UC Berkeley
Department of Electrical Engineering and Computer Science

EECS 227A
NONLINEAR AND CONVEX OPTIMIZATION

Solutions 1
Fall 2009

Review 1.1

For future reference, the function f has the following first and second derivatives:

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y) &= 4x^3 - 4yx - 20x \\
\frac{\partial f}{\partial y}(x,y) &= -2x^2 + 2y \\
\frac{\partial^2 f}{\partial y^2}(x,y) &= -4x \\
\frac{\partial^2 f}{\partial x \partial y}(x,y) &= 2
\end{aligned}$$

(a) Note that $f(1,0) = f(-1,0) = -9$, so that $\frac{1}{2}f(1,0) + \frac{1}{2}f(-1,0) = -9$. On the other hand, we have $f(0,0) = 0 > -9$, showing that f is not convex.

(b) We need to solve $\nabla f(x,y) = 0$ for $(x,y) \in \mathbb{R}^2$. Taking derivatives yields

$$\nabla f(x,y) = \begin{bmatrix} 4x\,(x^2 - y) - 20x & -2(x^2 - y) \end{bmatrix}^T$$

By inspection, $(x,y) = (0,0)$ is the only stationary point. It is neither a local maximum nor a local minimum, since $f(\delta,0) = \delta^4 - 10\delta^2 < 0$ for δ small enough, and $f(0,\epsilon) = \epsilon^2 > 0$ for $\epsilon \neq 0$. Alternatively, we may compute the Hessian matrix $\nabla^2 f(0,0) = \begin{pmatrix} -20 & 0 \\ 0 & 2 \end{pmatrix}$ which is indefinite, showing $(0,0)$ is neither a local minimum nor a local maximum.

(c) The function f is convex. The set C is bounded in the y-direction, and the function value tends to infinity as $|x|$ increases, so that a global minimum exists and is attained. By necessary optimality condition for constrained optimization to solve this problem, if (x^*, y^*) is global minimum it needs to satisfy

$$\nabla f(x^*, y^*)^T \left[\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x^* \\ y^* \end{pmatrix} \right] \geq 0, \forall (x,y) \in X$$

In part (b), we showed that f has only one stationary point in the interior of the given set, and it is not a minimum. Hence, the minimum must occur on the boundary of the set—i.e., either at some point $(x^*, 1)$ or some point $(x^*, 0)$, and we must have $\frac{\partial f}{\partial x}(x^*, 1) = 0$ or $\frac{\partial f}{\partial x}(x^*, 0) = 0$. For $y = 0$, we have $f(x,0) = x^4 - 10x^2$. Since $\frac{\partial f}{\partial x}(x,0) = 4x^3 - 20x^2$, the relevant stationary points are at $x = \pm\sqrt{5}$, with $f(\pm\sqrt{5}, 0) = 25 - 50 = -25$. Otherwise, for $y = 1$, we have $f(x,1) = x^4 - 12x^2 + 1$. We have $\frac{\partial f}{\partial x}(x,1) = 4x^3 - 24x$, so stationary points are found at $x = \pm\sqrt{6}$ with $f(\pm\sqrt{6}, 0) = -35$. Checking second derivatives shows that this is the global optimum.

Review 1.2

Using KKT conditions, we have

$$\begin{aligned} &g_j(x^*) \le 0, j = 1, \dots, m \\ &\lambda_j^* \ge 0, j = 1, \dots, m \\ &\lambda_j^* g_j(x^*) = 0, j = 1, \dots, m \\ &\nabla f(x^*) + \sum_{j=1}^m \lambda_j^* \nabla g_j(x^*) = 0 \end{aligned}$$

So

$$\nabla f(x^*)^T (x - x^*) = -\sum_{j=1}^m \lambda_j^* \nabla g_j(x^*)^T (x - x^*)$$

Since $g_j, j = 1, \dots, m$ are convex functions: $\nabla g_j(x^*)^T (x - x^*) \le g_j(x) - g_j(x^*)$, therefore

$$\begin{aligned} \nabla f(x^*)^T (x - x^*) &\ge -\sum_{j=1}^m \lambda_j^* (g_j(x) - g_j(x^*)) \\ &= -\sum_{j=1}^m \lambda_j^* g_j(x) + \sum_{j=1}^m \lambda_j^* g_j(x^*) \\ &= -\sum_{j=1}^m \lambda_j^* g_j(x) \end{aligned}$$

For any $x \in C$ we have $g_j(x) \le 0, j = 1, \dots, m$. Since $\lambda_j^* \ge 0$, we conclude that $\nabla f(x^*)^T (x - x^*) \ge 0$, as claimed.

