

Exercise 7.1

- (a) For $x, y \in (0, \infty)$, $x \neq y$, $\lambda \in (0, 1)$ we know

$$\begin{aligned}
 0 &< (1 - \lambda)\lambda(x - y)^2 \\
 &= (1 - \lambda)\lambda(x^2 + y^2) + 2(\lambda - 1)\lambda xy \\
 &= (1 - \lambda)\lambda(x^2 + y^2) + (\lambda^2 + 1 - 2\lambda + \lambda^2 - 1)xy \\
 \implies xy &< (1 - \lambda)\lambda(x^2 + y^2) + (\lambda^2 + (1 - \lambda)^2)xy \\
 &= ((1 - \lambda)y + \lambda x)((1 - \lambda)x + \lambda y) \\
 \implies \frac{xy}{(1 - \lambda)x + \lambda y} &< (1 - \lambda)y + \lambda x \\
 \implies \frac{1}{(1 - \lambda)x + \lambda y} &< \frac{(1 - \lambda)}{x} + \frac{\lambda}{y}
 \end{aligned}$$

so that $x \mapsto 1/x$ is strictly convex.

- (b) Suppose Q is positive semidefinite. For $\lambda = 0 \vee 1$ it's trivial to see the convexity inequality holds (with equality). Let $x, y \in \mathbb{R}^n$, and notice for $x = 0 \vee y = 0 \vee x = y$ the claim is again trivial with equality holding, so supposing $x \neq y$ and $x, y \neq 0$ we have

$$\begin{aligned}
 ((1 - \lambda)x + \lambda y) \cdot Q((1 - \lambda)x + \lambda y) &= (1 - \lambda)x \cdot Q((1 - \lambda)x + \lambda y) + \lambda y \cdot Q((1 - \lambda)x + \lambda y) \\
 &= (1 - \lambda)^2 x \cdot Qx + (1 - \lambda)\lambda x \cdot Qy + (1 - \lambda)\lambda y \cdot Qx + \lambda^2 y \cdot Qy \\
 &= (1 - \lambda)^2 x \cdot Qx + \lambda^2 y \cdot Qy + (1 - \lambda)\lambda(x \cdot Qy + y \cdot Qx) \\
 &= (1 - \lambda)^2 x \cdot Qx + \lambda^2 y \cdot Qy + (1 - \lambda)\lambda(x \cdot Qx + y \cdot Qy - (x - y) \cdot Q(x - y)) \\
 &= (1 - \lambda)x \cdot Qx + \lambda y \cdot Qy - (1 - \lambda)\lambda(x - y) \cdot Q(x - y) \\
 &\leq (1 - \lambda)x \cdot Qx + \lambda y \cdot Qy
 \end{aligned}$$

when Q is positive semidefinite. When Q is positive definite the above \leq becomes $<$ showing strict convexity then holds. Since all the above statements are equality, and the last inequality holds $\iff Q$ is positive semidefinite/definite we've also shown Q is positive semidefinite/definite when $x \mapsto x \cdot Qx$ is convex/strictly convex.

Exercise 7.3

- (a) For the positive scalar property simply multiply the (strict) convexity inequality by α to get the desired property. For $f + g$ we have, for $\gamma := (1 - \lambda)$

$$(f + g)(\gamma x + \lambda y) = f(\gamma x + \lambda y) + g(\gamma x + \lambda y) \leq \gamma f(x) + \lambda f(y) + \gamma g(x) + \lambda g(y) = \gamma(f + g)(x) + \lambda(f + g)(y)$$

- (b) Again putting $\gamma = (1 - \lambda)$ we have

$$g(f(\gamma x + \lambda y)) \leq g(\gamma f(x) + \lambda f(y)) \leq \gamma g(f(x)) + \lambda g(f(y))$$

where the first inequality comes from f 's convexity and g 's nondecreasing-ness, and the second from g 's convexity. The strict convexity case with g increasing is identical to the above except \leq is replaced with $<$, giving us our desired property.

(c) For $\gamma = (1 - \lambda)$

$$f(A(\gamma x + \lambda y) + b) = f(\gamma Ax + \gamma b + \lambda Ay + \lambda b) \leq \gamma f(Ax + b) + \lambda f(Ay + b)$$

Exercise 7.7 By Prop. 7.2 $d_C(\cdot)$ is convex, and combining that with Exercise 7.3 (b), and the fact that x^2 is convex, we know $d_C(\cdot)^2$ is convex.

