Notes on Convex Analysis and Nonlinear Optimization by Borwein and Lewis

Forest Yang

May 30 2019

0.1 4.1 Continuity of Convex Functions

For a real $L \geq 0$, we say that a function $f : \mathbb{E} \to (-\infty, +\infty]$ is *Lipschitz (with constant L)* on a subset C of $\operatorname{dom} f$ if $|f(x) - f(y)| \leq L\|x - y\|$ for any $x, y \in C$. If f is Lipschitz on a neighborhood of a point z then we say f is *locally Lipschitz around z*. If $F : \mathbb{E} \to \mathbb{Y}$ then replace $|f(x) - f(y)|$ with $\|F(x) - F(y)\|$.

Theorem 0.1 (4.1.1 (Local boundedness)). *Let $f : \mathbb{E} \to (-\infty, +\infty]$ be a convex function. Then f is locally Lipschitz around a point z in its domain if and only if it is bounded above on a neighborhood of z.*

Lemma 0.2 (4.1.2). *Let Δ be the **simplex** $\{x \in \mathbb{R}^n_+ \mid \sum x_i \leq 1\}$. If the function $g : \Delta \to \mathbb{R}$ is convex then it is continuous on $\operatorname{int} \Delta$.*

Theorem 0.3 (4.1.3 (Convexity and continuity)). *Let $f : \mathbb{E} \to (-\infty, +\infty]$ be a convex function. Then f is continuous (in fact locally Lipschitz) on the interior of its domain.*

The *gauge function* $\gamma_C : \mathbb{E} \to (-\infty, +\infty]$ associated with a nonempty set $C \subset \mathbb{E}$ is defined as $\gamma_C(x) = \inf\{\lambda \in \mathbb{R}_+ \mid x \in \lambda C\}$ and is sublinear when C is convex.

Theorem 0.4 (4.1.4 (Core and interior)). *The core and interior of any convex set in $\mathbb{E}$ are identical and convex.*

The conjugate of the gauge function γ_C is the indicator function of a set $C^\circ \subset \mathbb{E}$ defined by

$$C^\circ = \{\phi \in \mathbb{E} \mid \langle \phi, x \rangle \leq 1 \text{ for all } x \in C\}.$$

We call C° the *polar set* for C. It is a closed convex set containing 0.

Theorem 0.5 (4.1.5 (Bipolar set). *The bipolar set of any subset C of $\mathbb{E}$ is given by*

$$C^{\circ\circ} = \operatorname{cl}(\operatorname{conv}(C \cup \{0\})).$$

Theorem 0.6 (4.1.6 (Supporting hyperplane)). *Suppose that the convex set $C \subset \mathbb{E}$ has nonempty interior and that the point $\bar{x}$ lies on the boundary of C. Then there is a **supporting hyperplane** to C at $\bar{x}$: there is a nonzero element a of $\mathbb{E}$ satisfying $\langle a, x \rangle \geq \langle a, \bar{x} \rangle$ for all points x in C.*

An *extreme point* of a convex set $C \subset \mathbb{E}$ is a point x in C whose complement $C \setminus \{x\}$ is convex. We denote the set of extreme points by $\operatorname{ext} C$.

Lemma 0.7 (4.1.7)**.** *Given a supporting hyperplane H of a convex set $C \subset \mathbb{E}$, any extreme point of $C \cap H$ is also an extreme point of C.*

Define the *dimension* of a set $C \subset \mathbb{E}$, $\dim C$, as the dimension of $\operatorname{Span}(C - x)$ for any point $x \in C$.

Theorem 0.8 (4.1.8 (Minkowski))**.** *Any compact convex set $C \subset \mathbb{E}$ is the convex hull of its extreme points.*

Proof that if C is compact and convex, then $\operatorname{conv}(\operatorname{bd} C) = C$:

Proof. $\operatorname{conv}(\operatorname{bd} C) \subset C$ because $\operatorname{bd} C \subset C$ because C is closed and $\operatorname{conv} C = C$ because C is convex.
To prove $C \subset \operatorname{conv}(\operatorname{bd} C)$: take $x \in \operatorname{int} C$, and shift C by $-x$ so x becomes 0. For every $i \in [n]$ and $\operatorname{sgn} \in \{+, -\}$, there exists $c_{i,\text{sgn}} > 0$ such that $c_{i,\text{sgn}} e_i \in \operatorname{bd} C$ where $\{e_1, \ldots, e_n\}$ is the standard basis. We can express 0 as a convex combination of $c_{i,+}e_i$ and $-c_{i,-}e_i$ for any i. □

0.2 Exercises for 4.1

1 * (Points of continuity) Suppose the function $f : \mathbb{E} \to (-\infty, +\infty]$ is convex.

(a) Use the Local boundedness theorem (0.1) to prove that f is continuous and finite at x if and only if it minorizes a function $g : \mathbb{E} \to (-\infty, +\infty]$ which is continuous and finite at x.
If f is continuous and finite at x, then an obvious function that is continuous and finite at x and minorized by f is itself.
If f minorizes a function that is continuous and finite at x, then f is bounded above on a neighborhood of x, which by Theorem 4.1.1 (0.1) implies that f is continuous at x.

(b) Suppose f is continuous at some point $y \in \operatorname{dom} f$. Use part (a) to prove directly that f is continuous at any point z in $\operatorname{core}(\operatorname{dom} f)$. (Hint: Pick a point $u \in \operatorname{dom} f$ such that $z = \delta y + (1-\delta)u$ for some real $\delta \in (0, 1)$; now observe that the function

$$x \in \mathbb{E} \mapsto \delta^{-1}(f(\delta x + (1-\delta)u) - (1-\delta)f(u))$$

minorizes f.)
Since $z \in \operatorname{core}(\operatorname{dom} f)$, $u = z + (z-y)\epsilon \in \operatorname{dom} f$ for some $\epsilon > 0$. We have $z = \delta y + (1-\delta)u$ for $\delta = \frac{\epsilon}{1+\epsilon} \in (0, 1)$. Now,

$$f(\delta x + (1-\delta)u) \le \delta f(x) + (1-\delta)f(u)$$
$$\implies g(x) := \frac{f(\delta x + (1-\delta)u) - (1-\delta)f(u)}{\delta} \le f(x).$$

Since f is finite and continuous at y, by part (a), g, being minorized by f, is finite and continuous at y. Because the product and addition of functions continuous at a point is continuous at that point, $h(x) := \delta g(x) + \frac{1-\delta}{\delta} f(u) = f(\delta x + (1-\delta)u)$ is continuous at y. Notice that $h(\frac{x-(1-\delta)}{\delta}) = f(x)$. Since h is continuous at y and $\frac{\cdot-(1-\delta)}{\delta}$ is continuous everywhere, in particular at z where it evaluates to y, f is continuous at z.
This is a really interesting technique; I think one can interpret g as a "flattening" of f centered at u, as $g(u) = f(u)$. To help see the "flattening," notice that $g'(x) = f'(\delta x + (1-\delta)u)$. That is, the derivative of g is the derivative of f, pulled towards u, which by convexity is smaller than the derivative of f. Intuitively, this makes g seem like a flattened version of f.

(c) Prove that f is continuous at a point x in $\operatorname{dom} f$ if and only if

$$(x, f(x) + \epsilon) \in \operatorname{int}(\operatorname{epi} f)$$

for some (all) real $\epsilon > 0$.
If f is continuous at x, then there exists $\delta > 0$ such that $\|y - x\| \leq \delta \implies f(y) \leq f(x) + \frac{\epsilon}{2}$. Thus, $(x, f(x)+\epsilon) + \min\{\delta, \frac{\epsilon}{2}\}B \subset \operatorname{epi} f$. Therefore, $(x, f(x)+\epsilon) \in \operatorname{int}(\operatorname{epi} f)$.
Now suppose $(x, f(x) + \epsilon) \in \operatorname{int}(\operatorname{epi} f)$. Then there exists $\delta > 0$ such that $(x, f(x) + \epsilon) + \delta B \subset \operatorname{epi} f$, i.e. for every $(y, r) \in (x, f(x) + \epsilon) + \delta B$, we have $f(y) \leq r$. Note $r \leq f(x) + \epsilon + \delta$. Thus, for every y where $\|x - y\| \leq \delta$, we have $f(y) \leq f(x) + \epsilon + \delta$. In other words, f is bounded on a neighborhood of x, which by Theorem 4.1.1 (0.1) implies f is continuous at x.

(d) Assuming $0 \in \operatorname{cont} f$, prove f^* has bounded level sets. Deduce that the function $X \in \mathbb{S}^n \mapsto \langle C, X\rangle + \operatorname{ld}(X)$ has compact level sets for any matrix $C \in \mathbb{S}^n_{++}$.
If $0 \in \operatorname{cont} f$, then given $\epsilon > 0$, there exists $\delta > 0$ such that for all x with $\|x\| \leq \delta$, we have $f(x) \leq f(0) + \epsilon$. Now if $\phi \neq 0$,

$$f^*(\phi) = \sup_{x \in \mathbb{E}} \langle \phi, x\rangle - f(x) \geq \left\langle \phi, \frac{\delta\phi}{\|\phi\|}\right\rangle - f\left(\frac{\delta\phi}{\|\phi\|}\right) \geq \delta\|\phi\| - f(0) - \epsilon.$$

Since the RHS goes to infinity as $\|\phi\|$ does, f^* has bounded level sets.
Now recall that $\operatorname{ld}(X) = -\log\det(X)$ if $X \in \mathbb{S}^n_{++}$ and $+\infty$ otherwise. Furthermore, $\operatorname{ld}^{**}(X) = -n + \operatorname{ld}(-X)$. Furthermore, for any f, $(\langle c, x\rangle + f(x))^*(\phi) = f^*(\phi - c)$. Thus, $(\langle C, X\rangle + \operatorname{ld}(X))^*(D) = -n + \operatorname{ld}(C - D)$. Since for a constant $c \in \mathbb{R}$, $(f + c)^* = f^* - c$, and $(f(ax))^*(\phi) = f^*(\frac{\phi}{a})$ for $a \neq 0$, and $(f(x+b))^*(\phi) = f^*(\phi) - \langle \phi, b\rangle$, we have

$$\begin{aligned}(-n + \operatorname{ld}(C - \cdot))^*(X) &= n + (\operatorname{ld}(C - \cdot))^*(X)\\ &= n + (\operatorname{ld}(-\cdot))^*(X) + \langle C, X\rangle = \langle C, X\rangle + \operatorname{ld}(X).\end{aligned}$$

Since $-n + \operatorname{ld}(C - \cdot)$ is continuous at 0 (as ld is continuous at any PD $C \in \mathbb{S}^n_{++}$), the level sets of $\langle C, X\rangle + \operatorname{ld}(X)$ are bounded and thus compact.

(e) Assuming $x \in \operatorname{cont} f$, prove $\partial f(x)$ is a nonempty compact convex set.
By Section 3.1, Exercise 3, $\partial f(x)$ is closed and convex. By Theorem 3.1.8 (Max

formula), since $\operatorname{cont} f \subset \operatorname{int}(\operatorname{dom} f) \subset \operatorname{core}(\operatorname{dom} f)$ (in fact we have equalities by Theorems 4.1.3 and 4.1.4), $\partial f(x)$ is nonempty. $x \in \operatorname{cont} f$ implies f is locally Lipschitz around x with some Lipschitz constant L. We see that by Proposition 3.1.6 (Subgradients and directional derivatives), for any $\phi \in \partial f(x)$ and $d \in \mathbb{E}$,

$$\langle \phi, d \rangle \le f'(x; d) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} \le L\|d\|.$$

This implies that $\|\phi\| \le L$, i.e. $\partial f(x)$ is compact.

2 (Equivalent norms). A *norm* is a sublinear function $|||\cdot||| : \to \mathbb{R}_+$ that satisfies $|||x||| = |||-x||| > 0$ for all nonzero points x in $\mathbb{E}$. By considering the function $|||\cdot|||$ on the standard unit ball B, prove any norm $|||\cdot|||$ is *equivalent* to the Euclidean norm $\|\cdot\|$: that is, there are constants $K \ge k > 0$ with $k\|x\| \le |||x||| \le K\|x\|$ for all x.

Proof. Since $|||\cdot|||$ is finite and convex, by Theorem 4.1.1 (0.1), it is locally Lipschitz at 0 with some constant L. This implies that $|||x||| \le L\|x\|$ for all x in a neighborhood of 0, which by scaling implies the inequality for all $x \in \mathbb{E}$.
Now define $k = \inf_{x:\|x\|=1} |||x|||$. By Theorem 4.1.3, $|||\cdot|||$ is continuous, and therefore the inf is achieved and positive. For all $x \in \mathbb{E}$ where $\|x\| = 1$, we have $\frac{|||x|||}{k} \ge 1 = \|x\|$. By scaling (more precisely, positive homogeneity), $\frac{|||x|||}{k} \ge \|x\|$ holds for all $x \in \mathbb{E}$. Thus, for all $x \in \mathbb{E}$, $k\|x\| \le |||x||| \le L\|x\|$. $\square$

3. (Examples of polars) Calculate the polars of the following sets:

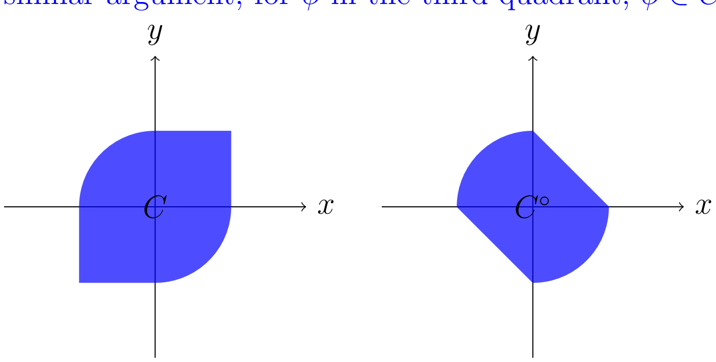
(a) $C = \operatorname{conv}(B \cup \{(1,1), (-1,-1)\}) \subset \mathbb{R}^2$.
If ϕ is in the second or fourth quadrant, $x \in B$, and $\lambda_i \ge 0$ ($i \in [3]$), $\sum_{i=1}^{3} \lambda_i = 1$,

$$\langle \phi_1, \lambda_1(1,1) + \lambda_2(-1,-1) + \lambda_3 x \rangle \le (\lambda_1 + \lambda_2)|\phi_1 - \phi_2| + \lambda_3 \langle \phi, x \rangle \le \|\phi\|.$$

Furthermore, there exists $x \in B \subset C$ such that $\langle \phi, x \rangle = \|\phi\|$. Thus, for ϕ in the second or fourth quadrant, $\phi \in C^\circ$ iff $\|\phi\| \le 1$. For ϕ in the first quadrant, $\lambda \in [0,1]$, and $x \in B$,

$$\langle \phi, \lambda(1,1) + x \rangle = \lambda\|\phi\|_1 + (1-\lambda)\langle \phi, x \rangle \le \lambda\|\phi\|_1 + (1-\lambda)\|\phi\| \le \|\phi\|_1,$$

with equality when $\lambda = 1$. Thus, for ϕ in the first quadrant, $\phi \in C^\circ$ iff $\|\phi\| = 1$. By a similar argument, for ϕ in the third quadrant, $\phi \in C^\circ$ iff $\|\phi\| = 1$.



