thanks: Vris CHEUNG
Department of Combinatorics and Optimization
University of Waterloo

November 12, 2009

CONVEX OPT. AND ANALYSIS — Assignment 3

1 Convex Functions, Convex Sets, Fenchel Conjugates

Question 1. We prove that (a) $\Longrightarrow$ (c) $\Longrightarrow$ (b) $\Longrightarrow$ (a).

(a) $\Longrightarrow$ (c) : Let $\{x_n, \alpha_n\}_n \subset \text{epi}(f)$ be a sequence that converges to (x, α). Since $f(x_n) \leq \alpha_n \ \forall\ n$, we have

$$f(x) \leq \liminf f(x_n) \leq \liminf \alpha_n = \lim \alpha_n = \alpha \quad .$$

Thus $(x, \alpha) \in \text{epi}(f)$, showing that $\text{epi}(f)$ is closed.

(c) $\Longrightarrow$ (b) : Let $\{x_n\}_n \subset L_\alpha$ be a sequence that converges to x. Then $(x_n, \alpha) \subset \text{epi}(f)$ converges to (x, α). By the closedness of $\text{epi}(f)$, $(x, \alpha) \in \text{epi}(f)$, so $f(x) \leq \alpha$, that is, $x \in L_\alpha$. This shows that L_α is closed.

(b) $\Longrightarrow$ (a) : We shall prove the contrapositive argument. Suppose f is not lower semi-continuous at some $x \in \mathbb{E}$. Then there exists a sequence $\{x_n\}_n$ that converges to x but $\liminf f(x_n) < f(x)$. This means we can pick a subsequence $\{x_{n_k}\}_k$ of $\{x_n\}_n$ such that $\lim_k f(x_{n_k}) = \liminf f(x_n)$, and an $\alpha \in \mathbb{R}$ such that $\liminf f(x_n) < \alpha < f(x)$. Then $\lim_k f(x_{n_k}) < \alpha < f(x)$, so there exists k_0 such that for all $k \geq k_0$, $f(x_{n_k}) < \alpha < f(x)$. Consequently, $x_{n_k} \in L_\alpha$ converges to x but $x \notin L_\alpha$. Hence, L_α is not closed.

Question 2. Let $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by $\phi(x, x^*) = x \cdot x^* - |x| = |x|\big(x^* \text{sgn}(x) - 1\big)$. Note that $f^*(x^*) = \sup_x \phi(x, x^*)$. Fix any x^*; two things could happen:

Case 1: $|x^*| - 1 > 0$. Let $x = \text{sgn}(x^*)$. Then for all $\lambda > 0$, $\text{sgn}(\lambda x) = \text{sgn}(x^*)$, and

$$\phi(\lambda x, x^*) = |\lambda x|\big(x^* \text{sgn}(\lambda x) - 1\big) = \lambda(|x^*| - 1) \to +\infty \quad \text{as } \lambda \to +\infty.$$

Thus $f^*(x^*) = \sup_x \phi(x, x^*) = +\infty$ if $|x^*| - 1 > 0$.

Case 2: $|x^*| - 1 \le 0$**.** In this case,

$$\phi(x, x^*) = x \cdot x^* - |x| = |x|\big(x^*\mathrm{sgn}(x) - 1\big) \le |x|\big(|x^*| - 1\big) \le 0 \quad ,$$

and $\phi(0, x^*) = 0$. Thus $f^*(x^*) = \sup_x \phi(x, x^*) = 0$ if $|x^*| - 1 \le 0$.

Therefore $f^* = \delta_{[-1,1]}$.

Question 3. Given Euclidean spaces $\mathbb{E}$, Y and a linear map $A : \mathbb{E} \to Y$, we first prove the following key lemma:

Lemma 1 *If* $f : \mathbb{E} \to (-\infty, +\infty]$ *and* $g : Y \to (-\infty, +\infty]$ *satisfy*

$$0 \in int(dom(g) - Adom(f)) \quad , \tag{1.1}$$

then $0 \in \partial(f + g \circ A)(\bar{x})$ *implies* $0 \in \partial f(\bar{x}) + A^* \partial g(A\bar{x})$.

Proof It immediately follows from the definition of subdifferentials that $0 \in \partial(f + g \circ A)(\bar{x})$ implies $\bar{x}$ is a global minimizer of $f + g \circ A$ on $\mathbb{E}$. If (1.1) holds, by Theorem 3.3.5 of Borwein and Lewis, we have that

$$f(\bar{x}) + g(A\bar{x}) = \inf_{\mathbb{E}} \{f + g \circ A\} = \sup_{\phi \in Y} \{-f^*(A^*\phi) - g^*(-\phi)\} = -f^*(A^*\bar{\phi}) - g^*(-\bar{\phi})$$

for some $\bar{\phi} \in Y$. Hence $[f(\bar{x}) + f^*(A^*\bar{\phi})] + [g(A\bar{x}) + g^*(-\bar{\phi})] = 0$.