In order to avoid 80000 parens, I'll be using $\langle a, b \rangle := a \cdot b$. We need to show, where e_i is the standard basis on \mathbb{R}^n ,

$$\frac{\partial (d_C(x)^2)}{\partial x_i} = \lim_{t \rightarrow 0} \frac{d_C(x + te_i)^2 - d_C(x)^2}{t} = 2\langle e_i, x - P_C(x) \rangle$$

because, then since each of these partials is continuous, we would show the desired differentiability with the specified gradient. Using the hint, we know

$$d_D(x + te_i)^2 \leq d_C(x + te_i)^2 \leq d_{\{P_C(x)\}}(x + te_i)^2$$

With these inequalities I aim to use the squeeze theorem, i.e. I want to show

$$\lim_{t \rightarrow 0} \frac{d_D(x + te_i)^2 - d_C(x)^2}{t} = \lim_{t \rightarrow 0} \frac{d_{\{P_C(x)\}}(x + te_i)^2 - d_C(x)^2}{t} = 2\langle x - P_C(x), e_i \rangle$$

the second equality holds since

$$\begin{aligned} d_{\{P_C(x)\}}(x + te_i)^2 - d_C(x)^2 &= \langle x + te_i - P_C(x), x + te_i - P_C(x) \rangle - \langle x - P_C(x), x - P_C(x) \rangle \\ &= 2t\langle e_i, x - P_C(x) \rangle \end{aligned}$$

Next we have

$$\begin{aligned} d_D(x + te_i)^2 - d_C(x)^2 &= \langle x + te_i - P_D(x + te_i), x + te_i - P_D(x + te_i) \rangle - \langle x - P_C(x), x - P_C(x) \rangle \\ &= t^2 + 2t\langle e_i, x - P_D(x + te_i) \rangle + \langle P_D(x + te_i), P_D(x + te_i) \rangle - 2\langle x, P_D(x + te_i) - P_C(x) \rangle - \langle P_C(x), P_C(x) \rangle \end{aligned}$$

Pausing on this calculation, notice by construction of D

$$\begin{aligned} &\langle P_D(x + te_i) - P_C(x), x - P_C(x) \rangle \leq 0 \\ \implies &-2\langle x, P_D(x + te_i) - P_C(x) \rangle \geq -2\langle P_D(x + te_i) - P_C(x), P_C(x) \rangle \\ \implies &-2\langle x, P_D(x + te_i) - P_C(x) \rangle - \langle P_C(x), P_C(x) \rangle \geq -2\langle P_D(x + te_i), P_C(x) \rangle + \langle P_C(x), P_C(x) \rangle \\ \implies &\langle P_D(x + te_i), P_D(x + te_i) \rangle - 2\langle x, P_D(x + te_i) - P_C(x) \rangle - \langle P_C(x), P_C(x) \rangle \geq \|P_D(x + te_i) - P_C(x)\|^2 \end{aligned}$$

Combining the computation of the right limit above, along with the above bounds we can see

$$\begin{aligned} 2\langle e_i, x - P_C(x) \rangle &\geq \lim_{t \rightarrow 0} \frac{d_D(x + te_i)^2 - d_C(x)^2}{t} \\ &\geq \lim_{t \rightarrow 0} \frac{t^2 + 2t\langle e_i, x - P_D(x + te_i) \rangle + \|P_D(x + te_i) - P_C(x)\|^2}{t} \\ &= \lim_{t \rightarrow 0} 2\langle e_i, x - P_D(x + te_i) \rangle + \frac{\|P_D(x + te_i) - P_C(x)\|^2}{t} \end{aligned}$$

By construction of D $P_D(x) = P_C(x)$ so we know, using Prop 4.15

$$\lim_{t \rightarrow 0} \frac{\|P_D(x + te_i) - P_C(x)\|^2}{t} = \lim_{t \rightarrow 0} \frac{\|P_D(x + te_i) - P_D(x)\|^2}{t} \leq \lim_{t \rightarrow 0} \frac{\|x + te_i - x\|^2}{t} = 0$$

Again leveraging prop 4.15 we know $P_D(\cdot)$ is Lipschitz, hence continuous, so that

$$\lim_{t \rightarrow 0} 2\langle e_i, x - P_D(x + te_i) \rangle = 2\langle e_i, x - P_D(x) \rangle = 2\langle e_i, x - P_C(x) \rangle$$

Altogether that gives us

$$\lim_{t \rightarrow 0} \frac{d_D(x + te_i)^2 - d_C(x)^2}{t} = 2\langle e_i, x - P_C(x) \rangle \implies \nabla d_C(x)^2 = 2(x - P_C(x))$$

A final note here: I do think I could avoid the partials and work with differentiability directly... the proof would probably be mostly the same (just replacing t with $\|t\|$ in various places), but I didn't feel like rewriting what I already wrote.

