Corrections for the book CONVEX ANALYSIS AND OPTIMIZATION, Athena Scientific, 2003, by Dimitri P. Bertsekas

Last Changed: 5/3/04

p. 3 (+22) Change "as the union of the closures of all line segments" to "as the closure of the union of all line segments"

p. 37 (-2) Change "Every x" to "Every $x \neq 0$"

p. 38 (+1) Change "Every x in" to "Every $x \notin X$ that belongs to"

p. 38 (+19) Change "i.e.," to "with $x_1, \ldots, x_m \in \Re^n$ and $m \geq 2$, i.e.,"

p. 63 (+4, +6, +7, +19) Change four times "$c'\overline{y}$" to "$a'\overline{y}$"

p. 67 (+3) Change "$y \in AC$" to "$\overline{y} \in AC$"

p. 70 (+9) Change "[BeN02]" to "[NeB02]"

p. 110 (+3 after the figure caption) Change "... does not belong to the interior of C" to "... does not belong to the interior of C and hence does not belong to the interior of $\mathrm{cl}(C)$ [cf. Prop. 1.4.3(b)]"

p. 148 (-8) Change "$\{x \mid r(x) \leq \gamma\}$" to "$\{z \mid r(z) \leq \gamma\}$"

p. 213 (-6) Change "remaining vectors v_j, $j \neq i$." to "vectors v_j with $v_j \neq v_i$."

p. 219 (+3) Change "$f_i : C \mapsto \Re$" to "$f_i : \Re^n \mapsto \Re$"

p. 265 (+10) Change "$\overline{d}/\|\overline{d}\|$" to "$-\overline{d}/\|\overline{d}\|$"

p. 268 (-3) Change "$j \in A(x^*)$" to "$j \notin A(x^*)$"

p. 338 (+17) Change "Section 5.2" to "Section 5.3"

p. 384 (+6) Change "convex, possibly nonsmooth functions" to "smooth functions, and convex (possibly nonsmooth) functions"

p. 446 (+6 and +8) Interchange "... constrained problem (7.16)" and "... penalized problem (7.19)"

p. 458 (+13) Change "... as well real-valued" to "... as well as real-valued"

p. 458 (-10) Change "We will focus on this ... dual functions." to "In this case, the dual problem can be solved using gradient-like algorithms for differentiable optimization (see e.g., Bertsekas [Ber99a])."

Convex Analysis and Optimization

Chapter 1 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 1: SOLUTION MANUAL

1.1

Assume that C is convex. Then, clearly $(\lambda_1 + \lambda_2)C \subset \lambda_1 C + \lambda_2 C$; this is true even if C is not convex. To show the reverse inclusion, note that a vector x in $\lambda_1 C + \lambda_2 C$ is of the form $x = \lambda_1 x_1 + \lambda_2 x_2$, where $x_1, x_2 \in C$. By convexity of C, we have

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} x_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} x_2 \in C,$$

and it follows that

$$x = \lambda_1 x_1 + \lambda_2 x_2 \in (\lambda_1 + \lambda_2)C.$$

Hence $\lambda_1 C + \lambda_2 C \subset (\lambda_1 + \lambda_2)C$.

For a counterexample when C is not convex, let C be a set in $\Re^n$ consisting of two vectors, 0 and $x \neq 0$, and let $\lambda_1 = \lambda_2 = 1$. Then evidently C is not convex, and $(\lambda_1 + \lambda_2)C = 2C = \{0, 2x\}$ while $\lambda_1 C + \lambda_2 C = C + C = \{0, x, 2x\}$, showing that $(\lambda_1 + \lambda_2)C \neq \lambda_1 C + \lambda_2 C$.

1.2 (Properties of Cones)

(a) Let $x \in \cap_{i \in I} C_i$ and let α be a positive scalar. Since $x \in C_i$ for all $i \in I$ and each C_i is a cone, the vector αx belongs to C_i for all $i \in I$. Hence, $\alpha x \in \cap_{i \in I} C_i$, showing that $\cap_{i \in I} C_i$ is a cone.

(b) Let $x \in C_1 \times C_2$ and let α be a positive scalar. Then $x = (x_1, x_2)$ for some $x_1 \in C_1$ and $x_2 \in C_2$, and since C_1 and C_2 are cones, it follows that $\alpha x_1 \in C_1$ and $\alpha x_2 \in C_2$. Hence, $\alpha x = (\alpha x_1, \alpha x_2) \in C_1 \times C_2$, showing that $C_1 \times C_2$ is a cone.

(c) Let $x \in C_1 + C_2$ and let α be a positive scalar. Then, $x = x_1 + x_2$ for some $x_1 \in C_1$ and $x_2 \in C_2$, and since C_1 and C_2 are cones, $\alpha x_1 \in C_1$ and $\alpha x_2 \in C_2$. Hence, $\alpha x = \alpha x_1 + \alpha x_2 \in C_1 + C_2$, showing that $C_1 + C_2$ is a cone.

(d) Let $x \in \mathrm{cl}(C)$ and let α be a positive scalar. Then, there exists a sequence $\{x_k\} \subset C$ such that $x_k \to x$, and since C is a cone, $\alpha x_k \in C$ for all k. Furthermore, $\alpha x_k \to \alpha x$, implying that $\alpha x \in \mathrm{cl}(C)$. Hence, $\mathrm{cl}(C)$ is a cone.

(e) First we prove that $A \cdot C$ is a cone, where A is a linear transformation and $A \cdot C$ is the image of C under A. Let $z \in A \cdot C$ and let α be a positive scalar. Then, $Ax = z$ for some $x \in C$, and since C is a cone, $\alpha x \in C$. Because $A(\alpha x) = \alpha z$, the vector αz is in $A \cdot C$, showing that $A \cdot C$ is a cone.

Next we prove that the inverse image $A^{-1} \cdot C$ of C under A is a cone. Let $x \in A^{-1} \cdot C$ and let α be a positive scalar. Then $Ax \in C$, and since C is a cone, $\alpha A x \in C$. Thus, the vector $A(\alpha x)$ is in C, implying that $\alpha x \in A^{-1} \cdot C$, and showing that $A^{-1} \cdot C$ is a cone.

1.3 (Lower Semicontinuity under Composition)

(a) Let $\{x_k\} \subset \Re^n$ be a sequence of vectors converging to some $x \in \Re^n$. By continuity of f, it follows that $\big\{f(x_k)\big\} \subset \Re^m$ converges to $f(x) \in \Re^m$, so that by lower semicontinuity of g, we have

$$\liminf_{k\to\infty} g\big(f(x_k)\big) \geq g\big(f(x)\big).$$

Hence, h is lower semicontinuous.

(b) Assume, to arrive at a contradiction, that h is not lower semicontinuous at some $x \in \Re^n$. Then, there exists a sequence $\{x_k\} \subset \Re^n$ converging to x such that

$$\liminf_{k\to\infty} g\big(f(x_k)\big) < g\big(f(x)\big).$$

Let $\{x_k\}_\mathcal{K}$ be a subsequence attaining the above limit inferior, i.e.,

$$\lim_{k\to\infty,\ k\in\mathcal{K}} g\big(f(x_k)\big) = \liminf_{k\to\infty} g\big(f(x_k)\big) < g\big(f(x)\big). \tag{1.1}$$

Without loss of generality, we may assume that

$$g\big(f(x_k)\big) < g\big(f(x)\big), \qquad \forall\ k \in \mathcal{K}.$$

Since g is monotonically nondecreasing, it follows that

$$f(x_k) < f(x), \qquad \forall\ k \in \mathcal{K},$$

which together with the fact $\{x_k\}_\mathcal{K} \to x$ and the lower semicontinuity of f yields

$$f(x) \leq \liminf_{k\to\infty,\ k\in\mathcal{K}} f(x_k) \leq \limsup_{k\to\infty,\ k\in\mathcal{K}} f(x_k) \leq f(x),$$

showing that $\big\{f(x_k)\big\}_\mathcal{K} \to f(x)$. By our choice of the sequence $\{x_k\}_\mathcal{K}$ and by lower semicontinuity of g, it follows that

$$\lim_{k\to\infty,\ k\in\mathcal{K}} g\big(f(x_k)\big) = \liminf_{k\to\infty,\ k\in\mathcal{K}} g\big(f(x_k)\big) \geq g\big(f(x)\big),$$

contradicting Eq. (1.1). Hence, h is lower semicontinuous.

As an example showing that the assumption that g is monotonically nondecreasing is essential, consider the functions

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0, \end{cases}$$

and $g(x) = -x$. Then

$$g\big(f(x)\big) = \begin{cases} 0 & \text{if } x \leq 0, \\ -1 & \text{if } x > 0, \end{cases}$$

which is not lower semicontinuous at 0.

1.4 (Convexity under Composition)

(a) Let $x, y \in C$ and let $\alpha \in [0,1]$. Then we have

$$\begin{aligned} h\big(\alpha x + (1-\alpha)y\big) &= g\Big(f\big(\alpha x + (1-\alpha)y\big)\Big) \\ &\le g\big(\alpha f(x) + (1-\alpha)f(y)\big) \\ &\le \alpha g\big(f(x)\big) + (1-\alpha)g\big(f(y)\big) \\ &= \alpha h(x) + (1-\alpha)h(y), \end{aligned}$$

where the first inequality above follows from the convexity of f and the monotonicity of g, while the second inequality follows from the convexity of g. If g is monotonically increasing and f is strictly convex, then the first inequality in the preceding relation is strict whenever $x \neq y$ and $\alpha \in (0,1)$, showing that h is strictly convex.

(b) Let $x, y \in \Re^n$ and let $\alpha \in [0,1]$. Then, by the definitions of h and f, we have

$$\begin{aligned} h\big(\alpha x + (1-\alpha)y\big) &= g\Big(f\big(\alpha x + (1-\alpha)y\big)\Big) \\ &= g\Big(f_1\big(\alpha x + (1-\alpha)y\big), \ldots, f_m\big(\alpha x + (1-\alpha)y\big)\Big) \\ &\le g\big(\alpha f_1(x) + (1-\alpha)f_1(y), \ldots, \alpha f_m(x) + (1-\alpha)f_m(y)\big) \\ &= g\Big(\alpha\big(f_1(x), \ldots, f_m(x)\big) + (1-\alpha)\big(f_1(y), \ldots, f_m(y)\big)\Big) \\ &\le \alpha g\big(f_1(x), \ldots, f_m(x)\big) + (1-\alpha)g\big(f_1(y), \ldots, f_m(y)\big) \\ &= \alpha g\big(f(x)\big) + (1-\alpha)g\big(f(y)\big) \\ &= \alpha h(x) + (1-\alpha)h(y), \end{aligned}$$

where the first inequality follows by convexity of each f_i and monotonicity of g, while the second inequality follows by convexity of g.

1.5 (Examples of Convex Functions)

(a) It can be seen that f_1 is twice continuously differentiable over X and its Hessian matrix is given by

$$\nabla^2 f_1(x) = \frac{f_1(x)}{n^2}\begin{bmatrix} \frac{1-n}{x_1^2} & \frac{1}{x_1 x_2} & \cdots & \frac{1}{x_1 x_n} \\ \frac{1}{x_2 x_1} & \frac{1-n}{x_2^2} & \cdots & \frac{1}{x_2 x_n} \\ & & \vdots & \\ \frac{1}{x_n x_1} & \frac{1}{x_1 x_2} & \cdots & \frac{1-n}{x_n^2} \end{bmatrix}$$

for all $x = (x_1, \ldots, x_n) \in X$. From this, direct computation shows that for all $z = (z_1, \ldots, z_n) \in \Re^n$ and $x = (x_1, \ldots, x_n) \in X$, we have

$$z'\nabla^2 f_1(x)z = \frac{f_1(x)}{n^2}\left(\left(\sum_{i=1}^n \frac{z_i}{x_i}\right)^2 - n\sum_{i=1}^n \left(\frac{z_i}{x_i}\right)^2\right).$$

Note that this quadratic form is nonnegative for all $z \in \Re^n$ and $x \in X$, since $f_1(x) < 0$, and for any real numbers $\alpha_1, \ldots, \alpha_n$, we have

$$(\alpha_1 + \cdots + \alpha_n)^2 \leq n(\alpha_1^2 + \cdots + \alpha_n^2),$$

in view of the fact that $2\alpha_j\alpha_k \leq \alpha_j^2 + \alpha_k^2$. Hence, $\nabla^2 f_1(x)$ is positive semidefinite for all $x \in X$, and it follows from Prop. 1.2.6(a) that f_1 is convex.

(b) We show that the Hessian of f_2 is positive semidefinite at all $x \in \Re^n$. Let $\beta(x) = e^{x_1} + \cdots + e^{x_n}$. Then a straightforward calculation yields

$$z'\nabla^2 f_2(x)z = \frac{1}{\beta(x)^2} \sum_{i=1}^{n} \sum_{j=1}^{n} e^{(x_i + x_j)}(z_i - z_j)^2 \geq 0, \qquad \forall\ z \in \Re^n.$$

Hence by Prop. 1.2.6, f_2 is convex.

(b) The function $f_3(x) = \|x\|^p$ can be viewed as a composition $g\big(f(x)\big)$ of the scalar function $g(t) = t^p$ with $p \geq 1$ and the function $f(x) = \|x\|$. In this case, g is convex and monotonically increasing over the nonnegative axis, the set of values that f can take, while f is convex over $\Re^n$ (since any vector norm is convex, see the discussion preceding Prop. 1.2.4). Using Exercise 1.4, it follows that the function $f_3(x) = \|x\|^p$ is convex over $\Re^n$.

(c) The function $f_4(x) = \frac{1}{f(x)}$ can be viewed as a composition $g\big(h(x)\big)$ of the function $g(t) = -\frac{1}{t}$ for $t < 0$ and the function $h(x) = -f(x)$ for $x \in \Re^n$. In this case, the g is convex and monotonically increasing in the set $\{t \mid t < 0\}$, while h is convex over $\Re^n$. Using Exercise 1.4, it follows that the function $f_4(x) = \frac{1}{f(x)}$ is convex over $\Re^n$.

(d) The function $f_5(x) = \alpha f(x) + \beta$ can be viewed as a composition $g\big(f(x)\big)$ of the function $g(t) = \alpha t + \beta$, where $t \in \Re$, and the function $f(x)$ for $x \in \Re^n$. In this case, g is convex and monotonically increasing over $\Re$ (since $\alpha \geq 0$), while f is convex over $\Re^n$. Using Exercise 1.4, it follows that the function $f_5(x) = \alpha f(x) + \beta$ is convex over $\Re^n$.

(e) The function $f_6(x) = e^{\beta x'Ax}$ can be viewed as a composition $g\big(f(x)\big)$ of the function $g(t) = e^{\beta t}$ for $t \in \Re$ and the function $f(x) = x'Ax$ for $x \in \Re^n$. In this case, g is convex and monotonically increasing over $\Re$, while f is convex over $\Re^n$ (since A is positive semidefinite). Using Exercise 1.4, it follows that the function $f_6(x) = e^{\beta x'Ax}$ is convex over $\Re^n$.

(f) This part is straightforward using the definition of a convex function.

1.6 (Ascent/Descent Behavior of a Convex Function)

(a) Let x_1, x_2, x_3 be three scalars such that $x_1 < x_2 < x_3$. Then we can write x_2 as a convex combination of x_1 and x_3 as follows

$$x_2 = \frac{x_3 - x_2}{x_3 - x_1} x_1 + \frac{x_2 - x_1}{x_3 - x_1} x_3,$$

so that by convexity of f, we obtain

$$f(x_2) \le \frac{x_3 - x_2}{x_3 - x_1} f(x_1) + \frac{x_2 - x_1}{x_3 - x_1} f(x_3).$$

This relation and the fact

$$f(x_2) = \frac{x_3 - x_2}{x_3 - x_1} f(x_2) + \frac{x_2 - x_1}{x_3 - x_1} f(x_2),$$

imply that

$$\frac{x_3 - x_2}{x_3 - x_1}\big(f(x_2) - f(x_1)\big) \le \frac{x_2 - x_1}{x_3 - x_1}\big(f(x_3) - f(x_2)\big).$$

By multiplying the preceding relation with $x_3 - x_1$ and by dividing it with $(x_3 - x_2)(x_2 - x_1)$, we obtain

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \le \frac{f(x_3) - f(x_2)}{x_3 - x_2}.$$

(b) Let $\{x_k\}$ be an increasing scalar sequence, i.e., $x_1 < x_2 < x_3 < \cdots$. Then according to part (a), we have for all k

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \le \frac{f(x_3) - f(x_2)}{x_3 - x_2} \le \cdots \le \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}. \tag{1.2}$$

Since $\big(f(x_k) - f(x_{k-1})\big)/(x_k - x_{k-1})$ is monotonically nondecreasing, we have

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \to \gamma, \tag{1.3}$$

where γ is either a real number or ∞. Furthermore,

$$\frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} \le \gamma, \qquad \forall\ k. \tag{1.4}$$

We now show that γ is independent of the sequence $\{x_k\}$. Let $\{y_j\}$ be any increasing scalar sequence. For each j, choose x_{k_j} such that $y_j < x_{k_j}$ and $x_{k_1} < x_{k_2} < \cdots < x_{k_j}$, so that we have $y_j < y_{j+1} < x_{k_{j+1}} < x_{k_{j+2}}$. By part (a), it follows that

$$\frac{f(y_{j+1}) - f(y_j)}{y_{j+1} - y_j} \le \frac{f(x_{k_{j+2}}) - f(x_{k_{j+1}})}{x_{k_{j+2}} - x_{k_{j+1}}},$$

and letting $j \to \infty$ yields

$$\lim_{j\to\infty} \frac{f(y_{j+1}) - f(y_j)}{y_{j+1} - y_j} \le \gamma.$$

Similarly, by exchanging the roles of $\{x_k\}$ and $\{y_j\}$, we can show that

$$\lim_{j\to\infty} \frac{f(y_{j+1}) - f(y_j)}{y_{j+1} - y_j} \ge \gamma.$$

Thus the limit in Eq. (1.3) is independent of the choice for $\{x_k\}$, and Eqs. (1.2) and (1.4) hold for any increasing scalar sequence $\{x_k\}$.

We consider separately each of the three possibilities $\gamma < 0, \gamma = 0$, and $\gamma > 0$. First, suppose that $\gamma < 0$, and let $\{x_k\}$ be any increasing sequence. By using Eq. (1.4), we obtain

$$\begin{aligned} f(x_k) &= \sum_{j=1}^{k-1} \frac{f(x_{j+1}) - f(x_j)}{x_{j+1} - x_j}(x_{j+1} - x_j) + f(x_1) \\ &\leq \sum_{j=1}^{k-1} \gamma(x_{j+1} - x_j) + f(x_1) \\ &= \gamma(x_k - x_1) + f(x_1), \end{aligned}$$

and since $\gamma < 0$ and $x_k \to \infty$, it follows that $f(x_k) \to -\infty$. To show that f decreases monotonically, pick any x and y with $x < y$, and consider the sequence $x_1 = x$, $x_2 = y$, and $x_k = y + k$ for all $k \geq 3$. By using Eq. (1.4) with $k = 1$, we have

$$\frac{f(y) - f(x)}{y - x} \leq \gamma < 0,$$

so that $f(y) - f(x) < 0$. Hence f decreases monotonically to $-\infty$, corresponding to case (1).

Suppose now that $\gamma = 0$, and let $\{x_k\}$ be any increasing sequence. Then, by Eq. (1.4), we have $f(x_{k+1}) - f(x_k) \leq 0$ for all k. If $f(x_{k+1}) - f(x_k) < 0$ for all k, then f decreases monotonically. To show this, pick any x and y with $x < y$, and consider a new sequence given by $y_1 = x$, $y_2 = y$, and $y_k = x_{K+k-3}$ for all $k \geq 3$, where K is large enough so that $y < x_K$. By using Eqs. (1.2) and (1.4) with $\{y_k\}$, we have

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(x_{K+1}) - f(x_K)}{x_{K+1} - x_K} < 0,$$

implying that $f(y) - f(x) < 0$. Hence f decreases monotonically, and it may decrease to $-\infty$ or to a finite value, corresponding to cases (1) or (2), respectively.

If for some K we have $f(x_{K+1}) - f(x_K) = 0$, then by Eqs. (1.2) and (1.4) where $\gamma = 0$, we obtain $f(x_k) = f(x_K)$ for all $k \geq K$. To show that f stays at the value $f(x_K)$ for all $x \geq x_K$, choose any x such that $x > x_K$, and define $\{y_k\}$ as $y_1 = x_K$, $y_2 = x$, and $y_k = x_{N+k-3}$ for all $k \geq 3$, where N is large enough so that $x < x_N$. By using Eqs. (1.2) and (1.4) with $\{y_k\}$, we have

$$\frac{f(x) - f(x_K)}{x - x_K} \leq \frac{f(x_N) - f(x)}{x_N - x} \leq 0,$$

so that $f(x) \leq f(x_K)$ and $f(x_N) \leq f(x)$. Since $f(x_K) = f(x_N)$, we have $f(x) = f(x_K)$. Hence $f(x) = f(x_K)$ for all $x \geq x_K$, corresponding to case (3).

Finally, suppose that $\gamma > 0$, and let $\{x_k\}$ be any increasing sequence. Since $\big(f(x_k) - f(x_{k-1})\big)/(x_k - x_{k-1})$ is nondecreasing and tends to γ [cf. Eqs. (1.3) and (1.4)], there is a positive integer K and a positive scalar ϵ with $\epsilon < \gamma$ such that

$$\epsilon \leq \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}, \qquad \forall\, k \geq K. \tag{1.5}$$

Therefore, for all $k > K$

$$f(x_k) = \sum_{j=K}^{k-1} \frac{f(x_{j+1}) - f(x_j)}{x_{j+1} - x_j}(x_{j+1} - x_j) + f(x_K) \geq \epsilon(x_k - x_K) + f(x_K),$$

implying that $f(x_k) \to \infty$. To show that $f(x)$ increases monotonically to ∞ for all $x \geq x_K$, pick any $x < y$ satisfying $x_K < x < y$, and consider a sequence given by $y_1 = x_K$, $y_2 = x$, $y_3 = y$, and $y_k = x_{N+k-4}$ for $k \geq 4$, where N is large enough so that $y < x_N$. By using Eq. (1.5) with $\{y_k\}$, we have

$$\epsilon \leq \frac{f(y) - f(x)}{y - x}.$$

Thus $f(x)$ increases monotonically to ∞ for all $x \geq x_K$, corresponding to case (4) with $\overline{x} = x_K$.

1.7 (Characterization of Differentiable Convex Functions)

If f is convex, then by Prop. 1.2.5(a), we have

$$f(y) \geq f(x) + \nabla f(x)'(y - x), \qquad \forall\, x, y \in C.$$

By exchanging the roles of x and y in this relation, we obtain

$$f(x) \geq f(y) + \nabla f(y)'(x - y), \qquad \forall\, x, y \in C,$$

and by adding the preceding two inequalities, it follows that

$$\big(\nabla f(y) - \nabla f(x)\big)'(x - y) \geq 0. \tag{1.6}$$

Conversely, let Eq. (1.6) hold, and let x and y be two points in C. Define the function $h : \Re \mapsto \Re$ by

$$h(t) = f\big(x + t(y - x)\big).$$

Consider some $t, t' \in [0, 1]$ such that $t < t'$. By convexity of C, we have that $x + t(y - x)$ and $x + t'(y - x)$ belong to C. Using the chain rule and Eq. (1.6), we have

$$\begin{aligned}
&\left(\frac{dh(t')}{dt} - \frac{dh(t)}{dt}\right)(t' - t) \\
&\qquad = \Big(\nabla f\big(x + t'(y - x)\big) - \nabla f\big(x + t(y - x)\big)\Big)'(y - x)(t' - t) \\
&\qquad \geq 0.
\end{aligned}$$

Thus, dh/dt is nondecreasing on $[0, 1]$ and for any $t \in (0, 1)$, we have

$$\frac{h(t) - h(0)}{t} = \frac{1}{t}\int_0^t \frac{dh(\tau)}{d\tau}\, d\tau \leq h(t) \leq \frac{1}{1 - t}\int_t^1 \frac{dh(\tau)}{d\tau}\, d\tau = \frac{h(1) - h(t)}{1 - t}.$$

Equivalently,

$$th(1) + (1 - t)h(0) \geq h(t),$$

and from the definition of h, we obtain

$$tf(y) + (1 - t)f(x) \geq f\big(ty + (1 - t)x\big).$$

Since this inequality has been proved for arbitrary $t \in [0, 1]$ and $x, y \in C$, we conclude that f is convex.

1.8 (Characterization of Twice Continuously Differentiable Convex Functions)

Suppose that $f : \Re^n \mapsto \Re$ is convex over C. We first show that for all $x \in \mathrm{ri}(C)$ and $y \in S$, we have $y'\nabla^2 f(x)y \geq 0$. Assume to arrive at a contradiction, that there exists some $\overline{x} \in \mathrm{ri}(C)$ such that for some $y \in S$, we have

$$y'\nabla^2 f(\overline{x})y < 0.$$

Without loss of generality, we may assume that $\|y\| = 1$. Using the continuity of $\nabla^2 f$, we see that there is an open ball $B(\overline{x}, \epsilon)$ centered at $\bar{x}$ with radius ϵ such that $B(\overline{x}, \epsilon) \cap \mathrm{aff}(C) \subset C$ [since $\overline{x} \in \mathrm{ri}(C)$], and

$$y'\nabla^2 f(x)y < 0, \qquad \forall\, x \in B(\overline{x}, \epsilon). \tag{1.7}$$

By Prop. 1.1.13(a), for all positive scalars α with $\alpha < \epsilon$, we have

$$f(\bar{x} + \alpha y) = f(\bar{x}) + \alpha \nabla f(\bar{x})'y + \frac{1}{2} y'\nabla^2 f(\bar{x} + \bar{\alpha} y)y,$$

for some $\bar{\alpha} \in [0, \alpha]$. Furthermore, $\|(\overline{x} + \overline{\alpha} y) - \overline{x}\| \leq \epsilon$ [since $\|y\| = 1$ and $\bar{\alpha} < \epsilon$]. Hence, from Eq. (1.7), it follows that

$$f(\bar{x} + \alpha y) < f(\bar{x}) + \alpha \nabla f(\bar{x})'y, \qquad \forall\, \alpha \in [0, \epsilon).$$

On the other hand, by the choice of ϵ and the assumption that $y \in S$, the vectors $\bar{x} + \alpha y$ are in C for all α with $\alpha \in [0, \epsilon)$, which is a contradiction in view of the convexity of f over C. Hence, we have $y'\nabla^2 f(x)y \geq 0$ for all $y \in S$ and all $x \in \mathrm{ri}(C)$.

Next, let $\overline{x}$ be a point in C that is not in the relative interior of C. Then, by the Line Segment Principle, there is a sequence $\{x_k\} \subset \mathrm{ri}(C)$ such that $x_k \to \overline{x}$. As seen above, $y'\nabla^2 f(x_k)y \geq 0$ for all $y \in S$ and all k, which together with the continuity of $\nabla^2 f$ implies that

$$y'\nabla^2 f(\overline{x})y = \lim_{k \to \infty} y'\nabla^2 f(x_k)y \geq 0, \qquad \forall\, y \in S.$$

It follows that $y'\nabla^2 f(x)y \geq 0$ for all $x \in C$ and $y \in S$.

Conversely, assume that $y'\nabla^2 f(x)y \geq 0$ for all $x \in C$ and $y \in S$. By Prop. 1.1.13(a), for all $x, z \in C$ we have

$$f(z) = f(x) + (z - x)'\nabla f(x) + \tfrac{1}{2}(z - x)'\nabla^2 f\big(x + \alpha(z - x)\big)(z - x)$$

for some $\alpha \in [0, 1]$. Since $x, z \in C$, we have that $(z - x) \in S$, and using the convexity of C and our assumption, it follows that

$$f(z) \geq f(x) + (z - x)'\nabla f(x), \qquad \forall\, x, z \in C.$$

From Prop. 1.2.5(a), we conclude that f is convex over C.

1.9 (Strong Convexity)

(a) Fix some $x, y \in \Re^n$ such that $x \neq y$, and define the function $h : \Re \mapsto \Re$ by $h(t) = f\big(x + t(y - x)\big)$. Consider scalars t and s such that $t < s$. Using the chain rule and the equation

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) \geq \alpha \|x - y\|^2, \qquad \forall\, x, y \in \Re^n, \tag{1.8}$$

for some $\alpha > 0$, we have

$$\begin{aligned}
&\left(\frac{dh(s)}{dt} - \frac{dh(t)}{dt}\right)(s - t) \\
&\qquad = \Big(\nabla f\big(x + s(y - x)\big) - \nabla f\big(x + t(y - x)\big)\Big)'(y - x)(s - t) \\
&\qquad \geq \alpha (s - t)^2 \|x - y\|^2 > 0.
\end{aligned}$$

Thus, dh/dt is strictly increasing and for any $t \in (0, 1)$, we have

$$\frac{h(t) - h(0)}{t} = \frac{1}{t}\int_0^t \frac{dh(\tau)}{d\tau}\, d\tau < \frac{1}{1 - t}\int_t^1 \frac{dh(\tau)}{d\tau}\, d\tau = \frac{h(1) - h(t)}{1 - t}.$$

Equivalently, $th(1) + (1 - t)h(0) > h(t)$. The definition of h yields $tf(y) + (1 - t)f(x) > f\big(ty + (1 - t)x\big)$. Since this inequality has been proved for arbitrary $t \in (0, 1)$ and $x \neq y$, we conclude that f is strictly convex.

(b) Suppose now that f is twice continuously differentiable and Eq. (1.8) holds. Let c be a scalar. We use Prop. 1.1.13(b) twice to obtain

$$f(x + cy) = f(x) + cy'\nabla f(x) + \frac{c^2}{2} y' \nabla^2 f(x + tcy) y,$$

and

$$f(x) = f(x + cy) - cy'\nabla f(x + cy) + \frac{c^2}{2} y' \nabla^2 f(x + scy) y,$$

for some t and s belonging to $[0, 1]$. Adding these two equations and using Eq. (1.8), we obtain

$$\frac{c^2}{2} y' \big(\nabla^2 f(x + scy) + \nabla^2 f(x + tcy)\big) y = \big(\nabla f(x + cy) - \nabla f(x)\big)'(cy) \geq \alpha c^2 \|y\|^2.$$

We divide both sides by c^2 and then take the limit as $c \to 0$ to conclude that $y'\nabla^2 f(x) y \geq \alpha \|y\|^2$. Since this inequality is valid for every $y \in \Re^n$, it follows that $\nabla^2 f(x) - \alpha I$ is positive semidefinite.

For the converse, assume that $\nabla^2 f(x) - \alpha I$ is positive semidefinite for all $x \in \Re^n$. Consider the function $g : \Re \mapsto \Re$ defined by

$$g(t) = \nabla f\big(tx + (1 - t)y\big)'(x - y).$$

Using the Mean Value Theorem (Prop. 1.1.12), we have

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) = g(1) - g(0) = \frac{dg(t)}{dt}$$

for some $t \in [0,1]$. On the other hand,

$$\frac{dg(t)}{dt} = (x-y)'\nabla^2 f\big(tx + (1-t)y\big)(x-y) \geq \alpha\|x-y\|^2,$$

where the last inequality holds because $\nabla^2 f\big(tx+(1-t)y\big)-\alpha I$ is positive semidefinite. Combining the last two relations, it follows that f is strongly convex with coefficient α.

1.10 (Posynomials)

(a) Consider the following posynomial for which we have $n = m = 1$ and $\beta = \frac{1}{2}$,

$$g(y) = y^{\frac{1}{2}}, \qquad \forall\ y > 0.$$

This function is not convex.

(b) Consider the following change of variables, where we set

$$f(x) = \ln\big(g(y_1,\ldots,y_n)\big), \qquad b_i = \ln \beta_i,\ \forall\ i, \qquad x_j = \ln y_j,\ \forall\ j.$$

With this change of variables, $f(x)$ can be written as

$$f(x) = \ln\left(\sum_{i=1}^{m} e^{b_i + a_{i1}x_1 + \cdots + a_{in}x_n}\right).$$

Note that $f(x)$ can also be represented as

$$f(x) = \ln\exp(Ax+b), \qquad \forall\ x \in \Re^n,$$

where $\ln\exp(z) = \ln\big(e^{z_1} + \cdots + e^{z_m}\big)$ for all $z \in \Re^m$, A is an $m \times n$ matrix with entries a_{ij}, and $b \in \Re^m$ is a vector with components b_i. Let $f_2(z) = \ln(e^{z_1} + \cdots + e^{z_m})$. This function is convex by Exercise 1.5(b). With this identification, $f(x)$ can be viewed as the composition $f(x) = f_2(Ax+b)$, which is convex by Exercise 1.5(g).

(c) Consider the function $g : \Re^n \mapsto \Re$ of the form

$$g(y) = g_1(y)^{\gamma_1} \cdots g_r(y)^{\gamma_r},$$

where g_k is a posynomial and $\gamma_k > 0$ for all k. Using a change of variables similar to part (b), we see that we can represent the function $f(x) = \ln g(y)$ as

$$f(x) = \sum_{k=1}^{r} \gamma_k \ln\exp(A_k x + b_k),$$

with the matrix A_k and the vector b_k being associated with the posynomial g_k for each k. Since $f(x)$ is a linear combination of convex functions with nonnegative coefficients [part (b)], it follows from Prop. 1.2.4(a) that $f(x)$ is convex.

1.11 (Arithmetic-Geometric Mean Inequality)

Consider the function $f(x) = -\ln(x)$. Since $\nabla^2 f(x) = 1/x^2 > 0$ for all $x > 0$, the function $-\ln(x)$ is strictly convex over $(0, \infty)$. Therefore, for all positive scalars $x_1, \ldots, x_n \in (0, \infty)$ and $\alpha_1, \ldots \alpha_n$ with $\sum_{i=1}^n \alpha_i = 1$, we have

$$-\ln(\alpha_1 x_1 + \cdots + \alpha_n x_n) \leq -\alpha_1 \ln(x_1) - \cdots - \alpha_n \ln(x_n),$$

which is equivalent to

$$e^{\ln(\alpha_1 x_1 + \cdots + \alpha_n x_n)} \geq e^{\alpha_1 \ln(x_1) + \cdots + \alpha_n \ln(x_n)} = e^{\alpha_1 \ln(x_1)} \cdots e^{\alpha_n \ln(x_n)},$$

or

$$\alpha_1 x_1 + \cdots + \alpha_n x_n \geq x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

as desired. Since $-\ln(x)$ is strictly convex, the above inequality is satisfied with equality if and only if $x_1, \ldots, x_n$ are all equal.

1.12 (Young and Holder Inequalities)

According to Exercise 1.11, we have

$$u^{\frac{1}{p}} v^{\frac{1}{q}} \leq \frac{u}{p} + \frac{v}{q}, \qquad \forall\ u > 0, \quad \forall\ v > 0,$$

where $1/p + 1/q = 1$, $p > 0$, and $q > 0$. The above relation also holds if $u = 0$ or $v = 0$. By setting $u = x^p$ and $v = y^q$, we obtain Young's inequality

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}, \qquad \forall\ x \geq 0, \quad \forall\ y \geq 0.$$

To show Holder's inequality, note that it holds if $x_1 = \cdots = x_n = 0$ or $y_1 = \cdots = y_n = 0$. If $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are such that $(x_1, \ldots, x_n) \neq 0$ and $(y_1, \ldots, y_n) \neq 0$, then by using

$$x = \frac{|x_i|}{\left(\sum_{j=1}^n |x_j|^p\right)^{1/p}} \quad \text{and} \quad y = \frac{|y_i|}{\left(\sum_{j=1}^n |y_j|^q\right)^{1/q}}$$

in Young's inequality, we have for all $i = 1, \ldots, n$,

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

By adding these inequalities over $i = 1, \ldots, n$, we obtain

$$\frac{\sum_{i=1}^n |x_i| \cdot |y_i|}{\left(\sum_{j=1}^n |x_j|^p\right)^{1/p} \left(\sum_{j=1}^n |y_j|^q\right)^{1/q}} \leq \frac{1}{p} + \frac{1}{q} = 1,$$

which implies Holder's inequality.

1.13

Let (x, w) and (y, v) be two vectors in $\text{epi}(f)$. Then $f(x) \leq w$ and $f(y) \leq v$, implying that there exist sequences $\big\{(x, \overline{w}_k)\big\} \subset C$ and $\big\{(y, \overline{v}_k)\big\} \subset C$ such that for all k,

$$\overline{w}_k \leq w + \frac{1}{k}, \qquad \overline{v}_k \leq v + \frac{1}{k}.$$

By the convexity of C, we have for all $\alpha \in [0, 1]$ and all k,

$$\big(\alpha x + (1 - \alpha y),\ \alpha \overline{w}_k + (1 - \alpha)\overline{v}_k\big) \in C,$$

so that for all k,

$$f\big(\alpha x + (1 - \alpha)y\big) \leq \alpha \overline{w}_k + (1 - \alpha)\overline{v}_k \leq \alpha w + (1 - \alpha)v + \frac{1}{k}.$$

Taking the limit as $k \to \infty$, we obtain

$$f\big(\alpha x + (1 - \alpha)y\big) \leq \alpha w + (1 - \alpha)v,$$

so that $\alpha(x, w) + (1 - \alpha)(y, v) \in \text{epi}(f)$. Hence, $\text{epi}(f)$ is convex, implying that f is convex.

1.14

The elements of X belong to $\text{conv}(X)$, so all their convex combinations belong to $\text{conv}(X)$ since $\text{conv}(X)$ is a convex set. On the other hand, consider any two convex combinations of elements of X, $x = \lambda_1 x_1 + \cdots + \lambda_m x_m$ and $y = \mu_1 y_1 + \cdots + \mu_r y_r$, where $x_i \in X$ and $y_j \in X$. The vector

$$(1 - \alpha)x + \alpha y = (1 - \alpha)\left(\lambda_1 x_1 + \cdots + \lambda_m x_m\right) + \alpha\left(\mu_1 y_1 + \cdots + \mu_r y_r\right),$$

where $0 \leq \alpha \leq 1$, is another convex combination of elements of X. Thus, the set of convex combinations of elements of X is itself a convex set. It contains X, and is contained in $\text{conv}(X)$, so it must coincide with $\text{conv}(X)$.

1.15

Let $y \in \text{cone}(C)$. If $y = 0$, then $y \in \cup_{x \in C}\{\gamma x \mid \gamma \geq 0\}$ and we are done. If $y \neq 0$, then by definition of $\text{cone}(C)$, we have

$$y = \sum_{i=1}^{m} \lambda_i x_i,$$

for some positive integer m, nonnegative scalars λ_i, and vectors $x_i \in C$. Since $y \neq 0$, we cannot have all λ_i equal to zero, implying that $\sum_{i=1}^{m} \lambda_i > 0$. Because $x_i \in C$ for all i and C is convex, the vector

$$x = \sum_{i=1}^{m} \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} x_i$$

belongs to C. For this vector, we have

$$y = \left(\sum_{i=1}^m \lambda_i \right) x,$$

with $\sum_{i=1}^m \lambda_i > 0$, implying that $y \in \cup_{x \in C} \big\{ \gamma x \mid \gamma \geq 0 \big\}$ and showing that

$$\mathrm{cone}(C) \subset \cup_{x \in C} \{ \gamma x \mid \gamma \geq 0 \}.$$

The reverse inclusion follows directly from the definition of cone(C).

1.16 (Convex Cones)

(a) Let $x \in C$ and let λ be a positive scalar. Then

$$a_i'(\lambda x) = \lambda a_i' x \leq 0, \qquad \forall\, i \in I,$$

showing that $\lambda x \in C$ and that C is a cone. Let $x, y \in C$ and let $\lambda \in [0, 1]$. Then

$$a_i'\big(\lambda x + (1 - \lambda) y\big) = \lambda a_i' x + (1 - \lambda) a_i' y \leq 0, \qquad \forall\, i \in I,$$

showing that $\big(\lambda x + (1-\lambda) y\big) \in C$ and that C is convex. Let a sequence $\{x_k\} \subset C$ converge to some $\bar{x} \in \Re^n$. Then

$$a_i' \bar{x} = \lim_{k \to \infty} a_i' x_k \leq 0, \qquad \forall\, i \in I,$$

showing that $\bar{x} \in C$ and that C is closed.

(b) Let C be a cone such that $C + C \subset C$, and let $x, y \in C$ and $\alpha \in [0, 1]$. Then since C is a cone, $\alpha x \in C$ and $(1 - \alpha) y \in C$, so that $\alpha x + (1 - \alpha) y \in C + C \subset C$, showing that C is convex. Conversely, let C be a convex cone and let $x, y \in C$. Then, since C is a cone, $2x \in C$ and $2y \in C$, so that by the convexity of C, $x + y = \frac{1}{2}(2x + 2y) \in C$, showing that $C + C \subset C$.

(c) First we prove that $C_1 + C_2 \subset \mathrm{conv}(C_1 \cup C_2)$. Choose any $x \in C_1 + C_2$. Since $C_1 + C_2$ is a cone [see Exercise 1.2(c)], the vector $2x$ is in $C_1 + C_2$, so that $2x = x_1 + x_2$ for some $x_1 \in C_1$ and $x_2 \in C_2$. Therefore,

$$x = \frac{1}{2} x_1 + \frac{1}{2} x_2,$$

showing that $x \in \mathrm{conv}(C_1 \cup C_2)$.

Next, we show that $\mathrm{conv}(C_1 \cup C_2) \subset C_1 + C_2$. Since $0 \in C_1$ and $0 \in C_2$, it follows that

$$C_i = C_i + 0 \subset C_1 + C_2, \qquad i = 1, 2,$$

implying that

$$C_1 \cup C_2 \subset C_1 + C_2.$$

By taking the convex hull of both sides in the above inclusion and by using the convexity of $C_1 + C_2$, we obtain

$$\mathrm{conv}(C_1 \cup C_2) \subset \mathrm{conv}(C_1 + C_2) = C_1 + C_2.$$

We finally show that

$$C_1 \cap C_2 = \bigcup_{\alpha \in [0,1]} \big(\alpha C_1 \cap (1-\alpha) C_2\big).$$

We claim that for all α with $0 < \alpha < 1$, we have

$$\alpha C_1 \cap (1-\alpha) C_2 = C_1 \cap C_2.$$

Indeed, if $x \in C_1 \cap C_2$, it follows that $x \in C_1$ and $x \in C_2$. Since C_1 and C_2 are cones and $0 < \alpha < 1$, we have $x \in \alpha C_1$ and $x \in (1-\alpha) C_2$. Conversely, if $x \in \alpha C_1 \cap (1-\alpha) C_2$, we have

$$\frac{x}{\alpha} \in C_1,$$

and

$$\frac{x}{(1-\alpha)} \in C_2.$$

Since C_1 and C_2 are cones, it follows that $x \in C_1$ and $x \in C_2$, so that $x \in C_1 \cap C_2$.

If $\alpha = 0$ or $\alpha = 1$, we obtain

$$\alpha C_1 \cap (1-\alpha) C_2 = \{0\} \subset C_1 \cap C_2,$$

since C_1 and C_2 contain the origin. Thus, the result follows.

1.17

By Exercise 1.14, C is the set of all convex combinations $x = \alpha_1 y_1 + \cdots + \alpha_m y_m$, where m is a positive integer, and the vectors $y_1, \ldots, y_m$ belong to the union of the sets C_i. Actually, we can get C just by taking those combinations in which the vectors are taken from different sets C_i. Indeed, if two of the vectors, y_1 and y_2 belong to the same C_i, then the term $\alpha_1 y_1 + \alpha_2 y_2$ can be replaced by αy, where $\alpha = \alpha_1 + \alpha_2$ and

$$y = (\alpha_1/\alpha) y_1 + (\alpha_2/\alpha) y_2 \in C_i.$$

Thus, C is the union of the vector sums of the form

$$\alpha_1 C_{i_1} + \cdots + \alpha_m C_{i_m},$$

with

$$\alpha_i \geq 0,\ \forall\ i = 1, \ldots, m, \quad \sum_{i=1}^m \alpha_i = 1,$$

and the indices $i_1, \ldots, i_m$ are all different, proving our claim.