On the other hand, by the Fenchel-Young inequality, we have

$$\begin{aligned} f(\bar{x}) + f^*(A^*\bar{\phi}) &\geq \langle \bar{x}, A^*\bar{\phi} \rangle = \langle A\bar{x}, \bar{\phi} \rangle \quad \text{(equality holds iff } A^*\bar{\phi} \in \partial f(\bar{x});) \\ g(A\bar{x}) + g^*(-\bar{\phi}) &\geq \langle A\bar{x}, -\bar{\phi} \rangle \quad \text{(equality holds iff } -\bar{\phi} \in \partial g(A\bar{x}).) \end{aligned}$$

Summing the two inequalities gives $[f(\bar{x}) + f^*(A^*\bar{\phi})] + [g(A\bar{x}) + g^*(-\bar{\phi})] \geq 0$. But the strong duality theorem mentioned above says that we have equality. This implies that

$$\begin{aligned} f(\bar{x}) + f^*(A^*\bar{\phi}) = \langle A\bar{x}, \bar{\phi} \rangle \quad &\Longrightarrow \quad A^*\bar{\phi} \in \partial f(\bar{x}); \text{ and} \\ g(A\bar{x}) + g^*(-\bar{\phi}) = \langle A\bar{x}, -\bar{\phi} \rangle \quad &\Longrightarrow \quad -\bar{\phi} \in \partial g(A\bar{x}) \\ &\Longrightarrow \quad -A^*\bar{\phi} \in A^*\partial g(A\bar{x}). \end{aligned}$$

This shows that $0 = A^*\bar{\phi} - A^*\bar{\phi} \in \partial f(\bar{x}) + A^*\partial g(A\bar{x})$. $\square$

Given any $f : \mathbb{E} \to (-\infty, +\infty]$, $g : Y \to (-\infty, +\infty]$ and any linear map $A : \mathbb{E} \to Y$, we have $\partial f(x) + A^*\partial g(Ax) \subseteq \partial(f + g \circ A)(x)$ for any fixed x: suppose $x^* \in \partial f(x)$ and $y^* \in \partial g(Ax)$. Then we have that for any $u \in \mathbb{E}$,

$$\begin{aligned} &\langle x^*, u - x \rangle \leq f(u) - f(x) \quad \text{, and} \\ &\langle A^*y^*, u - x \rangle = \langle y^*, Au - Ax \rangle \leq g(Au) - g(Ax) \\ \Longrightarrow\ &\langle x^* + A^*y^*, u - x \rangle \leq (f + g \circ A)(u) - (f + g \circ A)(x) \end{aligned}$$

so $x^* + A^*y^* \in \partial(f + g \circ A)(x)$.

Now suppose that the constraint qualification

$$0 \in \operatorname{int}(\operatorname{dom}(g) - A\operatorname{dom}(f))$$

holds. We show that $\partial f(x) + A^*\partial g(Ax) = \partial(f + g \circ A)(x)$. In fact, if $x^* \in \partial(f + g \circ A)(x)$, then for all $u \in \mathbb{E}$,

$$\begin{aligned} \langle x^*, u - x\rangle &\leq (f + g \circ A)(u) - (f + g \circ A)(x) \\ \implies 0 &\leq (\tilde{f} + g \circ A)(u) - (\tilde{f} + g \circ A)(x) \quad , \end{aligned}$$

where $\tilde{f} := f + \langle -x^*, \cdot\rangle$. Therefore $0 \in \partial(\tilde{f} + g \circ A)(x)$.

Since $\langle -x^*, \cdot\rangle$ is a real-valued function, the domain of $\tilde{f} : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is the same as that of f. In particular, the constraint qualification $0 \in \operatorname{int}(\operatorname{dom}(g) - A\operatorname{dom}(\tilde{f}))$ holds. Therefore, by Lemma 1, we have that there exists some $\tilde{x}^* \in A^*\partial g(Ax)$ such that $-\tilde{x}^* \in \partial \tilde{f}(x)$. Then for any $u \in \mathbb{E}$,

$$\begin{aligned} \langle -\tilde{x}^*, u - x\rangle &\leq f(u) - f(x) - \langle x^*, u - x\rangle \\ \implies \langle -\tilde{x}^* + x^*, u - x\rangle &\leq f(u) - f(x) \quad , \end{aligned}$$

so $\tilde{x}^* + x^* \in \partial f(x)$. Consequently,

$$x^* = (-\tilde{x}^* + x^*) + \tilde{x}^* \in \partial f(x) + A^*\partial g(Ax) \quad . \qquad \square$$

Question 4(a). Given S is non-empty, open and convex, and $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is such that $\text{dom}(f) = S$.

