QUESTION BOOKLET
EECS 227A Fall 2009
Midterm
Tuesday, Ocotober 20, 11:10-12:30pm

DO NOT OPEN THIS QUESTION BOOKLET UNTIL YOU ARE TOLD TO DO SO

- You have 80 minutes to complete the midterm.
- The midterm consists of three problems, provided in the question booklet (THIS BOOKLET), that are in no particular order of difficulty.
- Write your solution to each problem in the space provided in the solution booklet (THE OTHER BOOKLET). Try to be neat! If we can't read it, we can't grade it.
- You may give an answer in the form of an arithmetic expression (sums, products, ratios, factorials) that could be evaluated using a calculator. Expressions like $\binom{8}{3}$ or $\sum_{k=0}^{5}(1/2)^k$ are also fine.
- A correct answer does not guarantee full credit and a wrong answer does not guarantee loss of credit. You should concisely explain your reasoning and show all relevant work. The grade on each problem is based on our judgment of your understanding as reflected by what you have written.
- This is a closed-book exam except for one $8.5'' \times 11''$ page of notes.

Problem 1:
For *all parts of this problem*, let f be a convex and twice continuously differentiable function. Consider the steepest descent method

$$x^{k+1} = x^k - \alpha^k \nabla f(x^k).$$

(a) Suppose that $\nabla^2 f(x) \succ 0$ for all $x \in \mathbb{R}^n$. How many local minima does f have? How many global minima does f have?

(b) State a convergence result for the steepest descent method. Be sure to specify explicitly how your step size is chosen.

(c) Consider the condition

$$\exists\ \gamma > 0 \text{ such that } \quad \frac{\|\nabla f(x) - \nabla f(y)\|_2^2}{\gamma} \leq (\nabla f(x) - \nabla f(y))^T (x-y) \quad \text{for all } x, y \in \mathbb{R}^n. \quad (1)$$

Given a quadratic function $f(x) = \frac{1}{2} x^T Q x - b^T x$ where Q is symmetric, state the weakest sufficient conditions on (Q, b) under which condition (1) guaranteed to hold.

(d) Assume that there exists some $x^* \in \mathbb{R}^n$ such that $\nabla f(x^*) = 0$. Assuming that condition (1) holds, prove that $x^k \to x^*$ as long as there is some $\epsilon \in (0, 2)$ such that the step sizes α^k satisfy

$$\epsilon \leq \alpha^k \leq \frac{2-\epsilon}{\gamma} \quad \text{for all } k = 1, 2, \ldots.$$

Solution:

(a) Under the given condition, any local minimum must be a global minimum. It is possible that f does not have a global minimum (e.g., $f(x) = \exp(-x)$ for $x \in \mathbb{R}$.) If it has a global minimum, then it must be unique since the assumption implies that f is strictly convex on $\mathbb{R}^n$.

(b) Suppose that f achieves its global minimum at x^*; from part (a), x^* is the only possible stationary point. Therefore, if we implement steepest descent with α^k chosen by the Armijo rule (as stated in class or in the Bertsekas book), then the sequence $\{x^k\}$ will converge to the global minimum x^*.

(c) The weakest sufficient conditions are that $0 \preceq Q \preceq \gamma I$. No conditions on b are needed. We have $\nabla f(x) = Qx - b$ for all $x \in \mathbb{R}^n$, and therefore we want to show that there exists $\gamma > 0$ such that

$$\frac{\|Q(x-y)\|_2^2}{\gamma} \leq (x-y)^T Q (x-y)$$

for all $x, y \in \mathbb{R}^n$. This will hold if and only if $\gamma Q - Q^2 \succeq 0$. Let $z \in \mathbb{R}^n$ be a unit norm eigenvector of Q associated with a non-zero eigenvalue λ. We then need

$$\gamma z^T Q z - z^T Q^2 z = \gamma\lambda - \lambda^2 \ = \ \lambda\,(\gamma - \lambda) \geq 0.$$

This inequality holds if and only if $\lambda \in [0, \gamma]$ for all eigenvalues λ, as claimed.

(d) We claim that condition (1) implies that the gradient is Lipschitz with parameter $\gamma > 0$. Indeed, we have

$$\begin{aligned}\frac{\|\nabla f(x) - \nabla f(y)\|_2^2}{\gamma} &\leq (\nabla f(x) - \nabla f(y))^T (x - y) \\ &\leq \|\nabla f(x) - \nabla f(y)\|_2 \, \|x - y\|_2,\end{aligned}$$

using the Cauchy-Schwartz inequality. Therefore, we have $\|\nabla f(x) - \nabla f(y)\|_2 \leq \gamma \|x - y\|_2$ for all $x, y \in \mathbb{R}^n$. The claimed convergence result thus follows from a proposition stated in class (see Prop 1.2.3 Bertsekas).