(b) $C = \left\{ (x,y) \in \mathbb{R}^2 \,\middle|\, y \geq b + \frac{x^2}{2} \right\}$ $(b \in \mathbb{R})$.

Notice that $C = \operatorname{epi} f$ where $f = b + \frac{x^2}{2}$. We have $\bar{x} = \nabla f(\bar{x})$. If $c < 0$, then $(d,c) = -c(-d/c, -1) = -c(\nabla f(-d/c), -1) \in N_{\operatorname{epi} f}(-d/c, f(-d/c))$ (see Section 3.2, Exercise 11). Thus, for all $(x,r) \in \operatorname{epi} f$,

$$d(x + \frac{d}{c}) + c(r - f\left(-\frac{d}{c}\right)) \leq 0.$$

In other words,

$$\max_{(x,r)\in \operatorname{epi} f} (d,c)^\top (x,r) = -\frac{d^2}{c} + cf\left(-\frac{d}{c}\right) = c\left(b - \frac{d^2}{2c^2}\right).$$

Thus, when $c < 0$, we have

$$(d,c) \in C^\circ \iff c\left(b - \frac{d^2}{2c^2}\right) \leq 1 \iff d^2 \leq 2c(bc-1).$$

When $c > 0$, then we can take (x,r) with $r \to \infty$ to get that $(d,c) \notin C^\circ$, as $(d,c)^\top(x,r) > 1$. When $c = 0$, if $d \neq 0$, then we can take $x \to \infty$ with a suitable choice of r or likewise with $x \to -\infty$ to make $(d,c)^\top(x,r) = dx \to +\infty$, so that $(d,0) \notin C^\circ$. On the other hand, if $(d,c) = 0$, then $(d,c)^\top(x,r) = 0 \leq 1$ for all (x,r), so $0 \in C^\circ$. Thus,

$$C^\circ = \{0\} \cup \{(d,c) : c < 0,\, d^2 \leq 2c(bc-1)\} = \{(d,c) : c \leq 0,\, d^2 \leq 2c(bc-1)\}.$$

4 (Polar sets and cones). Suppose the set $C \subset \mathbb{E}$ is closed, convex, and contains 0. Prove the convex cones in $\mathbb{E} \times \mathbb{R}$

$$\operatorname{cl} \mathbb{R}_+(C \times \{1\}) \text{ and } \operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\})$$

are mutually polar.

Proof. Notice the above two sets are cones, which means we are verifying they are each others' polar cones. As $S := \operatorname{cl} \mathbb{R}_+(C \times \{1\})$ is a closed convex cone, by Theorem 3.1.8 (Bipolar cone), the bipolar equals itself, so we just need to verify that $T := \operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\})$ equals $S^\circ = S^-$.

For any set A, it turns out $A^\circ = (\operatorname{cl} A)^\circ$. Clearly $(\operatorname{cl} A)^\circ \subset A^\circ$. Now take $\phi \in A^\circ$. If $\tilde{a} \in \operatorname{cl} A$, then there is a sequence $a^i \to \tilde{a}$ in A, so $\langle \phi, a^i \rangle \leq 1$ for all $i \in \mathbb{N}$. Taking the limit, $\langle \phi, \tilde{a} \rangle \leq 1$, and since $\tilde{a}$ was arbitrary $\phi \in (\operatorname{cl} A)^\circ$. So $A^\circ \subset (\operatorname{cl} A)^\circ$, ergo $A^\circ = (\operatorname{cl} A)^\circ$.

Actually, here is a proof that polar cones of A, B with $\operatorname{cl}\operatorname{conv} \mathbb{R}_+ A = \operatorname{cl}\operatorname{conv} \mathbb{R}_+ B$ equal. $A^{--} = \operatorname{cl}\operatorname{conv} \mathbb{R}_+ A$, and since A^- is a closed convex cone, $A^- = A^{---} = (\operatorname{cl}\operatorname{conv} \mathbb{R}_+ A)^-$.

Thus, denoting an element of $\mathbb{R}_+(C \times \{1\})$ as $c(x,1)$, where $c \geq 0$ and $x \in C$,

$$(y,r) \in S^\circ \iff \forall c(x,1) \in \mathbb{R}_+(C \times \{1\}),\ c(\langle x,y \rangle + r) \leq 0 \iff \forall x \in C,\ \langle x,y \rangle + r \leq 0.$$

If $r \neq 0$, we get for all $x \in C$, $-r(\langle x, -y/r\rangle - 1) \leq 0$. If $r > 0$, then we get $\forall x \in C$, $\langle x, -y/r\rangle \geq 1$. But this gives a contradiction, since we can take $x = 0$ (Kind of amazing that this is the only place where the assumption $0 \in C$ is used). If $r < 0$, then we have $\forall x \in C$, $\langle x, -y/r\rangle \leq 1$, i.e. $-y/r \in C^\circ$. Moreover, if $-y/r \in C^\circ$, $\forall x \in C$, $\langle x, y\rangle + r = -r(\langle x, -y/r\rangle - 1) \leq 0$. Thus,

$$r \neq 0 : \ (y, r) \in X^\circ \iff r < 0, \ -y/r \in C^\circ \iff (y, r) \in \mathbb{R}_{++}(C^\circ \times \{-1\}).$$

On the other hand, if $r = 0$, then we get $\forall x \in C$, $\langle x, y\rangle \leq 0$, i.e. $y \in C^-$. Thus,

$$S^\circ = \mathbb{R}_{++}(C^\circ \times \{-1\}) \cup (C^-, 0).$$

Furthermore, by the inclusions

$$\operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\}) \subset \mathbb{R}_{++}(C^\circ \times \{-1\}) \cup (C^-, 0) \subset \operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\}),$$

we have $S^\circ = T$ as desired. To verify the left inclusion, note that $\mathbb{R}_+(C^\circ \times \{-1\}) = \mathbb{R}_{++}(C^\circ \times \{-1\}) \cup \{0\}$, and $0 \in (C^-, 0)$. Furthermore, S° is closed, which gives the left inclusion. For the right inclusion, note that $\mathbb{R}_+ C^- = C^- \subset C^\circ$. Therefore, for $x^- \in C^-$, we can take $c^\circ(\frac{x^-}{c^\circ}, -1) = (x^-, c^\circ) \in \mathbb{R}_+(C^\circ \times \{-1\})$ and take $c^\circ \to 0$ to obtain $(x^-, 0) \in (C^-, 0)$. Shorter proof, more opaque but maybe more elegant (wait actually after writing it I think it's basically a more concise version of the above proof): By the statement $\forall x \in C$, $\langle x, x^\circ\rangle - 1 \leq 0$ iff $x^\circ \in C^\circ$, we obtain

$$C^\circ \times \{-1\} = \{(y, -1) : (y, -1) \in (C \times \{1\})^-\}.$$

Note that no element of the form $(y, 1)$ exists in $(C \times \{1\})^-$, since $(0, 1) \in C \times \{1\}$. Furthermore, for every element of the form $(y, 0) \in (C \times \{1\})^-$, we can use convex conity of $(C \times \{1\})^-$ on $(y, 0)$ and $(0, -1)$ and the fact that $\mathbb{R}_+ y \subset C^- \subset C^\circ$ to obtain a sequence in $\mathbb{R}_+(C^\circ \times \{-1\})$ converging to $(y, 0)$. Therefore, $(C \times \{1\})^- = \operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\})$. By an earlier comment, $(\operatorname{cl} \mathbb{R}_+(C \times \{1\}))^\circ = (C \times \{1\})^- = \operatorname{cl} \mathbb{R}_+(C^\circ \times \{-1\})$, and $\operatorname{cl} \mathbb{R}_+(C \times \{1\})$ is self-bipolar, which finishes the proof. $\square$

5 * (Polar sets). Suppose C is a nonempty subset of $\mathbb{E}$.

(a) Prove $\gamma_C^* = \delta_{C^\circ}$.
Let $y \notin C^\circ$. Thus, there exists $x \in C$ such that $\langle y, x\rangle > 1$. Now, for all $\lambda \in \mathbb{R}_+$,

$$(\gamma_C)^*(y) = \sup_{x' \in \mathbb{E}} \langle y, x'\rangle - \gamma_C(x') \geq \langle y, \lambda x\rangle - \gamma_C(\lambda x) = \lambda(\langle y, x\rangle - \gamma_C(x)).$$

By taking $\lambda \to +\infty$, we see $(\gamma_C)^*(y) = +\infty$. Now suppose $y \in C^\circ$ and take $x \in \mathbb{E}$. If $\nexists \lambda \in \mathbb{R}_+$ such that $x \in \lambda C$, then $\langle y, x\rangle - \gamma_C(x) = -\infty$. Otherwise, let $\bar{\lambda} = \gamma_C(x)$. If $x = \bar{\lambda} x_C$ for some $x_C \in C$,

$$\langle y, \bar{\lambda} x_C\rangle - \gamma_C(\bar{\lambda} x_C) = \bar{\lambda}(\langle y, x_C\rangle - \gamma_C(x_C)) \leq 0,$$

because $\langle y, x_C\rangle \leq 1$, and if $\gamma_C(x_C) < 1$, then for some $\lambda < 1$, $x_C \in \lambda C \implies x \in \bar{\lambda}\lambda C$, contradicting minimality of $\bar{\lambda}$. Now if $x \neq \bar{\lambda} x_C$ for any $x_C \in C$, there is still a sequence

λ^i, x_C^i where $\lambda^i > \bar{\lambda}$, $\lambda^i \to \bar{\lambda}$, and $x = \lambda^i x_C^i$. Now $\gamma_C(x_C^i) \geq \frac{\bar{\lambda}}{\lambda^i}$, because if $\gamma_C(x_C^i) < \frac{\bar{\lambda}}{\lambda^i}$, then $x = \lambda^i x_C^i = \lambda x_C'$ for some $\lambda < \bar{\lambda}$ and $x_C' \in C$, contradicting minimality of $\bar{\lambda}$. Thus,

$$\langle y, x\rangle - \gamma_C(x) = \lambda_i(\langle y, x^i\rangle - \gamma_C(x^i)) \leq \lambda_i(1 - \frac{\bar{\lambda}}{\lambda^i}).$$

If $\bar{\lambda} > 0$, then since $\lambda^i \to \bar{\lambda}$, the RHS goes to 0. If $\bar{\lambda} = 0$, then the RHS still goes to 0 because of the factor $\lambda_i \to 0$. Thus, $\langle y, x\rangle - \gamma_C(x) \leq 0$. Since x was arbitrary, we have shown

$$y \in C^\circ \implies (\gamma_C)^*(y) = 0.$$

This completes the proof that $(\gamma_C)^* = \delta_{C^\circ}$. God this question was annoying.

(b) Prove C° is a closed convex set containing 0.
C° is closed because it is an intersection of closed sets (specifically, halfspaces):

$$C^\circ = \bigcap_{x \in \mathbb{E}} \{\phi \in \mathbb{E} : \langle \phi, x\rangle \leq 1\}.$$

Finally, by the above, it is also an intersection of convex sets containing 0.

(c) Prove $C \subset C^{\circ\circ}$.
Take $\phi \in C^\circ$ and $x \in C$. By definition, $\langle \phi, x\rangle \leq 1$. But, ϕ was arbitrary, so in fact $x \in C^{\circ\circ}$.

(d) If C is a cone, prove $C^\circ = C^-$.
If $\langle \phi, x\rangle > 0$ for some $x \in C$, by scaling we have $\langle \phi, cx\rangle > 1$ for some $cx \in C$. Thus, if $\phi \in C^\circ$, $\phi \in C^-$, so $C^\circ \subset C^-$. The reverse inclusion is clear.

(e) For a subset D of $\mathbb{E}$, prove $C \subset D$ implies $D^\circ \subset C^\circ$.
Suppose $\phi \in D^\circ$. Then, if $x \in C$, since $x \in D$, we have $\langle \phi, x\rangle \leq 1$. Since $x \in C$ was arbitrary, $\phi \in C^\circ$.

(f) Prove C is bounded if and only if $0 \in \operatorname{int} C^\circ$.
If C is bounded, then there exists $M > 0$ such that $\|x\| \leq M$ for all $x \in C$. Then, if $\|\phi\| \leq M^{-1}$, we have $\langle \phi, x\rangle \leq \|\phi\|\|x\| \leq 1$ for all $x \in C$, i.e. $\phi \in C^\circ$. Thus, $M^{-1}B \in C^\circ$, i.e. $0 \in \operatorname{int} C^\circ$.
Now if C is unbounded, there exists a sequence x^i with unbounded norm. For any $\epsilon > 0$, $\sup_{\phi \in \epsilon B} \langle \phi, x^i\rangle = \epsilon\|x^i\| \to +\infty$. Therefore, $0 \notin \operatorname{int} C^\circ$.