Another aside: I didn't really understand the dot product construction of half planes until this problem, so while it took me a while to figure out, I've come out a little bit better at geometry. I really do struggle with geometry...

Exercise 7.9

- By definition

$$T_C((0,1)) = \left\{ v \in \mathbb{R}^2 \mid \exists (x_i, y_i) \in C \rightarrow (0,1), \lambda_i \downarrow 0 \text{ so that } \frac{(x_i, y_i) - (0,1)}{\lambda_i} \rightarrow v \right\}$$

To begin let $v \in \mathbb{R} \times (-\infty, 0]$. Put

$$x_i = \frac{-\operatorname{sgn} v_x}{i v_y}, \quad y_i = 1 - \frac{1}{|v_x| i}, \quad \lambda_i = \frac{-1}{i |v_x| v_y}$$

then it's clear $x_i \rightarrow 0$, $y_i \rightarrow 1$, $\lambda_i \downarrow 0$ and

$$\frac{x_i}{\lambda_i} = v_x, \quad \frac{y_i - 1}{\lambda_i} = v_y$$

All that's left to show is whether $(x_i, y_i) \in C$ for large enough i , indeed for

$$i \geq \max \left\{ \frac{1}{|v_y|}, \frac{1}{|v_x|}, \frac{|v_x|}{2} \left(\frac{1}{v_x^2} + \frac{1}{v_y^2} \right) \right\}$$

we know

$$i \geq \frac{1}{|v_y|} \implies 1 \geq \frac{1}{i |v_y|} \implies x_i \in [-1, 1] \text{ and } i \geq \frac{1}{|v_x|} \implies y_i = 1 - \frac{1}{|v_x| i} \geq 0$$

and

$$\begin{aligned} i &\geq \frac{|v_x|}{2} \left(\frac{1}{v_x^2} + \frac{1}{v_y^2} \right) \\ \implies \frac{2}{|v_x| i} &\geq \frac{1}{i^2} \left(\frac{1}{v_x^2} + \frac{1}{v_y^2} \right) \\ \implies 1 - \frac{1}{v_y^2 i^2} &\geq 1 + \frac{1}{v_x^2 i^2} - \frac{2}{|v_x| i} = \left(1 - \frac{1}{|v_x| i} \right)^2 \\ \implies \sqrt{1 - x_i^2} &\geq \left| 1 - \frac{1}{|v_x| i} \right| = y_i \end{aligned}$$

where the last equality comes from $y_i \geq 0$ as shown above, so that $\mathbb{R} \times (-\infty, 0] \subset T_C((0, 1))$. Suppose the opposite inclusion doesn't hold so that $\exists v \in T_C((0, 1))$, $v \notin \mathbb{R} \times (-\infty, 0]$, then $v \in \mathbb{R} \times (0, \infty)$, and $\exists (x_i, y_i), \lambda_i$, so that in particular $y_i \in [0, 1], \lambda_i > 0$ and so

$$\frac{y_i - 1}{\lambda_i} \leq 0 \implies v_y \leq 0$$

a contradiction.

- By definition

$$N((0, 1)) = \{w \in \mathbb{R}^2 \mid w \cdot ((x, y) - (0, 1)) \leq 0 \ \forall (x, y) \in C\}$$

Begin by letting $w \in \{0\} \times [0, \infty)$ so that $w_x = 0, w_y \geq 0$. For any $(x, y) \in C$ $y \in [0, 1]$ so

$$w \cdot ((x, y) - (0, 1)) = w_y \cdot (y - 1) \leq 0 \implies w \in N((0, 1))$$

For the reverse inclusion notice $(0, 0) \in C$ so that for $w \in N((0, 1))$

$$0 \geq w \cdot ((0, 0) - (0, 1)) = -w_y \implies w_y \geq 0$$

Suppose for contradiction $w_x \neq 0$ and w.l.o.g. $w_x > 0$ (symmetry allows us this). Consider $x < \frac{w_x}{w_y}$ then

$$\frac{w_y}{w_x} < \frac{1}{x} = \frac{x}{x^2} < \frac{x}{1 - \sqrt{1 - x^2}} \implies w_y(1 - \sqrt{1 - x^2}) < w_x x \implies 0 < w \cdot ((x, \sqrt{1 - x^2}) - (0, 1)) \implies w \notin N((0, 1))$$

so that $w_x = 0$ and thus $N((0, 1)) \subset \{0\} \times [0, \infty)$

- By definition

$$T_C((1, 0)) = \left\{ v \in \mathbb{R}^2 \mid \exists (x_i, y_i) \in C \rightarrow (1, 0), \lambda_i \downarrow 0 \text{ so that } \frac{(x_i, y_i) - (1, 0)}{\lambda_i} \rightarrow v \right\}$$