1.18 (Convex Hulls, Affine Hulls, and Generated Cones)

(a) We first show that X and $\mathrm{cl}(X)$ have the same affine hull. Since $X \subset \mathrm{cl}(X)$, there holds

$$\mathrm{aff}(X) \subset \mathrm{aff}\big(\mathrm{cl}(X)\big).$$

Conversely, because $X \subset \mathrm{aff}(X)$ and $\mathrm{aff}(X)$ is closed, we have $\mathrm{cl}(X) \subset \mathrm{aff}(X)$, implying that

$$\mathrm{aff}\big(\mathrm{cl}(X)\big) \subset \mathrm{aff}(X).$$

We now show that X and $\mathrm{conv}(X)$ have the same affine hull. By using a translation argument if necessary, we assume without loss of generality that X contains the origin, so that both $\mathrm{aff}(X)$ and $\mathrm{aff}\big(\mathrm{conv}(X)\big)$ are subspaces. Since $X \subset \mathrm{conv}(X)$, evidently $\mathrm{aff}(X) \subset \mathrm{aff}\big(\mathrm{conv}(X)\big)$. To show the reverse inclusion, let the dimension of $\mathrm{aff}\big(\mathrm{conv}(X)\big)$ be m, and let $x_1, \ldots, x_m$ be linearly independent vectors in $\mathrm{conv}(X)$ that span $\mathrm{aff}\big(\mathrm{conv}(X)\big)$. Then every $x \in \mathrm{aff}\big(\mathrm{conv}(X)\big)$ is a linear combination of the vectors $x_1, \ldots, x_m$, i.e., there exist scalars $\beta_1, \ldots, \beta_m$ such that

$$x = \sum_{i=1}^{m} \beta_i x_i.$$

By the definition of convex hull, each x_i is a convex combination of vectors in X, so that x is a linear combination of vectors in X, implying that $x \in \mathrm{aff}(X)$. Hence, $\mathrm{aff}\big(\mathrm{conv}(X)\big) \subset \mathrm{aff}(X)$.

(b) Since $X \subset \mathrm{conv}(X)$, clearly $\mathrm{cone}(X) \subset \mathrm{cone}\big(\mathrm{conv}(X)\big)$. Conversely, let $x \in \mathrm{cone}\big(\mathrm{conv}(X)\big)$. Then x is a nonnegative combination of some vectors in $\mathrm{conv}(X)$, i.e., for some positive integer p, vectors $x_1, \ldots, x_p \in \mathrm{conv}(X)$, and nonnegative scalars $\alpha_1, \ldots, \alpha_p$, we have

$$x = \sum_{i=1}^{p} \alpha_i x_i.$$

Each x_i is a convex combination of some vectors in X, so that x is a nonnegative combination of some vectors in X, implying that $x \in \mathrm{cone}(X)$. Hence $\mathrm{cone}\big(\mathrm{conv}(X)\big) \subset \mathrm{cone}(X)$.

(c) Since $\mathrm{conv}(X)$ is the set of all convex combinations of vectors in X, and $\mathrm{cone}(X)$ is the set of all nonnegative combinations of vectors in X, it follows that $\mathrm{conv}(X) \subset \mathrm{cone}(X)$. Therefore

$$\mathrm{aff}\big(\mathrm{conv}(X)\big) \subset \mathrm{aff}\big(\mathrm{cone}(X)\big).$$

As an example showing that the above inclusion can be strict, consider the set $X = \big\{(1,1)\big\}$ in $\Re^2$. Then $\mathrm{conv}(X) = X$, so that

$$\mathrm{aff}\big(\mathrm{conv}(X)\big) = X = \big\{(1,1)\big\},$$

and the dimension of $\mathrm{conv}(X)$ is zero. On the other hand, $\mathrm{cone}(X) = \big\{(\alpha, \alpha) \mid \alpha \geq 0\big\}$, so that

$$\mathrm{aff}\big(\mathrm{cone}(X)\big) = \big\{(x_1, x_2) \mid x_1 = x_2\big\},$$

and the dimension of cone(X) is one.

(d) In view of parts (a) and (c), it suffices to show that

$$\text{aff}\big(\text{cone}(X)\big) \subset \text{aff}\big(\text{conv}(X)\big) = \text{aff}(X).$$

It is always true that $0 \in \text{cone}(X)$, so $\text{aff}\big(\text{cone}(X)\big)$ is a subspace. Let the dimension of $\text{aff}\big(\text{cone}(X)\big)$ be m, and let $x_1, \ldots, x_m$ be linearly independent vectors in cone(X) that span $\text{aff}\big(\text{cone}(X)\big)$. Since every vector in $\text{aff}\big(\text{cone}(X)\big)$ is a linear combination of $x_1, \ldots, x_m$, and since each x_i is a nonnegative combination of some vectors in X, it follows that every vector in $\text{aff}\big(\text{cone}(X)\big)$ is a linear combination of some vectors in X. In view of the assumption that $0 \in \text{conv}(X)$, the affine hull of conv(X) is a subspace, which implies by part (a) that the affine hull of X is a subspace. Hence, every vector in $\text{aff}\big(\text{cone}(X)\big)$ belongs to aff(X), showing that $\text{aff}\big(\text{cone}(X)\big) \subset \text{aff}(X)$.

1.19

By definition, $f(x)$ is the infimum of the values of w such that $(x, w) \in C$, where C is the convex hull of the union of nonempty convex sets epi(f_i). By Exercise 1.17, $(x, w) \in C$ if and only if (x, w) can be expressed as a convex combination of the form

$$(x, w) = \sum_{i \in \overline{I}} \alpha_i (x_i, w_i) = \left(\sum_{i \in \overline{I}} \alpha_i x_i, \sum_{i \in \overline{I}} \alpha_i w_i \right),$$

where $\overline{I} \subset I$ is a finite set and $(x_i, w_i) \in \text{epi}(f_i)$ for all $i \in \overline{I}$. Thus, $f(x)$ can be expressed as

$$f(x) = \inf \left\{ \sum_{i \in \overline{I}} \alpha_i w_i \;\middle|\; (x, w) = \sum_{i \in \overline{I}} \alpha_i (x_i, w_i), \right.$$
$$\left. (x_i, w_i) \in \text{epi}(f_i),\ \alpha_i \geq 0,\ \forall\, i \in \overline{I},\ \sum_{i \in \overline{I}} \alpha_i = 1 \right\}.$$

Since the set $\big\{\big(x_i, f_i(x_i)\big) \mid x_i \in \Re^n\big\}$ is contained in epi(f_i), we obtain

$$f(x) \leq \inf \left\{ \sum_{i \in \overline{I}} \alpha_i f_i(x_i) \;\middle|\; x = \sum_{i \in \overline{I}} \alpha_i x_i,\ x_i \in \Re^n,\ \alpha_i \geq 0,\ \forall\, i \in \overline{I},\ \sum_{i \in \overline{I}} \alpha_i = 1 \right\}.$$

On the other hand, by the definition of epi(f_i), for each $(x_i, w_i) \in \text{epi}(f_i)$ we have $w_i \geq f_i(x_i)$, implying that

$$f(x) \geq \inf \left\{ \sum_{i \in \overline{I}} \alpha_i f_i(x_i) \;\middle|\; x = \sum_{i \in \overline{I}} \alpha_i x_i,\ x_i \in \Re^n,\ \alpha_i \geq 0,\ \forall\, i \in \overline{I},\ \sum_{i \in \overline{I}} \alpha_i = 1 \right\}.$$

By combining the last two relations, we obtain

$$f(x) = \inf \left\{ \sum_{i \in \overline{I}} \alpha_i f_i(x_i) \;\middle|\; x = \sum_{i \in \overline{I}} \alpha_i x_i,\ x_i \in \Re^n,\ \alpha_i \geq 0,\ \forall\, i \in \overline{I},\ \sum_{i \in \overline{I}} \alpha_i = 1 \right\},$$

where the infimum is taken over all representations of x as a convex combination of elements x_i such that only finitely many coefficients α_i are nonzero.

1.20 (Convexification of Nonconvex Functions)

(a) Since $\mathrm{conv}\big(\mathrm{epi}(f)\big)$ is a convex set, it follows from Exercise 1.13 that F is convex over $\mathrm{conv}(X)$. By Caratheodory's Theorem, it can be seen that $\mathrm{conv}\big(\mathrm{epi}(f)\big)$ is the set of all convex combinations of elements of $\mathrm{epi}(f)$, so that

$$\begin{aligned} F(x) = \inf \Bigg\{ \sum_i \alpha_i w_i \;\Bigg|\; & (x, w) = \sum_i \alpha_i (x_i, w_i), \\ & (x_i, w_i) \in \mathrm{epi}(f),\ \alpha_i \geq 0,\ \sum_i \alpha_i = 1 \Bigg\}, \end{aligned}$$

where the infimum is taken over all representations of x as a convex combination of elements of X. Since the set $\big\{\big(z, f(z)\big) \mid z \in X\big\}$ is contained in $\mathrm{epi}(f)$, we obtain

$$F(x) \leq \inf \left\{ \sum_i \alpha_i f(x_i) \;\middle|\; x = \sum_i \alpha_i x_i,\ x_i \in X,\ \alpha_i \geq 0,\ \sum_i \alpha_i = 1 \right\}.$$

On the other hand, by the definition of $\mathrm{epi}(f)$, for each $(x_i, w_i) \in \mathrm{epi}(f)$ we have $w_i \geq f(x_i)$, implying that

$$\begin{aligned} F(x) \geq \inf \Bigg\{ \sum_i \alpha_i f(x_i) \;\Bigg|\; & (x, w) = \sum_i \alpha_i (x_i, w_i), \\ & (x_i, w_i) \in \mathrm{epi}(f),\ \alpha_i \geq 0,\ \sum_i \alpha_i = 1 \Bigg\}, \\ = \inf \Bigg\{ \sum_i \alpha_i f(x_i) \;\Bigg|\; & x = \sum_i \alpha_i x_i,\ x_i \in X,\ \alpha_i \geq 0,\ \sum_i \alpha_i = 1 \Bigg\}, \end{aligned}$$

which combined with the preceding inequality implies the desired relation.

(b) By using part (a), we have for every $x \in X$

$$F(x) \leq f(x),$$

since $f(x)$ corresponds to the value of the function $\sum_i \alpha_i f(x_i)$ for a particular representation of x as a finite convex combination of elements of X, namely $x = 1 \cdot x$. Therefore, we have

$$\inf_{x \in X} F(x) \le \inf_{x \in X} f(x),$$

and since $X \subset \mathrm{conv}(X)$, it follows that

$$\inf_{x \in \mathrm{conv}(X)} F(x) \le \inf_{x \in X} f(x).$$

Let $f^* = \inf_{x \in X} f(x)$. If $\inf_{x \in \mathrm{conv}(X)} F(x) < f^*$, then there exists $z \in \mathrm{conv}(X)$ with $F(z) < f^*$. According to part (a), there exist points $x_i \in X$ and nonnegative scalars α_i with $\sum_i \alpha_i = 1$ such that $z = \sum_i \alpha_i x_i$ and

$$F(z) \le \sum_i \alpha_i f(x_i) < f^*,$$

implying that

$$\sum_i \alpha_i \big(f(x_i) - f^*\big) < 0.$$

Since each α_i is nonnegative, for this inequality to hold, we must have $f(x_i) - f^* < 0$ for some i, but this cannot be true because $x_i \in X$ and f^* is the optimal value of f over X. Therefore

$$\inf_{x \in \mathrm{conv}(X)} F(x) = \inf_{x \in X} f(x).$$

(c) If $x^* \in X$ is a global minimum of f over X, then x^* also belongs to $\mathrm{conv}(X)$, and by part (b)

$$\inf_{x \in \mathrm{conv}(X)} F(x) = \inf_{x \in X} f(x) = f(x^*) \ge F(x^*),$$

showing that x^* is also a global minimum of F over $\mathrm{conv}(X)$.

1.21 (Minimization of Linear Functions)

Let $f : X \mapsto \Re$ be the function $f(x) = c'x$, and define

$$F(x) = \inf\big\{w \mid (x, w) \in \mathrm{conv}\big(\mathrm{epi}(f)\big)\big\},$$

as in Exercise 1.20. According to this exercise, we have

$$\inf_{x \in \mathrm{conv}(X)} F(x) = \inf_{x \in X} f(x),$$

and

$$\begin{aligned}F(x) &= \inf\left\{\sum_i \alpha_i c'x_i \;\middle|\; \sum_i \alpha_i x_i = x,\, x_i \in X,\, \sum_i \alpha_i = 1,\, \alpha_i \geq 0\right\}\\ &= \inf\left\{c'\left(\sum_i \alpha_i x_i\right) \;\middle|\; \sum_i \alpha_i x_i = x,\, x_i \in X,\, \sum_i \alpha_i = 1,\, \alpha_i \geq 0\right\}\\ &= c'x,\end{aligned}$$

showing that

$$\inf_{x\in \mathrm{conv}(X)} c'x = \inf_{x\in X} c'x.$$

According to Exercise 1.20(c), if $\inf_{x\in X} c'x$ is attained at some $x^* \in X$, then $\inf_{x\in\mathrm{conv}(X)} c'x$ is also attained at x^*. Suppose now that $\inf_{x\in\mathrm{conv}(X)} c'x$ is attained at some $x^* \in \mathrm{conv}(X)$, i.e., there is $x^* \in \mathrm{conv}(X)$ such that

$$\inf_{x\in \mathrm{conv}(X)} c'x = c'x^*.$$

Then, by Caratheodory's Theorem, there exist vectors $x_1, \ldots, x_{n+1}$ in X and nonnegative scalars $\alpha_1, \ldots, \alpha_{n+1}$ with $\sum_{i=1}^{n+1} \alpha_i = 1$ such that $x^* = \sum_{i=1}^{n+1} \alpha_i x_i$, implying that

$$c'x^* = \sum_{i=1}^{n+1} \alpha_i c'x_i.$$

Since $x_i \in X \subset \mathrm{conv}(X)$ for all i and $c'x \geq c'x^*$ for all $x \in \mathrm{conv}(X)$, it follows that

$$c'x^* = \sum_{i=1}^{n+1} \alpha_i c'x_i \geq \sum_{i=1}^{n+1} \alpha_i c'x^* = c'x^*,$$

implying that $c'x_i = c'x^*$ for all i corresponding to $\alpha_i > 0$. Hence, $\inf_{x\in X} c'x$ is attained at the x_i's corresponding to $\alpha_i > 0$.

1.22 (Extension of Caratheodory's Theorem)

The proof will be an application of Caratheodory's Theorem [part (a)] to the subset of $\Re^{n+1}$ given by

$$Y = \big\{(x, 1) \mid x \in X_1\big\} \cup \big\{(y, 0) \mid y \in X_2\big\}.$$

If $x \in X$, then

$$x = \sum_{i=1}^{k} \gamma_i x_i + \sum_{i=k+1}^{m} \gamma_i y_i,$$

where the vectors $x_1, \ldots, x_k$ belong to X_1, the vectors $y_{k+1}, \ldots, y_m$ belong to X_2, and the scalars $\gamma_1, \ldots, \gamma_m$ are nonnegative with $\gamma_1 + \cdots + \gamma_k = 1$. Equivalently, $(x, 1) \in \mathrm{cone}(Y)$. By Caratheodory's Theorem part (a), we have that

$$(x, 1) = \sum_{i=1}^{k} \alpha_i (x_i, 1) + \sum_{i=k+1}^{m} \alpha_i (y_i, 0),$$

for some positive scalars $\alpha_1, \ldots, \alpha_m$ and vectors

$$(x_1, 1), \ldots (x_k, 1),\ (y_{k+1}, 0), \ldots,\ (y_m, 0),$$

which are linearly independent (implying that $m \leq n+1$) or equivalently,

$$x = \sum_{i=1}^{k} \alpha_i x_i + \sum_{i=k+1}^{m} \alpha_i y_i, \qquad 1 = \sum_{i=1}^{k} \alpha_i.$$

Finally, to show that the vectors $x_2 - x_1, \ldots,\ x_k - x_1,\ y_{k+1}, \ldots, y_m$ are linearly independent, assume to arrive at a contradiction, that there exist $\lambda_2, \ldots, \lambda_m$, not all 0, such that

$$\sum_{i=2}^{k} \lambda_i (x_i - x_1) + \sum_{i=k+1}^{m} \lambda_i y_i = 0.$$

Equivalently, defining $\lambda_1 = -(\lambda_2 + \cdots + \lambda_m)$, we have

$$\sum_{i=1}^{k} \lambda_i (x_i, 1) + \sum_{i=k+1}^{m} \lambda_i (y_i, 0) = 0,$$

which contradicts the linear independence of the vectors

$$(x_1, 1), \ldots, (x_k, 1), (y_{k+1}, 0), \ldots, (y_m, 0).$$

1.23

The set $\mathrm{cl}(X)$ is compact since X is bounded by assumption. Hence, by Prop. 1.3.2, its convex hull, $\mathrm{conv}\big(\mathrm{cl}(X)\big)$, is compact, and it follows that

$$\mathrm{cl}\big(\mathrm{conv}(X)\big) \subset \mathrm{cl}\Big(\mathrm{conv}\big(\mathrm{cl}(X)\big)\Big) = \mathrm{conv}\big(\mathrm{cl}(X)\big).$$

It is also true in general that

$$\mathrm{conv}\big(\mathrm{cl}(X)\big) \subset \mathrm{conv}\Big(\mathrm{cl}\big(\mathrm{conv}(X)\big)\Big) = \mathrm{cl}\big(\mathrm{conv}(X)\big),$$

since by Prop. 1.2.1(d), the closure of a convex set is convex. Hence, the result follows.

1.24 (Radon's Theorem)

Consider the system of $n+1$ equations in the m unknowns $\lambda_1, \ldots, \lambda_m$

$$\sum_{i=1}^m \lambda_i x_i = 0, \qquad \sum_{i=1}^m \lambda_i = 0.$$

Since $m > n+1$, there exists a nonzero solution, call it λ^*. Let

$$I = \{i \mid \lambda_i^* \geq 0\}, \qquad J = \{j \mid \lambda_j^* < 0\},$$

and note that I and J are nonempty, and that

$$\sum_{k \in I} \lambda_k^* = \sum_{k \in J} (-\lambda_k^*) > 0.$$

Consider the vector

$$x^* = \sum_{i \in I} \alpha_i x_i,$$

where

$$\alpha_i = \frac{\lambda_i^*}{\sum_{k \in I} \lambda_k^*}, \qquad i \in I.$$

In view of the equations $\sum_{i=1}^m \lambda_i^* x_i = 0$ and $\sum_{i=1}^m \lambda_i^* = 0$, we also have

$$x^* = \sum_{j \in J} \alpha_j x_j,$$

where

$$\alpha_j = \frac{-\lambda_j^*}{\sum_{k \in J} (-\lambda_k^*)}, \qquad j \in J.$$

It is seen that the α_i and α_j are nonnegative, and that

$$\sum_{i \in I} \alpha_i = \sum_{j \in J} \alpha_j = 1,$$

so x^* belongs to the intersection

$$\mathrm{conv}\big(\{x_i \mid i \in I\}\big) \cap \mathrm{conv}\big(\{x_j \mid j \in J\}\big).$$

1.25 (Helly's Theorem [Hel21])

Let B_j be defined as in the hint, and for each j, let x_j be a vector in B_j. Since $M+1 \geq n+2$, we can apply Radon's Theorem to the vectors $x_1, \ldots, x_{M+1}$. Thus, there exist nonempty and disjoint index subsets I and J such that $I \cup J = \{1, \ldots, m\}$, nonnegative scalars $\alpha_1, \ldots, \alpha_{m+1}$, and a vector x^* such that

$$x^* = \sum_{i \in I} \alpha_i x_i = \sum_{j \in J} \alpha_j x_j, \qquad \sum_{i \in I} \alpha_i = \sum_{j \in J} \alpha_j = 1.$$

It can be seen that for every $i \in I$, a vector in B_i belongs to the intersection $\cap_{j \in J} C_j$. Therefore, since x^* is a convex combination of vectors in B_i, $i \in I$, x^* also belongs to the intersection $\cap_{j \in J} C_j$. Similarly, by reversing the role of I and J, we see that x^* belongs to the intersection $\cap_{i \in I} C_I$. Thus, x^* belongs to the intersection of the entire collection $C_1, \ldots, C_{M+1}$.

1.26

Assume the contrary, i.e., that for every index set $I \subset \{1, \ldots, M\}$, which contains no more than $n+1$ indices, we have

$$\inf_{x \in \Re^n} \left\{ \max_{i \in I} f_i(x) \right\} < f^*.$$

This means that for every such I, the intersection $\cap_{i \in I} X_i$ is nonempty, where

$$X_i = \bigl\{ x \mid f_i(x) < f^* \bigr\}.$$

From Helly's Theorem, it follows that the entire collection $\{X_i \mid i = 1, \ldots, M\}$ has nonempty intersection, thereby implying that

$$\inf_{x \in \Re^n} \left\{ \max_{i=1,\ldots,M} f_i(x) \right\} < f^*.$$

This contradicts the definition of f^*. *Note*: The result of this exercise relates to the following question: what is the minimal number of functions f_i that we need to include in the cost function $\max_i f_i(x)$ in order to attain the optimal value f^*? According to the result, the number is no more than $n+1$. For applications of this result in structural design and Chebyshev approximation, see Ben Tal and Nemirovski [BeN01].

1.27

Let $\overline{x}$ be an arbitrary vector in $\mathrm{cl}(C)$. If $f(\overline{x}) = \infty$, then we are done, so assume that $f(\overline{x})$ is finite. Let x be a point in the relative interior of C. By the Line Segment Principle, all the points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to $\mathrm{ri}(C)$ and therefore, belong to C. From this, the given property of f, and the convexity of f, we obtain for all $\alpha \in (0, 1]$,

$$\alpha f(x) + (1-\alpha) f(\overline{x}) \geq f\bigl(\alpha x + (1-\alpha)\overline{x}\bigr) \geq \gamma.$$

By letting $\alpha \to 0$, it follows that $f(\overline{x}) \geq \gamma$. Hence, $f(\overline{x}) \geq \gamma$ for all $x \in \mathrm{cl}(C)$.

1.28

From Prop. 1.4.5(b), we have that for any vector $a \in \Re^n$, $\mathrm{ri}(C + a) = \mathrm{ri}(C) + a$. Therefore, we can assume without loss of generality that $0 \in C$, and $\mathrm{aff}(C)$ coincides with S. We need to show that

$$\mathrm{ri}(C) = \mathrm{int}(C + S^{\perp}) \cap C.$$

Let $x \in \mathrm{ri}(C)$. By definition, this implies that $x \in C$ and there exists some open ball $B(x, \epsilon)$ centered at x with radius $\epsilon > 0$ such that

$$B(x, \epsilon) \cap S \subset C. \tag{1.9}$$

We now show that $B(x,\epsilon) \subset C + S^\perp$. Let z be a vector in $B(x,\epsilon)$. Then, we can express z as $z = x + \alpha y$ for some vector $y \in \Re^n$ with $\|y\| = 1$, and some $\alpha \in [0,\epsilon)$. Since S and $S^\perp$ are orthogonal subspaces, y can be uniquely decomposed as $y = y_S + y_{S^\perp}$, where $y_S \in S$ and $y_{S^\perp} \in S^\perp$. Since $\|y\| = 1$, this implies that $\|y_S\| \leq 1$ (Pythagorean Theorem), and using Eq. (1.9), we obtain

$$x + \alpha y_S \in B(x,\epsilon) \cap S \subset C,$$

from which it follows that the vector $z = x + \alpha y$ belongs to $C + S^\perp$, implying that $B(x,\epsilon) \subset C + S^\perp$. This shows that $x \in \mathrm{int}(C + S^\perp) \cap C$.

Conversely, let $x \in \mathrm{int}(C + S^\perp) \cap C$. We have that $x \in C$ and there exists some open ball $B(x,\epsilon)$ centered at x with radius $\epsilon > 0$ such that $B(x,\epsilon) \subset C + S^\perp$. Since C is a subset of S, it can be seen that $(C + S^\perp) \cap S = C$. Therefore,

$$B(x,\epsilon) \cap S \subset C,$$

implying that $x \in \mathrm{ri}(C)$.

1.29

(a) Let C be the given convex set. The convex hull of any subset of C is contained in C. Therefore, the maximum dimension of the various simplices contained in C is the largest m for which C contains $m+1$ vectors $x_0, \ldots, x_m$ such that $x_1 - x_0, \ldots, x_m - x_0$ are linearly independent.

Let $K = \{x_0, \ldots, x_m\}$ be such a set with m maximal, and let $\mathrm{aff}(K)$ denote the affine hull of set K. Then, we have $\dim\big(\mathrm{aff}(K)\big) = m$, and since $K \subset C$, it follows that $\mathrm{aff}(K) \subset \mathrm{aff}(C)$.

We claim that $C \subset \mathrm{aff}(K)$. To see this, assume that there exists some $x \in C$, which does not belong to $\mathrm{aff}(K)$. This implies that the set $\{x, x_0, \ldots, x_m\}$ is a set of $m+2$ vectors in C such that $x - x_0,\ x_1 - x_0, \ldots, x_m - x_0$ are linearly independent, contradicting the maximality of m. Hence, we have $C \subset \mathrm{aff}(K)$, and it follows that

$$\mathrm{aff}(K) = \mathrm{aff}(C),$$

thereby implying that $\dim(C) = m$.

(b) We first consider the case where C is n-dimensional with $n > 0$ and show that the interior of C is not empty. By part (a), an n-dimensional convex set contains an n-dimensional simplex. We claim that such a simplex S has a nonempty interior. Indeed, applying an affine transformation if necessary, we can assume that the vertices of S are the vectors $(0,0,\ldots,0), (1,0,\ldots,0), \ldots, (0,0,\ldots,1)$, i.e.,

$$S = \left\{ (x_1, \ldots, x_n) \;\middle|\; x_i \geq 0,\ \forall\ i = 1, \ldots, n,\ \ \sum_{i=1}^n x_i \leq 1 \right\}.$$

The interior of the simplex S,

$$\mathrm{int}(S) = \left\{ (x_1, \ldots, x_n) \mid x_i > 0,\ \forall\ i = 1, \ldots, n,\ \ \sum_{i=1}^n x_i < 1 \right\},$$

is nonempty, which in turn implies that $\mathrm{int}(C)$ is nonempty.

For the case where $\dim(C) < n$, consider the n-dimensional set $C + S^\perp$, where $S^\perp$ is the orthogonal complement of the subspace parallel to $\mathrm{aff}(C)$. Since $C + S^\perp$ is a convex set, it follows from the above argument that $\mathrm{int}(C + S^\perp)$ is nonempty. Let $x \in \mathrm{int}(C + S^\perp)$. We can represent x as $x = x_C + x_{S^\perp}$, where $x_C \in C$ and $x_{S^\perp} \in S^\perp$. It can be seen that $x_C \in \mathrm{int}(C + S^\perp)$. Since

$$\mathrm{ri}(C) = \mathrm{int}(C + S^\perp) \cap C,$$

(cf. Exercise 1.28), it follows that $x_c \in \mathrm{ri}(C)$, so $\mathrm{ri}(C)$ is nonempty.

1.30

(a) Let C_1 be the segment $\big\{(x_1, x_2) \mid 0 \le x_1 \le 1,\ x_2 = 0\big\}$ and let C_2 be the box $\big\{(x_1, x_2) \mid 0 \le x_1 \le 1,\ 0 \le x_2 \le 1\big\}$. We have

$$\mathrm{ri}(C_1) = \big\{(x_1, x_2) \mid 0 < x_1 < 1,\ x_2 = 0\big\},$$

$$\mathrm{ri}(C_2) = \big\{(x_1, x_2) \mid 0 < x_1 < 1,\ 0 < x_2 < 1\big\}.$$

Thus $C_1 \subset C_2$, while $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \varnothing$.

(b) Let $x \in \mathrm{ri}(C_1)$, and consider a open ball B centered at x such that $B \cap \mathrm{aff}(C_1) \subset C_1$. Since $\mathrm{aff}(C_1) = \mathrm{aff}(C_2)$ and $C_1 \subset C_2$, it follows that $B \cap \mathrm{aff}(C_2) \subset C_2$, so $x \in \mathrm{ri}(C_2)$. Hence $\mathrm{ri}(C_1) \subset \mathrm{ri}(C_2)$.

(c) Because $C_1 \subset C_2$, we have

$$\mathrm{ri}(C_1) = \mathrm{ri}(C_1 \cap C_2).$$

Since $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \varnothing$, there holds

$$\mathrm{ri}(C_1 \cap C_2) = \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$$

[Prop. 1.4.5(a)]. Combining the preceding two relations, we obtain $\mathrm{ri}(C_1) \subset \mathrm{ri}(C_2)$.

(d) Let x_2 be in the intersection of C_1 and $\mathrm{ri}(C_2)$, and let x_1 be in the relative interior of C_1 [$\mathrm{ri}(C_1)$ is nonempty by Prop. 1.4.1(b)]. If $x_1 = x_2$, then we are done, so assume that $x_1 \neq x_2$. By the Line Segment Principle, all the points on the line segment connecting x_1 and x_2, except possibly x_2, belong to the relative interior of C_1. Since $C_1 \subset C_2$, the vector x_1 is in C_2, so that by the Line Segment Principle, all the points on the line segment connecting x_1 and x_2, except possibly x_1, belong to the relative interior of C_2. Hence, all the points on the line segment connecting x_1 and x_2, except possibly x_1 and x_2, belong to the intersection $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$, showing that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty.

1.31

(a) Let $x \in \text{ri}(C)$. We will show that for every $\overline{x} \in \text{aff}(C)$, there exists a $\gamma > 1$ such that $x + (\gamma - 1)(x - \overline{x}) \in C$. This is true if $\overline{x} = x$, so assume that $\overline{x} \neq x$. Since $x \in \text{ri}(C)$, there exists $\epsilon > 0$ such that

$$\big\{z \mid \|z - x\| < \epsilon\big\} \cap \text{aff}(C) \subset C.$$

Choose a point $\overline{x}_\epsilon \in C$ in the intersection of the ray $\big\{x + \alpha(\overline{x} - x) \mid \alpha \geq 0\big\}$ and the set $\big\{z \mid \|z - x\| < \epsilon\big\} \cap \text{aff}(C)$. Then, for some positive scalar α_ϵ,

$$x - \overline{x}_\epsilon = \alpha_\epsilon(x - \overline{x}).$$

Since $x \in \text{ri}(C)$ and $\overline{x}_\epsilon \in C$, by Prop. 1.4.1(c), there is $\gamma_\epsilon > 1$ such that

$$x + (\gamma_\epsilon - 1)(x - \overline{x}_\epsilon) \in C,$$

which in view of the preceding relation implies that

$$x + (\gamma_\epsilon - 1)\alpha_\epsilon(x - \overline{x}) \in C.$$

The result follows by letting $\gamma = 1 + (\gamma_\epsilon - 1)\alpha_\epsilon$ and noting that $\gamma > 1$, since $(\gamma_\epsilon - 1)\alpha_\epsilon > 0$. The converse assertion follows from the fact $C \subset \text{aff}(C)$ and Prop. 1.4.1(c).

(b) The inclusion $\text{cone}(C) \subset \text{aff}(C)$ always holds if $0 \in C$. To show the reverse inclusion, we note that by part (a) with $x = 0$, for every $\overline{x} \in \text{aff}(C)$, there exists $\gamma > 1$ such that $\tilde{x} = (\gamma - 1)(-\overline{x}) \in C$. By using part (a) again with $x = 0$, for $\tilde{x} \in C \subset \text{aff}(C)$, we see that there is $\tilde{\gamma} > 1$ such that $z = (\tilde{\gamma} - 1)(-\tilde{x}) \in C$, which combined with $\tilde{x} = (\gamma - 1)(-\overline{x})$ yields $z = (\tilde{\gamma} - 1)(\gamma - 1)\overline{x} \in C$. Hence

$$\overline{x} = \frac{1}{(\tilde{\gamma} - 1)(\gamma - 1)} z$$

with $z \in C$ and $(\tilde{\gamma} - 1)(\gamma - 1) > 0$, implying that $\overline{x} \in \text{cone}(C)$ and, showing that $\text{aff}(C) \subset \text{cone}(C)$.

(c) This follows by part (b), where $C = \text{conv}(X)$, and the fact

$$\text{cone}\big(\text{conv}(X)\big) = \text{cone}(X)$$

[Exercise 1.18(b)].

1.32

(a) If $0 \in C$, then $0 \in \text{ri}(C)$ since 0 is not on the relative boundary of C. By Exercise 1.31(b), it follows that $\text{cone}(C)$ coincides with $\text{aff}(C)$, which is a closed set. If $0 \notin C$, let y be in the closure of $\text{cone}(C)$ and let $\{y_k\} \subset \text{cone}(C)$ be a sequence converging to y. By Exercise 1.15, for every y_k, there exists a nonnegative scalar α_k and a vector $x_k \in C$ such that $y_k = \alpha_k x_k$. Since $\{y_k\} \to y$, the sequence $\{y_k\}$ is bounded, implying that

$$\alpha_k \|x_k\| \leq \sup_{m \geq 0} \|y_m\| < \infty, \qquad \forall\ k.$$

We have $\inf_{m \geq 0} \|x_m\| > 0$, since $\{x_k\} \subset C$ and C is a compact set not containing the origin, so that

$$0 \leq \alpha_k \leq \frac{\sup_{m \geq 0} \|y_m\|}{\inf_{m \geq 0} \|x_m\|} < \infty, \qquad \forall\ k.$$

Thus, the sequence $\{(\alpha_k, x_k)\}$ is bounded and has a limit point (α, x) such that $\alpha \geq 0$ and $x \in C$. By taking a subsequence of $\{(\alpha_k, x_k)\}$ that converges to (α, x), and by using the facts $y_k = \alpha_k x_k$ for all k and $\{y_k\} \to y$, we see that $y = \alpha x$ with $\alpha \geq 0$ and $x \in C$. Hence, $y \in \text{cone}(C)$, showing that $\text{cone}(C)$ is closed.

(b) To see that the assertion in part (a) fails when C is unbounded, let C be the line $\big\{(x_1, x_2) \mid x_1 = 1,\ x_2 \in \Re\big\}$ in $\Re^2$ not passing through the origin. Then, $\text{cone}(C)$ is the nonclosed set $\big\{(x_1, x_2) \mid x_1 > 0,\ x_2 \in \Re\big\} \cup \big\{(0, 0)\big\}$.

To see that the assertion in part (a) fails when C contains the origin on its relative boundary, let C be the closed ball $\big\{(x_1, x_2) \mid (x_1 - 1)^2 + x_2^2 \leq 1\big\}$ in $\Re^2$. Then, $\text{cone}(C)$ is the nonclosed set $\big\{(x_1, x_2) \mid x_1 > 0,\ x_2 \in \Re\big\} \cup \big\{(0, 0)\big\}$ (see Fig. 1.3.2).

(c) Since C is compact, the convex hull of C is compact (cf. Prop. 1.3.2). Because $\text{conv}(C)$ does not contain the origin on its relative boundary, by part (a), the cone generated by $\text{conv}(C)$ is closed. By Exercise 1.18(b), $\text{cone}\big(\text{conv}(C)\big)$ coincides with $\text{cone}(C)$ implying that $\text{cone}(C)$ is closed.

1.33

(a) By Prop. 1.4.1(b), the relative interior of a convex set is a convex set. We only need to show that $\text{ri}(C)$ is a cone. Let $y \in \text{ri}(C)$. Then, $y \in C$ and since C is a cone, $\alpha y \in C$ for all $\alpha > 0$. By the Line Segment Principle, all the points on the line segment connecting y and αy, except possibly αy, belong to $\text{ri}(C)$. Since this is true for every $\alpha > 0$, it follows that $\alpha y \in \text{ri}(C)$ for all $\alpha > 0$, showing that $\text{ri}(C)$ is a cone.

(b) Consider the linear transformation A that maps $(\alpha_1, \ldots, \alpha_m) \in \Re^m$ into $\sum_{i=1}^m \alpha_i x_i \in \Re^n$. Note that C is the image of the nonempty convex set

$$\big\{(\alpha_1, \ldots, \alpha_m) \mid \alpha_1 \geq 0, \ldots, \alpha_m \geq 0\big\}$$

under the linear transformation A. Therefore, by using Prop. 1.4.3(d), we have

$$
\begin{aligned}
\mathrm{ri}(C) &= \mathrm{ri}\Big(A \cdot \big\{(\alpha_1, \ldots, \alpha_m) \mid \alpha_1 \geq 0, \ldots, \alpha_m \geq 0\big\}\Big) \\
&= A \cdot \mathrm{ri}\Big(\big\{(\alpha_1, \ldots, \alpha_m) \mid \alpha_1 \geq 0, \ldots, \alpha_m \geq 0\big\}\Big) \\
&= A \cdot \big\{(\alpha_1, \ldots, \alpha_m) \mid \alpha_1 > 0, \ldots, \alpha_m > 0\big\} \\
&= \left\{ \sum_{i=1}^{m} \alpha_i x_i \mid \alpha_1 > 0, \ldots, \alpha_m > 0 \right\}.
\end{aligned}
$$

1.34

Define the sets

$$D = \Re^n \times C, \qquad S = \big\{(x, Ax) \mid x \in \Re^n\big\}.$$

Let T be the linear transformation that maps $(x, y) \in \Re^{n+m}$ into $x \in \Re^n$. Then it can be seen that

$$A^{-1} \cdot C = T \cdot (D \cap S). \tag{1.10}$$

The relative interior of D is given by $\mathrm{ri}(D) = \Re^n \times \mathrm{ri}(C)$, and the relative interior of S is equal to S (since S is a subspace). Hence,

$$A^{-1} \cdot \mathrm{ri}(C) = T \cdot \big(\mathrm{ri}(D) \cap S\big). \tag{1.11}$$

In view of the assumption that $A^{-1} \cdot \mathrm{ri}(C)$ is nonempty, we have that the intersection $\mathrm{ri}(D) \cap S$ is nonempty. Therefore, it follows from Props. 1.4.3(d) and 1.4.5(a) that

$$\mathrm{ri}\big(T \cdot (D \cap S)\big) = T \cdot \big(\mathrm{ri}(D) \cap S\big). \tag{1.12}$$

Combining Eqs. (1.10)-(1.12), we obtain

$$\mathrm{ri}(A^{-1} \cdot C) = A^{-1} \cdot \mathrm{ri}(C).$$

Next, we show the second relation. We have

$$A^{-1} \cdot \mathrm{cl}(C) = \big\{x \mid Ax \in \mathrm{cl}(C)\big\} = T \cdot \big\{(x, Ax) \mid Ax \in \mathrm{cl}(C)\big\} = T \cdot \big(\mathrm{cl}(D) \cap S\big).$$

Since the intersection $\mathrm{ri}(D) \cap S$ is nonempty, it follows from Prop. 1.4.5(a) that $\mathrm{cl}(D) \cap S = \mathrm{cl}(D \cap S)$. Furthermore, since T is continuous, we obtain

$$A^{-1} \cdot \mathrm{cl}(C) = T \cdot \mathrm{cl}(D \cap S) \subset \mathrm{cl}\big(T \cdot (D \cap S)\big),$$

which combined with Eq. (1.10) yields

$$A^{-1} \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A^{-1} \cdot C).$$

To show the reverse inclusion, $\mathrm{cl}(A^{-1} \cdot C) \subset A^{-1} \cdot \mathrm{cl}(C)$, let $\overline{x}$ be some vector in $\mathrm{cl}(A^{-1} \cdot C)$. This implies that there exists some sequence $\{x_k\}$ converging to $\overline{x}$ such that $Ax_k \in C$ for all k. Since x_k converges to $\overline{x}$, we have that Ax_k converges to $A\overline{x}$, thereby implying that $A\overline{x} \in \mathrm{cl}(C)$, or equivalently, $\overline{x} \in A^{-1} \cdot \mathrm{cl}(C)$.

1.35 (Closure of a Convex Function)

(a) Let $g : \Re^n \mapsto [-\infty, \infty]$ be such that $g(x) \le f(x)$ for all $x \in \Re^n$. Choose any $x \in \mathrm{dom}(\mathrm{cl}\, f)$. Since $\mathrm{epi}(\mathrm{cl}\, f) = \mathrm{cl}\big(\mathrm{epi}(f)\big)$, we can choose a sequence $\big\{(x_k, w_k)\big\} \in \mathrm{epi}(f)$ such that $x_k \to x$, $w_k \to (\mathrm{cl}\, f)(x)$. Since g is lower semicontinuous at x, we have

$$g(x) \le \liminf_{k\to\infty} g(x_k) \le \liminf_{k\to\infty} f(x_k) \le \liminf_{k\to\infty} w_k = (\mathrm{cl}\, f)(x).$$

Note also that since $\mathrm{epi}(f) \subset \mathrm{epi}(\mathrm{cl}\, f)$, we have $(\mathrm{cl}\, f)(x) \le f(x)$ for all $x \in \Re^n$.

(b) For the proof of this part and the next, we will use the easily shown fact that for any convex function f, we have

$$\mathrm{ri}\big(\mathrm{epi}(f)\big) = \big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(f)\big),\ f(x) < w\big\}.$$

Let $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, and consider the vertical line $L = \big\{(x, w) \mid w \in \Re\big\}$. Then there exists $\hat{w}$ such that $(x, \hat{w}) \in L \cap \mathrm{ri}\big(\mathrm{epi}(f)\big)$. Let $\overline{w}$ be such that $(x, \overline{w}) \in L \cap \mathrm{cl}\big(\mathrm{epi}(f)\big)$. Then, by Prop. 1.4.5(a), we have $L \cap \mathrm{cl}\big(\mathrm{epi}(f)\big) = \mathrm{cl}\big(L \cap \mathrm{epi}(f)\big)$, so that $(x, \overline{w}) \in \mathrm{cl}\big(L \cap \mathrm{epi}(f)\big)$. It follows from the Line Segment Principle that the vector $\big(x, \hat{w} + \alpha(\overline{w} - \hat{w})\big)$ belongs to $\mathrm{epi}(f)$ for all $\alpha \in [0, 1)$. Taking the limit as $\alpha \to 1$, we see that $f(x) \le \overline{w}$ for all $\overline{w}$ such that $(x, \overline{w}) \in L \cap \mathrm{cl}\big(\mathrm{epi}(f)\big)$, implying that $f(x) \le (\mathrm{cl}\, f)(x)$. On the other hand, since $\mathrm{epi}(f) \subset \mathrm{epi}(\mathrm{cl}\, f)$, we have $(\mathrm{cl}\, f)(x) \le f(x)$ for all $x \in \Re^n$, so $f(x) = (\mathrm{cl}\, f)(x)$.

We know that a closed convex function that is improper cannot take a finite value at any point. Since $\mathrm{cl}\, f$ is closed and convex, and takes a finite value at all points of the nonempty set $\mathrm{ri}\big(\mathrm{dom}(f)\big)$, it follows that $\mathrm{cl}\, f$ must be proper.