(g) For any closed halfspace $H \subset \mathbb{E}$ containing 0, prove $H^{\circ\circ} = H$.
H takes the form $\{x : \langle a, x\rangle \leq b\}$ for some $b \geq 0$. If $\phi \neq 0$ is not a multiple of a, then it has a component perpendicular to a, v. The point $cv \in H$ for $c > 0$ large enough gives $\langle \phi, cv\rangle > 1$, meaning $\phi \notin H^\circ$. Now if $\phi = ca$ for some $c \geq 0$, for any $x \in H$ we have $\langle ca, x\rangle \leq cb$, with equality when $x = \frac{ba}{\|a\|^2}$. If $c < 0$ then we can take $x = -la \in H$, $l \to +\infty$ to get $\langle ca, -la\rangle > 1$. Thus, $C^\circ = \{ca : 0 \leq c \leq b^{-1}\}$ If $b = 0$, then replace b^{-1} with $+\infty$.
Now take $x \in \mathbb{E}$.

$$\langle x, ca\rangle = c\langle x, a\rangle \leq 1 \; \forall 0 \leq c \leq b^{-1} \iff \langle x, a\rangle \leq b.$$

Thus, $H^{\circ\circ} = H$.

(h) Prove Theorem 4.1.5 (Bipolar set).
By parts (b) and (c), $C^{\circ\circ}$ is closed, convex, and contains 0, and $C \subset C^{\circ\circ}$. Therefore, $\mathrm{cl}(\mathrm{conv}(C \cup \{0\})) \subset C^{\circ\circ}$.
Now take $y \notin \mathrm{cl}(\mathrm{conv}(C \cup \{0\}))$. By Theorem 2.1.6 (Basic separation), there exist $a \in \mathbb{E}$ and $b \in \mathbb{R}$ such that

$$\forall x \in \mathrm{cl}(\mathrm{conv}(C \cup \{0\})),\ \langle a, x\rangle \leq b < \langle a, y\rangle .$$

Now if $b < 0$, then we get a contradiction since we can take $x = 0$. Thus, $b \geq 0$. If $b = 0$, then we can replace b with $2^{-1}(b + \langle a, y\rangle)$, allowing us to assume $b > 0$. Then,

$$\forall x \in \mathrm{cl}(\mathrm{conv}(C \cup \{0\})),\ \left\langle \frac{a}{b}, x\right\rangle \leq 1 < \left\langle \frac{a}{b}, y\right\rangle .$$

Since $C \subset \mathrm{cl}(\mathrm{conv}(C \cup \{0\}))$, we have $\frac{a}{b} \in C^{\circ}$. Yet, $\left\langle \frac{a}{b}, y\right\rangle > 1$. Thus, $y \notin C^{\circ\circ}$. This completes the proof that $C^{\circ\circ} = \mathrm{cl}(\mathrm{conv}(C \cup \{0\}))$.

6 * (Polar sets and strict separation). Fix a nonempty set $C \in \mathbb{E}$.

(a) For points $x \in \mathrm{int}\, C$ and $\phi \in C^{\circ}$, prove $\langle \phi, x\rangle \leq 1$.
Since $x \in \mathrm{int}\, C$, for some $\epsilon > 0$, $x + \epsilon B \subset C$. Thus, if $\phi = 0$, the inequality is obviously true, and otherwise,

$$\left\langle x + \epsilon\frac{\phi}{\|\phi\|}, \phi\right\rangle \leq 1 \implies \langle x, \phi\rangle \leq 1 - \epsilon\|\phi\| < 1.$$

(b) Assume further that C is a convex set. Prove γ_C is sublinear.
First we prove that $\gamma(\mu x) = \mu\gamma(x)$ for $\mu \in \mathbb{R}_+$. If $\mu > 0$, then for any $\lambda \in \mathbb{R}_+$,

$$x \in \lambda C \iff \mu x \in \lambda\mu C \implies \gamma_C(\mu x) = \mu\gamma_C(x).$$

If $\mu = 0$, then $\gamma(\mu x) = \gamma(0) = 0$, since $0 \in 0C$. Note that as we just showed, positive homogeneity does not require convexity of C.
Now we show subadditivity, which goes on to imply sublinearity. For x_1 and x_2 in $\mathbb{E}$, suppose $x_1 \in \lambda_1 C$ and $x_2 \in \lambda_2 C$ for $\lambda_1, \lambda_2 \geq 0$. If $\lambda_1 = \lambda_2 = 0$, then $x_1 = x_2 = 0$ (since $0C = \{0\}$), so $x_1 + x_2 = 0 \in (\lambda_1 + \lambda_2)C = 0C$. Otherwise,

$$x_1 + x_2 \in \lambda_1 C + \lambda_2 C = (\lambda_1 + \lambda_2)\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}C + \frac{\lambda_2}{\lambda_1 + \lambda_2}C\right) \subset (\lambda_1 + \lambda_2)C,$$

by convexity of C. Therefore, if $\gamma_C(x_1) < r_1$ and $\gamma_C(x_2) < r_2$, then for some $0 \leq \lambda_1 < r_1$, $0 \leq \lambda_2 < r_2$, $x_1 \in \lambda_1 C$ and $x_2 \in \lambda_2 C$, so $x_1 + x_2 \in (\lambda_1 + \lambda_2)C$. Thus, $\gamma_C(x_1 + x_2) < r_1 + r_2$. So, taking an infimum over r_1 and then r_2,

$$\begin{gathered}\forall r_2 > \gamma_C(x_2),\ \gamma_C(x_1 + x_2) \leq \inf\{r_1 > \gamma_C(x_1)\} + r_2 \\ \implies \gamma_C(x_1 + x_2) \leq \inf\{r_1 > \gamma_C(x_1)\} + \inf\{r_2 > \gamma_C(x_2)\} = \gamma_C(x_1) + \gamma_C(x_2).\end{gathered}$$

Surprisingly hard to justify that if $x \leq r_1 + r_2$ for all $r_1 \in R_1$ and $r_2 \in R_2$, then $x \leq \inf R_1 + \inf R_2$ XD

(c) Assume in addition $0 \in \operatorname{core} C$. Deduce

$$\operatorname{cl} C = \{x \mid \gamma_C(x) \leq 1\}.$$

Since $0 \in \operatorname{core} C$, γ_C is everywhere finite and therefore continuous, by Theorem 4.1.3 (Convexity and continuity). Suppose $x \in \operatorname{cl} C$. Then, there exists a sequence $x^i \to x$ in C. By continuity, $\gamma_C(x^i) \to \gamma_C(x)$. Since $\gamma_C(x^i) \leq 1$ for all $i \in \mathbb{N}$, $\gamma_C(x) \leq 1$.
Now suppose $\gamma_C(x) \leq 1$. Then, there is a decreasing sequence $\mu^i \to 1$ such that $x \in \mu^i C$ for every $i \in \mathbb{N}$, i.e. $\frac{x}{\mu^i} \in C$. Since $\mu^i \to 1$, we have $\frac{x}{\mu^i} \to x$, which proves $x \in \operatorname{cl} C$.

(d) Finally, suppose in addition that $D \subset \mathbb{E}$ is a convex set disjoint from the interior of C. By considering the Fenchel program $\inf\{\delta_D + \gamma_C\}$, prove there is a closed halfspace containing D but disjoint from the interior of C.
Note that $\inf\{\delta_D + \gamma_C\} \geq 1$, since if $x \in D$, $\gamma_C(x) \geq 1$ (the proof of Theorem 4.1.4 (Core and interior) shows that $\operatorname{int} C = \{x \mid \gamma_C(x) < 1\}$, which is disjoint from D by assumption).
Since $\operatorname{dom} \gamma_C = \mathbb{E}$, the primal and dual Fenchel problems have equal value, with achievement for the dual (see Theorem 3.3.5, Fenchel duality):

$$\begin{aligned} 1 \leq \inf_{x \in \mathbb{E}} \{\gamma_C(x) + \delta_D(x)\} &= \sup_{\phi \in \mathbb{E}} \{-\gamma_C^*(\phi) - \delta_D^*(-\phi)\} \\ &= \sup_{\phi \in \mathbb{E}} \{-\delta_{C^\circ}(\phi) - \sup_{x \in D} \langle -\phi, x \rangle\} \\ &= \sup_{\phi \in C^\circ} \inf_{x \in D} \langle \phi, x \rangle . \end{aligned}$$

In the second line we used $\gamma_C^* = \delta_{C^\circ}$, from part (a) of Exercise 5. The sup is achieved, so there exists $\phi \in C^\circ$ such that $\forall x \in D$, $\langle \phi, x \rangle \geq 1$. But since $\phi \in C^\circ$, for any $x \in \operatorname{int} C$, we have $\langle \phi, x \rangle < 1$. We have found our closed halfspace:

$$H = \{x \in \mathbb{E} : \langle \phi, x \rangle \geq 1\} \text{ satisfies } D \subset H \text{ and } D \cap \operatorname{int} C = \emptyset.$$

Wow Fenchel duality is op.

7 * (Polar calculus [23]). Suppose C and D are subsets of $\mathbb{E}$.

(a) Prove $(C \cup D)^\circ = C^\circ \cap D^\circ$.

$$\begin{aligned} \phi \in (C \cup D)^\circ &\iff \forall x \in C,\ \langle \phi, x \rangle \leq 1 \text{ and } \forall x \in D,\ \langle \phi, x \rangle \leq 1 \\ &\iff \phi \in C^\circ \cap D^\circ. \end{aligned}$$

(b) If C and D are convex, prove

$$\operatorname{conv}(C \cup D) = \bigcup_{\lambda \in [0,1]} (\lambda C + (1-\lambda) D).$$

The $\supset$ inclusion holds almost by definition: for any $\lambda \in [0,1]$ and $c \in C$, $d \in D$, $\lambda c + (1-\lambda)d \in \operatorname{conv}(C \cup D)$ because $c \in \operatorname{conv}(C \cup D)$ and $d \in \operatorname{conv}(C \cup D)$ and $\operatorname{conv}(C \cup D)$ is convex.
For $\subset$, we show that $\bigcup_{\lambda \in [0,1]}(\lambda C + (1-\lambda)D)$ is convex. Clearly, it contains $C \cup D$. Since $\operatorname{conv}(C \cup D)$ is the smallest convex set containing $C \cup D$, the inclusion follows. To see convexity, first note the sets $[0,1](C \times \{1\})$ and $[0,1](D \times \{1\})$ are convex: for any nonnegative interval $[a,b]$ and convex set $\bar{C}$, the set $[a,b]\bar{C}$ is convex, since $a \le \alpha, \beta \le b$ and $x_1, x_2 \in \bar{C}$, $\lambda \in [0,1]$ implies

$$\lambda\alpha x_1 + (1-\lambda)\beta x_2 = (\lambda\alpha + (1-\lambda)\beta)\left(\frac{\lambda\alpha}{\lambda\alpha + (1-\lambda)\beta}x_1 + \frac{(1-\lambda)\beta}{\lambda\alpha + (1-\lambda)\beta}x_2\right) \in [a,b]\bar{C}.$$

If $\lambda\alpha + (1-\lambda)\beta = 0$, then $\lambda\alpha = (1-\lambda)\beta = 0$ and either α or $\beta = 0$, so $\lambda\alpha x_1 + (1-\lambda)\beta x_2 = 0 \in [a,b]\bar{C} = [0,b]\bar{C}$.
Now, notice that $g := \delta_{[0,1](C\times\{1\})} \odot \delta_{[0,1](D\times\{1\})}$ is convex (3.3 Exercise 12(a)) and

$$\begin{aligned} g(y,1) &= \inf_{(x,\lambda)\in\mathbb{E}\times\mathbb{R}} \delta_{[0,1](C\times\{1\})}(x,\lambda) + \delta_{[0,1](D\times\{1\})}(y-x, 1-\lambda) \\ &= \delta_{\bigcup_{\lambda\in[0,1]}(\lambda C + (1-\lambda)D)}(y). \end{aligned}$$

In other words, $\bigcup_{\lambda\in[0,1]} \lambda C + (1-\lambda)D$ is the $\mathbb{E}$ part of $\operatorname{dom} g$, which is convex. Thus, $\bigcup_{\lambda\in[0,1]} \lambda C + (1-\lambda)D$ is convex.

(c) If C is a convex cone and the convex set D contains 0, prove

$$C + D \subset \operatorname{cl}\operatorname{conv}(C \cup D).$$

Consider $c+d \in C+D$. Take a sequence $\lambda^i \to 0$. We have $\lambda^i \frac{c}{\lambda^i} + (1-\lambda^i)d \to c+d$. Notice the LHS is in $\bigcup_{\lambda\in[0,1]}(\lambda C + (1-\lambda)D) = \operatorname{conv}(C \cup D)$. Therefore, $c+d \in \operatorname{cl}(\operatorname{conv}(C \cup D))$.

Now suppose the closed convex sets K and H of $\mathbb{E}$ both contain 0.

(d) Prove $(K \cap H)^\circ = \operatorname{cl}\operatorname{conv}(K^\circ \cup H^\circ)$.
By part (a), $(K^\circ \cup H^\circ)^\circ = K^{\circ\circ} \cap D^{\circ\circ} = K \cap H$. The second equality is by Theorem 4.1.5 (Bipolar set) (0.5) Since K and H already are closed, convex, and contain 0, they are self bipolar. Now we take another polar and apply Theorem 4.1.5 again:

$$(K^\circ \cup H^\circ)^{\circ\circ} = \operatorname{cl}\operatorname{conv}(K^\circ \cup H^\circ) = (K \cap H)^\circ.$$

(e) If furthermore K is a cone, prove $(K \cap H)^\circ = \operatorname{cl}(K^\circ + H^\circ)$.
Note $\operatorname{conv}(K^\circ \cup H^\circ) \subset K^\circ + H^\circ$. This is because by part (b), an element of $\operatorname{conv}(K^\circ \cup H^\circ)$ is $\lambda k + (1-\lambda)h$ for some $\lambda \in [0,1]$, $k \in K^\circ$, $h \in H^\circ$, which is in $K^\circ + H^\circ$ because both sets contain 0. Thus, $\operatorname{cl}\operatorname{conv}(K^\circ \cup H^\circ) \subset \operatorname{cl}(K^\circ + H^\circ)$. Part (c) shows the reverse inclusion. Thus, by the previous part,

$$(K \cap H)^\circ = \operatorname{cl}(\operatorname{conv}(K^\circ \cup H^\circ)) = \operatorname{cl}(K^\circ + H^\circ).$$

8 ** (Polar calculus [23]). Suppose P is a cone in $\mathbb{E}$ and C is a nonempty subset of a Euclidean space $\mathbb{Y}$.