Let $v \in (-\infty, 0] \times [0, \infty)$ and put

$$\lambda_i = \frac{-1}{iv_x v_y}, x_i = 1 - \frac{1}{iv_y}, y_i = \frac{-1}{iv_x}$$

then clearly $\lambda_i \downarrow 0, x_i \rightarrow 1, y_i \rightarrow 0$ and

$$\frac{x_i - 1}{\lambda_i} = v_x, \frac{y_i}{\lambda_i} = v_y$$

Similar to above large enough i give us that $(x_i, y_i) \in C$ so that $v \in T((1, 0))$.

Now for $v \in T((1, 0))$ $\exists (x_i, y_i) \in C \rightarrow (1, 0), \lambda_i \downarrow 0$ so that, since $|x| \leq 1, y \in [0, 1]$

$$\frac{x_i - 1}{\lambda_i} \leq 0 \implies v_x \leq 0, \frac{y_i}{\lambda_i} \geq 0 \implies v_y \geq 0 \implies v \in (-\infty, 0] \times [0, \infty)$$

- By definition

$$N((1, 0)) = \{w \in \mathbb{R}^2 \mid w \cdot ((x, y) - (1, 0)) \leq 0 \ \forall (x, y) \in C\}$$

Let $w \in -T_C((1, 0)) = [0, \infty) \times (-\infty, 0]$ and $(x, y) \in C$ then $x - 1 \leq 0$ and $y \geq 0$ so $w \cdot (x - 1, y) = w_x(x - 1) + w_y y \leq 0$. For the opposite inclusion let $w \in N((1, 0))$ so that for every $(x, y) \in C$

$$w \cdot (x - 1, y) \leq 0$$

then in particular $0 \geq w \cdot (0, 0) = -w_x \implies w_x \in [0, \infty)$. Assume for contradiction that $w_y > 0$ then $w \cdot (0, 1) = w_x + w_y > 0$ so that $w \notin N((1, 0))$, thus $w_y \in (-\infty, 0]$.

A final note here: this problem took longer than I expected it to, but on the plus side I have a better intuition about these cones, I think?

Exercise 7.10 (a) Non-emptiness comes from the fact that $0 \in T_C(x) \cap N_C(x)$ (take the constant sequence $x_i = x, \lambda_i = 1/i$ to see the first inclusion, and the second inclusion is trivial since $0 \leq 0$). To show $T_C(\bar{x})$ is closed, fix a sequence $v_i \in T_C(\bar{x}) \rightarrow v$. By construction for each $i \exists x_{i,j} \in C \rightarrow \bar{x}, \lambda_{i,j} \downarrow 0$ and $\frac{x_{i,j} - \bar{x}}{\lambda_{i,j}} \rightarrow v_i$. Diagonalizing gives us $v \in T_C(\bar{x})$ (am I skipping too many details here?).

Now consider $w_i \in N(\bar{x}) \rightarrow w$ then $\forall x \in C w_i \cdot (x - \bar{x}) \leq 0$, so since the dot product is continuous $w \cdot (x - \bar{x}) \leq 0$.

To demonstrate convexity I'll use the fact that $T_C(\bar{x})$ and $N_C(\bar{x})$ are cones, and a cone D is convex $\iff x + y \in D$ for every $x, y \in D$ (which I did convince myself of, as you/the book suggested). Take $x, y \in T_C(\bar{x})$ then $\exists x_i, y_i \in C \rightarrow \bar{x}$ and $\lambda_i, \gamma_i \downarrow 0$ such that

$$\frac{x_i - \bar{x}}{\lambda_i} \rightarrow x, \quad \frac{y_i - \bar{x}}{\gamma_i} \rightarrow y$$

Fix $v_i = \frac{1}{\lambda_i + \gamma_i}(\gamma_i x_i + \lambda_i y_i)$. First we must show $v_i \rightarrow \bar{x}$, so let $\epsilon > 0$ and consider i large enough so that $|x_i - \bar{x}| < \epsilon$ and $|y_i - \bar{x}| < \epsilon$ then

$$\left| \frac{1}{\lambda_i + \gamma_i}(\gamma_i x_i + \lambda_i y_i) - \bar{x} \right| \leq \frac{\gamma_i}{\lambda_i + \gamma_i} |x_i - \bar{x}| + \frac{\lambda_i}{\lambda_i + \gamma_i} |y_i - \bar{x}| < \epsilon \left(\frac{\gamma_i}{\lambda_i + \gamma_i} + \frac{\lambda_i}{\lambda_i + \gamma_i} \right) = \epsilon$$