Suppose f is a convex function. Then $\text{epi}(f)$ is a convex set. Fix any $x \in S$. Then $(x, f(x))$ is on the boundary of the closure $\text{cl}(\text{epi}(f))$, and since S is open, $\text{int}(\text{cl}(\text{epi}(f))) \neq \emptyset$. The supporting hyperplane theorem implies that $\exists\ (\alpha, \eta) \in (\mathbb{E} \times \mathbb{R}) \backslash \{(0,0)\}$ such that

$$\alpha^T x + \eta f(x) \geq \alpha^T y + \eta r, \quad \forall\ (y, r) \in \text{epi}(f).$$

First observe that since $(x, r) \in \text{epi}(f)$ for all $r \geq f(x)$, r can be arbitrarily large and the above inequality implies that $\eta \leq 0$. In fact, $\eta < 0$: if on the contrary $\eta = 0$, we have that $\alpha^T x \geq \alpha^T y$ for all $y \in S$. Since S is open, we can pick sufficiently small $\varepsilon > 0$ such that $x \pm \varepsilon\alpha \in S$. Then the above inequality implies that $\varepsilon\|\alpha\|^2 = 0$, so $\alpha = 0$, which contradicts the earlier result that $(\alpha, \eta) \neq (0, 0)$.

Now that $\eta < 0$, we may assume without loss of generality that $\eta = -1$, so we have

$$\alpha^T x - f(x) \geq \alpha^T y - r \quad (y, r) \in \text{epi}(f) \quad \Longrightarrow \quad f(y) - f(x) \geq \alpha^T (y - x) \quad \forall\, y \in \text{dom}(f) \quad .$$

In other words, $\alpha \in \partial f(x)$. This shows that $\partial f(x) \neq \emptyset$.

Conversely, if $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ is not a convex function, there exists some $x, y \in S$, $\lambda \in (0, 1)$ such that $f(z) > \lambda f(x) + (1 - \lambda) f(y)$, where $z := \lambda x + (1 - \lambda) y$. We show that $\partial f(z)$ is an empty set. Suppose on the contrary that there exists some $d \in \partial f(z)$. Then

$$\begin{aligned} d^T(x - z) \leq f(x) - f(z) \quad &\Longrightarrow \quad d^T(\lambda x - \lambda z) \leq \lambda f(x) - \lambda f(z) \\ d^T(y - z) \leq f(y) - f(z) \quad &\Longrightarrow \quad d^T[(1 - \lambda) y - (1 - \lambda) z] \leq (1 - \lambda) f(y) - (1 - \lambda) f(z) \quad . \end{aligned}$$

Summing the two inequalities on the right, we have $\lambda f(x) + (1 - \lambda) f(y) - f(z) \geq 0$, which contradicts the choice of z that $f(z) > \lambda f(x) + (1 - \lambda) f(y)$. Hence we must have $\partial f(z) = \emptyset$.

Question 4(b). If $h : \text{cl}S \to \mathbb{R}$ is convex, then h is certainly convex on S.

Conversely, suppose $h : \text{cl}S \to \mathbb{R}$ is continuous and $h|_S$ is a convex function on S. Pick any $x, y \in \text{cl}S$ and $\lambda \in [0, 1]$; then there exist sequences $\{x_n\}_n \subset S$ and $\{y_n\}_n \subset S$ such that $x_n \to x$ and $y_n \to y$. For each n, we have

$$h(\lambda x_n + (1 - \lambda) y_n) \leq \lambda h(x_n) + (1 - \lambda) h(y_n) \quad .$$

Taking $n \to \infty$, by continuity of h we have $h(\lambda x + (1 - \lambda) y) \leq \lambda h(x) + (1 - \lambda) h(y)$. Thus h is convex on $\text{cl}S$.

But the statement " $h : \text{cl}S \to \mathbb{R}$ being continuous and $h|_S$ being a *strictly* convex function on S

imply that h is *strictly* convex on $\mathrm{cl}S$" is *not* true. Consider the function $h : \mathbb{R}^2_+ \to \mathbb{R}$ defined by $h(x, y) = -x^\alpha y^\beta$, where $\alpha, \beta \in (0, 1/2)$. The function is smooth on $\mathbb{R}^2_{++}$: for any $x, y > 0$,

$$\nabla h(x, y) = \begin{pmatrix} -\alpha x^{\alpha-1} y^\beta \\ -\beta x^\alpha y^{\beta-1} \end{pmatrix} \quad \text{and} \quad \nabla^2 h(x, y) = \begin{pmatrix} \alpha(1-\alpha) x^{\alpha-2} y^\beta & -\alpha\beta x^{\alpha-1} y^{\beta-1} \\ -\alpha\beta x^{\alpha-1} y^{\beta-1} & \beta(1-\beta) x^\alpha y^{\beta-2} \end{pmatrix}$$

which has a positive trace $\alpha(1-\alpha)x^{\alpha-2}y^\beta + \beta(1-\beta)x^\alpha y^{\beta-2}$ and determinant $\alpha\beta[(1-\alpha)(1-\beta) - \alpha\beta]x^{2(\alpha-1)}y^{2(\beta-1)}$ which is positive because $1 - \alpha > 1/2 > \alpha$ and $1 - \beta > 1/2 > \beta$. Thus h is strictly convex on $\mathbb{R}^2_{++}$. But h is not strictly convex on $\mathbb{R}^2_+$ because h is identically zero on its boundary.

