Solutions to Selected Exercises in Chapter Two

Exercises from Section 2.1

2.1.1. See [410, Theorem 4.43]. □

2.1.2. For $x > x_0$, $x \in I$, the three-slope inequality (2.1.1) implies

$$\frac{f(x) - f(x_0)}{x - x_0} \geq f'_+(x_0).$$

For $x < x_0$, $x \in I$, the three-slope inequality (2.1.1) implies

$$\frac{f(x) - f(x_0)}{x - x_0} \leq f'_-(x_0).$$

Hence $f(x) - f(x_0) \geq \lambda(x - x_0)$ for all $x \in I$ whenever $f'_-(x_0) \leq \lambda \leq f'_+(x_0)$. Considering $x > x_0$, the three-slope inequality (2.1.1) implies $f'_+(x_0) = \sup\{\lambda : \lambda(x - x_0) \leq f(x) - f(x_0) \text{ for all } x \in I\}$, and by the above, $f'(x_0)$ is the maximum such λ. □

2.1.3. Suppose $T : E \to F$ is a linear mapping, and let $A := y_0 + T$ where $y_0 \in F$. Then for any $\lambda \in \mathbb{R}$, we have

$$\begin{aligned}
A(\lambda x + (1 - \lambda)y) &= T(\lambda x + (1 - \lambda)y) + y_0 \\
&= \lambda(Tx) + (1 - \lambda)(Ty) + y_0 \\
&= \lambda(Tx + y_0) + (1 - \lambda)(Ty + y_0) \\
&= \lambda Ax + (1 - \lambda)Ay
\end{aligned}$$

For the converse, suppose $A : E \to F$ is an affine mapping. Let $y_0 = A(0)$, $T(x) := Ax - y_0$ and $k \in \mathbb{R}$. Then

$$\begin{aligned}
T(kx) &= A(kx) - y_0 = A(kx + (1 - k)0) - y_0 \\
&= kA(x) + (1 - k)A(0) - y_0 = kA(x) - ky_0 = kT(x).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
T(x + y) &= A(x + y) - y_0 = A\left(\frac{1}{2}(2x) + \frac{1}{2}(2y)\right) - y_0 \\
&= \frac{1}{2}A(2x) + \frac{1}{2}A(2y) - y_0 = \frac{1}{2}[A(2x) - y_0] + \frac{1}{2}[A(2y) - y_0] \\
&= \frac{1}{2}T(2x) + \frac{1}{2}T(2y) = T(x) + T(y).
\end{aligned}$$

□

2.1.4. Suppose f is positively homogeneous and subadditive, then for $x, y \in X$, $\alpha \geq 0$ and $\beta \geq 0$, we have

$$f(\alpha x + \beta y) \leq f(\alpha x) + f(\beta y) = \alpha f(x) + \beta f(y).$$

This shows f is sublinear. Conversely, suppose f is sublinear. Then clearly it is subadditive by choosing $\alpha = 1$ and $\beta = 1$ in the previous inequality. Now $f(0) \leq 0 \cdot f(x) + 0 \cdot f(-x) = 0$ and

$f(0) \leq f(0) + f(0)$ implies $f(0) \geq 0$, and so $f(0) = 0$. When $\lambda = 0$, $0 = f(0) = f(\lambda x) = \lambda f(x)$, now for $\lambda > 0$,

$$f(\lambda x) = f(\lambda x + 0) \leq \lambda f(x) + 1 \cdot f(0) = \lambda f(x) = \lambda f(\lambda^{-1} \lambda x) \leq \lambda \cdot \lambda^{-1} f(\lambda x)$$

where the second inequality follows from the first. $\square$

2.1.5. Suppose f is a convex function. Consider the set $F = \{x \in E : f(x) \leq \alpha\}$. In the case F is not empty, let $u, v \in F$. Then for $0 \leq \lambda \leq 1$,

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda) f(v) \leq \lambda \alpha + (1 - \lambda)\alpha = \alpha.$$

Thus $\lambda u + (1 - \lambda)v \in F$ whenever $0 \leq \lambda \leq 1$, and so f is quasi-convex. Analogously, suppose $u, v \in \operatorname{dom} f$. Then we can choose $\alpha \in \mathbb{R}$ so that $\max\{f(u), f(v)\} \leq \alpha$. It follows from the previous reasoning that $\lambda u + (1 - \lambda)v \in \operatorname{dom} f$ whenever $0 \leq \lambda \leq 1$ and so $\operatorname{dom} f$ is convex. $\square$

2.1.6. Observe that a pointwise suprema of convex functions is convex, because the epigraph is an intersection of convex sets which is convex. Notice that the function need not be proper. Moreover, the epigraph will be closed if all of the functions are closed.

(a) Suppose $f : X \to [-\infty, +\infty]$ is convex. If $f \equiv +\infty$, then $\operatorname{epi} f = \emptyset$ is convex. Otherwise, let $(x, t), (y, s) \in \operatorname{epi} f$. Then for $0 \leq \lambda \leq 1$ we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y) \leq \lambda t + (1 - \lambda)s.$$

Therefore, $\lambda(x, t) + (1 - \lambda)(y, s) \in \operatorname{epi} f$ as desired.

Conversely, suppose $\operatorname{epi} f$ is convex, if $\operatorname{epi} f = \emptyset$ then $f \equiv +\infty$ is convex. Otherwise, suppose $f(x), f(y) < +\infty$. Then $(x, t), (y, s) \in \operatorname{epi} f$ where $f(x) < t$ and $f(y) < s$. Thus for $0 \leq \lambda \leq 1$, we have $\lambda(x, t)) + (1 - \lambda)(y, s)) \in \operatorname{epi} f$. This implies

$$f(\lambda x + (1 - \lambda)y) \leq \lambda t + (1 - \lambda)s,$$

for all $t > f(x)$ and $s > f(y)$. It follows that f is convex. (Hence this confirms that we can define improper convex functions via their epigraphs as remarked earlier) $\square$

(c) Let $x, y \in E$, and $0 \leq \lambda \leq 1$. Then

$$m(g(\lambda x + (1 - \lambda)y) \leq m(\lambda(g(x)) + (1 - \lambda)g(y)) \leq \lambda m(g(x)) + (1 - \lambda)m(g(y)),$$

as desired.

For a nice alternative geometric approach and explanation to (d), the reader is encouraged to consult [255, Proposition 2.2.1]. For the converse, the convexity of g follows by restricting to the case $t = 1$. $\square$

2.1.7. It is clear that if $x_0 \in \operatorname{int} C$, then $x_0 \in \operatorname{core} C$. For the converse, suppose C is convex and $x_0 \in \operatorname{core} C$. Then there exist $\delta_i > 0$ so that $x_0 + te_i \in C$ for all $|t| \leq \delta_i$ and $i = 1, 2 \ldots, n$ where $\{e_i\}_{i=1}^n$ is the usual basis of $\mathbb{R}^n$. Now let $\delta = \min\{\delta_1, \delta_2, \ldots, \delta_n\}$. Because C is convex, it follows that $x_0 + h \in C$ whenever $h = a_1 e_1 + a_2 e_2 + \ldots a_n e_n$ where $|a_i| \leq \delta/n$ for $i = 1, 2, \ldots, n$. Consequently, $x_0 \in \operatorname{int} C \neq \emptyset$. A conventional example of a nonconvex set $F \subset \mathbb{R}^2$ with $(0, 0) \in \operatorname{core} F \setminus \operatorname{int} F$ is $F = \{(x, y) \in \mathbb{R}^2 : |y| \geq x^2 \text{ or } y = 0\}$; see also Figure 2.4 for another example. $\square$

2.1.8. Let $x \in E$ be a point of continuity of the convex function f. The max formula (2.1.19) ensures that $\partial f(x) \neq \emptyset$. Moreover, f has Lipschitz constant K on some neighborhood U of x (Theorem 2.1.10). Because $\langle v, y-x\rangle \leq f(y)-f(x)$ for all $y \in U$ and $v \in \partial f(x)$, it follows that $\|v\| \leq K$. Thus, $\partial f(x) \subset KB_E$. Now suppose $(v_n) \subset \partial f(x)$ and $v_n \to v$. Then

$$\langle v, y-x\rangle = \lim_{n\to\infty} \langle v_n, y-x\rangle \leq f(y)-f(x), \quad \text{for all } y \in E.$$

Therefore, $\partial f(x)$ is closed. Finally, suppose $u, v \in \partial f(x)$ and $0 \leq \lambda \leq 1$. Writing $w = \lambda u + (1-\lambda)v$, for each $y \in E$ we have

$$\begin{aligned} \langle w, y-x\rangle &= \lambda\langle u, y-x\rangle + (1-\lambda)\langle v, y-x\rangle \\ &\leq \lambda[f(y)-f(x)] + (1-\lambda)[f(y)-f(x)] = f(y)-f(x). \end{aligned}$$

This shows $w \in \partial f(x)$. Therefore, $\partial f(x)$ is a nonempty, closed, convex and bounded subset of E. □

2.1.9. Let $x \in \operatorname{dom} f$ and $d \in E$. If $x+d \notin \operatorname{dom} f$, $f(x+d)-f(x) = \infty$ so the inequality is clear. In the case $x+d \in \operatorname{dom} f$, the three slope inequality implies for $0 < t < 1$,

$$f'(x;d) \leq \frac{f(x+td)-f(x)}{t} \leq \frac{f(x+d)-f(x)}{1}.$$

as desired. Thus $f(x+d) \geq f(x)$ for all $d \in E$ if and only if $f'(x;d) \geq 0$ for all $d \in E$. Also, $0 \in \partial f(x)$ if and only if $\langle 0, d\rangle \leq f(x+d)-f(x)$ for all $d \in E$ if and only if $f(x+d) \geq f(x)$ for all $d \in E$. □

2.1.10. The measurability of $\phi \circ f$ follows because ϕ is continuous (see [384]). Let $a := \int f d\mu$ and note the bounds on f ensure $a \in I$. Apply Corollary 2.1.3 to obtain $\lambda \in \mathbb{R}$ such that $\phi(t) \geq \phi(a) + \lambda(t-a)$ for all $t \in I$. Then $\phi(f(t)) - \lambda(f(t)-a) - \phi(a) \geq 0$. Integrating we obtain

$$\int_\Omega [\phi(f(t)) - \lambda(f(t)-a) - \phi(a)]d\mu \geq 0.$$

Because $\mu(\Omega) = 1$ and because of the choice of a, this implies

$$\int_\Omega \phi(f(t))d\mu \geq \lambda(a-a) + \phi\left(\int_\Omega f d\mu\right) = \phi\left(\int_\Omega f d\mu\right),$$

as desired. □

2.1.11. (a) For arbitrary $a, b \in \mathbb{R}$ and $0 < \lambda < 1$, let f be defined by $f(x) := a$ for $0 \leq x \leq \lambda$ and $f(x) := b$ for $\lambda < x \leq 1$. Then

$$g\left(\int_0^1 f(x)\,dx\right) = g(\lambda a + (1-\lambda)b)$$

while

$$\int_0^1 g(f)\,dx = \lambda g(a) + (1-\lambda)g(b).$$

The original assumption

$$g\left(\int_0^1 f(x)dx\right) \leq \int_0^1 g(f)dx$$

then implies $g(\lambda a + (1-\lambda)b) \le \lambda g(a) + (1-\lambda)g(b)$. This establishes the convexity of g.
(b) Applying Jensen's inequality when $\phi := \exp(\cdot)$ we have

$$\exp\left[\int_\Omega f\,d\mu\right] \le \int_\Omega \exp(f)\,d\mu.$$

When Ω is the finite set $\{s_1, s_2, \ldots, s_n\}$ with $\mu(\{s_i\}) = 1/n$ and $f(s_i) = y_i$ for $i = 1, 2, \ldots, n$, this becomes

$$\exp\left[\frac{1}{n}(y_1 + \ldots + y_n)\right] \le \frac{1}{n}(e^{y_1} + \ldots + e^{y_n}), \quad y_i \in \mathbb{R}.$$

When $x_i = e^{y_i}$ this becomes

$$(x_1 x_2 \cdots x_n)^{1/n} \le \frac{1}{n}(x_1 + x_2 + \ldots + x_n),$$

that is, the arithmetic-geometric mean inequality. $\square$

2.1.12. Suppose f is not identically $-\infty$. Then fix x_0 where f is real-valued. We may assume $f(x_0) = 0$. Suppose $f(x_0 + h) > 0$ for some $h \in E$. Then for $t > 1$,

$$0 < f(x_0 + h) \le (1 - t^{-1})f(x_0) + t^{-1}f(x_0 + th) = t^{-1}f(x_0 + th).$$

Thus $\lim_{t\to\infty} f(x_0 + th) = \lim_{t\to\infty} tf(x_0 + h) = \infty$ which is a contradiction. Finally, in the case $f(x_0 + h) < 0$ for some $h \in E$, we would deduce $f(x_0 - h) > 0$, and so this, too, is impossible. $\square$

2.1.13. (a) For $k \ge 0$, observe that $x \in \lambda C$ if and only if $kx \in k\lambda C$. Therefore γ_C is positively homogeneous (the definition of $\gamma_C := \inf\{\lambda \ge 0 : x \in \lambda C\}$ ensures $\gamma_C(0) = 0$ even when $0 \notin C$). Suppose $C \subset E$ is convex. Let $x, y \in E$. In the case $\gamma_C(x) = \infty$ or $\gamma_C(y) = \infty$ then it is clear

$$\gamma_C(\lambda x + (1-\lambda)y) \le \lambda\gamma_C(x) + (1-\lambda)\gamma_C(y) \quad \text{when } 0 < \lambda < 1.$$

Suppose $\gamma_C(x), \gamma_C(y)$ are real-valued and $\epsilon > 0$. Choose α, β so that $\gamma_C(x) \le \alpha < \gamma_C(x) + \epsilon$, $\gamma_C(y) \le \beta < \gamma_C(y) + \epsilon$ and $\frac{1}{\alpha}x, \frac{1}{\beta}y \in C$. So we choose $u, v \in C$ so that $x = \alpha u$ and $y = \alpha v$ (the case $x = 0$ or $y = 0$ are fine). Then

$$\frac{\lambda\alpha u + (1-\lambda)\beta v}{\lambda\alpha + (1-\lambda)\beta} \in C$$

and therefore

$$\gamma_C(\lambda\alpha u + (1-\lambda)\beta v) \le \lambda\alpha + (1-\lambda)\beta$$

or in other words, $\gamma_C(\lambda x + (1-\lambda)y) \le \lambda\gamma_C(x) + (1-\lambda)\gamma_C(y) + \epsilon$. This shows γ_C is convex when C is convex, in fact, γ_C is subadditive because

$$\gamma_C(x + y) = \gamma_C\left(\frac{1}{2}(2x) + \frac{1}{2}(2y)\right) \le \frac{1}{2}\gamma_C(2x) + \frac{1}{2}\gamma_C(2y) = \gamma_C(x) + \gamma_C(y).$$

Hence γ_C is sublinear when C is convex.
(b) Suppose $0 \in \operatorname{core} C$. Given $x \in E$, there exists $t > 0$ so that $tx \in C$ and then $x \in \frac{1}{t}C$. Thus $\gamma_C(x) \le 1/t$. Because γ_C is convex and everywhere finite on E, it is continuous everywhere by Theorem 2.1.12.

(c) Suppose $0 \in \operatorname{core} C$. Then γ_C is continuous, and therefore $\{x \in E : \gamma_C(x) \leq 1\}$ is closed. Observe that $\gamma_C(x) \leq 1$ for all $x \in C$, consequently $\operatorname{cl} C \subset \{x \in E : \gamma_C(x) \leq 1\}$.

