UC Berkeley
Department of Electrical Engineering and Computer Science

EECS 227A
Nonlinear and Convex Optimization

Solutions 1
Fall 2009

The first part of this problem set provides some practice on the mathematical pre-requisites for this course (vector calculus, elementary analysis, and linear algebra) and will calibrate your degree of preparation for the course.

Reading: Boyd and Vandenberghe, §3.1, 3.2 and Appendix A for background material.

Solution 1.1

Consider the eigenvector decomposition $Q = \sum_{i=1}^{n} \lambda_i u_i u_i^T$.

First, suppose that Q is positive semidefinite. In this case, we must have $u_k^T Q u_k \geq 0$ for all $k = 1, \ldots, n$. But since the eigenvectors $\{u_i\}$ are all orthonormal, we have $u_k^T Q u_k = \lambda_k \geq 0$ for all $k = 1, \ldots, n$.

In the other direction, suppose that all eigenvalues of Q are non-negative (i.e. $\lambda_k \geq 0$ for all $k = 1, \ldots, n$). Then for any $x \in \mathbb{R}^n$, we have

$$x^T Q x = \sum_{i=1}^{n} \lambda_i (x^T u_i)^2 \geq 0$$

so Q is positive semidefinite.

Solution 1.2

By decomposing $x = \begin{bmatrix} x_A \\ x_C \end{bmatrix}$, we have

$$x^T M x = x_A^T A x_A + 2x_A^T B x_C + x_C^T C x_C$$

(a) ⇒ (b)
Let $x_C = 0$, since $x^T M x > 0, \forall x \neq 0$ we also have $x_A^T A x_A > 0, \forall x_A \neq 0$, thus A is positive definite (so it is invertible).
Let $x_A = -A^{-1} B x_C$, then $x^T M x = x_C^T (C - B^T A^{-1} B) x_C$. Again, since $x^T M x > 0, \forall x \neq 0$ we also have $x_C^T (C - B^T A^{-1} B) x_C > 0, \forall x_C \neq 0$, thus $C - B^T A^{-1} B$ is positive definite

(b) ⇒ (a)

$$\begin{aligned} \min_{x \neq 0} x^T M x &= \min \left\{ \min_{x_C \neq 0} \min_{x_A} x_A^T A x_A + 2x_A^T B x_C + x_C^T C x_C, \min_{x_C = 0} \min_{x_A \neq 0} x_A^T A x_A + 2x_A^T B x_C + x_C^T C x_C \right\} \\ &= \min \left\{ \min_{x_C \neq 0} \min_{x_A} x_A^T A x_A + 2x_A^T B x_C + x_C^T C x_C, \min_{x_A \neq 0} x_A^T A x_A \right\} \end{aligned}$$

Since A is positive definite, assuming that x_C is fixed, $x_A^T A x_A + 2x_A^T B x_C + x_C^T C x_C$ achieves minimum at $x_A = -A^{-1} B x_C$, thus

$$\min_{x \neq 0} x^T M x = \min \left\{ \min_{x_C \neq 0} x_C^T (C - B^T A^{-1} B) x_C ; \min_{x_A \neq 0} x_A^T A x_A \right\}$$

As A and $C - B^T A^{-1} B$ is positive definite, we have $\min_{x \neq 0} x^T M x > 0$ therefore M is positive definite.

Solution 1.3

(a) False. A counter example is $\{a_i\}$ such that $a_i = \frac{1}{i}, \forall i$. Indeed,

$$s_{2^k-1} = \sum_{i=1}^{2^k-1} \frac{1}{i} = \sum_{j=1}^{k} \sum_{i=2^{j-1}}^{2^j-1} \frac{1}{i} \geq \sum_{j=1}^{k} 2^{j-1} \frac{1}{2^j - 1} \geq \frac{k}{2}$$

So $s_{2^k-1} \to +\infty$ as $k \to +\infty$, therefore s_n does not converge.

(b) True.
$x^T Q y = x^T(\mu y) = \mu x^T y$
$x^T Q y = x^T Q^T y = (Qx)^T y = \lambda x^T y$
So $\mu x^T y = \lambda x^T y$, but $\lambda \neq \mu$ we must have $x^T y = 0$, thus $x^T Q y = 0$

(c) False. Consider $M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, then $U^T M U = \begin{bmatrix} 0 & \alpha \\ -\alpha & 0 \end{bmatrix}$ for some α.
Thus $U^T M U$ is diagonal implies that $U^T M U = 0 \Rightarrow U U^T M U U^T = 0 \Rightarrow M = 0$ (contradiction)