Review 1.3

(a) Using projection theorem, we have

$$(y - \Pi_K(y))^T (x - \Pi_K(y)) \le 0, \forall x \in K$$

Since $\Pi_K(y) \in K$, $x = t\Pi_K(y) \in K$ for any $t \ge 0$, so

$$(t - 1)(y - \Pi_K(y))^T \Pi_K(y) \le 0, \forall t \ge 0$$

By replacing $t = 0, t = 2$ we must have $(y - \Pi_K(y))^T \Pi_K(y) = 0$

(b) From (a) we have $(\Pi_K(y) - y)^T x \ge 0, \forall x \in K$ so $\Pi_K(y) - y \in K^*$

(c) Using result from part (a) and (b), we have

$$\Pi_K(y) - y \in K^* \text{ and } \langle \Pi_K(y) - y, \Pi_K(y) \rangle = 0$$

We also note that there exists unique $z \in K$ such that

$$z - y \in K^* \text{ and } \langle z - y, z \rangle = 0$$

(i) The orthant cone $K = \{x \in \mathbb{R}^n \mid x_i \geq 0\}$ has the dual cone $K^* = K$.
From the above conditions, we must have $\Pi_K(y)_i - y_i \geq 0$, $(\Pi_K(y)_i - y_i)\Pi_K(y)_i = 0$, for any $i = 1, \ldots, n$. So $\Pi_K(y)_i = \max(0, y_i)$. Therefore, $\Pi_K(y) = y_+$

(ii) The semidefinite cone $K = \{X \in S^n \mid X \succeq 0\}$ has the dual cone $K^* = K$.
Suppose $X = \sum_{i=1}^n \lambda_i u_i u_i^T$ be the SVD of X, then the matrix $\sum_{i=1}^n \max(0, \lambda_i) u_i u_i^T$ satisfies the above conditions so $\Pi_K(y) = \sum_{i=1}^n \max(0, \lambda_i) u_i u_i^T$

Review 1.4

(a) It is not, since the constraint $\|x\|_2 = 1$ is non-convex. Conic programs are always convex.

(b) Consider $X = xx^T$ such that x is the optimal solution of (1), we then have

$$X \succeq 0 \text{ and } \operatorname{trace}(X) = \sum_{i=1}^n x_i^2 = 1 \text{ and } \sum_{i,j} |X_{ij}| = \sum_{i,j} |x_i x_j| = (\sum_i |x_i|)^2 \leq C^2$$

so X is feasible for (2), therefore

$$b^* \geq \operatorname{trace}(\Gamma X) = \operatorname{trace}(\Gamma x x^T) = x \Gamma x^T = a^*$$

(c) Consider the Lagrangian

$$\begin{aligned} L(X, Y, \nu, \mu) &= \operatorname{trace}(\Gamma X) + \operatorname{trace}(YX) + \nu(\operatorname{trace}(X) - 1) - \mu(\textstyle\sum_{i,j} |X_{ij}| - C^2) \\ &= \operatorname{trace}((\Gamma + Y + \nu I)X) - \mu \textstyle\sum_{i,j} |X_{ij}|) + C^2\mu - \nu \end{aligned}$$

where $Y \succeq 0, \nu \geq 0, \mu \in \mathbb{R}$ are dual variables. So the dual function is

$$\begin{aligned} q(Y, \nu, \mu) &= C^2\mu - \nu + \sup_X \left(\operatorname{trace}((\Gamma + Y + \nu I)X) - \mu \sum_{i,j} |X_{ij}| \right) \\ &= C^2\mu - \nu + \sum_{ij} \sup_{X_{ij}} ((\Gamma + Y + \nu I)_{ij} X_{ij} - \mu |X_{ij}|) \\ &= \begin{cases} C^2\mu - \nu & \text{if } |(\Gamma + Y + \nu I)_{ij}| \leq \mu \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

So the dual problem is

$$\begin{array}{ll} \min & C^2\mu - \nu \\ \text{subject to} & Y \succeq 0, \\ & -\mu \leq (\Gamma + Y + \nu I)_{ij} \leq \mu \end{array}$$

Letting $x = \begin{pmatrix} \mu \\ \nu \\ Y_{ij} \end{pmatrix}$, the above problem is of the form

$$\begin{array}{ll} \min & c^T x \\ \text{subject to} & \sum_{i=1}^m x_i F_i \preceq 0 \\ & Ax \leq b \end{array}$$

for some matrices A, F_i and vector b.