Problem 2:

(a) For some fixed integer $n \geq 1$, consider the set

$$A = \left\{ (c_0, c_1, \ldots, c_{2n}) \in \mathbb{R}^{2n+1} \mid \sum_{i=0}^{2n} c_i t^i \geq 0 \quad \text{for all } t \in [0,1] \right\}. \tag{2}$$

Show that A is a convex cone.

(b) Is A a pointed cone? Prove or disprove.

(c) Given a cone K, define the notion of its dual K^*. Show that

(i) K^* is convex (even if K is *not* convex).

(ii) If K has a non-empty interior, then K^* is pointed.

(d) Compute the dual cone A^* to the cone A from part (a).

Hint: you can use the following facts:

(i) given a set $S \subseteq \mathbb{R}^d$, every point in the conic hull of S can be written as a conic combination of some d points from S.

(ii) for any cone K, we have $K^{**} = \mathrm{cl}(\mathrm{convhull}(K))$, where cl means taking closure of a set, and convhull means taking the convex hull.

Solution:

(a) It is clear that A is a cone, since if $\sum_{i=0}^{2n} c_i t^i \geq 0$ for all $t \in [0,1]$, then certainly the same inequalities hold for $\lambda\, c$, for any $\lambda \geq 0$. To establish convexity, if c and d both belong to A, then we have

$$\sum_{i=0}^{2n} \big(\lambda c_i + (1-\lambda) d_i\big) t^i = \lambda \sum_{i=0}^{2n} c_i t^i + (1-\lambda) \sum_{i=0}^{2n} d_i t^i \;\geq\; 0.$$

(b) We claim that A is pointed. To prove this, we need to show that if $c \in A$ and $-c \in A$, then $c = 0$. Note that both c and $-c$ must satisfy the inequalities

$$\sum_{i=0}^{2n} c_i t^i \geq 0, \;\text{ and }\; \sum_{i=0}^{2n} (-c_i) t^i \geq 0, \;\text{ for all } t \in [0,1].$$

This implies that the polynomial $f(t) = \sum_{i=0}^{2n} c_i t^i$ is identically zero on $[0,1]$. Therefore, we must have $f(0) = c_0 = 0$. Moreover, we must have $f'(0) = c_1 = 0$. Continuing with this procedure (taking derivatives and evaluating at zero), we conclude that $c_i = 0$ for all $i = 0, 1, \ldots, 2n$, so that $c = 0$ as claimed.

(c) The dual cone $K^* = \left\{ y \in \mathbb{R}^n \mid y^T x \geq 0 \text{ for all } x \in K \right\}$.

(i) Suppose that $y, z \in K^*$. Then for any $x \in K$, we have

$$(\lambda y + (1-\lambda) z)^T x = \lambda y^T x + (1-\lambda) z^T x \geq 0,$$

so that $\lambda y + (1-\lambda) z \in K^*$ as well. Therefore, K^* is always convex.

(ii) Suppose that z and $-z$ are both elements of K^*; we want to show that $z = 0$. For any element x of K, we must have

$$z^T x \geq 0 \quad \text{and} \quad (-z)^T x \geq 0,$$

so that $z^T x = 0$ for all elements of K. But if K has a non-empty interior, then we can find a point $x \in K$ and a $\delta > 0$ such that the ball $\mathbb{B}_\delta(x) = \{y \in \mathbb{R}^n \mid \|y - x\|_2 \leq \delta\}$ is contained with K. If $z \neq 0$, then the vector $x + \delta \frac{z}{\|z\|_2}$ belongs to this ball, whence we must have

$$z^T\big(x + \delta \frac{z}{\|z\|_2}\big) = z^T x + \delta \|z\|_2 = 0.$$

Since $z^T x = 0$, this implies that $\delta \|z\|_2 = 0$, which is a contradiction.

(d) We claim that A^* is equivalent to the set

$$B = \text{conic hull}\big\{(1, t, t^2, \dots, t^{2n}) \mid t \in [0,1]\big\}.$$

To prove this claim, we first prove that $A = B^*$. Indeed, any $z \in B$ can be written as $z = \sum_{i=1}^m \lambda_i (1, t_i, \dots, t_i^{2n})$ for some $\lambda_i \geq 0$ and $t_i \in [0,1]$. Consequently, given any $c \in A$, we have

$$c^T z = \sum_{k=0}^{2n} c_k z_k = \sum_{i=1}^{m} \lambda_i (\sum_{k=0}^{2n} c_k t_i^k) \geq 0$$

which implies $c \in B^*$, and hence shows that $A \subseteq B^*$ For the reverse inclusion, suppose that $c \notin A$. Then there must exist some $t \in [0,1]$ such that $\sum_{k=0}^{2n} c_k t^k < 0$. Now if we choose $z = (1, t, \dots, t^{2n}) \in B$, we are guaranteed $c^T z < 0$, which implies that $c \notin B^*$. We have thus shown that $A^c \subseteq (B^*)^c$. Since $A \subseteq B^*$ and $A^c \subseteq (B^*)^c$, we must have $A = B^*$, therefore $A^* = B^{**}$.