(c) Since the function $\mathrm{cl}\, f$ is closed and is majorized by f, we have

$$(\mathrm{cl}\, f)(y) \le \liminf_{\alpha\downarrow 0} (\mathrm{cl}\, f)\big(y + \alpha(x - y)\big) \le \liminf_{\alpha\downarrow 0} f\big(y + \alpha(x - y)\big).$$

To show the reverse inequality, let w be such that $f(x) < w$. Then, $(x, w) \in \mathrm{ri}\big(\mathrm{epi}(f)\big)$, while $\big(y, (\mathrm{cl}\, f)(y)\big) \in \mathrm{cl}\big(\mathrm{epi}(f)\big)$. From the Line Segment Principle, it follows that

$$\big(\alpha x + (1 - \alpha)y, \alpha w + (1 - \alpha)(\mathrm{cl}\, f)(y)\big) \in \mathrm{ri}\big(\mathrm{epi}(f)\big), \qquad \forall\ \alpha \in (0, 1].$$

Hence,

$$f\big(\alpha x + (1 - \alpha)y\big) < \alpha w + (1 - \alpha)(\mathrm{cl}\, f)(y), \qquad \forall\ \alpha \in (0, 1].$$

By taking the limit as $\alpha \to 0$, we obtain

$$\liminf_{\alpha\downarrow 0} f\big(y + \alpha(x - y)\big) \le (\mathrm{cl}\, f)(y),$$

thus completing the proof.

(d) Let $x \in \cap_{i=1}^m \mathrm{ri}\big(\mathrm{dom}(f_i)\big)$. Since by Prop. 1.4.5(a), we have

$$\mathrm{ri}\big(\mathrm{dom}(f)\big) = \cap_{i=1}^m \mathrm{ri}\big(\mathrm{dom}(f_i)\big),$$

it follows that $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$. By using part (c), we have for every $y \in \mathrm{dom}(\mathrm{cl}\, f)$,

$$(\mathrm{cl}\, f)(y) = \lim_{\alpha\downarrow 0} f\big(y + \alpha(x - y)\big) = \sum_{i=1}^m \lim_{\alpha\downarrow 0} f_i\big(y + \alpha(x - y)\big) = \sum_{i=1}^m (\mathrm{cl}\, f_i)(y).$$

1.36

Assume first that C is closed. Since $C \cap M$ is bounded, by part (c) of the Recession Cone Theorem (cf. Prop. 1.5.1), $R_{C\cap M} = \{0\}$. This and the fact $R_{C\cap M} = R_C \cap R_M$, imply that $R_C \cap R_M = \{0\}$. Let S be a subspace such that $M = x + S$ for some $x \in M$. Then $R_M = S$, so that $R_C \cap S = \{0\}$. For every affine set $\overline{M}$ that is parallel to M, we have $R_{\overline{M}} = S$, so that

$$R_{C\cap\overline{M}} = R_C \cap R_{\overline{M}} = R_C \cap S = \{0\}.$$

Therefore, by part (c) of the Recession Cone Theorem, $C \cap \overline{M}$ is bounded.

In the general case where C is not closed, the assumption that $C \cap M$ is nonempty and bounded implies that $\mathrm{cl}(C) \cap M$ is nonempty and bounded. Therefore, by what has already been proved, $\mathrm{cl}(C) \cap \overline{M}$ is bounded, implying that $C \cap \overline{M}$ is bounded.

1.37 (Properties of Cartesian Products)

(a) We first show that the convex hull of X is equal to the Cartesian product of the convex hulls of the sets $X_i,\ i = 1, \ldots, m$. Let y be a vector that belongs to $\mathrm{conv}(X)$. Then, by definition, for some k, we have

$$y = \sum_{i=1}^{k} \alpha_i y_i, \qquad \text{with } \alpha_i \geq 0,\ i = 1, \ldots, m, \quad \sum_{i=1}^{k} \alpha_i = 1,$$

where $y_i \in X$ for all i. Since $y_i \in X$, we have that $y_i = (x_1^i, \ldots, x_m^i)$ for all i, with $x_1^i \in X_1, \ldots, x_m^i \in X_m$. It follows that

$$y = \sum_{i=1}^{k} \alpha_i (x_1^i, \ldots, x_m^i) = \left(\sum_{i=1}^{k} \alpha_i x_1^i, \ldots, \sum_{i=1}^{k} \alpha_i x_m^i \right),$$

thereby implying that $y \in \mathrm{conv}(X_1) \times \cdots \times \mathrm{conv}(X_m)$.

To prove the reverse inclusion, assume that y is a vector in $\mathrm{conv}(X_1) \times \cdots \times \mathrm{conv}(X_m)$. Then, we can represent y as $y = (y_1, \ldots, y_m)$ with $y_i \in \mathrm{conv}(X_i)$, i.e., for all $i = 1, \ldots, m$, we have

$$y_i = \sum_{j=1}^{k_i} \alpha_j^i x_j^i, \qquad x_j^i \in X_i,\ \forall\, j, \qquad \alpha_j^i \geq 0,\ \forall\, j, \quad \sum_{j=1}^{k_i} \alpha_j^i = 1.$$

First, consider the vectors

$$(x_1^1, x_{r_1}^2, \ldots, x_{r_{m-1}}^m),\ (x_2^1, x_{r_1}^2, \ldots, x_{r_{m-1}}^m), \ldots, (x_{k_i}^1, x_{r_1}^2, \ldots, x_{r_{m-1}}^m),$$

for all possible values of $r_1, \ldots, r_{m-1}$, i.e., we fix all components except the first one, and vary the first component over all possible x_j^1's used in the convex combination that yields y_1. Since all these vectors belong to X, their convex combination given by

$$\left(\left(\sum_{j=1}^{k_1}\alpha_j^1 x_j^1\right), x_{r_1}^2, \ldots, x_{r_{m-1}}^m\right)$$

belongs to the convex hull of X for all possible values of $r_1, \ldots, r_{m-1}$. Now, consider the vectors

$$\left(\left(\sum_{j=1}^{k_1}\alpha_j^1 x_j^1\right), x_1^2, \ldots, x_{r_{m-1}}^m\right), \ldots, \left(\left(\sum_{j=1}^{k_1}\alpha_j^1 x_j^1\right), x_{k_2}^2, \ldots, x_{r_{m-1}}^m\right),$$

i.e., fix all components except the second one, and vary the second component over all possible x_j^2's used in the convex combination that yields y_2. Since all these vectors belong to $\mathrm{conv}(X)$, their convex combination given by

$$\left(\left(\sum_{j=1}^{k_1}\alpha_j^1 x_j^1\right), \left(\sum_{j=1}^{k_2}\alpha_j^2 x_j^2\right), \ldots, x_{r_{m-1}}^m\right)$$

belongs to the convex hull of X for all possible values of $r_2, \ldots, r_{m-1}$. Proceeding in this way, we see that the vector given by

$$\left(\left(\sum_{j=1}^{k_1}\alpha_j^1 x_j^1\right), \left(\sum_{j=1}^{k_2}\alpha_j^2 x_j^2\right), \ldots, \left(\sum_{j=1}^{k_m}\alpha_j^m x_j^m\right)\right)$$

belongs to $\mathrm{conv}(X)$, thus proving our claim.

Next, we show the corresponding result for the closure of X. Assume that $y = (x_1, \ldots, x_m) \in \mathrm{cl}(X)$. This implies that there exists some sequence $\{y^k\} \subset X$ such that $y^k \to y$. Since $y^k \in X$, we have that $y^k = (x_1^k, \ldots, x_m^k)$ with $x_i^k \in X_i$ for each i and k. Since $y^k \to y$, it follows that $x_i \in \mathrm{cl}(X_i)$ for each i, and hence $y \in \mathrm{cl}(X_1) \times \cdots \times \mathrm{cl}(X_m)$. Conversely, suppose that $y = (x_1, \ldots, x_m) \in \mathrm{cl}(X_1) \times \cdots \times \mathrm{cl}(X_m)$. This implies that there exist sequences $\{x_i^k\} \subset X_i$ such that $x_i^k \to x_i$ for each $i = 1, \ldots, m$. Since $x_i^k \in X_i$ for each i and k, we have that $y^k = (x_1^k, \ldots, x_m^k) \in X$ and $\{y^k\}$ converges to $y = (x_1, \ldots, x_m)$, implying that $y \in \mathrm{cl}(X)$.

Finally, we show the corresponding result for the affine hull of X. Let's assume, by using a translation argument if necessary, that all the X_i's contain the origin, so that $\mathrm{aff}(X_1), \ldots, \mathrm{aff}(X_m)$ as well as $\mathrm{aff}(X)$ are all subspaces.

Assume that $y \in \mathrm{aff}(X)$. Let the dimension of $\mathrm{aff}(X)$ be r, and let $y^1, \ldots, y^r$ be linearly independent vectors in X that span $\mathrm{aff}(X)$. Thus, we can represent y as

$$y = \sum_{i=1}^{r} \beta^i y^i,$$

where $\beta^1, \ldots, \beta^r$ are scalars. Since $y^i \in X$, we have that $y^i = (x_1^i, \ldots, x_m^i)$ with $x_j^i \in X_j$. Thus,

$$y = \sum_{i=1}^r \beta^i (x_1^i, \ldots, x_m^i) = \left(\sum_{i=1}^r \beta^i x_1^i, \ldots, \sum_{i=1}^r \beta^i x_m^i \right),$$

implying that $y \in \text{aff}(X_1) \times \cdots \times \text{aff}(X_m)$. Now, assume that $y \in \text{aff}(X_1) \times \cdots \times \text{aff}(X_m)$. Let the dimension of $\text{aff}(X_i)$ be r_i, and let $x_i^1, \ldots, x_i^{r_i}$ be linearly independent vectors in X_i that span $\text{aff}(X_i)$. Thus, we can represent y as

$$y = \left(\sum_{j=1}^{r_1} \beta_1^j x_1^j, \ldots, \sum_{j=1}^{r_m} \beta_m^j x_m^j \right).$$

Since each X_i contains the origin, we have that the vectors

$$\left(\sum_{j=1}^{r_1} \beta_1^j x_1^j, 0, \ldots, 0 \right), \left(0, \sum_{j=1}^{r_2} \beta_2^j x_2^j, 0, \ldots, 0 \right), \ldots, \left(0, \ldots, \sum_{j=1}^{r_m} \beta_m^j x_m^j \right),$$

belong to $\text{aff}(X)$, and so does their sum, which is the vector y. Thus, $y \in \text{aff}(X)$, concluding the proof.

(b) Assume that $y \in \text{cone}(X)$. We can represent y as

$$y = \sum_{i=1}^r \alpha^i y^i,$$

for some r, where $\alpha^1, \ldots, \alpha^r$ are nonnegative scalars and $y_i \in X$ for all i. Since $y^i \in X$, we have that $y^i = (x_1^i, \ldots, x_m^i)$ with $x_j^i \in X_j$. Thus,

$$y = \sum_{i=1}^r \alpha^i (x_1^i, \ldots, x_m^i) = \left(\sum_{i=1}^r \alpha^i x_1^i, \ldots, \sum_{i=1}^r \alpha^i x_m^i \right),$$

implying that $y \in \text{cone}(X_1) \times \cdots \times \text{cone}(X_m)$.

Conversely, assume that $y \in \text{cone}(X_1) \times \cdots \times \text{cone}(X_m)$. Then, we can represent y as

$$y = \left(\sum_{j=1}^{r_1} \alpha_1^j x_1^j, \ldots, \sum_{j=1}^{r_m} \alpha_m^j x_m^j \right),$$

where $x_i^j \in X_i$ and $\alpha_i^j \geq 0$ for each i and j. Since each X_i contains the origin, we have that the vectors

$$\left(\sum_{j=1}^{r_1} \alpha_1^j x_1^j, 0, \ldots, 0 \right), \left(0, \sum_{j=1}^{r_2} \alpha_2^j x_2^j, 0, \ldots, 0 \right) \ldots, \left(0, \ldots, \sum_{j=1}^{r_m} \alpha_m^j x_m^j \right),$$

belong to the cone(X), and so does their sum, which is the vector y. Thus, $y \in \text{cone}(X)$, concluding the proof.

Finally, consider the example where

$$X_1 = \{0, 1\} \subset \Re, \qquad X_2 = \{1\} \subset \Re.$$

For this example, $\text{cone}(X_1) \times \text{cone}(X_2)$ is given by the nonnegative quadrant, whereas cone(X) is given by the two halflines $\alpha(0,1)$ and $\alpha(1,1)$ for $\alpha \geq 0$ and the region that lies between them.

(c) We first show that

$$\text{ri}(X) = \text{ri}(X_1) \times \cdots \times \text{ri}(X_m).$$

Let $x = (x_1, \ldots, x_m) \in \text{ri}(X)$. Then, by Prop. 1.4.1 (c), we have that for all $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_m) \in X$, there exists some $\gamma > 1$ such that

$$x + (\gamma - 1)(x - \overline{x}) \in X.$$

Therefore, for all $\overline{x}_i \in X_i$, there exists some $\gamma > 1$ such that

$$x_i + (\gamma - 1)(x_i - \overline{x}_i) \in X_i,$$

which, by Prop. 1.4.1(c), implies that $x_i \in \text{ri}(X_i)$, i.e., $x \in \text{ri}(X_1) \times \cdots \times \text{ri}(X_m)$. Conversely, let $x = (x_1, \ldots, x_m) \in \text{ri}(X_1) \times \cdots \times \text{ri}(X_m)$. The above argument can be reversed through the use of Prop. 1.4.1(c), to show that $x \in \text{ri}(X)$. Hence, the result follows.

Finally, let us show that

$$R_X = R_{X_1} \times \cdots \times R_{X_m}.$$

Let $y = (y_1, \ldots, y_m) \in R_X$. By definition, this implies that for all $x \in X$ and $\alpha \geq 0$, we have $x + \alpha y \in X$. From this, it follows that for all $x_i \in X_i$ and $\alpha \geq 0$, $x_i + \alpha y_i \in X_i$, so that $y_i \in R_{X_i}$, implying that $y \in R_{X_1} \times \cdots \times R_{X_m}$. Conversely, let $y = (y_1, \ldots, y_m) \in R_{X_1} \times \cdots \times R_{X_m}$. By definition, for all $x_i \in X_i$ and $\alpha \geq 0$, we have $x_i + \alpha y_i \in X_i$. From this, we get for all $x \in X$ and $\alpha \geq 0$, $x + \alpha y \in X$, thus showing that $y \in R_X$.

1.38 (Recession Cones of Nonclosed Sets)

(a) Let $y \in R_C$. Then, by the definition of R_C, $x + \alpha y \in C$ for every $x \in C$ and every $\alpha \geq 0$. Since $C \subset \text{cl}(C)$, it follows that $x + \alpha y \in \text{cl}(C)$ for some $x \in \text{cl}(C)$ and every $\alpha \geq 0$, which, in view of part (b) of the Recession Cone Theorem (cf. Prop. 1.5.1), implies that $y \in R_{\text{cl}(C)}$. Hence

$$R_C \subset R_{\text{cl}(C)}.$$

By taking closures in this relation and by using the fact that $R_{\text{cl}(C)}$ is closed [part (a) of the Recession Cone Theorem], we obtain $\text{cl}(R_C) \subset R_{\text{cl}(C)}$.

To see that the inclusion $\mathrm{cl}(R_C) \subset R_{\mathrm{cl}(C)}$ can be strict, consider the set

$$C = \big\{(x_1, x_2) \mid 0 \le x_1,\ 0 \le x_2 < 1\big\} \cup \big\{(0, 1)\big\},$$

whose closure is

$$\mathrm{cl}(C) = \{(x_1, x_2) \mid 0 \le x_1,\ 0 \le x_2 \le 1\}.$$

The recession cones of C and its closure are

$$R_C = \big\{(0, 0)\big\}, \qquad R_{\mathrm{cl}(C)} = \big\{(x_1, x_2) \mid 0 \le x_1,\ x_2 = 0\big\}.$$

Thus, $\mathrm{cl}(R_C) = \big\{(0, 0)\big\}$, and $\mathrm{cl}(R_C)$ is a strict subset of $R_{\mathrm{cl}(C)}$.

(b) Let $y \in R_C$ and let x be a vector in C. Then we have $x + \alpha y \in C$ for all $\alpha \ge 0$. Thus for the vector x, which belongs to $\overline{C}$, we have $x + \alpha y \in \overline{C}$ for all $\alpha \ge 0$, and it follows from part (b) of the Recession Cone Theorem (cf. Prop. 1.5.1) that $y \in R_{\overline{C}}$. Hence, $R_C \subset R_{\overline{C}}$.

To see that the inclusion $R_C \subset R_{\overline{C}}$ can fail when $\overline{C}$ is not closed, consider the sets

$$C = \big\{(x_1, x_2) \mid x_1 \ge 0,\ x_2 = 0\big\}, \qquad \overline{C} = \big\{(x_1, x_2) \mid x_1 \ge 0,\ 0 \le x_2 < 1\big\}.$$

Their recession cones are

$$R_C = C = \big\{(x_1, x_2) \mid x_1 \ge 0,\ x_2 = 0\big\}, \qquad R_{\overline{C}} = \big\{(0, 0)\big\},$$

showing that R_C is not a subset of $R_{\overline{C}}$.

1.39 (Recession Cones of Relative Interiors)

(a) The inclusion $R_{\mathrm{ri}(C)} \subset R_{\mathrm{cl}(C)}$ follows from Exercise 1.38(b).

Conversely, let $y \in R_{\mathrm{cl}(C)}$, so that by the definition of $R_{\mathrm{cl}(C)}$, $x + \alpha y \in \mathrm{cl}(C)$ for every $x \in \mathrm{cl}(C)$ and every $\alpha \ge 0$. In particular, $x + \alpha y \in \mathrm{cl}(C)$ for every $x \in \mathrm{ri}(C)$ and every $\alpha \ge 0$. By the Line Segment Principle, all points on the line segment connecting x and $x + \alpha y$, except possibly $x + \alpha y$, belong to $\mathrm{ri}(C)$, implying that $x + \alpha y \in \mathrm{ri}(C)$ for every $x \in \mathrm{ri}(C)$ and every $\alpha \ge 0$. Hence, $y \in R_{\mathrm{ri}(C)}$, showing that $R_{\mathrm{cl}(C)} \subset R_{\mathrm{ri}(C)}$.

(b) If $y \in R_{\mathrm{ri}(C)}$, then by the definition of $R_{\mathrm{ri}(C)}$ for every vector $x \in \mathrm{ri}(C)$ and $\alpha \ge 0$, the vector $x + \alpha y$ is in $\mathrm{ri}(C)$, which holds in particular for some $x \in \mathrm{ri}(C)$ [note that $\mathrm{ri}(C)$ is nonempty by Prop. 1.4.1(b)].

Conversely, let y be such that there exists a vector $x \in \mathrm{ri}(C)$ with $x + \alpha y \in \mathrm{ri}(C)$ for all $\alpha \ge 0$. Hence, there exists a vector $x \in \mathrm{cl}(C)$ with $x + \alpha y \in \mathrm{cl}(C)$ for all $\alpha \ge 0$, which, by part (b) of the Recession Cone Theorem (cf. Prop. 1.5.1), implies that $y \in R_{\mathrm{cl}(C)}$. Using part (a), it follows that $y \in R_{\mathrm{ri}(C)}$, completing the proof.

(c) Using Exercise 1.38(c) and the assumption that $C \subset \overline{C}$ [which implies that $C \subset \overline{\mathrm{cl}}(C)$], we have

$$R_C \subset R_{\mathrm{cl}(\overline{C})} = R_{\mathrm{ri}(\overline{C})} = R_{\overline{C}},$$

where the equalities follow from part (a) and the assumption that $\overline{C} = \mathrm{ri}(\overline{C})$.

To see that the inclusion $R_C \subset R_{\overline{C}}$ can fail when $\overline{C} \neq \mathrm{ri}(\overline{C})$, consider the sets

$$C = \big\{(x_1, x_2) \mid x_1 \geq 0,\ 0 < x_2 < 1\big\}, \qquad \overline{C} = \big\{(x_1, x_2) \mid x_1 \geq 0,\ 0 \leq x_2 < 1\big\},$$

for which we have $C \subset \overline{C}$ and

$$R_C = \big\{(x_1, x_2) \mid x_1 \geq 0,\ x_2 = 0\big\}, \qquad R_{\overline{C}} = \big\{(0, 0)\big\},$$

showing that R_C is not a subset of $R_{\overline{C}}$.

1.40

For each k, consider the set $\overline{C}_k = X_k \cap C_k$. Note that $\{\overline{C}_k\}$ is a sequence of nonempty closed convex sets and X is specified by linear inequality constraints. We will show that, under the assumptions given in this exercise, the assumptions of Prop. 1.5.6 are satisfied, thus showing that the intersection $X \cap (\cap_{k=0}^{\infty} \overline{C}_k)$ [which is equal to the intersection $\cap_{k=0}^{\infty}(X_k \cap C_k)$] is nonempty.

Since $X_{k+1} \subset X_k$ and $C_{k+1} \subset C_k$ for all k, it follows that

$$\overline{C}_{k+1} \subset \overline{C}_k, \qquad \forall\ k,$$

showing that assumption (1) of Prop. 1.5.6 is satisfied. Similarly, since by assumption $X_k \cap C_k$ is nonempty for all k, we have that, for all k, the set

$$X \cap \overline{C}_k = X \cap X_k \cap C_k = X_k \cap C_k,$$

is nonempty, showing that assumption (2) is satisfied. Finally, let R denote the set $R = \cap_{k=0}^{\infty} R_{\overline{C}_k}$. Since by assumption $\overline{C}_k$ is nonempty for all k, we have, by part (e) of the Recession Cone Theorem, that $R_{\overline{C}_k} = R_{X_k} \cap R_{C_k}$ implying that

$$\begin{aligned} R &= \cap_{k=0}^{\infty} R_{\overline{C}_k} \\ &= \cap_{k=0}^{\infty} (R_{X_k} \cap R_{C_k}) \\ &= \big(\cap_{k=0}^{\infty} R_{X_k}\big) \cap \big(\cap_{k=0}^{\infty} R_{C_k}\big) \\ &= R_X \cap R_C. \end{aligned}$$

Similarly, letting L denote the set $L = \cap_{k=0}^{\infty} L_{\overline{C}_k}$, it can be seen that $L = L_X \cap L_C$. Since, by assumption $R_X \cap R_C \subset L_C$, it follows that

$$R_X \cap R = R_X \cap R_C \subset L_C,$$

which, in view of the assumption that $R_X = L_X$, implies that

$$R_X \cap R \subset L_C \cap L_X = L,$$

showing that assumption (3) of Prop. 1.5.6 is satisfied, and thus proving that the intersection $X \cap (\cap_{k=0}^{\infty} \overline{C}_k)$ is nonempty.

1.41

Let y be in the closure of $A \cdot C$. We will show that $y = Ax$ for some $x \in \mathrm{cl}(C)$. For every $\epsilon > 0$, the set

$$C_\epsilon = \mathrm{cl}(C) \cap \big\{x \mid \|y - Ax\| \leq \epsilon\big\}$$

is closed. Since $A \cdot C \subset A \cdot \mathrm{cl}(C)$ and $y \in \mathrm{cl}(A \cdot C)$, it follows that y is in the closure of $A \cdot \mathrm{cl}(C)$, so that C_ϵ is nonempty for every $\epsilon > 0$. Furthermore, the recession cone of the set $\big\{x \mid \|Ax - y\| \leq \epsilon\big\}$ coincides with the null space $N(A)$, so that $R_{C_\epsilon} = R_{\mathrm{cl}(C)} \cap N(A)$. By assumption we have $R_{\mathrm{cl}(C)} \cap N(A) = \{0\}$, and by part (c) of the Recession Cone Theorem (cf. Prop. 1.5.1), it follows that C_ϵ is bounded for every $\epsilon > 0$. Now, since the sets C_ϵ are nested nonempty compact sets, their intersection $\cap_{\epsilon > 0} C_\epsilon$ is nonempty. For any x in this intersection, we have $x \in \mathrm{cl}(C)$ and $Ax - y = 0$, showing that $y \in A \cdot \mathrm{cl}(C)$. Hence, $\mathrm{cl}(A \cdot C) \subset A \cdot \mathrm{cl}(C)$. The converse $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$ is clear, since for any $x \in \mathrm{cl}(C)$ and sequence $\{x_k\} \subset C$ converging to x, we have $Ax_k \to Ax$, showing that $Ax \in \mathrm{cl}(A \cdot C)$. Therefore,

$$\mathrm{cl}(A \cdot C) = A \cdot \mathrm{cl}(C). \tag{1.13}$$

We now show that $A \cdot R_{\mathrm{cl}(C)} = R_{A \cdot \mathrm{cl}(C)}$. Let $y \in A \cdot R_{\mathrm{cl}(C)}$. Then, there exists a vector $u \in R_{\mathrm{cl}(C)}$ such that $Au = y$, and by the definition of $R_{\mathrm{cl}(C)}$, there is a vector $x \in \mathrm{cl}(C)$ such that $x + \alpha u \in \mathrm{cl}(C)$ for every $\alpha \geq 0$. Therefore, $Ax + \alpha Au \in A \cdot \mathrm{cl}(C)$ for every $\alpha \geq 0$, which, together with $Ax \in A \cdot \mathrm{cl}(C)$ and $Au = y$, implies that y is a direction of recession of the closed set $A \cdot \mathrm{cl}(C)$ [cf. Eq. (1.13)]. Hence, $A \cdot R_{\mathrm{cl}(C)} \subset R_{A \cdot \mathrm{cl}(C)}$.

Conversely, let $y \in R_{A \cdot \mathrm{cl}(C)}$. We will show that $y \in A \cdot R_{\mathrm{cl}(c)}$. This is true if $y = 0$, so assume that $y \neq 0$. By definition of direction of recession, there is a vector $z \in A \cdot \mathrm{cl}(C)$ such that $z + \alpha y \in A \cdot \mathrm{cl}(C)$ for every $\alpha \geq 0$. Let $x \in \mathrm{cl}(C)$ be such that $Ax = z$, and for every positive integer k, let $x_k \in \mathrm{cl}(C)$ be such that $Ax_k = z + ky$. Since $y \neq 0$, the sequence $\{Ax_k\}$ is unbounded, implying that $\{x_k\}$ is also unbounded (if $\{x_k\}$ were bounded, then $\{Ax_k\}$ would be bounded, a contradiction). Because $x_k \neq x$ for all k, we can define

$$u_k = \frac{x_k - x}{\|x_k - x\|}, \qquad \forall\ k.$$

Let u be a limit point of $\{u_k\}$, and note that $u \neq 0$. It can be seen that u is a direction of recession of $\mathrm{cl}(C)$ [this can be done similar to the proof of part (c) of the Recession Cone Theorem (cf. Prop. 1.5.1)]. By taking an appropriate subsequence if necessary, we may assume without loss of generality that $\lim_{k\to\infty} u_k = u$. Then, by the choices of u_k and x_k, we have

$$Au = \lim_{k\to\infty} Au_k = \lim_{k\to\infty} \frac{Ax_k - Ax}{\|x_k - x\|} = \lim_{k\to\infty} \frac{k}{\|x_k - x\|} y,$$

implying that $\lim_{k\to\infty} \frac{k}{\|x_k - x\|}$ exists. Denote this limit by λ. If $\lambda = 0$, then u is in the null space $N(A)$, implying that $u \in R_{\mathrm{cl}(C)} \cap N(A)$. By the given condition $R_{\mathrm{cl}(C)} \cap N(A) = \{0\}$, we have $u = 0$ contradicting the fact $u \neq 0$. Thus, λ is positive and $Au = \lambda y$, so that $A(u/\lambda) = y$. Since $R_{\mathrm{cl}(C)}$ is a cone [part (a) of the Recession Cone Theorem] and $u \in R_{\mathrm{cl}(C)}$, the vector u/λ is in $R_{\mathrm{cl}(C)}$, so that y belongs to $A \cdot R_{\mathrm{cl}(C)}$. Hence, $R_{A\cdot \mathrm{cl}(C)} \subset A \cdot R_{\mathrm{cl}(C)}$, completing the proof.

As an example showing that $A\cdot R_{\mathrm{cl}(C)}$ and $R_{A\cdot \mathrm{cl}(C)}$ may differ when $R_{\mathrm{cl}(C)} \cap N(A) \neq \{0\}$, consider the set

$$C = \big\{(x_1, x_2) \mid x_1 \in \Re,\ x_2 \geq x_1^2\big\},$$

and the linear transformation A that maps $(x_1, x_2) \in \Re^2$ into $x_1 \in \Re$. Then, C is closed and its recession cone is

$$R_C = \big\{(x_1, x_2) \mid x_1 = 0,\ x_2 \geq 0\big\},$$

so that $A \cdot R_C = \{0\}$, where 0 is scalar. On the other hand, $A \cdot C$ coincides with $\Re$, so that $R_{A\cdot C} = \Re \neq A \cdot R_C$.

1.42

Let S be defined by

$$S = R_{\mathrm{cl}(C)} \cap N(A),$$

and note that S is a subspace of $L_{\mathrm{cl}(C)}$ by the given assumption. Then, by Lemma 1.5.4, we have

$$\mathrm{cl}(C) = \big(\mathrm{cl}(C) \cap S^\perp\big) + S,$$

so that the images of $\mathrm{cl}(C)$ and $\mathrm{cl}(C) \cap S^\perp$ under A coincide [since $S \subset N(A)$], i.e.,

$$A \cdot \mathrm{cl}(C) = A \cdot \big(\mathrm{cl}(C) \cap S^\perp\big). \tag{1.14}$$

Because $A \cdot C \subset A \cdot \mathrm{cl}(C)$, we have

$$\mathrm{cl}(A \cdot C) \subset \mathrm{cl}\big(A \cdot \mathrm{cl}(C)\big),$$

which in view of Eq. (1.14) gives

$$\mathrm{cl}(A \cdot C) \subset \mathrm{cl}\Big(A \cdot \big(\mathrm{cl}(C) \cap S^\perp\big)\Big).$$

Define

$$\overline{C} = \mathrm{cl}(C) \cap S^\perp$$

so that the preceding relation becomes

$$\mathrm{cl}(A \cdot C) \subset \mathrm{cl}(A \cdot \overline{C}). \tag{1.15}$$

The recession cone of $\overline{C}$ is given by

$$R_{\overline{C}} = R_{\mathrm{cl}(C)} \cap S^\perp, \tag{1.16}$$

[cf. part (e) of the Recession Cone Theorem, Prop. 1.5.1], for which, since $S = R_{\mathrm{cl}(C)} \cap N(A)$, we have

$$R_{\overline{C}} \cap N(A) = S \cap\ S^{\perp} = \{0\}.$$

Therefore, by Prop. 1.5.8, the set $A \cdot \overline{C}$ is closed, implying that $\mathrm{cl}(A \cdot \overline{C}) = A \cdot \overline{C}$. By the definition of $\overline{C}$, we have $A \cdot \overline{C} \subset A \cdot \mathrm{cl}(C)$, implying that $\mathrm{cl}(A \cdot \overline{C}) \subset A \cdot \mathrm{cl}(C)$ which together with Eq. (1.15) yields $\mathrm{cl}(A \cdot C) \subset A \cdot \mathrm{cl}(C)$. The converse $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$ is clear, since for any $x \in \mathrm{cl}(C)$ and sequence $\{x_k\} \subset C$ converging to x, we have $Ax_k \to Ax$, showing that $Ax \in \mathrm{cl}(A \cdot C)$. Therefore,

$$\mathrm{cl}(A \cdot C) = A \cdot \mathrm{cl}(C). \tag{1.17}$$

We next show that $A \cdot R_{\mathrm{cl}(C)} = R_{A \cdot \mathrm{cl}(C)}$. Let $y \in A \cdot R_{\mathrm{cl}(C)}$. Then, there exists a vector $u \in R_{\mathrm{cl}(C)}$ such that $Au = y$, and by the definition of $R_{\mathrm{cl}(C)}$, there is a vector $x \in \mathrm{cl}(C)$ such that $x + \alpha u \in \mathrm{cl}(C)$ for every $\alpha \geq 0$. Therefore, $Ax + \alpha Au \in A\mathrm{cl}(C)$ for some $x \in \mathrm{cl}(C)$ and for every $\alpha \geq 0$, which together with $Ax \in A \cdot \mathrm{cl}(C)$ and $Au = y$ implies that y is a recession direction of the closed set $A \cdot \mathrm{cl}(C)$ [Eq. (1.17)]. Hence, $A \cdot R_{\mathrm{cl}(C)} \subset R_{A \cdot \mathrm{cl}(C)}$.

Conversely, in view of Eq. (1.14) and the definition of $\overline{C}$, we have

$$R_{A \cdot \mathrm{cl}(C)} = R_{A \cdot \overline{C}}.$$

Since $R_{\overline{C}} \cap N(A) = \{0\}$ and $\overline{C}$ is closed, by Exercise 1.41, it follows that

$$R_{A \cdot \overline{C}} = A \cdot R_{\overline{C}},$$

which combined with Eq. (1.16) implies that

$$A \cdot R_{\overline{C}} \subset A \cdot R_{\mathrm{cl}(C)}.$$

The preceding three relations yield $R_{A \cdot \mathrm{cl}(C)} \subset A \cdot R_{\mathrm{cl}(C)}$, completing the proof.

1.43 (Recession Cones of Vector Sums)

(a) Let C be the Cartesian product $C_1 \times \cdots \times C_m$. Then, by Exercise 1.37, C is closed, and its recession cone and lineality space are given by

$$R_C = R_{C_1} \times \cdots \times R_{C_m}, \qquad L_C = L_{C_1} \times \cdots \times L_{C_m}.$$

Let A be a linear transformation that maps $(x_1, \ldots, x_m) \in \Re^{mn}$ into $x_1 + \cdots + x_m \in \Re^n$. The null space of A is the set of all $(y_1, \ldots, y_m)$ such that $y_1 + \cdots + y_m = 0$. The intersection $R_C \cap N(A)$ consists of points $(y_1, \ldots, y_m)$ such that $y_1 + \cdots + y_m = 0$ with $y_i \in R_{C_i}$ for all i. By the given condition, every vector $(y_1, \ldots, y_m)$ in the intersection $R_C \cap N(A)$ is such that $y_i \in L_{C_i}$ for all i, implying that $(y_1, \ldots, y_m)$ belongs to the lineality space L_C. Thus, $R_C \cap N(A) \subset L_C \cap N(A)$. On the other hand by definition of the lineality space, we have $L_C \subset R_C$, so that $L_C \cap N(A) \subset R_C \cap N(A)$. Therefore, $R_C \cap N(A) = L_C \cap N(A)$, implying that

$R_C \cap N(A)$ is a subspace of L_C. By Exercise 1.42, the set $A \cdot C$ is closed and $R_{A \cdot C} = A \cdot R_C$. Since $A \cdot C = C_1 + \cdots + C_m$, the assertions of part (a) follow.

(b) The proof is similar to that of part (a). Let C be the Cartesian product $C_1 \times \cdots \times C_m$. Then, by Exercise 1.37(a),

$$\mathrm{cl}(C) = \mathrm{cl}(C_1) \times \cdots \times \mathrm{cl}(C_m), \tag{1.18}$$

and its recession cone and lineality space are given by

$$R_{\mathrm{cl}(C)} = R_{\mathrm{cl}(C_1)} \times \cdots \times R_{\mathrm{cl}(C_m)}, \tag{1.19}$$

$$L_{\mathrm{cl}(C)} = L_{\mathrm{cl}(C_1)} \times \cdots \times L_{\mathrm{cl}(C_m)}.$$

Let A be a linear transformation that maps $(x_1, \ldots, x_m) \in \Re^{mn}$ into $x_1 + \cdots + x_m \in \Re^n$. Then, the intersection $R_{\mathrm{cl}}(C) \cap N(A)$ consists of points $(y_1, \ldots, y_m)$ such that $y_1 + \cdots + y_m = 0$ with $y_i \in R_{\mathrm{cl}(C_i)}$ for all i. By the given condition, every vector $(y_1, \ldots, y_m)$ in the intersection $R_{\mathrm{cl}(C)} \cap N(A)$ is such that $y_i \in L_{\mathrm{cl}(C_i)}$ for all i, implying that $(y_1, \ldots, y_m)$ belongs to the lineality space $L_{\mathrm{cl}(C)}$. Thus, $R_{\mathrm{cl}(C)} \cap N(A) \subset L_{\mathrm{cl}(C)} \cap N(A)$. On the other hand by definition of the lineality space, we have $L_{\mathrm{cl}(C)} \subset R_{\mathrm{cl}(C)}$, so that $L_{\mathrm{cl}(C)} \cap N(A) \subset R_{\mathrm{cl}(C)} \cap N(A)$. Hence, $R_{\mathrm{cl}(C)} \cap N(A) = L_{\mathrm{cl}(C)} \cap N(A)$, implying that $R_{\mathrm{cl}(C)} \cap N(A)$ is a subspace of $L_{\mathrm{cl}(C)}$. By Exercise 1.42, we have $\mathrm{cl}(A \cdot C) = A \cdot \mathrm{cl}(C)$ and $R_{A \cdot \mathrm{cl}(C)} = A \cdot R_{\mathrm{cl}(C)}$, from which by using the relation $A \cdot C = C_1 + \cdots + C_m$, and Eqs. (1.18) and (1.19), we obtain

$$\mathrm{cl}(C_1 + \cdots + C_m) = \mathrm{cl}(C_1) + \cdots + \mathrm{cl}(C_m),$$

$$R_{\mathrm{cl}(C_1 + \cdots + C_m)} = R_{\mathrm{cl}(C_1)} + \cdots + R_{\mathrm{cl}(C_m)}.$$

1.44

Let C be the Cartesian product $C_1 \times \cdots \times C_m$ viewed as a subset of $\Re^{mn}$, and let A be the linear transformation that maps a vector $(x_1, \ldots, x_m) \in \Re^{mn}$ into $x_1 + \cdots + x_m$. Note that set C can be written as

$$C = \left\{ x = (x_1, \ldots, x_m) \mid x' \overline{Q}_{ij} x + \overline{a}'_{ij} x + b_{ij} \le 0,\ i = 1, \ldots, m,\ j = 1, \ldots, r_i \right\},$$

where the $\overline{Q}_{ij}$ are appropriately defined symmetric positive semidefinite $mn \times mn$ matrices and the $\overline{a}_{ij}$ are appropriately defined vectors in $\Re^{mn}$. Hence, the set C is specified by convex quadratic inequalities. Thus, we can use Prop. 1.5.8(c) to assert that the set $AC = C_1 + \cdots + C_m$ is closed.

1.45 (Set Intersection and Helly's Theorem)

Helly's Theorem implies that the sets $\overline{C}_k$ defined in the hint are nonempty. These sets are also nested and satisfy the assumptions of Props. 1.5.5 and 1.5.6. Therefore, the intersection $\cap_{i=1}^{\infty} \overline{C}_i$ is nonempty. Since

$$\cap_{i=1}^{\infty} \overline{C}_i \subset \cap_{i=1}^{\infty} C_i,$$

the result follows.

Convex Analysis and Optimization

Chapter 2 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 2: SOLUTION MANUAL

2.1

(a) If x^* is a global minimum of f, evidently it is also a global minimum of f along any line passing through x^*.

Conversely, let x^* be a global minimum of f along any line passing through x^*, i.e., for all $d \in \Re^n$, the function $g : \Re \mapsto \Re$, defined by $g(\alpha) = f(x^* + \alpha d)$, has $\alpha^* = 0$ as its global minimum. Assume, to arrive at a contradiction, that x^* is not a global minimum of f. This implies that there exists some $\overline{x} \in \Re^n$ such that $f(\overline{x}) < f(x^*)$. Let $d = \overline{x} - x^*$. In view of the assumption that $g(\alpha)$ has $\alpha^* = 0$ as its global minimum, it follows that $g(0) \leq g(\frac{1}{2})$, or equivalently,

$$\begin{aligned} f(x^*) &\leq f\left(x^* + \frac{1}{2}(\overline{x} - x^*)\right), \\ &= f\left(\frac{1}{2}\overline{x} + \frac{1}{2}x^*\right), \\ &\leq \frac{1}{2}f(\overline{x}) + \frac{1}{2}f(x^*), \\ &< f(x^*), \end{aligned}$$

which is a contradiction. [Note that in the above string of equations, the second inequality follows by the convexity of f, and the strict inequality follows from the assumption that $f(\overline{x}) < f(x^*)$.] Hence, it follows that x^* is a global minimum of f.

(b) Consider the function $f(x_1, x_2) = (x_2 - px_1^2)(x_2 - qx_1^2)$, where $0 < p < q$ and let $x^* = (0, 0)$.

We first show that $g(\alpha) = f(x^* + \alpha d)$ is minimized at $\alpha = 0$ for all $d \in \Re^2$. We have

$$g(\alpha) = f(x^* + \alpha d) = (\alpha d_2 - p\alpha^2 d_1^2)(\alpha d_2 - q\alpha^2 d_1^2) = \alpha^2(d_2 - p\alpha d_1^2)(d_2 - q\alpha d_1^2).$$

Also,

$$g'(\alpha) = 2\alpha(d_2 - p\alpha d_1^2)(d_2 - q\alpha d_1^2) + \alpha^2(-pd_1^2)(d_2 - q\alpha d_1^2) + \alpha^2(d_2 - p\alpha d_1^2)(-qd_1^2).$$

Thus $g'(0) = 0$. Furthermore,

$$\begin{aligned} g''(\alpha) = {} & 2(d_2 - p\alpha d_1^2)(d_2 - q\alpha d_1^2) + 2\alpha(-pd_1^2)(d_2 - q\alpha d_1^2) \\ & + 2\alpha(d_2 - p\alpha d_1^2)(-qd_1^2) + 2\alpha(-pd_1^2)(d_2 - q\alpha d_1^2) + \alpha^2(-pd_1^2)(-qd_1^2) \\ & + 2\alpha(d_2 - p\alpha d_1^2)(-qd_1^2) + \alpha^2(-pd_1^2)(-qd_1^2). \end{aligned}$$

Thus $g''(0) = 2d_2^2$, which is greater than 0 if $d_2 \neq 0$. If $d_2 = 0$, $g(\alpha) = pq\alpha^4 d_1^4$, which is clearly minimized at $\alpha = 0$. Therefore, $(0,0)$ is a local minimum of f along every line that passes through $(0,0)$.

Let's now show that if $p < m < q$, $f(y, my^2) < 0$ if $y \neq 0$ and that $f(y, my^2) = 0$ otherwise. Consider a point of the form (y, my^2). We have $f(y, my^2) = y^4(m-p)(m-q)$. Clearly, $f(y, my^2) < 0$ if and only if $p < m < q$ and $y \neq 0$. In any $\epsilon-$neighborhood of $(0,0)$, there exists a $y \neq 0$ such that for some $m \in (p,q)$, (y, my^2) also belongs to the neighborhood. Since $f(0,0) = 0$, we see that $(0,0)$ is not a local minimum.

2.2 (Lipschitz Continuity of Convex Functions)

Let ϵ be a positive scalar and let C_ϵ be the set given by

$$C_\epsilon = \big\{ z \mid \|z - x\| \leq \epsilon, \text{ for some } x \in \mathrm{cl}(X) \big\}.$$

We claim that the set C_ϵ is compact. Indeed, since X is bounded, so is its closure, which implies that $\|z\| \leq \max_{x \in \mathrm{cl}(X)} \|x\| + \epsilon$ for all $z \in C_\epsilon$, showing that C_ϵ is bounded. To show the closedness of C_ϵ, let $\{z_k\}$ be a sequence in C_ϵ converging to some z. By the definition of C_ϵ, there is a corresponding sequence $\{x_k\}$ in $\mathrm{cl}(X)$ such that

$$\|z_k - x_k\| \leq \epsilon, \qquad \forall\, k. \tag{2.1}$$

Because $\mathrm{cl}(X)$ is compact, $\{x_k\}$ has a subsequence converging to some $x \in \mathrm{cl}(X)$. Without loss of generality, we may assume that $\{x_k\}$ converges to $x \in \mathrm{cl}(X)$. By taking the limit in Eq. (2.1) as $k \to \infty$, we obtain $\|z - x\| \leq \epsilon$ with $x \in \mathrm{cl}(X)$, showing that $z \in C_\epsilon$. Hence, C_ϵ is closed.