2.1.14. For example, let $f(x, y) = -\sqrt[4]{xy}$ when $x \geq 0$, and $y \geq 0$. The strict convexity assertion is probably best shown by computing the Hessian (as introduced in Section 2.2) of f at (x, y), which is

$$H = \begin{bmatrix} \frac{3}{16}x^{-7/4}y^{1/4} & -\frac{1}{16}x^{-3/4}y^{-3/4} \\ -\frac{1}{16}x^{-3/4}y^{-3/4} & \frac{3}{16}x^{1/4}y^{-7/4} \end{bmatrix}.$$

Then f is strictly convex on the interior of its domain because this matrix is positive definite for all (x, y) with $x > 0$ and $y > 0$ which follows because $h_{11} > 0$, and $|H| > 0$ at all such (x, y). $\square$

2.1.15. Let (f_i) be a family of proper functions and define $f : E \to [-\infty, +\infty]$ by

$$f(x) := \inf\left\{\sum \lambda_i f_i(x_i) : \sum \lambda_i = 1, \lambda_i \geq 0, \lambda_i \text{ finitely nonzero}, \ \sum \lambda_i x_i = x\right\}.$$

Let $u, v \in \operatorname{dom} f$ and and let α and β be any real numbers satisfying $f(u) < \alpha$ and $f(v) < \beta$. Now choose sums as in the definition above so

$$\sum \alpha_i f(u_i) < \alpha, \quad \sum \beta_i f(v_i) < \beta \text{ where } \ \sum \alpha_i u_i = u, \sum \beta_i v_i = v.$$

Then for any $0 \leq \lambda \leq 1$, we have $\sum[\lambda\alpha_i + (1-\lambda)\beta_i] = 1$ and $\sum(\lambda\alpha_i u_i + (1-\lambda)\beta_i v_i) = \lambda u + (1-\lambda)v$. Then by the definition of f,

$$f(\lambda u + (1-\lambda)v) \leq \sum \lambda\alpha_i f_i(u_i) + (1-\lambda)\beta_i f_i(v_i) < \lambda\alpha + (1-\lambda)\beta.$$

The convexity of f follows from this. Next we show that f is the largest convex function minorizing the family. Indeed suppose h is convex and h minorizes the family (f_i). Then for any x such that $x = \sum \lambda_i x_i$ where $\sum \lambda_i = 1$, $\lambda_i \geq 0$, and only finitely many of the λ_i are nonzero, by the convexity of h and minorization property we have

$$h(\sum \lambda_i x_i) \leq \sum \lambda_i h(x_i) \leq \sum \lambda_i f(x_i).$$

Taking the infimum over all such sums we see that $h \leq f$.

For the example, let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(t) := t^2$ if $t \neq 0$ and $f(0) = 1$. Notice that $\operatorname{epi}\operatorname{conv} f = \{(x, y) \in \mathbb{R}^2 : y \leq x^2\}$ and $\operatorname{conv}\operatorname{epi} f = \operatorname{epi}\operatorname{conv} f \setminus \{(0, 0)\}$. $\square$

2.1.16. Suppose $T : X \to Y$ is an open mapping. Then $T(U)$ is open where $U = \operatorname{int} B_X$. Then, $0 \in \operatorname{int} T(U)$, so we choose $r > 0$ so that $rB_Y \subset T(U)$. For any $y \in Y$, choose $n > 0$ so that $n^{-1}\|y\| < r$. Then $n^{-1}y \in T(U)$, and so we let $x \in X$ be chosen so $Tx = n^{-1}y$. Then $T(nx) = y$ and so T is onto as desired.

Conversely, suppose T is onto. Let $y \in T(U)$ where U is an open subset of X, and choose $x \in U$ so that $Tx = y$. Let V be an open convex set so that $x \in V \subset U$. Now fix $h \in Y$. Because T is onto, we fix $v \in X$ so that $Tv = h$. Because V is open, we choose $\delta > 0$ so that $x + tv \in V$ for all $0 \leq t \leq \delta$. Then $T(x + tv) = y + th \in T(V)$ for all $0 \leq t \leq \delta$. Thus $y \in \operatorname{core} T(V)$, and so $y \in \operatorname{int} T(V) \subset T(U)$. Thus $T(U)$ is open.

2.1.18. Let x be such that $f(x)$ is real-valued. Suppose $f(x) = (\operatorname{cl} f)(x)$ and let $x_n \to x$. Because the epigraph of $\operatorname{cl} f$ is closed and $f \geq \operatorname{cl} f$ we know

$$\liminf_{n\to\infty} f(x_n) \geq \liminf_{n\to\infty} (\operatorname{cl} f)(x_n) \geq (\operatorname{cl} f)(x) = f(x)$$

and so f is lower semicontinuous at x. Conversely, suppose f is lower semicontinuous at x. Let $(x_n, t_n) \in \operatorname{epi} f$ be such that $(x_n, t_n) \to (x, (\operatorname{cl} f)(x))$ (possibly in the extended sense in the second coordinate). Then

$$(\operatorname{cl} f)(x) = \lim_{n\to\infty} t_n \geq \liminf_{n\to\infty} f(x_n) \geq f(x).$$

Because $(\operatorname{cl} f) \leq f$, we conclude $(\operatorname{cl} f)(x) = f(x)$. According to the lower semicontinuity of f at x we know $f(u) > \alpha > -\infty$ on some neighborhood of x. It follows easily that f is proper, see the proof of Lemma 2.3.3. □

2.1.19. Suppose $\liminf_{\|x\|\to\infty} f(x)/\|x\| > \beta > 0$. Let $S = \{x : f(x) \leq \alpha\}$ for some $\alpha \in \mathbb{R}$. Find $k > 0$ so that $k\beta > \alpha$, and $f(x)/\|x\| > \beta$ for all $\|x\| \geq k$. If $\|x\| \geq k$, then $f(x) > k\beta > \alpha$. Therefore, $S \subset kB_E$.

Clearly $f : \mathbb{R} \to \mathbb{R}$ defined by $f(t) := \sqrt{|t|}$ is not convex, and $\{t : f(t) \leq \alpha\} = [-\alpha^2, \alpha^2]$ for each $\alpha > 0$. However,

$$\lim_{|t|\to\infty} f(t)/|t| = \lim_{|t|\to\infty} |t|^{-1/2} = 0.$$

□

2.1.22. (a) For each $n \in \mathbb{N}$, let $F_n := \{x \in S : |f_k(x)| \leq n\}$. Then $\bigcup_n F_n = S$ because $f_n(x) \to f(x)$ and $f(x) \in \mathbb{R}$. According to the Baire category theorem, there exist $N \in \mathbb{N}$ such that F_N contains a relatively open (in E) subset U. Then $|f_k(x)| \leq N$ for all $k \in \mathbb{N}$, $x \in U$.

(b) By shifting, we may assume $x = 0$ where x is some given point in $\operatorname{int} S$. Now $rB_E \subset \operatorname{int} S$ for some $r > 0$. By part (a), there is some $B_{r_1}(y) \subset rB_E$ for which $f_k(x) \leq M$ for all $k \in \mathbb{N}$, $x \in B_{r_1}(y)$. Replacing M with a larger number as necessary, we may also assume $f_k(-y) \leq M$ for all $k \in \mathbb{N}$. By the convexity of f_n, we have that $f_n \leq M$ on $\operatorname{conv}(\{-y\} \cup B_{r_1}(y))$ which contains $\frac{r_1}{2}B_E$. Thus f_n is uniformly bounded on $\frac{r_1}{2}B_E$.

Now let K be a compact subset of $\operatorname{int} S$. Suppose by way of contradiction there exists $\epsilon > 0$ and a subsequence $(x_{n_k}) \subset K$ such that

$$(1) \qquad |f_{n_k}(x_{n_k}) - f(x_{n_k})| > \epsilon \;\text{ for all }\; n_k.$$

By passing to a further subsequence as necessary, we may assume $x_{n_k} \to \bar{x}$ where $\bar{x} \in K$. By the previous paragraph, we find $r > 0$, so that f_{n_k} is uniformly bounded on $B_r(\bar{x})$. The proof of Theorem 2.1.10 shows that (f_{n_k}) is equi-Lipschitz on $B_{\frac{r}{2}}(\bar{x})$. Hence (f_{n_k}) converges uniformly to f on $B_{\frac{r}{2}}(\bar{x})$. This is a contraction with (1) because (x_{n_k}) is eventually in $B_{\frac{r}{2}}(\bar{x})$. □

2.1.23. (a) Suppose f does not have Lipschitz constant $K \geq 0$ on U. Fix $u, v \in U$ such that $f(v) - f(u) > K\|v - u\|$, and let $\phi \in \partial f(v)$. Then

$$\langle \phi, u - v\rangle \leq f(u) - f(v) < -K\|u - v\|$$

and so $\|\phi\| > K$.

Conversely, suppose f has Lipschitz constant $K \geq 0$ on U. Let $u \in U$. Then $\partial f(u)$ is not empty because f is continuous. Moreover, let $\phi \in \partial f(u)$. Then

$$\langle \phi, v - u\rangle \leq f(v) - f(u) \leq K\|v - u\|$$

for all $v \in U$. Because u is in the interior of U, it follows that $\|\phi\| \leq K$.

The "in particular"' part, follows from the first part because f is continuous on U, and therefore, locally Lipschitz on U.

(b) Because $\operatorname{dom} f = E$, f is continuous on E, and the extreme value theorem implies f is bounded on bounded subsets of E. Exercise 2.1.22 then ensures f is Lipschitz on bounded subsets of E, and then part (a) of this exercise ensures ∂f maps bounded subsets of E to bounded subsets of E. $\square$

Exercises from Section 2.2

2.2.1. A proof of the convexity of f using the Hessian is sketched in [369, pp. 27–28]. Because f is convex, we have

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y)$$

which means

$$-\sqrt[n]{\left(\frac{x_1+y_1}{2}\right)\cdots\left(\frac{x_n+y_n}{2}\right)} \leq -\frac{1}{2}\sqrt[n]{x_1x_2\cdots x_n} - \frac{1}{2}\sqrt[n]{y_1y_2\cdots y_n}.$$

Multiplying both sides of the previous inequality by -2 yields the result. $\square$

2.2.2. (a) $\Rightarrow$ (b): Let $g := -1/f$ and $\phi(t) := -\ln(-t)$ for $t \in (-\infty, 0)$. Then ϕ is convex and increasing, and g is convex. Therefore, $\ln \circ f = \phi \circ g$ is convex.
(b) $\Rightarrow$ (c): $g := \ln \circ f$ is convex, therefore, $f = \exp(\ln \circ f)$ is convex since exp is convex and increasing. $\square$

2.2.4. We provide details as in [34, Lemma 3.2]. For a function f on I consider the associated Bregman distance D_f defined by $D_f(x, y) := f(x) - f(y) - f'(y)(x - y)$. Let $g := -1/h$ so that $g' = h'/h^2$. (a): $1/h$ is concave if and only if g is convex if and only if D_g is nonnegative if and only if

$$0 \leq -1/h(x) + 1/h(y) - (h'(y)/h^2(y))(x - y) \text{ for all } x, y \in I$$

if and only if

$$0 \leq h(x)h(y) - h^2(y) - h(x)h'(y)(x - y) \text{ for all } x, y \in I.$$

Part (b) is similar, noting that $D_f \equiv 0$ if and only if f is affine. Part (c) was shown in Exercise 2.2.2. Part (d): $1/h$ is concave if and only if g is convex if and only if $g'' = (h^2h'' - 2h(h')^2)/h^4 \geq 0$ if and only if $hh'' \geq 2(h')^2$. $\square$

2.2.5. First, g is a real-valued convex function on $[0, 1]$ and so Theorem 2.1.2(d) ensures that g is differentiable except at possibly countably many $t \in [0, 1]$. Then Theorem 2.2.1 implies that at points of differentiability $\nabla g(t) = \{\partial g(t)\}$. Now let $t \in (0, 1)$ be a point of differentiability of of g. Observe that

$$\langle \phi_t, sh\rangle \leq f(x + (s + t)h) - f(x + th) = g(t + s) - g(t),$$

Hence $\langle \phi_t, h\rangle \in \partial g(t)$ and we conclude $\nabla g(t) = \langle \phi_t, h\rangle$. $\square$

2.2.8. Suppose f is Fréchet differentiable at x_0. Let $\phi = f'(x_0)$. Given $\epsilon > 0$ we choose $\delta > 0$ so that $\delta\|\phi\| < \epsilon$ and

$$|f(x_0 + h) - f(x_0) - \phi(h)| \leq \frac{\epsilon}{2}\|h\|$$

whenever $0 < \|h\| < \delta$. Now suppose $\|x - x_0\| < \delta$. Then $x = x_0 + h$ where $\|h\| < \delta$ and the previous inequality then implies

$$|f(x) - f(x_0)| \leq |f(x_0 + h) - f(x_0)| \leq (\|\phi\| + \epsilon/2)\|h\| < \epsilon.$$

Thus, f is continuous at x_0. $\square$

2.2.9. Suppose f has Lipschitz constant K in a neighborhood U of x_0 and Gâteaux differentiable at x_0 with Gâteaux derivative ϕ. Then $\|\phi\| \leq K$. Suppose f is not Fréchet differentiable at x_0. Then there exists $\epsilon > 0$ and $t_n \to 0^+$, $h_n \in S_E$ such that

$$|f(x_0 + t_n h_n) - f(x_0) - \phi(t_n h_n)| \geq t_n \epsilon$$

Because S_X is compact, we may replace (h_n) above with one of its convergent subsequences, so we suppose $h_n \to h \in S_X$. When $\|h_n - h\| < \epsilon/3K$. For n sufficiently large we have

$$\begin{aligned} |f(x_0 + t_n h) - f(x_0) - \phi(t_n h)| &\geq |f(x_0 + t_n h_n) - f(x_0) - \phi(t_n h_n)| - 2K t_n \|h_n - h\| \\ &\geq t_n \epsilon - 2K(\epsilon/3K) \geq t_n \epsilon/3 \end{aligned}$$

which contradicts the Gâteaux differentiability of f. For a slightly different proof of this, see the last part of the proof of Theorem 2.5.4. $\square$

2.2.13. Suppose f is convex on the inverval I, and suppose $J := [a, b]$ is a compact subinterval of I. Then for m affine and $0 \leq \lambda \leq 1$, we have

$$\begin{aligned} (f + m)(\lambda a + (1 - \lambda) b) &\leq \lambda (f + m)(a) + (1 - \lambda)(f + m)(b) \\ &\leq \max\{(f + m)(a), (f + m)(b)\}. \end{aligned}$$

Thus the supremum of $f + m$ is attained at one of the endpoints a or b.
Conversely, suppose $a, b \in I$. Now choose an affine function m such that $(f+m)(a) = (f+m)(b)$. Because $(f + m)$ attains its max on $[a, b]$ at an endpoint, we know it attains its max on $[a, b]$ at both a and b. Then, for $0 \leq \lambda \leq 1$, we have

$$\begin{aligned} (f + m)(\lambda a + (1 - \lambda) b) &= f(\lambda a + (1 - \lambda) b) + \lambda m(a) + (1 - \lambda) m(b) \\ &\leq \max_{[a,b]}(f + m) = \lambda(f(a) + m(a)) + (1 - \lambda)(f(b) + m(b)). \end{aligned}$$

Consequent, $f(\lambda a + (1 - \lambda) b) \leq \lambda f(a) + (1 - \lambda) f(b)$ and so f is convex as desired. $\square$

2.2.16. Suppose $a < b$ and M is an affine function through $(a, f(a))$ and $(b, f(b))$, and let m be an affine minorant of f passing through $((a + b)/2, f((a + b)/2)$ (using the max formula (2.1.19)). Then $m \leq f \leq M$, and thus

$$f\left(\frac{a + b}{2}\right) \leq \frac{1}{b - a} \int_a^b f(t)\, dt \leq \frac{f(a) + f(b)}{2}$$

since these quantities are the averages of m, f and M respectively on $[a, b]$. $\square$

2.2.20. First, for $(x, y) \in \operatorname{dom} f \setminus \{(0, 0)\}$, the Hessian of f at (x, y) is

$$H = \begin{bmatrix} 6xy^{-2} & -6x^2 y^{-3} \\ -6x^2 y^{-3} & 6x^3 y^{-4} \end{bmatrix}.$$

Then H is positive semidefinite for such (x, y) because $|H| = 0$ and $6xy^{-2} \geq 0$. Also, for $(x, y) \in \operatorname{dom} f$, we have

$$\lambda f(x, y) + (1 - \lambda) f(0, 0) = \lambda \frac{x^3}{y^2} = f(\lambda(x, y))$$

and together we deduce f is convex. It is also closed because: (i) its domain is closed (ii) x^3/y^2 is continuous when $y \neq 0$, (iii) $\liminf_{(x,y)\to(0,0)} f(x, y) \geq f(0, 0)$. However, f is not continuous at $(0, 0)$ even when considering that the underlying topological space is the domain. This is because $\lim_{(x,x^2)\to(0,0)} f(x, x^2) = +\infty$.