Next fix $\phi_i = \frac{\gamma_i \lambda_i}{\lambda_i + \gamma_i}$ so that $\phi_i \downarrow 0$ and finally

$$\frac{\frac{1}{\lambda_i + \gamma_i}(\gamma_i x_i + \lambda_i y_i) - \bar{x}}{\frac{\lambda_i \gamma_i}{\lambda_i + \gamma_i}} = \frac{\lambda_i + \gamma_i}{\lambda_i \gamma_i} \frac{\gamma_i(x_i - \bar{x}) + \lambda_i(y_i - \bar{x})}{\lambda_i + \gamma_i} = \frac{x_i - \bar{x}}{\lambda_i} + \frac{y_i - \bar{x}}{\gamma_i} \rightarrow x + y$$

Lastly with $a, b \in N(\bar{x})$ it's trivial to see $a + b \in N(\bar{x})$ (skipping too much here?)

- (b) To show N_C is osc take $x_i \rightarrow x$ and $y_i \in N_C(x_i)$ such that $y_i \rightarrow y$, then since the dot product is continuous, for any $v \in C$

$$\langle y_i, v - x_i \rangle \leq 0 \implies \langle y, v - x \rangle \leq 0$$

so that $y \in N_C(x)$.

Exercise 7.13

- (a) We only need to show the backward implication. Suppose $\text{epi } f$ is closed. Let $x_i \rightarrow x$ then $\exists x_{i_k}$ so that $f(x_{i_k}) \rightarrow \liminf_i f(x_i)$. We have $(x_{i_k}, f(x_{i_k})) \in \text{epi } f$, so since $\text{epi } f$ is closed $(x, \liminf_i f(x_i)) \in \text{epi } f$ so that $\liminf_i f(x_i) \geq f(x)$, hence f is lsc.
- (b) Suppose f is convex, then let $(x, y), (t, u) \in \text{epi } f$ and $\lambda \in [0, 1]$ we know

$$f(\lambda x + (1 - \lambda)t) \leq \lambda f(x) + (1 - \lambda)f(t) \leq \lambda y + (1 - \lambda)u \implies \lambda(x, y) + (1 - \lambda)(t, u) \in \text{epi } f$$

Conversely suppose $\text{epi } f$ is convex, let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$ then, since $(x, f(x)), (y, f(y)) \in \text{epi } f$

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in \text{epi } f \implies \lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y)$$

so that f is convex.

- (c) This case is identical to above, but replacing $\text{epi } f$ with the specified set and \leq with $< / \geq$ with $>$ where relevant.

By the way, is this problem supposed to say 'strictly convex function'?

As an aside, I wasn't careful with taking values of ∞ . After briefly double checking, I don't think I do? Maybe in (c), but not the other two? Let's chat about this next time we see each other.

Exercise 7.15 Before I begin, I'd like it to be known that I am in fact partially illiterate: for a non-trivial amount of time I missed the assumption that f_α were lsc. Without this assumption I was able to come up with a counter example that g is lsc. On the upside, now I think I can show that any proper convex function whose effective domain is open is lsc everywhere. Another way put: a convex function is lsc on the interior of its effective domain. Which, this is a pretty cool fact.

Now for the actual exercise... which is much easier to show than the above fact, in my opinion:

- First we'll prove from the definition. Let $x_i \rightarrow x$ then we know for each $\alpha \in \mathcal{A}$

$$g(x_i) \geq f_\alpha(x_i) \implies \liminf_i g(x_i) \geq \liminf_i f_\alpha(x_i) \geq f_\alpha(x)$$

Taking the sup of the rhs gives us our desired inequality.

- Using exercise 7.13, we know $\text{epi } f_\alpha$ is closed. We have $\text{epi } g = \bigcap_\alpha \text{epi } f_\alpha$ and so $\text{epi } g$ is closed, thus invoking that same exercise we see g is lsc.

EDIT: I read the next 2 pages in the book again... which shows the claim I stated at the top of this problem. Cool!