We now show that f has the Lipschitz property over X. Let x and y be two distinct points in X. Then, by the definition of C_ϵ, the point

$$z = y + \frac{\epsilon}{\|y - x\|}(y - x)$$

is in C_ϵ. Thus

$$y = \frac{\|y - x\|}{\|y - x\| + \epsilon} z + \frac{\epsilon}{\|y - x\| + \epsilon} x,$$

showing that y is a convex combination of $z \in C_\epsilon$ and $x \in C_\epsilon$. By convexity of f, we have

$$f(y) \leq \frac{\|y - x\|}{\|y - x\| + \epsilon} f(z) + \frac{\epsilon}{\|y - x\| + \epsilon} f(x),$$

implying that

$$f(y) - f(x) \leq \frac{\|y - x\|}{\|y - x\| + \epsilon}\big(f(z) - f(x)\big) \leq \frac{\|y - x\|}{\epsilon}\left(\max_{u \in C_\epsilon} f(u) - \min_{v \in C_\epsilon} f(v)\right),$$

where in the last inequality we use Weierstrass' theorem (f is continuous over $\Re^n$ by Prop. 1.4.6 and C_ϵ is compact). By switching the roles of x and y, we similarly obtain

$$f(x) - f(y) \leq \frac{\|x - y\|}{\epsilon}\left(\max_{u \in C_\epsilon} f(u) - \min_{v \in C_\epsilon} f(v)\right),$$

which combined with the preceding relation yields $\big|f(x) - f(y)\big| \leq L\|x - y\|$, where $L = \big(\max_{u \in C_\epsilon} f(u) - \min_{v \in C_\epsilon} f(v)\big)/\epsilon$.

2.3 (Exact Penalty Functions)

(a) By assumption, x^* minimizes f over X, so that $x^* \in X$, and we have for all $c > L$, $y \in X$, and $x \in Y$,

$$F_c(x^*) = f(x^*) \leq f(y) \leq f(x) + L\|y - x\| \leq f(x) + c\|y - x\|,$$

where we use the Lipschitz continuity of f to get the second inequality. Taking the infimum over all $y \in X$, we obtain

$$F_c(x^*) \leq f(x) + c \inf_{y \in X} \|y - x\| = F_c(x), \qquad \forall\, x \in Y.$$

Hence, x^* minimizes $F_c(x)$ over Y for all $c > L$. (Note that the infimum in the preceding relation is attained by Weierstrass' Theorem, since X is closed by assumption, and $\|\cdot\|$ is a continuous function that has compact level sets.)

(b) Suppose, to arrive at a contradiction, that x^* minimizes $F_c(x)$ over Y, but $x^* \notin X$.

We have that $F_c(x^*) = f(x^*) + c \min_{y \in X} \|y - x^*\|$. Using the argument given earlier, the minimum of $\|y - x\|$ over $y \in X$ is attained at some $\tilde{x} \in X$, which is not equal to x^*, and therefore,

$$\begin{aligned} F_c(x^*) &= f(x^*) + c\|\tilde{x} - x^*\| \\ &> f(x^*) + L\|\tilde{x} - x^*\| \\ &\geq f(\tilde{x}) \\ &= F_c(\tilde{x}), \end{aligned}$$

which contradicts the fact that x^* minimizes $F_c(x)$ over Y. (Note that the first inequality follows from $c > L$ and $\tilde{x} \neq x^*$. The second inequality follows from the Lipschitz continuity of function f.) Hence, if x^* minimizes $F_c(x)$ over Y, it follows that $x^* \in X$.

2.4 (Ekeland's Variational Principle [Eke74])

For some $\delta > 0$, define the function $F : \Re^n \mapsto (-\infty, \infty]$ by

$$F(x) = f(x) + \delta\|x - \overline{x}\|.$$

The function F is closed in view of the assumption that f is closed. Hence, by Prop. 1.2.2(b), it follows that all the level sets of F are closed. The level sets are also bounded, since for all $\gamma > f^*$, we have

$$\big\{x \mid F(x) \leq \gamma\big\} \subset \big\{x \mid f^* + \delta\|x - \overline{x}\| \leq \gamma\big\} = B\left(\overline{x}, \frac{\gamma - f^*}{\delta}\right), \tag{2.2}$$

where $B\big(\overline{x}, (\gamma - f^*)/\delta\big)$ denotes the closed ball centered at $\overline{x}$ with radius $(\gamma - f^*)/\delta$. Hence, it follows by Weierstrass' Theorem that F attains a minimum over $\Re^n$, i.e., the set $\arg\min_{x\in\Re^n} F(x)$ is nonempty and compact.

Consider now minimizing f over the set $\arg\min_{x\in\Re^n} F(x)$. Since f is closed by assumption, we conclude by using Weierstrass' Theorem that f attains a minimum at some $\tilde{x}$ over the set $\arg\min_{x\in\Re^n} F(x)$. Hence, we have

$$f(\tilde{x}) \le f(x), \qquad \forall\ x \in \arg\min_{x\in\Re^n} F(x). \tag{2.3}$$

Since $\tilde{x} \in \arg\min_{x\in\Re^n} F(x)$, it follows that $F(\tilde{x}) \le F(x)$, for all $x \in \Re^n$, and

$$F(\tilde{x}) < F(x), \qquad \forall\ x \notin \arg\min_{x\in\Re^n} F(x),$$

which by using the triangle inequality implies that

$$\begin{aligned} f(\tilde{x}) &< f(x) + \delta\|x - \overline{x}\| - \delta\|\tilde{x} - \overline{x}\| \\ &\le f(x) + \delta\|x - \tilde{x}\|, \qquad \forall\ x \notin \arg\min_{x\in\Re^n} F(x). \end{aligned} \tag{2.4}$$

Using Eqs. (2.3) and (2.4), we see that

$$f(\tilde{x}) < f(x) + \delta\|x - \tilde{x}\|, \qquad \forall\ x \ne \tilde{x},$$

thereby implying that $\tilde{x}$ is the unique optimal solution of the problem of minimizing $f(x) + \delta\|x - \tilde{x}\|$ over $\Re^n$.

Moreover, since $F(\tilde{x}) \le F(x)$ for all $x \in \Re^n$, we have $F(\tilde{x}) \le F(\overline{x})$, which implies that

$$f(\tilde{x}) \le f(\overline{x}) - \delta\|\tilde{x} - \overline{x}\| \le f(\overline{x}),$$

and also

$$F(\tilde{x}) \le F(\overline{x}) = f(\overline{x}) \le f^* + \epsilon.$$

Using Eq. (2.2), it follows that $\tilde{x} \in B(\overline{x}, \epsilon/\delta)$, proving the desired result.

2.5 (Approximate Minima of Convex Functions)

(a) Let $\epsilon > 0$ be given. Assume, to arrive at a contradiction, that for any sequence $\{\delta_k\}$ with $\delta_k \downarrow 0$, there exists a sequence $\{x_k\} \subset X$ such that for all k

$$f(x_k) \le f^* + \delta_k, \qquad \min_{x^*\in X^*} \|x_k - x^*\| \ge \epsilon.$$

It follows that, for some $\overline{k}$, x_k belongs to the set $\big\{x \in X \mid f(x) \le f^* + \delta_{\overline{k}}\big\}$, for all $k \ge \overline{k}$. Since by assumption f and X have no common nonzero direction of recession, by the Recession Cone Theorem, we have that the closed convex set $\big\{x \in X \mid f(x) \le f^* + \delta_{\overline{k}}\big\}$ is bounded. Therefore, the sequence $\{x_k\}$ is bounded and has a limit point $\overline{x} \in X$, which, in view of the preceding relations, satisfies

$$f(\overline{x}) \le f^*, \qquad \|\overline{x} - x^*\| \ge \epsilon, \quad \forall\ x^* \in X^*,$$

which is a contradiction. This proves that, for every $\epsilon > 0$, there exists a $\delta > 0$ such that every vector $x \in X$ with $f(x) \leq f^* + \delta$ satisfies $\min_{x^* \in X^*} \|x - x^*\| < \epsilon$.

(b) Fix $\epsilon > 0$. By part (a), there exists some $\delta_\epsilon > 0$ such that every vector $x \in X$ with $f(x) \leq f^* + \delta_\epsilon$ satisfies $\min_{x^* \in X^*} \|x - x^*\| \leq \epsilon$. Since $f(x_k) \to f^*$, there exists some K_ϵ such that

$$f(x_k) \leq f^* + \delta_\epsilon, \qquad \forall\ k \geq K_\epsilon.$$

By part (a), this implies that $\{x_k\}_{k \geq K_\epsilon} \subset X^* + \epsilon B$. Since X^* is nonempty and compact (cf. Prop. 2.3.2), it follows that every such sequence $\{x_k\}$ is bounded.

Let $\overline{x}$ be a limit point of the sequence $\{x_k\} \subset X$ satisfying $f(x_k) \to f^*$. By lower semicontinuity of the function f, we get that

$$f(\overline{x}) \leq \liminf_{k \to \infty} f(x_k) = f^*.$$

Because $\{x_k\} \subset X$ and X is closed, we have $\overline{x} \in X$, which in view of the preceding relation implies that $f(\overline{x}) = f^*$, i.e., $\overline{x} \in X^*$.

2.6 (Directions Along Which a Function is Flat)

(a) Let $d \in R_X \cap R_f$. If $d \notin F_f$, then we must have $\lim_{\alpha \to \infty} f(x + \alpha d) = -\infty$, for some $x \in \mathrm{dom}(f) \cap X$, implying that for each k, $x + \alpha d \in C_k$ for all sufficiently large α. Thus d is a direction of recession of C_k and hence also of f, i.e., $d \in R_f$. Therefore, we must have $R_X \cap F_f = R_X \cap R_f$, so using the hypothesis, we have $R_X \cap R_f \subset L_f$. From Prop. 1.5.6 [condition (3)], it follows that $X \cap \big(\cap_{k=0}^\infty C_k\big)$ is nonempty.

(b) Let $d \in R_X \cap R_f$. If $d \notin F_f$, then we must have $\lim_{\alpha \to \infty} f(x + \alpha d) = -\infty$, for some $x \in \mathrm{dom}(f) \cap X$. Since $d \in R_X$, we have $x + \alpha d \in X$ for all $x \in X$ and $\alpha \gamma e 0$. It follows that $\inf_{x \in X} f(x) = -\infty$, a contradiction. Therefore, we must have $R_X \cap F_f = R_X \cap R_f$, so using the hypothesis, we have $R_X \cap R_f \subset L_f$. From Prop. 2.3.3 [condition (2)], it follows that there exists at least one optimal solution.

(c) Let $X = \Re$ and $f(x) = x$. Then

$$F_f = L_f = \big\{y \mid y = 0\big\},$$

so the condition $R_X \cap F_f \subset L_f$ is satisfied. However, we have $\inf_{x \in X} f(x) = -\infty$ and f does not attain a minimum over X. Note that Prop. 2.3.3 [under condition (2)] does not apply here, because the relation $R_X \cap R_f \subset L_f$ is not satisfied.

2.7 (Bidirectionally Flat Functions)

(a) As a first step, we will show that either $\cap_{k=1}^\infty C_k \neq \emptyset$ or else

$$\text{there exists } \overline{j} \in \{1, \ldots, r\} \text{ and } y \in \cap_{j=0}^r R_{g_j} \text{ with } y \notin F_{g_{\overline{j}}}.$$

Let $\overline{x}$ be a vector in C_0, and for each $k \geq 1$, let x_k be the projection of $\overline{x}$ on C_k. If $\{x_k\}$ is bounded, then since the g_j are closed, any limit point $\tilde{x}$ of $\{x_k\}$ satisfies

$$g_j(\tilde{x}) \leq \liminf_{k\to\infty} g_j(x_k) \leq 0,$$

so $\tilde{x} \in \cap_{k=1}^{\infty} C_k$, and $\cap_{k=1}^{\infty} C_k \neq \varnothing$. If $\{x_k\}$ is unbounded, let y be a limit point of the sequence $\big\{(x_k - \overline{x})/\|x_k - \overline{x}\| \mid x_k \neq \overline{x}\big\}$, and without loss of generality, assume that

$$\frac{x_k - \overline{x}}{\|x_k - \overline{x}\|} \to y.$$

We claim that

$$y \in \cap_{j=0}^{r} R_{g_j}.$$

Indeed, if for some j, we have $y \notin R_{g_j}$, then there exists $\alpha > 0$ such that $g_j(\overline{x} + \alpha y) > w_0$. Let

$$z_k = \overline{x} + \alpha \frac{x_k - \overline{x}}{\|x_k - \overline{x}\|},$$

and note that for sufficiently large k, z_k lies in the line segment connecting $\overline{x}$ and x_k, so that $g_1(z_k) \leq w_0$. On the other hand, we have $z_k \to \overline{x} + \alpha y$, so using the closedness of g_j, we must have

$$g_j(\overline{x} + \alpha y) \leq \liminf_{k\to\infty} g_1(z_k) \leq w_0,$$

which contradicts the choice of α to satisfy $g_j(\overline{x} + \alpha y) > w_0$.

If $y \in \cap_{j=0}^{r} F_{g_j}$, since all the g_j are bidirectionally flat, we have $y \in \cap_{j=0}^{r} L_{g_j}$. If the vectors $\overline{x}$ and x_k, $k \geq 1$, all lie in the same line [which must be the line $\{\overline{x} + \alpha y \mid \alpha \in \Re\}$], we would have $g_j(\overline{x}) = g_j(x_k)$ for all k and j. Then it follows that $\overline{x}$ and x_k all belong to $\cap_{k=1}^{\infty} C_k$. Otherwise, there must be some x_k, with k large enough, and such that, by the Projection Theorem, the vector $x_k - \alpha y$ makes an angle greater than $\pi/2$ with $x_k - \overline{x}$. Since the g_j are constant on the line $\{x_k - \alpha y \mid \alpha \in \Re\}$, all vectors on the line belong to C_k, which contradicts the fact that x_k is the projection of $\overline{x}$ on C_k.

Finally, if $y \in R_{g_0}$ but $y \notin F_{g_0}$, we have $g_0(x + \alpha y) \to -\infty$ as $\alpha \to \infty$, so that $\cap_{k=1}^{\infty} C_k \neq \varnothing$. This completes the proof that

$$\cap_{k=1}^{\infty} C_k = \varnothing \quad \Rightarrow \quad \text{there exists } \overline{j} \in \{1, \ldots, r\} \text{ and } y \in \cap_{j=0}^{r} R_{g_j} \text{ with } y \notin F_{g_{\overline{j}}}. \tag{1}$$

We now use induction on r. For $r = 0$, the preceding proof shows that $\cap_{k=1}^{\infty} C_k \neq \varnothing$. Assume that $\cap_{k=1}^{\infty} C_k \neq \varnothing$ for all cases where $r < \overline{r}$. We will show that $\cap_{k=1}^{\infty} C_k \neq \varnothing$ for $r = \overline{r}$. Assume the contrary. Then, by Eq. (1), there exists $\overline{j} \in \{1, \ldots, r\}$ and $y \in \cap_{j=0}^{r} R_{g_j}$ with $y \notin F_{g_{\overline{j}}}$. Let us consider the sets

$$\overline{C}_k = \big\{x \mid g_0(x) \leq w_k,\, g_j(x) \leq 0,\, j = 1, \ldots, r,\, j \neq \overline{j}\big\}.$$

Since these sets are nonempty, by the induction hypothesis, $\cap_{k=1}^{\infty} \overline{C}_k \neq \varnothing$. For any $\tilde{x} \in \cap_{k=1}^{\infty} \overline{C}_k$, the vector $\tilde{x} + \alpha y$ belongs to $\cap_{k=1}^{\infty} \overline{C}_k$ for all $\alpha > 0$, since $y \in \cap_{j=0}^{r} R_{g_j}$. Since $g_0(\tilde{x}) \leq 0$, we have $\tilde{x} \in \mathrm{dom}(g_{\overline{j}})$, by the hypothesis regarding the domains of the g_j. Since $y \in \cap_{j=0}^{r} R_{g_j}$ with $y \notin F_{g_{\overline{j}}}$, it follows that $g_{\overline{j}}(\tilde{x} +$

$\alpha y) \to -\infty$ as $\alpha \to \infty$. Hence, for sufficiently large α, we have $g_{\overline{j}}(\tilde{x} + \alpha y) \leq 0$, so $\tilde{x} + \alpha y$ belongs to $\cap_{k=1}^{\infty} C_k$.

Note: To see that the assumption

$$\big\{x \mid g_0(x) \leq 0\big\} \subset \cap_{j=1}^{r} \mathrm{dom}(g_j)$$

is essential for the result to hold, consider an example in $\Re^2$. Let

$$g_0(x_1, x_2) = x_1, \qquad g_1(x_1, x_2) = \phi(x_1) - x_2,$$

where the function $\phi : \Re \mapsto (-\infty, \infty]$ is convex, closed, and coercive with $\mathrm{dom}(\phi) = (0, 1)$ [for example, $\phi(t) = -\ln t - \ln(1 - t)$ for $0 < t < 1$]. Then it can be verified that $C_k \neq \emptyset$ for every k and sequence $\{w_k\} \subset (0, 1)$ with $w_k \downarrow 0$ [take $x_1 \downarrow 0$ and $x_2 \geq \phi(x_1)$]. On the other hand, we have $\cap_{k=0}^{\infty} C_k = \emptyset$. The difficulty here is that the set $\big\{x \mid g_0(x) \leq 0\big\}$, which is equal to

$$\{x \mid x_1 \leq 0,\, x_2 \in \Re\},$$

is not contained in $\mathrm{dom}(g_1)$, which is equal to

$$\{x \mid 0 < x_1 < 1,\, x_2 \in \Re\}$$

(in fact the two sets are disjoint).

(b) We will use part (a) and the line of proof of Prop. 1.5.8(c). In particular, let $\{y_k\}$ be a sequence in $A\,C$ converging to some $\overline{y} \in \Re^n$. We will show that $\overline{y} \in A\,C$. We let

$$g_0(x) = \|Ax - \overline{y}\|^2, \qquad w_k = \|y_k - \overline{y}\|^2,$$

and

$$C_k = \big\{x \mid g_0(x) \leq w_k,\, g_j(x) \leq 0,\, j = 1, \ldots, r\big\}.$$

The functions involved in the definition of C_k are bidirectionally flat, and each C_k is nonempty by construction. By applying part (a), we see that the intersection $\cap_{k=0}^{\infty} C_k$ is nonempty. For any x in this intersection, we have $Ax = \overline{y}$ (since $y_k \to \overline{y}$), showing that $\overline{y} \in A\,C$.

(c) We will use part (a) and the line of proof of Prop. 2.3.3 [condition (3)]. Denote

$$f^* = \inf_{x \in C} f(x),$$

and assume without loss of generality that $f^* = 0$ [otherwise, we replace $f(x)$ by $f(x) - f^*$]. We choose a scalar sequence $\{w_k\}$ such that $w_k \downarrow f^*$, and we consider the (nonempty) sets

$$C_k = \big\{x \in \Re^n \mid f(x) \leq w_k,\, g_j(x) \leq 0,\, j = 1, \ldots, r\big\}.$$

By part (a), it follows that $\cap_{k=0}^{\infty} C_k$, the set of minimizers of f over C, is nonempty.

(d) Use the line of proof of Prop. 2.3.9.

2.8 (Minimization of Quasiconvex Functions)

(a) Let x^* be a local minimum of f over X and assume, to arrive at a contradiction, that there exists a vector $\overline{x} \in X$ such that $f(\overline{x}) < f(x^*)$. Then, $\overline{x}$ and x^* belong to the set $X \cap V_{\gamma^*}$, where $\gamma^* = f(x^*)$. Since this set is convex, the line segment connecting x^* and $\overline{x}$ belongs to the set, implying that

$$f\big(\alpha\overline{x} + (1-\alpha)x^*\big) \leq f(x^*), \qquad \forall\, \alpha \in [0,1]. \tag{1}$$

For each integer $k \geq 1$, there exists an $\alpha_k \in (0, 1/k]$ such that

$$f\big(\alpha_k\overline{x} + (1-\alpha_k)x^*\big) < f(x^*), \qquad \text{for some } \alpha_k \in (0,1/k]; \tag{2}$$

otherwise, in view of Eq. (1), we would have that $f(x)$ is constant for x on the line segment connecting x^* and $(1/k)\overline{x} + \big(1-(1/k)\big)x^*$. Equation (2) contradicts the local optimality of x^*.

(b) We consider the level sets

$$V_\gamma = \big\{x \mid f(x) \leq \gamma\big\}$$

for $\gamma > f^*$. Let $\{\gamma^k\}$ be a scalar sequence such that $\gamma^k \downarrow f^*$. Using the fact that for two nonempty closed convex sets C and D such that $C \subset D$, we have $R_C \subset R_D$, it can be seen that

$$R_f = \cap_{\gamma\in\Gamma} R_\gamma = \cap_{k=1}^\infty R_{\gamma^k}.$$

Similarly, L_f can be written as

$$L_f = \cap_{\gamma\in\Gamma} L_\gamma = \cap_{k=1}^\infty L_{\gamma^k}.$$

Under each of the conditions (1)-(4), we show that the set of minima of f over X, which is given by

$$X^* = \cap_{k=1}^\infty (X \cap V_{\gamma^k})$$

is nonempty.

Let condition (1) hold. The sets $X \cap V_{\gamma^k}$ are nonempty, closed, convex, and nested. Furthermore, for each k, their recession cone is given by $R_X \cap R_{\gamma^k}$ and their lineality space is given by $L_X \cap L_{\gamma^k}$. We have that

$$\cap_{k=1}^\infty (R_X \cap R_{\gamma^k}) = R_X \cap R_f,$$

and

$$\cap_{k=1}^\infty (L_X \cap L_{\gamma^k}) = L_X \cap L_f,$$

while by assumption $R_X \cap R_f = L_X \cap L_f$. Then it follows by Prop. 1.5.5 that X^* is nonempty.

Let condition (2) hold. The sets V_{γ^k} are nested and the intersection $X \cap V_{\gamma^k}$ is nonempty for all k. We also have by assumption that $R_X \cap R_f \subset L_f$ and X is specified by linear inequalities. By Prop. 1.5.6, it follows that X^* is nonempty.

Let condition (3) hold. The sets V_{γ^k} have the form

$$V_{\gamma^k} = \{x \in \Re^n \mid x'Qx + c'x + b(\gamma^k) \leq 0\}.$$

In view of the assumption that $b(\gamma)$ is bounded for $\gamma \in (f^*, \overline{\gamma}]$, we can consider a subsequence $\{b(\gamma^k)\}_{\mathcal{K}}$ that converges to a scalar. Furthermore, X is specified by convex quadratic inequalities, and the intersection $X \cap V_{\gamma^k}$ is nonempty for all $k \in \mathcal{K}$. By Prop. 1.5.7, it follows that X^* is nonempty.

Similarly, under condition (4), the result follows using Exercise 2.7(a).

2.9 (Partial Minimization)

(a) The epigraph of f is given by

$$\mathrm{epi}(f) = \big\{(x, w) \mid f(x) \le w\big\}.$$

If $(x, w) \in E_f$, then it follows that $(x, w) \in \mathrm{epi}(f)$, showing that $E_f \subset \mathrm{epi}(f)$. Next, assume that $(x, w) \in \mathrm{epi}(f)$, i.e., $f(x) \le w$. Let $\{w_k\}$ be a sequence with $w_k > w$ for all k , and $w_k \to w$. Then we have, $f(x) < w_k$ for all k, implying that $(x, w_k) \in E_f$ for all k, and that the limit $(x, w) \in \mathrm{cl}(E_f)$. Thus we have the desired relations,

$$E_f \subset \mathrm{epi}(f) \subset \mathrm{cl}(E_f). \tag{2.5}$$

We next show that f is convex if and only if E_f is convex. By definition, f is convex if and only if $\mathrm{epi}(f)$ is convex. Assume that $\mathrm{epi}(f)$ is convex. Suppose, to arrive at a contradiction, that E_f is not convex. This implies the existence of vectors $(x_1, w_1) \in E_f$, $(x_2, w_2) \in E_f$, and a scalar $\alpha \in (0, 1)$ such that $\alpha(x_1, w_1)+ (1 - \alpha)(x_2, w_2) \notin E_f$, from which we get

$$\begin{aligned} f\big(\alpha x_1 + (1-\alpha)x_2\big) &\ge \alpha w_1 + (1-\alpha)w_2 \\ &> \alpha f(x_1) + (1-\alpha)f(x_2), \end{aligned} \tag{2.6}$$

where the second inequality follows from the fact that (x_1, w_1) and (x_2, w_2) belong to E_f. We have $\big(x_1, f(x_1)\big) \in \mathrm{epi}(f)$ and $\big(x_2, f(x_2)\big) \in \mathrm{epi}(f)$. In view of the convexity assumption of $\mathrm{epi}(f)$, this yields $\alpha\big(x_1, f(x_1)\big) + (1-\alpha)\big(x_2, f(x_2)\big) \in \mathrm{epi}(f)$ and therefore,

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha)f(x_2).$$

Combined with Eq. (2.6), the preceding relation yields a contradiction, thus showing that E_f is convex.

Next assume that E_f is convex. We show that $\mathrm{epi}(f)$ is convex. Let (x_1, w_1) and (x_2, w_2) be arbitrary vectors in $\mathrm{epi}(f)$. Consider sequences of vectors (x_1, w_1^k) and (x_2, w_2^k) such that $w_1^k > w_1$, $w_2^k > w_2$, and $w_1^k \to w_1$, $w_2^k \to w_2$. It follows that for each k, (x_1, w_1^k) and (x_2, w_2^k) belong to E_f. Since E_f is convex by assumption, this implies that for each $\alpha \in [0, 1]$ and all k, the vector $\big(\alpha x_1 + (1-\alpha)x_2, \alpha w_1^k + (1-\alpha)w_2^k\big) \in E_f$, i.e., we have for each k

$$f\big(\alpha x_1 + (1-\alpha)x_2\big) < \alpha w_1^k + (1-\alpha)w_2^k.$$

Taking the limit in the preceding relation, we get

$$f\big(\alpha x_1 + (1-\alpha)x_2\big) \le \alpha w_1 + (1-\alpha)w_2,$$

showing that $\big(\alpha x_1 + (1-\alpha)x_2, \alpha w_1 + (1-\alpha)w_2\big) \in \mathrm{epi}(f)$. Hence $\mathrm{epi}(f)$ is convex.

(b) Let T denote the projection of the set $\big\{(x, z, w) \mid F(x, z) < w\big\}$ on the space of (x, w). We show that $E_f = T$. Let $(x, w) \in E_f$. By definition, we have

$$\inf_{z \in \Re^m} F(x, z) < w,$$

which implies that there exists some $\overline{z} \in \Re^m$ such that

$$F(x, \overline{z}) < w,$$

showing that $(x, \overline{z}, w)$ belongs to the set $\{(x, z, w) \mid F(x, z) < w\}$, and $(x, w) \in T$. Conversely, let $(x, w) \in T$. This implies that there exists some z such that $F(x, z) < w$, from which we get

$$f(x) = \inf_{z \in \Re^m} F(x, z) < w,$$

showing that $(x, w) \in E_f$, and completing the proof.

(c) Let F be a convex function. Using part (a), the convexity of F implies that the set $\{(x, z, w) \mid F(x, z) < w\}$ is convex. Since the projection mapping is linear, and hence preserves convexity, we have, using part (b), that the set E_f is convex, which implies by part (a) that f is convex.

2.10 (Partial Minimization of Nonconvex Functions)

(a) For each $u \in \Re^m$, let $f_u(x) = f(x, u)$. There are two cases; either $f_u \equiv \infty$, or f_u is lower semicontinuous with bounded level sets. The first case, which corresponds to $p(u) = \infty$, can't hold for every u, since f is not identically equal to ∞. Therefore, $\mathrm{dom}(p) \neq \emptyset$, and for each $u \in \mathrm{dom}(p)$, we have by Weierstrass' Theorem that $p(u) = \inf_x f_u(x)$ is finite [i.e., $p(u) > -\infty$ for all $u \in \mathrm{dom}(p)$] and the set $P(u) = \arg\min_x f_u(x)$ is nonempty and compact.

We now show that p is lower semicontinuous. By assumption, for all $\overline{u} \in \Re^m$ and for all $\alpha \in \Re$, there exists a neighborhood $\overline{N}$ of $\overline{u}$ such that the set $\{(x, u) \mid f(x, u) \leq \alpha\} \cap (\Re^n \times \overline{N})$ is bounded in $\Re^n \times \Re^m$. We can choose a smaller closed set N containing $\overline{u}$ such that the set $\{(x, u) \mid f(x, u) \leq \alpha\} \cap (\Re^n \times N)$ is closed (since f is lower semicontinuous) and bounded. In view of the assumption that f_u is lower semicontinuous with bounded level sets, it follows using Weierstrass' Theorem that for any scalar α,

$$p(u) \leq \alpha \text{ if and only if there exists } x \text{ such that } f(x, u) \leq \alpha.$$

Hence, the set $\{u \mid p(u) \leq \alpha\} \cap N$ is the image of the set $\{(x, u) \mid f(x, u) \leq \alpha\} \cap (\Re^n \times N)$ under the continuous mapping $(x, u) \mapsto u$. Since the image of a compact set under a continuous mapping is compact [cf. Prop. 1.1.9(d)], we see that $\{u \mid p(u) \leq \alpha\} \cap N$ is closed.

Thus, each $\overline{u} \in \Re^m$ is contained in a closed set whose intersection with $\{u \mid p(u) \leq \alpha\}$ is closed, so that the set $\{u \mid p(u) \leq \alpha\}$ itself is closed for all scalars α. It follows from Prop. 1.2.2 that p is lower semicontinuous.

(b) Consider the following example

$$f(x, u) = \begin{cases} \min\{|x - 1/u|, 1 + |x|\} & \text{if } u \neq 0,\ x \in \Re, \\ 1 + |x| & \text{if } u = 0,\ x \in \Re, \end{cases}$$

where x and u are scalars. This function is continuous in (x, u) and the level sets are bounded in x for each u, but not *locally uniformly* in u, i.e., there does not exists a neighborhood N of $u = 0$ such that the set $\big\{(x, u) \mid u \in N,\ f(x, u) \leq \alpha\big\}$ is bounded for some $\alpha > 0$.

For this function, we have

$$p(u) = \begin{cases} 0 & \text{if } u \neq 0, \\ 1 & \text{if } u = 0. \end{cases}$$

Hence, the function p is not lower semicontinuous at 0.

(c) Let $\{u_k\}$ be a sequence such that $u_k \to u^*$ for some $u^* \in \text{dom}(p)$, and also $p(u_k) \to p(u^*)$. Let α be any scalar such that $p(u^*) < \alpha$. Since $p(u_k) \to p(u^*)$, we obtain

$$f(x_k, u_k) = p(u_k) < \alpha, \tag{2.7}$$

for all k sufficiently large, where we use the fact that $x_k \in P(u_k)$ for all k. We take N to be a closed neighborhood of u^* as in part (a). Since $u_k \to u^*$, using Eq. (2.7), we see that for all k sufficiently large, the pair (x_k, u_k) lies in the compact set

$$\big\{(x, u) \mid f(x, u) \leq \alpha\big\} \cap (\Re^n \times N).$$

Hence, the sequence $\{x_k\}$ is bounded, and therefore has a limit point, call it x^*. It follows that

$$(x^*, u^*) \in \big\{(x, u) \mid f(x, u) \leq \alpha\big\}.$$

Since this is true for arbitrary $\alpha > p(u^*)$, we see that $f(x^*, u^*) \leq p(u^*)$, which, by the definition of $p(u)$, implies that $x^* \in P(u^*)$.

(d) By definition, we have $p(u) \leq f(x^*, u)$ for all u and $p(u^*) = f(x^*, u^*)$. Since $f(x^*, \cdot)$ is continuous at u^*, we have for any sequence $\{u_k\}$ converging to u^*

$$\limsup_{k \to \infty} p(u_k) \leq \limsup_{k \to \infty} f(x^*, u_k) = f(x^*, u^*) = p(u^*),$$

thereby implying that p is upper semicontinuous at u^*. Since p is also lower semicontinuous at u^* by part (a), we conclude that p is continuous at u^*.

2.11 (Projection on a Nonconvex Set)

We define the function f by

$$f(w, x) = \begin{cases} \|w - x\| & \text{if } w \in C, \\ \infty & \text{if } w \notin C. \end{cases}$$

With this identification, we get

$$d_C(x) = \inf_w f(w, x), \qquad P_C(x) = \arg\min_w f(w, x).$$

We now show that $f(w, x)$ satisfies the assumptions of Exercise 2.10, so that we can apply the results of this exercise to this problem.

Since the set C is closed by assumption, it follows that $f(w, x)$ is lower semicontinuous. Moreover, by Weierstrass' Theorem, we see that $f(w, x) > -\infty$ for all x and w. Since the set C is nonempty by assumption, we also have that dom(f) is nonempty. It is also straightforward to see that the function $\|\cdot\|$, and therefore the function f, satisfies the locally uniformly level-boundedness assumption of Exercise 2.10.

(a) Since the function $\|\cdot\|$ is lower semicontinuous and the set C is closed, it follows from Weierstrass' Theorem that for all $x^* \in \Re^n$, the infimum in $\inf_w f(w, x)$ is attained at some w^*, i.e., $P(x^*)$ is nonempty. Hence, we see that for all $x^* \in \Re^n$, there exists some $w^* \in P(x^*)$ such that $f(w^*, \cdot)$ is continuous at x^*, which follows by continuity of the function $\|\cdot\|$. Hence, the function $f(w, x)$ satisfies the sufficiency condition given in Exercise 2.10(d), and it follows that $d_C(x)$ depends continuously on x.

(b) This part follows from part (a) of Exercise 2.10.

(c) This part follows from part (c) of Exercise 2.10.

2.12 (Convergence of Penalty Methods [RoW98])

(a) We set $\overline{s} = 1/\overline{c}$ and consider the function $g(x, s) : \Re^n \times \Re \mapsto (-\infty, \infty]$ defined by

$$g(x, s) = f(x) + \tilde{\theta}\big(F(x), s\big),$$

with the function $\tilde{\theta}$ given by

$$\tilde{\theta}(u, s) = \begin{cases} \theta(u, 1/s) & \text{if } s \in (0, \overline{s}], \\ \delta_D(u) & \text{if } s = 0, \\ \infty & \text{if } s < 0 \text{ or } s > \overline{s}, \end{cases}$$

where

$$\delta_D(u) = \begin{cases} 0 & \text{if } u \in D, \\ \infty & \text{if } u \notin D. \end{cases}$$

We identify the original problem with that of minimizing $g(x, 0)$ in $x \in \Re^n$, and the approximate problem for parameter $s \in (0, \overline{s}]$ with that of minimizing $g(x, s)$ in $x \in \Re^n$ where $s = 1/c$. With the notation introduced in Exercise 2.10, the optimal value of the original problem is given by $p(0)$ and the optimal value of the approximate problem is given by $p(s)$. Hence, we have

$$p(s) = \inf_{x \in \Re^n} g(x, s).$$

We now show that, for the function $g(x, s)$, the assumptions of Exercise 2.10 are satisfied.

We have that $g(x, s) > -\infty$ for all (x, s), since by assumption $f(x) > -\infty$ for all x and $\theta(u, s) > -\infty$ for all (u, s). The function $\tilde{\theta}$ is such that $\tilde{\theta}(u, s) < \infty$ at least for one vector (u, s), since the set D is nonempty. Therefore, it follows that $g(x, s) < \infty$ for at least one vector (x, s), unless $g \equiv \infty$, in which case all the results of this exercise follow trivially.

We now show that the function $\tilde{\theta}$ is lower semicontinuous. This is easily seen at all points where $s \neq 0$ in view of the assumption that the function θ is lower semicontinuous on $\Re^m \times (0, \infty)$. We next consider points where $s = 0$. We claim that for any $\alpha \in \Re$,

$$\big\{u \mid \tilde{\theta}(u, 0) \leq \alpha\big\} = \bigcap_{s \in (0, \overline{s}]} \big\{u \mid \tilde{\theta}(u, s) \leq \alpha\big\}. \tag{2.8}$$

To see this, assume that $\tilde{\theta}(u, 0) \leq \alpha$. Since $\tilde{\theta}(u, s) \uparrow \tilde{\theta}(u, 0)$ as $s \downarrow 0$, we have $\tilde{\theta}(u, s) \leq \alpha$ for all $s \in (0, \overline{s}]$. Conversely, assume that $\tilde{\theta}(u, s) \leq \alpha$ for all $s \in (0, \overline{s}]$. By definition of $\tilde{\theta}$, this implies that

$$\theta(u, 1/s) \leq \alpha, \qquad \forall\ s \in (0, s_0].$$

Taking the limit as $s \to 0$ in the preceding relation, we get

$$\lim_{s \to 0} \theta(u, 1/s) = \delta_D(u) = \tilde{\theta}(u, 0) \leq \alpha,$$

thus, proving the relation in (2.8). Note that for all $\alpha \in \Re$ and all $s \in (0, \overline{s}]$, the set

$$\big\{u \mid \tilde{\theta}(u, s) \leq \alpha\big\} = \big\{u \mid \theta(u, 1/s) \leq \alpha\big\},$$

is closed by the lower semicontinuity of the function θ. Hence, the relation in Eq. (2.8) implies that the set $\big\{u \mid \tilde{\theta}(u, 0) \leq \alpha\big\}$ is closed for all $\alpha \in \Re$, thus showing that the function $\tilde{\theta}$ is lower semicontinuous everywhere (cf. Prop. 1.2.2). Together with the assumptions that f is lower semicontinuous and F is continuous, it follows that g is lower semicontinuous.

Finally, we show that g satisfies the locally uniform level boundedness property given in Exercise 2.10, i.e., for all $s^* \in \Re$ and for all $\alpha \in \Re$, there exists a neighborhood N of s^* such that the set $\big\{(x, s) \mid s \in N, g(x, s) \leq \alpha\big\}$ is bounded. By assumption, we have that the level sets of the function $g(x, \overline{s}) = f(x) + \tilde{\theta}\big(F(x), 1/\overline{s}\big)$ are bounded. The definition of $\tilde{\theta}$, together with the fact that $\tilde{\theta}(u, s)$ is monotonically increasing as $s \downarrow 0$, implies that g is indeed level-bounded in x locally uniformly in s.

Therefore, all the assumptions of Exercise 2.10 are satisfied and we get that the function p is lower semicontinuous in s. Since $\tilde{\theta}(u, s)$ is monotonically increasing as $s \downarrow 0$, it follows that p is monotonically nondecreasing as $s \downarrow 0$. This implies that

$$p(s) \to p(0), \qquad \text{as } s \downarrow 0.$$

Defining $s_k = 1/c_k$ for all k, where $\{c_k\}$ is the given sequence of parameter values, we get

$$p(s_k) \to p(0),$$

thus proving that the optimal value of the approximate problem converges to the optimal value of the original problem.

(b) We have by assumption that $s_k \to 0$ with $x_k \in P_{1/s_k}$. It follows from part (a) that $p(s_k) \to p(0)$, so Exercise 2.10(b) implies that the sequence $\{x_k\}$ is bounded and all its limit points are optimal solutions of the original problem.

2.13 (Approximation by Envelope Functions [RoW98])

(a) We fix a $c_0 \in (0, c_f)$ and consider the function

$$h(w,x,c) = \begin{cases} f(w) + (\frac{1}{2c})\|w-x\|^2 & \text{if } c \in (0, c_0], \\ f(x) & \text{if } c = 0 \text{ and } w = x, \\ \infty & \text{otherwise.} \end{cases}$$

We consider the problem of minimizing $h(w,x,c)$ in w. With this identification and using the notation introduced in Exercise 2.10, for some $c \in (0, c_0)$, we obtain

$$e_c f(x) = p(x,c) = \inf_w h(w,x,c),$$

and

$$P_c f(x) = P(x,c) = \arg\min_w h(w,x,c).$$

We now show that, for the function $h(w,x,c)$, the assumptions given in Exercise 2.10 are satisfied.

We have that $h(w,x,c) > -\infty$ for all (w,x,c), since by assumption $f(x) > -\infty$ for all $x \in \Re^n$. Furthermore, $h(w,x,c) < \infty$ for at least one vector (w,x,c), since by assumption $f(x) < \infty$ for at least one vector $x \in X$.

We next show that the function h is lower semicontinuous in (w,x,c). This is easily seen at all points where $c \in (0, c_0]$ in view of the assumption that f is lower semicontinuous and the function $\|\cdot\|^2$ is lower semicontinuous. We now consider points where $c = 0$ and $w \neq x$. Let $\big\{(w_k, x_k, c_k)\big\}$ be a sequence that converges to some $(w, x, 0)$ with $w \neq x$. We can assume without loss of generality that $w_k \neq x_k$ for all k. Note that for some k, we have

$$h(w_k, x_k, c_k) = \begin{cases} \infty & \text{if } c_k = 0, \\ f(w_k) + (\frac{1}{2c_k})\|w_k - x_k\|^2 & \text{if } c_k > 0. \end{cases}$$

Taking the limit as $k \to \infty$, we have

$$\lim_{k\to\infty} h(w_k, x_k, c_k) = \infty \geq h(w, x, 0),$$

since $w \neq x$ by assumption. This shows that h is lower semicontinuous at points where $c = 0$ and $w \neq x$. We finally consider points where $c = 0$ and $w = x$. At these points, we have $h(w,x,c) = f(x)$. Let $\big\{(w_k, x_k, c_k)\big\}$ be a sequence that converges to some $(w, x, 0)$ with $w = x$. Considering all possibilities, we see that the limit inferior of the sequence $\big\{h(w_k, x_k, c_k)\big\}$ cannot be less than $f(x)$, thus showing that h is also lower semicontinuous at points where $c = 0$ and $w = x$.

Finally, we show that h satisfies the locally uniform level-boundedness property given in Exercise 2.10, i.e., for all (x^*, c^*) and for all $\alpha \in \Re$, there exists a neighborhood N of (x^*, c^*) such that the set $\big\{(w,x,c) \mid (x,c) \in N, h(w,x,c) \leq \alpha\big\}$ is bounded. Assume, to arrive at a contradiction, that there exists a sequence $\big\{(w_k, x_k, c_k)\big\}$ such that

$$h(w_k, x_k, c_k) \leq \alpha < \infty, \tag{2.9}$$

for some scalar α, with $(x_k, c_k) \to (x^*, c^*)$, and $\|w_k\| \to \infty$. Then, for sufficiently large k, we have $w_k \neq x_k$, which in view of Eq. (2.9) and the definition of the function h, implies that $c_k \in (0, c_0]$ and

$$f(w_k) + \frac{1}{2c_k}\|w_k - x_k\|^2 \leq \alpha,$$

for all sufficiently large k. In particular, since $c_k \leq c_0$, it follows from the preceding relation that

$$f(w_k) + \frac{1}{2c_0}\|w_k - x_k\|^2 \leq \alpha. \tag{2.10}$$

The choice of c_0 ensures, through the definition of c_f, the existence of some $c_1 > c_0$, some $\overline{x} \in \Re^n$, and some scalar β such that

$$f(w) \geq -\frac{1}{2c_1}\|w - \overline{x}\|^2 + \beta, \qquad \forall\, w.$$

Together with Eq. (2.10), this implies that

$$-\frac{1}{2c_1}\|w_k - \overline{x}\|^2 + \frac{1}{2c_0}\|w_k - x_k\|^2 \leq \alpha - \beta,$$

for all sufficiently large k. Dividing this relation by $\|w_k\|^2$ and taking the limit as $k \to \infty$, we get

$$-\frac{1}{2c_1} + \frac{1}{2c_0} \leq 0,$$

from which it follows that $c_1 \leq c_0$. This is a contradiction by our choice of c_1. Hence, the function $h(w, x, c)$ satisfies all the assumptions of Exercise 2.10.