A further observation is that this type of example cannot occur on $\mathbb{R}$. Indeed, if $f : [a, b] \to \mathbb{R}$ is convex and lower semicontinuous then f is continuous as a function on $[a, b]$. Indeed, it is continuous on (a, b) and further $\liminf_{x\to b^-} f(x) \geq f(b)$ by lower semicontinuity while the convexity of f implies $\limsup_{x\to b^-} f(x) \leq f(b)$. Similarly, f is continuous from the right at a. $\square$

2.2.21. Suppose $x, y \in U$, and $x^* \in \partial f(x)$, $y^* \in \partial f(y)$ (which are not empty by the max formula (2.1.19)). Then by the subdifferential inequality

$$\begin{aligned} \langle y^* - x^*, y - x\rangle &= y^*(y - x) + x^*(x - y) \\ &\geq f(y) - f(x) + f(x) - f(y) = 0. \end{aligned}$$

Hence the subdifferential is a monotone mapping. The 'in particular' statement follows because $\partial f(x) = \{\nabla f(x)\}$ when f is differentiable at x. $\square$

2.2.22. (a) Suppose not, then there exists $x_n \to x_0$ and $\epsilon > 0$ so that $\phi_n \in \partial f(x_n)$, but $\phi_n \notin \partial f(x_0) + \epsilon B_E$. Use the local Lipschitz property of f (Theorem 2.1.12) to deduce that $(\|\phi_n\|)_n$ is bounded. Then use compactness to find convergent subsequence, say $\phi_{n_k} \to \phi$. Now fix $y \in E$. Then

$$\begin{aligned} \phi(y) - \phi(x_0) &= \phi(y) - \phi(x_{n_k}) + \phi(x_{n_k}) - \phi(x_0) \\ &= \lim_{k\to\infty} \phi_{n_k}(y - x_{n_k}) + \phi_{n_k}(x_{n_k} - x_0) \\ &\leq \lim_{k\to\infty} f(y) - f(x_{n_k}) + \phi_{n_k}(x_{n_k} - x_0) = f(y) - f(x). \end{aligned}$$

Therefore $\phi \in \partial f(x_0)$ which contradicts that $\phi_{n_k} \to \phi$.

(b) This follows from (a) and the fact $\partial f(x_0) = \{f'(x_0)\}$ (Theorem 2.2.1).

(c) Suppose not, then there is a subsequence (n_k) and $\epsilon > 0$ such that $\phi_{n_k} \in \partial f_{n_k}(w_{n_k})$, $w_{n_k} \in W$ but $\phi_{n_k} \notin \nabla f(w_{n_k}) + \epsilon B_E$. Because $f_n \to f$ uniformly on bounded sets, it follows that (f_n) is uniformly bounded on bounded sets, and thus (f_n) is eventually uniformly Lipschitz on bounded sets. Hence by passing to a further subsequence, if necessary, we may assume $w_{n_k} \to w_0$, and $\phi_{n_k} \to \phi$ for some $w_0, \phi \in E$. Now let $y \in E$, and observe

$$\begin{aligned} \phi(y) - \phi(x_0) &= \phi(y) - \phi(w_{n_k}) + \phi(w_{n_k}) - \phi(w_0) \\ &= \lim_{k\to\infty} \phi_{n_k}(y - w_{n_k}) + \phi_{n_k}(w_{n_k} - x_0) \\ &\leq \lim_{k\to\infty} f_{n_k}(y) - f_{n_k}(w_{n_k}) + \phi_{n_k}(w_{n_k} - w_0) = f(y) - f(w_0), \end{aligned}$$

where the last equality follows by the uniform convergence of f_{n_k} to f on bounded sets. Thus $\phi \in \partial f(w_0)$, that is $\phi = \nabla f(w_0)$. By (b), $\nabla f(w_k) \to \nabla f(w_0) = \phi$ which yields a contradiction because $\|\phi_{n_k} - \nabla f(w_{n_k})\| > \epsilon$.

(d) For example, let $f_n := \max\{|\cdot|-1/n, 0\}$ and $f := |\cdot|$ on $\mathbb{R}$. Then $\partial f_n(1/n) \not\subset \partial f(1/n)+\frac{1}{2}B_{\mathbb{R}}$ for any $n \in \mathbb{N}$. Indeed, $\partial f_n(1/n) = [0,1]$ while $\partial f(1/n) + \frac{1}{2}B_{\mathbb{R}} = [1/2, 3/2]$. For the remaining part, suppose no such N exists. As in (c), choose $\phi_{n_k} \in \partial f_{n_k}(w_{n_k})$ but $\phi_{n_k} \notin \partial f(w) + \epsilon B_E$ for $\|w - w_{n_k}\| < \delta$ and as in (c), $w_{n_k} \to w_0$ and $\phi_{n_k} \to \phi$ for some $w_0 \in E$ and $\phi \in E$. Again, as in (c), one can show that $\phi \in \partial f(w_0)$. However, for $\|w_{n_k} - w_0\| < \delta$, we have $\phi_{n_k} \notin \partial f(w_0) + \epsilon B_E$ which is a contradiction. $\square$

2.2.23. (a) We will use the max formula (2.1.19). Suppose $x_0 \in \operatorname{bndy} C$ and take $x_n \notin C$ such that $x_n \to x_0$. By the max formula (2.1.19), let $\phi_n \in \partial d_C(x_n)$. Then $\|\phi_n\| \leq 1$ because d_C has Lipschitz constant 1, and $\|\phi_n\| \geq 1$, because we choose $\bar{x}_n \in C$ such that $d_C(x_n) = \|x_n - \bar{x}_n\|$, and then $\langle \phi_n, \bar{x}_n - x_n\rangle \leq -d_C(x_n)$. By the compactness of B_E, we know $\phi_{n_k} \to \bar{\phi}$ for some $\bar{\phi}$, and $\|\bar{\phi}\| = 1$. Also, for any $x \in E$, we have

$$\langle \bar{\phi}, x - x_0\rangle = \lim_{k\to\infty} \langle \phi_{n_k}, x - x_{n_k}\rangle \leq \lim_{n_k} (d_C(x) - d_C(x_{n_k}) = d_C(x) - d_C(x_0).$$

Thus $\bar{\phi} \in \partial d_C(x_0)$. Then $n\bar{\phi} \in \partial n d_C(x_0)$ and so $n\bar{\phi} + \phi \in \partial f(x_0)$.

(b) Let $f(t) := -\sqrt{t}$ for $t \geq 0$ and $f(t) := +\infty$ when $t < 0$. Then $\partial f(0) = \emptyset$. Let $g := \delta_{[0,+\infty)}$. Then $\partial g(0) = (-\infty, 0]$.

Further notes. The proof of (a) shows that given any nonempty convex set A with $x_0 \in \operatorname{bndy} A$, that there exists $\bar{\phi} \in \partial d_C(x_0)$ with $\|\bar{\phi}\| = 1$, and C being the closure of A. It then follows that $\bar{\phi}(x_0) = \sup_C \bar{\phi}$. Had we done this separation theorem earlier, we could have more elegantly completed the proof of Theorem 2.2.1 and part (a) of this exercise.

Exercises from Section 2.3

2.3.2. Using calculus, one can show that for $f := |\cdot|^p/p$ on $\mathbb{R}$ one has $f^* = |\cdot|^q/q$. The Fenchel–Young inequality (2.3.1) then shows $f(x) + f^*(y) \geq xy$, that is,

$$|x|^p/p + |y|^q/q \geq xy \quad \text{for all real } x \text{ and } y,$$

as desired. $\square$

2.3.3. Let $\|f\|_p = \alpha$, $\|g\|_q = \beta$ where $\alpha, \beta > 0$ (if either $\alpha = 0$ or $\beta = 0$, then $fg = 0$ a.e. and so the inequality is trivially true). Now we integrate both sides of the Young inequality:

$$\int_X \frac{f(x)}{\alpha}\frac{g(x)}{\beta} d\mu \leq \int_X \frac{1}{p}\frac{|f(x)|^p}{\alpha^p} + \frac{1}{q}\frac{|g(x)|^q}{\beta^q} d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

Multiplying both sides by $\alpha\beta$ yields the result. $\square$

2.3.4. (a) To show that $x \mapsto \sum_{k=1}^N |x_k|^p$ is convex observe that $g := |\cdot|^p$ is a convex function $\mathbb{R}$, and then

$$x \mapsto \sum_{k=1}^N g(P_k(x)) \quad \text{where } P_k(x) = x_k$$

is a sum of convex functions since $g \circ P_k$ is a convex function for each k as it is a composition of a convex function with a linear function. Now use the gauge construction as suggested in the hint.

(b) Alternatively, one may apply the discrete form of Hölder's inequality of Exercise 2.3.3 as follows:

$$\begin{aligned}
\sum_{k=1}^{N}|x_k+y_k|^p &= \sum_{k=1}^{N}|x_k||x_k+y_k|^{p-1}+|y_k||x_k+y_k|^{p-1}\\
&\le \left(\sum_{k=1}^{N}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{N}|x_k+y_k|^{(p-1)q}\right)^{\frac{1}{q}}+\\
&\qquad \left(\sum_{k=1}^{N}|y_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{N}|x_k+y_k|^{(p-1)q}\right)^{\frac{1}{q}}\\
&= \left(\sum_{k=1}^{N}|x_k+y_k|^p\right)^{\frac{1}{q}}\left[\left(\sum_{k=1}^{N}|x_k|^p\right)^{\frac{1}{p}}+\left(\sum_{k=1}^{N}|y_k|^p\right)^{\frac{1}{p}}\right]
\end{aligned}$$

where we used $(p-1)q=p$. Now divide both sides by $\left(\sum_{k=1}^{N}|x_k+y_k|^p\right)^{\frac{1}{q}}$ using that $1-1/q=1/p$ to obtain

$$\left(\sum_{k=1}^{N}|x_k+y_k|^p\right)^{\frac{1}{p}} \le \left(\sum_{k=1}^{N}|x_k|^p\right)^{\frac{1}{p}}+\left(\sum_{k=1}^{N}|y_k|^p\right)^{\frac{1}{p}}$$

as desired. $\square$

2.3.12. Part (a).

(i) According to the Fenchel–Young inequality (2.3.1) we have

$$\begin{aligned}
f(x)+g(Ax) &\ge \langle A^*\phi, x\rangle - f^*(A\phi)+\langle -\phi, Ax\rangle - g^*(-\phi)\\
&= \langle \phi, Ax\rangle - f^*(A^*\phi) - \langle \phi, Ax\rangle - g^*(-\phi)\\
&= -f^*(A^*\phi)-g^*(-\phi).
\end{aligned}$$

Taking the infimum of the left-hand side over $x\in E$, and then taking the supremum of the right-hand side over $\phi\in Y$ establishes the weak duality inequality $p\ge d$.

(ii) (This part is for the subdifferential sum rule). Let $x\in E$ and suppose $\phi\in\partial f(x)$ and $\Lambda\in\partial g(Ax)$. Then for any $v\in E$ we have

$$\begin{aligned}
\langle \phi + A^*\Lambda, v-x\rangle &= \langle \phi, v-x\rangle + \langle \Lambda, A(v-x)\rangle\\
&\le f(v)-f(x)+g(Av)-g(Ax).
\end{aligned}$$

Thus $\phi + A^*\Lambda \in \partial(f+g\circ A)(x)$ from which the inclusion follows.

(iii) Fix $u\in Y$. Then $f(x)+g(Ax+u)<\infty$ for some $x\in X$ if and only if there exist $x\in \operatorname{dom} f$ such that $Ax+u\in\operatorname{dom} g$ if and only if $u\in \operatorname{dom} g - A\operatorname{dom} f$. Thus $\operatorname{dom} h = \operatorname{dom} g - A\operatorname{dom} f$. To check the convexity of h, suppose $u,v\in\operatorname{dom} h$ and let α,β be any numbers such that

$h(u) < \alpha$ and $h(v) < \beta$. Choose $x_1 \in E$ such that $f(x_1) + g(Ax_1 + u) < \alpha$ and $x_2 \in E$ such that $f(x_2) + g(Ax_2 + v) < \beta$. Then for any $0 \le \lambda \le 1$, we have

$$\begin{aligned} h(\lambda u + (1-\lambda)v) &= \inf_{x \in E}\{f(x) + g(Ax + \lambda u + (1-\lambda)v)\} \\ &\le f(\lambda x_1 + (1-\lambda)x_2) + g(A(\lambda x_1 + (1-\lambda)x_2) + \lambda u + (1-\lambda)v) \\ &\le \lambda f(x_1) + (1-\lambda)f(x_2) + \lambda g(Ax_1 + u) + (1-\lambda)g(Ax_2 + v) \\ &< \lambda\alpha + (1-\lambda)\beta. \end{aligned}$$

Thus, $h(\lambda u + (1-\lambda)v) \le \lambda h(u) + (1-\lambda)h(v)$ as desired.

(iv) Let $x_0 \in \operatorname{dom} f$ be such that $Ax_0 \in \operatorname{cont} g$. Let $y_0 = Ax_0$. Because g is continuous at y_0, this implies $y_0 + rB_Y \subset \operatorname{dom} g$ for some $r > 0$. Therefore,

$$rB_Y = (y_0 + rB_Y) - Ax_0 \subset \operatorname{dom} g - A \operatorname{dom} f$$

which implies $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$ as desired.

Part (b).

(i) First, inclusion was completed in (a)(ii) above. Conversely suppose $\phi \in \partial(f + g \circ A)(\bar{x})$. Applying the Fenchel–Young inequality (2.3.1), and then applying the Fenchel duality theorem (2.3.4), we obtain

$$\begin{aligned} f(\bar{x}) + g(A\bar{x}) - \langle \phi, \bar{x} \rangle &= \inf_{x \in E}\{f(x) + g(Ax) - \langle \phi, x \rangle\} = \inf_{x \in E}\{(f - \phi)(x) + g(Ax)\} \\ &= -(f - \phi)^*(A^*\bar{\phi}) - g^*(-\bar{\phi}), \end{aligned}$$

where $\bar{\phi} \in Y$ is a point where d in the Fenchel duality theorem (2.3.4) is attained. Therefore,

$$(f - \phi)(\bar{x}) - \langle A^*\bar{\phi}, \bar{x} \rangle + g(A\bar{x}) - \langle -\bar{\phi}, A\bar{x} \rangle = -(f - \phi)^*(A^*\bar{\phi}) - g^*(-\bar{\phi})$$

and by Fenchel–Young inequaltiy (2.3.1), $A^*\bar{\phi} \in \partial(f - \phi)(\bar{x})$ and $-\bar{\phi} \in \partial g(A\bar{x})$. The first inclusion implies $A^*\bar{\phi} + \phi \in \partial f(\bar{x})$, and using the second inclusion we check

$$\langle -A^*\bar{\phi}, u - \bar{x} \rangle = \langle -\bar{\phi}, A(u - \bar{x}) \rangle \le g(Au) - g(A\bar{x}) \quad \text{for all } u \in E;$$

thus $-\bar{\phi} \in \partial g(A\bar{x})$, consequently equality holds in the sum formula.

(ii) The previous part has proved the 'only if' assertion, and we can essentially reverse our steps to deduce the 'if' assertion. □

2.3.13. Suppose $f : E \to \mathbb{R}$ has Lipschitz constant k. Suppose $\phi \in E$ and $\|\phi\| > k$. Choose $x_0 \in E$ with $\|x_0\| = 1$ and $\phi(x_0) > k$. Then $\lim_{t \to \infty} \phi(tx) - f(tx) \to \infty$. So $\phi \notin \operatorname{dom} f^*$.
For the converse, assume $\operatorname{dom} f^* \subset kB_E$ is not empty, then f is bounded below by $\phi - a$ where $\phi \in \operatorname{dom} f^*$ and $a = f^*(\phi)$. (Using relative interior properties, one knows that the domain of the subdifferential of a proper convex function on E is nonempty, and hence the domain of the conjugate is not empty; see Theorem 2.4.8). Then if $\operatorname{dom} f \ne E$, one can find $y_n \in \operatorname{dom} f^*$ such that $\|y_n\| \to \infty$: for example, letting $f_n := \phi - a + nd_C$ where $C := \overline{\operatorname{dom} f}$, one has $f_n \le f$, but for $x \notin C$, and $y \in \partial f_n(x)$, one has $\|y\| \ge n - \|\phi\|$. Since $y \in \operatorname{dom} f^*$, this yields a contradiction.

Thus $\operatorname{dom} f = E$ and thus f is continuous. In the event f is not k-Lipschitz, one can choose $u, v \in E$ such that $f(v) - f(u) > k\|u - v\|$. Now let $y \in \partial f(v)$. Then

$$\langle y, u - v \rangle < -k\|u - v\|$$

and so $\|y\| > k$, but $y \in \operatorname{dom} f^*$ which is a contradiction.
This result fails if f is not convex: for example, consider $f(x) := \sqrt{|x|}$ on $\mathbb{R}$. Then $\operatorname{dom} f^* = \{0\}$, but f is not Lipschitz. □

2.3.14. Basic facts about infimal convolutions.

(a) Let $(x, s) \in \operatorname{epi} f$ and $(y, t) \in \operatorname{epi} g$. Then

$$(f \,\square\, g)(x + y) \leq f(x) + g(x + y - x) = f(x) + g(y) \leq s + t$$

and so $(x + y, s + t) \in \operatorname{epi}(f \,\square\, g)$. That is, $\operatorname{epi} f + \operatorname{epi} g \subset \operatorname{epi}(f \,\square\, g)$. Now suppose h is a function such that there exists $\bar{x} \in E$ with $h(\bar{x}) > (f \,\square\, g)(\bar{x})$. Choose $t \in \mathbb{R}$ such that $h(\bar{x}) > t > (f \,\square\, g)(\bar{x})$. Then we choose $y \in E$ such that $f(\bar{x}) + g(y - \bar{x}) < t$ Then $(\bar{x}, t) \notin \operatorname{epi} h$, but $(\bar{x}, t) \in \operatorname{epi}(f \,\square\, g)$. Thus $(f \,\square\, g)$ is the largest function whose epigraph contains $\operatorname{epi} f + \operatorname{epi} g$.