Exercise 7.20

- (a) Closedness comes from the continuity of the dot product. Let $v, w \in \partial f(x)$, $\lambda \in [0, 1]$ and $x' \in \mathbb{R}^n$ then with $\gamma = 1 - \lambda$ we know $\lambda f(x') \geq \lambda f(x) + \lambda v \cdot (x' - x)$, $\gamma f(x') \geq \gamma f(x) + \gamma w \cdot (x' - x)$ and so

$$f(x') = \lambda f(x') + \gamma f(x') \geq \lambda f(x) + \lambda v \cdot (x' - x) + \gamma f(x) + \gamma w \cdot (x' - x) = f(x) + (\lambda v + \gamma w) \cdot (x' - x)$$

so that $\partial f(x)$ is convex.

- (b) Let $x \in \mathbb{R}^n$ so that $f(x)$ is finite and lsc. Let $x_i \rightarrow x$ and fix $y_i \in \partial f(x_i)$ where $y_i \rightarrow y$ and let $x' \in \mathbb{R}^n$. If $f(x') = \infty$ then trivially $f(x') \geq f(x) + y \cdot (x' - x)$. For finite $f(x')$ we know

$$y_i \cdot (x' - x_i) \leq f(x') - f(x_i) \implies y \cdot (x' - x) = \liminf_i y_i \cdot (x' - x_i) \leq \liminf_i f(x') - f(x_i) \leq f(x') - f(x)$$

so that $y \in \partial f(x)$ hence ∂f is osc at x .

- (c) I'll start by showing $\partial f(x)$ is bounded. Since f is locally Lipschitz in the interior of the domain, there's $\epsilon > 0$ so that when $y, z \in x + \epsilon \mathbb{B}$ we have $|f(y) - f(z)| \leq K \|y - z\|$ for some $K > 0$. Let $v \in \partial f(x)$ we know

$$f\left(x + \frac{\epsilon}{\|v\|}v\right) - f(x) \geq \left\langle v, x + \frac{\epsilon}{\|v\|}v - x \right\rangle = \epsilon \|v\| \implies \|v\| \leq \frac{f\left(x + \frac{\epsilon}{\|v\|}v\right) - f(x)}{\epsilon} \leq K$$

where the last inequality comes because $\left\{x + \frac{\epsilon}{\|v\|}v, x\right\} \subset x + \epsilon \mathbb{B}$, and the fact that the numerator is positive.

To show $\partial f(x)$ is non-empty, I begin by claiming for each $i \in \mathbb{N}$ $(x, f(x) - 1/i) \notin \overline{\text{epi } f}$ since f is locally Lipschitz around x . Suppose otherwise, so that there's a sequence $(x_j, y_j) \in \text{epi } f \rightarrow (x, f(x) - 1/i)$. Choose N big enough so that any $j > N$ $\|x_j - x\| < \frac{1}{2Ki}$ where K is the local Lipschitz constant, then

$$|f(x_j) - f(x)| \leq K\|x_j - x\| < \frac{1}{2i}$$

so that $f(x_j) \in f(x) + \frac{1}{2i}\mathbb{B}$ so that $f(x_j) \not\rightarrow f(x) - 1/i$, hence we find a contradiction and so we can use the logic from Theorem 4.16 to see for $w \in \text{epi } f$, $v_i = (x, f(x) - 1/i) - P_{\text{epi } f}((x, f(x) - 1/i))$ with $\hat{v}_i = \frac{v_i}{\|v_i\|}$

$$\langle \hat{v}_i, w \rangle \leq \langle \hat{v}_i, (x, f(x) - 1/i) \rangle - \|v_i\|$$

Since $\hat{v}_i \in \mathbb{B}$ there's a convergent subsequence $\hat{v}_{i_k} \rightarrow \hat{v} \in \mathbb{B}$. Using the fact that

$$\|v_i\| = \|(x, f(x) - 1/i) - P_{\text{epi } f}((x, f(x) - 1/i))\| \rightarrow 0$$

we can see

$$\langle \hat{v}, w \rangle \leq \langle \hat{v}, (x, f(x)) \rangle \iff \langle (\hat{v}_1, \hat{v}_2), (w_1 - x, w_2 - f(x)) \rangle \leq 0$$

where $(\hat{v}_1, \hat{v}_2) \in \mathbb{R}^n \times \mathbb{R}$, $(w_1, w_2) \in \text{epi } f$. I claim $\hat{v}_2 \neq 0$; suppose for contradiction this isn't the case. By the construction $\|\hat{v}\| = 1$ so then $\hat{v}_1 \neq 0$. Since $x \in \text{int dom } f$ there's $\epsilon > 0$ so that f is finite on $x + \epsilon\mathbb{B}$. Notably

