UC Berkeley
Department of Electrical Engineering and Computer Science

EECS 227A
NONLINEAR AND CONVEX OPTIMIZATION

Solutions 5
Fall 2009

Reading: Boyd and Vandenberghe, Chapter 5

Solution 5.1
Note that we have $\nabla f(x) = \mathbf{1}$ for any $x \in \mathbb{R}^n$. Since for any feasible point $\nabla h(x) = 2x \neq 0$, any local minimum will be regular. From the Lagrange multiplier theorem, there exists λ^* such that

$$\nabla f(x^*) + \lambda^* \nabla h(x^*) = 0,$$

or equivalently $\mathbf{1} + 2\lambda^* x^* = 0$. Using the requirement that $h(x^*) = \|x^*\|^2 - 1 = 0$, this yields the two possibilities

$$(x_i^*, \lambda_i^*) = \pm\Big(\frac{1}{\sqrt{n}}, -\frac{\sqrt{n}}{2}\Big).$$

By using the second order optimality conditions, we have

$$y'\big(\nabla^2 f(x^*) + \lambda^* \nabla^2 h(x^*)\big)y \geq 0$$

for all $y \in V(x^*) := \{y \,|\, 2(x^*)^T y = 0\}$. Since $\nabla^2 f(x^*) = 0$ and $\nabla^2 h(x^*) = 2I$, the above condition is equivalent to $2\lambda^* y^T I y \geq 0$ for all $y \in V(x^*)$, which can hold only for non-negative λ^*. Hence the only point satisfying the first and second order conditions for a local minimum is $x^* = -\frac{1}{\sqrt{n}}\mathbf{1}$. Since the constraint set is compact and f is continuous, an optimal point x^* exists, and the above point must be the optimum.

One geometric interpretation of the Lagrange multiplier conditions is that the gradient direction $\nabla f(x^*) = \mathbf{1}$ is parallel with the gradient $\nabla h(x^*) = 2x^*$ associated with the constraint $h(x) = \|x\|^2 - 1 = 0$, which defines the boundary of a circle.

Solution 5.2

(a) Note that the constraint set for k^{th} problem is

$$C_k = \{x \in \mathbb{R}^n : \|x\|_2 = 1, e_i^T x = 0, i = 1, \ldots, k\}$$

satisfies $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_n$. Since all these problems have the same objective, we must have the optimal value satisfies $\lambda_1 \leq \lambda_2 \ldots \leq \lambda_n$

(b) Suppose $\alpha_1, \ldots, \alpha_n$ satisfies $\sum_{i=1}^n \alpha_i e_i = 0$. By multiplying both sides with e_k^T, note that $e_k^T e_i = 0, \forall i \neq k$ and $e_k^T e_k = 1$, we have $\alpha_k = 0, \forall k = 1, \ldots, n$. Therefore, $e_1, \ldots, e_n$ are linearly independent.

(c) First consider problem P_1. Since $e_1 \neq 0$, the optimum is regular, so that the Lagrange multiplier theorem guarantees the existence of some $\mu_{1,1}^* \in \mathbb{R}$ such that

$$2Qe_1 + 2\mu_{1,1}^* e_1 = 0.$$

Taking inner products of both sides of this equation with e_1 and using $\|e_1\|^2 = 1$ yields that $\lambda_1 = -\mu^*_{1,1}$ is a Lagrange multiplier, and (λ_1, e_1) are an eigenvalue/vector pair.

Now from the linear independence established in (b), the points $e_2, \dots e_n$ are regular optima for problems $P_2, P_3 \dots P_n$ respectively. So by our Lagrange multiplier condition, for $i = 2, \dots, n$, there exists a unique $\mu^*_i = \{\mu^*_{i,1}, \mu^*_{i,2}, \dots, \mu^*_{i,i}\} \in \mathbb{R}^i$ for problem P_i such that

$$2Qe_i + 2\mu^*_{i,i}e_i + \big[\sum_{j=1}^{i-1} \mu^*_{i,j}e_j\big] = 0. \tag{1}$$

Take inner products of both sides with e_i and use the orthogonality condition to obtain

$$2e_i^T Qe_i + 2\mu^*_{i,i}\|e_i\|^2 = 2\lambda_i + 2\mu^*_{i,i} = 0.$$

Thus, we can identify $\lambda_i = -\mu^*_{i,i}$ as the Lagrange multiplier for P_i associated with the constraint $\|x\|^2 - 1 = 0$.

