^T\boldsymbol{Q}(\boldsymbol{Q}\mathbf{x}^* + k\boldsymbol{Q}\mathbf{v} - b)}(\boldsymbol{Q}\mathbf{x}^* + k\boldsymbol{Q}\mathbf{v} - b) \\
&= x^* + k\mathbf{v} - k * \frac{(\boldsymbol{Q}\mathbf{v})^T(\boldsymbol{Q}\mathbf{v})}{(\boldsymbol{Q}\mathbf{v})^T\boldsymbol{Q}(\boldsymbol{Q}\mathbf{v})} \\
&= x^* + k\mathbf{v} - k * \frac{(\lambda\mathbf{v})^T(\lambda\mathbf{v})\boldsymbol{Q}}{(\lambda\mathbf{v})\boldsymbol{Q}(\lambda\mathbf{v})}\mathbf{v} \\
&= x^*
\end{aligned}$$

(3)

Thus we have shown that this optimization problem will converge in a single iteration with steepest gradient descent.

Problem 2: LPs in Standard Form

Show that we can solve:

$$\text{minimize} \quad \sum_i \max(\mathbf{a}_i^T\mathbf{x} - b_i, -0.5)$$

by constructing an LP in standard form.

We can reformulate this problem as a constrained problem:

$$min_{\mathbf{x}} \sum m \quad s.t. \;\; m_i \geq \mathbf{a}_i^T - b_i, m_i \geq -0.5 \tag{4}$$

by introducing the slack variable, we can construct an LP in standard form:

$$min_{\mathbf{x},\mathbf{s}} \sum m \quad s.t. \;\; \mathbf{l} \geq \boldsymbol{A}\mathbf{x} - \mathbf{b} + \mathbf{s}, m \geq -0.5, \mathbf{s} \geq 0 \tag{5}$$

Problem 3: Duality

Show that the these two problems are dual by showing the equivalence of the KKT conditions:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & \boldsymbol{A}\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq 0 \end{array} \quad \text{and} \quad \begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & \mathbf{b}^T\boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{A}^T\boldsymbol{\lambda} \leq \mathbf{c}, \boldsymbol{\lambda} \geq 0 \end{array}.$$

The KKT conditions for the first problem is:

$$\begin{aligned} & c - A^T\lambda - \mu = 0 \\ & b - Ax \leq 0 \\ & x \geq 0 \\ & \mu X = 0 \\ & \lambda^T(Ax - b) = 0 \\ & \lambda \geq 0 \\ & \mu \geq 0 \end{aligned} \tag{6}$$

The KKT conditions for the second problem is:

$$\begin{aligned} & Ax - b - \mu = 0 \\ & A^T\lambda - c \leq c \\ & \lambda \geq 0 \\ & x^T(c - A^T\lambda) = 0 \\ & \mu\lambda = 0 \\ & x \geq 0 \\ & \mu \geq 0 \end{aligned} \tag{7}$$

Clearly the KKT conditions for the two problems are equivalent. Thus these two problems are dual.

Problem 4: Geometry of LPs

(Griva, Sofer, and Nash, Problem 3.12) Consider the system of constraints $\boldsymbol{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0$ with

$$\boldsymbol{A} = \begin{bmatrix} 1 & 4 & 7 & 1 & 0 & 0 \\ 2 & 5 & 8 & 0 & 1 & 0 \\ 3 & 6 & 9 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} 12 \\ 15 \\ 18 \end{bmatrix}$$

Is $\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}^T$ a basic feasible point? Explain your answer precisely in terms of the definition.

$\mathbf{x}$ is not a basic feasible point, since the matrix

$$\boldsymbol{A}[:, 1:3] = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \tag{8}$$

is not invertible.

Problem 5: Using the geometry

(Griva, Sofer, and Nash, Section 4.3, problem 3.13. Suppose that a linear program originally included a free variable x_i where there were no upper-and-lower bounds on its values. As we described in class, this can be converted into a pair of variables x_i^+ and x_i^- such that $x_i^+, x_i^- \geq 0$ and x_i is replaced with the difference $x_i^+ - x_i^-$. Prove that a basic feasible point can have only one of x_i^+ or x_i^- different from zero. (Hint: this is basically a one-line proof once you see the right characterization. I would suggest trying an example.)

Let's assume that a basic feasible point can have both x_i^+ or x_i^- different from zero. Then the constraint matrix $(a_1, a_2, ..., a_n, 0, 0$ must satisfy $(a_{n1}, a_{n2}, ..., a_{nm}, 0, 0$ is nonsingular, where a_{ni} are the columns corresponding to the non-zero entries of x. This contradicts with the fact that the columns of B should be linearly independent. Therefore it is proved that a basic feasible point can have only one of x_i^+ or x_i^- different from zero.