The second step is to prove that B is a closed set.

Let z be a point in the boundary of B, that is there exists a sequence $z^k \to z$ as $k \to \infty$ and $z^k \in B, \forall k$. Since $S = \{(1, t, \dots, t^{2n}) | t \in [0,1]\} \subset \mathbb{R}^{2n+1}$, every point in the conic hull of S can be written as a conic combination of some $2n+1$ points from S. We thus can write $z^k = \sum_{i=0}^{2n} \lambda_{k,i}(1, t_{k,i}, \dots, t_{k,i}^{2n})$ where $\lambda_{k,i} \geq 0$ and $t_{k,i} \in [0,1]$.

Since $t_k = (t_{k,0}, \dots, t_{k,2n}) \in [0,1]^{2n+1}$ - compact set, we can chose a sub sequence $K \subseteq \{1, 2, \dots\}$ such that t_k converges to a point $\bar{t} \in [0,1]^{2n+1}$ as $k \to \infty, k \in K$.

Furthermore, $\sum_{i=1}^{2n} \lambda_{k,i} \to z_1$ as $k \to \infty, k \in K$ and $\lambda_{k,i} \geq 0$, there exists L such that $\lambda_{k,i} \leq L, \forall k, i$. So $\lambda_k = (\lambda_{k,0}, \dots, \lambda_{k,2n}) \in [0, L]^{2n+1}$ - compact set thus we can choose a subsequence $K' \subseteq K$ such that $\lambda_k \to \bar{\lambda} \in [0, L]^{2n+1}$ as $k \to \infty, k \in K'$

Since $z^k \to z$ as $k \to \infty, k \in K'$, we must have $z = \sum_{i=0}^{2n} \bar{\lambda}_i (1, \bar{t}_i, \dots, \bar{t}_i^{2n}) \in B$.

Finally, using the hint, B^{**} is the closure of the convex hull of B (see also B V, §2.6.1). In our case, B is closed and convex, so that we are guaranteed that $A^* = B^{**} = B$, as claimed.

Problem 3:

True or false? Justify your answer. (No points for simply writing down T or F; answers must be justified with an explicit argument.)

(a) Given a closed convex set $C \subseteq \mathbb{R}^n$, the projection of any point $z \in \mathbb{R}^n$ under the ℓ_1-norm onto C exists and is unique. (Recall that the ℓ_1-norm is $\|x\|_1 = \sum_{i=1}^n |x_i|$.)

(b) The set

$$C = \left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid \frac{\|x\|_2^2}{1+\|x\|_2^2} \leq t\right\}$$

is convex.

(c) Let $K_1 \subseteq \mathbb{R}^n$ and $K_2 \subseteq \mathbb{R}^n$ be two convex cones with non-empty interiors. If the interiors of the cones are disjoint, then there exists a non-zero vector $y \in \mathbb{R}^n$ such that $y \in K_1^*$ and $-y \in K_2^*$.

(d) Given a symmetric matrix $A \succeq 0$, vector $b \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$, the set

$$C = \left\{x \in \mathbb{R}^n \mid x^T A x + b^T x + c \leq 0\right\}.$$

is convex.

Solution:

(a) False: uniqueness need not hold because the ℓ_1-norm is not strictly convex. Consider the problem

$$\min_{x \in \mathbb{R}^2} |x_1| + |x_2| \quad \text{such that } x_1 + x_2 = 2.$$

This corresponds to projecting $(0,0)$ onto the given line. This linear program has infinitely many optima: e.g., $x^* = (1,1)$ and $y^* = (2,0)$ are both optimal solutions.

(b) False. Consider $n = 1$. Both of the points $(0,0)$ and $(2,4/5)$ belong to the set. The average of these two points is $(1,2/5)$. However $1/(1+1) = 1/2 > 2/5$ so it fails to belong to the set.

(c) True. Since $\text{int}(K_1)$ and $\text{int}(K_2)$ are convex and disjoint, they can be separated by a separating hyperplane of the form $\{y^T x = \alpha\}$ for some $y \neq 0$. (Explicitly, we have $y^T x \geq \alpha, \forall x \in \text{int}(K_1)$ and $y^T x \leq \alpha, \forall x \in \text{int}(K_2)$.) By taking limits for boundary points, we also have $y^T x \geq \alpha, \forall x \in K_1$ and $y^T x \leq \alpha, \forall x \in K_2$. We must have $\alpha = 0$ because K_1 and K_2 are cones (so if $x \in K_1$ then $tx \in K_1$ for all $t > 0$). This means that $y \in K_1^*$ and $-y \in K_2^*$, as claimed.

(d) True. If $A \succeq 0$, then the function $f(x) = x^T A x + b^T x + c$ is convex. The given set is the sub-level set $\{x \in \mathbb{R}^n \mid f(x) \leq 0\}$ and so must be convex (since f is a convex function).