By assumption, we have that $f(\overline{x}) < \infty$ for some $\overline{x} \in \Re^n$. Using the definition of $e_c f(x)$, this implies that

$$\begin{aligned} e_c f(x) &= \inf_w \left\{ f(w) + \frac{1}{2c}\|w - x\|^2 \right\} \\ &\leq f(\overline{x}) + \frac{1}{2c}\|\overline{x} - x\|^2 < \infty, \qquad \forall\, x \in \Re^n, \end{aligned}$$

where the first inequality is obtained by setting $w = \overline{x}$ in $f(w) + \frac{1}{2c}\|w - x\|^2$. Together with Exercise 2.10(a), this shows that for every $c \in (0, c_0)$ and all $x \in \Re^n$, the function $e_c f(x)$ is finite, and the set $P_c f(x)$ is nonempty and compact. Furthermore, it can be seen from the definition of $h(w, x, c)$, that for all $c \in (0, c_0)$, $h(w, x, c)$ is continuous in (x, c). Therefore, it follows from Exercise 2.10(d) that for all $c \in (0, c_0)$, $e_c f(x)$ is continuous in (x, c). In particular, since $e_c f(x)$ is a monotonically decreasing function of c, it follows that

$$e_c f(x) = p(x, c) \uparrow p(x, 0) = f(x), \quad \forall\, x \text{ as } c \downarrow 0.$$

This concludes the proof for part (a).
(b) Directly follows from Exercise 2.10(c).

2.14 (Envelopes and Proximal Mappings under Convexity [RoW98])

We consider the function g_c defined by

$$g_c(x, w) = f(w) + \frac{1}{2c}\|w - x\|^2.$$

In view of the assumption that f is lower semicontinuous, it follows that $g_c(x, w)$ is lower semicontinuous. We also have that $g_c(x, w) > -\infty$ for all (x, w) and $g_c(x, w) < \infty$ for at least one vector (x, w). Moreover, since $f(x)$ is convex by assumption, $g_c(x, w)$ is convex in (x, w), even strictly convex in w.

Note that by definition, we have

$$e_c f(x) = \inf_w g_c(x, w),$$
$$P_c f(x) = \arg\min_w g_c(x, w).$$

(a) In order to show that c_f is ∞, it suffices to show that $e_c f(0) > -\infty$ for all $c > 0$. This will follow from Weierstrass' Theorem, once we show the boundedness of the level sets of $g_c(0, \cdot)$. Assume the contrary, i.e., there exists some $\alpha \in \Re$ and a sequence $\{x_k\}$ such that $\|x_k\| \to \infty$ and

$$g_c(0, x_k) = f(x_k) + \frac{1}{2c}\|x_k\|^2 \leq \alpha, \qquad \forall\ k. \tag{2.11}$$

Assume without loss of generality that $\|x_k\| > 1$ for all k. We fix an x_0 with $f(x_0) < \infty$. We define

$$\tau_k = \frac{1}{\|x_k\|} \in (0, 1),$$

and

$$\overline{x}_k = (1 - \tau_k)x_0 + \tau_k x_k.$$

Since $\|x_k\| \to \infty$, it follows that $\tau_k \to 0$. Using Eq. (2.11) and the convexity of f, we obtain

$$\begin{aligned} f(\overline{x}_k) &\leq (1 - \tau_k)f(x_0) + \tau_k f(x_k) \\ &\leq (1 - \tau_k)f(x_0) + \tau_k\alpha - \frac{\tau_k}{2c}\|x_k\|^2. \end{aligned}$$

Taking the limit as $k \to \infty$ in the above equation, we see that $f(\overline{x}_k) \to -\infty$. It follows from the definitions of τ_k and $\overline{x}_k$ that

$$\begin{aligned} \|\overline{x}_k\| &\leq \|1 - \tau_k\|\|x_0\| + \|\tau_k\|\|x_k\| \\ &\leq \|x_0\| + 1. \end{aligned}$$

Therefore, the sequence $\{\overline{x}_k\}$ is bounded. Since f is lower semicontinuous, Weierstrass' Theorem suggests that f is bounded from below on every bounded subset of $\Re^n$. Since the sequence $\{\overline{x}_k\}$ is bounded, this implies that the sequence $f(\overline{x}_k)$ is bounded from below, which contradicts the fact that $f(\overline{x}_k) \to \infty$. This proves that the level sets of the function $g_c(0,\cdot)$ are bounded. Therefore, using Weierstrass' Theorem, we have that the infimum in $e_cf(0) = \inf_w g_c(0,w)$ is attained, and $e_cf(0) > -\infty$ for every $c > 0$. This shows that the supremum c_f of all $c > 0$, such that $e_cf(x) > -\infty$ for some $x \in \Re^n$, is ∞.

(b) Since the value c_f is equal to ∞ by part (a), it follows that e_cf and P_cf have all the properties given in Exercise 2.13 for all $c > 0$: The set $P_cf(x)$ is nonempty and compact, and the function $e_cf(x)$ is finite for all x, and is continuous in (x,c). Consider a sequence $\{w_k\}$ with $w_k \in P_{c_k}f(x_k)$ for some sequences $x_k \to x^*$ and $c_k \to c^* > 0$. Then, it follows from Exercise 2.13(b) that the sequence $\{w_k\}$ is bounded and all its limit points belong to the set $P_{c^*}f(x^*)$. Since $g_c(x,w)$ is strictly convex in w, it follows from Prop. 2.1.2 that the proximal mapping P_cf is single-valued. Hence, we have that $P_cf(x) \to P_{c^*}f(x^*)$ whenever $(x,c) \to (x^*,c^*)$ with $c^* > 0$.

(c) The envelope function e_cf is convex by Exercise 2.15 [since $g_c(x,w)$ is convex in (x,w)], and continuous by Exercise 2.13. We now prove that it is differentiable. Consider any point $\overline{x}$, and let $\overline{w} = P_cf(\overline{x})$. We will show that e_cf is differentiable at $\overline{x}$ with

$$\nabla e_cf(\overline{x}) = \frac{(\overline{x} - \overline{w})}{c}.$$

Equivalently, we will show that the function h given by

$$h(u) = e_cf(\overline{x} + u) - e_cf(\overline{x}) - \frac{(\overline{x} - \overline{w})'}{c}u \tag{2.12}$$

is differentiable at 0 with $\nabla h(0) = 0$. Since $\overline{w} = P_cf(\overline{x})$, we have

$$e_cf(\overline{x}) = f(\overline{w}) + \frac{1}{2c}\|\overline{w} - \overline{x}\|^2,$$

whereas

$$e_cf(\overline{x} + u) \le f(\overline{w}) + \frac{1}{2c}\|\overline{w} - (\overline{x} + u)\|^2, \qquad \forall\, u,$$

so that

$$h(u) \le \frac{1}{2c}\|\overline{w} - (\overline{x} + u)\|^2 - \frac{1}{2c}\|\overline{w} - \overline{x}\|^2 - \frac{1}{c}(\overline{x} - \overline{w})'u = \frac{1}{2c}\|u\|^2, \ \forall\, u. \tag{2.13}$$

Since e_cf is convex, it follows from Eq. (2.12) that h is convex, and therefore,

$$0 = h(0) = h\left(\frac{1}{2}u + \frac{1}{2}(-u)\right) \le \frac{1}{2}h(u) + \frac{1}{2}h(-u),$$

which implies that $h(u) \ge -h(-u)$. From Eq. (2.13), we obtain

$$-h(-u) \ge -\frac{1}{2c}\|-u\|^2 = -\frac{1}{2c}\|u\|^2, \qquad \forall\, u,$$

which together with the preceding relation yields

$$h(u) \ge -\frac{1}{2c}\|u\|^2, \qquad \forall\, u.$$

Thus, we have

$$|h(u)| \le \frac{1}{2c}\|u\|^2, \qquad \forall\, u,$$

which implies that h is differentiable at 0 with $\nabla h(0) = 0$. From the formula for $\nabla e_cf(\cdot)$ and the continuity of $P_cf(\cdot)$, it also follows that e_c is continuously differentiable.

2.15

(a) In view of the assumption that $\mathrm{int}(C_1)$ and C_2 are disjoint and convex [cf Prop. 1.2.1(d)], it follows from the Separating Hyperplane Theorem that there exists a vector $a \neq 0$ such that

$$a'x_1 \leq a'x_2, \qquad \forall\ x_1 \in \mathrm{int}(C_1), \quad \forall\ x_2 \in C_2.$$

Let $b = \inf_{x_2 \in C_2} a'x_2$. Then, from the preceding relation, we have

$$a'x \leq b, \qquad \forall\ x \in \mathrm{int}(C_1). \tag{2.14}$$

We claim that the closed halfspace $\{x \mid a'x \geq b\}$, which contains C_2, does not intersect $\mathrm{int}(C_1)$.

Assume to arrive at a contradiction that there exists some $\overline{x}_1 \in \mathrm{int}(C_1)$ such that $a'\overline{x}_1 \geq b$. Since $\overline{x}_1 \in \mathrm{int}(C_1)$, we have that there exists some $\epsilon > 0$ such that $\overline{x}_1 + \epsilon a \in \mathrm{int}(C_1)$, and

$$a'(\overline{x}_1 + \epsilon a) \geq b + \epsilon \|a\|^2 > b.$$

This contradicts Eq. (2.14). Hence, we have

$$\mathrm{int}(C_1) \subset \{x \mid a'x < b\}.$$

(b) Consider the sets

$$C_1 = \Big\{(x_1, x_2) \mid x_1 = 0\Big\},$$
$$C_2 = \Big\{(x_1, x_2) \mid x_1 > 0,\ x_2 x_1 \geq 1\Big\}.$$

These two sets are convex and C_2 is disjoint from $\mathrm{ri}(C_1)$, which is equal to C_1. The only separating hyperplane is the x_2 axis, which corresponds to having $a = (0, 1)$, as defined in part (a). For this example, there does not exist a closed halfspace that contains C_2 but is disjoint from $\mathrm{ri}(C_1)$.

2.16

If there exists a hyperplane H with the properties stated, the condition $M \cap \mathrm{ri}(C) = \emptyset$ clearly holds. Conversely, if $M \cap \mathrm{ri}(C) = \emptyset$, then M and C can be properly separated by Prop. 2.4.5. This hyperplane can be chosen to contain M since M is affine. If this hyperplane contains a point in $\mathrm{ri}(C)$, then it must contain all of C by Prop. 1.4.2. This contradicts the proper separation property, thus showing that $\mathrm{ri}(C)$ is contained in one of the open halfspaces.

2.17 (Strong Separation)

(a) We first show that (i) implies (ii). Suppose that C_1 and C_2 can be separated strongly. By definition, this implies that for some nonzero vector $a \in \Re^n$, $b \in \Re$, and $\epsilon > 0$, we have

$$C_1 + \epsilon B \subset \{x \mid a'x > b\},$$

$$C_2 + \epsilon B \subset \{x \mid a'x < b\},$$

where B denotes the closed unit ball. Since $a \neq 0$, we also have

$$\inf\{a'y \mid y \in B\} < 0, \qquad \sup\{a'y \mid y \in B\} > 0.$$

Therefore, it follows from the preceding relations that

$$b \leq \inf\{a'x + \epsilon a'y \mid x \in C_1, y \in B\} < \inf\{a'x \mid x \in C_1\},$$

$$b \geq \sup\{a'x + \epsilon a'y \mid x \in C_2, y \in B\} > \sup\{a'x \mid x \in C_2\}.$$

Thus, there exists a vector $a \in \Re^n$ such that

$$\inf_{x \in C_1} a'x > \sup_{x \in C_2} a'x,$$

proving (ii).

Next, we show that (ii) implies (iii). Suppose that (ii) holds, i.e., there exists some vector $a \in \Re^n$ such that

$$\inf_{x \in C_1} a'x > \sup_{x \in C_2} a'x, \tag{2.15}$$

Using the Schwartz inequality, we see that

$$\begin{aligned} 0 &< \inf_{x \in C_1} a'x - \sup_{x \in C_2} a'x \\ &= \inf_{x_1 \in C_1,\ x_2 \in C_2} a'(x_1 - x_2), \\ &\leq \inf_{x_1 \in C_1,\ x_2 \in C_2} \|a\| \|x_1 - x_2\|. \end{aligned}$$

It follows that

$$\inf_{x_1 \in C_1, x_2 \in C_2} \|x_1 - x_2\| > 0,$$

thus proving (iii).

Finally, we show that (iii) implies (i). If (iii) holds, we have for some $\epsilon > 0$,

$$\inf_{x_1 \in C_1, x_2 \in C_2} \|x_1 - x_2\| > 2\epsilon > 0.$$

From this we obtain for all $x_1 \in C_1$, all $x_2 \in C_2$, and for all $y_1,\ y_2$ with $\|y_1\| \leq \epsilon$, $\|y_2\| \leq \epsilon$,

$$\|(x_1 + y_1) - (x_2 + y_2)\| \geq \|x_1 - x_2\| - \|y_1\| - \|y_2\| > 0,$$

which implies that $0 \notin (C_1 + \epsilon B) - (C_2 + \epsilon B)$. Therefore, the convex sets $C_1 + \epsilon B$ and $C_2 + \epsilon B$ are disjoint. By the Separating Hyperplane Theorem, we see that $C_1 + \epsilon B$ and $C_2 + \epsilon B$ can be separated, i.e., $C_1 + \epsilon B$ and $C_2 + \epsilon B$ lie in opposite closed halfspaces associated with the hyperplane that separates them. Then, the sets $C_1 + (\epsilon/2)B$ and $C_2 + (\epsilon/2)B$ lie in opposite open halfspaces, which by definition implies that C_1 and C_2 can be separated strongly.

(b) Since C_1 and C_2 are disjoint, we have $0 \notin (C_1 - C_2)$. Any one of conditions (2)-(5) of Prop. 2.4.3 imply condition (1) of that proposition (see the discussion in the proof of Prop. 2.4.3), which states that the set $C_1 - C_2$ is closed, i.e.,

$$\mathrm{cl}(C_1 - C_2) = C_1 - C_2.$$

Hence, we have $0 \notin \mathrm{cl}(C_1 - C_2)$, which implies that

$$\inf_{x_1 \in C_1,\, x_2 \in C_2} \|x_1 - x_2\| > 0.$$

From part (a), it follows that there exists a hyperplane separating C_1 and C_2 strongly.

2.18

(a) If C_1 and C_2 can be separated properly, we have from the Proper Separation Theorem that there exists a vector $a \neq 0$ such that

$$\inf_{x \in C_1} a'x \geq \sup_{x \in C_2} a'x, \tag{2.16}$$

$$\sup_{x \in C_1} a'x > \inf_{x \in C_2} a'x. \tag{2.17}$$

Let

$$b = \sup_{x \in C_2} a'x. \tag{2.18}$$

and consider the hyperplane

$$H = \{x \mid a'x = b\}.$$

Since C_2 is a cone, we have

$$\lambda a'x = a'(\lambda x) \leq b < \infty, \qquad \forall\, x \in C_2,\ \forall\, \lambda > 0.$$

This relation implies that $a'x \leq 0$, for all $x \in C_2$, since otherwise it is possible to choose λ large enough and violate the above inequality for some $x \in C_2$. Hence, it follows from Eq. (2.18) that $b \leq 0$. Also, by letting $\lambda \to 0$ in the preceding relation, we see that $b \geq 0$. Therefore, we have that $b = 0$ and the hyperplane H contains the origin.

(b) If C_1 and C_2 can be separated strictly, we have by definition that there exists a vector $a \neq 0$ and a scalar β such that

$$a'x_2 < \beta < a'x_1, \qquad \forall\, x_1 \in C_1, \quad \forall\, x_2 \in C_2. \tag{2.19}$$

We choose b to be

$$b = \sup_{x \in C_2} a'x, \tag{2.20}$$

and consider the closed halfspace

$$K = \{x \mid a'x \leq b\},$$

which contains C_2. By Eq. (2.19), we have

$$b \leq \beta < a'x, \qquad \forall\ x \in C_1,$$

so the closed halfspace K does not intersect C_1.

Since C_2 is a cone, an argument similar to the one in part (a) shows that $b = 0$, and hence the hyperplane associated with the closed halfspace K passes through the origin, and has the desired properties.

2.19 (Separation Properties of Cones)

(a) C is contained in the intersection of the homogeneous closed halfspaces that contain C, so we focus on proving the reverse inclusion. Let $x \notin C$. Since C is closed and convex by assumption, by using the Strict Separation Theorem, we see that the sets C and $\{x\}$ can be separated strictly. From Exercise 2.18(c), this implies that there exists a hyperplane that passes through the origin such that one of the associated closed halfspaces contains C, but is disjoint from x. Hence, if $x \notin C$, then x cannot belong to the intersection of the homogeneous closed halfspaces containing C, proving that C contains that intersection.

(b) A homogeneous halfspace is in particular a closed convex cone containing the origin, and such a cone includes X if and only if it includes $\mathrm{cl}\big(\mathrm{cone}(X)\big)$. Hence, the intersection of all closed homogeneous halfspaces containing X and the intersection of all closed homogeneous halfspaces containing $\mathrm{cl}\big(\mathrm{cone}(X)\big)$ coincide. From what has been proved in part(a), the latter intersection is equal to $\mathrm{cl}\big(\mathrm{cone}(X)\big)$.

2.20 (Convex System Alternatives)

(a) Consider the set

$$C = \big\{u \mid \text{there exists an } x \in X \text{ such that } g_j(x) \leq u_j,\ j = 1, \ldots, r\big\},$$

which may be viewed as the projection of the set

$$M = \big\{(x, u) \mid x \in X,\ g_j(x) \leq u_j,\ j = 1, \ldots, r\big\}$$

on the space of u. Let us denote this linear transformation by A. It can be seen that

$$R_M \cap N(A) = \big\{(y, 0) \mid y \in R_X \cap R_{g_1} \cdots \cap R_{g_r}\big\},$$

where R_M denotes the recession cone of set M. Similarly, we have

$$L_M \cap N(A) = \big\{(y,0) \mid y \in L_X \cap L_{g_1} \cdots \cap L_{g_r}\big\},$$

where L_M denotes the lineality space of set M. Under conditions (1), (2), and (3), it follows from Prop. 1.5.8 that the set $AM = C$ is closed. Similarly, under condition (4), it follows from Exercise 2.7(b) that the set $AM = C$ is closed.

By assumption, there is no vector $x \in X$ such that

$$g_1(x) \le 0, \ldots, g_r(x) \le 0.$$

This implies that the origin does not belong to C. Therefore, by the Strict Separation Theorem, it follows that there exists a hyperplane that strictly separates the origin and the set C, i.e., there exists a vector μ such that

$$0 < \epsilon \le \mu' u, \qquad \forall\ u \in C. \tag{2.21}$$

This equation implies that $\mu \ge 0$ since for each $u \in C$, we have that $(u_1, \ldots, u_j + \gamma, \ldots, u_r) \in C$ for all j and $\gamma > 0$. Since $\big(g_1(x), \ldots, g_r(x)\big) \in C$ for all $x \in X$, Eq. (2.21) yields

$$\mu_1 g_1(x) + \cdots + \mu_r g_r(x) \ge \epsilon, \qquad \forall\ x \in X. \tag{2.22}$$

(b) Assume that there is no vector $x \in X$ such that

$$g_1(x) \le 0, \ldots, g_r(x) \le 0.$$

This implies by part (a) that there exists a positive scalar ϵ, and a vector $\mu \in \Re^r$ with $\mu \ge 0$, such that

$$\mu_1 g_1(x) + \cdots + \mu_r g_r(x) \ge \epsilon, \qquad \forall\ x \in X.$$

Let x be an arbitrary vector in X and let $j(x)$ be the smallest index that satisfies $j(x) = \arg\max_{j=1,\ldots,r} g_j(x)$. Then Eq. (2.22) implies that for all $x \in X$

$$\epsilon \le \sum_{j=1}^r \mu_j g_j(x) \le \sum_{j=1}^r \mu_j g_{j(x)}(x) = g_{j(x)}(x) \sum_{j=1}^r \mu_j.$$

Hence, for all $x \in X$, there exists some $j(x)$ such that

$$g_{j(x)}(x) \ge \frac{\epsilon}{\sum_{j=1}^r \mu_j} > 0.$$

This contradicts the statement that for every $\epsilon > 0$, there exists a vector $x \in X$ such that

$$g_1(x) < \epsilon, \ldots, g_r(x) < \epsilon,$$

and concludes the proof.

2.21

C is contained in the intersection of the closed halfspaces that contain C and correspond to nonvertical hyperplanes, so we focus on proving the reverse inclusion. Let $x \notin C$. Since by assumption C does not contain any vertical lines, we can apply Prop. 2.5.1, and we see that there exists a closed halfspace that correspond to a nonvertical hyperplane, containing C but not containing x. Hence, if $x \notin C$, then x cannot belong to the intersection of the closed halfspaces containing C and corresponding to nonvertical hyperplanes, proving that C contains that intersection.

2.22 (Min Common/Max Crossing Duality)

(a) Let us denote the optimal value of the min common point problem and the max crossing point problem corresponding to $\mathrm{conv}(M)$ by $w^*_{\mathrm{conv}(M)}$ and $q^*_{\mathrm{conv}(M)}$, respectively. In view of the assumption that M is compact, it follows from Prop. 1.3.2 that the set $\mathrm{conv}(M)$ is compact. Therefore, by Weierstrass' Theorem, $w^*_{\mathrm{conv}(M)}$, defined by

$$w^*_{\mathrm{conv}(M)} = \inf_{(0,w)\in \mathrm{conv}(M)} w$$

is finite. It can also be seen that the set

$$\overline{\mathrm{conv}(M)} = \big\{(u,w) \mid \text{there exists } \overline{w} \text{ with } \overline{w} \le w \text{ and } (u,\overline{w}) \in \mathrm{conv}(M)\big\}$$

is convex. Indeed, we consider vectors $(u,w) \in \overline{\mathrm{conv}(M)}$ and $(\tilde{u},\tilde{w}) \in \overline{\mathrm{conv}(M)}$, and we show that their convex combinations lie in $\overline{\mathrm{conv}(M)}$. The definition of $\overline{\mathrm{conv}(M)}$ implies that there exists some w_M and $\tilde{w}_M$ such that

$$w_M \le w, \qquad (u, w_M) \in \mathrm{conv}(M),$$

$$\tilde{w}_M \le \tilde{w}, \qquad (\tilde{u}, \tilde{w}_M) \in \mathrm{conv}(M).$$

For any $\alpha \in [0,1]$, we multiply these relations with α and $(1-\alpha)$, respectively, and add. We obtain

$$\alpha w_M + (1-\alpha)\tilde{w}_M \le \alpha w + (1-\alpha)\tilde{w}.$$

In view of the convexity of $\mathrm{conv}(M)$, we have $\alpha(u, w_M) + (1-\alpha)(\tilde{u}, \tilde{w}_M) \in \mathrm{conv}(M)$, so these equations imply that the convex combination of (u,w) and $(\tilde{u},\tilde{w})$ belongs to $\overline{\mathrm{conv}(M)}$. This proves the convexity of $\overline{\mathrm{conv}(M)}$.

Using the compactness of $\mathrm{conv}(M)$, it can be shown that for every sequence $\big\{(u_k, w_k)\big\} \subset \mathrm{conv}(M)$ with $u_k \to 0$, there holds $w^*_{\mathrm{conv}(M)} \le \liminf_{k\to\infty} w_k$. Let $\big\{(u_k, w_k)\big\} \subset \mathrm{conv}(M)$ be a sequence with $u_k \to 0$. Since $\mathrm{conv}(M)$ is compact, the sequence $\big\{(u_k, w_k)\big\}$ has a subsequence that converges to some $(0,\overline{w}) \in \mathrm{conv}(M)$. Assume without loss of generality that $\big\{(u_k, w_k)\big\}$ converges to $(0,\overline{w})$. Since $(0,\overline{w}) \in \mathrm{conv}(M)$, we get

$$w^*_{\mathrm{conv}(M)} \le \overline{w} = \liminf_{k\to\infty} w_k.$$

Therefore, by Min Common/Max Crossing Theorem I, we have

$$w^*_{\mathrm{conv}(M)} = q^*_{\mathrm{conv}(M)}. \tag{2.23}$$

Let q^* be the optimal value of the max crossing point problem corresponding to M, i.e.,

$$q^* = \sup_{\mu \in \Re^n} q(\mu),$$

where for all $\mu \in \Re^n$

$$q(\mu) = \inf_{(u,w)\in M} \{w + \mu' u\}.$$

We will show that $q^* = w^*_{\mathrm{conv}(M)}$. For every $\mu \in \Re^n$, $q(\mu)$ can be expressed as $q(\mu) = \inf_{x\in M} c'x$, where $c = (\mu, 1)$ and $x = (u, w)$. From Exercise 2.23, it follows that minimization of a linear function over a set is equivalent to minimization over its convex hull. In particular, we have

$$q(\mu) = \inf_{x\in X} c'x = \inf_{x\in \mathrm{conv}(X)} c'x,$$

from which using Eq. (2.23), we get

$$q^* = q^*_{\mathrm{conv}(M)} = w^*_{\mathrm{conv}(M)},$$

proving the desired claim.

(b) The function f is convex by the result of Exercise 2.23. Furthermore, for all $x \in \mathrm{dom}(f)$, the infimum in the definition of $f(x)$ is attained. The reason is that, for $x \in \mathrm{dom}(f)$, the set $\big\{w \mid (x,w) \in M\big\}$ is closed and bounded below, since M is closed and does not contain a halfline of the form $\big\{(x, w+\alpha) \mid \alpha \le 0\big\}$. Thus, we have $f(x) > -\infty$ for all $x \in \mathrm{dom}(f)$, while $\mathrm{dom}(f)$ is nonempty, since M is nonempty in the min common/max crossing framework. It follows that f is proper. Furthermore, by its definition, $\overline{M}$ is the epigraph of f. Finally, to show that f is closed, we argue by contradiction. If f is not closed, there exists a vector x and a sequence $\{x_k\}$ that converges to x and is such that

$$f(x) > \lim_{k\to\infty} f(x_k).$$

We claim that $\lim_{k\to\infty} f(x_k)$ is finite, i.e., that $\lim_{k\to\infty} f(x_k) > -\infty$. Indeed, by Prop. 2.5.1, the epigraph of f is contained in the upper halfspace of a nonvertical hyperplane of $\Re^{n+1}$. Since $\{x_k\}$ converges to x, the limit of $\big\{f(x_k)\big\}$ cannot be equal to $-\infty$. Thus the sequence $\big(x_k, f(x_k)\big)$, which belongs to M, converges to $\big(x, \lim_{k\to\infty} f(x_k)\big)$ Therefore, since M is closed, $\big(x, \lim_{k\to\infty} f(x_k)\big) \in M$. By the definition of f, this implies that $f(x) \le \lim_{k\to\infty} f(x_k)$, contradicting our earlier hypothesis.

(c) We prove this result by showing that all the assumptions of Min Common/Max Crossing Theorem I are satisfied. By assumption, $w^* < \infty$ and the set $\overline{M}$ is convex. Therefore, we only need to show that for every sequence $\{u_k, w_k\} \subset M$ with $u_k \to 0$, there holds $w^* \le \liminf_{k\to\infty} w_k$.

Consider a sequence $\{u_k, w_k\} \subset M$ with $u_k \to 0$. If $\liminf_{k\to\infty} w_k = \infty$, then we are done, so assume that $\liminf_{k\to\infty} w_k = \tilde{w}$ for some scalar $\tilde{w}$. Since $M \subset \overline{M}$ and $\overline{M}$ is closed by assumption, it follows that $(0, \tilde{w}) \in \overline{M}$. By the definition of the set $\overline{M}$, this implies that there exists some $\overline{w}$ with $\overline{w} \leq \tilde{w}$ and $(0, \overline{w}) \in M$. Hence we have

$$w^* = \inf_{(0,w)\in M} w \leq \overline{w} \leq \tilde{w} = \liminf_{k\to\infty} w_k,$$

proving the desired result, and thus showing that $q^* = w^*$.

2.23 (An Example of Lagrangian Duality)

(a) The corresponding max crossing problem is given by

$$q^* = \sup_{\mu\in\Re^m} q(\mu),$$

where $q(\mu)$ is given by

$$q(\mu) = \inf_{(u,w)\in M} \{w + \mu' u\} = \inf_{x\in X} \left\{ f(x) + \sum_{i=1}^{m} \mu_i (e_i' x - d_i) \right\}.$$

(b) Consider the set

$$\overline{M} = \Big\{ \big(u_1, \ldots, u_m, w\big) \mid \exists\, x \in X \text{ such that } e_i' x - d_i = u_i,\ \forall\, i,\ f(x) \leq w \Big\}.$$

We show that $\overline{M}$ is convex. To this end, we consider vectors $(u, w) \in \overline{M}$ and $(\tilde{u}, \tilde{w}) \in \overline{M}$, and we show that their convex combinations lie in $\overline{M}$. The definition of $\overline{M}$ implies that for some $x \in X$ and $\tilde{x} \in X$, we have

$$\begin{aligned} f(x) &\leq w, \qquad e_i' x - d_i = u_i, \quad i = 1, \ldots, m, \\ f(\tilde{x}) &\leq \tilde{w}, \qquad e_i' \tilde{x} - d_i = \tilde{u}_i, \quad i = 1, \ldots, m. \end{aligned}$$

For any $\alpha \in [0, 1]$, we multiply these relations with α and 1-α, respectively, and add. By using the convexity of f, we obtain

$$f\big(\alpha x + (1-\alpha)\tilde{x}\big) \leq \alpha f(x) + (1-\alpha) f(\tilde{x}) \leq \alpha w + (1-\alpha)\tilde{w},$$

$$e_i' \big(\alpha x + (1-\alpha)\tilde{x}\big) - d_i = \alpha u_i + (1-\alpha)\tilde{u}_i, \quad i = 1, \ldots, m.$$

In view of the convexity of X, we have $\alpha x + (1-\alpha)\tilde{x} \in X$, so these equations imply that the convex combination of (u, w) and $(\tilde{u}, \tilde{w})$ belongs to $\overline{M}$, thus proving that $\overline{M}$ is convex.

(c) We prove this result by showing that all the assumptions of Min Common/Max Crossing Theorem I are satisfied. By assumption, w^* is finite. It follows from part (b) that the set $\overline{M}$ is convex. Therefore, we only need to show that for every sequence $\big\{(u_k, w_k)\big\} \subset M$ with $u_k \to 0$, there holds $w^* \leq \liminf_{k\to\infty} w_k$.

Consider a sequence $\big\{(u_k, w_k)\big\} \subset M$ with $u_k \to 0$. Since X is compact and f is convex by assumption (which implies that f is continuous by Prop. 1.4.6), it follows from Prop. 1.1.9(c) that set M is compact. Hence, the sequence $\big\{(u_k, w_k)\big\}$ has a subsequence that converges to some $(0, \overline{w}) \in M$. Assume without loss of generality that $\big\{(u_k, w_k)\big\}$ converges to $(0, \overline{w})$. Since $(0, \overline{w}) \in M$, we get

$$w^* = \inf_{(0,w)\in M} w \leq \overline{w} = \liminf_{k\to\infty} w_k,$$

proving the desired result, and thus showing that $q^* = w^*$.

(d) We prove this result by showing that all the assumptions of Min Common/Max Crossing Theorem II are satisfied. By assumption, w^* is finite. It follows from part (b) that the set $\overline{M}$ is convex. Therefore, we only need to show that the set

$$D = \big\{(e_1'x - d_1, \ldots, e_m'x - d_m) \mid x \in X\big\}$$

contains the origin in its relative interior. The set D can equivalently be written as

$$D = E \cdot X - d,$$

where E is a matrix, whose rows are the vectors e_i', $i = 1, \ldots, m$, and d is a vector with entries equal to d_i, $i = 1, \ldots, m$. By Prop. 1.4.4 and Prop. 1.4.5(b), it follows that

$$\mathrm{ri}(D) = E \cdot \mathrm{ri}(X) - d.$$

Hence the assumption that there exists a vector $\overline{x} \in \mathrm{ri}(X)$ such that $E\overline{x} - d = 0$ implies that 0 belongs to the relative interior of D, thus showing that $q^* = w^*$ and that the max crossing problem has an optimal solution.

2.24 (Saddle Points in Two Dimensions)

We consider a function ϕ of two real variables x and z taking values in compact intervals X and Z, respectively. We assume that for each $z \in Z$, the function $\phi(\cdot, z)$ is minimized over X at a unique point denoted $\hat{x}(z)$, and for each $x \in X$, the function $\phi(x, \cdot)$ is maximized over Z at a unique point denoted $\hat{z}(x)$,

$$\hat{x}(z) = \arg\min_{x\in X} \phi(x, z), \qquad \hat{z}(x) = \arg\max_{z\in Z} \phi(x, z).$$

Consider the composite function $f : X \mapsto X$ given by

$$f(x) = \hat{x}\big(\hat{z}(x)\big),$$

which is a continuous function in view of the assumption that the functions $\hat{x}(z)$ and $\hat{z}(x)$ are continuous over Z and X, respectively. Assume that the compact interval X is given by $[a, b]$. We now show that the function f has a fixed point, i.e., there exists some $x^* \in [a, b]$ such that

$$f(x^*) = x^*.$$

Define the function $g : X \mapsto X$ by

$$g(x) = f(x) - x.$$

Assume that $f(a) > a$ and $f(b) < b$, since otherwise we are done. We have

$$g(a) = f(a) - a > 0,$$

$$g(b) = f(b) - b < 0.$$

Since g is a continuous function, the preceding relations imply that there exists some $x^* \in (a, b)$ such that $g(x^*) = 0$, i.e., $f(x^*) = x^*$. Hence, we have

$$\hat{x}\big(\hat{z}(x^*)\big) = x^*.$$

Denoting $\hat{z}(x^*)$ by z^*, we get

$$x^* = \hat{x}(z^*), \qquad z^* = \hat{z}(x^*). \tag{2.24}$$

By definition, a pair $(\overline{x}, \overline{z})$ is a saddle point if and only if

$$\max_{z \in Z} \phi(\overline{x}, z) = \phi(\overline{x}, \overline{z}) = \min_{x \in X} \phi(x, \overline{z}),$$

or equivalently, if $\overline{x} = \hat{x}(\overline{z})$ and $\overline{z} = \hat{z}(\overline{x})$. Therefore, from Eq. (2.24), we see that (x^*, z^*) is a saddle point of ϕ.

We now consider the function $\phi(x, z) = x^2 + z^2$ over $X = [0, 1]$ and $Z = [0, 1]$. For each $z \in [0, 1]$, the function $\phi(\cdot, z)$ is minimized over $[0, 1]$ at a unique point $\hat{x}(z) = 0$, and for each $x \in [0, 1]$, the function $\phi(x, \cdot)$ is maximized over $[0, 1]$ at a unique point $\hat{z}(x) = 1$. These two curves intersect at $(x^*, z^*) = (0, 1)$, which is the unique saddle point of ϕ.

2.25 (Saddle Points of Quadratic Functions)

Let X and Z be closed and convex sets. Then, for each $z \in Z$, the function $t_z : \Re^n \mapsto (-\infty, \infty]$ defined by

$$t_z(x) = \begin{cases} \phi(x, z) & \text{if } x \in X, \\ \infty & \text{otherwise,} \end{cases}$$

is closed and convex in view of the assumption that Q is a positive semidefinite symmetric matrix. Similarly, for each $x \in X$, the function $r_x : \Re^m \mapsto (-\infty, \infty]$ defined by

$$r_x(z) = \begin{cases} -\phi(x, z) & \text{if } z \in Z, \\ \infty & \text{otherwise,} \end{cases}$$

is closed and convex in view of the assumption that R is a positive semidefinite symmetric matrix. Hence, Assumption 2.6.1 is satisfied. Let also Assumptions 2.6.2 and 2.6.3 hold, i.e,

$$\inf_{x \in X} \sup_{z \in Z} \phi(x, z) < \infty,$$

and

$$-\infty < \sup_{z \in Z} \inf_{x \in X} \phi(x, z).$$

By the positive semidefiniteness of Q, it can be seen that, for each $z \in Z$, the recession cone of the function t_z is given by

$$R_{t_z} = R_X \cap N(Q) \cap \{y \mid y' Dz \leq 0\},$$

where R_X is the recession cone of the convex set X and $N(Q)$ is the null space of the matrix Q. Similarly, for each $z \in Z$, the constancy space of the function t_z is given by

$$L_{t_z} = L_X \cap N(Q) \cap \{y \mid y' Dz = 0\},$$

where L_X is the lineality space of the set X. By the positive semidefiniteness of R, for each $x \in X$, it can be seen that the recession cone of the function r_x is given by

$$R_{r_x} = R_Z \cap N(R) \cap \{y \mid x' Dy \geq 0\},$$

where R_Z is the recession cone of the convex set Z and $N(R)$ is the null space of the matrix R. Similarly, for each $x \in X$, the constancy space of the function r_x is given by

$$L_{r_x} = L_Z \cap N(R) \cap \{y \mid x' Dy = 0\},$$

where L_Z is the lineality space of the set Z.

If

$$\bigcap_{z \in Z} R_{t_z} = \{0\}, \quad \text{and} \quad \bigcap_{x \in X} R_{r_x} = \{0\}, \tag{2.25}$$

then it follows from the Saddle Point Theorem part (a), that the set of saddle points of ϕ is nonempty and compact. [In particular, the condition given in Eq. (2.25) holds when Q and R are positive definite matrices, or if X and Z are compact.]

Similarly, if

$$\bigcap_{z \in Z} R_{t_z} = \bigcap_{z \in Z} L_{t_z}, \quad \text{and} \quad \bigcap_{x \in X} R_{r_x} = \bigcap_{x \in X} L_{r_x},$$

then it follows from the Saddle Point Theorem part (b), that the set of saddle points of ϕ is nonempty.

Convex Analysis and Optimization

Chapter 3 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 3: SOLUTION MANUAL

3.1 (Cone Decomposition Theorem)

(a) Let $\hat{x}$ be the projection of x on C, which exists and is unique since C is closed and convex. By the Projection Theorem (Prop. 2.2.1), we have

$$(x - \hat{x})'(y - \hat{x}) \le 0, \qquad \forall\ y \in C.$$

Since C is a cone, we have $(1/2)\hat{x} \in C$ and $2\hat{x} \in C$, and by taking $y = (1/2)\hat{x}$ and $y = 2\hat{x}$ in the preceding relation, it follows that

$$(x - \hat{x})'\hat{x} = 0.$$

By combining the preceding two relations, we obtain

$$(x - \hat{x})'y \le 0, \qquad \forall\ y \in C,$$

implying that $x - \hat{x} \in C^*$.

Conversely, if $\hat{x} \in C$, $(x - \hat{x})'\hat{x} = 0$, and $x - \hat{x} \in C^*$, then it follows that

$$(x - \hat{x})'(y - \hat{x}) \le 0, \qquad \forall\ y \in C,$$

and by the Projection Theorem, $\hat{x}$ is the projection of x on C.

(b) Suppose that property (i) holds, i.e., x_1 and x_2 are the projections of x on C and C^*, respectively. Then, by part (a), we have

$$x_1 \in C, \qquad (x - x_1)'x_1 = 0, \qquad x - x_1 \in C^*.$$

Let $y = x - x_1$, so that the preceding relation can equivalently be written as

$$x - y \in C = (C^*)^*, \qquad y'(x - y) = 0, \qquad y \in C^*.$$

By using part (a), we conclude that y is the projection of x on C^*. Since by the Projection Theorem, the projection of a vector on a closed convex set is unique, it follows that $y = x_2$. Thus, we have $x = x_1 + x_2$ and in view of the preceding two relations, we also have $x_1 \in C$, $x_2 \in C^*$, and $x_1'x_2 = 0$. Hence, property (ii) holds.

Conversely, suppose that property (ii) holds, i.e., $x = x_1 + x_2$ with $x_1 \in C$, $x_2 \in C^*$, and $x_1'x_2 = 0$. Then, evidently the relations

$$x_1 \in C, \qquad (x - x_1)'x_1 = 0, \qquad x - x_1 \in C^*,$$

$$x_2 \in C^*, \qquad (x - x_2)'x_2 = 0, \qquad x - x_2 \in C$$

are satisfied, so that by part (a), x_1 and x_2 are the projections of x on C and C^*, respectively. Hence, property (i) holds.

3.2

If $a \in C^* + \big\{x \mid \|x\| \leq \gamma/\beta\big\}$, then

$$a = \hat{a} + \overline{a} \qquad \text{with} \quad \hat{a} \in C^* \quad \text{and} \quad \|\overline{a}\| \leq \gamma/\beta.$$

Since C is a closed convex cone, by the Polar Cone Theorem (Prop. 3.1.1), we have $(C^*)^* = C$, implying that for all x in C with $\|x\| \leq \beta$,

$$\hat{a}'x \leq 0 \qquad \text{and} \qquad \overline{a}'x \leq \|\overline{a}\| \cdot \|x\| \leq \gamma.$$

Hence,

$$a'x = (\hat{a} + \overline{a})'x \leq \gamma, \qquad \forall\ x \in C \ \text{ with } \ \|x\| \leq \beta,$$

thus implying that

$$\max_{\|x\| \leq \beta,\ x \in C} a'x \leq \gamma.$$

Conversely, assume that $a'x \leq \gamma$ for all $x \in C$ with $\|x\| \leq \beta$. Let $\hat{a}$ and $\overline{a}$ be the projections of a on C^* and C, respectively. By the Cone Decomposition Theorem (cf. Exercise 3.1), we have $a = \hat{a} + \overline{a}$ with $\hat{a} \in C^*$, $\overline{a} \in C$, and $\hat{a}'\overline{a} = 0$. Since $a'x \leq \gamma$ for all $x \in C$ with $\|x\| \leq \beta$ and $\overline{a} \in C$, we obtain

$$a' \frac{\overline{a}}{\|\overline{a}\|} \beta = (\hat{a} + \overline{a})' \frac{\overline{a}}{\|\overline{a}\|} \beta = \|\overline{a}\| \beta \leq \gamma,$$

implying that $\|\overline{a}\| \leq \gamma/\beta$, and showing that $a \in C^* + \big\{x \mid \|x\| \leq \gamma/\beta\big\}$.

3.3

Note that $\mathrm{aff}(C)$ is a subspace of $\Re^n$ because C is a cone in $\Re^n$. We first show that

$$L_{C^*} = \big(\mathrm{aff}(C)\big)^\perp.$$

Let $y \in L_{C^*}$. Then, by the definition of the lineality space (see Chapter 1), both vectors y and $-y$ belong to the recession cone R_{C^*}. Since $0 \in C^*$, it follows that $0 + y$ and $0 - y$ belong to C^*. Therefore,

$$y'x \leq 0, \qquad (-y)'x \leq 0, \qquad \forall\ x \in C,$$

implying that

$$y'x = 0, \qquad \forall\ x \in C. \tag{3.1}$$

Let the dimension of the subspace $\mathrm{aff}(C)$ be m. By Prop. 1.4.1, there exist vectors $x_0, x_1, \ldots, x_m$ in $\mathrm{ri}(C)$ such that $x_1 - x_0, \ldots, x_m - x_0$ span $\mathrm{aff}(C)$. Thus, for any $z \in \mathrm{aff}(C)$, there exist scalars $\beta_1, \ldots, \beta_m$ such that

$$z = \sum_{i=1}^{m} \beta_i (x_i - x_0).$$

By using this relation and Eq. (3.1), for any $z \in \mathrm{aff}(C)$, we obtain

$$y'z = \sum_{i=1}^{m} \beta_i y'(x_i - x_0) = 0,$$

implying that $y \in \big(\mathrm{aff}(C)\big)^{\perp}$. Hence, $L_{C^*} \subset \big(\mathrm{aff}(C)\big)^{\perp}$.