We have already shown that (λ_1, e_1) are an eigenvalue/vector pair for Q. We now proceed by induction on i: suppose that for $j = 1, \dots, i-1$, the corresponding (λ_j, e_j) are eigenvalue/vector pairs for Q. Consider equation (1) for P_i. Taking inner products with e_j for $j = 1, \dots, i-1$, using the orthogonality relations and the fact that (by the induction hypothesis) e_j is an eigenvector with eigenvalue λ_j yields

$$2e_j^T Qe_i + \mu^*_{i,j}\|e_j\|^2 = 2e_j^T \lambda_j e_i + \mu^*_{i,j} = \mu^*_{i,j} = 0.$$

Thus, substituting $\mu^*_{i,i} = -\lambda_i$ and $\mu^*_{i,j} = 0$ for all $j = 1, \dots, i-1$ into equation (1) yields that $2Qe_i = 2\lambda e_i$, so that (λ_i, e_i) are another eigenvalue/vector pair.

Solution 5.3

(a) Translating the condition for regularity to linear constraints, the row vectors $a_1^T, \dots, a_m^T$ forming A be linearly independent. If $\text{rank}(A) < m$, this is not posssible.

(b) If A is rank-deficient (say $\text{rank}(A) = k < m$), then we can form a reduced matrix of the form

$$B = \begin{bmatrix} a_{i_1}^T \\ \dots \\ a_{i_k}^T \end{bmatrix}$$

where $\{i_1, \dots, i_k\}$ are a subset of k indices from $\{1, \dots, m\}$, such that B has rank k, and imposes the same set of constraints as A. Thus, the original problem is equivalent to $\min f(x)$ subject to $Bx = 0$. Any optimum for the reduced problem will be regular by construction, so that the Lagrange multiplier theorem yields that there exist scalars $\lambda^*_{i_j} \in \mathbb{R}$ for $j = 1, \dots, k$ such that

$$\nabla f(x^*) + \sum_{j=1}^{k} \lambda^*_{i_j} a_{i_j} = 0.$$

We can now add Lagrange multipliers $\lambda_\ell^* = 0$ for any index ℓ not included in $\{i_1, \ldots, i_k\}$, which leads to the Lagrange multiplier condition for the original problem.

(c) We lose uniqueness of the Lagrange multiplier, because there is freedom in which subset of k linearly independent rows that we choose to form the reduced problem.

Solution 5.4

The given problem is equivalent to minimizing $-y^T x$ subject to $g(x) = x^T Q x - 1 \leq 0$. The constraint set is closed and bounded, and the cost function is continuous, so that a global minimum exists.

If $y = 0$, then $f(x) = 0$ for all x in the constraint set, and the result follows. Otherwise, for $y \neq 0$, we have

$$\nabla f(x) = -y, \qquad \nabla g(x) = 2Qx.$$

Since $Q \succ 0$, any feasible point is regular. Assume that x^* is a local minimum. By the KKT conditions, there exists a $\mu^* \geq 0$ such that

$$-y + \mu^* 2Qx^* = 0, \qquad \mu^* = 0 \text{ if } (x^*)^T Q x^* < 1.$$

Since $y \neq 0$, we must have $\mu^* > 0$ and thus $(x^*)^T Q x^* = 1$, and $y = 2\mu^* Q x^*$. Multiplying both sides of this latter equation by $(x^*)^T$ yields

$$(x^*)^T y = 2\mu^* (x^*)^T Q x^* \;=\; 2\mu^*.$$

But we also have

$$x^* = \frac{Q^{-1} y}{2\mu^*} = \frac{Q^{-1} y}{y^T x^*}$$

Take inner products of both sides with y^T to obtain

$$y^T x^* = \frac{y^T Q^{-1} y}{y^T x^*}$$

which is equivalent to $y^T x^* = \pm\sqrt{y^T Q^{-1} y}$. Therefore, the optimal minimum is $-\sqrt{y^T Q^{-1} y}$, and so the optimal value of the original maximization problem is $\sqrt{y^T Q^{-1} y}$.