(d) Here we describe a barrier method for solving the dual problem of the SDP. The barrier function corresponding to the above problem is

$$\phi_t(x) = tc^T x - \log\det(-\sum_{i=1}^{m} x_i F_i) - \sum_{j=1}^{k} \log(b_j - a_j^T x)$$

The gradient and Hessian of $\phi_t(x)$ are given by

$$\frac{\partial \phi_t}{\partial x_i} = tc_i + \text{trace}(F(x)^{-1} F_i) + \sum_{j=1}^{k} \frac{a_{ji}}{b_j - a_j^T x}$$

$$\frac{\partial^2 \phi_t}{\partial x_i \partial x_l} = \text{trace}(F(x)^{-1} F_i F(x)^{-1} F_l) + \sum_{j=1}^{k} \frac{a_{ji} a_{jl}}{(b_j - a_j^T x)^2}$$

An interior point method would follow the central path by minimizing the function $\phi_t(x)$ for a sequence of steps $t^0, t^1, t^2, \ldots$, where $t^{k+1} = \mu t^k$ for some parameter $\mu > 1$. Each of the minimizations could be solved by Newton's method, using the previous optimal solution as the initial starting point.

Review 1.5

(a) $f(x) = \exp(x), x \in \mathbb{R}$

(i) We calculate $f''(x) = \exp(x) > 0, \forall x$ so f is convex. By definition, ew have $\text{dom}(f) = \mathbb{R}$.
Also, f is differentiable everywhere so $\partial f(x) = \{f'(x)\} = \{\exp(x)\}$

(ii) $f^*(x) = \sup_{y \in \mathbb{R}}(yx - \exp(y))$.
If $x \leq 0$ then $f^*(x) = \infty$. Otherwise, taking first derivative, we have $y^* = \log(x)$, so $f^*(x) = x\log(x) - x$ if $x > 0$, ∞ otherwise. $\text{dom}(f^*) = \{x \in \mathbb{R} | x > 0\}$

(iii) $f^{**}(x) = \sup_{y \in \text{dom}(f^*)}(yx - y\log(y) + y)$, taking first derivative we have $y^* = \exp(x)$ so $f^{**}(x) = \exp(x) = f(x)$.
Since f is convex and closed, we do expect $f^{**} = f$

(b) $f(x) = x\log(x) - x$ for $x > 0$, ∞ otherwise

(i) Using result from (a), we have that $f(x)$ is the conjugate of $\exp(x)$ so it is convex. $\text{dom}(f) = \{x \in \mathbb{R} | x > 0\}$. f is differentiable so $\partial f(x) = f'(x) = \log(x)$

(ii) Also from (a), $f^*(x) = \exp(x)$, $\text{dom}(f^*) = \mathbb{R}$

(iii) Also from (a), $f^{**}(x) = f(x)$.

(c) $f(x) = \frac{1}{2} x^T \Gamma x$ where $\Gamma \succeq 0$

(i) Since f is twice continously differentiable and $\nabla^2 f(x) = \Gamma \succeq 0$, f is convex. $\text{dom}(f) = \mathbb{R}^n$ and $\partial f(x) = \nabla f(x) = \Gamma x$

(ii) $f^*(y) = \sup_{x \in \mathbb{R}^n}(y^T x - \frac{1}{2} x^T \Gamma x)$

- If $y \in range(\Gamma)$, by taking first derivative we have $y = \Gamma x^*$, so $x^* = \Gamma^+ y + z$ where Γ^+ is the pseudo inverse of Γ and $z \in null(\Gamma)$, thus $f^*(y) = \frac{1}{2} y^T \Gamma^+ y$
- If $y \notin range(\Gamma)$, then there exists $v \in null(\Gamma)$ such that $y^T v \neq 0$, by choosing $x = tv$, we have $y^T x - \frac{1}{2} x^T \Gamma x = t y^T v \to \infty$ as $t \to \infty$.