Conversely, let $y \in \big(\mathrm{aff}(C)\big)^{\perp}$, so that in particular, we have

$$y'x = 0, \qquad (-y)'x = 0, \qquad \forall\ x \in C.$$

Therefore, $0 + \alpha y \in C^*$ and $0 + \alpha(-y) \in C^*$ for all $\alpha \geq 0$, and since C^* is a closed convex set, by the Recession Cone Theorem(b) [Prop. 1.5.1(b)], it follows that y and $-y$ belong to the recession cone R_{C^*}. Hence, y belongs to the lineality space of C^*, showing that $\big(\mathrm{aff}(C)\big)^{\perp} \subset L_{C^*}$ and completing the proof of the equality $L_{C^*} = \big(\mathrm{aff}(C)\big)^{\perp}$.

By definition, we have $\dim(C) = \dim\big(\mathrm{aff}(C)\big)$ and since $L_{C^*} = \big(\mathrm{aff}(C)\big)^{\perp}$, we have $\dim(L_{C^*}) = \dim\Big(\big(\mathrm{aff}(C)\big)^{\perp}\Big)$. This implies that

$$\dim(C) + \dim(L_{C^*}) = n.$$

By replacing C with C^* in the preceding relation, and by using the Polar Cone Theorem (Prop. 3.1.1), we obtain

$$\dim(C^*) + \dim\big(L_{(C^*)^*}\big) = \dim(C^*) + \dim\big(L_{\mathrm{cl}(\mathrm{conv}(C))}\big) = n.$$

Furthermore, since

$$L_{\mathrm{conv}(C)} \subset L_{\mathrm{cl}(\mathrm{conv}(C))},$$

it follows that

$$\dim(C^*) + \dim\big(L_{\mathrm{conv}(C)}\big) \leq \dim(C^*) + \dim\big(L_{\mathrm{cl}(\mathrm{conv}(C))}\big) = n.$$

3.4 (Polar Cone Operations)

(a) It suffices to consider the case where $m = 2$. Let $(y_1, y_2) \in (C_1 \times C_2)^*$. Then, we have $(y_1, y_2)'(x_1, x_2) \leq 0$ for all $(x_1, x_2) \in C_1 \times C_2$, or equivalently

$$y_1'x_1 + y_2'x_2 \leq 0, \qquad \forall\ x_1 \in C_1, \quad \forall\ x_2 \in C_2.$$

Since C_2 is a cone, 0 belongs to its closure, so by letting $x_2 \to 0$ in the preceding relation, we obtain $y_1'x_1 \leq 0$ for all $x_1 \in C_1$, showing that $y_1 \in C_1^*$. Similarly, we obtain $y_2 \in C_2^*$, and therefore $(y_1, y_2) \in C_1^* \times C_2^*$, implying that $(C_1 \times C_2)^* \subset C_1^* \times C_2^*$.

Conversely, let $y_1 \in C_1^*$ and $y_2 \in C_2^*$. Then, we have

$$(y_1, y_2)'(x_1, x_2) = y_1'x_1 + y_2'x_2 \leq 0, \qquad \forall\, x_1 \in C_1, \quad \forall\, x_2 \in C_2,$$

implying that $(y_1, y_2) \in (C_1 \times C_2)^*$, and showing that $C_1^* \times C_2^* \subset (C_1 \times C_2)^*$.

(b) A vector y belongs to the polar cone of $\cup_{i\in I} C_i$ if and only if $y'x \leq 0$ for all $x \in C_i$ and all $i \in I$, which is equivalent to having $y \in C_i^*$ for every $i \in I$. Hence, y belongs to $\big(\cup_{i\in I} C_i\big)^*$ if and only if y belongs to $\cap_{i\in I} C_i^*$.

(c) Let $y \in (C_1 + C_2)^*$, so that

$$y'(x_1 + x_2) \leq 0, \qquad \forall\, x_1 \in C_1, \qquad \forall\, x_2 \in C_2. \tag{3.2}$$

Since the zero vector is in the closures of C_1 and C_2, by letting $x_2 \to 0$ with $x_2 \in C_2$ in Eq. (3.2), we obtain

$$y'x_1 \leq 0, \qquad \forall\, x_1 \in C_1,$$

and similarly, by letting $x_1 \to 0$ with $x_1 \in C_1$ in Eq. (3.2), we obtain

$$y'x_2 \leq 0, \qquad \forall\, x_2 \in C_2.$$

Thus, $y \in C_1^* \cap C_2^*$, showing that $(C_1 + C_2)^* \subset C_1^* \cap C_2^*$.

Conversely, let $y \in C_1^* \cap C_2^*$. Then, we have

$$y'x_1 \leq 0, \qquad \forall\, x_1 \in C_1,$$

$$y'x_2 \leq 0, \qquad \forall\, x_2 \in C_2,$$

implying that

$$y'(x_1 + x_2) \leq 0, \qquad \forall\, x_1 \in C_1, \quad \forall\, x_2 \in C_2.$$

Hence $y \in (C_1 + C_2)^*$, showing that $C_1^* \cap C_2^* \subset (C_1 + C_2)^*$.

(d) Since C_1 and C_2 are closed convex cones, by the Polar Cone Theorem (Prop. 3.1.1) and by part (b), it follows that

$$C_1 \cap C_2 = (C_1^*)^* \cap (C_2^*)^* = (C_1^* + C_2^*)^*.$$

By taking the polars and by using the Polar Cone Theorem, we obtain

$$(C_1 \cap C_2)^* = \big((C_1^* + C_2^*)^*\big)^* = \mathrm{cl}\big(\mathrm{conv}(C_1^* + C_2^*)\big).$$

The cone $C_1^* + C_2^*$ is convex, so that

$$(C_1 \cap C_2)^* = \mathrm{cl}(C_1^* + C_2^*).$$

Suppose now that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$. We will show that $C_1^* + C_2^*$ is closed by using Exercise 1.43. According to this exercise, if for any nonempty closed convex sets $\overline{C}_1$ and $\overline{C}_2$ in $\Re^n$, the equality $y_1 + y_2 = 0$ with $y_1 \in R_{\overline{C}_1}$ and

$y_2 \in R_{\overline{C}_2}$ implies that y_1 and y_2 belong to the lineality spaces of $\overline{C}_1$ and $\overline{C}_2$, respectively, then the vector sum $\overline{C}_1 + \overline{C}_2$ is closed.

Let $y_1 + y_2 = 0$ with $y_1 \in R_{C_1^*}$ and $y_2 \in R_{C_2^*}$. Because C_1^* and C_2^* are closed convex cones, we have $R_{C_1^*} = C_1^*$ and $R_{C_2^*} = C_2^*$, so that $y_1 \in C_1^*$ and $y_2 \in C_2^*$. The lineality space of a cone is the set of vectors y such that y and $-y$ belong to the cone, so that in view of the preceding discussion, to show that $C_1^* + C_2^*$ is closed, it suffices to prove that $-y_1 \in C_1^*$ and $-y_2 \in C_2^*$.

Since $y_1 = -y_2$ and $y_1 \in C_1^*$, it follows that

$$y_2' x \geq 0, \qquad \forall\ x \in C_1, \tag{3.3}$$

and because $y_2 \in C_2^*$, we have

$$y_2' x \leq 0, \qquad \forall\ x \in C_2,$$

which combined with the preceding relation yields

$$y_2' x = 0, \qquad \forall\ x \in C_1 \cap C_2. \tag{3.4}$$

In view of the fact $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$, and Eqs. (3.3) and (3.4), it follows that the linear function $y_2' x$ attains its minimum over the convex set C_1 at a point in the relative interior of C_1, implying that $y_2' x = 0$ for all $x \in C_1$ (cf. Prop. 1.4.2). Therefore, $y_2 \in C_1^*$ and since $y_2 = -y_1$, we have $-y_1 \in C_1^*$. By exchanging the roles of y_1 and y_2 in the preceding analysis, we similarly show that $-y_2 \in C_2^*$, completing the proof.

(e) By drawing the cones C_1 and C_2, it can be seen that $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$ and

$$C_1 \cap C_2 = \big\{(x_1, x_2, x_3) \mid x_1 = 0,\ x_2 = -x_3,\ x_3 \leq 0\big\},$$

$$C_1^* = \big\{(y_1, y_2, y_3) \mid y_1^2 + y_2^2 \leq y_3^2,\ y_3 \geq 0\big\},$$

$$C_2^* = \big\{(z_1, z_2, z_3) \mid z_1 = 0,\ z_2 = z_3\big\}.$$

Clearly, $x_1 + x_2 + x_3 = 0$ for all $x \in C_1 \cap C_2$, implying that $(1, 1, 1) \in (C_1 \cap C_2)^*$. Suppose that $(1, 1, 1) \in C_1^* + C_2^*$, so that $(1, 1, 1) = (y_1, y_2, y_3) + (z_1, z_2, z_3)$ for some $(y_1, y_2, y_3) \in C_1^*$ and $(z_1, z_2, z_3) \in C_2^*$, implying that $y_1 = 1$, $y_2 = 1 - z_2$, $y_3 = 1 - z_2$ for some $z_2 \in \Re$. However, this point does not belong to C_1^*, which is a contradiction. Therefore, $(1, 1, 1)$ is not in $C_1^* + C_2^*$. Hence, when $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$, the relation

$$(C_1 \cap C_2)^* = C_1^* + C_2^*$$

may fail.

3.5 (Linear Transformations and Polar Cones)

We have $y \in (AC)^*$ if and only if $y'Ax \le 0$ for all $x \in C$, which is equivalent to $(A'y)'x \le 0$ for all $x \in C$. This is in turn equivalent to $A'y \in C^*$. Hence, $y \in (AC)^*$ if and only if $y \in (A')^{-1} \cdot C^*$, showing that

$$(AC)^* = (A')^{-1} \cdot C^*. \tag{3.5}$$

We next show that for a closed convex cone $K \subset \Re^m$, we have

$$\big(A^{-1} \cdot K\big)^* = \mathrm{cl}(A'K^*).$$

Let $y \in \big(A^{-1} \cdot K\big)^*$ and to arrive at a contradiction, assume that $y \notin \mathrm{cl}(A'K^*)$. By the Strict Separation Theorem (Prop. 2.4.3), the closed convex cone $\mathrm{cl}(A'K^*)$ and the vector y can be strictly separated, i.e., there exist a vector $a \in \Re^n$ and a scalar b such that

$$a'x < b < a'y, \qquad \forall\, x \in \mathrm{cl}(A'K^*).$$

If $a'x > 0$ for some $x \in \mathrm{cl}(A'K^*)$, then since $\mathrm{cl}(A'K^*)$ is a cone, we would have $\lambda x \in \mathrm{cl}(A'K^*)$ for all $\lambda > 0$, implying that $a'(\lambda x) \to \infty$ when $\lambda \to \infty$, which contradicts the preceding relation. Thus, we must have $a'x \le 0$ for all $x \in \mathrm{cl}(A'K^*)$, and since $0 \in \mathrm{cl}(A'K^*)$, it follows that

$$\sup_{x \in \mathrm{cl}(A'K^*)} a'x = 0 \le b < a'y. \tag{3.6}$$

Therefore, $a \in \big(\mathrm{cl}(A'K^*)\big)^*$, and since $\big(\mathrm{cl}(A'K^*)\big)^* \subset (A'K^*)^*$, it follows that $a \in (A'K^*)^*$. In view of Eq. (3.5) and the Polar Cone Theorem (Prop. 3.1.1), we have

$$(A'K^*)^* = A^{-1}(K^*)^* = A^{-1} \cdot K,$$

implying that $a \in A^{-1} \cdot K$. Because $y \in \big(A^{-1} \cdot K\big)^*$, it follows that $y'a \le 0$, contradicting Eq. (3.6). Hence, we must have $y \in \mathrm{cl}(A'K^*)$, showing that

$$\big(A^{-1} \cdot K\big)^* \subset \mathrm{cl}(A'K^*).$$

To show the reverse inclusion, let $y \in A'K^*$ and assume, to arrive at a contradiction, that $y \notin (A^{-1} \cdot K)^*$. By the Strict Separation Theorem (Prop. 2.4.3), the closed convex cone $(A^{-1} \cdot K)^*$ and the vector y can be strictly separated, i.e., there exist a vector $\overline{a} \in \Re^n$ and a scalar $\overline{b}$ such that

$$\overline{a}'x < \overline{b} < \overline{a}'y, \qquad \forall\, x \in (A^{-1} \cdot K)^*.$$

Similar to the preceding analysis, since $(A^{-1} \cdot K)^*$ is a cone, it can be seen that

$$\sup_{x \in (A^{-1} \cdot K)^*} \overline{a}'x = 0 \le \overline{b} < \overline{a}'y, \tag{3.7}$$

implying that $\overline{a} \in \big((A^{-1}\cdot K)^*\big)^*$. Since K is a closed convex cone and A is a linear (and therefore continuous) transformation, the set $A^{-1}\cdot K$ is a closed convex cone. Furthermore, by the Polar Cone Theorem, we have that $\big((A^{-1}\cdot K)^*\big)^* = A^{-1}\cdot K$. Therefore, $\overline{a} \in A^{-1}\cdot K$, implying that $A\overline{a} \in K$. Since $y \in A'K^*$, we have $y = A'v$ for some $v \in K^*$, and it follows that

$$y'\overline{a} = (A'v)'\overline{a} = v'A\overline{a} \leq 0,$$

contradicting Eq. (3.7). Hence, we must have $y \in (A^{-1} \cdot K)^*$, implying that

$$A'K^* \subset (A^{-1} \cdot K)^*.$$

Taking the closure of both sides of this relation, we obtain

$$\mathrm{cl}(A'K^*) \subset (A^{-1} \cdot K)^*,$$

completing the proof.

Suppose that $\mathrm{ri}(K^*) \cap R(A) \neq \emptyset$. We will show that the cone $A'K^*$ is closed by using Exercise 1.42. According to this exercise, if $R_{K^*} \cap N(A')$ is a subspace of the lineality space L_{K^*} of K^*, then

$$\mathrm{cl}(A'K^*) = A'K^*.$$

Thus, it suffices to verify that $R_{K^*} \cap N(A')$ is a subspace of L_{K^*}. Indeed, we will show that $R_{K^*} \cap N(A') = L_{K^*} \cap N(A')$.

Let $y \in K^* \cap N(A')$. Because $y \in K^*$, we obtain

$$(-y)'x \geq 0, \qquad \forall\ x \in K. \tag{3.8}$$

For $y \in N(A')$, we have $-y \in N(A')$ and since $N(A') = R(A)^\perp$, it follows that

$$(-y)'z = 0, \qquad \forall\ z \in R(A). \tag{3.9}$$

In view of the relation $\mathrm{ri}(K) \cap R(A) \neq \emptyset$, and Eqs. (3.8) and (3.9), the linear function $(-y)'x$ attains its minimum over the convex set K at a point in the relative interior of K, implying that $(-y)'x = 0$ for all $x \in K$ (cf. Prop. 1.4.2). Hence $(-y) \in K^*$, so that $y \in L_{K^*}$ and because $y \in N(A')$, we see that $y \in L_{K^*} \cap N(A')$. The reverse inclusion follows directly from the relation $L_{K^*} \subset R_{K^*}$, thus completing the proof.

3.6 (Pointed Cones and Bases)

(a) $\Rightarrow$ (b) Since C is a pointed cone, $C \cap (-C) = \{0\}$, so that

$$\big(C \cap (-C)\big)^* = \Re^n.$$

On the other hand, by Exercise 3.4, it follows that

$$\big(C \cap (-C)\big)^* = \mathrm{cl}(C^* - C^*),$$

which when combined with the preceding relation yields $\mathrm{cl}(C^* - C^*) = \Re^n$.

(b) $\Rightarrow$ (c) Since C is a closed convex cone, by the polar cone operations of Exercise 3.4, it follows that

$$\big(C \cap (-C)\big)^* = \mathrm{cl}(C^* - C^*) = \Re^n.$$

By taking the polars and using the Polar Cone Theorem (Prop. 3.1.1), we obtain

$$\Big(\big(C \cap (-C)\big)^*\Big)^* = C \cap (-C) = \{0\}. \tag{3.10}$$

Now, to arrive at a contradiction assume that there is a vector $\hat{x} \in \Re^n$ such that $\hat{x} \notin C^* - C^*$. Then, by the Separating Hyperplane Theorem (Prop. 2.4.2), there exists a nonzero vector $a \in \Re^n$ such that

$$a'\hat{x} \geq a'x, \qquad \forall\, x \in C^* - C^*.$$

If $a'x > 0$ for some $x \in C^* - C^*$, then since $C^* - C^*$ is a cone, the right hand-side of the preceding relation can be arbitrarily large, a contradiction. Thus, we have $a'x \leq 0$ for all $x \in C^* - C^*$, implying that $a \in (C^* - C^*)^*$. By the polar cone operations of Exercise 3.4(b) and the Polar Cone Theorem, it follows that

$$(C^* - C^*)^* = (C^*)^* \cap (-C^*)^* = C \cap (-C).$$

Thus, $a \in C \cap (-C)$ with $a \neq 0$, contradicting Eq. (3.10). Hence, we must have $C^* - C^* = \Re^n$.

(c) $\Rightarrow$ (d) Because $C^* \subset \mathrm{aff}(C^*)$ and $-C^* \subset \mathrm{aff}(C^*)$, we have $C^* - C^* \subset \mathrm{aff}(C^*)$ and since $C^* - C^* = \Re^n$, it follows that $\mathrm{aff}(C^*) = \Re^n$, showing that C^* has nonempty interior.

(d) $\Rightarrow$ (e) Let v be a vector in the interior of C^*. Then, there exists a positive scalar δ such that the vector $v + \delta\frac{y}{\|y\|}$ is in C^* for all $y \in \Re^n$ with $y \neq 0$, i.e.,

$$\left(v + \delta\frac{y}{\|y\|}\right)' x \leq 0, \qquad \forall\, x \in C, \quad \forall\, y \in \Re^n,\ y \neq 0.$$

By taking $y = x$, it follows that

$$\left(v + \delta\frac{x}{\|x\|}\right)' x \leq 0, \qquad \forall\, x \in C,\ x \neq 0,$$

implying that

$$v'x + \delta\|x\| \leq 0, \qquad \forall\, x \in C,\ x \neq 0.$$

Clearly, this relation holds for $x = 0$, so that

$$v'x \leq -\delta\|x\|, \qquad \forall\, x \in C.$$

Multiplying the preceding relation with -1 and letting $\hat{x} = -v$, we obtain

$$\hat{x}'x \geq \delta\|x\|, \qquad \forall\, x \in C.$$

(e) $\Rightarrow$ (f) Let

$$D = \big\{y \in C \mid \hat{x}'y = 1\big\}.$$

Then, D is a closed convex set since it is the intersection of the closed convex cone C and the closed convex set $\{y \mid \hat{x}'y = 1\}$. Obviously, $0 \notin D$. Thus, to show that D is a base for C, it remains to prove that $C = \text{cone}(D)$. Take any $x \in C$. If $x = 0$, then $x \in \text{cone}(D)$ and we are done, so assume that $x \neq 0$. We have by hypothesis

$$\hat{x}'x \geq \delta\|x\| > 0, \qquad \forall\, x \in C,\ x \neq 0,$$

so we may define $\hat{y} = \frac{x}{\hat{x}'x}$. Clearly, $\hat{y} \in D$ and $x = (\hat{x}'x)\hat{y}$ with $\hat{x}'x > 0$, showing that $x \in \text{cone}(D)$ and that $C \subset \text{cone}(D)$. Since $D \subset C$, the inclusion $\text{cone}(D) \subset C$ is obvious. Thus, $C = \text{cone}(D)$ and D is a base for C. Furthermore, for every y in D, since y is also in C, we have

$$1 = \hat{x}'y \geq \delta\|y\|,$$

showing that D is bounded and completing the proof.

(f) $\Rightarrow$ (a) Since C has a bounded base, $C = \text{cone}(D)$ for some bounded convex set D with $0 \notin \text{cl}(D)$. To arrive at a contradiction, we assume that the cone C is not pointed, so that there exists a nonzero vector $d \in C \cap (-C)$, implying that d and $-d$ are in C. Let $\{\lambda_k\}$ be a sequence of positive scalars. Since $\lambda_k d \in C$ for all k and D is a base for C, there exist a sequence $\{\mu_k\}$ of positive scalars and a sequence $\{y_k\}$ of vectors in D such that

$$\lambda_k d = \mu_k y_k, \qquad \forall\, k.$$

Therefore, $y_k = \frac{\lambda_k}{\mu_k} d \in D$ for all k and because D is bounded, the sequence $\big\{y_k\big\}$ has a subsequence converging to some $y \in \text{cl}(D)$. Without loss of generality, we may assume that $y_k \to y$, which in view of $y_k = \frac{\lambda_k}{\mu_k} d$ for all k, implies that $y = \alpha d$ and $\alpha d \in \text{cl}(D)$ for some $\alpha \geq 0$. Furthermore, by the definition of base, we have $0 \notin \text{cl}(D)$, so that $\alpha > 0$. Similar to the preceding, by replacing d with $-d$, we can show that $\tilde{\alpha}(-d) \in \text{cl}(D)$ for some positive scalar $\tilde{\alpha}$. Therefore, $\alpha d \in \text{cl}(D)$ and $\tilde{\alpha}(-d) \in \text{cl}(D)$ with $\alpha > 0$ and $\tilde{\alpha} > 0$. Since D is convex, its closure $\text{cl}(D)$ is also convex, implying that $0 \in \text{cl}(D)$, contradicting the definition of a base. Hence, the cone C must be pointed.

3.7

Let the closed convex cone C be polyhedral, and of the form

$$C = \big\{x \mid a_j'x \leq 0,\ j = 1, \ldots, r\big\},$$

for some vectors a_j in $\Re^n$. By Farkas' Lemma [Prop. 3.2.1(b)], we have

$$C^* = \mathrm{cone}\big(\{a_1, \ldots, a_r\}\big),$$

so the polar cone of a polyhedral cone is finitely generated. Conversely, using the Polar Cone Theorem, we have

$$\mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)^* = \big\{x \mid a_j'x \le 0,\ j = 1, \ldots, r\big\},$$

so the polar of a finitely generated cone is polyhedral. Thus, a closed convex cone is polyhedral if and only if its polar cone is finitely generated. By the Minkowski-Weyl Theorem [Prop. 3.2.1(c)], a cone is finitely generated if and only if it is polyhedral. Therefore, a closed convex cone is polyhedral if and only if its polar cone is polyhedral.

3.8

(a) We first show that C is a subset of R_P, the recession cone of P. Let $\overline{y} \in C$, and choose any $\alpha \ge 0$ and $x \in P$ of the form $x = \sum_{j=1}^m \mu_j v_j$. Since C is a cone, $\alpha\overline{y} \in C$, so that $x + \alpha\overline{y} \in P$ for all $\alpha \ge 0$. It follows that $\overline{y} \in R_P$. Hence $C \subset R_P$.

Conversely, to show that $R_P \subset C$, let $\overline{y} \in R_P$ and take any $x \in P$. Then $x + k\overline{y} \in P$ for all $k \ge 1$. Since $P = V + C$, where $V = \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big)$, it follows that

$$x + k\overline{y} = v^k + y^k, \qquad \forall\ k \ge 1,$$

with $v^k \in V$ and $y^k \in C$ for all $k \ge 1$. Because V is compact, the sequence $\{v^k\}$ has a limit point $v \in V$, and without loss of generality, we may assume that $v^k \to v$. Then

$$\lim_{k\to\infty} \|k\overline{y} - y^k\| = \lim_{k\to\infty} \|v^k - x\| = \|v - x\|,$$

implying that

$$\lim_{k\to\infty} \big\|\overline{y} - (1/k)y^k\big\| = 0.$$

Therefore, the sequence $\big\{(1/k)y^k\big\}$ converges to $\overline{y}$. Since $y^k \in C$ for all $k \ge 1$, the sequence $\big\{(1/k)y^k\big\}$ is in C, and by the closedness of C, it follows that $\overline{y} \in C$. Hence, $R_P \subset C$.

(b) Any point in P has the form $v + y$ with $v \in \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big)$ and $y \in C$, or equivalently

$$v + y = \frac{1}{2}v + \frac{1}{2}(v + 2y),$$

with v and $v + 2y$ being two distinct points in P if $y \ne 0$. Therefore, none of the points $v + y$, with $v \in \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big)$ and $y \in C$, is an extreme point of P if $y \ne 0$. Hence, an extreme point of P must be in the set $\{v_1, \ldots, v_m\}$. Since by definition, an extreme point of P is not a convex combination of points in P, an extreme point of P must be equal to some v_i that cannot be expressed as a convex combination of the remaining vectors v_j, $j \ne i$.

3.9 (Polyhedral Cones and Sets under Linear Transformations)

(a) Let A be an $m \times n$ matrix and let C be a polyhedral cone in $\Re^n$. By the Minkowski-Weyl Theorem [Prop. 3.2.1(c)], C is finitely generated, so that

$$C = \left\{ x \;\middle|\; x = \sum_{j=1}^{r} \mu_j a_j,\ \mu_j \geq 0,\ j = 1, \ldots, r \right\},$$

for some vectors $a_1, \ldots, a_r$ in $\Re^n$. The image of C under A is given by

$$AC = \{y \mid y = Ax,\ x \in C\} = \left\{ y \;\middle|\; y = \sum_{j=1}^{r} \mu_j A a_j,\ \mu_j \geq 0,\ j = 1, \ldots, r \right\},$$

showing that AC is a finitely generated cone in $\Re^m$. By the Minkowski-Weyl Theorem, the cone AC is polyhedral.

Let now K be a polyhedral cone in $\Re^m$ given by

$$K = \big\{ y \mid d_j' y \leq 0,\ j = 1, \ldots, r \big\},$$

for some vectors $d_1, \ldots, d_r$ in $\Re^m$. Then, the inverse image of K under A is

$$\begin{aligned} A^{-1} \cdot K &= \{x \mid Ax \in K\} \\ &= \big\{ x \mid d_j' A x \leq 0,\ j = 1, \ldots, r \big\} \\ &= \big\{ x \mid (A' d_j)' x \leq 0,\ j = 1, \ldots, r \big\}, \end{aligned}$$

showing that $A^{-1} \cdot K$ is a polyhedral cone in $\Re^n$.

(b) Let P be a polyhedral set in $\Re^n$ with Minkowski-Weyl Representation

$$P = \left\{ x \;\middle|\; x = \sum_{j=1}^{m} \mu_j v_j + y,\ \sum_{j=1}^{m} \mu_j = 1,\ \mu_j \geq 0,\ j = 1, \ldots, m,\ y \in C \right\},$$

where $v_1, \ldots, v_m$ are some vectors in $\Re^n$ and C is a finitely generated cone in $\Re^n$ (cf. Prop. 3.2.2). The image of P under A is given by

$$\begin{aligned} AP &= \{z \mid z = Ax,\ x \in P\} \\ &= \left\{ z \;\middle|\; z = \sum_{j=1}^{m} \mu_j A v_j + Ay,\ \sum_{j=1}^{m} \mu_j = 1,\ \mu_j \geq 0,\ j = 1, \ldots, m,\ Ay \in AC \right\}. \end{aligned}$$

By setting $Av_j = w_j$ and $Ay = u$, we obtain

$$\begin{aligned} AP &= \left\{ z \;\middle|\; z = \sum_{j=1}^{m} \mu_j w_j + u,\ \sum_{j=1}^{m} \mu_j = 1,\ \mu_j \geq 0,\ j = 1, \ldots, m,\ u \in AC \right\} \\ &= \operatorname{conv}\big(\{w_1, \ldots, w_m\}\big) + AC, \end{aligned}$$

where $w_1, \ldots, w_m \in \Re^m$. By part (a), the cone AC is polyhedral, implying by the Minkowski-Weyl Theorem [Prop. 3.2.1(c)] that AC is finitely generated. Hence, the set AP has a Minkowski-Weyl representation and therefore, it is polyhedral (cf. Prop. 3.2.2).

Let also Q be a polyhedral set in $\Re^m$ given by

$$Q = \big\{ y \mid d_j' y \le b_j,\ j = 1, \ldots, r \big\},$$

for some vectors $d_1, \ldots, d_r$ in $\Re^m$. Then, the inverse image of Q under A is

$$\begin{aligned} A^{-1} \cdot Q &= \{ x \mid Ax \in Q \} \\ &= \big\{ x \mid d_j' Ax \le b_j,\ j = 1, \ldots, r \big\} \\ &= \big\{ x \mid (A' d_j)' x \le b_j,\ j = 1, \ldots, r \big\}, \end{aligned}$$

showing that $A^{-1} \cdot Q$ is a polyhedral set in $\Re^n$.

3.10

It suffices to show the assertions for $m = 2$.

(a) Let C_1 and C_2 be polyhedral cones in $\Re^{n_1}$ and $\Re^{n_2}$, respectively, given by

$$C_1 = \big\{ x_1 \in \Re^{n_1} \mid \overline{a}_j' x_1 \le 0,\ j = 1, \ldots, r_1 \big\},$$

$$C_2 = \big\{ x_2 \in \Re^{n_2} \mid \tilde{a}_j' x_2 \le 0,\ j = 1, \ldots, r_2 \big\},$$

where $\overline{a}_1, \ldots, \overline{a}_{r_1}$ and $\tilde{a}_1, \ldots, \tilde{a}_{r_2}$ are some vectors in $\Re^{n_1}$ and $\Re^{n_2}$, respectively. Define

$$a_j = (\overline{a}_j, 0), \qquad \forall\ j = 1, \ldots, r_1,$$

$$a_j = (0, \tilde{a}_j), \qquad \forall\ j = r_1 + 1, \ldots, r_1 + r_2.$$

We have $(x_1, x_2) \in C_1 \times C_2$ if and only if

$$\overline{a}_j' x_1 \le 0, \qquad \forall\ j = 1, \ldots, r_1,$$

$$\tilde{a}_j' x_2 \le 0, \qquad \forall\ j = r_1 + 1, \ldots, r_1 + r_2,$$

or equivalently

$$a_j'(x_1, x_2) \le 0, \qquad \forall\ j = 1, \ldots, r_1 + r_2.$$

Therefore,

$$C_1 \times C_2 = \big\{ x \in \Re^{n_1 + n_2} \mid a_j' x \le 0,\ j = 1, \ldots, r_1 + r_2 \big\},$$

showing that $C_1 \times C_2$ is a polyhedral cone in $\Re^{n_1 + n_2}$.

(b) Let C_1 and C_2 be polyhedral cones in $\Re^n$. Then, straightforwardly from the definition of a polyhedral cone, it follows that the cone $C_1 \cap C_2$ is polyhedral.

By part (a), the Cartesian product $C_1 \times C_2$ is a polyhedral cone in $\Re^{n+n}$. Under the linear transformation A that maps $(x_1, x_2) \in \Re^{n+n}$ into $x_1 + x_2 \in \Re^n$, the image $A \cdot (C_1 \times C_2)$ is the set $C_1 + C_2$, which is a polyhedral cone by Exercise 3.9(a).

(c) Let P_1 and P_2 be polyhedral sets in $\Re^{n_1}$ and $\Re^{n_2}$, respectively, given by

$$P_1 = \big\{x_1 \in \Re^{n_1} \mid \overline{a}_j' x_1 \leq \overline{b}_j,\ j = 1, \ldots, r_1\big\},$$

$$P_2 = \big\{x_2 \in \Re^{n_2} \mid \tilde{a}_j' x_2 \leq \tilde{b}_j,\ j = 1, \ldots, r_2\big\},$$

where $\overline{a}_1, \ldots, \overline{a}_{r_1}$ and $\tilde{a}_1, \ldots, \tilde{a}_{r_2}$ are some vectors in $\Re^{n_1}$ and $\Re^{n_2}$, respectively, and $\overline{b}_j$ and $\tilde{b}_j$ are some scalars. By defining

$$a_j = (\overline{a}_j, 0), \qquad b_j = \overline{b}_j, \qquad \forall\ j = 1, \ldots, r_1,$$

$$a_j = (0, \tilde{a}_j), \qquad b_j = \tilde{b}_j, \qquad \forall\ j = r_1 + 1, \ldots, r_1 + r_2,$$

similar to the proof of part (a), we see that

$$P_1 \times P_2 = \big\{x \in \Re^{n_1 + n_2} \mid a_j' x \leq b_j,\ j = 1, \ldots, r_1 + r_2\big\},$$

showing that $P_1 \times P_2$ is a polyhedral set in $\Re^{n_1+n_2}$.

(d) Let P_1 and P_2 be polyhedral sets in $\Re^n$. Then, using the definition of a polyhedral set, it follows that the set $P_1 \cap P_2$ is polyhedral.

By part (c), the set $P_1 \times P_2$ is polyhedral. Furthermore, under the linear transformation A that maps $(x_1, x_2) \in \Re^{n+n}$ into $x_1 + x_2 \in \Re^n$, the image $A \cdot (P_1 \times P_2)$ is the set $P_1 + P_2$, which is polyhedral by Exercise 3.9(b).

3.11

We give two proofs. The first is based on the Minkowski-Weyl Representation of a polyhedral set P (cf. Prop. 3.2.2), while the second is based on a representation of P by a system of linear inequalities.

Let P be a polyhedral set with Minkowski-Weyl representation

$$P = \left\{x \;\middle|\; x = \sum_{j=1}^{m} \mu_j v_j + y,\ \sum_{j=1}^{m} \mu_j = 1,\ \mu_j \geq 0,\ j = 1, \ldots, m,\ y \in C\right\},$$

where $v_1, \ldots, v_m$ are some vectors in $\Re^n$ and C is a finitely generated cone in $\Re^n$. Let C be given by

$$C = \left\{y \;\middle|\; y = \sum_{i=1}^{r} \lambda_i a_i,\ \lambda_i \geq 0,\ i = 1, \ldots, r\right\},$$

where $a_1, \ldots, a_r$ are some vectors in $\Re^n$, so that

$$P = \left\{x \;\middle|\; x = \sum_{j=1}^{m} \mu_j v_j + \sum_{i=1}^{r} \lambda_i a_i,\ \sum_{j=1}^{m} \mu_j = 1,\ \mu_j \geq 0,\ \forall\ j,\ \lambda_i \geq 0,\ \forall\ i\right\}.$$

We claim that

$$\text{cone}(P) = \text{cone}\big(\{v_1, \ldots, v_m, a_1, \ldots, a_r\}\big).$$

Since $P \subset \text{cone}\big(\{v_1, \ldots, v_m, a_1, \ldots, a_r\}\big)$, it follows that

$$\text{cone}(P) \subset \text{cone}\big(\{v_1, \ldots, v_m, a_1, \ldots, a_r\}\big).$$

Conversely, let $y \in \text{cone}\big(\{v_1, \ldots, v_m, a_1, \ldots, a_r\}\big)$. Then, we have

$$y = \sum_{j=1}^{m} \overline{\mu}_j v_j + \sum_{i=1}^{r} \overline{\lambda}_i a_i,$$

with $\overline{\mu}_j \geq 0$ and $\overline{\lambda}_i \geq 0$ for all i and j. If $\overline{\mu}_j = 0$ for all j, then $y = \sum_{i=1}^{r} \overline{\lambda}_i a_i \in C$, and since $C = R_P$ (cf. Exercise 3.8), it follows that $y \in R_P$. Because the origin belongs to P and $y \in R_P$, we have $0 + y \in P$, implying that $y \in P$, and consequently $y \in \text{cone}(P)$. If $\overline{\mu}_j > 0$ for some j, then by setting $\overline{\mu} = \sum_{j=1}^{m} \overline{\mu}_j$, $\mu_j = \overline{\mu}_j / \overline{\mu}$ for all j, and $\lambda_i = \overline{\lambda}_i / \overline{\mu}$ for all i, we obtain

$$y = \overline{\mu} \left(\sum_{j=1}^{m} \mu_j v_j + \sum_{i=1}^{r} \lambda_i a_i \right),$$

where $\overline{\mu} > 0$, $\mu_j \geq 0$ with $\sum_{j=1}^{m} \mu_j = 1$, and $\lambda_i \geq 0$. Therefore $y = \overline{\mu}\, \overline{x}$ with $\overline{x} \in P$ and $\overline{\mu} > 0$, implying that $y \in \text{cone}(P)$ and showing that

$$\text{cone}\big(\{v_1, \ldots, v_m, a_1, \ldots, a_r\}\big) \subset \text{cone}(P).$$

We now give an alternative proof using the representation of P by a system of linear inequalities. Let P be given by

$$P = \big\{x \mid a_j' x \leq b_j,\ j = 1, \ldots, r\big\},$$

where $a_1, \ldots, a_r$ are vectors in $\Re^n$ and $b_1, \ldots, b_r$ are scalars. Since P contains the origin, it follows that $b_j \geq 0$ for all j. Define the index set J as follows

$$J = \{j \mid b_j = 0\}.$$

We consider separately the two cases where $J \neq \emptyset$ and $J = \emptyset$. If $J \neq \emptyset$, then we will show that

$$\text{cone}(P) = \big\{x \mid a_j' x \leq 0,\ j \in J\big\}.$$

To see this, note that since $P \subset \big\{x \mid a_j' x \leq 0,\ j \in J\big\}$, we have

$$\text{cone}(P) \subset \big\{x \mid a_j' x \leq 0,\ j \in J\big\}.$$

Conversely, let $\overline{x} \in \big\{x \mid a_j'x \le 0,\ j \in J\big\}$. We will show that $\overline{x} \in \text{cone}(P)$. If $\overline{x} \in P$, then $\overline{x} \in \text{cone}(P)$ and we are done, so assume that $\overline{x} \notin P$, implying that the set

$$\overline{J} = \{j \notin J \mid a_j'\overline{x} > b_j\} \tag{3.11}$$

is nonempty. By the definition of J, we have $b_j > 0$ for all $j \notin J$, so let

$$\mu = \min_{j \in \overline{J}} \frac{b_j}{a_j'\overline{x}},$$

and note that $0 < \mu < 1$. We have

$$a_j'(\mu\overline{x}) \le 0, \qquad \forall\ j \in J,$$

$$a_j'(\mu\overline{x}) \le b_j, \qquad \forall\ j \in \overline{J}.$$

For $j \notin \overline{J} \cup J$ and $a_j'\overline{x} \le 0 < b_j$, since $\mu > 0$, we still have $a_j'(\mu\overline{x}) \le 0 < b_j$. For $j \notin \overline{J} \cup J$ and $0 < a_j'\overline{x} \le b_j$, since $\mu < 1$, we have $0 < a_j'(\mu\overline{x}) < b_j$. Therefore, $\mu\overline{x} \in P$, implying that $\overline{x} = \frac{1}{\mu}(\mu\overline{x}) \in \text{cone}(P)$. It follows that

$$\big\{x \mid a_j'x \le 0,\ j \in J\big\} \subset \text{cone}(P),$$

and hence, $\text{cone}(P) = \big\{x \mid a_j'x \le 0,\ j \in J\big\}$.

If $J = \emptyset$, then we will show that $\text{cone}(P) = \Re^n$. To see this, take any $\overline{x} \in \Re^n$. If $\overline{x} \in P$, then clearly $\overline{x} \in \text{cone}(P)$, so assume that $\overline{x} \notin P$, implying that the set $\overline{J}$ as defined in Eq. (3.11) is nonempty. Note that $b_j > 0$ for all j, since J is empty. The rest of the proof is similar to the preceding case.

As an example, where cone(P) is not polyhedral when P does not contain the origin, consider the polyhedral set $P \subset \Re^2$ given by

$$P = \big\{(x_1, x_2) \mid x_1 \ge 0,\ x_2 = 1\big\}.$$

Then, we have

$$\text{cone}(P) = \big\{(x_1, x_2) \mid x_1 > 0,\ x_2 > 0\big\} \cup \big\{(x_1, x_2) \mid x_1 = 0, x_2 \ge 0\big\},$$

which is not closed and therefore not polyhedral.

3.12 (Properties of Polyhedral Functions)

(a) Let f_1 and f_2 be polyhedral functions such that $\text{dom}(f_1) \cap \text{dom}(f_2) \ne \emptyset$. By Prop. 3.2.3, $\text{dom}(f_1)$ and $\text{dom}(f_2)$ are polyhedral sets in $\Re^n$, and

$$f_1(x) = \max\big\{a_1'x + b_1, \ldots, a_m'x + b_m\big\}, \qquad \forall\ x \in \text{dom}(f_1),$$

$$f_2(x) = \max\big\{\overline{a}_1'x + \overline{b}_1, \ldots, \overline{a}_m'x + \overline{b}_{\overline{m}}\big\}, \qquad \forall\ x \in \text{dom}(f_2),$$

where a_i and $\overline{a}_i$ are vectors in $\Re^n$, and b_i and $\overline{b}_i$ are scalars. The domain of f_1+f_2 coincides with $\mathrm{dom}(f_1)\cap\mathrm{dom}(f_2)$, which is polyhedral by Exercise 3.10(d). Furthermore, we have for all $x \in \mathrm{dom}(f_1 + f_2)$,

$$\begin{aligned} f_1(x) + f_2(x) &= \max\Big\{a_1'x + b_1, \ldots, a_m'x + b_m\Big\} + \max\Big\{\overline{a}_1'x + \overline{b}_1, \ldots, \overline{a}_{\overline{m}}'x + \overline{b}_{\overline{m}}\Big\} \\ &= \max_{1\le i\le m,\ 1\le j\le \overline{m}} \Big\{a_i'x + b_i + \overline{a}_j'x + \overline{b}_j\Big\} \\ &= \max_{1\le i\le m,\ 1\le j\le \overline{m}} \Big\{(a_i + \overline{a}_j)'x + (b_i + \overline{b}_j)\Big\}. \end{aligned}$$

Therefore, by Prop. 3.2.3, the function $f_1 + f_2$ is polyhedral.

(b) Since $g : \Re^m \mapsto (-\infty, \infty]$ is a polyhedral function, by Prop. 3.2.3, $\mathrm{dom}(g)$ is a polyhedral set in $\Re^m$ and g is given by

$$g(y) = \max\Big\{a_1'y + b_1, \ldots, a_m'y + b_m\Big\}, \qquad \forall\ y \in \mathrm{dom}(g),$$

for some vectors a_i in $\Re^m$ and scalars b_i. The domain of f can be expressed as

$$\mathrm{dom}(f) = \Big\{x \mid f(x) < \infty\Big\} = \Big\{x \mid g(Ax) < \infty\Big\} = \Big\{x \mid Ax \in \mathrm{dom}(g)\Big\}.$$

Thus, $\mathrm{dom}(f)$ is the inverse image of the polyhedral set $\mathrm{dom}(g)$ under the linear transformation A. By the assumption that $\mathrm{dom}(g)$ contains a point in the range of A, it follows that $\mathrm{dom}(f)$ is nonempty, while by Exercise 3.9(b), the set $\mathrm{dom}(f)$ is polyhedral. Furthermore, for all $x \in \mathrm{dom}(f)$, we have

$$\begin{aligned} f(x) &= g(Ax) \\ &= \max\Big\{a_1'Ax + b_1, \ldots, a_m'Ax + b_m\Big\} \\ &= \max\Big\{(A'a_1)'x + b_1, \ldots, (A'a_m)'x + b_m\Big\}. \end{aligned}$$

Thus, by Prop. 3.2.3, it follows that the function f is polyhedral.