$$q := x + \epsilon \frac{\hat{v}_1}{\|\hat{v}_1\|} \in x + \epsilon\mathbb{B}$$

Since $f(q)$ is finite, that means $(q, f(q)) \in \text{epi } f$ and so

$$0 \geq \langle (\hat{v}_1, \hat{v}_2), (q - x, f(q) - f(x)) \rangle = \left\langle \hat{v}_1, x + \epsilon \frac{\hat{v}_1}{\|\hat{v}_1\|} - x \right\rangle = \epsilon \|\hat{v}_1\| > 0$$

drawing a contradiction, so indeed $\hat{v}_2 \neq 0$. Additionally, since $f(x)$ is finite $(x, f(x) + 1) \in \text{epi } f$ so we know

$$0 \geq \langle (\hat{v}_1, \hat{v}_2), (x - x, f(x) + 1 - f(x)) \rangle = \hat{v}_2$$

and so $\hat{v}_2 < 0$. This means we can divide the limiting separation inequality to find

$$\frac{1}{\hat{v}_2} \langle \hat{v}_1, w_1 - x \rangle + (f(w_1) - f(x)) \geq 0 \implies f(w_1) \geq f(x) + \left\langle \frac{-\hat{v}_1}{\hat{v}_2}, w_1 - x \right\rangle$$

for any $w_1 \in \text{dom } f$. If $w_1 \notin \text{dom } f$ then $f(w_1) = \infty$ and so the above inequality also holds, so that indeed

$$\frac{-\hat{v}_1}{\hat{v}_2} \in \partial f(x)$$

- (d) If f is differentiable at \bar{x} then it's continuous and so $\bar{x} \in \text{int dom } f$, hence by above $\partial f(x)$ is non-empty. Let $v \in \partial f(\bar{x})$ and suppose $v \neq \nabla f(\bar{x})$, then in particular the directional derivative in the direction of $v - \nabla f(\bar{x})$ exists, i.e.

$$\langle \nabla f(\bar{x}), v - \nabla f(\bar{x}) \rangle = \lim_{\epsilon \rightarrow 0} \frac{f(\bar{x} + \epsilon(v - \nabla f(\bar{x}))) - f(\bar{x})}{\epsilon}$$

On the other hand, we know for small $\epsilon > 0$

$$f(\bar{x} + \epsilon(v - \nabla f(\bar{x}))) - f(\bar{x}) \geq \langle v, \bar{x} + \epsilon(v - \nabla f(\bar{x})) - \bar{x} \rangle = \epsilon \langle v, v - \nabla f(\bar{x}) \rangle$$

so that for every $\epsilon > 0$

$$\frac{f(\bar{x} + \epsilon(v - \nabla f(\bar{x}))) - f(\bar{x})}{\epsilon} \geq \langle v, v - \nabla f(\bar{x}) \rangle$$

Taking the limit as $\epsilon \rightarrow 0$ gives us from the above definition of directional derivative

$$\langle \nabla f(\bar{x}), v - \nabla f(\bar{x}) \rangle \geq \langle v, v - \nabla f(\bar{x}) \rangle \implies \langle \nabla f(\bar{x}) - v, v - \nabla f(\bar{x}) \rangle \geq 0 \implies \|v - \nabla f(\bar{x})\| \leq 0$$

so that $v = \nabla f(\bar{x})$.

Exercise 7.22

$$f(x) = \max\{x_1 + x_2, x_1 - x_2, -x_1 + x_2, -x_1 - x_2\}$$

- At $(0, 0)$ the maximum is attained by all four functions so

$$\partial f((0, 0)) = \text{con}\{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$$

which is a square with sides of length 2 centered at the origin.

- At $(5, 0)$ the maximum of 5 is attained by the first two functions so

$$\partial f((5, 0)) = \text{con}\{(1, 1), (1, -1)\}$$

which is the line segment from $(1, 1)$ to $(1, -1)$.

- At $(0, 7)$ the maximum of 7 is achieved by the first and third functions so

$$\partial f((0, 7)) = \text{con}\{(1, 1), (-1, 1)\}$$

which is the line segment from $(1, 1)$ to $(-1, 1)$.

Exercise 7.24

- (a) First order of business: I saw your message in the email: I greatly appreciate your candor. I probably have seen this proof before, but I don't remember it. I've been feeling dumb math-wise recently, but I don't believe myself to be that dumb, mathwise at least, so I felt it necessary to prove to myself I could whip up this proof. Showing x is a minimizer of f over all of C is the same as showing x is a global minimizer on all of \mathbb{R}^n of f extended to all of \mathbb{R}^n by taking ∞ outside of C . Notably this extension is still convex because C is convex. For the below I'll work with this extended f .