Question 5(a). $\psi : \mathbb{R}_+ \to \mathbb{R}$ defined by $\psi(x) = 1 - \sqrt{x}$ is convex on $\mathbb{R}_{++}$, because

$$\psi'(x) = -\frac{1}{2\sqrt{x}} \quad \text{and} \quad \psi''(x) = \frac{3}{4x^{3/2}} > 0$$

for $x > 0$. Following from Q.4(b), (right-)continuity of ψ at 0 implies that ψ is indeed convex on $\mathbb{R}_+$. Thus the set $\{(x_1, x_2; r) \in \mathbb{R}^2 \times \mathbb{R} : 1 - \sqrt{x_1} \le r\}$ is convex. Similarly, convexity of the absolute value function ensures that $\{(x_1, x_2; r) \in \mathbb{R}^2 \times \mathbb{R} : |x_2| \le r\}$ is a convex set. Next,

$$\begin{aligned} \operatorname{epi}(f) &= \{(x_1, x_2; r) \in \mathbb{R}^2 \times \mathbb{R} : 1 - \sqrt{x_1} \le r \text{ and } |x_2| \le r\} \\ &= \{(x_1, x_2; r) \in \mathbb{R}^2 \times \mathbb{R} : 1 - \sqrt{x_1} \le r\} \cup \{(x_1, x_2; r) \in \mathbb{R}^2 \times \mathbb{R} : |x_2| \le r\} \quad , \end{aligned}$$

meaning that $\operatorname{epi}(f)$ as an intersection of two convex sets is convex. Hence f is convex.

Question 5(b). We show that $(0,1), (0,-1) \in \operatorname{dom}(\partial f)$, but $(0,0) = 1/2[(0,1) + (0,-1)]$ does not lie in $\operatorname{dom}(\partial f)$.

$(0, \pm 1) \in \mathbf{dom}(\partial f)$**:** For any $(x_1, x_2) \in \mathbb{R}_+ \times \mathbb{R}$,

$$f(x_1, x_2) - f(0, \pm 1) = \max\{1 - \sqrt{x_1}, |x_2|\} - \max\{1 - \sqrt{0}, |\pm 1|\} = \max\{-\sqrt{x_1}, |x_2| - 1\} \ge |x_2| - 1 \quad .$$

Note that $|x_2| - 1 \ge \pm x_2 - 1$. Consequently,

$$\begin{aligned} f(x_1, x_2) - f(0, 1) \ge x_2 - 1 = 0 \cdot (x_1 - 0) + 1 \cdot (x_2 - 1) \quad \forall\, (x_1, x_2) \in \mathbb{R}^2 &\implies (0, 1) \in \partial f(0, 1) \\ f(x_1, x_2) - f(0, -1) \ge -x_2 - 1 = 0 \cdot (x_1 - 0) + 1 \cdot [x_2 - (-1)] \quad \forall\, (x_1, x_2) \in \mathbb{R}^2 &\implies (0, -1) \in \partial f(0, -1) \quad , \end{aligned}$$

so both $\partial f(0, 1)$ and $\partial f(0, -1)$ are non-empty.

$(0, 0) \notin \mathbf{dom}(\partial f)$**:** If $(d_1, d_2) \in \partial f(0, 0)$, then for all $x_1 \ge 0$ (taking x_2 to be constantly 0),

$$d_1 x_1 \le f(x_1, 0) - f(0, 0) = \max\{-\sqrt{x_1}, -1\} = -\sqrt{x_1}$$

for $x_1 \in (0, 1)$. Dividing both sides by x_1 (which can be done for $x_1 > 0$), we obtain $d_1 \le -1/\sqrt{x_1}$, which goes to $-\infty$ as $x \searrow 0^+$. This absurd result indicates that such (d_1, d_2) does not exist. Hence $\partial f(0, 0) = \emptyset$.