(d) True. Using mean value theorem, for all $s, t \in [0,1]$ and $s \neq t$, there exists $\xi \in [s,t] \subseteq [0,1]$ such that

$$f'(\xi) = \frac{f(t) - f(s)}{t - s}$$

Since $|f'(\xi)| \leq 0.1$, we have $|f(t) - f(s)| \leq 0.1|t - s|$

(e) False. Consider $\mathbb{R}^1$, we construct a counter example as following: for each $k = 0, 1, \ldots,$

$$x_{2^k+i} = (-1)^k + i\frac{(-1)^{k+1} - (-1)^k}{2^k}, \forall 0 \leq i \leq 2^k - 1$$

Clearly $||x_i - x_{i+1}||_2 \leq \frac{1}{2^{[\log_2 i]-1}}$ so for all $\epsilon > 0$, there exists some $N(\epsilon)$ such that for all $i \geq N(\epsilon)$, we have $||x_i - x_{i+1}||_2 < \epsilon$. However x_n does not converge because $x_{2^{2k}} = 1$ and $x_{2^{2k+1}} = -1$ for all k

Note: If we had assumed that there exists an $N(\epsilon)$ such that $\|x_i - x_j\|_2 \leq \epsilon$ *for all* $i, j \geq N(\epsilon)$ (not just consecutive ones), then the statement would be true, since any Cauchy sequence in $\mathbb{R}$ converges to some element of $\mathbb{R}$.

Solution 1.4

(a) Since x^* is a local minimum of $g_d(\alpha) = f(x^* + \alpha d)$, we can apply the first-order necessary optimality conditions to g_d so as to conclude $g_d'(0) = \nabla f(x^*)^T d = 0$. Since this statement holds for all $d \in \mathbb{R}^n$, we have have $\nabla f(x^*) = 0$.

(b) Using the hint, we first show that $f(0,0)$ is a local minimum along all lines. For $\beta \neq 0$, let $x_2 = \beta x_1$ be an arbitrary line through $(0,0)$. We have

$$h(x_1) := f(x_1, \beta x_1) = (\beta x_1 - p x_1^2)\,(\beta x_1 - q x_1^2) \;=\; pq x_1^4 - \beta\,(p+q) x_1^3 + \beta^2 x_1^2$$

Taking first and second derivatives:

$$\begin{aligned} h'(x_1) &= 4pq x_1^3 - 3\beta\,(p+q) x_1^2 + 2\beta^2 x_1 \\ h''(x_1) &= 12pq x_1^2 - 6\beta\,(p+q) x_1 + 2\beta^2 \end{aligned}$$

so that $h'(0) = 0$ and $h''(0) = 2\beta^2 > 0$ for $\beta \neq 0$. Thus, by the sufficient optimality conditions, the point $(0,0)$ is a local minimum on every line through the origin.

Now suppose that we take a parabolic path to the origin — say $x_2 = \gamma x_1^2$. Then

$$f(x_1, \gamma x_1^2) = \left(\gamma x_1^2 - p x_1^2\right)\left(\gamma x_1^2 - q x_1^2\right) \;=\; (\gamma - p)\,(\gamma - q) x_1^4$$

For any $\gamma \in (p, q)$, there holds $f(x_1, \gamma x_1^2) < 0$ for all $x_1 \neq 0$, so that $(0,0)$ cannot be a local minimum.

Solution 1.5

(a) To prove $(i) \implies (ii)$: Say x^* is a global minimum. From the necessary first-order condition $\nabla f(x^*) = 0$, we have $Qx^* = b$, whence $b \in \text{Range}(Q)$. From the necessary second-order condition $\nabla^2 f(x^*) \succeq 0$, we obtain $Q \succeq 0$.
Turning to $(ii) \implies (i)$: suppose that $b = Qx^*$ for some $x^* \in \mathbb{R}^n$ and $Q \succeq 0$. Then we can write

$$\begin{aligned} f(x) &= \frac{1}{2} x^T Q x - x^T Q x^* \\ &= \frac{1}{2}\,[x - x^*]^T\, Q\, [x - x^*] \;-\; \frac{1}{2} (x^*)^T Q x^* \\ &\geq -\frac{1}{2} (x^*)^T Q x^*, \end{aligned}$$

with equality holding for $x = x^*$. The final inequality relies on the fact $Q \succeq 0$.

(b) For (i), set $Q = I$ and $b = 0$. For (ii), set

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and } b = 0.$$

For (iii) set $Q = -I$ and $b = 0$.