3.13 (Partial Minimization of Polyhedral Functions)

As shown at the end of Section 2.3, we have

$$P\big(\mathrm{epi}(F)\big) \subset \mathrm{epi}(f) \subset \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big).$$

Since the function F is polyhedral, its epigraph

$$\mathrm{epi}(F) = \Big\{(x, z, w) \mid F(x, z) \le w,\ (x, w) \in \mathrm{dom}(F)\Big\}$$

is a polyhedral set in $\Re^{n+m+1}$. The set $P\big(\mathrm{epi}(F)\big)$ is the image of the polyhedral set $\mathrm{epi}(F)$ under the linear transformation P, and therefore, by Exercise 3.9(b), the set $P\big(\mathrm{epi}(F)\big)$ is polyhedral. Furthermore, a polyhedral set is always closed, and hence

$$P\big(\mathrm{epi}(F)\big) = \mathrm{cl}\Big(P\big(\mathrm{epi}(F)\big)\Big).$$

The preceding two relations yield

$$\mathrm{epi}(f) = P\big(\mathrm{epi}(F)\big),$$

implying that the function f is polyhedral.

3.14 (Existence of Minima of Polyhedral Functions)

If the set of minima of f over P is nonempty, then evidently $\inf_{x\in P} f(x)$ must be finite.

Conversely, suppose that $\inf_{x\in P} f(x)$ is finite. Since f is a polyhedral function, by Prop. 3.2.3, we have

$$f(x) = \max\bigl\{a_1'x + b_1, \ldots, a_m'x + b_m\bigr\}, \qquad \forall\, x \in \mathrm{dom}(f),$$

where $\mathrm{dom}(f)$ is a polyhedral set. Therefore,

$$\inf_{x\in P} f(x) = \inf_{x\in P\cap\mathrm{dom}(f)} f(x) = \inf_{x\in P\cap\mathrm{dom}(f)} \max\bigl\{a_1'x + b_1, \ldots, a_m'x + b_m\bigr\}.$$

Let $\overline{P} = P\cap\mathrm{dom}(f)$ and note that $\overline{P}$ is nonempty by assumption. Since $\overline{P}$ is the intersection of the polyhedral sets P and $\mathrm{dom}(f)$, the set $\overline{P}$ is polyhedral. The problem

$$\begin{aligned} &\text{minimize} \quad \max\bigl\{a_1'x + b_1, \ldots, a_m'x + b_m\bigr\} \\ &\text{subject to} \ \ x \in \overline{P} \end{aligned}$$

is equivalent to the following linear program

$$\begin{aligned} &\text{minimize} \quad y \\ &\text{subject to} \ \ a_j'x + b_j \le y,\ j = 1, \ldots, m,\ x \in \overline{P},\ y \in \Re. \end{aligned}$$

By introducing the variable $z = (x, y) \in \Re^{n+1}$, the vector $c = (0, \ldots, 0, 1) \in \Re^{n+1}$, and the set

$$\hat{P} = \bigl\{(x, y) \mid a_j'x + b_j \le y,\ j = 1, \ldots, m,\ x \in \overline{P},\ y \in \Re\bigr\},$$

we see that the original problem is equivalent to

$$\begin{aligned} &\text{minimize} \quad c'z \\ &\text{subject to} \ \ z \in \hat{P}, \end{aligned}$$

where $\hat{P}$ is polyhedral ($\hat{P} \ne \emptyset$ since $\overline{P} \ne \emptyset$). Furthermore, because $\inf_{x\in P} f(x)$ is finite, it follows that $\inf_{z\in\hat{P}} c'z$ is also finite. Thus, by Prop. 2.3.4 of Chapter 2, the set Z^* of minimizers of $c'z$ over $\hat{P}$ is nonempty, and the nonempty set $\bigl\{x \mid z = (x, y),\ z \in Z^*\bigr\}$ is the set of minimizers of f over P.

3.15 (Existence of Solutions of Quadratic Nonconvex Programs [FrW56])

We use induction on the dimension of the set X. Suppose that the dimension of X is 0. Then, X consists of a single point, which is the global minimum of f over X.

Assume that, for some $l < n$, f attains its minimum over every set $\overline{X}$ of dimension less than or equal to l that is specified by linear inequality constraints, and is such that f is bounded over $\overline{X}$. Let X be of the form

$$X = \{x \mid a_j'x \leq b_j,\ j = 1, \ldots, r\},$$

have dimension $l + 1$, and be such that f is bounded over X. We will show that f attains its minimum over X.

If X is a bounded polyhedral set, f attains a minimum over X by Weierstrass' Theorem. We thus assume that X is unbounded. Using the the Minkowski-Weyl representation, we can write X as

$$X = \{x \mid x = v + \alpha y,\ v \in V,\ y \in C,\ \alpha \geq 0\},$$

where V is the convex hull of finitely many vectors and C is the intersection of a finitely generated cone with the surface of the unit sphere $\{x \mid \|x\| = 1\}$. Then, for any $x \in X$ and $y \in C$, the vector $x + \alpha y$ belongs to X for every positive scalar α and

$$f(x + \alpha y) = f(x) + \alpha(c' + x'Q)y + \alpha^2 y'Qy.$$

In view of the assumption that f is bounded over X, this implies that $y'Qy \geq 0$ for all $y \in C$.

If $y'Qy > 0$ for all $y \in C$, then, since C and V are compact, there exist some $\delta > 0$ and $\gamma > 0$ such that $y'Qy > \delta$ for all $y \in C$, and $(c' + v'Q)y > -\gamma$ for all $v \in V$ and $y \in C$. It follows that for all $v \in V$, $y \in C$, and $\alpha \geq \gamma/\delta$, we have

$$\begin{aligned} f(v + \alpha y) &= f(v) + \alpha(c' + v'Q)y + \alpha^2 y'Qy \\ &> f(v) + \alpha(-\gamma + \alpha\delta) \\ &\geq f(v), \end{aligned}$$

which implies that

$$\inf_{x \in X} f(x) = \inf_{\substack{x \in (V + \alpha C) \\ 0 \leq \alpha \leq \frac{\gamma}{\delta}}} f(x).$$

Since the minimization in the right hand side is over a compact set, it follows from Weierstrass' Theorem and the preceding relation that the minimum of f over X is attained.

Next, assume that there exists some $\overline{y} \in C$ such that $\overline{y}'Q\overline{y} = 0$. From Exercise 3.8, it follows that $\overline{y}$ belongs to the recession cone of X, denoted by R_X. If $\overline{y}$ is in the lineality space of X, denoted by L_X, the vector $x + \alpha\overline{y}$ belongs to X for every $x \in X$ and every scalar α, and we have

$$f(x + \alpha\overline{y}) = f(x) + \alpha(c' + x'Q)\overline{y}.$$

This relation together with the boundedness of f over X implies that

$$(c' + x'Q)\overline{y} = 0, \qquad \forall\ x \in X. \tag{3.12}$$

Let $S = \{\gamma\overline{y} \mid \gamma \in \Re\}$ be the subspace generated by $\overline{y}$ and consider the following decomposition of X:

$$X = S + (X \cap S^{\perp}),$$

(cf. Prop. 1.5.4). Then, we can write any $x \in X$ as $x = z + \alpha\overline{y}$ for some $z \in X \cap S^{\perp}$ and some scalar α, and it follows from Eq. (3.12) that $f(x) = f(z)$, which implies that

$$\inf_{x \in X} f(x) = \inf_{x \in X \cap S^{\perp}} f(x).$$

It can be seen that the dimension of set $X \cap S^{\perp}$ is smaller than the dimension of set X. To see this, note that $S^{\perp}$ contains the subspace parallel to the affine hull of $X \cap S^{\perp}$. Therefore, $\overline{y}$ does not belong to the subspace parallel to the affine hull of $X \cap S^{\perp}$. On the other hand, $\overline{y}$ belongs to the subspace parallel to the affine hull of X, hence showing that the dimension of set $X \cap S^{\perp}$ is smaller than the dimension of set X. Since $X \cap S^{\perp} \subset X$, f is bounded over $X \cap S^{\perp}$, so by using the induction hypothesis, it follows that f attains its minimum over $X \cap S^{\perp}$, which, in view of the preceding relation, is also the minimum of f over X.

Finally, assume that $\overline{y}$ is not in L_X, i.e., $\overline{y} \in R_X$, but $-\overline{y} \notin R_X$. The recession cone of X is of the form

$$R_X = \{y \mid a_j'y \leq 0,\ j = 1, \ldots, r\}.$$

Since $\overline{y} \in R_X$, we have

$$a_j'\overline{y} \leq 0, \qquad \forall\ j = 1, \ldots, r,$$

and since $-\overline{y} \notin R_X$, the index set

$$J = \{j \mid a_j'\overline{y} < 0\}$$

is nonempty.

Let $\{x_k\}$ be a minimizing sequence, i.e.,

$$f(x_k) \to f^*,$$

where $f^* = \inf_{x\ inX} f(x)$. Suppose that for each k, we start at x_k and move along $-\overline{y}$ as far as possible without leaving the set X, up to the point where we encounter the vector

$$\overline{x}_k = x_k - \beta_k\overline{y},$$

where β_k is the nonnegative scalar given by

$$\beta_k = \min_{j \in J} \frac{a_j'x_k - b_j}{a_j'\overline{y}}.$$

Since $\overline{y} \in R_X$ and f is bounded over X, we have $(c' + x'Q)\overline{y} \geq 0$ for all $x \in X$, which implies that

$$f(\overline{x}_k) \leq f(x_k), \qquad \forall\ k. \tag{3.13}$$

By construction of the sequence $\{\overline{x}_k\}$, it follows that there exists some $j_0 \in J$ such that $a_{j_0}'\overline{x}_k = b_{j_0}$ for all k in an infinite index set $\mathcal{K} \subset \{0, 1, \ldots\}$. By reordering the linear inequalities if necessary, we can assume that $j_0 = 1$, i.e.,

$$a_1'\overline{x}_k = b_1, \qquad \forall\ k \in \mathcal{K}.$$

To apply the induction hypothesis, consider the set

$$\overline{X} = \{x \mid a_1'x = b_1,\ a_j'x \leq b_j,\ j = 2, \ldots, r\},$$

and note that $\{\overline{x}_k\}_{\mathcal{K}} \subset \overline{X}$. The dimension of $\overline{X}$ is smaller than the dimension of X. To see this, note that the set $\{x \mid a_1'x = b_1\}$ contains $\overline{X}$, so that a_1 is orthogonal to the subspace $S_{\overline{X}}$ that is parallel to $\text{aff}(\overline{X})$. Since $a_1'\overline{y} < 0$, it follows that $\overline{y} \notin S_{\overline{X}}$. On the other hand, $\overline{y}$ belongs to S_X, the subspace that is parallel to $\text{aff}(X)$, since for all k, we have $x_k \in X$ and $x_k - \beta_k\overline{y} \in X$.

Since $\overline{X} \subset X$, f is also bounded over $\overline{X}$, so it follows from the induction hypothesis that f attains its minimum over $\overline{X}$ at some x^*. Because $\{\overline{x}_k\}_{\mathcal{K}} \subset \overline{X}$, and using also Eq. (3.13), we have

$$f(x^*) \leq f(\overline{x}_k) \leq f(x_k), \qquad \forall\ k \in \mathcal{K}.$$

Since $f(x_k) \to f^*$, we obtain

$$f(x^*) \leq \lim_{k \to \infty,\ k \in \mathcal{K}} f(x_k) = f^*,$$

and since $x^* \in \overline{X} \subset X$, this implies that f attains the minimum over X at x^*, concluding the proof.

3.16

Assume that P has an extreme point, say v. Then, by Prop. 3.3.3(a), the set

$$A_v = \Big\{a_j \mid a_j'v = b_j,\ j \in \{1, \ldots, r\}\Big\}$$

contains n linearly independent vectors, so the set of vectors $\{a_j \mid j = 1, \ldots, r\}$ contains a subset of n linearly independent vectors.

Assume now that the set $\{a_j \mid j = 1, \ldots, r\}$ contains a subset of n linearly independent vectors. Suppose, to obtain a contradiction, that P does not have any extreme points. Then, by Prop. 3.3.1, P contains a line

$$L = \{x + \lambda d \mid \lambda \in \Re\},$$

where $x \in P$ and $d \in \Re^n$ is a nonzero vector. Since $L \subset P$, it follows that $a_j'd = 0$ for all $j = 1, \ldots, r$. Since $d \neq 0$, this implies that the set $\{a_1, \ldots, a_r\}$ cannot contain a subset of n linearly independent vectors, a contradiction.

3.17

Suppose that x is not an extreme point of C. Then $x = \alpha x_1 + (1-\alpha)x_2$ for some $x_1, x_2 \in C$ with $x_1 \neq x$ and $x_2 \neq x$, and a scalar $\alpha \in (0,1)$, so that $Ax = \alpha Ax_1 + (1-\alpha)Ax_2$. Since the columns of A are linearly independent, we have $Ay_1 = Ay_2$ if and only if $y_1 = y_2$. Therefore, $Ax_1 \neq Ax$ and $Ax_2 \neq Ax$, implying that Ax is a convex combination of two distinct points in AC, i.e., Ax is not an extreme point of AC.

Suppose now that Ax is not an extreme point of AC, so that $Ax = \alpha Ax_1 + (1-\alpha)Ax_2$ for some $x_1, x_2 \in C$ with $Ax_1 \neq Ax$ and $Ax_2 \neq Ax$, and a scalar $\alpha \in (0,1)$. Then, $A\big(x - \alpha x_1 - (1-\alpha)x_2\big) = 0$ and since the columns of A are linearly independent, it follows that $x = \alpha x_1 - (1-\alpha)x_2$. Furthermore, because $Ax_1 \neq Ax$ and $Ax_2 \neq Ax$, we must have $x_1 \neq x$ and $x_2 \neq x$, implying that x is not an extreme point of C.

As an example showing that if the columns of A are linearly dependent, then Ax can be an extreme point of AC, for some non-extreme point x of C, consider the 1×2 matrix $A = [1\ 0]$, whose columns are linearly dependent. The polyhedral set C given by

$$C = \big\{(x_1, x_2) \mid x_1 \geq 0,\ 0 \leq x_2 \leq 1\big\}$$

has two extreme points, (0,0) and (0,1). Its image $AC \subset \Re$ is given by

$$AC = \{x_1 \mid x_1 \geq 0\},$$

whose unique extreme point is $x_1 = 0$. The point $x = (0, 1/2) \in C$ is not an extreme point of C, while its image $Ax = 0$ is an extreme point of AC. Actually, all the points in C on the line segment connecting (0,0) and (0,1), except for (0,0) and (0,1), are non-extreme points of C that are mapped under A into the extreme point 0 of AC.

3.18

For the sets C_1 and C_2 as given in this exercise, the set $C_1 \cup C_2$ is compact, and its convex hull is also compact by Prop. 1.3.2 of Chapter 1. The set of extreme points of $\mathrm{conv}(C_1 \cup C_2)$ is not closed, since it consists of the two end points of the line segment C_1, namely $(0,0,-1)$ and $(0,0,1)$, and all the points $x = (x_1, x_2, x_3)$ such that

$$x \neq 0, \qquad (x_1 - 1)^2 + x_2^2 = 1, \qquad x_3 = 0.$$

3.19

By Prop. 3.3.2, a polyhedral set has a finite number of extreme points. Conversely, let P be a compact convex set having a finite number of extreme points $\{v_1, \ldots, v_m\}$. By the Krein-Milman Theorem (Prop. 3.3.1), a compact convex set is equal to the convex hull of its extreme points, so that $P = \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big)$, which is a polyhedral set by the Minkowski-Weyl Representation Theorem (Prop. 3.2.2).

As an example showing that the assertion fails if compactness of the set is replaced by a weaker assumption that the set is closed and contains no lines, consider the set $D \subset \Re^3$ given by

$$D = \{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \le 1,\ x_3 = 1\}.$$

Let $C = \mathrm{cone}(D)$. It can seen that C is not a polyhedral set. On the other hand, C is closed, convex, does not contain a line, and has a unique extreme point at the origin.

[For a more formal argument, note that if C were polyhedral, then the set

$$D = C \cap \{(x_1, x_2, x_3) \mid x_3 = 1\}$$

would also be polyhedral by Exercise 3.10(d), since both C and $\{(x_1, x_2, x_3) \mid x_3 = 1\}$ are polyhedral sets. Thus, by Prop. 3.2.2, it would follow that D has a finite number of extreme points. But this is a contradiction because the set of extreme points of D coincides with $\{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 = 1,\ x_3 = 1\}$, which contains an infinite number of points. Thus, C is not a polyhedral cone, and therefore not a polyhedral set, while C is closed, convex, does not contain a line, and has a unique extreme point at the origin.]

3.20 (Faces)

(a) Let P be a polyhedral set in $\Re^n$, and let $F = P \cap H$ be a face of P, where H is a hyperplane passing through some boundary point $\overline{x}$ of P and containing P in one of its halfspaces. Then H is given by $H = \{x \mid a'x = a'\overline{x}\}$ for some nonzero vector $a \in \Re^n$. By replacing $a'x = a'\overline{x}$ with two inequalities $a'x \le a'\overline{x}$ and $-a'x \le -a'\overline{x}$, we see that H is a polyhedral set in $\Re^n$. Since the intersection of two nondisjoint polyhedral sets is a polyhedral set [cf. Exercise 3.10(d)], the set $F = P \cap H$ is polyhedral.

(b) Let P be given by

$$P = \{x \mid a_j'x \le b_j,\ j = 1, \ldots, r\},$$

for some vectors $a_j \in \Re^n$ and scalars b_j. Let v be an extreme point of P, and without loss of generality assume that the first n inequalities define v, i.e., the first n of the vectors a_j are linearly independent and such that

$$a_j'v = b_j, \qquad \forall\ j = 1, \ldots, n$$

[cf. Prop. 3.3.3(a)]. Define the vector $a \in \Re^n$, the scalar b, and the hyperplane H as follows

$$a = \frac{1}{n}\sum_{j=1}^{n} a_j, \qquad b = \frac{1}{n}\sum_{j=1}^{n} b_j, \qquad H = \{x \mid a'x = b\}.$$

Then, we have

$$a'v = b,$$

so that H passes through v. Moreover, for every $x \in P$, we have $a_j'x \leq b_j$ for all j, implying that $a'x \leq b$ for all $x \in P$. Thus, H contains P in one of its halfspaces.

We will next prove that $P \cap H = \{v\}$. We start by showing that for every $\overline{v} \in P \cap H$, we must have

$$a_j'\overline{v} = b_j, \qquad \forall\ j = 1, \ldots, n. \tag{3.14}$$

To arrive at a contradiction, assume that $a_j'\overline{v} < b_j$ for some $\overline{v} \in P \cap H$ and $j \in \{1, \ldots, n\}$. Without loss of generality, we can assume that the strict inequality holds for $j = 1$, so that

$$a_1'\overline{v} < b_1, \qquad a_j'\overline{v} \leq b_j, \qquad \forall\ j = 2, \ldots, n.$$

By multiplying each of the above inequalities with $1/n$ and by summing the obtained inequalities, we obtain

$$\frac{1}{n}\sum_{j=1}^{n} a_j'\overline{v} < \frac{1}{n}\sum_{j=1}^{n} b_j,$$

implying that $a'\overline{v} < b$, which contradicts the fact that $\overline{v} \in H$. Hence, Eq. (3.14) holds, and since the vectors $a_1, \ldots, a_n$ are linearly independent, it follows that $v = \overline{v}$, showing that $P \cap H = \{v\}$.

As discussed in Section 3.3, every extreme point of P is a relative boundary point of P. Since every relative boundary point of P is also a boundary point of P, it follows that every extreme point of P is a boundary point of P. Thus, v is a boundary point of P, and as shown earlier, H passes through v and contains P in one of its halfspaces. By definition, it follows that $P \cap H = \{v\}$ is a face of P.

(c) Since P is not an affine set, it cannot consist of a single point, so we must have $\dim(P) > 0$. Let P be given by

$$P = \left\{x \mid a_j'x \leq b_j,\ j = 1, \ldots, r\right\},$$

for some vectors $a_j \in \Re^n$ and scalars b_j. Also, let A be the matrix with rows a_j' and b be the vector with components b_j, so that

$$P = \{x \mid Ax \leq b\}.$$

An inequality $a_j'x \leq b_j$ of the system $Ax \leq b$ is *redundant* if it is implied by the remaining inequalities in the system. If the system $Ax \leq b$ has no redundant inequalities, we say that the system is *nonredundant*. An inequality $a_j'x \leq b_j$ of the system $Ax \leq b$ is an *implicit equality* if $a_j'x = b_j$ for all x satisfying $Ax \leq b$.

By removing the redundant inequalities if necessary, we may assume that the system $Ax \leq b$ defining P is nonredundant. Since P is not an affine set, there exists an inequality $a_{j_0}'x \leq b_{j_0}$ that is not an implicit equality of the system $Ax \leq b$. Consider the set

$$F = \{x \in P \mid a_{j_0}'x = b_{j_0}\}.$$

Note that $F \neq \emptyset$, since otherwise $a_{j_0}'x \leq b_{j_0}$ would be a redundant inequality of the system $Ax \leq b$, contradicting our earlier assumption that the system is nonredundant. Note also that every point of F is a boundary point of P. Thus, F is the intersection of P and the hyperplane $\{x \mid a_{j_0}'x = b_{j_0}\}$ that passes through a boundary point of P and contains P in one of its halfspaces, i.e., F is a face of P. Since $a_{j_0}'x \leq b_{j_0}$ is not an implicit equality of the system $Ax \leq b$, the dimension of F is $\dim(P) - 1$.

(d) Let P be a polyhedral set given by

$$P = \{x \mid a_j'x \leq b_j,\ j = 1, \ldots, r\},$$

with $a_j \in \Re^n$ and $b_j \in \Re$, or equivalently

$$P = \{x \mid Ax \leq b\},$$

where A is an $r \times n$ matrix and $b \in \Re^r$. We will show that F is a face of P if and only if F is nonempty and

$$F = \{x \in P \mid a_j'x = b_j,\ j \in J\},$$

where $J \subset \{1, \ldots, r\}$. From this it will follow that the number of distinct faces of P is finite.

By removing the redundant inequalities if necessary, we may assume that the system $Ax \leq b$ defining P is nonredundant. Let F be a face of P, so that $F = P \cap H$, where H is a hyperplane that passes through a boundary point of P and contains P in one of its halfspaces. Let $H = \{x \mid c'x = c\overline{x}\}$ for a nonzero vector $c \in \Re^n$ and a boundary point $\overline{x}$ of P, so that

$$F = \{x \in P \mid c'x = c\overline{x}\}$$

and

$$c'x \leq c\overline{x}, \qquad \forall\ x \in P.$$

These relations imply that the set of points x such that $Ax \leq b$ and $c'x \leq c\overline{x}$ coincides with P, and since the system $Ax \leq b$ is nonredundant, it follows that $c'x \leq c\overline{x}$ is a redundant inequality of the system $Ax \leq b$ and $c'x \leq c\overline{x}$. Therefore, the inequality $c'x \leq c\overline{x}$ is implied by the inequalities of $Ax \leq b$, so that there exists some $\mu \in \Re^r$ with $\mu \geq 0$ such that

$$\sum_{j=1}^{r} \mu_j a_j = c, \qquad \sum_{j=1}^{r} \mu_j b_j = c'\overline{x}.$$

Let $J = \{j \mid \mu_j > 0\}$. Then, for every $x \in P$, we have

$$c'x = c\overline{x} \quad \Longleftrightarrow \quad \sum_{j \in J} \mu_j a_j'x = \sum_{j \in J} \mu_j b_j \quad \Longleftrightarrow \quad a_j'x = b_j,\ j \in J, \qquad (3.15)$$

implying that

$$F = \big\{x \in P \mid a_j'x = b_j, j \in J\big\}.$$

Conversely, let F be a nonempty set given by

$$F = \big\{x \in P \mid a_j'x = b_j, j \in J\big\},$$

for some $J \subset \{1, \ldots, r\}$. Define

$$c = \sum_{j \in J} a_j, \qquad \beta = \sum_{j \in J} b_j.$$

Then, we have

$$\big\{x \in P \mid a_j'x = b_j, j \in J\big\} = \big\{x \in P \mid c'x = \beta\big\},$$

[cf. Eq. (3.15) where $\mu_j = 1$ for all $j \in J$]. Let $H = \big\{x \mid c'x = \beta\big\}$, so that in view of the preceding relation, we have that $F = P \cap H$. Since every point of F is a boundary point of P, it follows that H passes through a boundary point of P. Furthermore, for every $x \in P$, we have $a_j'x \leq b_j$ for all $j \in J$, implying that $c'x \leq \beta$ for every $x \in P$. Thus, H contains P in one of its halfspaces. Hence, F is a face.

3.21 (Isomorphic Polyhedral Sets)

(a) Let P and Q be isomorhic polyhedral sets, and let $f : P \mapsto Q$ and $g : Q \mapsto P$ be affine functions such that

$$x = g\big(f(x)\big), \quad \forall\, x \in P, \qquad y = f\big(g(y)\big), \quad \forall\, y \in Q.$$

Assume that x^* is an extreme point of P and let $y^* = f(x^*)$. We will show that y^* is an extreme point of Q. Since x^* is an extreme point of P, by Exercise 3.20(b), it is also a face of P, and therefore, there exists a vector $c \in \Re^n$ such that

$$c'x < c'x^*, \qquad \forall\, x \in P,\ x \neq x^*.$$

For any $y \in Q$ with $y \neq y^*$, we have

$$f\big(g(y)\big) = y \neq y^* = f(x^*),$$

implying that

$$g(y) \neq g(y^*) = x^*, \qquad \text{with } \ g(y) \in P.$$

Hence,

$$c'g(y) < c'g(y^*), \qquad \forall\, y \in Q,\ y \neq y^*.$$

Let the affine function g be given by $g(y) = By + d$ for some $n \times m$ matrix B and vector $d \in \Re^n$. Then, we have

$$c'(By + d) < c'(By^* + d), \qquad \forall\, y \in Q,\ y \neq y^*,$$

implying that

$$(B'c)'y < (B'c)'y^*, \qquad \forall\, y \in Q,\ y \neq y^*.$$

If y^* were not an extreme point of Q, then we would have $y^* = \alpha y_1 + (1-\alpha)y_2$ for some distinct points $y_1, y_2 \in Q$, $y_1 \neq y^*$, $y_2 \neq y^*$, and $\alpha \in (0,1)$, so that

$$(B'c)'y^* = \alpha(B'c)'y_1 + (1-\alpha)(B'c)'y_2 < (B'c)'y^*,$$

which is a contradiction. Hence, y^* is an extreme point of Q.

Conversely, if y^* is an extreme point of Q, then by using a symmetrical argument, we can show that x^* is an extreme point of P.

(b) For the sets

$$P = \{x \in \Re^n \mid Ax \leq b,\ x \geq 0\},$$

$$Q = \big\{(x,z) \in \Re^{n+r} \mid Ax + z = b,\ x \geq 0,\ z \geq 0\big\},$$

let f and g be given by

$$f(x) = (x, b - Ax), \qquad \forall\, x \in P,$$

$$g(x,z) = x, \qquad \forall\, (x,z) \in Q.$$

Evidently, f and g are affine functions. Furthermore, clearly

$$f(x) \in Q, \qquad g\big(f(x)\big) = x, \qquad \forall\, x \in P,$$

$$g(x,z) \in P, \qquad f\big(g(x,z)\big) = x, \qquad \forall\, (x,z) \in Q.$$

Hence, P and Q are isomorphic.

3.22 (Unimodularity I)

Suppose that the system $Ax = b$ has integer components for every vector $b \in \Re^n$ with integer components. Since A is invertible, it follows that the vector $A^{-1}b$ has integer components for every $b \in \Re^n$ with integer components. For $i = 1, \ldots, n$, let e_i be the vector with ith component equal to 1 and all other components equal to 0. Then, for $b = e_i$, the vectors $A^{-1}e_i$, $i = 1, \ldots, n$, have integer components, implying that the columns of A^{-1} are vectors with integer components, so that A^{-1} has integer entries. Therefore, $\det(A^{-1})$ is integer, and since $\det(A)$ is also integer and $\det(A) \cdot \det(A^{-1}) = 1$, it follows that either $\det(A) = 1$ or $\det(A) = -1$, showing that A is unimodular.

Suppose now that A is unimodular. Take any vector $b \in \Re^n$ with integer components, and for each $i \in \{1, \ldots, n\}$, let A_i be the matrix obtained from A by replacing the ith column of A with b. Then, according to Cramer's rule, the components of the solution $\hat{x}$ of the system $Ax = b$ are given by

$$\hat{x}_i = \frac{\det(A_i)}{\det(A)}, \qquad i = 1, \ldots, n.$$

Since each matrix A_i has integer entries, it follows that $\det(A_i)$ is integer for all $i = 1, \ldots, n$. Furthermore, because A is invertible and unimodular, we have either $\det(A) = 1$ or $\det(A) = -1$, implying that the vector $\hat{x}$ has integer components.

3.23 (Unimodularity II)

(a) The proof is straightforward from the definition of the totally unimodular matrix and the fact that B is a submatrix of A if and only if B' is a submatrix of A'.

(b) Suppose that A is totally unimodular. Let J be a subset of $\{1,\ldots,n\}$. Define z by $z_j = 1$ if $j \in J$, and $z_j = 0$ otherwise. Also let $w = Az$, $c_i = d_i = \frac{1}{2}w_i$ if w_i is even, and $c_i = \frac{1}{2}(w_i - 1)$ and $d_i = \frac{1}{2}(w_i + 1)$ if w_i is odd. Consider the polyhedral set

$$P = \{x \mid c \le Ax \le d,\ 0 \le x \le z\},$$

and note that $P \neq \emptyset$ because $\frac{1}{2}z \in P$. Since A is totally unimodular, the polyhedron P has integer extreme points. Let $\hat{x} \in P$ be one of them. Because $0 \le \hat{x} \le z$ and $\hat{x}$ has integer components, it follows that $\hat{x}_j = 0$ for $j \notin J$ and $\hat{x}_j \in \{0,1\}$ for $j \in J$. Therefore, $z_j - 2\hat{x}_j = \pm 1$ for $j \in J$. Define $J_1 = \{j \in J \mid z_j - 2\hat{x}_j = 1\}$ and $J_2 = \{j \in J \mid z_j - 2\hat{x}_j = -1\}$. We have

$$\begin{aligned}
\sum_{j \in J_1} a_{ij} - \sum_{j \in J_2} a_{ij} &= \sum_{j \in J} a_{ij}(z_j - 2\hat{x}_j) \\
&= \sum_{j=1}^{n} a_{ij}(z_j - 2\hat{x}_j) \\
&= [Az]_i - 2[A\hat{x}]_i \\
&= w_i - 2[A\hat{x}]_i,
\end{aligned}$$

where $[Ax]_i$ denotes the ith component of the vector Ax. If w_i is even, then since $c_i \le [A\hat{x}]_i \le d_i$ and $c_i = d_i = \frac{1}{2}w_i$, it follows that $[A\hat{x}]_i = w_i$, so that

$$w_i - 2[A\hat{x}]_i = 0, \qquad \text{when } w_i \text{ is even.}$$

If w_i is odd, then since $c_i \le [A\hat{x}]_i \le d_i$, $c_i = \frac{1}{2}(w_i - 1)$, and $d_i = \frac{1}{2}(w_i + 1)$, it follows that

$$\frac{1}{2}(w_i - 1) \le [A\hat{x}]_i \le \frac{1}{2}(w_i + 1),$$

implying that

$$-1 \le w_i - 2[A\hat{x}]_i \le 1.$$

Because $w_i - 2[A\hat{x}]_i$ is integer, we conclude that

$$w_i - 2[A\hat{x}]_i \in \{-1, 0, 1\}, \qquad \text{when } w_i \text{ is odd.}$$

Therefore,

$$\left| \sum_{j \in J_1} a_{ij} - \sum_{j \in J_2} a_{ij} \right| \le 1, \quad \forall\ i = 1,\ldots,m. \tag{3.16}$$

Suppose now that the matrix A is such that any $J \subset \{1,\ldots,n\}$ can be partitioned into two subsets so that Eq. (3.16) holds. We prove that A is totally unimodular, by showing that each of its square submatrices is unimodular, i.e., the determinant of every square submatrix of A is -1, 0, or 1. We use induction on the size of the square submatrices of A.

To start the induction, note that for $J \subset \{1, \ldots, n\}$ with J consisting of a single element, from Eq. (3.16) we obtain $a_{ij} \in \{-1, 0, 1\}$ for all i and j. Assume now that the determinant of every $(k-1) \times (k-1)$ submatrix of A is -1, 0, or 1. Let B be a $k \times k$ submatrix of A. If $\det(B) = 0$, then we are done, so assume that B is invertible. Our objective is to prove that $|\det B| = 1$. By Cramer's rule and the induction hypothesis, we have $B^{-1} = \frac{B^*}{\det(B)}$, where $b^*_{ij} \in \{-1, 0, 1\}$. By the definition of B^*, we have $Bb^*_1 = \det(B)e_1$, where b^*_1 is the first column of B^* and $e_1 = (1, 0, \ldots 0)'$.

Let $J = \{j \mid b^*_{j1} \neq 0\}$ and note that $J \neq \emptyset$ since B is invertible. Let $\overline{J}_1 = \{j \in J \mid b^*_{j1} = 1\}$ and $\overline{J}_2 = \{j \in J \mid j \notin \overline{J}_1\}$. Then, since $[Bb^*_1]_i = 0$ for $i = 2, \ldots, k$, we have

$$[Bb^*_1]_i = \sum_{j=1}^{k} b_{ij} b^*_{j1} = \sum_{j \in \overline{J}_1} b_{ij} - \sum_{j \in \overline{J}_2} b_{ij} = 0, \qquad \forall\ i = 2, \ldots, k.$$

Thus, the cardinality of the set J is even, so that for any partition $(\tilde{J}_1, \tilde{J}_2)$ of J, it follows that $\sum_{j \in \tilde{J}_1} b_{ij} - \sum_{j \in \tilde{J}_2} b_{ij}$ is even for all $i = 2, \ldots, k$. By assumption, there is a partition (J_1, J_2) of J such that

$$\left| \sum_{j \in J_1} b_{ij} - \sum_{j \in J_2} b_{ij} \right| \leq 1 \qquad \forall\ i = 1, \ldots, k, \tag{3.17}$$

implying that

$$\sum_{j \in J_1} b_{ij} - \sum_{j \in J_2} b_{ij} = 0, \qquad \forall\ i = 2, \ldots, k. \tag{3.18}$$

Consider now the value $\alpha = \left| \sum_{j \in J_1} b_{1j} - \sum_{j \in J_2} b_{1j} \right|$, for which in view of Eq. (3.17), we have either $\alpha = 0$ or $\alpha = 1$. Define $y \in \Re^k$ by $y_i = 1$ for $i \in J_1$, $y_i = -1$ for $i \in J_2$, and $y_i = 0$ otherwise. Then, we have $\big|[By]_1\big| = \alpha$ and by Eq. (3.18), $[By]_i = 0$ for all $i = 2, \ldots, k$. If $\alpha = 0$, then $By = 0$ and since B is invertible, it follows that $y = 0$, implying that $J = \emptyset$, which is a contradiction. Hence, we must have $\alpha = 1$ so that $By = \pm e_1$. Without loss of generality assume that $By = e_1$ (if $By = -e_1$, we can replace y by $-y$). Then, since $Bb^*_1 = \det(B)e_1$, we see that $B\big(b^*_1 - \det(B)y\big) = 0$ and since B is invertible, we must have $b^*_1 = \det(B)y$. Because y and b^*_1 are vectors with components -1, 0, or 1, it follows that $b^*_1 = \pm y$ and $\big|\det(B)\big| = 1$, completing the induction and showing that A is totally unimodular.

3.24 (Unimodularity III)

(a) We show that the determinant of any square submatrix of A is -1, 0, or 1. We prove this by induction on the size of the square submatrices of A. In particular, the 1×1 submatrices of A are the entries of A, which are -1, 0, or 1. Suppose that the determinant of each $(k-1) \times (k-1)$ submatrix of A is -1, 0, or 1, and consider a $k \times k$ submatrix B of A. If B has a zero column, then $\det(B) = 0$ and we are done. If B has a column with a single nonzero component (1 or -1), then by expanding its determinant along that column and by using the induction hypothesis, we see that $\det(B) = 1$ or $\det(B) = -1$. Finally, if each column of B has exactly two nonzero components (one 1 and one -1), the sum of its rows is zero, so that B is singular and $\det(B) = 0$, completing the proof and showing that A is totally unimodular.

(b) The proof is based on induction as in part (a). The 1×1 submatrices of A are the entries of A, which are 0 or 1. Suppose now that the determinant of each $(k-1) \times (k-1)$ submatrix of A is -1, 0, or 1, and consider a $k \times k$ submatrix B of A. Since in each column of A, the entries that are equal to 1 appear consecutively, the same is true for the matrix B. Take the first column b_1 of B. If $b_1 = 0$, then B is singular and $\det(B) = 0$. If b_1 has a single nonzero component, then by expanding the determinant of B along b_1 and by using the induction hypothesis, we see that $\det(B) = 1$ or $\det(B) = -1$. Finally, let b_1 have more than one nonzero component (its nonzero entries are 1 and appear consecutively). Let l and p be rows of B such that $b_{i1} = 0$ for all $i < l$ and $i > p$, and $b_{i1} = 1$ for all $l \leq i \leq p$. By multiplying the lth row of B with (-1) and by adding it to the $l+1$st, $l+2$nd, $\ldots$, kth row of B, we obtain a matrix $\overline{B}$ such that $\det(B) = \det(\overline{B})$ and the first column $\overline{b}_1$ of $\overline{B}$ has a single nonzero component. Furthermore, the determinant of every square submatrix of $\overline{B}$ is -1, 0, or 1 (this follows from the fact that the determinant of a square matrix is unaffected by adding a scalar multiple of a row of the matrix to some of its other rows, and from the induction hypothesis). Since $\overline{b}_1$ has a single nonzero component, by expanding the determinant of $\overline{B}$ along $\overline{b}_1$, it follows that $\det(\overline{B}) = 1$ or $\det(\overline{B}) = -1$, implying that $\det(B) = 1$ or $\det(B) = -1$, completing the induction and showing that A is totally unimodular.

3.25 (Unimodularity IV)

If A is totally unimodular, then by Exercise 3.23(a), its transpose A' is also totally unimodular and by Exercise 3.23(b), the set $I = \{1, \ldots, m\}$ can be partitioned into two subsets I_1 and I_2 such that

$$\left| \sum_{i \in I_1} a_{ij} - \sum_{i \in I_2} a_{ij} \right| \leq 1, \qquad \forall\, j = 1, \ldots, n.$$

Since $a_{ij} \in \{-1, 0, 1\}$ and exactly two of $a_{1j}, \ldots, a_{mj}$ are nonzero for each j, it follows that

$$\sum_{i \in I_1} a_{ij} - \sum_{i \in I_2} a_{ij} = 0, \qquad \forall\, j = 1, \ldots, n.$$

Take any $j \in \{1, \ldots, n\}$, and let l and p be such that $a_{ij} = 0$ for all $i \neq l$ and $i \neq p$, so that in view of the preceding relation and the fact $a_{ij} \in \{-1, 0, 1\}$, we see that: if $a_{lj} = -a_{pj}$, then both l and p are in the same subset (I_1 or I_2); if $a_{lj} = a_{pj}$, then l and p are not in the same subset.

Suppose now that the rows of A can be divided into two subsets such that for each column the following property holds: if the two nonzero entries in the column have the same sign, they are in different subsets, and if they have the opposite sign, they are in the same subset. By multiplying all the rows in one of the subsets by -1, we obtain the matrix $\overline{A}$ with entries $\overline{a}_{ij} \in \{-1, 0, 1\}$, and exactly one 1 and exactly one -1 in each of its columns. Therefore, by Exercise 3.24(a), $\overline{A}$ is totally unimodular, so that every square submatrix of $\overline{A}$ has determinant -1, 0, or 1. Since the determinant of a square submatrix of $\overline{A}$ and the determinant of the corresponding submatrix of A differ only in sign, it follows that every square submatrix of A has determinant -1, 0, or 1, showing that A is totally unimodular.

3.26 (Gordan's Theorem of the Alternative [Gor73])

(a) Assume that there exist $\hat{x} \in \Re^n$ and $\mu \in \Re^r$ such that both conditions (i) and (ii) hold, i.e.,

$$a_j'\hat{x} < 0, \qquad \forall\ j = 1, \ldots, r, \tag{3.19}$$

$$\mu \neq 0, \qquad \mu \geq 0, \qquad \sum_{j=1}^{r} \mu_j a_j = 0. \tag{3.20}$$

By premultiplying Eq. (3.19) with $\mu_j \geq 0$ and summing the obtained inequalities over j, we have

$$\sum_{j=1}^{r} \mu_j a_j'\hat{x} < 0.$$

On the other hand, from Eq. (3.20), we obtain

$$\sum_{j=1}^{r} \mu_j a_j'\hat{x} = 0,$$

which is a contradiction. Hence, both conditions (i) and (ii) cannot hold simultaneously.

The proof will be complete if we show that the conditions (i) and (ii) cannot *fail* to hold simultaneously. Assume that condition (i) fails to hold, and consider the sets given by

$$C_1 = \big\{w \in \Re^r \mid a_j'x \leq w_j,\ j = 1, \ldots, r,\ x \in \Re^n\big\},$$

$$C_2 = \{\xi \in \Re^r \mid \xi_j < 0,\ j = 1, \ldots, r\}.$$

It can be seen that both C_1 and C_2 are convex. Furthermore, because the condition (i) does not hold, C_1 and C_2 are disjoint sets. Therefore, by the Separating Hyperplane Theorem (Prop. 2.4.2), C_1 and C_2 can be separated, i.e., there exists a nonzero vector $\mu \in \Re^r$ such that

$$\mu'w \geq \mu'\xi, \qquad \forall\ w \in C_1, \qquad \forall\ \xi \in C_2,$$

implying that

$$\inf_{w \in C_1} \mu' w \geq \mu' \xi, \qquad \forall\, \xi \in C_2.$$

Since each component ξ_j of $\xi \in C_2$ can be any negative scalar, for the preceding relation to hold, μ_j must be nonnegative for all j. Furthermore, by letting $\xi \to 0$, in the preceding relation, it follows that

$$\inf_{w \in C_1} \mu' w \geq 0,$$

implying that

$$\mu_1 w_1 + \cdots + \mu_r w_r \geq 0, \qquad \forall\, w \in C_1.$$

By setting $w_j = a_j' x$ for all j, we obtain

$$(\mu_1 a_1 + \cdots + \mu_r a_r)' x \geq 0, \qquad \forall\, x \in \Re^n,$$

and because this relation holds for all $x \in \Re^n$, we must have

$$\mu_1 a_1 + \cdots + \mu_r a_r = 0.$$

Hence, the condition (ii) holds, showing that the conditions (i) and (ii) cannot fail to hold simultaneously.

Alternative proof: We will show the equivalent statement of part (b), i.e., that a polyhedral cone contains an interior point if and only if the polar C^* does not contain a line. This is a special case of Exercise 3.2 (the dimension of C plus the dimension of the lineality space of C^* is n), as well as Exercise 3.6(d), but we will give an independent proof.