Question 6. (c.f. Prop 2.1.7 of Borwein and Lewis) We will prove some slightly more general results (at the expense of having a slightly longer proof). First we need the following lemma:

Lemma 2 *If $h : \mathbb{E} \to \mathbb{R}$ is a continuous function with bounded level sets, then* $\arg\min_{\mathbb{E}} h$ *is nonempty.*

Proof First note that h must be bounded below on $\mathbb{E}$: if there exists a sequence $\{x_n\}_n$ such that $h(x_n) \to \infty$ as $n \to +\infty$, we may assume without loss of generality that the sequence $\{h(x_n)\}_n$ is strictly decreasing. Then $\{x_n\}_n \subseteq \{x \in \mathbb{E} : h(x) \leq h(x_1)\}$ which is a bounded set, so $\{x_n\}_n$ is a bounded sequence in $\mathbb{E}$. By Weierstrass Theorem, this sequence has a convergent subsequence; by passing to that subsequence, we may assume without loss of generality that $\{x_n\}_n$ converges to some $\bar{x} \in \mathbb{E}$. By continuity of h, $h(x_n) \to h(\bar{x}) \in \mathbb{R}$ as $n \to +\infty$, contradicting the given condition that $h(x_n) \to -\infty$ as $n \to +\infty$. Therefore h must be bounded below on $\mathbb{E}$.

h being bounded below on $\mathbb{E}$ implies that $\inf_{\mathbb{E}} h \in \mathbb{R}$. Consider any minimizing sequence $\{x_n\}_n$ satisfying $h(x_n) < \inf h + n^{-1}$. Then $\{x_n\}_n \subseteq \{x \in \mathbb{E} : h(x) \leq \inf h + 1\}$ which is bounded by assumption. Again, by passing to subsequence we may assume that the sequence $\{x_n\}_n$ converges to some $\bar{x} \in \mathbb{E}$. By continuity of h,

$$\inf h \leq h(\bar{x}) = \lim_n h(x_n) \leq \lim_n \left(\inf h + \frac{1}{n}\right) = \inf h \quad ,$$

which shows that $\bar{x} \in \arg\min_{\mathbb{E}} h$. □

Remark This proof shows that any limit point of a minimizing sequence of such function h is indeed a global minimizer.

Now we prove that if $f : \mathbb{E} \to \mathbb{R}$ is differentiable and is bounded below on $\mathbb{E}$ by some $m \in \mathbb{R}$, then for any $\varepsilon > 0$, there exists some $\bar{x}_\varepsilon \in \mathbb{E}$ such that $\|\nabla f(\bar{x}_\varepsilon)\| \leq \varepsilon$. (In this question the vector norm is always assumed to be ℓ_2 norm.)

For any fixed $\varepsilon > 0$, define the function $f_\varepsilon : \mathbb{E} \to \mathbb{R}$ by $f_\varepsilon = f + \varepsilon\|\cdot\|$. This function is continuous, and any level set $S_\alpha(f_\varepsilon) = \{x \in \mathbb{E} : f_\varepsilon(x) \leq \alpha\} = \{x : \|x\| \leq \varepsilon^{-1}(\alpha - f(x)) \leq \varepsilon^{-1}(\alpha - m)\}$ is bounded. By the lemma, f_ε must have a global minimizer $\bar{x}_\varepsilon$. It follows that for any $t > 0$,

$$\begin{aligned}
& f_\varepsilon(\bar{x}_\varepsilon) \leq f_\varepsilon(\bar{x}_\varepsilon - t\nabla f(\bar{x}_\varepsilon)) \\
\implies \quad & -\varepsilon\|t\nabla f(\bar{x}_\varepsilon)\| \leq -\varepsilon(\|\bar{x}_\varepsilon\| - \|\bar{x}_\varepsilon - t\nabla f(\bar{x}_\varepsilon))\|) \leq f(\bar{x}_\varepsilon - t\nabla f(\bar{x}_\varepsilon)) - f(\bar{x}_\varepsilon) \\
\implies \quad & -\varepsilon\|\nabla f(\bar{x}_\varepsilon)\| \leq \frac{f(\bar{x}_\varepsilon - t\nabla f(\bar{x}_\varepsilon)) - f(\bar{x}_\varepsilon)}{t} \to \nabla f(\bar{x}_\varepsilon)^T[-\nabla f(\bar{x}_\varepsilon)] \quad \text{as } t \searrow 0 \\
\implies \quad & \|\nabla f(\bar{x}_\varepsilon)\| \leq \varepsilon
\end{aligned}$$

As for convex function $f : \mathbb{E} \to \mathbb{R}$ that is bounded below on $\mathbb{E}$, we have the following result:

Claim 1 *For any $\varepsilon > 0$, there exists $\bar{x}_\varepsilon$, $\bar{\phi}_\varepsilon \in \mathbb{E}$ such that $\bar{\phi}_\varepsilon \in \partial f(\bar{x}_\varepsilon)$ and $\|\bar{\phi}_\varepsilon\| \leq \varepsilon$.*

Before proving Claim 1, we need the following lemma:

