EE/M 520, Spring 2005
Exam 2: Due start of Session 26

Solutions (version: April 18, 2005, 15:15) 75 mins.; Total 50 pts.

1. (11 pts.) Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $m \geq n$, rank $\boldsymbol{A} = n$, $\boldsymbol{b}_1, \dots, \boldsymbol{b}_p \in \mathbb{R}^m$, and $\alpha_1, \dots, \alpha_p \in \mathbb{R}$, consider the problem

$$\text{minimize} \quad \alpha_1\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_1\|^2 + \alpha_2\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_2\|^2 + \cdots + \alpha_p\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_p\|^2. \tag{1}$$

Suppose that $\boldsymbol{x}_i^*$ is the solution to the problem

$$\text{minimize}\, \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_i\|^2,$$

where $i = 1, \dots, p$. Assuming that $\alpha_1 + \cdots + \alpha_p > 0$, derive a simple expression for the solution to (1) in terms of $\boldsymbol{x}_1^*, \dots, \boldsymbol{x}_p^*$ and $\alpha_1, \dots, \alpha_p$.

Ans.: Write

$$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}_i\|^2 = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b}_i + \|\boldsymbol{b}_i\|^2$$

Therefore, the given objective function can be written as

$$(\alpha_1 + \cdots + \alpha_p)\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T(\alpha_1\boldsymbol{b}_1 + \cdots + \alpha_p\boldsymbol{b}_p) + \alpha_1\|\boldsymbol{b}_1\|^2 + \cdots + \alpha_p\|\boldsymbol{b}_i\|^2.$$

The solution is therefore (by inspection)

$$\boldsymbol{x}^* = ((\alpha_1 + \cdots + \alpha_p)\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T(\alpha_1\boldsymbol{b}_1 + \cdots + \alpha_p\boldsymbol{b}_p) = \frac{1}{\alpha_1 + \cdots + \alpha_p}\sum_{i=1}^{p}\alpha_i\boldsymbol{x}_i^* = \sum_{i=1}^{p}\beta_i\boldsymbol{x}_i^*,$$

where $\beta_i = \alpha_i/(\alpha_1 + \cdots + \alpha_p)$.

Note that the original problem can be written as the least squares problem

$$\text{minimize}\, \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2,$$

where

$$\boldsymbol{b} = \frac{\alpha_1\boldsymbol{b}_1 + \cdots + \alpha_p\boldsymbol{b}_p}{\alpha_1 + \cdots + \alpha_p}.$$

2. (12 pts.) Consider the problem

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \\ \text{subject to} & \|\boldsymbol{x}\|^2 = 1, \end{array}$$

where $\boldsymbol{Q} = \boldsymbol{Q}^T > 0$.

a. Using the penalty function $P(\boldsymbol{x}) = (\|\boldsymbol{x}\|^2 - 1)^2$ and penalty parameter γ, write down an unconstrained optimization problem whose solution $\boldsymbol{x}_\gamma$ approximates the solution to the above problem.

b. Show that for any γ, $\boldsymbol{x}_\gamma$ is an eigenvector of $\boldsymbol{Q}$.

c. Show that $\|\boldsymbol{x}_\gamma\|^2 - 1 = O(1/\gamma)$ as $\gamma \to \infty$.

Ans.: a. The unconstrained problem based on the given penalty function is

$$\text{minimize}\, \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \gamma(\|\boldsymbol{x}\|^2 - 1)^2.$$

b. By the FONC, $\boldsymbol{x}_\gamma$ satisfies

$$2\boldsymbol{Q}\boldsymbol{x}_\gamma + 4\gamma(\|\boldsymbol{x}_\gamma\|^2 - 1)\boldsymbol{x}_\gamma = 0.$$

Rearranging, we obtain

$$\boldsymbol{Q}\boldsymbol{x}_\gamma = 2\gamma(1 - \|\boldsymbol{x}_\gamma\|^2)\boldsymbol{x}_\gamma = \lambda_\gamma \boldsymbol{x}_\gamma,$$

where λ_γ is a scalar. Hence, $\boldsymbol{x}_\gamma$ is an eigenvector of $\boldsymbol{Q}$.

c. Now, $\lambda_\gamma = 2\gamma(1 - \|\boldsymbol{x}_\gamma\|^2) \leq \lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{Q}$. Hence, $\|\boldsymbol{x}_\gamma\|^2 - 1 = -\lambda_{\max}/(2\gamma) = O(1/\gamma)$ as $\gamma \to \infty$.