(a) Prove $(P \times C)^\circ = P^\circ \times C^\circ$.
$P^\circ \times C^\circ \subset (P \times C)^\circ$, since if $\langle p^\circ, p\rangle \le 1$ for all $p \in P$ and $\langle c^\circ, c\rangle \le 1$ for all $c \in C$, then since P is a cone (cones are nonempty, see page 1 of the textbook), $\langle p^\circ, p\rangle \le 0$ for all $p \in P$, and

$$\forall p, c \in P \times C,\ \langle (p^\circ, c^\circ), (p, c)\rangle = \langle p^\circ, p\rangle + \langle c^\circ, c\rangle \le 1.$$

$(P \times C)^\circ \subset P^\circ \times C^\circ$ because if $(a, b) \in (P \times C)^\circ$, then since $0 \in P$,

$$\forall c \in C,\ \langle (a, b), (0, c)\rangle = \langle b, c\rangle \le 1,$$

so $b \in C^\circ$. Furthermore,

$$\forall p \in P,\ \langle (a, b), (p, c)\rangle = \langle a, p\rangle + \langle b, c\rangle \le 1.$$

If P contains a nonzero element p, then, we must have $\langle a, p\rangle \le 0$. Thus, $(a, b) \in P^\circ \times C^\circ$.

(b) If furthermore C is compact and convex (possibly not containing 0), and K is a cone in $\mathbb{E} \times \mathbb{Y}$, prove

$$(K \cap (P \times C))^\circ = (K \cap (P \times C^{\circ\circ}))^\circ.$$

$K \cap (P \times C) \subset K \cap (P \times C^{\circ\circ})$, so $(K \cap (P \times C^{\circ\circ}))^\circ \subset (K \cap (P \times C))^\circ$ (See Exercise 5 (e)).
Now take $(\phi, \psi) \in (K \cap (P \times C))^\circ$. Consider an element $(a, b) \in K \cap (P \times C^{\circ\circ})$. By compactness of C and Section 2.2 Exercise 5 (d) which says that the convex hull of a compact set is compact,

$$C^{\circ\circ} = \mathrm{cl}(\mathrm{conv}(C \cup \{0\})) = \mathrm{conv}(C \cup \{0\}) = \bigcup_{\lambda \in [0,1]} \lambda C.$$

The last equality follows from Exercise 7, part (b). Therefore, $(a, b) = (p, \lambda c)$ for some $\lambda \in [0, 1]$ and $c \in C$. If $\lambda > 0$, then $(\frac{p}{\lambda}, c) \in K \cap (P \times C)$. Thus,

$$\langle (\phi, \psi), (p, \lambda c)\rangle = \lambda \left\langle (\phi, \psi), (\frac{p}{\lambda}, c)\right\rangle \le \lambda \le 1.$$

If $\lambda = 0$, we can take a sequence $\lambda^i \to \lambda$ such that $\langle (\phi, \psi), (p, \lambda^i c)\rangle \le 1$, which proves that $\langle (\phi, \psi), (p, 0)\rangle \le 1$. Therefore, $(\phi, \psi) \in (K \cap (P \times C^{\circ\circ}))^\circ$.

(c) If furthermore K and P are closed and convex, use Exercise 7 to prove

$$(K \cap (P \times C))^\circ = \mathrm{cl}(K^\circ + (P^\circ \times C^\circ)).$$

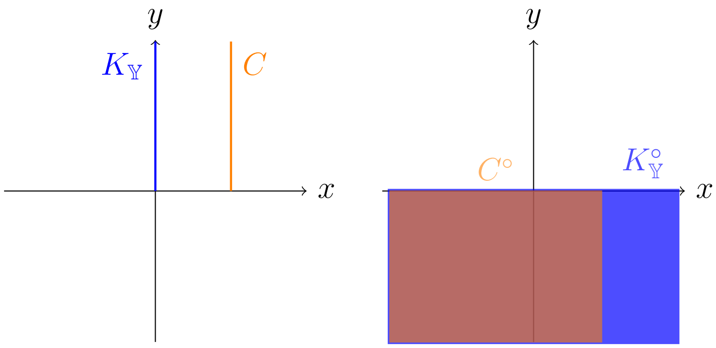
By Exercise 7(e), since $(P \times C^{\circ\circ})$ is now closed and convex and contains 0, and K is a closed convex cone,

$$(K \cap (P \times C))^\circ = (K \cap (P \times C^{\circ\circ})) = \mathrm{cl}(K^\circ + (P \times C^{\circ\circ})^\circ) = \mathrm{cl}(K^\circ + (P^\circ \times C^\circ)),$$

using parts (a) and (b).

(d) Find a counterexample to part (c) when C is unbounded.

Let $P = \{0\}, C = \{(1, r) : r \geq 0\}$, and $K = \{0\} \times \{(0, r) : r \geq 0\}$. P and K are closed, convex cones, and C is a closed, convex set. Since basically only the $\mathbb{Y}$ space is relevant here, let's plot C and $K_{\mathbb{Y}}$, the projection of K onto its $\mathbb{Y}$ component, as well as their polars.



As shown, $C^\circ = \{(1, y) : y \geq 0\}^\circ = \{(x, y) : x \leq 1,\ y \leq 0\}$. If $y \geq 0$ and $x \leq 1$, $y' \leq 0$, then $\langle (1, y), (x, y') \rangle = x + yy' \leq x \leq 1$. Conversely, if $x > 1$, then $\langle (1, 0), (x, y) \rangle = x > 1$ and if $y' > 0$, then by choosing y large enough, $\langle (1, y), (x, y') \rangle = x + yy' > 1$.

Now notice $(K \cap (P \times C))^\circ = (\emptyset)^\circ = \mathbb{E} \times \mathbb{Y}$. On the other hand,

$$\begin{aligned} \operatorname{cl}(K^\circ + (P^\circ \times C^\circ)) &= \operatorname{cl}((\mathbb{E} \times \{(x, y) : y \leq 0\}) + (\mathbb{E} \times \{(x, y) : x \leq 1,\ y \leq 0\})) \\ &\subset \mathbb{E} \times \{(x, y) : y \leq 0\} \neq \mathbb{E} \times \mathbb{Y} = (K \cap (P \times C))^\circ. \end{aligned}$$

9 * (Open mapping theorem). Suppose the linear map $A : \mathbb{E} \to \mathbb{Y}$ is surjective.

(a) Prove any set $C \in \mathbb{E}$ satisfies $A \operatorname{core} C \subset \operatorname{core} AC$.

Take $x \in \operatorname{core} C$. Since A is surjective, for any $d \in \mathbb{Y}$, there exists $y \in \mathbb{E}$ such that $Ay = d$. For some $\epsilon > 0$, $x + \epsilon y \in C$. Thus, $A(x + \epsilon y) = Ax + \epsilon d \in AC$. Since d was arbitrary, $Ax \in \operatorname{core} AC$. Thus, $A \operatorname{core} C \subset \operatorname{core} AC$.

(b) Deduce A is an open map: that is, the image of any open set is open.

Linear maps map convex sets to convex sets: if $z_1 = Ax_1$ and $z_2 = Ax_2$, then $\lambda z_1 + (1 - \lambda) z_2 = A(\lambda x_1 + (1 - \lambda) x_2)$.

Now if C is an arbitrary open set, take a point $x \in \operatorname{int} C$. For some $\epsilon > 0$, $\tilde{B} := x + \epsilon B \subset C$. $\tilde{B}$ is convex, and so $A\tilde{B} \subset AC$ is convex. Since $x \in \operatorname{core} \tilde{B}$, $Ax \in \operatorname{core} A\tilde{B} = \operatorname{int} A\tilde{B}$ by Theorem 4.4 (Core and interior). There is a ball around Ax in $A\tilde{B}$, and thus AC. Thus, AC is open.

(c) Prove another condition ensuring condition (3.3.8) in the Fenchel theorem is that there is a point $\hat{x}$ in $\operatorname{int}(\operatorname{dom} f)$ with $A\hat{x}$ in $\operatorname{dom} g$ and A is surjective. Prove similarly that a sufficient condition for Fenchel duality with linear constraints (Corollary 3.3.11) to hold is A surjective and $b \in A(\operatorname{int}(\operatorname{dom} f))$.

The condition (3.3.8) is $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$. Since by part (a), $A \operatorname{core}(\operatorname{dom} f) = \operatorname{core}(A \operatorname{dom} f)$, $A\hat{x} \in \operatorname{core}(A \operatorname{dom} f) \cap \operatorname{dom} g$. Thus, we can take this to be our point in $\operatorname{dom} g$, and for any direction find a point in $A \operatorname{dom} f$ which is strictly in that direction relative to $A\hat{x}$, proving $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$.

Now the sufficient condition of Corollary 3.3.11 is $b \in \operatorname{core}(A \operatorname{dom} f)$. But by part (a), this is the same as $b \in A \operatorname{core}(\operatorname{dom} f)$. By Theorem 4.4 (Core and interior) and convexity of $\operatorname{dom} f$, this is the same as $b \in A \operatorname{int}(\operatorname{dom} f)$.

(d) Deduce that any cones $H \subset \mathbb{Y}$ and $K \subset \mathbb{E}$, and any surjective linear map $A : \mathbb{E} \to \mathbb{Y}$ satisfy $(K \cap A^{-1}H)^- = A^*H^- + K^-$, providing $H \cap A(\operatorname{int} K) \neq \emptyset$.
I think this should really have the hypothesis that H and K are convex, or else you can make $(K \cap H)^- = \{(x, y) : y \leq 0\}$ and $H^- + K^- = \{(x, y) : |x| \leq -y\}$, by making K a convex cone centered on the positive y axis of angle less than 90, and H a cone which has a ray along the positive y axis and the two rays whose union is $\{(x, y) : |x| = y\}$. $H \cap K$ here would just be the positive y axis.
Anyways, a sufficient condition for Theorem 3.3.13 (Krein-Rutman polar cone calculus) is $H - AK = \mathbb{Y}$. If $H \cap A(\operatorname{int} K) \neq \emptyset$, then $H \cap \operatorname{int}(AK) \neq \emptyset$. By choosing an element of AK which differs from the interior element in an arbitrary direction and scaling, we get $H - AK = \mathbb{Y}$.

10 * (Conical absorption)

(a) If the set $A \subset \mathbb{E}$ is convex, the set $C \subset \mathbb{E}$ is bounded, and $\mathbb{R}_+ A = \mathbb{E}$, prove there exists a real $\delta > 0$ such that $\delta C \subset A$.
If $\mathbb{R}_+ A = \mathbb{E}$, then the gauge function γ_A is finite everywhere and continuous. Because C is bounded, it is contained in some closed ball MB, on which γ_A obtains a maximum. Note that if $\gamma_A(x) \leq \mu$, then $x \in \mu A$, since there exists $\lambda < \mu$ where $x \in \lambda A$, and by $\mathbb{R}_+ A = \mathbb{E}$ and A convex, A must contain 0, so $\lambda A \subset \mu A$. Thus, $C \subset MB \subset (\max_{x \in MB} \gamma_A(x))A$. If $(\max_{x \in MB} \gamma_A(x)) = 0$ we can set δ to anything, otherwise take $\delta = (\max_{x \in MB} \gamma_A(x))^{-1}$.

Now define two sets in $\mathbb{S}^2_+$ by

$$A = \left\{ \begin{bmatrix} y & x \\ x & z \end{bmatrix} \in \mathbb{S}^2_+ \,\middle|\, |x| \leq y^{2/3} \right\}, \text{ and}$$
$$C = \{X \in \mathbb{S}^2_+ \mid \operatorname{Tr} X \leq 1\}.$$

(b) Prove that both A and C are closed, convex, and contain 0, and that C is bounded.
A is convex: note $y \geq 0$, because the corresponding matrix is PSD. Take two matrices in A, differentiating their entries by indexing those of one by 1 and those of the other by 2. $\mathbb{S}^2_+$ is convex, so any convex combination of them is still PSD. Also, for $\lambda \in [0, 1]$, by convexity of $x \mapsto |x|^{3/2}$,

$$|\lambda x_1 + (1 - \lambda)x_2|^{3/2} \leq \lambda |x_1|^{3/2} + (1 - \lambda)|x_2|^{3/2} = \lambda y_1 + (1 - \lambda) y_2.$$

Thus, raising both sides to the power 2/3, we get that the convex combination is in A. Thus, A is convex. It contains 0 as $|0| \leq 0^{2/3}$ and $0 \in \mathbb{S}^2_+$. It is closed because $f\left(\begin{bmatrix} y & x \\ x & z \end{bmatrix}\right) = |x| - y^{2/3}$ is continuous, and so $\{X : f(X) \leq 0\}$ is closed.
Tr is a linear functional, so C is the intersection of $\mathbb{S}^2_+$ with the closed halfspace

$\{X \in \mathbb{S}^2 | \operatorname{Tr} X \leq 1\}$, which makes it closed and convex. It contains 0 because $0 \in \mathbb{S}^2_+$ and $\operatorname{Tr} 0 = 0 \leq 1$. C is bounded because

$$X \in C \implies \|X\| = \|\lambda(X)\| \leq \sum_{i=1}^{n} |\lambda_i(X)| = \operatorname{Tr} X \leq 1.$$

(c) Prove $\mathbb{R}_+ A = \mathbb{S}^2_+ = \mathbb{R}_+ C$.
For any nonzero $X \in \mathbb{S}^2_+$, $X / \operatorname{Tr} X \in C$. Further, any positive scaling of C is in $\mathbb{S}^2_+$. Thus, $\mathbb{S}^2_+ = \mathbb{R}_+ C$. Likewise, any positive scaling of A is in $\mathbb{S}^2_+$ (this follows from the fact that $\mathbb{S}^2_+$ is a cone).
Now given $X = \begin{bmatrix} y & x \\ x & z \end{bmatrix} \in \mathbb{S}^2_+$, for any $c > 0$,

$$|cx| \leq (cy)^{2/3} \iff c^{1/3} x \leq y^{2/3}.$$

We can clearly take a c small enough so that $c^{1/3} x \leq y^{2/3}$. Thus, $cX \in A$. Thus, $X \in c^{-1} A$. This proves that $\mathbb{S}^2_+ = \mathbb{R}_+ A$.