Since $x \in \text{int dom } f$, Exercise 7.20 (c) tells us $\exists v \in \partial f(x)$ which means for every $y \in \mathbb{R}^n$

$$\langle v, y \rangle \leq f(x + y) - f(x)$$

For each $y \exists \epsilon \in (0, 1)$ so that $x + \epsilon y \in U$. By local minimality we know

$$0 \leq \frac{1}{2\epsilon}(f(x + \epsilon y) - f(x))$$

Dividing the initial inequality by 2 and using the previous we see

$$\left\langle \frac{v}{2}, y \right\rangle \leq \frac{1}{2}(f(x + y) - f(x)) \leq \frac{1}{2\epsilon}(\epsilon f(x + y) - (\epsilon + 1)f(x) + f(x + \epsilon y))$$

Convexity of f tells us

$$f(x + \epsilon y) = f((1 - \epsilon)x + \epsilon(x + y)) \leq (1 - \epsilon)f(x) + \epsilon f(x + y)$$

Combining the previous two inequalities gives us

$$\left\langle \frac{v}{2}, y \right\rangle \leq \frac{1}{2\epsilon}(\epsilon f(x + y) - (\epsilon + 1)f(x) + (1 - \epsilon)f(x) + \epsilon f(x + y)) = \frac{1}{2\epsilon}(2\epsilon f(x + y) - 2\epsilon f(x)) = f(x + y) - f(x)$$

Hence $2^{-1}v \in \partial f(x)$. Since v was arbitrary we can find a sequence $v_n := 2^{-n}v \in \partial f(x)$ and so Exercise 7.20 (a) tells us $0 = \lim_n v_n \in \partial f(x)$ so that

$$f(x + y) \geq f(x) \quad \forall y \in \mathbb{R}^n,$$

or in other words $f(x)$ is a global minimizer on \mathbb{R}^n and thus on C .

And as has been demonstrated, I'm not the best at reading because I found your hint after coming up with the above. At least the above argument is fun, for with an appropriate definition of fun.

- (b) If there's less than 2 minimizers then the set of minimizers is trivially/vacuously convex. Otherwise let x, y minimize f and $\lambda \in [0, 1]$ then $\lambda x + (1 - \lambda)y \in C$ and

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) = \min_{x \in C} f(x)$$

so that $\lambda x + (1 - \lambda)y$ is also a minimizer.

If f is strictly convex and the set of minimizers isn't empty, then suppose there're 2 minimizers x, y . Similar to above we know

$$f(2^{-1}x + 2^{-1}y) < 2^{-1}f(x) + 2^{-1}f(y) = \min_{x \in C} f(x)$$

which is a contradiction because $f(2^{-1}x + 2^{-1}y)$ is smaller than the supposed minimizers.

Exercise 7.27 Skipped (thank you, seems complicated. Would love a quick 'this is the point' in person or over email if that's possible to do succinctly)

Exercise 7.31 Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. Construct $w_i \in U(x), v_i \in U(y)$ so that

$$p(x, w_i) \rightarrow \inf_{u \in U(x)} p(x, u), \quad p(y, v_i) \rightarrow \inf_{u \in U(y)} p(y, u)$$

Since the graph of U is convex we know for every i $\lambda(x, w_i) + (1 - \lambda)(y, v_i) \in \text{gph } U$ so that $\lambda w_i + (1 - \lambda)v_i \in U(\lambda x + (1 - \lambda)y)$. This means

$$\inf_{u \in U(\lambda x + (1 - \lambda)y)} p(\lambda x + (1 - \lambda)y, u) \leq p(\lambda x + (1 - \lambda)y, \lambda w_i + (1 - \lambda)v_i) \leq \lambda p(x, w_i) + (1 - \lambda)p(y, v_i)$$

where the last inequality used convexity of p . Finally taking $i \rightarrow \infty$ gives us the desired property.

To show $M(x)$ is convex: if $p(x, u) = \infty$ then $M(x) = U(x)$ and so convexity of $\text{gph } U$ gives us our result. Otherwise if $p(x, u) \neq \infty$ for some $u \in U(x)$, define $f : U(x) \rightarrow \mathbb{R} \cup \{\infty\}$ by $f(u) = p(x, u)$, so that f is convex (from p 's convexity). Because $U(x)$ is convex, restricting f to its effective domain keeps it convex, and its effective domain also convex. Now we can apply the first half of Exercise 7.24 (b) to see the set of minimizers of restricted f is convex. This set is the same as the set of minimizers of unrestricted f , which in turn is exactly $M(x)$.

After writing all that I see it may have been quicker just to copy and paste the proof from Exercise 7.24 (b) instead of setting up the right f ...