Therefore, $f^*(y) = \frac{1}{2} y^T \Gamma^+ y$ if $y \in range(\Gamma), \infty$ otherwise. $\text{dom}(f^*) = range(\Gamma)$

(iii) $f^{**}(x) = \sup_{y \in range(\Gamma)} (x^T y - \frac{1}{2} y^T \Gamma^+ y) = \sup_{z \in \mathbb{R}^n} (x^T \Gamma z - \frac{1}{2} z^T \Gamma \Gamma^+ \Gamma z)$ (let $y = \Gamma z$). Since $\Gamma \Gamma^+ \Gamma = \Gamma$, we have

$$f^{**}(x) = \sup_{z \in \mathbb{R}^n} (x^T \Gamma z - \frac{1}{2} z^T \Gamma z)$$

Taking the first derivative, $\Gamma x = \Gamma z^*$, so $z^* = x + w$ for some $w \in null(\Gamma)$, thus

$$f^{**}(x) = \frac{1}{2} x^T \Gamma x = f(x)$$

(d) $f(x) = \sup_{y \in C} y^T x$ where C is a non-empty, closed and convex set. We first observe that $f(x) = \mathbb{I}^*_C(x)$

(i) f is the pointwise supremum over a set of linear functions so it is convex. WLOG, assume x is unit vector, $f(x) < \infty$ if and only if there exists $y_0 \in C$: $y^T x \leq y_0^T x, \forall y \in C$ so C is contained in the halfspace $\{y | y^T x \leq y_0^T x\}$ which means that x has to be a normal vector of some supporting hyperplane of C. So $\text{dom}(f) = \{tx | x \text{ is a normal vector of some supporting hyperplane of } C, t \geq 0\}$. Using the property of sub-gradients of pointwise supremum function we have

$$\begin{aligned} \partial f(x) &= cl(conv(y \in C | y^T x = f(x))) \\ &= cl(conv(y \in C | x \text{ defines a supporting hyperplane of } C \text{ at y})) \end{aligned}$$

where $cl(A)$ is the closure of A, $conv(A)$ is the convex hull of A. Since C is closed and convex, this set can be simplified as

$$\partial f(x) = \{y \in C | x \text{ defines a supporting hyperplane of } C \text{ at y}\}$$

(ii) Define the indicator function $g(z) = \mathbb{I}_C(z)$. Note that

$$g^*(x) \;=\; \sup_z \{x^T z - g(z)\} \;=\; \sup_{z \in C} x^T z = f(x).$$

Therefore, we have $g^{**} = f^*$. But g is a closed and convex function, so that $g^{**} = g$, and hence $f^* = \mathbb{I}_C$. The domain of f^* is C.

(iii) From part (ii), the dual function of f^*—that is, the function f^{**}—is given by f. This is to be expected since f is a closed and convex function.

Review 1.6

(a) True if $n = 1$. False if $n \geq 2$, since both of the vectors $(1, 1, 1, \ldots, 1)$ and $(-1, -1, 1, \ldots, 1)$ belong to C but the center of these points $(0, 0, 1, \ldots, 1)$ is not in C. The statement would be true for arbitrary $n \geq 1$ if the additional constraints $x_i > 0$ were added, and credit would be given if this extra constraint was given, and the validity was justified.

(b) From results in class, the statement is true if the domain of f is all of $\mathbb{R}^n$. (Without this condition, it could be false. Consider functions of the form $f(x) = c^T x + \mathbb{I}_{g(x)\leq 0}(x)$, where $\mathbb{I}$ is the indicator function for the set $\{g(x) \leq 0\}$. This effectively introduces a non-linear constraint into our problem, and we know that some form of constraint qualification (e.g., Slater) is needed in general to ensure strong duality in such a setting.)

(c) True as long as $\nabla h(x^*) \neq 0$, but false in general. We proved in class that Lagrange multipliers exist assuming that the set of constraint gradients are not linearly independent. Here there is only one set of constraint gradients, so the result would be true if $\nabla h(x^*) \neq 0$. To see what can go wrong otherwise, consider minimizing $f(x, y) = x$ subject to $h(x, y) = x^2 - y^3 = 0$. Here the global optimum is $(x^*, y^*) = (0, 0)$ and $\nabla h(0, 0) = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$. However, we have $\nabla f(0, 0) = \begin{bmatrix} 1 & 0 \end{bmatrix}^T$, so that the given relation cannot hold. Credit would be given for both answers, assuming that the appropriate justification was given.

(d) For the univariate version of Newton's method (finding a zero of a function $g : \mathbb{R} \to \mathbb{R}$), one correct statement that if x^* is a stationary point (i.e., $g(x^*) = 0$), and g is continuously differentiable in a ball around x^* with $g'(x^*) \neq 0$, then there is a neighborhood of x^* such that if Newton's method is initialized inside it, then it will converge to x^*. Credit would be awarded for true with this justification, or for answering false, and pointing out that the condition $g'(x^*) \neq 0$ is needed.