(d) Prove there is no real $\delta > 0$ such that $\delta C \subset A$.
Let $\delta > 0$ be arbitrary and consider the following matrix in δC for some $\lambda \in (0, 1)$:

$$X = \begin{bmatrix} \delta\lambda & \delta\sqrt{\lambda(1-\lambda)} \\ \delta\sqrt{\lambda(1-\lambda)} & \delta(1-\lambda). \end{bmatrix}$$

Its determinant is 0 and has trace δ, so it belongs to δC. Let us compute when X is not in A:

$$\delta\sqrt{\lambda(1-\lambda)} > (\delta\lambda)^{2/3} \iff \delta^6(\lambda(1-\lambda))^3 > \delta^4\lambda^4 \iff \delta^2 > \frac{\lambda}{(1-\lambda)^3}.$$

If we take $\lambda \to 0$, then $\frac{\lambda}{(1-\lambda)^3} \to 0 < \delta$. Therefore, there exists a setting of λ for which $X \in \delta C$ yet $X \notin A$, proving $\delta C \not\subset A$.

11 * (Hölder's inequality). This question develops an alternative approach to the theory of the p-norm $\|\cdot\|_p$ defined in Section 2.3, Exercise 6.

(a) Prove $p^{-1}\|x\|_p^p$ is a convex function, and deduce the set

$$B_p = \{x \mid \|x\|_p \leq 1\}$$

is convex.
Since the function $x \mapsto |x|^p$ is convex with derivative $x|x|^{p-2}$ (see Section 3.1 Exercise 14), the function

$$\frac{\|x\|_p^p}{p} = \frac{1}{p}\sum_{i=1}^{n} |x_i|^p$$

is a sum of convex function and is thus convex, and has gradient $x|x|^{p-2}$ (where the multiplication and absolute value are element-wise). Then, B_p is a level set of a convex function, which makes it convex.

(b) Prove the gauge function $\gamma_{B_p}(\cdot)$ is exactly $\|\cdot\|_p$, and deduce $\|\cdot\|_p$ is convex.

$$\|\lambda x\|_p = \left(\sum_{i=1}^n |\lambda x_i|^p\right)^{1/p} = \left(|\lambda|^p \sum_{i=1}^n |x_i|^p\right)^{1/p} = |\lambda|\|x\|_p,$$

which proves positive homogeneity of $\|\cdot\|_p$. In particular, $\lambda B_p = \{x \mid \|x\|_p \leq \lambda\}$. Therefore, $\gamma_{B_p}(x) = \inf\{\lambda \geq 0 \mid \|x\|_p \leq \lambda\} = \|x\|_p$. By convexity of the gauge function, $\|x\|_p$ is convex. Since it is convex and positive homogeneous, it is in fact sublinear.

(c) Use the Fenchel-Young inequality (3.3.4) to prove that any vectors x and $\phi \in \mathbb{R}^n$ satisfy the inequality

$$p^{-1}\|x\|_p^p + q^{-1}\|\phi\|_q^q \geq \langle \phi, x \rangle .$$

By Section 3.3, Exercise 1, the convex conjugate of $p^{-1}|x|^p$ is $q^{-1}|y|^q$, where $p^{-1}+q^{-1} = 1$. We have

$$\begin{aligned}(\|\cdot\|_p^p)^*(y) &= \sup_{x\in\mathbb{E}} \langle y, x\rangle - \frac{\|x\|_p^p}{p} = \sup_{x\in\mathbb{E}} \sum_{i=1}^n x_i y_i - \frac{|x_i|^p}{p} \\ &= \sum_{i=1}^n \sup_{x_i\in\mathbb{R}} x_i y_i - \frac{|x_i|^p}{p} = \sum_{i=1}^n \frac{|y_i|^q}{q} = \frac{\|y\|_q^q}{q}.\end{aligned}$$

Then, by the Fenchel-Young inequality (Theorem 3.3.4), for any $x, y \in \mathbb{E}$,

$$\frac{\|x\|_p^p}{p} + \frac{\|y\|_q^q}{q} \geq \langle x, y\rangle .$$

(d) Assuming $\|u\|_p = \|v\|_q = 1$, deduce $\langle u, v\rangle \leq 1$, and hence prove that any vectors x and $\phi \in \mathbb{R}^n$ satisfy the inequality

$$\langle \phi, x\rangle \leq \|\phi\|_q \|x\|_p.$$

By part (c), $\langle u, v\rangle \leq p^{-1}\|u\|_p^p + q^{-1}\|v\|_q^q = p^{-1} + q^{-1} = 1$. Now if $\phi = 0$ or $x = 0$, the inequality holds because both sides are 0. Otherwise, $\phi \neq 0$ and $x \neq 0$. Then, by the above, $\left\langle \|x\|_p^{-1}x, \|\phi\|_q^{-1}\phi\right\rangle \leq 1$, which by multiplying both sides by $\|x\|_p\|\phi\|_q$ implies $\langle x, \phi\rangle \leq \|x\|_p\|\phi\|_q$.

(e) Calculate B_p°.
$B_p^\circ = \{\phi \in \mathbb{E} : \forall x \text{ s.t. } \|x\|_p \leq 1,\ \langle \phi, x\rangle \leq 1\}$. By part (d), $B_q \subset B_p^\circ$, as if $\|\phi\|_q \leq 1$, then $\langle \phi, x\rangle \leq \|\phi\|_q\|x\|_p \leq \|x\|_p \leq 1$ assuming $x \in B_p$ for the last step.
To show the reverse inclusion, we have to show that for any ϕ, equality is obtained in $\langle \phi, x\rangle \leq \|\phi\|_q\|x\|_p$ by some x for any fixed p-norm. Suppose $\|x\|_p = \lambda$. Set

$$x = \lambda \frac{\phi|\phi|^{q-2}}{\|\phi\|_q^{q-1}}.$$

We'll verify that $\|x\|_p = \lambda$. To this end, we use the facts that $(p-1)(q-1) = 1$ and $q/p = q-1$.

$$\begin{aligned}
\|x\|_p &= \frac{\lambda}{\|\phi\|_q^{q/p}} \|\phi|\phi|^{q-2}\|_p \\
&= \frac{\lambda}{\|\phi\|_q^{q-1}} \left(\sum_{i=1}^n |\phi|^{p(q-1)}\right)^{1/p} \\
&= \frac{\lambda}{\|\phi\|_q^{q-1}} \left(\sum_{i=1}^n |\phi|^{q}\right)^{1/p} \\
&= \frac{\lambda}{\|\phi\|_q^{q-1}} \|\phi\|_q^{q/p} = \frac{\lambda}{\|\phi\|_q^{q-1}} \|\phi\|_q^{q-1} = \lambda.
\end{aligned}$$

Now we have

$$\langle \phi, x \rangle = \frac{\lambda}{\|\phi\|_q^{q-1}} \left\langle \phi, \phi|\phi|^{q-2} \right\rangle = \frac{\lambda}{\|\phi\|_q^{q-1}} \|\phi\|_q^q = \lambda \|\phi\|_q.$$

Now for $\phi \notin B_q$, there exists $x \in B_p$ such that $\langle x, \phi \rangle = \|\phi\|_q > 1$. Thus, $\phi \notin B_p^\circ$. Thus, $B_p^\circ \subset B_q$. We have completed the proof that $B_p^\circ = B_q$.

12 * (Pareto minimization). We use the notation of Section 3.3, Exercise 18 (Order convexity), and we assume the cone S is pointed and has nonempty interior. Given a set $D \subset \mathbb{Y}$, we say a point y in D is a *Pareto minimum of D (with respect to S)* if

$$(y - D) \cap S = \{0\},$$

and a *weak minimum* if

$$(y - D) \cap \operatorname{int} S = \emptyset.$$

(a) Prove y is a Pareto (respectively weak) minimum of D if and only if it is a Pareto (respectively weak) minimum of $D + S$.
Suppose that y is a Pareto minimum of D. Let $d \in D$ and $x \in S$. If $y - d - x \in S$, then $y - d \in x + S \subset S$ because S is a cone. Then $y - d = 0$. Then $-x \in S$. Then $x = 0$, because S is pointed ($S \cap -S = \{0\}$).
If y is a (weak) Pareto minimum of $D+S$, then since $D \subset D+S$, we have $(y-D) \cap S \subset (y - (D+S)) \cap S$. This implies y is a (weak) Pareto minimum of D.
If y is a weak Pareto minimum of D, suppose $y - d - x \in \operatorname{int} S$. Note that if $z \in \operatorname{int} S$ and $x \in S$, then $z + x \in \operatorname{int} S$. This holds because $z + \epsilon B \in S \implies z + x + \epsilon B \in S$ by convex conity of S. Then, $y - d \in \operatorname{int} S$. But this is a contradiction; therefore, $(y - (D+S)) \cap \operatorname{int} S = \emptyset$.

(b) The map $X \in \mathbb{S}^n_+ \mapsto X^{1/2}$ is $\mathbb{S}^n_+$-order-preserving (Section 1.2, Exercise 5). Use this fact to prove, for any matrix $Z \in \mathbb{S}^n_+$, the unique Pareto minimum of the set

$$\{X \in \mathbb{S}^n \mid X^2 \succeq Z^2\}$$

with respect to $\mathbb{S}^n_+$ is Z.

I think inside the set notation it should be $\mathbb{S}^n_+$, not $\mathbb{S}^n$, because if you take $X = -\epsilon I$ with $\epsilon > 0$ large enough, you'd have $\epsilon^2 I \succeq Z^2$ and $Z - (-\epsilon I) = Z + \epsilon I \in \operatorname{int} \mathbb{S}^n_+$.

Assuming $X \in \{X \in \mathbb{S}^n_+ \mid X^2 \succeq Z^2\}$, by order-preservingness of $X \mapsto X^{1/2}$, we have $X \succeq Z$. Thus, $Z - X \in -\mathbb{S}^n_+$. If $Z - X \in \mathbb{S}^n_+$ as well, $Z - X = 0$. This proves that Z is a Pareto minimum of the set with respect to $\mathbb{S}^n_+$.

Now if X is any Pareto minimum, we have $X - Z \in \mathbb{S}^n_+$ by $X \succeq Z$ and by definition of Pareto minimum, $X - Z = 0$. Thus, $X = Z$.

For a convex set $C \subset \mathbb{E}$ and an S-convex function $F : C \to \mathbb{Y}$, we say a point $\bar{x} \in C$ is a *Pareto* (respectively, *weak*) *minimum* of the *vector optimization problem*

$$\inf\{F(x) \mid x \in C\} \tag{0.2.1}$$

if $F(\bar{x})$ is a Pareto (respectively weak) minimum of $F(C)$.

(c) Prove $F(C) + S$ is convex.

Since F is S-convex, for any $x_1, x_2 \in C$,

$$\begin{aligned} &\lambda F(x_1) + (1-\lambda)F(x_2) - F(\lambda x_1 + (1-\lambda)x_2) \in S \\ \implies\ &\lambda F(x_1) + (1-\lambda)F(x_2) = F(\lambda x_1 + (1-\lambda)x_2) + s, \quad \text{for some } s \in S. \end{aligned}$$

Thus, given $F(x_1) + s_1 \in F(C) + S$ and $F(x_2) + s_2 \in F(C) + S$, we have

$$\begin{aligned} \lambda(F(x_1) + s_1) + (1-\lambda)(F(x_2) + s_2) &= \lambda F(x_1) + (1-\lambda)F(x_2) + \lambda s_1 + (1-\lambda)s_2 \\ &= F(\lambda x_1 + (1-\lambda)x_2) + s + \lambda s_1 + (1-\lambda)s_2 \in F(C) + S. \end{aligned}$$

Thus, $F(C) + S$ is convex.

(d) **(Scalarization).** Suppose $\bar{x}$ is a weak minimum of the problem (0.2.1). By separating $F(\bar{x}) - F(C) - S$ and $\operatorname{int} S$ (using Exercise 6), prove there is a nonzero element ϕ of $-S^-$ such that $\bar{x}$ solves the *scalarized* convex optimization problem

$$\inf\{\langle \phi, F(x)\rangle \mid x \in C\}.$$

Conversely, show any solution of this problem is a weak minimum of (0.2.1).

By part (a), $F(\bar{x})$ is a weak minimum of $F(C)$ iff $F(\bar{x})$ is a weak minimum of $F(C)+S$. Also, by part (c), $F(C) + S$ is convex. Therefore, $F(\bar{x}) - F(C) - S$ is a convex set disjoint from $\operatorname{int} S$. Thus, by Exercise 6, there exists a closed hyperplane containing $F(\bar{x}) - F(C) - S$ but not $\operatorname{int} S$, i.e. $\phi \in \mathbb{E}, b \in \mathbb{R}$ such that

$$\forall z \in F(\bar{x}) - F(C) - S,\ y \in \operatorname{int} S, \quad \langle \phi, z\rangle \geq b > \langle \phi, y\rangle .$$

Since we can take $z = 0$ and $y/\lambda \in \operatorname{int} S$ for any $\lambda > 0$, we have $b = 0$. Further note that $\phi \in S^-$. We have $\operatorname{cl} \operatorname{int} S = S$ by Section 1.1, Exercise 11 (e), and $\langle \phi, y^i\rangle < 0$ for all $i \in \mathbb{N}$ implies $\langle \phi, \lim_i y^i\rangle \leq 0$. So, $-\phi \in -S^-$ and satisfies

$$\begin{gathered} \forall z \in F(\bar{x}) - F(C) - S,\ \langle -\phi, z\rangle \leq 0, \\ \text{i.e. } \forall x \in C,\ s \in S,\ \langle -\phi, \bar{x}\rangle \leq \langle -\phi, x + s\rangle = \langle -\phi, x\rangle \ \text{ for } s = 0. \end{gathered}$$

This proves that $\bar{x}$ is also a solution to problem scalarized by $-\phi \in -S^-$.
We can see that any solution $\bar{x}$ to the scalarized problem satisfies $\langle -\phi, z \rangle \le 0$ for all $z \in F(\bar{x}) - F(C) - S$, and since $-\phi \in -S^-$ is nonzero, $\langle -\phi, y \rangle > 0$ for all $y \in \operatorname{int} S$. So, $F(\bar{x}) - F(C) - S$ is disjoint from $\operatorname{int} S$. In other words, $F(\bar{x})$ is a weak minimum of $F(C)$.