(b) As suggested, let $f(x) := e^x$ and $g(x) := 0$. Then f and g are continuous and convex, but $\operatorname{epi} f + \operatorname{epi} g = \{(x, y) \in \mathbb{R}^2 : y > 0\}$.

(c) As suggested, let $f(x) := x$ and $g(x) := 0$. For any $u \in \mathbb{R}$,

$$(f \,\square\, g)(u) \leq f(-e^n) + g(u + e^n) = -e^n \quad \text{for all } n \in \mathbb{N}.$$

Thus $(f \,\square\, g)(u) = -\infty$ for all $u \in \mathbb{R}$.

(d) As suggested let $C := \{(x, y) : y \geq e^x\}$ and $D := \{(x, y) : y \geq 0\}$. Then $\delta_C \,\square\, \delta_D = \delta_{\{(x,y):y>0\}}$ which is not closed.

(e) Suppose f and g are convex functions. Let $u, v \in \operatorname{dom}(f \,\square\, g)$. Let α and β be any real numbers satisfying $(f \,\square\, g)(u) < \alpha$ and $(f \,\square\, g)(v) < \beta$. Now choose $x_1, x_2 \in E$ so that

$$f(x_1) + g(u - x_1) < \alpha \quad \text{and} \quad f(x_2) + g(v - x_2) < \beta.$$

Then

$$\begin{aligned}
(f \,\square\, g)(\lambda u + (1 - \lambda)v) &\leq f(\lambda x_1 + (1 - \lambda)x_2) + g(\lambda(u - x_1) + (1 - \lambda)(v - x_2)) \\
&\leq \lambda f(x_1) + (1 - \lambda)f(x_2) + \lambda g(u - x_1) + (1 - \lambda)g(v - x_2) \\
&< \lambda\alpha + (1 - \lambda)\beta.
\end{aligned}$$

It follows that $f \,\square\, g$ is convex.

(f) Notice that (c) already shows this may fail if one of the functions is not bounded below, and we need to explicitly assume g is proper and let $x_0 \in \operatorname{dom} g$. Then

$$\inf_E f + \inf_E g \leq (f \,\square\, g)(x) \leq g(x_0) + f(x - x_0).$$

When f is continuous, this implies $(f \,\square\, g)$ is real-valued and hence continuous. When f is bounded on bounded sets, so is $(f \,\square\, g)$. When f is Lipschitz with Lipschitz constant $k \geq 0$, then $f \leq k\|\cdot\| + f(0)$ and so $f \,\square\, g \leq k\|\cdot\| + b$ where $b := g(x_0) + f(0) + k\|x_0\|$ which implies $(f \,\square\, g)$ is Lipschitz with Lipschitz constant k (see Exercise 4.1.28). See also the note following (g).

(g) Observe that

$$(\|\cdot\| \,\square\, \delta_C)(x) = \inf_{y \in X} \|x-y\| + \delta_C(y) = \inf_{y \in C} \|x-y\| = d_C(x).$$

As in the proof of (f), the convolution has Lipschitz constant 1 because the norm has Lipschitz constant 1.

Further notes. One may prefer a more explicit argument in (f). Indeed, once we have established $(f \,\square\, g)$ is real-valued, suppose $(f \,\square\, g)(\bar{x}) = \bar{t}$. Given $\epsilon > 0$, choose $y \in E$ such that $f(y)+g(\bar{x}-y) < \bar{t}+\epsilon$. Then for any $h \in E$,

$$\begin{aligned}(f \,\square\, g)(h+\bar{x}) &\leq f(y+h)+g(\bar{x}-y) \leq |f(y+h)-f(y)| + f(y) + g(\bar{x}-y) \\ &< (f \,\square\, g)(h+\bar{x}) + |f(y+h)-f(y)| + \epsilon.\end{aligned}$$

From here, local/global Lipschitz properties of the convolution then follow directly from the local/global Lipschitz properties of f (this argument works just as well in any normed linear space irrespective of the dimension). $\square$

2.3.15. For (a) see the proof of Lemma 4.4.15; for (b) see the proof of Lemma 4.4.16; and for (c), use (a) and (b), c.f. Corollary 4.4.17.

(d) Observe that if f and g are closed, then $(f^* \,\square\, g^*)^* = f^{**} + g^{**} = f + g$, where the first equality follows from (a). Then $(f+g)^* = (f^* \,\square\, g^*)$. The result as stated follows because $\mathrm{cl}(f+g) = \mathrm{cl}\, f + \mathrm{cl}\, g$ under the condition $\mathrm{dom}\, f \cap \mathrm{cont}\, g \neq \emptyset$ as we now sketch.

Clearly, $\mathrm{cl}\, f + \mathrm{cl}\, g \leq f+g$ and so $\mathrm{cl}\, f + \mathrm{cl}\, g \leq \mathrm{cl}(f+g)$. Let $\bar{x} \in \mathrm{dom}(\mathrm{cl}\, f + \mathrm{cl}\, g)$. We will show $\mathrm{cl}(f+g)(\bar{x}) \leq (\mathrm{cl}\, f + \mathrm{cl}\, g)(\bar{x})$. Fix $v \in \mathrm{int\,dom}\, g \cap \mathrm{dom}\, f$ and choose $r > 0$ so that g is bounded on $v + rB_X \subset \mathrm{int\,dom}\, g$. Now choose $u_n \in \mathrm{dom}\, f$ with $u_n \to \bar{x}$ and $f(u_n) \to \mathrm{cl}\, f(\bar{x})$. For $0 < \lambda < 1$, $\lambda\bar{x} + (1-\lambda)(v + rB_X) \subset \mathrm{int\,dom}\, g$. Because $u_n \to \bar{x}$, we fix $\lambda_n \to 1$ so that

$$\lambda_n u_n + (1-\lambda_n)\bar{v} \in \lambda_n \bar{x} + (1-\lambda_n)(\bar{v} + rB_X) \subset$$

Now,

$$\begin{aligned}g(\lambda_n u_n + (1-\lambda_n)(\bar{v})) &= \mathrm{cl}\, g(\lambda_n u_n + (1-\lambda_n)(\bar{v})) \\ &= \mathrm{cl}\, g(\lambda_n \bar{x} + (1-\lambda_n) v_n) \text{ where } v_n \in \bar{v} + rB_X \\ &\leq \lambda_n \,\mathrm{cl}\, g(\bar{x}) + (1-\lambda_n)\,\mathrm{cl}\, g(u_n) \to \mathrm{cl}\, g(\bar{x}).\end{aligned}$$

Similarly, $f(\lambda_n u_n + (1-\lambda_n)(\bar{v}) \leq \lambda_n f(u_n) + (1-\lambda_n) f(\bar{v}) \to \mathrm{cl}\, f(\bar{x})$. Altogether, we conclude $\mathrm{cl}(f+g)(\bar{x}) \leq \mathrm{cl}\, f(\bar{x}) + \mathrm{cl}\, g(\bar{x})$. $\square$

2.3.16. Suppose $f : C \to \mathbb{R}$ is Lipschitz with Lipschitz constant k, and let $\tilde{f}(x) := \inf\{f(y) + k\|x-y\| : y \in C\}$. For $x_0 \in C$, taking $y = x_0$ clearly shows $\tilde{f}(x_0) \leq f(x_0)$, hence $\tilde{f} \leq f$ on C. Also, fix $x_0 \in C$. Then $f(x) \leq f(x_0) + k\|x - x_0\|$, and so $f(x) < \infty$ for all $x \in X$. Moreover, $f(x) \geq f(x_0) - k\|x-x_0\|$ for $x \in C$. Now fix $x \in X$, then $f(y) + k\|x-y\| \geq f(x_0) - k\|y-x_0\| + k\|x-y\| \geq f(x_0) - k\|x-x_0\|$ for all $y \in C$. This shows $\tilde{f}(x) > -\infty$ for all $x \in X$, i.e. $\tilde{f}$ is real-valued.

Suppose $\tilde{f}(x_0) < f(x_0)$ for some $x_0 \in C$. Then there exists $y \in C$ such that $f(y) + k\|x_0 - y\| < f(x_0)$. This violates the Lipschitz constant of f on C. Hence $\tilde{f}(x) = f(x)$ for all $x \in C$. Similarly, one can see that $\tilde{f}$ is globally Lipschitz with Lipschitz constant k. Indeed, suppose $\tilde{f}(u) - \tilde{f}(v) >$

$k\|u-v\|+\epsilon$ for some $u, v \in X$ and $\epsilon > 0$. Choose $x_0 \in C$ such that $\tilde{f}(v) \le f(x_0) + k\|v - x_0\| + \epsilon$. Then

$$\tilde{f}(u) \le f(x_0) + k\|u - x_0\| \le f(x_0) + k\|v - x_0\| + k\|u - v\|$$

which is a contradiction. □

2.3.17. (a) Suppose x and y are global minimizers of f. Then $f(\lambda x + (1-\lambda)y) \ge f(x) = \lambda f(x) + (1-\lambda) f(y)$ for $0 < \lambda < 1$. Because f is strictly convex, $x = y$.

(b) It suffices to show this for $y = 0$, and it suffices to show

$$f\left(\frac{u+v}{2}\right) < \frac{1}{2} f(u) + \frac{1}{2} f(v) \text{ for all distinct } u, v \in E. \tag{2}$$

Indeed, if $f(\lambda u + (1-\lambda)v) = \lambda f(u) + (1-\lambda) f(v)$ for some $0 < \lambda < 1$ and distinct u and v, then by the convexity of f equality holds for all $0 < \lambda < 1$. We now check that (2) is an easy consequence of the parallelogram identity:

$$\begin{aligned} f\left(\frac{u+v}{2}\right) &= \frac{1}{2}\left\|\frac{u+v}{2}\right\|^2 = \frac{1}{2}\|u\|^2 + \frac{1}{2}\|v\|^2 - \frac{1}{2}\|u-v\|^2 \\ &< \frac{1}{2} f(u) + \frac{1}{2} f(v) \text{ when } u \ne v. \end{aligned}$$

(c) (i) Let $y \in E$ and define f by $f(x) := \frac{1}{2}\|x - y\|^2 + \delta_C$. The strict convexity of f follows from (b), and f attains its minimum on C because any minimizing sequence is bounded and a convergent subsequence converges to the unique, by part (a), minimizer.

Now let $y \in E$, and suppose $\bar{y} \in C$ satisfies $\langle y - \bar{y}, x - \bar{y}\rangle \le 0$ for all $x \in C$. Then for $x \in C$

$$\begin{aligned} \|y - \bar{y}\|^2 &= \langle y - \bar{y}, y - \bar{y}\rangle = \langle y - \bar{y}, y - x\rangle + \langle y - \bar{y}, x - \bar{y}\rangle \\ &\le \langle y - \bar{y}, y - x\langle \le \|y - \bar{y}\|\,\|y - x\| \end{aligned}$$

and so $\|y - \bar{y}\| \le \|y - x\|$ for all $x \in C$, thus $\bar{y} = P_C(y)$.

Conversely, suppose $\bar{y} \in C$ satisfies $\langle y - \bar{y}, x - \bar{y}\rangle > 0$ for some $x \in C$. Then for each $0 < \lambda < 1$, the convexity of C implies the point $x_\lambda := \lambda x + (1-\lambda)\bar{y}$ is in C. Now compute

$$\begin{aligned} \|y - x_\lambda\|^2 &= \langle y - x_\lambda, y - x_\lambda\rangle \\ &= \langle y - \bar{y} - \lambda(x - \bar{y}), y - \bar{y} - \lambda(x - \bar{y})\rangle \\ &= \|y - \bar{y}\|^2 - 2\lambda\langle y - \bar{y}, x - \bar{y}\rangle + \lambda^2\|x - \bar{y}\|^2 \\ &= \|y - \bar{y}\|^2 - \lambda[2\langle y - \bar{y}, x - \bar{y}\rangle - \lambda\|x - \bar{y}\|^2]. \end{aligned}$$

For $\lambda > 0$ sufficiently small, the term in the brackets is positive and then $\|y - \bar{y}\|^2 > \|y - x_\lambda\|^2$, and so $\bar{y} \ne P_C(y)$.

(ii) Let $\bar{x} \in C$. Then $d \in N_C(\bar{x})$ if and only if $d \in \partial\delta_C(\bar{x})$ if and only if

$$\langle \bar{x} + d - \bar{x}, x - \bar{x} \le \delta_C(x) - \delta_C(\bar{x}) = 0 \text{ for all } x \in C$$

if and only if $P_C(\bar{x} + d) = \bar{x}$ (by part (i)).

(iii) In fact one can show

$$\|P_C(x) - P_C(y)\|^2 + \|x - P_C(x) - (y - P_C(y))\|^2 \le \|x - y\|^2 \text{ for all } x, y \in E. \tag{3}$$

Indeed, expanding and rearranging using the inner product, the left-hand side of (3) is equal to

$$\langle x-y, x-y\rangle + 2\langle y-P_C(y), P_C(x)-P_C(y)\rangle + 2\langle x-P_C(x), P_C(y)-P_C(x)\rangle \tag{4}$$

and by part (i) the last two inner products in (4) are less than or equal to 0 which provides the desired conclusion.

(d) For example $S = \{-1,1\}$, then $P_S(0)$ is multi-valued, and $\lim_{x\to 0^+} P(x) = \{1\}$ while $\lim_{x\to 0^-} P(x) = \{-1\}$ so there is no single-valued selection of P that is continuous at $\{0\}$. $\square$

2.3.20. (a) Let $f(x) := \frac{\|x\|^2}{2} - \delta_S(x)$. The proof of Fact 4.5.6 shows

$$\frac{1}{2}d_S^2(y) = \frac{\|y\|^2}{2} - f^*(y)$$

or $d_S(\cdot) = \|\cdot\|^2 - 2f^*(\cdot)$ is a difference of convex functions as desired.

(b) $d_C = \|\cdot\| \,\square\, \delta_C$ and thus $d_C^* = (\|\cdot\|)^* + \delta_C^* = \delta_{B_E} + \sigma_C$.

(c) Let $x \in C$. By Fenchel–Young (Proposition 2.3.1), $\phi \in \partial d_C(x)$ if and only if $d_C^*(\phi) = \phi(x) - d_C(x)$ if and only if $\delta_{B_E}(\phi) + \sigma_C(\phi) = \phi(x) - d_C(x) = \phi(x)$ if and only if $\phi \in B_E$ and $\phi(x) = \sigma_C(\phi)$ if and only if $\phi \in \partial\delta_C$ and $\phi \in B_E$ if and only if $\phi \in N_C(x)$ and $\phi \in B_E$.

(d) Suppose $x \notin C$. Then $\phi \in \partial d_C(x)$ if and only if $\phi \in B_E$ and

$$\langle \phi, x - P_C(x)\rangle \geq d_C(x) - d_C(P_C(x)) = d_C(x) = \|x - P_C(x)\|.$$

Therefore, $\phi = \dfrac{1}{\|x - P_C(x)}(x - P_C(x)) = \dfrac{1}{d_C(x)}(x - P_C(x))$.

(e) For $x \in C$, we obtain that $\nabla d_C^2(x) = 0$ because

$$\lim_{t\to 0} |\nabla d_C^2(x+th) - d_C^2(x) - \langle 0, th\rangle| \leq t^2\|h\|.$$

For $x \notin C$, the chain rule implies $\nabla d_C^2(x) = 2d_C(x)\nabla d_C(x) = 2(x - P_C(x))$ (by part (c)). Thus (e) follows. $\square$

2.3.21. Let $D := \{x \in E : Ax = b\}$. Let $x \in D$, then $\phi \in \partial D(x)$ if and only if $\phi(y - x) \leq 0$ for all $y \in D$ if and only if $\phi(u) \leq 0$ for all $u \in \ker A$ if and only if $\phi(u) = 0$ for all $u \in \ker A$.