Now for any $x \neq 0$, let $\bar{x} = x / \sqrt{x^T Q x}$. We have $\bar{x}^T Q \bar{x} = 1$ so that $\bar{x}$ is feasible for the maximization problem and

$$y^T \bar{x} = \frac{y^T x}{\sqrt{x^T Q x}} \leq \sqrt{y^T Q^{-1} y}$$

as required.

Solution 5.5

(a) Simple manipulation yeilds

$$\|Ax_{cheb} - b\|_\infty \geq \frac{1}{\sqrt{m}} \|Ax_{cheb} - b\|_2 \geq \frac{1}{\sqrt{m}} \|Ax_{ls} - b\|_2 \geq \frac{1}{\sqrt{m}} \|Ax_{ls} - b\|_\infty$$

(b) From the expression $x_{ls} = (A^TA)^{-1}A^Tb$ we note that

$$A^T r_{ls} = A^T(b - A(A^TA)^{-1}b) = A^Tb - A^Tb = 0$$

therefore $A^T\hat{\nu} = 0$ and $A^T\tilde{\nu} = 0$. Obviously we also have $\|\hat{\nu}\|_1 = \|\tilde{\nu}\|_1 = 1$ so $\hat{\nu}$ and $\tilde{\nu}$ are dual feasible. We can write the dual objective value at $\hat{\nu}$ as

$$b^T\hat{\nu} = \frac{-b^T r_{ls}}{\|r_{ls}\|_1} = \frac{(Ax_{ls} - b)^T r_{ls}}{\|r_{ls}\|_1} = -\frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1}$$

and similarly,

$$b^T = \frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1}$$

Therefore, $\tilde{\nu}$ gives a better bound than $\hat{\nu}$ Finally to show that the resulting lower bound is better than the bound in part (a), we have to verify that

$$\frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1} \geq \frac{1}{\sqrt{m}}\|r_{ls}\|_\infty$$

This follows from the inequalities

$$\|x\|_1 \leq \sqrt{m}\|x\|_2, \|x\|_\infty \geq \|x\|_2$$

which hold for general $x \in \mathbb{R}^m$.

Solution 5.6

The Lagrangian is

$$L(x, z_1, \ldots, z_N) = \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2 - \sum_{i=1}^{N} z_i^T(y_i - A_ix - b_i)$$

We first minimize over y_i. We have

$$\inf_{y_i} \|y_i\|_2 + z_i^T y_i = \begin{cases} 0 & \text{if } \|z_i\|_2 \leq 1 \\ -\infty & \text{otherwise} \end{cases}$$

(If $\|z_i\|_2 > 1$, choose $y_i = -tz_i$ and let $t \to \infty$, to show that the function is unbounded below. If $\|z_i\|_2 \leq 1$, it follows from the Cauchy-Schwarz inequality that $\|y_i\|_2 + z_i^T y_i \geq 0$, so the minimum is reached when $y_i = 0$).

We can minimize over x by setting the gradient with respect to x equal to zero. This yields

$$x = x_0 + \sum_{i=1}^{N} A_i^T z$$

Substituting in the Lagrangian gives the dual function

$$g(z_1, \ldots, z_N) = \begin{cases} \sum_{i=1}^{N}(A_ix_0 + b_i)^T z_i - \frac{1}{2}\|\sum_{i=1}^{N} A_i^T z_i\|_2^2 & \text{if } \|z_i\|_2 \leq 1, i = 1, \ldots, N \\ -\infty & \text{otherwise} \end{cases}$$

Therefore the dual problem is

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^N (A_i x_0 + b_i)^T z_i - \frac{1}{2} \| \sum_{i=1}^N A_i^T z_i \|_2^2 \\ \text{subject to} & \|z_i\|_2 \leq 1, i = 1, \ldots, N \end{array}$$

Solution 5.7

If $\tilde{x}$ minimizes ϕ, then

$$\nabla f_0(\tilde{x}) + 2\alpha A^T (A\tilde{x} - b) = 0$$

Therefore $\tilde{x}$ is also a minimizer of

$$f_0(x) + \nu^T (Ax - b)$$

where $\nu = 2\alpha(A\tilde{x} - b)$. Therefore ν is dual feasible with

$$\begin{array}{rl} g(\nu) & = \inf_x f_0(x) + \nu^T (Ax - b) \\ & = f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2 \end{array}$$

Therefore,

$$f_0(x) \geq f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2$$

for all x such that $Ax = b$.