Lemma 3

$$\partial(\varepsilon\|\cdot\|)(\bar{x}_\varepsilon) = \begin{cases} \left\{\varepsilon\frac{\bar{x}_\varepsilon}{\|\bar{x}_\varepsilon\|}\right\} & \textit{if } \bar{x}_\varepsilon \neq 0 \\ B(0,\varepsilon) & \textit{if } \bar{x}_\varepsilon = 0 \end{cases} ,$$

Proof First observe that for any $\lambda > 0$, any function $h : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}$ and $x \in \mathrm{dom}(h)$, $\partial(\lambda h)(x) = \lambda\partial h(x)$:

$$\begin{aligned}\partial(\lambda h)(x) &= \{d \in \mathbb{E} : \langle d, y - x\rangle \leq \lambda[h(y) - h(x)]\ \forall y \in \mathbb{E}\} \\ &= \{\lambda d \in \mathbb{E} : \langle d, y - x\rangle \leq h(y) - h(x)\ \forall y \in \mathbb{E}\} = \lambda\partial h(x)\end{aligned}$$

Therefore to prove the claim, it suffices to show that

$$\partial\|\cdot\|(\bar{x}_\varepsilon) = \begin{cases} \left\{\frac{\bar{x}_\varepsilon}{\|\bar{x}_\varepsilon\|}\right\} & \text{if } \bar{x}_\varepsilon \neq 0 \\ B(0,1) & \text{if } \bar{x}_\varepsilon = 0 \end{cases} .$$

When $x \neq 0$, since $\|\cdot\| = \sqrt{\|\cdot\|^2}$, $x \mapsto \|x\|$ is indeed differentiable:

$$\nabla\|\cdot\|(x) = \nabla\sqrt{\|\cdot\|^2}(x) = \frac{2x}{2\sqrt{\|x\|^2}} = \frac{x}{\|x\|} ,$$

so $\partial\|\cdot\|(x) = \{\frac{x}{\|x\|}\}$ when $x \neq 0$.

When $x = 0$, using the variational form $\|d\|_2 = \sup\{\langle d, x\rangle / \|x\|_2 : x \neq 0\}$,[1] we have that

$$\begin{aligned}\partial\|\cdot\|(0) &= \{d \in \mathbb{R}^n : \langle d, x\rangle \leq \|x\|\ \forall x \in \mathbb{R}^n\} \\ &= \left\{d \in \mathbb{R}^n : \|d\|_2 = \sup_{x\neq 0}\frac{\langle d, x\rangle}{\|x\|} \leq 1\right\} \\ &= B(0,1) \quad . \quad \square\end{aligned}$$

Proof of Claim 1. We define the same f_ε for any $\varepsilon > 0$ and, by continuity and boundedness of f,[2] f_ε enjoys the same properties as described above, that is, there exists some global minimizer $\bar{x}_\varepsilon$ of f_ε.

[1] In general, for $p, q \in [1, +\infty]$ satisfying $p^{-1} + q^{-1} = 1$ (with the convention $(+\infty)^{-1} = 0$), we have that for any $d \in \mathbb{R}^n$,

$$\|d\|_p = \sup\left\{\frac{\langle d, x\rangle}{\|x\|_q} : x \in \mathbb{R}^n\backslash\{0\}\right\} .$$

[2] Recall that f as a real-valued convex function is locally Lipschitz on $\mathbb{E}$, so it is continuous on $\mathbb{E}$.

Since f_ε is also convex, we have that $0 \in \partial f_\varepsilon(\bar{x}_\varepsilon) = \partial(f + \varepsilon\|\cdot\|)(\bar{x}_\varepsilon) = \partial f(\bar{x}_\varepsilon) + \partial(\varepsilon\|\cdot\|)(\bar{x}_\varepsilon)$.[3] Since

$$\partial(\varepsilon\|\cdot\|)(\bar{x}_\varepsilon) = \begin{cases} \left\{\varepsilon \frac{\bar{x}_\varepsilon}{\|\bar{x}_\varepsilon\|}\right\} & \text{if } \bar{x}_\varepsilon \neq 0 \\ B(0, \varepsilon) & \text{if } \bar{x}_\varepsilon = 0 \end{cases},$$

there exists some $\bar{\phi}_\varepsilon$ of norm ε lying in $\partial f(\bar{x}_\varepsilon)$. □

Remark The function f_ε defined in the question is indeed a "regularized" version of f. While f may not have a global minimizer, such regularization of f could give us a new function that has a minimizer. This question shows that under some assumptions on the function f, the global minimizer from the regularized function can serve as a good proxy, in a sense that it approximately satisfies the first order necessary condition of optimality.

[3] The sum rule applies at the last equality because both f and $\varepsilon\|\cdot\|$ have the whole space $\mathbb{E}$ as their domains.