Suppose $\phi \in A^*Y$, that is $\phi = A^*y$ for some $y \in Y$. Fix $u \in D$, and let $v \in D$ be arbitrary, then

$$\langle A^*y, v - u\rangle = \langle y, A(v-u)\rangle = \langle y, b - b\rangle = 0.$$

Therefore $\phi \in \partial D(u)$, that is $\phi \in N_C(u)$

Conversely, suppose $\phi \in \partial\delta_D(x)$. Suppose $\phi \neq 0$, then fix $x_0 \in X$ such that $\phi(x_0) = 1$. We now express E as the direct sum $\ker\phi \oplus \mathbb{R}x_0$. Observe that $Ax_0 \notin A(\ker\phi)$. Indeed, suppose $Ax_0 = Ax_1$ for some $x_1 \in \ker\phi$. Then $A(x_0 - x_1) = 0$ and so by the previous paragraph $\phi(x_0 - x_1) = 0$. Thus, by the basic separation theorem (2.1.21), we choose $y \in Y$ such that $y(Ax_0) = 1$ and $y(A(\ker\phi)) = \{0\}$. Given $\bar{x} \in E$, we write $\bar{x} = h + \phi(\bar{x})x_0$ where $k \in \ker\phi$. Then

$$\langle A^*y, \bar{x}\rangle = \langle y, A(h + \phi(\bar{x})x_0)\rangle = \langle y, Ah\rangle + \phi(\bar{x})\langle y, Ax_0\rangle = \phi(\bar{x}).$$

Because $\bar{x} \in E$ was arbitrary, we have $\phi = A^*y$, and $A^*Y \subset \delta_D(x)$ as desired.

(a) Suppose $\bar{x}$ is a local minimizer as specified. Then $\nabla f(\bar{x})| \in \partial\delta_D(\bar{x})$ and so $\nabla f(\bar{x}) \in A^*Y$.

(b) Suppose $\nabla f(\bar{x}) \in A^*Y$ and f is convex. Then $\nabla f(\bar{x}) \in \partial\delta_D(\bar{x})$, and because f is convex, $\bar{x}$ is a global minimizer of $f|_D$. $\square$

Exercises from Section 2.4

2.4.1. Part(a).

(i) As in the proof of the Fenchel duality theorem (2.3.4), let $h : Y \to [-\infty, +\infty]$ be defined by

$$h(u) := \inf_{x\in E}\{f(x) + g(Ax + u)\},$$

then h is convex and $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f) = \operatorname{dom} h$, thus by the max formula (2.1.19) $\partial h(0)$ is not empty, so we let $-\phi \in \partial h(0)$. Now for all $x \in E$ and $u \in Y$ with $u = Av$ where $v \in E$, we have

$$\begin{aligned} 0 &\leq h(0) \leq h(u) + \langle \phi, u\rangle \\ &\leq f(x) + g(A(x + v)) + \langle \phi, Av\rangle \\ &= [f(x) - \langle A^*\phi, x\rangle] - [-g(A(x + v)) - \langle \phi, A(x + v)\rangle]. \end{aligned}$$

Define

$$b := \inf_{x\in E}\{f(x) - \langle A^*\phi, x\rangle\} \qquad a := \sup_{z\in E}\{-g(A(z)) - \langle A^*\phi, z\rangle\}. \tag{5}$$

Then $a \leq b$ and thus for any $r \in [a, b]$ we have

$$f(x) \geq r + \langle A^*\phi, x\rangle \geq -(g \circ A)(x) \quad \text{for all } x \in E.$$

(ii) With notation as in the Fenchel duality theorem (2.3.4), observe $p \geq 0$ because $f(x) \geq -g(Ax)$, and then the Fenchel duality theorem (2.3.4) says $d = p$ and because the supremum in d is attained, we choose $\phi \in Y$ such that

$$\begin{aligned} 0 \leq p &= -f^*(A^*\phi) - g^*(-\phi) \\ &\leq [f(x) - \langle \phi, Ax\rangle] + [g(y) + \langle \phi, y\rangle] \ \text{ for all } x \in X,\ y \in Y, \end{aligned}$$

where the second inequality is a direct consequence of the definitions of $f^*(A^*\phi)$ and $g^*(-\phi)$. Then for any $z \in E$, setting $y = Az$, in the previous inequality, we deduce $a \leq b$ where a and b are as in (5). Now choose $r \in [a, b]$ and let $\alpha(x) = \langle A^*\phi, x\rangle + r$.

(iii) The inclusion is straightforward (Exercise 2.3.12 (a)(ii)), so we prove the reverse inclusion. Suppose $\phi \in \partial(f + g \circ A)(\bar{x})$. Because shifting by a constant does not change the subdifferential, we may assume without loss of generality that

$$x \mapsto f(x) + g(Ax) - \phi(x)$$

attains its minimum of 0 at $\bar{x}$. According to the sandwich theorem (2.4.1) there exists an affine function $\alpha := \langle A^*y, \cdot\rangle + r$ with $-y \in \partial g(A\bar{x})$ such that

$$f(x) - \phi(x) \geq \alpha(x) \geq -g(Ax) \quad \text{for all } x \in E, \text{ with equality when } x = \bar{x}$$

Then $f(x) \geq \langle \phi + A^*y, x\rangle + r$ and $f(\bar{x}) = \langle \phi + A^*y, \bar{x}\rangle + r$. Thus $\phi + A^*y \in \partial f(\bar{x})$, and as a consequence, we have $\phi \in \partial f(\bar{x}) + A^*\partial g(A\bar{x})$ as desired.

(iv) Let the notation be as in the Hahn–Banach extension theorem (2.1.18). Then $-g \le p$ where $g = -f + \delta_S$. Because p is everywhere continuous, we can apply the sandwich theorem (2.4.1) to find an affine mapping α such that $-g \le \alpha \le p$, that is $f \le \alpha \le p$. Then $\alpha = \alpha(0) + \phi$ where $\phi \in E$. We know $\alpha(0) \ge f(0) = 0$ and so $\phi \le p$. We claim $\phi(s) = f(s)$ for all s in the linear subspace S. Indeed, if this were not true, then $f(x_0) - \phi(x_0) \ne 0$ for some $x_0 \in S$. Then choose $k \in \mathbb{R}$ so that $k(f(x_0) - \phi(x_0)) > \alpha(0)$. This implies $f(kx_0) > \phi(kx_0) + \alpha(0)$ which is impossible. Thus $\phi|_S = f$ as desired.

Part (b). Several connections were outlined in (a) and earlier, for now we'll derive a couple additional easy relations, to sketch one complete circle.

(i) Suppose the subdifferential sum rule is valid, and $x_0 \in \operatorname{core} \operatorname{dom} f$ where $f : E \to (-\infty, +\infty]$ is convex. Then $E = \partial(f + \delta_{\{x_0\}})(x_0) = \partial f(x_0) + \partial \delta_{\{x_0\}}(x_0)$ and so $\partial f(x_0)$ is not empty.

(ii) Suppose the subdifferential of a convex function on E at a point of continuity is not empty. Now consider a linear function f on a subspace Y of E and $f|_Y \le p$ for some sublinear function p on E. Consider $h = f \,\square\, p$. Then $h \le p$, $h|_Y = f$, h is continuous with $h(0) = 0$. Consider $\phi \in \partial h(0)$. Then $\phi|_Y = f$, and $\phi \le h$. Thus the Hahn-Banach extension theorem (2.1.18) follows.

So one of the circles of implications we have sketched is: Hahn–Banach extension $\Rightarrow$ max formula $\Rightarrow$ Fenchel duality theorem $\Rightarrow$ sandwich theorem $\Rightarrow$ nonemptiness of subdifferential $\Rightarrow$ Hahn–Banach extension theorem. Where the proofs of the respective implications are given in: proof of max formula (2.1.19), proof of the Fenchel duality theorem (2.3.4), Part(a)(ii), Part(a)(iii), Part(b)(i), Part(b)(ii).

<u>Further notes.</u> (I) The Fenchel duality and the sandwich theorems are most easily visualized and understood in the classical case $Y = E$ where A is the identity map, and yet still very powerful. In this case, the primal and dual problems are:

$$p := \inf_{x \in E}\{f(x) + g(x)\} \quad \text{and} \quad d := \sup_{y \in E}\{-f^*(y) - g^*(-y)\}.$$

As before, $p \ge d$ by the Fenchel–Young inequality (2.3.1). We derive $p = d$ using the sandwich theorem (2.4.1) when $0 \in \operatorname{core}(\operatorname{dom} g - \operatorname{dom} f)$. Indeed, when $p > -\infty$, we know $f(x) \ge p - g(x)$ for all $x \in E$, and thus there is an affine function $\alpha := \phi + r$ such that

$$f(x) \ge \langle \phi, x \rangle + r \ge -g(x) + p \quad \text{for all} \quad x \in E.$$

Then $\langle \phi, x\rangle - f(x) \le -r$ for all $x \in E$ and $\langle -\phi, x\rangle - g(x) \le r - p$ for all $x \in E$. Thus $f^*(\phi) \le -r$ and $g^*(-\phi) \le r - p$. In other words, $p \ge d \ge -f^*(\phi) - g^*(-\phi) \ge r + p - r = p$ and so $p = d$ and the sup is attained at ϕ as desired.

(II) Of course, one can derive the Fenchel dualilty theorem (2.3.4) from the sandwich theorem (2.4.1) by slightly modifying the proof as presented in the text. Indeed, let h be as defined in the proof of the Fenchel dualilty theorem (2.3.4), and observe $h \ge p - \delta_{\{0\}}$, and $h(0) = p$. By the sandwich theorem (2.4.1), there is an affine function, say $\alpha := p - \phi$ such that $h \ge \alpha \ge p - \delta_{\{0\}}$, and thus $-\phi \in \partial h(0)$. Now proceed as in the proof for the Fenchel dualilty theorem (2.3.4).

(III) A more direct derivation of the Fenchel dualilty theorem (2.3.4) from the sandwich theorem (2.4.1) is as follows. Let $h : E \times Y \to (-\infty, +\infty]$ be defined by $h(x, y) := f(x) + g(y)$. Then $h \ge p - \delta_{G(A)}$, where $G(A) := \{(x, y) : y = Ax\}$ is the graph of $A : E \to Y$, and we apply the sandwich theorem (2.4.1) to obtain $\Lambda \in E$, $\phi \in Y$ and $r \in \mathbb{R}$ such that

$$g(y) + f(x) \ge \Lambda(x) - \phi(y) + r \ge p - \delta_{G(A)}, \text{ for all } x \in X, y \in Y. \tag{6}$$

When $x = 0$ and $y = 0$, (6) implies $r \geq p$, and (6) further implies $\Lambda(x) - \phi(Ax) \geq p - r$ for all $x \in E$, and so $\Lambda = A^*\phi$. Lastly, we rewrite the left inequality of (6) as

$$-[-\phi(y) - g(y)] - [A^*\phi(x) - f(x)] \geq r \ (\geq p)$$

and take the infimum over $y \in Y$ and then over $x \in E$ to deduce $-g^*(\phi) - f^*(A^*\phi) \geq p$ which together with the weak duality inequality provides the result. $\square$

2.4.3. (a) Fix $\epsilon > 0$ and let $\bar{x} \in \operatorname{cl} C$. Then there exists a sequence $(x_n) \subset C$ such that $x_n \to \bar{x}$. Fix n_0 such that $\|x_{n_0} - \bar{x}\| < \epsilon$. Then $\bar{x} \in x_{n_0} + \epsilon B_X \subset C + \epsilon B_E$.
(b) Let $x + y \in D + F$ where $x \in D$, $y \in F$. Because D is open, we choose $\epsilon > 0$ so that $x + \epsilon B_E \subset D$. Then $x + y + \epsilon B_E \subset D + F$. Therefore, $x + y \in \operatorname{int}(D + F)$ and so $D + F$ is open.
(c) Let $x \in \operatorname{int} C$ and choose $\epsilon > 0$ so that $x + \epsilon B_E \subset C$. By part (a), for each $\lambda > 0$, $\operatorname{cl} C \subset C + \frac{\lambda\epsilon}{1-\lambda} B_E$. Therefore,

$$\begin{aligned} \lambda x + (1-\lambda) \operatorname{cl} C &\subset \lambda x + (1-\lambda)\left(C + \frac{\lambda\epsilon}{1-\lambda} B_E\right) \\ &= \lambda(x - \epsilon B_E) + (1-\lambda)C \\ &= \lambda(x + \epsilon B_E) + (1-\lambda)C \subset C. \end{aligned}$$

Because $x \in \operatorname{int} C$ was arbitrary it follows that

$$\lambda \operatorname{int} C + (1-\lambda) \operatorname{cl} C \subset C, \quad \text{for each } 0 < \lambda \leq 1,$$

and by part (b), the sum on the left hand side is open, and so

$$\lambda \operatorname{int} C + (1-\lambda) \operatorname{cl} C \subset \operatorname{int} C, \quad \text{for each } 0 < \lambda \leq 1,$$

as desired.
(d) Because $\operatorname{int} C \subset \operatorname{cl} C$, the previous inequality implies (trivially) $\lambda x + (1-\lambda)y \in \operatorname{int} C$ for all $x, y \in \operatorname{int} C$ and $0 < \lambda < 1$, and so $\operatorname{int} C$ is convex.
(e) For any fixed $x \in \operatorname{int} C$, $\lambda x + (1-\lambda) \operatorname{cl} C \subset \operatorname{int} C$. Letting $\lambda \to 0^+$, we deduce that $\operatorname{cl} C \subset \operatorname{cl}(\operatorname{int} C)$.
Clearly this can fail without convexity. For example, let $S := \mathbb{Q} \cup (0, 1) \subset \mathbb{R}$. Then $\operatorname{cl} S = \mathbb{R}$, but $\operatorname{cl}(\operatorname{int} S) = [0, 1]$.
Further notes: in any normed linear space it is straightforward to show the interior of a convex set is convex: Let $x, y \in \operatorname{int} C$, and choose $r > 0$ so that $x + rB_X \subset C$ and $y + rB_X \subset C$. Then for $0 \leq \lambda \leq 1$, one has

$$\lambda(x + rB_X) + (1-\lambda)(y + rB_X) \subset C.$$

Then, $\lambda x + (1-\lambda)y + rB_X \subset C$ and $\lambda x + (1-\lambda)y \subset \operatorname{int} C$ as desired. It is also easy to see that the closure of a convex set is closed (see solution to Exercise 2.4.8). See also [383, Theorem 1.13] for more. $\square$

2.4.4.(a) Let $(A_i)_{i \in I}$ be a collection of affine sets. Let $A = \bigcap_{i \in I} A_i$, and let $x, y \in A$ and $\lambda \in \mathbb{R}$. Then $x, y \in A_i$ for each $i \in I$ and so $\lambda x + (1-\lambda)y \in A_i$ for each $i \in I$. Therefore, $\lambda x + (1-\lambda)y \in A$, and so A is affine.
(b) Suppose A is a nonempty affine subset of E. Fix $x_0 \in A$. We claim that $Y := A - x_0$ is linear. Indeed, then for $\alpha, \beta \in \mathbb{R}$ and $x, y \in L$ we have $x = u - x_0$ and $y = v - x_0$ where $u, v \in A$, and

$$\alpha x + \beta y + x_0 = \alpha u - \alpha x_0 + \beta v - \beta x_0 + 1x_0 \in A$$

because $\alpha - \alpha + \beta - \beta + 1 = 1$ (see part (c)). Therefore, $\alpha x + \beta y \in A - x_0$, and $A - x_0$ is a linear subspace. Conversely, if Y is linear, and $x_0 \in E$, then $A := Y + x_0$ is affine. Indeed, for $x, y \in Y$, and $\alpha + \beta = 1$, we have

$$\alpha(x + x_0) + \beta(y + y_0) = \alpha x + \beta y + x_0 \in A,$$

which completes the proof of (b). (Alternatively, one can directly deduce this from Lemma 2.4.5).

(c) This part follows immediately from Lemma 2.4.5 which shows $\operatorname{aff} D = x + \operatorname{span}(D - x)$ for any $x \in D$. Indeed, suppose $x_1, x_2, \ldots, x_m \in D$. Then for $\sum_{i=1}^m \lambda_i = 1$, we have

$$\sum_{i=1}^m \lambda_i x_i = x + \sum_{i=1}^m \lambda_i (x_i - x) \in \operatorname{aff} D.$$

Conversely, suppose $u \in \operatorname{aff} D$. Then $u \in x + \operatorname{span}(D - x)$ and so

$$u = x + \sum_{i=1}^m \alpha_i (x_i - x) = \sum_{i=1}^m \alpha_i x_i + (1 - \sum_{i=1}^m \alpha_i) x.$$

Thus $u \in \operatorname{aff} D$ and this proves (c).