Solution 1.6

Test solution

(a) We know that Jensen's inequality holds when $n = 2$. Now suppose that it is true up to n, we will prove that if $f : \mathbb{R}^d \to \mathbb{R}$ is convex then

$$f(\sum_{i=1}^{n+1} \lambda_i x_i) \leq \sum_{i=1}^{n+1} \lambda_i f(x_i), \forall x_i \in \mathbb{R}^d, \lambda_i \geq 0 \text{ and } \sum_{i=1}^{n+1} \lambda_i = 1$$

Indeed, WLOG assume $\lambda_{n+1} > 0$, let $\bar{x}_n = \dfrac{\lambda_n x_n + \lambda_{n+1} x_{n+1}}{\lambda_n + \lambda_{n+1}}$ and $\bar{\lambda}_n = (\lambda_n + \lambda_{n+1})$, we have

$$f(\sum_{i=1}^{n+1} \lambda_i x_i) = f(\sum_{i=1}^{n-1} \lambda_i x_i + \bar{\lambda}_n \bar{x}_n) \leq \sum_{i=1}^{n-1} \lambda_i f(x_i) + \bar{\lambda}_n f(\bar{x}_n) \text{ (by induction hypothesis)}$$

Furthermore,

$$f(\bar{x}_n) = f\left(\frac{\lambda_n}{\lambda_n + \lambda_{n+1}} x_n + \frac{\lambda_{n+1}}{\lambda_n + \lambda_{n+1}} x_{n+1}\right) \leq \frac{\lambda_n}{\bar{\lambda}_n} f(x_n) + \frac{\lambda_{n+1}}{\bar{\lambda}_n} f(x_{n+1})$$

Therefore

$$f(\sum_{i=1}^{n+1} \lambda_i x_i) \leq \sum_{i=1}^{n+1} \lambda_i f(x_i)$$

(b) Let $f(x) = e^x$, and $\lambda_i = \frac{1}{n}, \forall i$. Since $f(x)$ is convex, we can use Jensen's inequality

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = f\left(\sum_{i=1}^{n} \frac{1}{n} \log x_i\right) \leq \sum_{i=1}^{n} \frac{1}{n} f(\log x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

(c) Let $f(x) = -\log x$

$$h(x) = -\sum_{i=1}^{n} x_i \log x_i = -\sum_{i=1}^{n} x_i (-\log \frac{1}{x_i}) = -\sum_{i=1}^{n} x_i f\left(\frac{1}{x_i}\right)$$

Since $f(x)$ is a convex and $x_i \geq 0, \sum_{i=1}^{n} x_i = 1$, using Jensen's inequality we have

$$\sum_{i=1}^{n} x_i f\left(\frac{1}{x_i}\right) \geq f\left(\sum_{i}^{n} x_i \frac{1}{x_i}\right) = f(n) = -\log n$$

Therefore $h(x) \leq \log n$

(d) Consider convex function $f(x) = e^x$, using Jensen's inequality we have

$$f\left(\frac{p \log |a|}{p} + \frac{q \log |b|}{q}\right) \leq \frac{f(p \log |a|)}{p} + \frac{f(q \log |b|)}{q}$$

This proves Young's inequality: $ab \leq \dfrac{|a|^p}{p} + \dfrac{|b|^q}{q}$.
Now by Young's inequality,

$$\sum_{i=1}^{n} \frac{x_i}{(\sum_{i=1}^{n} |x_i|^p)^{1/p}} . \frac{y_i}{(\sum_{i=1}^{n} |y_i|^q)^{1/q}} \leq \sum_{i=1}^{n} \left(\frac{1}{p} \frac{|x_i|^p}{\sum_{i=1}^{n} |x_i|^p} + \frac{1}{q} \frac{|y_i|^q}{\sum_{i=1}^{n} |y_i|^q}\right) = 1$$

Therefore,

$$x^T y = \sum_{i=1}^{n} x_i y_i \leq ||x||_p ||y||_q$$

Solution 1.7
(a) Using the fact that the tangent approximation is an underestimate for any convex differentiable f, we have for any $x, y \in C$,

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^T (y - x) \\ f(x) &\geq f(y) + \nabla f(y)^T (x - y) \end{aligned}$$

Adding these two inequalities and re-arranging yields the monotonicity of the gradient mapping.
(b) A counterexample is given by the function $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$G(x) \; := \begin{bmatrix} x_1 + 2x_2 \\ x_2 \end{bmatrix}$$

If G were the gradient of any function f, then f would be twice continuously differentiable, implying that its second order mixed derivatives would agree (i.e., $\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}$). In terms of G, this would imply that $\frac{\partial G_1}{\partial x_2} = \frac{\partial G_2}{\partial x_1}$, which does not hold in this case. Therefore, G cannot be the gradient of any function. Moreover,

$$\begin{aligned} \left[G(x) - G(y)\right]^T (x - y) &= [x_1 + 2x_2 - y_1 - 2y_2, \; x_2 - y_2]^T [x_1 - y_1, x_2 - y_2] \\ &= (x_1 - y_1)^2 + 2(x_2 - y_2)(x_1 - y_1) + (x_2 - y_2)^2 \\ &= [(x_1 - y_1) + (x_2 - y_2)]^2 \geq 0, \end{aligned}$$

so that G is monotonic.