Question 7(a). Consider the closed convex cone $K = \{x \in \mathbb{R}^n : x_1 \geq \sqrt{x_2^2 + \cdots + x_n^2}\}$. First we note that for any $\hat{x} \in \mathbb{R}^{n-1}$, the vector $(\|\hat{x}\|_2, \hat{x}^T)^T$ lies in K.

Let $d = (d_1, \ldots, d_n) \in N_K(0)$. Define $\hat{d} := (d_2, \ldots, d_n)^T \in \mathbb{R}^{n-1}$. Then for any $\hat{x} \in \mathbb{R}^{n-1} \backslash \{0\}$,

$$0 \geq d^T(\|\hat{x}\|_2, x^T)^T = d_1\|\hat{x}\|_2 + \hat{d}^T\hat{x}$$
$$\Longrightarrow -d_1 \geq \frac{\hat{d}^T\hat{x}}{\|\hat{x}\|_2} \quad .$$

Taking supremum over all nonzero $\hat{x} \in \mathbb{R}^{n-1}$ and using the variational form of vector norm, we have

$$-d_1 \geq \sup_{\hat{x} \in \mathbb{R}^{n-1} \backslash \{0\}} \frac{\hat{d}^T\hat{x}}{\|\hat{x}\|_2} = \|\hat{d}\|_2 = \sqrt{(-d_2)^2 + \cdots + (-d_n)^2} \quad ,$$

that is, $-d \in K$.

Conversely, let $d = (d_1, \ldots, d_n) \in K$. For any $x = (x_1, \ldots, x_n) \in K$, by Cauchy-Schwartz inequality,

$$\begin{aligned} -d^T(x - 0) &= -d_1x_1 - \sum_{i=2}^n d_ix_i \\ &\leq -d_1x_1 + \sqrt{\sum_{i=2}^n d_i^2}\sqrt{\sum_{i=2}^n x_i^2} \\ &\leq -d_1x_1 + d_1x_1 = 0 \quad . \end{aligned}$$

Therefore $-d \in N_K(0)$. Consequently, $N_K(0) = -K$.

Question 7(b). Consider the closed convex cone $K = \mathcal{S}_+^n$ in the Euclidean space $(\mathcal{S}^n, \langle\cdot,\cdot\rangle_F)$. (Recall that the Frobenius norm is defined by $\langle X, Y\rangle_F = \text{trace}(X^TY)$.) Before proving $N_K(0) = -K$, we recall the following lemma which follows easily from linear algebra:

Lemma 4

$X \in \mathcal{S}_+^n$ if and only if trace$(XY) \geq 0$ for all $Y \in \mathcal{S}_+^n$.

Proof If $X \in \mathcal{S}_+^n$, then for any $Y = \sum_{i=1}^n \lambda_i q_i q_i^T \in \mathcal{S}_+^n$ (here λ_i is the i-th largest eigenvalue of Y and q_i is the corresponding normalized eigenvector), since $\text{trace}(Xq_iq_i^T) = q_i^TXq_i \geq 0$ and $\lambda_i \geq 0$ for all i, it follows that

$$\text{trace}(XY) = \text{trace}\left[X\left(\sum_{i=1}^n \lambda_i q_i q_i^T\right)\right] = \sum_{i=1}^n \lambda_i \text{trace}\left(Xq_iq_i^T\right) \geq 0 \; .$$

Conversely, if $X \notin \mathcal{S}_+^n$, then there exists some $q \in \mathbb{R}^n$ such that $\text{trace}(Xqq^T) = q^TXq < 0$. □

From Lemma 4,

$$
\begin{aligned}
X \in K \quad &\iff \quad \langle -X, Y - 0 \rangle_F = \text{trace}(-XY) \leq 0 \ \text{ for all } Y \in K \\
&\iff \quad -X \subseteq N_K(0).
\end{aligned}
$$

Therefore $N_K(0) = -K$.

2 Convex Optimization Problems

Question 1. We restate a special case of Theorem 3.3.5 of Borwein and Lewis:

> For any $f, g : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$, let
>
> $$p = \inf_{x \in \mathbb{E}} \{f(x) + g(x)\} \quad , \text{ and}$$
> $$d = \sup_{y \in \mathbb{E}} \{-f^*(y) - g^*(-y)\} \quad .$$
>
> Then *weak duality* holds: $p \geq d$.
>
> If, in addition, f and g are convex and $\mathrm{dom}(f) \cap \mathrm{int}\,\mathrm{dom}(g) \neq \emptyset$, then *strong duality* holds: $p = d$ and there exists some $\bar{y} \in \arg\max_{y \in \mathbb{E}} -f^*(y) - g^*(-y)$.