(d) Suppose D is nonempty. Linear subspaces of E are closed, therefore $\operatorname{aff} D = x + \operatorname{span}(D - x)$ is closed for any $x \in D$. Consequently, $\operatorname{cl} D \subset \operatorname{aff} D$. It then follows that $\operatorname{aff}(\operatorname{cl} D) \subset \operatorname{aff} D$. Because the reverse inclusion is clear, we deduce $\operatorname{aff}(\operatorname{cl} D) = \operatorname{aff} D$. $\square$

2.4.5. (a) Consider $C_1 = [0, 1] \times \{0\}$ and $C_2 = [0, 1] \times [0, 1]$ as subsets of $\mathbb{R}^2$. Then $\operatorname{ri} C_1 = (0, 1) \times \{0\}$ while $\operatorname{ri} C_2 = (0, 1) \times (0, 1)$. Thus $C_1 \subset C_2$, but $\operatorname{ri} C_1 \not\subset \operatorname{ri} C_2$.

(b) By translating C, we may assume that $0 \in C$, then $\operatorname{aff} C = Y$ is a linear space, and so $\operatorname{ri} C$ is the interior of C relative to Y, thus we may apply Exercise 2.4.3 using Y as the overspace to derive the conclusion.

(c) (i) $\Rightarrow$ (ii): Let $x \in \operatorname{ri} C$ then there exists $r > 0$ so that $x + \epsilon B_E \cap \operatorname{aff} C \subset C$. In particular, for $y \in C$ choosing $\epsilon > 0$ so that $\epsilon\|y - x\| < r$, we have $x + \epsilon(x - y) \in C$.

(ii) $\Rightarrow$ (iii): Let $Y := \{\lambda(c - x) : \lambda \geq 0, x \in C\}$. Certainly $0 \in Y$. Moreover, let $\lambda(c - x) \in Y$. Then $\alpha\lambda(c - x) \in Y$ if $\alpha \geq 0$. Suppose $\alpha < 0$, then choose $\epsilon > 0$ so that $\epsilon(x - c) + x \in C$. Then

$$\begin{aligned} \alpha\lambda(c - x) &= |\alpha|\lambda(x - c) = \frac{|\alpha|\lambda}{\epsilon}\epsilon(x - c) \\ &= \frac{|\alpha|\lambda}{\epsilon}[(\epsilon(x - c) + x) - x], \end{aligned}$$

and so Y is closed under scalar multiplication. We now show Y is closed under addition. Indeed, for the nontrivial case $\lambda_1, \lambda_2 > 0$ we have

$$\begin{aligned} \lambda_1(c_1 - x) + \lambda_2(c_2 - x) &= (\lambda_1 + \lambda_2)\left[\frac{\lambda_1}{\lambda_1 + \lambda_2}c_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2}c_2 - x\right] \\ &= (\lambda_1 + \lambda_2)(\bar{c} - x), \end{aligned}$$

where $\bar{c} \in C$ by the convexity of C as desired. Thus Y is a linear subspace.

(iii) $\Rightarrow$ (i) Let $Y = \{\lambda(c - x) : c \in C\}$. It follows from Lemma 2.4.5 that $\operatorname{aff} C = x + Y$. For a basis $\{y_i\}$ of Y, we can choose each of y_i and $-y_i$ can be written as $\lambda(c - x)$ for some $\lambda > 0$ and $c \in C$ from which it follows that x is in the interior of C relative to $\operatorname{aff} C$.

(d) In fact, $\operatorname{ri}(TC) = T(\operatorname{ri} C)$; see [369, Theorem 6.6]. $\square$

2.4.6. By shifting f we may assume $0 \in \operatorname{dom} f$, and let $Y = \operatorname{span}(\operatorname{dom} f)$. Let $x \in \operatorname{ri}(\operatorname{dom} f)$, the x is in the interior of the domain of f relatively to Y. By the max formula (2.1.19), $\partial f|_Y(x) \neq \emptyset$, that is, there exists $\phi \in Y$ such that

$$\langle \phi, y - x \rangle \leq f(y) - f(x) \quad \text{for all} \quad y \in Y. \tag{7}$$

Now write $E = Y + Z$ as a direct sum, and define $\tilde{\phi}$ on E by $\tilde{\phi}(y+z) = \phi(y)$. Because $\operatorname{dom} f \subset Y$, it follows from (7) that $\tilde{\phi} \in \partial f(x)$.

Notice that this result ensures the subdifferential of any proper convex function on E has nonempty domain and range. This is because the domain of a convex function is convex, and every nonempty convex subset of E has nonempty relative interior. $\square$

2.4.7. Suppose $x_0 \in \operatorname{dom} f$ and $\partial f(x_0) = \emptyset$. Because $\operatorname{cl}(\operatorname{ri} \operatorname{dom} f) = \operatorname{cl}(\operatorname{dom} f)$, there exists a sequence $(x_n) \subset \operatorname{ri}(\operatorname{dom} f)$ converging to x_0, and hence by Exercise 2.4.6 there exist $\phi_n \in \partial f(x_n)$. Suppose by way of contradiction that $\|\phi_n\| \not\to \infty$, hence it has a bounded subsequence, and then by compactness a convergent subsequence. Thus we suppose $(\phi_{n_j}) \to \phi$. Then for $y \in E$, we have

$$\phi(y) - \phi(x_0) = \lim_j \phi_{n_j}(y - x_{n_j}) \leq \liminf_j f(y) - f(x_{n_j}) \leq f(y) - f(x_0),$$

and so $\phi \in \partial f(x_0)$. This provides our desired contradiction. Thus $\|\phi_n\| \to \infty$. Furthermore, Exercise 2.2.23 shows that $\partial f(x_0)$ is unbounded whenever it is not empty and x_0 is in the boundary of the domain of f. Thus ∂f is not bounded on any neighborhood of a boundary point of the domain of f.

Further notes. Closedness is necessary as simple examples illustrate. Indeed, let $f(t) := 0$ if $t < 1$, $f(1) := 1$ and $f(t) := +\infty$ for $t > 1$. Then $\partial f(1) = \emptyset$ and $\partial f(t) = \{0\}$ for all $t < 1$. $\square$

2.4.8. Suppose f is proper. Fix $x_0 \in \operatorname{ri}(\operatorname{dom} f)$ and let $\phi \in \partial f(x_0)$. Then $f(x) \geq \operatorname{cl} f(x) \geq f(x_0) + \phi(x - x_0)$ for all $x \in X$ and so $\operatorname{cl} f$ is proper. One can verify $\operatorname{cl} f$ is convex because its epigraph is convex as the closure of a convex set; that a closure of a convex set is convex is elementary to verify. Indeed, suppose $D = \operatorname{cl} C$ where C is convex. Suppose $x, y \in D$, and $0 \leq \lambda \leq 1$. Choose $(x_n), (y_n) \subset C$ so that $x_n \to x$ and $y_n \to y$. Then

$$\lambda x + (1 - \lambda) y = \lim_n (\lambda x_n + (1 - \lambda) y_n).$$

Hence $\lambda x + (1 - \lambda) y \in D$ as it is a limit of elements from C.

Further notes. Hence, one can use the Hessian to check convexity of convex functions on closed domains. For example suppose $f : C \to \mathbb{R}$ is continuous and $\operatorname{int} C \neq \emptyset$. Suppose f is twice Gâteaux differentiable on $\operatorname{int} C$ whose Hessian is positive semidefinite at all $x \in \operatorname{int} C$. Then f is convex, because $f|_{\operatorname{int} C}$ is convex, and f is the closure of $f|_{\operatorname{int} C}$. $\square$

2.4.9. The set C is closed by Carathéodory's theorem (1.2.5) because it is the convex hull of a compact set. The set of extreme points of C is not closed because $(1, 0, 0)$ is not an extreme point of C but every other point on the circle $\{(x, y, z) : x^2 + y^2 = 1, z = 0\}$ is an extreme point of C. $\square$

2.4.10. (a) If (2.4.12) has a solution, then clearly (2.4.13) does not, so at most one of (2.4.12) and (2.4.13) has a solution. Let $C := \{x \in E : x = \sum_{i=0}^m \lambda_i x_i, \lambda_i \geq 0, \sum \lambda_i = 1\}$. Then C is a closed convex set. In the case $0 \in C$, then (2.4.12) has a solution. In the event $0 \notin C$, we apply the basic separation theorem (2.1.21) to find $x \in E$ so that $\sup_C x < \langle x, 0\rangle = 0$. In particular, $\langle x_i, x\rangle < 0$ for $i = 0, 1, \ldots, m$ and so x is a solution to (2.4.13).

(b) Clearly, (2.4.14) and (2.4.15) cannot simultaneously have solutions. Consider the cone

$$C := \{x : x = \sum_{i=1}^{m} \mu_i x_i,\ \mu_i \geq 0\}.$$

Then C is convex and it is a finitely generated cone which is thus closed by Carathéodory's theorem (1.2.5). In the event $c \in C$, then (2.4.14) has a solution. In the event $c \notin C$, we apply the basic separation theorem (2.1.21) to find $x \in E$ so that $\langle x, c\rangle > \sup_{u \in C}\langle x, u\rangle = 0$ (note $\sup_C x = 0$ because $0 \in C$, and if $\langle x, u\rangle > 0$ for some $u \in C$, then $nu \in C$ and so $\sup_C x > n\langle x, u\rangle > \langle x, c\rangle$). Therefore, (2.4.15) is satisfied by x because $\langle c, x\rangle > 0$ and $\langle x_i, x\rangle \leq \sup_C x = 0$ for $i = 1, \ldots, m$. $\square$

2.4.11. (a) Suppose $\{a_j\}_{j=1}^m$ is linearly dependent, and $x = \sum_{j=1}^m \mu_j a_j$ where $\mu_j \geq 0$ for $j = 1, 2, \ldots, m$. We will show that x can be written in this form using at most $m-1$ elements, from which the first statement in (a) will follow.

Using the linear dependence we can write

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m = 0, \quad \text{where } \sum_{j=1}^{m} \lambda_j \geq 0$$

and not all λ_j are 0. Now for any $t \in \mathbb{R}$ we have

$$x = \sum_{j=1}^{m} (\mu_j - t\lambda_j) a_j.$$

Let $J_+ := \{j : \lambda_j > 0\}$. Then $J_+ \neq \emptyset$ because $\sum \lambda_j \geq 0$ and not all $\lambda_j = 0$. Let j_0 denote an index in J_+ such that

$$\frac{\mu_{j_0}}{\lambda_{j_0}} = \min\{\mu_j/\lambda_j : j \in J_+\}.$$

Set $\bar{t} := \mu_{j_0}/\lambda_{j_0}$. Then $\bar{t} \geq 0$ and for $j \in J_+$ one has

$$(\mu_j - \bar{t}\lambda_j) = \lambda_j \left(\frac{\mu_j}{\lambda_j} - \frac{\mu_{j_0}}{\lambda_{j_0}}\right) \geq 0,$$

which equality when $j = j_0$. When $j \notin J_+$, then $\lambda_j \leq 0$ and so $\mu_j - \bar{t}\lambda_j \geq 0$. Therefore, $\mu_j - \bar{t}\lambda_j \geq 0$ for $j = 1, 2, \ldots, m$ with equality when $j = j_0$, and so we write

$$x = \sum_{1 \leq j \leq m, j \neq j_0} (\mu_j - \bar{t}\lambda_j) a_j$$

as desired.

For the second statement in (a), let $\{a_j : j \in J\}$ be a linearly independent set, and let $N := |J|$ and define the linear mapping $A : \mathbb{R}^N \to \mathbb{R}^N$ by $A(c_1, c_2, \ldots, c_N) = \sum_{i=1}^N c_i a_{j_i}$ where $J = \{j_1, j_2, \ldots, j_N\}$. Then A is an isomorphism and so it maps closed sets onto closed sets, in particular, $A(R_+^N) = C_J$ is closed.

(b) A finitely generated cone is thus closed as a union of finitely many closed sets.

(c) This can be proved along the same lines as (a), but with more care: see p. 41–42 in L.D. Berkowivitz, *Convexity and Optimization in* $\mathbb{R}^n$, Wiley, 2002. We will use the result of (a) to derive (c). Indeed, suppose $A \subset \mathbb{R}^m$, and suppose $x \in \operatorname{conv} A$. Then there exist $a_1, \dots, a_m$ in A such that $x = \sum_{i=1}^m \lambda_i a_i$ where $\lambda_i \geq 0$ and $\sum_{i=1}^m \lambda_i = 1$. Now consider the cone $C_I \subset \mathbb{R}^n \times \mathbb{R}$ generated by $\{(a_i, 1)\}_{i=1}^m$. Then $(x, 1) = \sum_{i=1}^m \lambda_i (a_i, 1) \in C_I$, and by part (a), $(x, 1) \in C_J$ where $\{a_j, 1\}_{j \in J}$ is linearly independent in $\mathbb{R}^n \times \mathbb{R}$, and so $|J| \leq n+1$. Now $(x, 1) = \sum_{j \in J} \mu_j (a_j, 1)$ where $\mu_j \geq 0$ for $j \in J$. Consequently, $x = \sum_{j \in J} \mu_j a_j$ and $1 = \sum_{j \in J} \mu_j$ which shows the desired result.

(d) Let A be a compact set in $\mathbb{R}^n$ and let f be a function from $\mathbb{R}^{(n+1)^2}$ to $\mathbb{R}^n$ defined by $f(y, x_1, x_2, \dots, x_{n+1}) = \sum_{i=1}^{n+1} y_i x_i$ where $y = (y_1, y_2, \dots, y_{n+1}) \in \mathbb{R}^{n+1}$ and $x_i \in \mathbb{R}^n$. Then f is a continuous function and $\operatorname{conv} A$ is the image of the compact set $\Delta \times A \times A \times \ldots \times A$ under the mapping f where Δ is the simplex in $\mathbb{R}^{n+1}$. $\square$

2.4.12. Let $\{x_1, x_2, \dots, x_{n+2}\} \subset \mathbb{R}^n$. The collection $\{x_i - x_1\}_{i=2}^{n+2}$ is linearly dependent in $\mathbb{R}^n$, hence we find $\{a_i\}_{i=2}^{n+2}$ not all 0 so that $\sum_{i=2}^{n+2} a_i (x_i - x_1) = 0$. Now set set $a_1 := -\sum_{i=2}^{n+2} a_i$. Then

$$\sum_{i=1}^{n+2} a_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^{n+2} a_i = 0. \tag{8}$$

Let $I_1 := \{i : a_i > 0\}$ and $I_2 := \{i : a_i \leq 0\}$, and let

$$C_1 := \operatorname{conv}\{x_i : i \in I_1\} \quad \text{and} \quad C_2 := \operatorname{conv}\{x_i : i \in I_2\}.$$

Let $a := \sum_{i \in I_1} a_i$ and let $\bar{x} := \sum_{i \in I_1} \frac{a_i}{a} x_i$. Then $\sum_{i \in I_1} \frac{a_i}{a} = 1$ and so $\bar{x} \in C_1$, and it follows from (8) that $\bar{x} = \sum_{i \in I_2} -\frac{a_i}{a} x_i$ and $\sum_{i \in I_2} -\frac{a_i}{a} = 1$. Consequently $\bar{x} \in C_2$ as well, and we are done. $\square$

2.4.13. We first establish the case when I is finite (in this case we need not assume the sets C_i are closed and bounded). The case $|I| \leq n+1$ is trivial, so we suppose $|I| = n+2$ and that the sets $C_1, C_2, \dots, C_{n+2}$ are such that that every subcollection of $n+1$ or fewer sets has nonempty intersection. For each $1 \leq i \leq n+2$, we fix $\bar{x}_i \in \bigcap_{j \in I, j \neq i} C_j$. In the case $\bar{x}_{j_1} = \bar{x}_{j_2}$ for some $j_1 \neq j_2$, then $\bar{x}_{j_1} \in \bigcap_{i \in I} C_i$ and we are done. So we suppose the $\bar{x}_i' s$ are all distinct. According to Radon's theorem (1.2.3) we can partition $I = I_1 \cup I_2$ so that $D_1 := \operatorname{conv}\{\bar{x}_i : i \in I_1\}$ and $D_2 := \operatorname{conv}\{\bar{x}_i : i \in I_2\}$ have nonempty intersection, say $\bar{x} \in D_1 \cap D_2$. Now $\bar{x} \in D_1$ ensures $\bar{x} \in \bigcap_{i \in I_2} C_i$ and $\bar{x} \in D_2$ ensures $\bar{x} \in \bigcap_{i \in I_1} C_i$ and so $\bar{x} \in \bigcap_{1 \leq i \leq n+2} C_i$ as desired.