3. (14 pts.) Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \in \Omega \end{array}$$

where $f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x}$ and $\boldsymbol{c} \in \mathbb{R}^n$ is a given nonzero vector. (Linear programming is a special case of this problem.) We wish to apply a fixed step-size projected gradient algorithm

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{\Pi}[\boldsymbol{x}^{(k)} - \nabla f(\boldsymbol{x}^{(k)})],$$

where, as usual, $\boldsymbol{\Pi}$ is the projection operator onto Ω (assume that for any $\boldsymbol{y}$, $\boldsymbol{\Pi}[\boldsymbol{y}] = \arg\min_{\boldsymbol{x}\in\Omega} \|\boldsymbol{y} - \boldsymbol{x}\|^2$ is unique).

a. Suppose that for some k, $\boldsymbol{x}^{(k)}$ is a global minimizer of the given problem. Is it necessarily the case that $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)}$? Explain fully.

b. Suppose that for some k, $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)}$. Is it necessarily the case that $\boldsymbol{x}^{(k)}$ is a local minimizer of the given problem? Explain fully.

Ans.: a. Yes. To show: Suppose that $\boldsymbol{x}^{(k)}$ is a global minimizer of the given problem. Then, for all $\boldsymbol{x} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^{(k)}$, we have $\boldsymbol{c}^T\boldsymbol{x} \geq \boldsymbol{c}^T\boldsymbol{x}^{(k)}$. Rewriting, we obtain $\boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \geq 0$. Recall that

$$\begin{aligned} \boldsymbol{\Pi}[\boldsymbol{x}^{(k)} - \nabla f(\boldsymbol{x}^{(k)})] &= \arg\min_{\boldsymbol{x}\in\Omega} \|\boldsymbol{x} - (\boldsymbol{x}^{(k)} - \nabla f(\boldsymbol{x}^{(k)}))\|^2 \\ &= \arg\min_{\boldsymbol{x}\in\Omega} \|\boldsymbol{x} - \boldsymbol{x}^{(k)} + \boldsymbol{c}\|^2. \end{aligned}$$

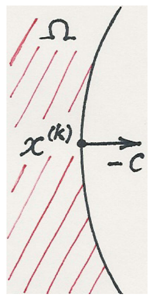
But, for any $\boldsymbol{x} \in \Omega$, $\boldsymbol{x} \neq \boldsymbol{x}^{(k)}$,

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{x}^{(k)} + \boldsymbol{c}\|^2 &= \|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|^2 + \|\boldsymbol{c}\|^2 + 2\boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \\ &> \|\boldsymbol{c}\|^2, \end{aligned}$$

where we used the facts that $\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|^2 > 0$ and $\boldsymbol{c}^T(\boldsymbol{x} - \boldsymbol{x}^{(k)}) \geq 0$. On the other hand, $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k)} + \boldsymbol{c}\|^2 = \|\boldsymbol{c}\|^2$. Hence,

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{\Pi}[\boldsymbol{x}^{(k)} - \nabla f(\boldsymbol{x}^{(k)})] = \boldsymbol{x}^{(k)}.$$

b. No. Counterexample:



4. (13 pts.) Given vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p \in \mathbb{R}^n$ and scalars $u_1, \ldots, u_p$, consider the problem

$$\begin{array}{ll} \text{minimize} & \max\{\boldsymbol{v}_1^T\boldsymbol{x} + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x} + u_p\} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \end{array}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Call this problem P1.

a. Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & y \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & y \geq \boldsymbol{v}_i^T\boldsymbol{x} + u_i, \quad i = 1, \ldots, p, \end{array}$$

where the decision variable is the vector $[\boldsymbol{x}^T, y]^T$. Call this problem P2. Show that $\boldsymbol{x}^*$ solves P1 if and only if $[\boldsymbol{x}^{*T}, y^*]^T$ with $y^* = \max\{\boldsymbol{v}_1^T\boldsymbol{x}^* + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x}^* + u_p\}$ solves P2.
Hint: $y \geq \max\{a, b, c\}$ if and only if $y \geq a$, $y \geq b$, and $y \geq c$.

b. Use part a to derive a linear programming problem

$$\begin{array}{ll} \text{minimize} & \hat{\boldsymbol{c}}^T\boldsymbol{z} \\ \text{subject to} & \hat{\boldsymbol{A}}\boldsymbol{z} \leq \hat{\boldsymbol{b}} \end{array}$$

that is equivalent to P1 (by "equivalent" we mean that the solution to one gives us the solution to the other). Explain how a solution to the linear programming problem above gives a solution to P1.

Ans.: a. First suppose that $\boldsymbol{x}^*$ is optimal in P1. Let $y^* = \max\{\boldsymbol{v}_1^T\boldsymbol{x}^* + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x}^* + u_p\}$. Then, $[\boldsymbol{x}^{*T}, y^*]^T$ is feasible in P2. Let $[\boldsymbol{x}^T, y]^T$ be any feasible point in P2. Then (by the hint),

$$y \geq \max\{\boldsymbol{v}_1^T\boldsymbol{x} + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x} + u_p\}.$$

Moreover, $\boldsymbol{x}$ is feasible in P1, and hence

$$y \geq \max\{\boldsymbol{v}_1^T\boldsymbol{x} + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x} + u_p\} \geq \max\{\boldsymbol{v}_1^T\boldsymbol{x}^* + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x}^* + u_p\} = y^*.$$

Hence, $[\boldsymbol{x}^{*T}, y^*]^T$ is optimal in the LP.

To prove the converse, suppose that $\boldsymbol{x}^*$ is not optimal in P1. Then, there is some $\boldsymbol{x}'$ that is feasible in P1 such that

$$y' = \max\{\boldsymbol{v}_1^T\boldsymbol{x}' + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x}' + u_p\} < \max\{\boldsymbol{v}_1^T\boldsymbol{x}^* + u_1, \ldots, \boldsymbol{v}_p^T\boldsymbol{x}^* + u_p\} = y^*.$$

But $[\boldsymbol{x}'^T, y']^T$ is evidently feasible in P2, and has objective function value (y') that is lower than that of $[\boldsymbol{x}^{*T}, y^*]^T$. Hence, $[\boldsymbol{x}^{*T}, y^*]^T$ is not optimal in the P2.

b. Define

$$\boldsymbol{z} = \begin{bmatrix} \boldsymbol{x} \\ y \end{bmatrix}, \qquad \hat{\boldsymbol{c}} = \begin{bmatrix} \boldsymbol{0} \\ 1 \end{bmatrix}, \qquad \hat{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{v}_1^T & -1 \\ \vdots & \vdots \\ \boldsymbol{v}_p^T & -1 \end{bmatrix}, \qquad \hat{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{b} \\ -u_1 \\ \vdots \\ -u_p \end{bmatrix}.$$

Then the equivalent problem can be written as

$$\begin{array}{rl} \text{minimize} & \hat{\boldsymbol{c}}^T\boldsymbol{z} \\ \text{subject to} & \hat{\boldsymbol{A}}\boldsymbol{z} \leq \hat{\boldsymbol{b}}. \end{array}$$

By part a, if we obtain a solution to this LP problem, then the first n components forms a solution to the original problem.