Let A and B be any nonempty compact convex sets in $\mathbb{E}$. The map δ_A is proper (because $A \neq \emptyset$) and convex (because its domain, which equals A, is convex). By compactness of A, the "sup" in the definition of δ_A is actually attained and can be replaced by "max". The map δ_B^* is proper—and is indeed real-valued: for any $x \in \mathbb{E}$,

$$\delta_B^*(x) = \sup_{y \in \mathbb{E}} \{\langle x, y \rangle - \delta_B(y)\} = \sup_{y \in B} \langle x, y \rangle \quad ,$$

which is attained by some $\bar{y} \in B$ because B is compact and $y \mapsto \langle x, y \rangle$ is a continuous map. As mentioned in class (and as will be proved at the end of the question), δ_B^* is a sublinear (and hence convex) map. In particular, δ_B^* being real-valued and convex must be continuous on $\mathbb{E}$. Moreover, $\delta_B^{**} = \delta_B$ (which holds essentially because B is closed and convex and can be shown by using separation theorem).

Also, observe that $\mathrm{dom}(\delta_A) \cap \mathrm{int}\,\mathrm{dom}(\delta_B^*) = A \cap \mathbb{E} = A \neq \emptyset$. Hence strong duality holds for the following primal-dual pair:

$$p = \inf_{x \in \mathbb{E}} \{\delta_A(x) + \delta_B^*(x)\} \text{ , and } d = \sup_{y \in \mathbb{E}} \{-\delta_A^*(y) - \delta_B^{**}(-y)\} \quad .$$

Now we simplify p and d:

$$\begin{aligned} p &= \inf_{x\in\mathbb{E}} \{\delta_A(x) + \delta_B^*(x)\} \\ &= \inf_{x\in A} \delta_B^*(x) = \min_{x\in A} \delta_B^*(x) \qquad \text{(by continuity of } \delta_B^* \text{ and compactness of } A) \\ &= \min_{x\in A} \max_{y\in B} \langle x, y\rangle \qquad ;\text{and} \end{aligned}$$

$$\begin{aligned} d &= \sup_{y\in\mathbb{E}} \{-\delta_A^*(y) - \delta_B^{**}(-y)\} \\ &= \sup_{y\in\mathbb{E}} \{-\delta_A^*(y) - \delta_B(-y)\} \\ &= \sup_{-y\in B} \{-\delta_A^*(y)\} = \max_{y\in B} \{-\delta_A^*(-y)\} \\ &= \max_{y\in B} \left\{ -\sup_{x\in A} \langle x, -y\rangle \right\} \\ &= \max_{y\in B} \min_{x\in A} \langle x, y\rangle \qquad . \end{aligned}$$

Therefore the strong duality implies that

$$\min_{x\in A} \max_{y\in B} \langle x, y\rangle = \max_{y\in B} \min_{x\in A} \langle x, y\rangle \quad .$$

Finally we prove the earlier claims about some basic properties of δ_B^*:

Claim 2 *If $B \subseteq \mathbb{E}$ is closed and convex, then δ_B^* is a sublinear (and hence convex) map, and $\delta_B^{**} = \delta_B$.*

Proof For any $\alpha, \beta \geq 0$ and $x, u \in \mathbb{E}$,

$$\delta_B^*(\alpha x + \beta u, y) = \sup_{y\in B} \{\langle \alpha x + \beta u, y\rangle\} \leq \sup_{y\in B} \alpha \langle x, y\rangle + \sup_{y\in B} \beta \langle u, y\rangle = \alpha\delta_B^*(x) + \beta\delta_B^*(u) \quad ,$$

which shows that δ_B^* is a sublinear (and hence convex) map.

Now we prove that $\delta_B^{**} = \delta_B$. Fix any $x \in \mathbb{E}$.

If $x \notin B$, then by separation theorem, there exists some non-zero $\alpha \in \mathbb{E}$ such that $\langle \alpha, x\rangle > \sup_{u\in B} \langle \alpha, u\rangle = \delta_B^*(\alpha)$. Since we saw that δ_B^* is positively homogeneous, $\langle \lambda\alpha, x\rangle - \delta_B^*(\lambda\alpha) \to +\infty$ as $\lambda \to +\infty$. Therefore $\delta_B^{**}(x) = +\infty = \delta_B(x)$.

If $x \in B$, by Fenchel-Young inequality, $\langle y, x\rangle - \delta_B^*(y) \leq \delta_B(x) = 0$ for all $y \in \mathbb{E}$, so $\delta_B^{**}(x) = \sup_{y\in\mathbb{E}}\big\{\langle x, y\rangle - \delta_B^*(y)\big\} \leq 0$. But since $\langle x, 0\rangle - \delta_B^*(0) = 0$, we have $\sup_{y\in\mathbb{E}}\big\{\langle x, y\rangle - \delta_B^*(y)\big\} = 0$, so $\delta_B^{**}(x) = \delta_B(x)$.

Therefore $\delta_B = \delta_B^{**}$. □

Submitted on November 12, 2009.