Now suppose $|I| = k > n+2$, and the assertion is true whenever $|I| \leq k-1$ the argument in the previous paragraph shows every subcollection of $n+2$ sets on $\{C_i\}_{i \in I}$ will have nonempty intersection. Now consider the collection $D_1 := C_1 \cap C_2$ and $D_i := C_{i+1}$ for $i = 2, \dots, k$. Then $D_1, D_2, \dots, D_k - 1$ is a collection of closed convex sets such that every subcollection of $n+1$ or fewer sets has nonempty intersection. By the induction hypothesis, $\bigcap_{i=1}^{k-1} D_k$ has nonempty intersection, that is $\bigcap_{i \in I} C_i$ is not empty as desired. By mathematical induction, the result is true for every $|I| \in \mathbb{N}$.

Now suppose $\{C_i\}_{i \in I}$ is as in the statement of Helly's theorem. According to the previous paragraph, every finite subcollection has nonempty intersection. By the finite intersection property for compact sets, $\bigcap_{i \in I} C_i$ is not empty. $\square$

2.4.20. Let $m := \inf_C f$. Then $f \geq -g$ where $g := \delta_C - m$. The conditions imply we can apply the sandwich theorem (2.4.1) to find an affine function α such that

$$m - \delta_C \leq \alpha \leq f.$$

Then $m \le \inf_C \alpha \le \inf_C f$ as desired.

Suppose $\bar{x}$ minimizes f on C, and write the affine separating function as $\alpha = \phi + r$. Then $\phi \in \partial f(\bar{x})$ and $-\phi \in \partial \delta_C(\bar{x})$ and so $0 = \phi - \phi \in \partial f(\bar{x}) + N_C(\bar{x})$ as desired.

Conversely, suppose $0 \in \partial f(\bar{x}) + \delta_C(\bar{x})$. Then $0 \in \partial(f + \delta_C)(\bar{x})$, and so $f + \delta_C$ attains its minimum at $\bar{x}$. Therefore f attains its minimum on C at $\bar{x}$.

Further notes. This last part could have been completed equally easily using the subdifferential sum rule because

$$0 \in \partial(f + \delta_C)(\bar{x}) \text{ if and only if } 0 \in \partial f(\bar{x}) + \partial \delta_C(\bar{x}) \text{ if and only if } 0 \in \partial f(\bar{x}) + N_C(\bar{x}).$$

□

2.4.21. (a) Applying the Fenchel duality theorem (2.3.4) we obtain

$$\begin{aligned}
\inf_{x \in E} \{\delta_C(x) + \sigma_D(Ax)\} &= \sup_{\phi \in Y} \{-\delta_C^*(A^*\phi) - \delta_D(-\phi)\} \\
&= \sup_{\phi \in Y} \{-\sigma_C(A^*\phi) - \delta_D(-\phi)\} \\
&= \sup_{\phi \in Y} \{-\sup_{x \in C} \langle \phi, Ax \rangle - \delta_D(-\phi)\} \\
&= \sup_{\phi \in Y} \{\inf_{x \in C} \langle -\phi, Ax \rangle - \delta_D(-\phi)\} \\
&= \sup_{y \in D} \{\inf_{x \in C} \{\langle y, Ax \rangle\}.
\end{aligned}$$

Further, the supremum is attained when finite according to the Fenchel duality theorem (2.3.4), and so we have

$$\inf_{x \in C} \sup_{y \in D} \langle y, Ax \rangle = \max_{y \in D} \inf_{x \in C} \langle y, Ax \rangle. \tag{9}$$

(b) In case (i) when D is bounded, σ_D has full domain and AC is not empty and so (2.4.17) holds. In case (ii), when A is surjective and $0 \in \operatorname{int} C$, then $0 \in \operatorname{int} AC$ because A is open. Clearly $0 \in \operatorname{dom} \sigma_D$, and thus (2.4.17) holds in this case as well.

(c) When D is compact, (2.4.17) holds by case part (b)(i). The compactness of D and C then allow the replacing of sup and inf with max and min. □

2.4.23. Let K be a nonempty subset of E. Then

$$K^- := \{\phi \in E : \langle \phi, x \rangle \le 0, \text{ for all } x \in K\}$$

Given any $x \in K$, $t \ge 0$ and $\phi \in K^-$ we have

$$\langle \phi, tx \rangle = t\langle \phi, x \rangle \le 0.$$

Therefore, $K \subset (K^-)^-$. Also, observe that for any set S, S^- is a closed convex set because it is an intersection of closed half-spaces. In particular, K^{--} is a closed convex set containing $\mathbb{R}_+ K$. Now let C be the closed convex hull of $\mathbb{R}_+ K$, and suppose $x_0 \notin C$. By the basic separation theorem (2.1.21), we choose $\phi \in E$ such that $\phi(x_0) > \sup_C \phi$. Observe that $\sup_C \phi = 0$ because $0 \in C$, and if $\sup_C \phi > 0$, then there exists $t > 0$ and $\bar{x} \in K$ such that $\phi(t\bar{x}) > 0$. Then $\lim_{n \to \infty} \phi(n\bar{x}) = \infty$, and so ϕ would not be bounded above by $\phi(x_0)$ on C. Since $\sup_C \phi = 0$, then $\phi \in K^-$, and so $x_0 \notin (K^-)^-$, and thus $K^{--} \subset C$ as desired. □

2.4.24. We will prove the assertions for $D+C$, the other case follows by considering $D+(-C)$. Suppose $d_1+c_1, d_2+c_2 \in D+C$. Then for $0 \le \lambda \le 1$, using the convexity of D and C we obtain

$$\lambda(d_1+c_1)+(1-\lambda)(d_2+c_2) = \lambda d_1 + (1-\lambda)d_2 + \lambda c_1 + (1-\lambda)c_2 \in D+C.$$

Now suppose $(d_n+c_n)_{n=1}^\infty \subset D+C$ and $d_n+c_n \to \bar{x}$. Let (c_{n_k}) be a convergent subsequence of (c_n), say, $c_{n_k} \to \bar{c} \in C$. Then $d_{n_k} \to \bar{x}-\bar{c}$. Thus $\bar{x}-\bar{c} \in D$, and so $\bar{x} \in D+C$ as desired. □

Exercises from Section 2.5

2.5.1. Suppose f is differentiable at x_0 (by Theorem 2.2.1 f is automatically Fréchet differentiable), so given $\epsilon > 0$ we choose $\delta > 0$ so that

$$|f(x+h)-f(x)-\langle f'(x_0),h\rangle| \le \frac{\epsilon}{2}\|h\| \text{ whenever } 0 \le \|h\| < \delta.$$

Therefore, for $\|h\| < \delta$, using the triangle inequality we obtain

$$\begin{aligned} |f(x+h)+f(x-h)-2f(x)| &= |f(x+h)-f(x)-\langle f'(x),h\rangle + f(x-h)-f(x)-\langle f'(x),-h\rangle| \\ &\le \frac{\epsilon}{2}\|h\| + \frac{\epsilon}{2}\|h\| = \epsilon\|h\|. \end{aligned}$$

This implies $\displaystyle\lim_{\|h\|\to 0} \frac{f(x+h)+f(x-h)-2f(x)}{\|h\|} = 0.$

Conversely, suppose f is not differentiable at x. Because f is continuous at x, the max formula (2.1.19) ensures $\partial f(x) \ne \emptyset$, thus we use Theorem 2.2.1 to deduce that there are distinct $\phi, \Lambda \in \partial f(x)$. Thus we choose $h_0 \in S_E$ so that $(\phi-\Lambda)(h_0) > 0$. Now let $\epsilon = (\phi-\Lambda)$. By the subdifferential inequality

$$f(x+th_0)-f(x)-\phi(th_0)+f(x-th_0)-f(x)-\Lambda(-th_0) \ge 0, \quad \text{for all } t.$$

In particular, $f(x+th_0)+f(x-th_0)-2f(x) \ge (\phi-\Lambda)(th_0) \ge \epsilon t$ for all $t > 0$ and so

$$\limsup_{\|h\|\to 0} \frac{f(x+h)+f(x-h)-2f(x)}{\|h\|} \ge \epsilon$$

which establishes the 'if' assertion. □

2.5.2. Let

$$G_{n,m} := \left\{ x \in U : \sup_{\|h\| \le \frac{1}{m}} f(x+h)+f(x-h)-2f(x) < \frac{1}{nm} \right\},$$

and $O_n := \bigcup_{m\ge 1} G_{n,m}$ and $G := \bigcap_{n \ge 1} O_n$. Suppose $x \in G$ and let $\epsilon > 0$. Choose n such that $1/n < \epsilon$. Now find m such that $x \in G_{n,m}$, and choose δ so that $0 < \delta < 1/m$. The convexity of f implies

$$\sup_{\|h\|=\alpha} f(x+h)+f(x-h)-2f(x) < \frac{1}{n}\alpha \text{ whenever } 0 < \alpha \le \frac{1}{m}.$$

Indeed, for $0 < \lambda \le 1$ we have

$$\begin{aligned} f(x+\lambda h)+f(x-\lambda h)-2f(x) &\le \lambda f(x+h)+(1-\lambda)f(x)+\lambda f(x-h)+(1-\lambda)f(x)-2f(x) \\ &= \lambda[f(x+h)+f(x-h)-2f(x)]. \end{aligned}$$

Exercise 2.5.1 implies that f is differentiable at x. Conversely, suppose f is differentiable at x, and fix $n \in \mathbb{N}$. Choose $0 < \epsilon < 1/n$ and use Exercise 2.5.1 to find $\delta > 0$ so that

$$f(x+h) + f(x-h) - 2f(x) \leq \epsilon\|h\| \text{ whenever } \|h\| \leq \delta.$$

Then $x \in G_{n,m}$ for all $m > 1/\delta$. It follows that $x \in G$ as desired.

It remains to verify that O_n is open for each n. Indeed, fix n and suppose $x \in O_n$. Then for some $m_0 \in \mathbb{N}$, $x \in G_{nm}$ and, as above, the convexity of f implies that $x \in G_{nm}$ for all $m \geq m_0$. Now fix $m > m_0$ sufficiently large so that f has Lipschitz constant, say $K > 0$ on $B_{2/m}(x)$. Then choose $\epsilon > 0$ so that

$$\sup_{\|h\| \leq \frac{1}{m}} f(x+h) + f(x-h) - 2f(x) < \frac{1}{mn} - \epsilon.$$

Now we choose $\alpha > 0$ so that $4K\alpha < \epsilon$ and $\alpha < 1/m$. Now suppose $\|u - x\| < \alpha$, then because f has Lipschitz constant K on $B_{2/m}(x)$, for $\|h\| < 1/m$ we have

$$f(u+h) + f(u-h) - 2f(u) \leq f(x+h) + f(x-h) - 2f(x) + 4K\|x-u\| < \frac{1}{mn}.$$

This shows O_n is open as desired. $\square$

2.5.3. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz. Let p_j be the j-th coordinate projection from $\mathbb{R}^m$ to $\mathbb{R}$. Define $f_j = p_j \circ f$. Then f_j is locally Lipschitz. Let $D = \{x \in \mathbb{R}^n : f_j \text{ is differentiable at } x,\ j = 1, 2, \ldots, m\}$. It follows from Rademacher's theorem, that D^c is a union of finitely many null sets, and thus has measure 0. It remains to show that f is differentiable at each $x \in D$. Indeed, fix $x \in D$ and let A be the m by n matrix whose j-th row is $\nabla f_j(x)$. It is not hard to verify that $\nabla f(x) = A$, indeed for $\epsilon > 0$, choose $\delta > 0$ so that

$$\left| \frac{f_i(x+th) - f_i(x)}{t} - \langle \nabla f_i(x), h \rangle \right| < \frac{\epsilon}{\sqrt{m}} \quad \text{whenever} \quad h \in S_{R^n},\ 0 < |t| < \delta.$$

Then for $h \in S_{R^n}$ and $0 < |t| < \delta$, we have

$$\begin{aligned} \left\| \frac{f(x+th) - f(x)}{t} - Ah \right\| &= \sqrt{\sum_{i=1}^m \left(\frac{f_i(x+th) - f_i(x)}{t} - \langle \nabla f_i(x), h \rangle \right)^2} \\ &< \sqrt{\sum_{i=1}^m \frac{\epsilon^2}{m}} = \epsilon. \end{aligned}$$

Hence $\nabla f(x) = A$ as desired. $\square$

2.5.4. Let $\epsilon > 0$, and let $h \in S_X$. Let $K > 0$ be chosen so that f satisfies Lipschitz constant K in a neighborhood $\delta B_r(\bar{x})$ of $\bar{x}$ and $\|y\| \leq K$. Now fix $k \in \mathbb{N}$ with $\|h_k - h\| < \epsilon/4K$. Now choose $0 < \delta < r$ so that

$$|f(\bar{x} + th_k) - f(\bar{x}) - \langle y, th_k \rangle| < \frac{\epsilon}{4} t \text{ whenever } |t| < \delta. \tag{10}$$

Now for $|t| < \delta$ we have

$$\begin{aligned} |f(\bar{x} + th) - f(\bar{x}) - \langle y, th \rangle| &\leq |f(\bar{x} + th_k) - f(\bar{x}) - \langle y, th_k \rangle| + 2K\|th - th_k\| \\ &\leq \frac{\epsilon}{2}|t| + 2K|t|\frac{\epsilon}{4K} = \epsilon|t|. \end{aligned}$$

This shows f is Gâteaux differentiable at $\bar{x}$ with $\nabla f(\bar{x}) = y$. So far, we didn't use that X is finite-dimensional. However, because X is finite-dimensional, and f is Lipschitz in a neighborhood of $\bar{x}$, Exercise 2.2.9 implies f is Féchet-differentiable at $\bar{x}$ as desired.

Further comments. The reader will notice that last part of the proof of Rademacher's theorem also proves this fact. Observe further that even the assertion f is Gâteaux differentiable at $\bar{x}$ may fail for continuous functions (the estimate with the Lipschitz constant above was crucial). Indeed, on $\mathbb{R}^2$ let $f(x,y) = 0$ whenever $y \le \sqrt{|x|}$ and $f(x,y) = y - \sqrt{|x|}$ for $y \ge \sqrt{|x|}$. The hypothesis of the exercise are satisfied at $\bar{x} := (0,0)$ for every direction $h \in S_{\mathbb{R}^2}$ with $y := (0,0)$ except for the direction $h := (0,1)$, however, f fails to be Gâteaux differentiable at $(0,0)$. $\square$

2.5.5. Let x be a boundary point of C. Because C has nonempty interior, we choose $\phi \in S_{X^*}$ so that $\phi(x) = \sup_C \phi$ and $\phi(x) > \phi(y)$ whenever $y \in \operatorname{int} C$. It follows that $\phi \in \partial d_C(x)$. Indeed, if $y \in C$, then $\phi(y-x) \le 0 = d_C(y) - d_C(x)$. If $y \notin C$, then $\phi(y-x) \le \inf\{\|y-u\| : \phi(u) \le \phi(x)\} \le d_C(y) = d_C(y) - d_C(x)$. Because $0 \in d_C(x)$, it follows that $\partial d_C(x)$ is not a singleton, so d_C is not differentiable at x. Thus d_C is a convex function that is not differentiable at the boundary points of C, and consequently, the boundary of C is both first category and Lebesgue-null. $\square$

2.5.7. (a)$\Rightarrow$(b): g is almost everywhere differentiable by Theorem 2.5.1. On the other hand, the subgradient inequality holds on U. Together, these facts imply (b).

(b)$\Rightarrow$(c): trivial.