13 (Existence of extreme points). Prove any nonempty compact convex set $C \subset \mathbb{E}$ has an extreme point, without using Minkowski's theorem, by considering the furthest point in C from the origin.

Proof. By the strict convexity of $x \mapsto \|x\|^2$ and compactness of C, the infimum $\inf_{x \in C} \|x\|^2$ is uniquely obtained at some $\bar{x} \in C$. For any $x, y \in C$ both not equal to $\bar{x}$ and $\lambda \in [0, 1]$,

$$\|\lambda x + (1-\lambda) y\|^2 \le \lambda \|x\|^2 + (1-\lambda)\|y\|^2 < \|\bar{x}\|^2.$$

Thus, $\lambda x + (1-\lambda) y \neq \bar{x}$. This proves $C - \bar{x}$ is convex. So, $\bar{x}$ is an extreme point. ☐

14. Prove Lemma 4.1.7.

Proof. Suppose $\bar{x}$ is an extreme point of $C \cap H$; that is, $(C \cap H) \setminus \{\bar{x}\}$ is convex. Let us show that $C - \{\bar{x}\}$ is convex. If it is not, then $\bar{x} = \lambda z + (1-\lambda) y$ for $z, y \in C$ and $\lambda \in (0, 1)$. H has the form $H = \{\langle \phi, x \rangle = b : x \in \mathbb{E}\}$ for some $\phi \in \mathbb{E}$, $b \in \mathbb{R}$. Thus, $\langle \phi, \bar{x} \rangle = b$, and $\langle \phi, x \rangle \ge b$ for all $x \in C$. If $\langle \phi, z \rangle > b$ or $\langle \phi, y \rangle > b$, then $\langle \phi, \lambda z + (1-\lambda) y \rangle > b$, contradicting the vact that $\bar{x} = \lambda z + (1-\lambda) y$. Thus, $z \in C \cap H$ and $y \in C \cap H$. But this contradicts the fact $(C \cap H) \setminus \{\bar{x}\}$ is convex. ☐

15. For any compact convex set $C \subset \mathbb{E}$, prove $C = \operatorname{conv}(\operatorname{bd} C)$.

Proof. $\operatorname{conv}(\operatorname{bd} C) \subset C$ because C is convex and $\operatorname{bd} C \subset C$. For the other direction, take $x \in \operatorname{int} C$. For an arbitrary $d \in C$, consider

$$c^+ = \sup\{\epsilon > 0 : x + \epsilon d \in C\}, \quad c^- = \sup\{\epsilon > 0 : x - \epsilon d \in C\}.$$

Since $x \in \operatorname{int} C$, $c^+ > 0$ and $c^- > 0$, and since C is bounded, $c^+ < \infty$ and $c^- < \infty$. We'll prove that $x + c^+ d \in \operatorname{bd} C$. There exist $c^i \to c^+$ such that $x + c^i d \in C$. Thus, since C is closed, $x + c^+ d \in \operatorname{bd} C$. Furthermore, for any $\epsilon > 0$, $x + (c^+ + \epsilon) d \notin C$. Thus, $x + c^+ d \in \operatorname{bd} C$. Similarly, $x - c^- d \in \operatorname{bd} C$. We can write

$$x = \frac{c^-}{c^+ + c^-}(x + c^+ d) + \frac{c^+}{c^+ + c^-}(x - c^- d).$$

Therefore, $x \in \operatorname{conv}(\operatorname{bd} C)$. Since $x \in \operatorname{int} C$ was arbitrary, $\operatorname{int} C \subset \operatorname{conv}(\operatorname{bd} C)$. Thus, $C = (\operatorname{int} C) \cup \operatorname{bd} C \subset \operatorname{conv}(\operatorname{bd} C)$. ☐

16 * (A converse of Minkowski's theorem). Suppose D is a subset of a compact convex set $C \subset \mathbb{E}$ satisfying $\operatorname{cl}(\operatorname{conv} D) = C$. Prove $\operatorname{ext} C \subset \operatorname{cl} D$.

Proof. Suppose $x \in \operatorname{ext} C$ but $x \notin \operatorname{cl} D$. Note $\operatorname{conv}(\operatorname{cl} D) = \operatorname{cl}(\operatorname{conv} D)$. This holds because $\operatorname{cl} D$ is compact (C is compact, so $D \subset C$ must be bounded), and the convex hull of a compact set is compact (Section 2.2, Exercise 5 (d)). Thus, $\operatorname{cl} D \subset \operatorname{cl}(\operatorname{conv} D)$ implies $\operatorname{conv}(\operatorname{cl} D) \subset \operatorname{cl}(\operatorname{conv} D)$, and $\operatorname{cl} \operatorname{conv} \operatorname{cl} D = \operatorname{conv} \operatorname{cl} D$ implies $\operatorname{cl}(\operatorname{conv} D) \subset \operatorname{conv} \operatorname{cl} D$.
Then, we get $C = \operatorname{conv}(\operatorname{cl} D) \subset \operatorname{conv}(C \setminus \{x\}) = C \setminus \{x\}$, a contradiction. Therefore, $x \in \operatorname{ext} C \implies x \in \operatorname{cl} D$, i.e. $\operatorname{ext} C \subset \operatorname{cl} D$. $\square$

17 * (Extreme points). Consider a compact convex set $C \subset \mathbb{E}$.

(a) If $\dim \mathbb{E} \le 2$, prove the set $\operatorname{ext} C$ is closed.
If $\dim \mathbb{E} = 0$, then $\operatorname{ext} C = \{0\}$ or $\emptyset$, both of which are closed. If $\dim \mathbb{E} = 1$, then C is empty, a point, or a line segment, in which $\operatorname{ext} C$ is empty, a point, or two isolated points, which are all closed. Now let $\dim \mathbb{E} = 2$. Suppose x^i is a sequence in $\operatorname{ext} C$ converging to $\bar{x}$. Note $\operatorname{ext} C \subset \operatorname{bd} C$, because if $x \in \operatorname{int} C$, then $C \setminus \{x\}$ is clearly nonconvex. Furthermore, by Section 1.1 Exercise 11 (c) (Accessibility lemma), if $\lambda \in (0,1)$ and $x \in \operatorname{int} C$ and $y \in C$, then $\lambda x + (1-\lambda)y \in \operatorname{int} C$.
$\bar{x} \in \operatorname{bd} C$ because $\operatorname{bd} C$ is closed: $\overline{\operatorname{bd} C} = \bar{C} \cup \operatorname{int} C$, which is a union of open sets and is thus open.
We assume $\bar{x} \notin \operatorname{ext} C$ and derive a contradiction. By this assumtion, there exist $x, y \in C$ and $\lambda \in (0,1)$ such that $\bar{x} = \lambda x + (1-\lambda)y$, which means $x, y \in \operatorname{bd} C$ by an above comment. If x^i is in the line segment $L = \{\lambda x + (1-\lambda)y : \lambda \in (0,1)\}$ we have a contradiction, because $x^i \in \operatorname{ext} C$. Thus, the sequence (x^i) is outside of the above line segment. Now by Theorem 4.1.6, there exists $\phi \in \mathbb{E}$ and $q \in \mathbb{R}$ such that $L \subset H = \{x \in \mathbb{E} : \langle \phi, x \rangle \le q\}$, and for all $x \in C \setminus H$, $\langle \phi, x \rangle > q$. Because H is one-dimensional, $H = \operatorname{aff} L$. For i large enough, $\|x - x^i\|$ is small enough to where $x^i \in H$ implies $x^i \in L$. Therefore, for all i large enough, $x^i \notin H$, that is $\langle \phi, x^i \rangle > q$ so that $\langle \phi, x^i - \bar{x} \rangle > 0$.
By the fundamental theorem of linear algebra, $x^i - \bar{x} = c_1 \hat{\phi} + c_2 \hat{\phi}_\perp$ where $\hat{\phi} = \frac{\phi}{\|\phi\|}$ and $\phi_\perp = \frac{y - \bar{x}}{\|y - \bar{x}\|} = -\frac{x - \bar{x}}{\|x - \bar{x}\|}$ is orthogonal to $\hat{\phi}$. We must have $c_1 > 0$ for $\langle \phi, x^i - \bar{x} \rangle > 0$. Note $-c_2 \hat{\phi}_\perp = -\frac{c_2}{\|y - \bar{x}\|}(y - \bar{x}) = \frac{c_2}{\|x - \bar{x}\|}(x - \bar{x})$. If $c_2 < 0$, set $c = \frac{-c_2}{\|y - \bar{x}\|}$ and $v = y - \bar{x}$, otherwise set $c = \frac{c_2}{\|x - \bar{x}\|}$ and $v = x - \bar{x}$. We have $c \ge 0$, $v + \bar{x} \in \{x, y\}$, $c_2 \hat{\phi}_\perp + cv = 0$.

$$
\begin{aligned}
C \ni \frac{1}{1+c} x^i + \frac{c}{1+c}(v + \bar{x}) &= \frac{1}{1+c}\left(\bar{x} + c_1 \hat{\phi} + c_2 \hat{\phi}_\perp\right) + \frac{c}{1+c}(v + \bar{x}) \\
&= \bar{x} + \frac{c_1}{1+c}\hat{\phi} + \frac{1}{1+c}(c_2 \hat{\phi}_\perp + cv) = \bar{x} + \frac{c_1}{1+c}\hat{\phi}.
\end{aligned}
$$

Therefore, $\mu \hat{\phi} \in C - \bar{x}$ for some $\mu > 0$. Furthermore, for some $\nu > 0$ small enough, $\nu \hat{\phi}_\perp$ and $-\nu \hat{\phi}_\perp$ are both in $C - \bar{x}$. For i large enough, $x^i - \bar{x} = c_1 \hat{\phi} + c_2 \hat{\phi}_\perp$ satisfies $\frac{c_1}{\mu} + \frac{|c_2|}{\nu} < 1$. That is, $x^i - \bar{x} = \left(1 - \frac{c_1}{\mu} - \frac{|c_2|}{\nu}\right) 0 + \frac{c_1}{\mu}\mu\hat{\phi} + \frac{|c_2|}{\nu}\operatorname{sgn}(c_2)\nu\hat{\phi}_\perp \in C - \bar{x}$. In fact, for any $z = c_1' \hat{\phi} + c_2' \hat{\phi}_\perp$ where $c_1' \ge 0$ and $\frac{c_1'}{\mu} + \frac{|c_2'|}{\nu} \le 1$, the previous representation of $x^i - \bar{x}$ implies $z \in C = \bar{x}$. Therefore, $x^i - \bar{x} \in \operatorname{int}(C - \bar{x})$, i.e. $x^i \in \operatorname{int} C$. But this contradicts $x^i \in \operatorname{ext} C \subset \operatorname{bd} C$. Thus, $\bar{x} \in \operatorname{ext} C$, i.e. $\operatorname{ext} C$ is closed.
I absolutely HATED proving this.

(b) If $\mathbb{E}$ is $\mathbb{R}^3$ and C is the convex hull of the set

$$\{(x,y,0) \mid x^2+y^2=1\} \cup \{(1,0,1),(1,0,-1)\},$$

prove $\operatorname{ext} C$ is not closed.
The point $(1,0,0) \notin \operatorname{ext} C$, because it is is equal to $\frac{1}{2}(1,0,-1)+\frac{1}{2}(1,0,1)$. Now we prove that any other $(x,y,0) \in \operatorname{ext} C$ where $(x,y) \in B$.
Suppose that $(x,y,0) = \lambda_1(1,0,1)+\lambda_2(1,0,-1)+(1-\lambda_1-\lambda_2)(x',y',0)$ where $0 \le \lambda_i$, $i \in [3]$ and $\|(x',y')\| = 1$. We must have $\lambda_1 = \lambda_2$ because the third component is 0. Thus, we get

$$(x,y) = \lambda(1,0)+(1-\lambda)(x',y')$$

where $\lambda \in (0,1)$. However, by strict convexity of $\|\cdot\|^2$,

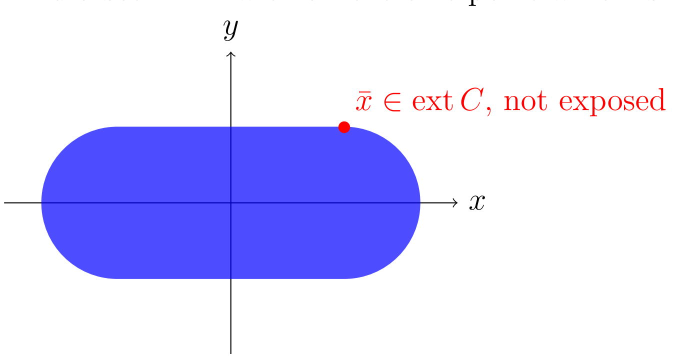
$$\|\lambda(1,0)+(1-\lambda)(x',y')\|^2 < \lambda\|(1,0)\|^2+(1-\lambda)\|(x',y')\|^2 = 1 = \|(x,y)\|^2,$$

unless $(x',y') = (1,0)$. But then $(x,y) = (1,0)$, a contradiction. Thus, $(x,y,0)$ is not writable as a strict convex combination of other points in C, which means it is in $\operatorname{ext} C$. Now, we see $(1,0,0) \in \operatorname{cl} \operatorname{ext} C \setminus \operatorname{ext} C$. Thus, $\operatorname{ext} C$ is not closed.

18 * (Exposed points). A point x in a convex set $C \subset \mathbb{E}$ is called *exposed* if there is an element ϕ of $\mathbb{E}$ such that $\langle \phi, x\rangle > \langle \phi, z\rangle$ for all points $z \neq x$ in C.

(a) Prove any exposed point is an extreme point.
Let $x \in C$ be exposed. Take $y,z \in C$. For any $\lambda \in [0,1]$, $\langle \phi, \lambda y+(1-\lambda)z\rangle < \langle \phi, x\rangle$. Thus, $\lambda y+(1-\lambda)z \neq x$. Therefore, $x \in \operatorname{ext} C$.

(b) Find a set in $\mathbb{R}^2$ with an extreme point which is not exposed.



19 ** (Tangency conditions). Let $\mathbb{Y}$ be a Euclidean space. Fix a convex set C in $\mathbb{E}$ and a point x in C.

(a) Show $x \in \operatorname{core} C$ if and only if $T_C(x) = \mathbb{E}$. (You may use Exercise 20(a).)
$x \in \operatorname{core} C \iff \mathbb{R}_+(C-x)\mathbb{E} \iff \operatorname{cl} \mathbb{R}_+(C-x) = T_C(x) = \mathbb{E}$.

(b) For a linear map $A : \mathbb{E} \to \mathbb{Y}$, prove $AT_C(x) \subset T_{AC}(Ax)$.

$$\begin{aligned} AT_C(x) &= A \operatorname{cl} \mathbb{R}_+(C-x) \subset \operatorname{cl} A\mathbb{R}_+(C-x) \\ &= \operatorname{cl} \mathbb{R}_+ A(C-x) = \operatorname{cl} \mathbb{R}_+(AC-Ax) = T_{AC}(Ax). \end{aligned}$$

For any set S, $A \operatorname{cl} S \subset \operatorname{cl} AS$ because $s^i \to s \implies As^i \to As$.

(c) For another convex set D in $\mathbb{Y}$ and a point $y \in D$, prove

$$N_{C\times D}(x,y) = N_C(x) \times N_D(y) \text{ and}$$
$$T_{C\times D}(x,y) = T_C(x) \times T_D(y).$$

Suppose $(a,b) \in N_{C\times D}(x,y)$. Then, since $x \in C$, for any $y' \in D$, $\langle (a,b),(x,y')-(x,y)\rangle = \langle b, y'-y\rangle \leq 0$. Similarly because $y \in D$, $\langle a, x-x'\rangle \leq 0$ for any $x' \in C$. Thus, $(a,b) \in N_C(x) \times N_D(y)$. If $a \in N_C(x)$ and $b \in N_D(y)$, then

$$\forall (x',y') \in C \times D, \quad \langle (a,b),(x',y')-(x,y)\rangle = \langle a, x'-x\rangle + \langle b, y'-y\rangle \leq 0.$$

Therefore, $N_{C\times D}(x,y) = N_C(x) \times N_D(y)$.

$$\begin{aligned} T_{C\times D}(x,y) &= \operatorname{cl}\mathbb{R}_+(C\times D-(x,y)) = \operatorname{cl}\mathbb{R}_+((C-x)\times(D-y)) \\ &= \operatorname{cl}(\mathbb{R}_+(C-x)\times\mathbb{R}_+(D-y)) = (\operatorname{cl}\mathbb{R}_+(C-x))\times(\operatorname{cl}\mathbb{R}_+(D-y)) = T_C(x)\times T_D(y). \end{aligned}$$

For the third equality, clearly $\operatorname{cl}\mathbb{R}_+(C-x)\times(D-y) \subset \operatorname{cl}(\mathbb{R}_+(C-x)\times\mathbb{R}_+(D-y))$. For the other direction, if $\mu u \in \mathbb{R}_+(C-x)$ and $\nu v \in \mathbb{R}_+(D-y)$, then if $\mu = \nu = 0$ clearly $(0,0) \in \mathbb{R}_+(C-x)\times(D-y)$. Otherwise, $\frac{\mu}{\mu+\nu}u \in C-x$ because $0, u \in C-x$ and $\frac{\nu}{\mu+\nu}v \in D-y$ because $0, v \in D-y$. Thus, $(\mu u, \nu v) \in \mathbb{R}_+(C-x)\times(D-y)$, multiplying the previous items by $\mu+\nu \in \mathbb{R}_+$.
Alternatively, by Theorem 3.3.14 (Bipolar cone), $T_C(x) = \operatorname{cl}\mathbb{R}_+(C-x) = (C-x)^{--} = N_C(x)^-$. By Exercise 8 (a), $N_{C\times D}(x,y)^- = N_C(x)^-\times N_D(y)^-$. Therefore, $T_{C\times D}(x,y) = T_C(x)\times T_D(y)$.

(d) Suppose the point x also lies in the convex set $G \subset \mathbb{E}$. Prove $T_C(x) - T_G(x) \subset T_{C-G}(0)$, and deduce

$$0 \in \operatorname{core}(C-G) \iff T_C(x) - T_G(x) = \mathbb{E}.$$

For any sets S, T, $\operatorname{cl} S - \operatorname{cl} T \subset \operatorname{cl}(S-T)$. This is because if $s \in \operatorname{cl} S$ and $t \in \operatorname{cl} T$, there are sequences $s^i \in S$ and $t^i \in T$ such that $s^i - t^i \to s - t$, i.e. $s - t \in \operatorname{cl}(S-T)$. Thus, $T_C(x) - T_G(x) = \operatorname{cl}\mathbb{R}_+(C-x) - \operatorname{cl}\mathbb{R}_+(G-x) \subset \operatorname{cl}(\mathbb{R}_+(C-x)-\mathbb{R}_+(G-x))$.
For any convex set S containing 0, $x \in S \implies \lambda x \in S$ if $\lambda \in [0,1]$. $C-x, G-s$ are convex sets containing 0. Therefore, given $u \in C-x$, $v \in G-x$ and $\mu, \nu \in \mathbb{R}_+$, if $\mu = \nu = 0$ then clearly $\mu u - \nu v = 0 \in \mathbb{R}_+(C-G)$. Otherwise,

$$\mu u - \nu v = (\mu+\nu)\left(\frac{\mu}{\mu+\nu}u - \frac{\nu}{\mu+\nu}v\right) \in \mathbb{R}_+(C-x-(G-x)) = \mathbb{R}_+(C-G).$$

Therefore, $T_C(x) - T_G(x) \subset \operatorname{cl}(\mathbb{R}_+(C-x)-\mathbb{R}_+(G-x)) = \operatorname{cl}\mathbb{R}_+(C-G) = T_{C-G}(0)$.
Alternatively, using parts (b) and (c), define the map $A: \mathbb{E}\times\mathbb{E}$, $(x,y)\mapsto x-y$.

$$T_C(x) - T_G(x) = A(T_C(x)\times T_G(x)) = A(T_{C\times G}(x,x)) \subset T_{A(C\times G)}(x-x) = T_{C-G}(0).$$

Now if $T_C(x) - T_G(x) = \mathbb{E}$, then $T_{C-G}(0) = \mathbb{E}$, so $0 \in \operatorname{core}(C-G)$. If the latter, then $T_{C-G}(0) = \mathbb{E}$. Inspecting (b), $\operatorname{cl} AT_C(x) = T_{AC}(Ax)$ for linear A and convex C, $x \in C$, so $\operatorname{cl}(T_C(x) - T_G(x)) = T_{C-G}(0) = \mathbb{E}$, so $T_C(x) - T_G(x) = \mathbb{E}$ by (a).

(e) Show that the condition (3.3.8) in the Fenchel theorem can be replaced by the condition

$$T_{\operatorname{dom} g}(Ax) - AT_{\operatorname{dom} f}(x) = \mathbb{Y}$$

for an arbitrary point x in $\operatorname{dom} f \cap A^{-1} \operatorname{dom} g$.
Since by part (b), $AT_{\operatorname{dom} f}(x) \subset T_{A \operatorname{dom} f}(Ax)$, we have

$$\mathbb{Y} = T_{\operatorname{dom} g}(Ax) - AT_{\operatorname{dom} f}(x) \subset T_{\operatorname{dom} g}(Ax) - T_{A \operatorname{dom} f}(Ax) \implies T_{\operatorname{dom} g}(Ax) - T_{A \operatorname{dom} f}(Ax) = \mathbb{Y}.$$

By part (d), this implies $0 \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$.

20 ** (Properties of the relative interior). (We use Exercise 9 (Open mapping theorem), as well as Section 1.1, Exercise 13.)

(a) Let D be a nonempty convex set in $\mathbb{E}$. Prove D is a linear subspace if and only if $\operatorname{cl} D$ is a linear subspace. (Hint: $\operatorname{ri} D \neq \emptyset$.)
If D is a linear subspace, then since linear subspaces are closed, $\operatorname{cl} D$ is a linear subspace. Now suppose $\operatorname{cl} D$ is a linear subspace, and suppose $v \in \operatorname{cl} D$. If $v \notin \operatorname{aff} D$, then since $\operatorname{aff} D$ is closed and convex, v has a positive distance to $\operatorname{aff} D$. But, $D \subset \operatorname{aff} D$, and there is a sequence in D converging to v, contradiction. Thus, $v \in \operatorname{aff} D$. In particular, $0 \in \operatorname{aff} D$, so $\operatorname{aff} D = \operatorname{Span} D$.
By section 1.1 Exercise 13 (b), $\operatorname{ri} D$ is nonempty. Let $x \in \operatorname{ri} D$. There exists $\varepsilon > 0$ such that for any $v \in \operatorname{aff}(D) - x = \operatorname{aff}(D) - 0 = \operatorname{Span}(D)$ where $\|v\| \le \varepsilon$, $x + v \in D$. Now let $y \in \operatorname{cl} D$ be arbitrary. Since $x \in \operatorname{cl} D$ and $\operatorname{cl} D$ is linear, $x + y \in \operatorname{cl} D$. Then, there is an element $u \in D$ ε close to $x + y$. Define $\eta = u - (x + y)$. Since $\|\eta\| \le \varepsilon$, we have $x - \eta \in D$. By convexity of D,

$$\frac{1}{2}u + \frac{1}{2}(x - \eta) = \frac{1}{2}(x + y + \eta) + \frac{1}{2}(x - \eta) = x + \frac{1}{2}y \in D.$$

Since $y \in \operatorname{cl} D$, a linear space containing x, was arbitrary, $y = 2(z - x)$ for arbitrary $z \in \operatorname{cl} D$ shows that $\operatorname{cl} D \subset D$. I.e., D is a linear subspace. This finishes the proof.

(b) For a point x in a convex set $C \subset \mathbb{E}$, prove the following properties are equivalent:

(i) $x \in \operatorname{ri} C$.
(i) $\implies$ (ii): By Section 1.1 Exercise 13 (d), $x \in \operatorname{ri} C$ implies $\mathbb{R}_+(C - x)$ is a linear subspace. Since linear subspaces are closed $\operatorname{cl} \mathbb{R}_+(C - x) = \mathbb{R}_+(C - x)$ is a linear subspace.

(ii) The tangent cone $\operatorname{cl} \mathbb{R}_+(C - x)$ is a linear subspace.
(ii) $\implies$ (iii): Since $N_C(x) = (\operatorname{cl} \mathbb{R}_+(C - x))^- = (T_C(x))^-$, $N_C(x) = \{\phi : \langle \phi, z \rangle \le 0 \text{ for all } z \in T_C(x)\}$. If $\langle \phi, z \rangle < 0$, then $\langle \phi, -z \rangle > 0$ and $-z \in T_C(x)$ since $T_C(x)$ is a linear subspace, i.e. $\phi \notin N_C(x)$. Thus, $N_C(x) = \{\phi : \langle \phi, z \rangle = 0 \text{ for all } z \in T_C(x)\} = T_C(x)^\perp$, the orthogonal complement of $T_C(x)$, which is a linear subspace.

(iii) The normal cone $N_C(x)$ is a linear subspace.
(iii) $\implies$ (iv): Being a convex cone closed under negation implies being a linear subspace, since convex cones are closed under addition of nonnegative multiples.

(iv) $y \in N_C(x) \Rightarrow -y \in N_C(x)$.
(iv) $\implies$ (i): (iv) implies that $N_C(x) = \{\phi \in \mathbb{E} : \langle \phi, z\rangle = 0 \;\forall z \in C - x\}$. Since the condition $\langle \cdot, z\rangle = 0$ is closed under linear combinations, $N_C(x)$ is linear. Thus, $\mathrm{cl}(\mathbb{R}_+(C-x)) = N_C(x)^- = N_C(x)^\perp$ is linear. Thus, by part (a) $\mathbb{R}_+(C-x)$ is linear. Thus, by Section 1.1 Exercise 13 (d), $x \in \mathrm{ri}\, C$.