(c)$\Rightarrow$(a): Fix u, v in U, $u \ne v$, and $t \in (0,1)$. It is not hard to see that there exist sequences (u_n) in U, (v_n) in U, (t_n) in $(0,1)$ with $u_n \to u$, $v_n \to v$, $t_n \to t$, and $x_n := t_n u_n + (1-t_n)v_n \in A$, for every n. By assumption, $\nabla g(x_n)(u_n - x_n) \le g(u_n) - g(x_n)$ and $\nabla g(x_n)(v_n - x_n) \le g(v_n) - g(x_n)$. Equivalently, $(1-t_n)\nabla g(x_n)(v_n - u_n) \le g(u_n) - g(x_n)$ and $t_n \nabla g(x_n)(u_n - v_n) \le g(v_n) - g(x_n)$. Multiply the former inequality by t_n, the latter by $1 - t_n$ and adding we obtain

$$0 \le t_n g(u_n) - t_n(g(x_n) + (1-t_n)g(v_n) - (1-t_n)g(x_n),$$

or in other words $g(x_n) \le t_n g(u_n) + (1-t_n)g(v_n)$. Now let n tend to $+\infty$, and deduce that $g\big(tu + (1-t)v\big) \le tg(u) + (1-t)g(v)$. The convexity of g follows and the proof is complete. $\square$

2.5.8. Let $f_n(x) := n[f(x+1/n) - f(x)]$. Let $G := \{x : f'(x) \text{ exists}\}$; then $f_n(x) \to f'(x)$ for $x \in G$. Let $F_n := \{x : |f_j(x) - f_k(x)| \le 1/2\}$ for all $j, k \ge n$. Then F_n is closed because it is an intersection of closed sets since $f_j - f_k$ is continuous for each $k, j \in \mathbb{N}$. Clearly $(f_n(x))$ is convergent for $x \in G$ and so $\bigcup F_n \supset G$. Suppose F_{n_0} contains an open interval I for some $n_0 \in \mathbb{N}$. Then $|f_{n_0}(x) - f'(x)| \le 1/2$ almost everywhere on I. By the Fundamental theorem of calculus, $f'(x) = \chi_S - \chi_{\mathbb{R}\setminus S}$ almost everywhere, and so $f'(x) = 1$ on a dense subset of I and $f'(x) = -1$ on another dense subset of I we conclude that $f_{n_0} \ge 1/2$ on a dense subset of I, and $f_{n_0} \le -1/2$ on a dense subset of I to contradict the continuity of f_{n_0}. Hence F_n is nowhere dense for each $n \in \mathbb{N}$ and so G is a set of first category. $\square$

Exercises from Section 2.6

2.6.1. Using the bilinear property of ϕ, and then the symmetric property we compute

$$\begin{aligned}
\phi(x+y, x+y) &= \phi(x, x+y) + \phi(y, x+y) \\
&= \phi(x,x) + \phi(x,y) + \phi(y,x) + \phi(y,y) \\
&= \phi(x,x) + 2\phi(x,y) + \phi(y,y).
\end{aligned}$$

Therefore,

$$\phi(x, y) = \frac{1}{2}[\phi(x + y, x + y) - \phi(x, x) - \phi(y, y)].$$

That is, ϕ is uniquely determined by the values $\phi(h, h)$ such that $h \in E$. $\square$

2.6.3. (a) This part is essentially a restatement of the definitions. Indeed, assume that

$$\Delta_t^2 f(x) : h \mapsto \frac{f(x + th) - f(x) - t\langle \nabla f(x), h\rangle}{\frac{1}{2}t^2},$$

converges uniformly on bounded sets to the function $h \mapsto \langle Ah, h\rangle$ as $t \to 0$ for some matrix A. Then given $\epsilon > 0$, choose $\delta > 0$ so that

$$\left|\frac{f(x + ty) - f(x) - t\langle \nabla f(x), y\rangle}{\frac{1}{2}t^2} - \langle Ay, y\rangle\right| < \epsilon \;\text{ when } 0 < |t| < \delta, y \in S_X.$$

Now letting $h := ty$ where $y \in S_X$ and $0 < |t| < \delta$ in the preceding, implies

$$f(x + h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle Ah, h\rangle + o(\|h\|^2),\ \|h\| \to 0$$

as desired. The converse implication follows essentially by reversing the preceding steps. Indeed, suppose f has a strong second-order Taylor expansion at x. Given $\epsilon > 0$, choose $\delta > 0$ so that

$$\left|f(x + h) - \left(f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle Ah, h\rangle\right)\right| \le \frac{1}{2}\epsilon\|h\|^2, \;\text{ when } \|h\| < \delta.$$

Then given any $r > 0$ for $y \in rB_X$ and $|t| < \delta/r$, we have

$$\left|f(x + ty) - \left(f(x) + t\langle \nabla f(x), y\rangle + \frac{1}{2}t^2\langle Ay, y\rangle\right)\right| \le \frac{1}{2}\epsilon|t|^2.$$

Dividing both sides by $\frac{1}{2}t^2$ when $t \neq 0$, we obtain

$$\left|\frac{f(x + ty) - f(x) - t\langle \nabla f(x), y\rangle}{\frac{1}{2}t^2} - \langle Ay, y\rangle\right| \le \epsilon.$$

Thus $\Delta_t^2 f(x) \to \langle Ah, h\rangle$ uniformly on bounded sets.
The argument for pointwise convergence is analogous.

(b) As in the proof of Theorem 2.6.1, let A be a symmetric matrix. By part (a), $q_t := \frac{1}{2}\Delta_t^2 f(x)$ converges pointwise to $q(h) := \frac{1}{2}\langle Ah, h\rangle$. According to Proposition 2.6.3 , the functions q_t are closed and convex. Because they converge pointwise to q, it follows that q is convex. It follows from Exercise 2.1.22 that the convergence is uniform on bounded sets. $\square$

2.6.4. (a) The definition of generalized Fréchet derivative implies $\phi_n \to \phi$ whenever $\phi_n \in \partial f(x_n)$ and $x_n \to x$. Then Corollary 2.5.3 implies $\partial f(x) = \{\phi\}$. Thus f is Fréchet differentiable at x with $\nabla f(x) = \phi$ (Theorem 2.2.1).

(b) Suppose by way of contradiction that the definition of the generalized second-order Gâteaux derivative at x works with distinct matrices A and B. Then we choose $h \in S_X$ such that $Ah \neq Bh$. Thus we set $\epsilon := \|(B-A)h\|$. Then choose $\delta > 0$ so that

$$\partial f(x+th) \subset \phi + A(th) + \frac{\epsilon}{3}|t|B_E \quad \text{and} \quad \partial f(x+th) \subset \phi + B(th) + \frac{\epsilon}{3}|t|B_E \quad \text{for } |t| < \delta.$$

Now for fixed $0 < t < \delta$, we let $\Lambda \in \partial f(x+th)$ and write

$$\Lambda = \phi + A(th) + y_1 \quad \text{and} \quad \Lambda = \phi + B(th) + y_2 \quad \text{where } \|y_1\|,\ \|y_2\| \le \frac{\epsilon t}{3}.$$

Then $\|(A-B)(th)\| \le \frac{2\epsilon}{3} t$ which contradicts $\epsilon := \|(B-A)h\|$.

(c) Suppose f has a generalized second-order Fréchet derivative at x. Then given $\epsilon > 0$, there exists $\delta > 0$ so that

$$\partial f(x+h) \subset \nabla f(x) + Ah + \epsilon\delta B_E \text{ whenever } 0 < \|h\| \le \delta.$$

Therefore, if $|t| < \delta$, $h \in B_E$ and $\phi_t \in \partial f(x+th)$ we have

$$\|\phi_t - \nabla f(x) - A(th)\| \le \epsilon|t|, \text{ and so } \left\|\frac{\phi_t - \nabla f(x)}{t} - Ah\right\| \le \epsilon.$$

Therefore, $\lim_{t\to 0} \frac{\phi_t - \nabla f(x)}{t} = Ah$ uniformly for $h \in B_E$ as desired.

Conversely, suppose $\lim_{t\to 0} \frac{\phi_t - \nabla f(x)}{t} = Ah$ uniformly for h in B_E. Given $\epsilon > 0$, there exists $\delta > 0$ such that

$$\left\|\frac{\phi_t - \nabla f(x)}{t} - Ah\right\| < \epsilon \text{ whenever } 0 < |t| \le \delta, \|h\| = 1, \phi_t \in \partial f(x+th).$$

Thus $\|\phi_t - \nabla f(x) - A(th)\| < \epsilon|t|$, or in other words,

$$\|\phi - \nabla f(x) - Ah\| < \epsilon\|h\| \quad \text{whenever } \phi \in \partial f(x+h),\ 0 < \|h\| \le \delta.$$

Because $\epsilon > 0$ was arbitrary, this implies

$$\partial f(x+h) \subset \nabla f(x) + Ah + o(\|h\|)B_E$$

as desired. □

2.6.5. Suppose that for some matrix $A : E \to E$, given any $\epsilon > 0$ and bounded set $W \subset E$ there exists $\delta > 0$ we so that

$$\Delta_t[\partial f](x)(h) - Ah \subset \epsilon B_E \quad \text{for all } h \in W, \quad t \in (0,\delta).$$

Applying this with $W = S_E$ and arbitrary $\epsilon > 0$ we find $\delta > 0$ so that

$$\frac{\partial f(x+th) - \nabla f(x)}{t} - Ah \subset \epsilon B_E \quad \text{for all } h \in B_E, \quad t \in (0,\delta).$$

In particular, when $0 < \|u\| < \delta$ we write $u = th$ where $t = \|u\|$ and $h \in S_E$ and mutiplying both sides of the previous inclusion by t we obtain

$$\partial f(x+u) \subset \nabla f(x) + Au + \epsilon\delta B_E$$

and thus f has a generalized second-order Fréchet derivative at x.

Conversely suppose f has a generalized second-order derivative at x. Let $W \subset E$ be bounded and let $\epsilon > 0$. Choose $K > 0$ so that $W \subset KB_E$. Using the definition of generalized derivative, choose $\eta > 0$ so that

$$\partial f(x+h) \subset \nabla f(x) + Ah + \frac{\epsilon}{K}\|h\|B_E \quad \text{whenever} \quad 0 < \|h\| < \eta.$$

Now set $\delta := \eta/K$ and suppose $0 < t < \delta$ and $h \in W$. Then $\|th\| < \eta$ and using the previous inclusion we obtain

$$\partial f(x+th) - \nabla f(x) \subset A(th) + \frac{\epsilon}{K}\|th\|B_E;$$

and then dividing both sides by t and noting $\|h\| \le K$ we obtain

$$\Delta_t[\partial f](x)(h) - Ah \subset \epsilon B_E \quad \text{for all } h \in W,\ t \in (0,\delta),$$

as desired. $\square$

2.6.6. The subgradient inequality ensures Δ_t^2 is nonnegative. Hence the convexity, closedness and properness of Δ_t^2 thus follows because f possesses those properties. We next verify $\partial\left[\frac{1}{2}\Delta_t^2 f(x)\right] = \Delta_t[\partial f](x)$. Indeed, suppose $y \in \partial\left[\frac{1}{2}\Delta_t^2 f(x)\right](h)$, then for $u \in E$,

$$\begin{aligned}\langle y,u\rangle &\le \frac{f(x+t(h+u)) - f(x) - t\langle\nabla f(x), h+u\rangle - [f(x+th) - f(x) - t\langle\nabla f(x),h\rangle]}{t^2}\\ &= \frac{f(x+t(h+u)) - f(x+th) - \langle\nabla f(x), tu\rangle}{t^2}\end{aligned}$$

Multiplying both sides of the previous inequality by t (note: $t > 0$), we obtain

$$\langle y, tu\rangle + \frac{1}{t}\nabla f(x), tu\rangle \le \frac{f(x+t(h+u)) - f(x+th)}{t}.$$

Thus $y + \frac{1}{t}\nabla f(x) \in \frac{1}{t}\partial f(x+th)$, that is $y \in \frac{\partial f(x+th) - \nabla f(x)}{t}$. Therefore,

$$\partial\left[\frac{1}{2}\Delta_t^2 f(x)\right] \subset \Delta_t[\partial f](x).$$

The reverse inclusion follows by roughly tracing the steps backwards. $\square$

2.6.8. First, $f'(0) = \lim\limits_{h\to 0}\dfrac{h^3\cos(1/h) - 0}{h} = 0$ and $f'(t) = 3t^2\cos(1/t) + t\sin(1/t)$ when $t \ne 0$ and so f is continuously differentiable. Moreover,

$$t^3\cos(1/t) = 0 + 0t + \frac{1}{2}0t^2 + o(t^2) = f(0) + f'(0)t + \frac{1}{2}0t^2 + o(t^2)$$

and so f has a second-order Taylor expansion at 0. However, f'' does not exist at 0 since

$$\lim_{h\to 0}\frac{f'(h) - f'(0)}{h} = \lim_{h\to 0}\frac{3h^2\cos(1/h) + h\sin(1/h) - 0}{h}$$

does not exist. $\square$

Exercises from Section 2.7

2.7.1. Note that given a nonempty set S, $\mathrm{conv}(S)$ is the collection of convex combinations of elements in S, that is element of the form $\sum_{i=1}^m \lambda_i s_i$ where $m \in \mathbb{N}$, $s_i \in S$, $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for all $1 \leq i \leq m$.

By the extreme value theorem, f attains its maximum at some $\bar{x} \in C$. By Minkowski's theorem (2.7.2), we may write $\bar{x} = \sum_{i=1}^m \lambda_i x_i$ where x_i is an extreme point of C, $\lambda_i \geq 0$ for $1 \leq i \leq m$ and $\sum_{i=1}^m \lambda_i = 1$. By the convexity of f, $f(\bar{x}) \leq \sum_{i=1}^m \lambda_i f(x_i)$. Because f attains it maximum at $\bar{x}$, this implies $f(x_i) = f(\bar{x})$ for each $1 \leq i \leq n$.

2.7.2.(Exposed Points)

(a) Suppose x_0 is an exposed point of a convex set C. Choose $\phi \in E$ such that $\langle \phi, x_0 \rangle = \sup_C \phi$ and $\langle \phi, x \rangle < \langle \phi, x_0 \rangle$ for all $x \in C \setminus \{x_0\}$. Let $x, y \in C \setminus \{x_0\}$. Then for any $0 \leq \lambda \leq 1$,

$$\phi(\lambda x + (1-\lambda)y) = \lambda\phi(x) + (1-\lambda)\phi(y) < \phi(x_0).$$

Thus x_0 is not a convex combination of x and y.

(b) Let C be a compact convex subset of E. Suppose $x_0 \in C$ is an exposed point, exposed by $\phi \in E$. Now suppose $(x_n) \subset C$ is a sequence such that $\phi(x_n) \to \phi(x_0)$, but $x_n \not\to x_0$. By the compactness of C, there is a convergent subsequence of (x_n) such that $(x_{n_k}) \to \bar{x}$ where $\bar{x} \neq x_0$ and $\bar{x} \in C$. Then $\phi(\bar{x}) = \lim \phi(x_n) = \phi(x_0)$. This contradicts the fact ϕ exposes x_0 in C.

(c) A proof that the exposed points are dense in the extreme points of a compact convex (or any closed convex set in E) can be found in [369, Theorem 18.6, p. 167–68], the proof uses the basic separation theorem (2.1.21) and Carathéodory's theorem (1.2.5). We will outline another proof that every compact convex subset of E is the closed convex hull of its strongly exposed points that mimics techniques that will be used in Section 6.6.

Indeed, let C be a compact convex subset of E. Then $\sigma_C : E \to \mathbb{R}$ is a continuous convex function. By Theorem 2.5.1, σ_C is differentiable on a dense subset of E. Now let D be the closed convex hull of the exposed points of C. Suppose by way of contradiction that $D \neq C$, that is, we fix $\bar{x} \in C \setminus D$. According to the basic separation theorem (2.1.21), we choose $y \in E$ such that $\langle y, \bar{x} \rangle > \sup_D y$. Because D is bounded, and the points of differentiability of σ_C are dense in E, we may choose $\phi \in E$ such that σ_C is differentiable at ϕ, and $\phi(\bar{x}) > \sup_D \phi$.

Now let $x_0 = \nabla\sigma_C(\phi)$ and so $\partial\sigma_C(\phi) = \{x_0\}$. It is easy to check that $\phi(x_0) = \sigma_C(\phi)$. Indeed, $\langle x_0, 2\phi - \phi \rangle \leq \sigma_C(2\phi) - \sigma_C(\phi) = \sigma_C(\phi)$ and $\langle x_0, \phi - 0 \rangle \geq \sigma_C(\phi) - \sigma_C(0) = -\sigma_C(\phi)$. Thus $\phi(x_0) = \sigma_C(\phi)$. Further, $x_0 \in C$, for otherwise we would use the basic separation theorem (2.1.21) to find $\Lambda \in E$ so that $\Lambda(x_0) > \sigma_C(\Lambda)$. This provides the immediate contradiction $\langle x, \Lambda - \phi \rangle > \sigma_C(\Lambda) - \sigma_C(\phi)$. Finally, if $u \in C$ is such that $\phi(u) = \sigma_C(\phi)$, then

$$\langle u, \Lambda - \sigma \rangle \leq \sigma_C(\Lambda) - \sigma_C(\phi) \quad \text{for all } \Lambda \in E$$

and so $u \in \partial\sigma_C(\phi) = \{x_0\}$. Thus ϕ attains its supremum on C uniquely at x_0, and this yields the contradiction that x_0 is an exposed point of C which is not in D.

Further notes. The set $C := \{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, -\sqrt{1-x^2} \leq y \leq 1 + \sqrt{1-x^2}\}$ is a compact convex set that is not the convex hull of its exposed points. Indeed, any exposed point (x, y) of C satisfies $|x| < 1$. Thus the closure is needed in (c). $\square$