(c) For a convex set $C \subset \mathbb{E}$ and a linear map $A : \mathbb{E} \to \mathbb{Y}$, prove $A\,\mathrm{ri}\, C \supset \mathrm{ri}\, AC$, and deduce

$$A\,\mathrm{ri}\, C = \mathrm{ri}\, AC.$$

I found this question difficult...
Suppose $y \in \mathrm{ri}\, AC$. Denote $C \cap A^{-1}y = \{x \in C : Ax = y\}$ and take $\bar{x} \in \mathrm{ri}(C \cap A^{-1}y)$. We will show that $\bar{x} \in \mathrm{ri}\, C$, which would imply $y \in A\,\mathrm{ri}\, C$ since $A\bar{x} = y$ by definition of $C \cap A^{-1}y$. We will do so by showing condition (iv) from the previous part.
Using the fundamental theorem of linear algebra, we can express an arbitrary element of $N_C(\bar{x})$ as $A^*\phi + w$, where $w \in \mathrm{null}(A)$. For any $x' \in C$, we have

$$\langle \phi, Ax' - y\rangle + \langle w, x' - \bar{x}\rangle = \langle A^*\phi + w, x' - \bar{x}\rangle \le 0. \tag{0.2.2}$$

For any $x' \in C \cap A^{-1}y$, this says $\langle w, x' - \bar{x}\rangle \le 0$, i.e. $w \in N_{C\cap A^{-1}y}(\bar{x})$. Since $\bar{x} \in \mathrm{ri}(C \cap A^{-1})$, $-w \in N_{C\cap A^{-1}y}(\bar{x})$, i.e. $\langle w, x' - \bar{x}\rangle = 0$ for every $x' \in C \cap A^{-1}y$.
Now since $y \in \mathrm{ri}\, AC$, for any $z \in AC$, there exists $\epsilon > 0$ such that $y + \epsilon(y - z) \in AC$ (See Section 1.1 Exercise 13 (d)). So there exist $x', \tilde{x} \in C$ where $Ax' = z$ and $A\tilde{x} = y + \epsilon(y - z)$. Now observe that

$$A\left(\frac{\epsilon}{1+\epsilon}x' + \frac{1}{1+\epsilon}\tilde{x}\right) = \frac{\epsilon}{1+\epsilon}z + \frac{1}{1+\epsilon}(y + \epsilon(y-z)) = y \implies \frac{\epsilon}{1+\epsilon}x' + \frac{1}{1+\epsilon}\tilde{x} \in C \cap A^{-1}y.$$

Furthermore, applying (0.2.2) to x' and $\tilde{x}$, we get

$$Q_1 := \langle \phi, z - y\rangle + \langle w, x' - \bar{x}\rangle \le 0, \qquad Q_2 := \epsilon\,\langle \phi, y - z\rangle + \langle w, \tilde{x} - \bar{x}\rangle \le 0.$$

But in fact, if we add the left hand sides, scaled by $\epsilon/(1+\epsilon)$ and $1/(1+\epsilon)$,

$$\frac{\epsilon}{1+\epsilon}Q_1 + \frac{1}{1+\epsilon}Q_2 = \left\langle w, \frac{\epsilon}{1+\epsilon}x' + \frac{1}{1+\epsilon}\tilde{x}\right\rangle = 0$$

because $\epsilon x'/(1+\epsilon) + \tilde{x}/(1+\epsilon) \in C \cap A^{-1}y$. We added together two nonpositive things and produced zero; therefore, they must both equal 0. Since x' was essentially arbitrary (we took arbitrary $z \in AC$, and arbitrary $x' \in C$ where $Ax' = z$), we have $\langle A^*\phi + w, x' - \bar{x}\rangle = 0$ for all $x' \in C$. Thus, $-(A^*\phi + w) \in N_C(\bar{x})$. Thus, by part (b), $\bar{x} \in \mathrm{ri}\, C$. Since part (e) of Section 1.1, Exercise 13 shows $A\,\mathrm{ri}\, C \subset \mathrm{ri}\, AC$, $A\,\mathrm{ri}\, C = \mathrm{ri}\, AC$.

(d) Suppose U and V are convex sets in $\mathbb{E}$. Deduce

$$\mathrm{ri}(U - V) = \mathrm{ri}\, U - \mathrm{ri}\, V.$$

Use part (c) with the linear map $A : \mathbb{E} \times \mathbb{E}$, $(x, y) \mapsto x - y$ and convex set $U \times V$.

(e) Apply Section 3.1, Exercise 29 (Relativizing the Max formula) to conclude that the condition (3.3.8) in the Fenchel theorem (3.3.5) can be replaced by

$$\operatorname{ri}(\operatorname{dom} g) \cap A \operatorname{ri}(\operatorname{dom} f) \neq \emptyset.$$

The proof of the Fenchel theorem hinges on the existence of a subgradient of $h(u) = \inf_{x \in \mathbb{E}}\{f(x) + g(Ax+u)\}$ at 0, which holds if $0 \in \operatorname{core}(\operatorname{dom} h) = \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} f)$. Section 3.1, Exercise 29 shows that a subgradient exists if $0 \in \operatorname{ri}(\operatorname{dom} h) = \operatorname{ri}(\operatorname{dom} g - A \operatorname{dom} f)$. By part (d) and (c), this equals $0 \in \operatorname{ri}(\operatorname{dom} g) - \operatorname{ri}(A \operatorname{dom} f) = \operatorname{ri}(\operatorname{dom} g) - A \operatorname{ri}(\operatorname{dom} f)$. I.e., $\operatorname{ri}(\operatorname{dom} g) \cap A \operatorname{ri}(\operatorname{dom} f) \neq \emptyset$.

(f) Suppose the function $f : \mathbb{E} \to (-\infty, +\infty]$ is bounded below on the convex set $C \subset \mathbb{E}$, and $\operatorname{ri} C \cap \operatorname{ri}(\operatorname{dom} f) \neq \emptyset$. Prove there is an affine function $\alpha \leq f$ with $\inf_C f = \inf_C \alpha$.
By $\operatorname{ri} C \cap \operatorname{ri}(\operatorname{dom} f) \neq \emptyset$ and f being bounded below on C, the following problems equal and have finite value:

$$\begin{aligned}
\inf_{x \in C}\{f(x)\} &= \sup_{\phi \in \mathbb{E}}\{-f^*(\phi) - \delta_C^*(-\phi)\} \\
&= \sup_{\phi \in \mathbb{E}}\{-\sup_{x' \in \mathbb{E}}\{\langle \phi, x' \rangle - f(x')\} - \sup_{x \in C} \langle -\phi, x \rangle\} \\
&= \sup_{\phi \in \mathbb{E}}\{\inf_{x' \in \mathbb{E}}\{f(x') - \langle \phi, x' \rangle\} + \inf_{x \in C} \langle \phi, x \rangle\}.
\end{aligned}$$

Moreover, the sup is obtained by some $\phi \in \mathbb{E}$, so that

$$\inf_{x \in C} f(x) = \inf_{x' \in \mathbb{E}}\{f(x') - \langle \phi, x' \rangle\} + \inf_{x \in C} \langle \phi, x \rangle .$$

In other words, $\inf_C f = \inf_C \alpha$ for the affine $\alpha = \inf_{x' \in \mathbb{E}}\{f(x') - \langle \phi, x' \rangle\} + \langle \phi, \cdot \rangle$. Moreover, by definition

$$\alpha(x) = \langle \phi, x \rangle + \inf_{x' \in \mathbb{E}}\{f(x') - \langle \phi, x' \rangle\} \leq \langle \phi, x \rangle + f(x) - \langle \phi, x \rangle = f(x).$$

21 ** (Essential smoothness). For any convex f and any point $x \in \operatorname{bd}(\operatorname{dom} f)$, prove $\partial f(x)$ is either empty or unbounded. Deduce that a function is essentially smooth if and only if its subdifferential is always singleton or empty.

Proof. If $\phi \in \partial f(x)$, then for any $d \in N_{\operatorname{dom} f}(x)$ and $x' \in \operatorname{dom} f$,

$$\langle \phi + d, x' - x \rangle \leq \langle \phi, x' - x \rangle \leq f(x') - f(x),$$

proving that $\phi + d \in \partial f(x)$. If $x \in \operatorname{bd}(\operatorname{dom} f)$, then $x \notin \operatorname{int}(\operatorname{dom} f) = \operatorname{core}(\operatorname{dom} f)$ by Theorem 4.1.4 (Core and interior). Thus, $\mathbb{R}_+(\operatorname{dom} f - x) \neq \mathbb{E}$. Thus, $T_{\operatorname{dom} f}(x) \neq \mathbb{E}$. Thus, $N_{\operatorname{dom} f}(x) \neq \{0\}$. Thus, $N_{\operatorname{dom} f}$ is unbounded, which implies $\partial f(x)$ is unbounded if nonempty. Since $\operatorname{dom} \partial f \subset \operatorname{core}(\operatorname{dom} f)$ for essentially smooth f, (see proof of Section 3.1 Exercise 24), we can apply the max formula (Theorem 3.1.8) to any point $\bar{x} \in \operatorname{dom} \partial f$ to show that $\partial f(\bar{x})$ is a singleton. Did I overkill the convex analysis? XD $\square$

22 ** (Birkhoff's theorem [15]) We use the notation of Section 1.2.

(a) Prove $\mathbb{P}^n = \{(z_{ij}) \in \Gamma^n \mid z_{ij} = 0 \text{ or } 1 \text{ for all } i, j\}$.
The provided definition of $\mathbb{P}^n$ is that each entry is 0 or 1 and each row and each column contains one 1. If $X \in \mathbb{P}^n$, clearly it is doubly stochastic and 0-1, showing $\subset$. If on the other hand, X is 0-1 and doubly stochastic, each row and column contains exactly one 1, since if some row/column contains 0 ones, the sum is $0 < 1$, and if it contains more than one 1, the sum is strictly greater than 1. This proves $\supset$.

(b) Prove $\mathbb{P}^n \subset \operatorname{ext}(\Gamma^n)$.
If $\lambda \in (0, 1)$ and $X_1 \neq X_2 \in \Gamma^n$, $X \in \mathbb{P}^n$, $\lambda X_1 + (1-\lambda)X_2 \neq X$ because $\lambda X_1 + (1-\lambda)X_2 \notin \mathbb{P}^n$: if $\lambda X_1 + (1-\lambda)X_2$ is 0-1, then X_1 and X_2 are both 0-1 (or else some entry is in $(0, 1)$), and have ones in the same places for the same reason. In other words, $X_1 = X_2$, a contradiction. Thus, $X \in \operatorname{ext}(\Gamma^n)$.

(c) Suppose $(z_{ij}) \in \Gamma^n \setminus \mathbb{P}^n$. Prove there exist sequences of distinct indices $i_1, i_2, \ldots, i_m$ and $j_1, j_2, \ldots, j_m$ such that

$$0 < z_{i_r, j_r}, z_{i_{r+1}, j_r} < 1 \quad (r = 1, 2, \ldots, m)$$

(where $i_{m+1} = i_1$)). For these sequences, show the matrix $(z'_{i,j})$ defined by

$$z'_{ij} - z_{ij} = \begin{cases} \epsilon & \text{if } (i, j) = (i_r, j_r) \text{ for some } r \\ -\epsilon & \text{if } (i, j) = (i_{r+1}, j_r) \text{ for some } r \\ 0 & \text{otherwise} \end{cases}$$

is doubly stochastic for all small real ϵ. Deduce $(z_{ij}) \notin \operatorname{ext}(\Gamma^n)$.
Since $(z_{ij}) \notin \mathbb{P}^n$, we can find (i_1, j_1) such that $z_{i_1, j_1} \in (0, 1)$. Since $(z_{ij}) \in \Gamma^n$, there must be (i_2, j_1) with $z_{i_2, j_1} \in (0, 1)$ (the j_1 column sums to 1). Since the i_2 row sums to 1, there must be (i_2, j_2) with $z_{i_2, j_2} \in (0, 1)$. Since the j_2 column sums to 1, we can find $(i_3, j_2), \ldots$. This gives a sequence

$$S = (i_1, j_1), (i_2, j_1), (i_2, j_2), (i_3, j_2), (i_3, j_3), \ldots \qquad (i, j) \in S \implies z_{ij} \in (0, 1).$$

Suppose that (i_n) repeats at or before (j_n) repeats. That is, there exist $k < l$ with $i_k = i_l$ and $j_n \neq j_m$ for all $n, m < l$. Consider the subsequence

$$(i_k, j_k), (i_{k+1}, j_k), \ldots, (i_{l-1}, j_{l-1}), (i_k = i_l, j_{l-1}).$$

By construction, $i_{k+1} \neq i_k$, so $l > k + 1$ and $i_k, \ldots, i_{l-1}$ and $j_k, \ldots, j_{l-1}$ are sequences of distinct elements of length at least 2. The exhibited subsequence demonstrates that these sequences fit the desired requirements.
Now suppose that (j_n) repeats before (i_n) does. Then for some $k < l$, $j_k = j_l$ and $i_1, \ldots, i_l$ are distinct. Similarly by construction $j_k \neq j_{k+1}$ so $l > k + 1$. Consider

$$(i_{k+1}, j_k = j_l), (i_{k+1}, j_{k+1}), \ldots, (i_l, j_l = j_k).$$

This subsequence shows that $i_{k+1}, \ldots, i_l$ and $j_{k+1}, \ldots, j_l$ are sequences of distinct elements of length at least 2 satisfying the desired properties.
Now since each row and column of $z'_{ij} - z_{ij}$ sums to 0 (they are either 0 or have an ϵ and a $-\epsilon$), for $|\epsilon|$ small enough z'_{ij} is doubly stochastic. Thus, $(z_{ij}) \notin \operatorname{ext}(\Gamma^n)$.

(d) Deduce $\mathrm{ext}(\Gamma^n) = \mathbb{P}^n$. Deduce Birkhoff's theorem (1.2.5).
By part (b), $\mathbb{P}^n \subset \mathrm{ext}(\Gamma^n)$ and by part (c), $\mathrm{ext}(\Gamma^n) \subset \mathbb{P}^n$. Thus $\mathrm{ext}(\Gamma^n) = \mathbb{P}^n$. By Theorem 4.1.8 (Minkowski) and compact convexity of Γ^n, $\Gamma^n = \mathrm{conv}(\mathbb{P}^n)$, which is Birkhoff's theorem.

(e) Use Caratheodory's theorem (Section 2.2, Exercise 5) to bound the number of permutation matrices needed to represent a doubly stochastic matrix in Birkhoff's theorem.
Caratheodory's theorem states that any element in the convex hull of $\langle a^i \mid i \in I \rangle$ can be expressed as a convex combination of elements in some $J \subset I$ where $|J| \leq 1 + \dim \mathbb{E}$. The dimension of $\mathbb{R}^{n \times n} \supset \mathbb{P}^n$ is n^2 and $\mathrm{conv}(\mathbb{P}^n) = \Gamma^n$, so any doubly stochastic matrix can be represented by $n^2 + 1$ permutation matrices. (Actually $\dim \mathrm{Span}(\mathbb{P}^2) = 2$, and any linear combination of permutation matrices have row and column sums identical, i.e. $X\mathbf{1} = X^\top \mathbf{1} = c\mathbf{1}$ for some $c \in \mathbb{R}$, so it's smaller XD).