Let

$$C = \big\{ x \mid a_j' x \leq 0, j = 1, \ldots, r \big\},$$

where $a_j \neq 0$ for all j. Assume that C contains an interior point, and to arrive at a contradiction, assume that C^* contains a line. Then there exists a $d \neq 0$ such that d and $-d$ belong to C^*, i.e., $d'x \leq 0$ and $-d'x \leq 0$ for all $x \in C$, so that $d'x = 0$ for all $x \in C$. Thus for the interior point $\overline{x} \in C$, we have $d'\overline{x} = 0$, and since $d \in C^*$ and $d = \sum_{j=1}^r \mu_j a_j$ for some $\mu_j \geq 0$, we have

$$\sum_{j=1}^r \mu_j a_j' \overline{x} = 0.$$

This is a contradiction, since $\overline{x}$ is an interior point of C, and we have $a_j' \overline{x} < 0$ for all j.

Conversely, assume that C^* does not contain a line. Then by Prop. 3.3.1(b), C^* has an extreme point, and since the origin is the only possible extreme point of a cone, it follows that the origin is an extreme point of C^*, which is the cone generated by $\{a_1, \ldots, a_r\}$. Therefore $0 \notin \mathrm{conv}\big(\{a_1, \ldots, a_r\}\big)$, and there exists a hyperplane that strictly separates the origin from $\mathrm{conv}\big(\{a_1, \ldots, a_r\}\big)$. Thus, there exists a vector x such that $y'x < 0$ for all $y \in \mathrm{conv}\big(\{a_1, \ldots, a_r\}\big)$, so in particular,

$$a_j' x < 0, \qquad \forall\, j = 1, \ldots, r,$$

and x is an interior point of C.

(b) Let C be a polyhedral cone given by

$$C = \left\{x \mid a_j'x \le 0, j = 1, \ldots, r\right\},$$

where $a_j \neq 0$ for all j. The interior of C is given by

$$\text{int}(C) = \left\{x \mid a_j'x < 0,\ j = 1, \ldots, r\right\},$$

so that C has nonempty interior if and only if the condition (i) of part (a) holds.

By Farkas' Lemma [Prop. 3.2.1(b)], the polar cone of C is given by

$$C^* = \left\{x \;\middle|\; x = \sum_{j=1}^{r} \mu_j a_j, \mu_j \ge 0,\ j = 1, \ldots, r\right\}.$$

We now show that C^* contains a line if and only if there is a $\mu \in \Re^r$ such that $\mu \neq 0$, $\mu \geq 0$, and $\sum_{j=1}^r \mu_j a_j = 0$ [condition (ii) of part (a) holds]. Suppose that C^* contains a line, i.e., a set of the form $\{x + \alpha z \mid \alpha \in \Re\}$, where $x \in C^*$ and z is a nonzero vector. Since C^* is a closed convex cone, by the Recession Cone Theorem (Prop. 1.5.1), it follows that z and $-z$ belong to R_{C^*}. This, implies that $0 + z = z \in C^*$ and $0 - z = -z \in C^*$, and therefore z and $-z$ can be represented as

$$z = \sum_{j=1}^{r} \mu_j a_j, \qquad \forall\ j,\ \mu_j \ge 0,\ \mu_j \neq 0 \text{ for some } j,$$

$$-z = \sum_{j=1}^{r} \overline{\mu}_j a_j, \qquad \forall\ j,\ \overline{\mu}_j \ge 0,\ \overline{\mu}_j \neq 0 \text{ for some } j.$$

Thus, $\sum_{j=1}^r (\mu_j + \overline{\mu}_j) a_j = 0$, where $(\mu_j + \overline{\mu}_j) \geq 0$ for all j and $(\mu_j + \overline{\mu}_j) \neq 0$ for at least one j, showing that the condition (ii) of part (a) holds.

Conversely, suppose that $\sum_{j=1}^r \mu_j a_j = 0$ with $\mu_j \geq 0$ for all j and $\mu_j \neq 0$ for some j. Assume without loss of generality that $\mu_1 > 0$, so that

$$-a_1 = \sum_{j \neq 1} \frac{\mu_j}{\mu_1} a_j,$$

with $\mu_j/\mu_1 \geq 0$ for all j, which implies that $-a_1 \in C^*$. Since $a_1 \in C^*$, $-a_1 \in C^*$, and $a_1 \neq 0$, it follows that C^* contains a line, completing the proof.

3.27 (Linear System Alternatives)

Assume that there exist $\hat{x} \in \Re^n$ and $\mu \in \Re^r$ such that both conditions (i) and (ii) hold, i.e.,

$$a_j'\hat{x} \le b_j, \qquad \forall\ j = 1, \ldots, r, \tag{3.21}$$

$$\mu \ge 0, \qquad \sum_{j=1}^{r} \mu_j a_j = 0, \qquad \sum_{j=1}^{r} \mu_j b_j < 0. \tag{3.22}$$

By premultiplying Eq. (3.21) with $\mu_j \ge 0$ and summing the obtained inequalities over j, we have

$$\sum_{j=1}^{r} \mu_j a_j'\hat{x} \le \sum_{j=1}^{r} \mu_j b_j.$$

On the other hand, by using Eq. (3.22), we obtain

$$\sum_{j=1}^{r} \mu_j a_j'\hat{x} = 0 > \sum_{j=1}^{r} \mu_j b_j,$$

which is a contradiction. Hence, both conditions (i) and (ii) cannot hold simultaneously.

The proof will be complete if we show that conditions (i) and (ii) cannot *fail* to hold simultaneously. Assume that condition (i) fails to hold, and consider the sets given by

$$P_1 = \{\xi \in \Re^r \mid \xi_j \le 0,\ j = 1, \ldots, r\},$$

$$P_2 = \big\{w \in \Re^r \mid a_j'x - b_j = w_j,\ j = 1, \ldots, r,\ x \in \Re^n\big\}.$$

Clearly, P_1 is a polyhedral set. For the set P_2, we have

$$P_2 = \{w \in \Re^r \mid Ax - b = w,\ x \in \Re^n\} = R(A) - b,$$

where A is the matrix with rows a_j' and b is the vector with components b_j. Thus, P_2 is an affine set and is therefore polyhedral. Furthermore, because the condition (i) does not hold, P_1 and P_2 are disjoint polyhedral sets, and they can be strictly separated [Prop. 2.4.3 under condition (5)]. Hence, there exists a vector $\mu \in \Re^r$ such that

$$\sup_{\xi \in P_1} \mu'\xi < \inf_{w \in P_2} \mu'w.$$

Since each component ξ_j of $\xi \in P_1$ can be any negative scalar, for the preceding relation to hold, μ_j must be nonnegative for all j. Furthermore, since $0 \in P_1$, it follows that

$$0 < \inf_{w \in P_2} \mu'w,$$

implying that

$$0 < \mu_1 w_1 + \cdots + \mu_r w_r, \qquad \forall\ w \in P_2.$$

By setting $w_j = a_j'x - b_j$ for all j, we obtain

$$\mu_1 b_1 + \cdots + \mu_r b_r < (\mu_1 a_1 + \cdots + \mu_r a_r)'x, \qquad \forall\ x \in \Re^n.$$

Since this relation holds for all $x \in \Re^n$, we must have

$$\mu_1 a_1 + \cdots + \mu_r a_r = 0,$$

implying that

$$\mu_1 b_1 + \cdots + \mu_r b_r < 0.$$

Hence, the condition (ii) holds, showing that the conditions (i) and (ii) cannot fail to hold simultaneously.

3.28 (Convex System Alternatives [FGH57])

Assume that there exist $\hat{x} \in C$ and $\mu \in \Re^r$ such that both conditions (i) and (ii) hold, i.e.,

$$f_j(\hat{x}) < 0, \qquad \forall\, j = 1, \ldots, r, \tag{3.23}$$

$$\mu \neq 0, \qquad \mu \geq 0, \qquad \sum_{j=1}^{r} \mu_j f_j(\hat{x}) \geq 0. \tag{3.24}$$

By premultiplying Eq. (3.23) with $\mu_j \geq 0$ and summing the obtained inequalities over j, we obtain, using the fact $\mu \neq 0$,

$$\sum_{j=1}^{r} \mu_j f_j(\hat{x}) < 0,$$

contradicting the last relation in Eq. (3.24). Hence, both conditions (i) and (ii) cannot hold simultaneously.

The proof will be complete if we show that conditions (i) and (ii) cannot *fail* to hold simultaneously. Assume that condition (i) fails to hold, and consider the sets given by

$$P = \{\xi \in \Re^r \mid \xi_j \leq 0,\ j = 1, \ldots, r\},$$

$$C_1 = \big\{w \in \Re^r \mid f_j(x) < w_j,\ j = 1, \ldots, r,\ x \in C\big\}.$$

The set P is polyhedral, while C_1 is convex by the convexity of C and f_j for all j. Furthermore, since condition (i) does not hold, P and C_1 are disjoint, implying that $\mathrm{ri}(C_1) \cap P = \emptyset$. By the Polyhedral Proper Separation Theorem (cf. Prop. 3.5.1), the polyhedral set P and convex set C_1 can be properly separated by a hyperplane that does not contain C_1, i.e., there exists a vector $\mu \in \Re^r$ such that

$$\sup_{\xi \in P} \mu'\xi \leq \inf_{w \in C_1} \mu' w, \qquad \inf_{w \in C_1} \mu' w < \sup_{w \in C_1} \mu' w.$$

Since each component ξ_j of $\xi \in P$ can be any negative scalar, the first relation implies that $\mu_j \geq 0$ for all j, while the second relation implies that $\mu \neq 0$. Furthermore, since $\mu'\xi \leq 0$ for all $\xi \in P$ and $0 \in P$, it follows that

$$\sup_{\xi \in P} \mu'\xi = 0 \leq \inf_{w \in C_1} \mu' w,$$

implying that

$$0 \leq \mu_1 w_1 + \cdots + \mu_r w_r, \qquad \forall\ w \in C_1.$$

By letting $w_j \to f_j(x)$ for all j, we obtain

$$0 \leq \mu_1 f_1(x) + \cdots + \mu_r f_r(x), \qquad \forall\ x \in C \cap \mathrm{dom}(f_1) \cap \cdots \cap \mathrm{dom}(f_r).$$

Thus, the convex function

$$f = \mu_1 f_1 + \cdots + \mu_r f_r$$

is finite and nonnegative over the convex set

$$\tilde{C} = C \cap \mathrm{dom}(f_1) \cap \cdots \cap \mathrm{dom}(f_r).$$

By Exercise 1.27, the function f is nonnegative over $\mathrm{cl}\big(\tilde{C}\big)$. Given that $\mathrm{ri}(C) \subset \mathrm{dom}(f_i)$ for all i, we have $\mathrm{ri}(C) \subset \tilde{C}$, and therefore

$$C \subset \mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}\big(\tilde{C}\big).$$

Hence, f is nonnegative over C and condition (ii) holds, showing that the conditions (i) and (ii) cannot fail to hold simultaneously.

3.29 (Convex-Affine System Alternatives)

Assume that there exist $\hat{x} \in C$ and $\mu \in \Re^r$ such that both conditions (i) and (ii) hold, i.e.,

$$f_j(\hat{x}) < 0, \quad \forall\ j = 1, \ldots, \overline{r}, \qquad f_j(\hat{x}) \leq 0, \quad \forall\ j = \overline{r}+1, \ldots, r, \tag{3.25}$$

$$(\mu_1, \ldots, \mu_{\overline{r}}) \neq 0, \qquad \mu \geq 0, \qquad \sum_{j=1}^{r} \mu_j f_j(\hat{x}) \geq 0. \tag{3.26}$$

By premultiplying Eq. (3.25) with $\mu_j \geq 0$ and by summing the obtained inequalities over j, since not all $\mu_1, \ldots, \mu_{\overline{r}}$ are zero, we obtain

$$\sum_{j=1}^{r} \mu_j f_j(\hat{x}) < 0,$$

contradicting the last relation in Eq. (3.26). Hence, both conditions (i) and (ii) cannot hold simultaneously.

The proof will be complete if we show that conditions (i) and (ii) cannot *fail* to hold simultaneously. Assume that condition (i) fails to hold, and consider the sets given by

$$P = \{\xi \in \Re^r \mid \xi_j \leq 0,\ j = 1, \ldots, r\},$$

$$C_1 = \big\{w \in \Re^r \mid f_j(x) < w_j,\ j = 1, \ldots, \overline{r},\ f_j(x) = w_j,\ j = \overline{r}+1, \ldots, r,\ x \in C\big\}.$$

The set P is polyhedral, and it can be seen that C_1 is convex, since C and $f_1, \ldots, f_{\overline{r}}$ are convex, and $f_{\overline{r}+1}, \ldots, f_r$ are affine. Furthermore, since the condition (i) does not hold, P and C_1 are disjoint, implying that $\mathrm{ri}(C_1) \cap P = \emptyset$. Therefore, by the Polyhedral Proper Separation Theorem (cf. Prop. 3.5.1), the polyhedral set P and convex set C_1 can be properly separated by a hyperplane that does not contain C_1, i.e., there exists a vector $\mu \in \Re^r$ such that

$$\sup_{\xi \in P} \mu'\xi \leq \inf_{w \in C_1} \mu'w, \qquad \inf_{w \in C_1} \mu'w < \sup_{w \in C_1} \mu'w. \tag{3.27}$$

Since each component ξ_j of $\xi \in P$ can be any negative scalar, the first relation implies that $\mu_j \geq 0$ for all j. Therefore, $\mu'\xi \leq 0$ for all $\xi \in P$ and since $0 \in P$, it follows that

$$\sup_{\xi \in P} \mu'\xi = 0 \leq \inf_{w \in C_1} \mu'w.$$

This implies that

$$0 \leq \mu_1 w_1 + \cdots + \mu_r w_r, \qquad \forall\ w \in C_1,$$

and by letting $w_j \to f_j(x)$ for $j = 1, \ldots, \overline{r}$, we have

$$0 \leq \mu_1 f_1(x) + \cdots + \mu_r f_r(x), \qquad \forall\ x \in C \cap \mathrm{dom}(f_1) \cap \cdots \cap \mathrm{dom}(f_r).$$

Thus, the convex function

$$f = \mu_1 f_1 + \cdots + \mu_r f_r$$

is finite and nonnegative over the convex set

$$\overline{C} = C \cap \mathrm{dom}(f_1) \cap \cdots \cap \mathrm{dom}(f_{\overline{r}}).$$

By Exercise 1.27, f is nonnegative over $\mathrm{cl}\big(\overline{C}\big)$. Given that $\mathrm{ri}(C) \subset \mathrm{dom}(f_i)$ for all $i = 1, \ldots, \overline{r}$, we have $\mathrm{ri}(C) \subset \overline{C}$, and therefore

$$C \subset \mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}\big(\overline{C}\big).$$

Hence, f is nonnegative over C.

We now show that not all $\mu_1, \ldots, \mu_{\overline{r}}$ are zero. To arrive at a contradiction, suppose that all $\mu_1, \ldots, \mu_{\overline{r}}$ are zero, so that

$$0 \leq \mu_{\overline{r}+1} f_{\overline{r}+1}(x) + \cdots + \mu_r f_r(x), \qquad \forall\ x \in C.$$

Since the system

$$f_{\overline{r}+1}(x) \leq 0, \ldots, f_r(x) \leq 0,$$

has a solution $\overline{x} \in \mathrm{ri}(C)$, it follows that

$$\mu_{\overline{r}+1} f_{\overline{r}+1}(\overline{x}) + \cdots + \mu_r f_r(\overline{x}) = 0,$$

so that

$$\inf_{x \in C} \left\{ \mu_{\overline{r}+1} f_{\overline{r}+1}(x) + \cdots + \mu_r f_r(x) \right\} = \mu_{\overline{r}+1} f_{\overline{r}+1}(\overline{x}) + \cdots + \mu_r f_r(\overline{x}) = 0,$$

with $\overline{x} \in \mathrm{ri}(C)$. Thus, the affine function $\mu_{\overline{r}+1} f_{\overline{r}+1} + \cdots + \mu_r f_r$ attains its minimum value over C at a point in the relative interior of C. Hence, by Prop. 1.4.2 of Chapter 1, the function $\mu_{\overline{r}+1} f_{\overline{r}+1} + \cdots + \mu_r f_r$ is constant over C, i.e.,

$$\mu_{\overline{r}+1} f_{\overline{r}+1}(x) + \cdots + \mu_r f_r(x) = 0, \qquad \forall\ x \in C.$$

Furthermore, we have $\mu_j = 0$ for all $j = 1, \ldots, \overline{r}$, while by the definition of C_1, we have $f_j(x) = w_j$ for $j = \overline{r} + 1, \ldots, r$, which combined with the preceding relation yields

$$\mu_1 w_1 + \cdots + \mu_r w_r = 0, \qquad \forall\ w \in C_1,$$

implying that

$$\inf_{w \in C_1} \mu' w = \sup_{w \in C_1} \mu' w.$$

This contradicts the second relation in (3.27). Hence, not all $\mu_1, \ldots, \mu_{\overline{r}}$ are zero, showing that the condition (ii) holds, and proving that the conditions (i) and (ii) cannot fail to hold simultaneously.

3.30 (Elementary Vectors [Roc69])

(a) If two elementary vectors z and $\overline{z}$ had the same support, the vector $z - \gamma \overline{z}$ would be nonzero and have smaller support than z and $\overline{z}$ for a suitable scalar γ. If z and $\overline{z}$ are not scalar multiples of each other, then $z - \gamma \overline{z} \neq 0$, which contradicts the definition of an elementary vector.

(b) We note that either y is elementary or else there exists a nonzero vector $\overline{z}$ with support strictly contained in the support of y. Repeating this argument for at most $n - 1$ times, we must obtain an elementary vector.

(c) We first show that every nonzero vector $y \in S$ has the property that there exists an elementary vector of S that is in harmony with y and has support that is contained in the support of y.

We show this by induction on the number of nonzero components of y. Let V_k be the subset of nonzero vectors in S that have k or less nonzero components, and let $\overline{k}$ be the smallest k for which V_k is nonempty. Then, by part (b), every vector $y \in V_{\overline{k}}$ must be elementary, so it has the desired property. Assume that all vectors in V_k have the desired property for some $k \geq \overline{k}$. We let y be a vector in V_{k+1} and we show that it also has the desired property. Let z be an elementary vector whose support is contained in the support of y. By using the negative of z if necessary, we can assume that $y_j z_j > 0$ for at least one index j. Then there exists a largest value of γ, call it $\overline{\gamma}$, such that

$$y_j - \gamma z_j \geq 0, \qquad \forall\ j \text{ with } y_j > 0,$$

$$y_j - \gamma z_j \leq 0, \qquad \forall\ j \text{ with } y_j < 0.$$

The vector $y - \overline{\gamma} z$ is in harmony with y and has support that is strictly contained in the support of y. Thus either $y - \overline{\gamma} z = 0$, in which case the elementary vector z is in harmony with y and has support equal to the support of y, or else $y - \overline{\gamma} z$ is nonzero. In the latter case, we have $y - \overline{\gamma} z \in V_k$, and by the induction hypothesis, there exists an elementary vector $\overline{z}$ that is in harmony with $y - \overline{\gamma} z$ and has support that is contained in the support of $y - \overline{\gamma} z$. The vector $\overline{z}$ is also in harmony with y and has support that is contained in the support of y. The induction is complete.

Consider now the given nonzero vector $x \in S$, and choose any elementary vector $\overline{z}^1$ of S that is in harmony with x and has support that is contained in the support of x (such a vector exists by the property just shown). By using the negative of $\overline{z}^1$ if necessary, we can assume that $x_j \overline{z}_j^1 > 0$ for at least one index j. Let $\overline{\gamma}$ be the largest value of γ such that

$$x_j - \gamma \overline{z}_j^1 \geq 0, \qquad \forall\ j \text{ with } x_j > 0,$$

$$x_j - \gamma \overline{z}_j^1 \leq 0, \qquad \forall\ j \text{ with } x_j < 0.$$

The vector $x - z^1$, where

$$z^1 = \overline{\gamma}\, \overline{z}^1,$$

is in harmony with x and has support that is strictly contained in the support of x. There are two cases: (1) $x = z^1$, in which case we are done, or (2) $x \neq z^1$, in which case we replace x by $x - z^1$ and we repeat the process. Eventually, after m steps where $m \leq n$ (since each step reduces the number of nonzero components by at least one), we will end up with the desired decomposition $x = z^1 + \cdots + z^m$.

3.31 (Combinatorial Separation Theorem [Cam68], [Roc69])

For simplicity, assume that B is the Cartesian product of bounded open intervals, so that B has the form

$$B = \{t \mid \underline{b}_j < t_j < \overline{b}_j,\ j = 1, \ldots, n\},$$

where $\underline{b}_j$ and $\overline{b}_j$ are some scalars. The proof is easily modified for the case where B has a different form.

Since $B \cap S^\perp = \emptyset$, there exists a hyperplane that separates B and $S^\perp$. The normal of this hyperplane is a nonero vector $d \in S$ such that

$$t'd \leq 0, \qquad \forall\ t \in B.$$

Since B is open, this inequality implies that actually

$$t'd < 0, \qquad \forall\ t \in B.$$

Equivalently, we have

$$\sum_{\{j \mid d_j > 0\}} (\overline{b}_j - \epsilon) d_j + \sum_{\{j \mid d_j < 0\}} (\underline{b}_j + \epsilon) d_j < 0, \tag{3.28}$$

for all $\epsilon > 0$ such that $\underline{b}_j + \epsilon < \overline{b}_j - \epsilon$. Let

$$d = z^1 + \cdots + z^m,$$

be a decomposition of d, where $z^1, \ldots, z^m$ are elementary vectors of S that are in harmony with x, and have supports that are contained in the support of d [cf. part (c) of the Exercise 3.30]. Then the condition (3.28) is equivalently written as

$$\begin{aligned}
0 &> \sum_{\{j \mid d_j > 0\}} (\overline{b}_j - \epsilon) d_j + \sum_{\{j \mid d_j < 0\}} (\underline{b}_j + \epsilon) d_j \\
&= \sum_{\{j \mid d_j > 0\}} (\overline{b}_j - \epsilon) \left(\sum_{i=1}^m z_j^i \right) + \sum_{\{j \mid d_j < 0\}} (\underline{b}_j + \epsilon) \left(\sum_{i=1}^m z_j^i \right) \\
&= \sum_{i=1}^m \left(\sum_{\{j \mid z_j^i > 0\}} (\overline{b}_j - \epsilon) z_j^i + \sum_{\{j \mid z_j^i < 0\}} (\underline{b}_j + \epsilon) z_j^i \right),
\end{aligned}$$

where the last equality holds because the vectors z^i are in harmony with d and their supports are contained in the support of d. From the preceding relation, we see that for at least one elementary vector z^i, we must have

$$0 > \sum_{\{j \mid z_j^i > 0\}} (\overline{b}_j - \epsilon) z_j^i + \sum_{\{j \mid z_j^i < 0\}} (\underline{b}_j + \epsilon) z_j^i,$$

for all $\epsilon > 0$ that are sufficiently small and are such that $\underline{b}_j + \epsilon < \overline{b}_j - \epsilon$, or equivalently

$$0 > t' z^i, \qquad \forall\, t \in B.$$

3.32 (Tucker's Complementarity Theorem)

(a) Fix an index k and consider the following two assertions:

(1) There exists a vector $x \in S$ with $x_i \geq 0$ for all i, and $x_k > 0$.

(2) There exists a vector $y \in S^\perp$ with $y_i \geq 0$ for all i, and $y_k > 0$.

We claim that one and only one of the two assertions holds. Clearly, assertions (1) and (2) cannot hold simultaneously, since then we would have $x'y > 0$, while $x \in S$ and $y \in S^\perp$. We will show that they cannot fail simultaneously. Indeed, if (1) does not hold, the Cartesian product $B = \Pi_{i=1}^n B_i$ of the intervals

$$B_i = \begin{cases} (0, \infty) & \text{if } i = k, \\ [0, \infty) & \text{if } i \neq k, \end{cases}$$

does not intersect the subspace S, so by the result of Exercise 3.31, there exists a vector z of $S^\perp$ such that $x'z < 0$ for all $x \in B$. For this to hold, we must have $z \in B^*$ or equivalently $z \leq 0$, while by choosing $x = (0, \ldots, 0, 1, 0, \ldots, 0) \in B$, with the 1 in the kth position, the inequality $x'z < 0$ yields $z_k < 0$. Thus assertion (2) holds with $y = -z$. Similarly, we show that if (2) does not hold, then (1) must hold.

Let now I be the set of indices k such that (1) holds, and for each $k \in I$, let $x(k)$ be a vector in S such that $x(k) \geq 0$ and $x_k(k) > 0$ (note that we do not exclude the possibility that one of the sets I and $\overline{I}$ is empty). Let $\overline{I}$ be the set of indices such that (2) holds, and for each $k \in \overline{I}$, let $y(k)$ be a vector in $S^\perp$ such that $y(k) \geq 0$ and $y_k(k) > 0$. From what has already been shown, I and $\overline{I}$ are disjoint, $I \cup \overline{I} = \{1, \ldots, n\}$, and the vectors

$$x = \sum_{k \in I} x(k), \qquad y = \sum_{k \in \overline{I}} y(k),$$

satisfy

$$x_i > 0, \quad \forall\, i \in I, \qquad x_i = 0, \quad \forall\, i \in \overline{I},$$

$$y_i = 0, \quad \forall\, i \in I, \qquad y_i > 0, \quad \forall\, i \in \overline{I}.$$

The uniqueness of I and $\overline{I}$ follows from their construction and the preceding arguments. In particular, if for some $k \in \overline{I}$, there existed a vector $x \in S$ with $x \geq 0$ and $x_k > 0$, then since for the vector $y(k)$ of $S^\perp$ we have $y(k) \geq 0$ and $y_k(k) > 0$, assertions (a) and (b) must hold simultaneously, which is a contradiction.

The last assertion follows from the fact that for each k, exactly one of the assertions (1) and (2) holds.

(b) Consider the subspace

$$S = \big\{(x, w) \mid Ax - bw = 0,\, x \in \Re^n,\, w \in \Re\big\}.$$

Its orthogonal complement is the range of the transpose of the matrix $[A \;\; -b]$, so it has the form

$$S^\perp = \big\{(A'z, -b'z) \mid z \in \Re^m\big\}.$$

By applying the result of part (a) to the subspace S, we obtain a partition of the index set $\{1, \ldots, n+1\}$ into two subsets. There are two possible cases:

(1) The index $n+1$ belongs to the first subset.

(2) The index $n+1$ belongs to the second subset.

In case (2), the two subsets are of the form I and $\overline{I} \cup \{n+1\}$ with $I \cup \overline{I} = \{1, \ldots, n\}$, and by the last assertion of part (a), we have $w = 0$ for all (x, w) such that $x \geq 0$, $w \geq 0$ and $Ax - bw = 0$. This, however, contradicts the fact that the set $F = \{x \mid Ax = b,\, x \geq 0\}$ is nonempty. Therefore, case (1) holds, i.e., the index $n+1$ belongs to the first index subset. In particular, we have that there exist disjoint index sets I and $\overline{I}$ with $I \cup \overline{I} = \{1, \ldots, n\}$, and vectors (x, w) with $Ax - bw = 0$, and $z \in \Re^m$ such that

$$w > 0, \qquad b'z = 0,$$

$$x_i > 0, \quad \forall\, i \in I, \qquad x_i = 0, \quad \forall\, i \in \overline{I},$$

$$y_i = 0, \quad \forall\, i \in I, \qquad y_i > 0, \quad \forall\, i \in \overline{I},$$

where $y = A'z$. By dividing (x, w) with w if needed, we may assume that $w = 1$ so that $Ax - b = 0$, and the result follows.

Convex Analysis and Optimization

Chapter 4 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 4: SOLUTION MANUAL

4.1 (Directional Derivative of Extended Real-Valued Functions)

(a) Since $f'(x;0) = 0$, the relation $f'(x;\lambda y) = \lambda f'(x;y)$ clearly holds for $\lambda = 0$ and all $y \in \Re^n$. Choose $\lambda > 0$ and $y \in \Re^n$. By the definition of directional derivative, we have

$$f'(x;\lambda y) = \inf_{\alpha>0} \frac{f\big(x + \alpha(\lambda y)\big) - f(x)}{\alpha} = \lambda \inf_{\alpha>0} \frac{f\big(x + (\alpha\lambda)y\big) - f(x)}{\alpha\lambda}.$$

By setting $\beta = \lambda\alpha$ in the preceding relation, we obtain

$$f'(x;\lambda y) = \lambda \inf_{\beta>0} \frac{f(x + \beta y) - f(x)}{\beta} = \lambda f'(x;y).$$

(b) Let (y_1, w_1) and (y_2, w_2) be two points in $\mathrm{epi}\big(f'(x;\cdot)\big)$, and let γ be a scalar with $\gamma \in (0,1)$. Consider a point (y_γ, w_γ) given by

$$y_\gamma = \gamma y_1 + (1-\gamma)y_2, \qquad w_\gamma = \gamma w_1 + (1-\gamma)w_2.$$

Since for all $y \in \Re^n$, the ratio

$$\frac{f(x + \alpha y) - f(x)}{\alpha}$$

is monotonically nonincreasing as $\alpha \downarrow 0$, we have

$$\frac{f(x + \alpha y_1) - f(x)}{\alpha} \leq \frac{f(x + \alpha_1 y_1) - f(x)}{\alpha_1}, \qquad \forall\ \alpha, \alpha_1,\ \text{with } 0 < \alpha \leq \alpha_1,$$

$$\frac{f(x + \alpha y_2) - f(x)}{\alpha} \leq \frac{f(x + \alpha_2 y_2) - f(x)}{\alpha_2}, \qquad \forall\ \alpha, \alpha_2,\ \text{with } 0 < \alpha \leq \alpha_2.$$

Multiplying the first relation by γ and the second relation by $1-\gamma$, and adding, we have for all α with $0 < \alpha \leq \alpha_1$ and $0 < \alpha \leq \alpha_2$,

$$\begin{aligned}\frac{\gamma f(x + \alpha y_1) + (1-\gamma) f(x + \alpha y_2) - f(x)}{\alpha} \leq\ & \gamma \frac{f(x + \alpha_1 y_1) - f(x)}{\alpha_1} \\ & + (1-\gamma)\frac{f(x + \alpha_2 y_2) - f(x)}{\alpha_2}.\end{aligned}$$

From the convexity of f and the definition of y_γ, it follows that

$$f(x + \alpha y_\gamma) \leq \gamma f(x + \alpha\gamma y_1) + (1-\gamma) f(x + \alpha y_2).$$

Combining the preceding two relations, we see that for all $\alpha \le \alpha_1$ and $\alpha \le \alpha_2$,

$$\frac{f(x+\alpha y_\gamma)-f(x)}{\alpha} \le \gamma \frac{f(x+\alpha_1 y_1)-f(x)}{\alpha_1} + (1-\gamma)\frac{f(x+\alpha_2 y_2)-f(x)}{\alpha_2}.$$

By taking the infimum over α, and then over α_1 and α_2, we obtain

$$f'(x;y_\gamma) \le \gamma f'(x;y_1) + (1-\gamma)f'(x;y_2) \le \gamma w_1 + (1-\gamma)w_2 = w_\gamma,$$

where in the last inequality we use the fact $(y_1,w_1),(y_2,w_2) \in \text{epi}\big(f'(x;\cdot)\big)$ Hence the point (y_γ, w_γ) belongs to $\text{epi}\big(f'(x;\cdot)\big)$, implying that $f'(x;\cdot)$ is a convex function.

(c) Since $f'(x;0)=0$ and $(1/2)y+(1/2)(-y)=0$, it follows that

$$f'\big(x;(1/2)y+(1/2)(-y)\big)=0, \qquad \forall\ y \in \Re^n.$$

By part (b), the function $f'(x;\cdot)$ is convex, so that

$$0 \le (1/2)f'(x;y)+(1/2)f'(x;-y),$$

and

$$-f'(x;-y) \le f'(x;y).$$

(d) Let a vector $\overline{y}$ be in the level set $\big\{y \mid f'(x;y) \le 0\big\}$, and let $\lambda > 0$. By part (a),

$$f'(x;\lambda\overline{y}) = \lambda f'(x;\overline{y}) \le 0,$$

so that $\lambda\overline{y}$ also belongs to this level set, which is therefore a cone. By part (b), the function $f'(x;\cdot)$ is convex, implying that the level set $\big\{y \mid f'(x;y) \le 0\big\}$ is convex.

Since $\text{dom}(f)=\Re^n$, $f'(x;\cdot)$ is a real-valued function, and since it is convex, by Prop. 1.4.6, it is also continuous over $\Re^n$. Therefore the level set $\big\{y \mid f'(x;y) \le 0\big\}$ is closed.

We now show that

$$\Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^* = \text{cl}\Big(\text{cone}\big(\partial f(x)\big)\Big).$$

By Prop. 4.2.2, we have

$$f'(x;y) = \max_{d \in \partial f(x)} y'd,$$

implying that $f'(x;y) \le 0$ if and only if $\max_{d \in \partial f(x)} y'd \le 0$. Equivalently, $f'(x;y) \le 0$ if and only if

$$y'd \le 0, \qquad \forall\ d \in \partial f(x).$$

Since

$$y'd \le 0, \qquad \forall\ d \in \partial f(x) \qquad \Longleftrightarrow \qquad y'd \le 0, \qquad \forall\ d \in \text{cone}\big(\partial f(x)\big),$$

it follows from Prop. 3.1.1(a) that $f'(x;y) \le 0$ if and only if

$$y'd \le 0, \qquad \forall\ d \in \text{cone}\big(\partial f(x)\big).$$

Therefore

$$\big\{y \mid f'(x;y) \le 0\big\} = \Big(\text{cone}\big(\partial f(x)\big)\Big)^*,$$

and the desired relation follows by the Polar Cone Theorem [Prop. 3.1.1(b)].

4.2 (Chain Rule for Directional Derivatives)

For any $d \in \Re^n$, by using the directional differentiability of f at x, we have

$$\begin{aligned} F(x+\alpha d) - F(x) &= g\big(f(x+\alpha d)\big) - g\big(f(x)\big) \\ &= g\big(f(x) + \alpha f'(x;d) + o(\alpha)\big) - g\big(f(x)\big). \end{aligned}$$

Let $z_\alpha = f'(x;d) + o(\alpha)/\alpha$ and note that $z_\alpha \to f'(x;d)$ as $\alpha \downarrow 0$. By using this and the assumed property of g, we obtain

$$\lim_{\alpha \downarrow 0} \frac{F(x+\alpha d) - F(x)}{\alpha} = \lim_{\alpha \downarrow 0} \frac{g\big(f(x) + \alpha z_\alpha\big) - g\big(f(x)\big)}{\alpha} = g'\big(f(x); f'(x;d)\big),$$

showing that F is directionally differentiable at x and that the given chain rule holds.

4.3

By definition, a vector $d \in \Re^n$ is a subgradient of f at x if and only if

$$f(y) \geq f(x) + d'(y-x), \qquad \forall\, y \in \Re^n,$$

or equivalently

$$d'x - f(x) \geq d'y - f(y), \qquad \forall\, y \in \Re^n.$$

Therefore, $d \in \Re^n$ is a subgradient of f at x if and only if

$$d'x - f(x) = \max_y \big\{d'y - f(y)\big\}.$$

4.4

(a) For $x \neq 0$, the function $f(x) = \|x\|$ is differentiable with $\nabla f(x) = x/\|x\|$, so that $\partial f(x) = \big\{\nabla f(x)\big\} = \big\{x/\|x\|\big\}$. Consider now the case $x = 0$. If a vector d is a subgradient of f at $x = 0$, then $f(z) \geq f(0) + d'z$ for all z, implying that

$$\|z\| \geq d'z, \qquad \forall\, z \in \Re^n.$$

By letting $z = d$ in this relation, we obtain $\|d\| \leq 1$, showing that $\partial f(0) \subset \big\{d \mid \|d\| \leq 1\big\}$.

On the other hand, for any $d \in \Re^n$ with $\|d\| \leq 1$, we have

$$d'z \leq \|d\| \cdot \|z\| \leq \|z\|, \qquad \forall\, z \in \Re^n,$$

which is equivalent to $f(0) + d'z \leq f(z)$ for all z, so that $d \in \partial f(0)$, and therefore $\big\{d \mid \|d\| \leq 1\big\} \subset \partial f(0)$.

Note that an alternative proof is obtained by writing

$$\|x\| = \max_{\|z\| \leq 1} x'z,$$

and by using Danskin's Theorem (Prop. 4.5.1).

(b) By convention $\partial f(x) = \emptyset$ when $x \notin \mathrm{dom}(f)$, and since here $\mathrm{dom}(f) = C$, we see that $\partial f(x) = \emptyset$ when $x \notin C$. Let now $x \in C$. A vector d is a subgradient of f at x if and only if

$$d'(z - x) \leq f(z), \qquad \forall\, z \in \Re^n.$$

Because $f(z) = \infty$ for all $z \notin C$, the preceding relation always holds when $z \notin C$, so the points $z \notin C$ can be ignored. Thus, $d \in \partial f(x)$ if and only if

$$d'(z - x) \leq 0, \qquad \forall\, z \in C.$$

Since C is convex, by Prop. 4.6.3, the preceding relation is equivalent to $d \in N_C(x)$, implying that $\partial f(x) = N_C(x)$ for all $x \in C$.

4.5

When f is defined on the real line, by Prop. 4.2.1, $\partial f(x)$ is a compact interval of the form

$$\partial f(x) = [\alpha, \beta].$$

By Prop. 4.2.2, we have

$$f'(x; y) = \max_{d \in \partial f(x)} y'd, \qquad \forall\, y \in \Re^n,$$

from which we see that

$$f'(x; 1) = \alpha, \qquad f'(x; -1) = \beta.$$

Since

$$f'(x; 1) = f^+(x), \qquad f'(x; -1) = f^-(x),$$

we have

$$\partial f(x) = [f^-(x), f^+(x)].$$

4.6

We can view the function

$$\varphi(t) = f\big(tx + (1 - t)y\big), \qquad t \in \Re$$

as the composition of the form

$$\varphi(t) = f\big(g(t)\big), \qquad t \in \Re,$$

where $g(t) : \Re \mapsto \Re^n$ is an affine function given by

$$g(t) = y + t(x - y), \qquad t \in \Re.$$

By using the Chain Rule [Prop. 4.2.5(a)], where $A = (x - y)$, we obtain

$$\partial \varphi(t) = A'\, \partial f\big(g(t)\big), \qquad \forall\, t \in \Re,$$

or equivalently

$$\partial \varphi(t) = \big\{(x - y)'d \mid d \in \partial f\big(tx + (1 - t)y\big)\big\}, \qquad \forall\, t \in \Re.$$

4.7

Let x and y be any two points in the set X. Since $\partial f(x)$ is nonempty, by using the subgradient inequality, it follows that

$$f(x) + d'(x - y) \le f(y), \qquad \forall\, d \in \partial f(x),$$

implying that

$$f(x) - f(y) \le \|d\| \cdot \|x - y\|, \qquad \forall\, d \in \partial f(x).$$

According to Prop. 4.2.3, the set $\cup_{x \in X} \partial f(x)$ is bounded, so that for some constant $L > 0$, we have

$$\|d\| \le L, \qquad \forall\, d \in \partial f(x), \quad \forall\, x \in X, \tag{4.1}$$

and therefore,

$$f(x) - f(y) \le L\, \|x - y\|.$$

By exchanging the roles of x and y, we similarly obtain

$$f(y) - f(x) \le L\, \|x - y\|,$$

and by combining the preceding two relations, we see that

$$\big|f(x) - f(y)\big| \le L\, \|x - y\|,$$

showing that f is Lipschitz continuous over X.

Also, by using Prop. 4.2.2(b) and the subgradient boundedness [Eq. (4.1)], we obtain

$$f'(x; y) = \max_{d \in \partial f(x)} d'y \le \max_{d \in \partial f(x)} \|d\| \cdot \|y\| \le L\, \|y\|, \qquad \forall\, x \in X, \quad \forall\, y \in \Re^n.$$

4.8 (Nonemptiness of Subdifferential)

Suppose that $\partial f(x)$ is nonempty, and let $z \in \mathrm{dom}(f)$. By the definition of subgradient, for any $d \in \partial f(x)$, we have

$$\frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha} \ge d'(z - x), \qquad \forall\, \alpha > 0,$$

implying that

$$f'(x; z - x) = \inf_{\alpha > 0} \frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha} \ge d'(z - x) > -\infty.$$

Furthermore, for all $\alpha \in (0,1)$, and $x, z \in \text{dom}(f)$, the vector $x+\alpha(z-x)$ belongs to $\text{dom}(f)$. Therefore, for all $\alpha \in (0,1)$,

$$\frac{f\big(x+\alpha(z-x)\big)-f(x)}{\alpha} < \infty,$$

implying that

$$f'(x; z-x) < \infty.$$

Hence, $f'(x; z-x)$ is finite.

Converseley, suppose that $f'(x; z-x)$ is finite for all $z \in \text{dom}(f)$. Fix a vector $\overline{x}$ in the relative interior of $\text{dom}(f)$. Consider the set

$$C = \big\{(z,\nu) \mid z \in \text{dom}(f),\ f(x) + f'(x; z-x) < \nu\big\},$$

and the halfline

$$P = \big\{(u,\zeta) \mid u = x + \beta(\overline{x}-x),\, \zeta = f(x) + \beta f'(x; \overline{x}-x),\ \beta \geq 0\big\}.$$

By Exercise 4.1(b), the directional derivative function $f'(x;\cdot)$ is convex, implying that $f'(x; z-x)$ is convex in z. Therefore, the set C is convex. Furthermore, being a halfline, the set P is polyhedral.

Suppose that C and P have a point (z,ν) in common, so that we have

$$z \in \text{dom}(f), \qquad f(x) + f'(x; z-x) < \nu, \tag{4.2}$$

$$z = x + \beta(\overline{x}-x), \qquad \nu = f(x) + \beta f'(x; \overline{x}-x),$$

for some scalar $\beta \geq 0$. Because $\beta f'(x; y) = f'(x; \beta y)$ for all $\beta \geq 0$ and $y \in \Re^n$ [see Exercise 4.1(a)], it follows that

$$\nu = f(x) + f'\big(x; \beta(\overline{x}-x)\big) = f(x) + f'\big(x; z-x\big),$$

contradicting Eq. (4.2), and thus showing that C and P do not have any common point. Hence, $\text{ri}(C)$ and P do not have any common point, so by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), the polyhedral set P and the convex set C can be properly separated by a hyperplane that does not contain C, i.e., there exists a vector $(a,\gamma) \in \Re^{n+1}$ such that

$$a'z + \gamma\nu \geq a'\big(x + \beta(\overline{x}-x)\big) + \gamma\big(f(x) + \beta f'(x; \overline{x}-x)\big), \quad \forall\ (z,\nu) \in C, \quad \forall\ \beta \geq 0, \tag{4.3}$$

$$\inf_{(z,\nu)\in C} \big\{a'z + \gamma\nu\big\} < \sup_{(z,\nu)\in C} \big\{a'z + \gamma\nu\big\}, \tag{4.4}$$

We cannot have $\gamma < 0$ since then the left-hand side of Eq. (4.3) could be made arbitrarily small by choosing ν sufficiently large. Also if $\gamma = 0$, then for $\beta = 1$, from Eq. (4.3) we obtain

$$a'(z - \overline{x}) \geq 0, \qquad \forall\ z \in \text{dom}(f).$$

Since $\overline{x} \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, we have that the linear function $a'z$ attains its minimum over $\mathrm{dom}(f)$ at a point in the relative interior of $\mathrm{dom}(f)$. By Prop. 1.4.2, it follows that $a'z$ is constant over $\mathrm{dom}(f)$, i.e., $a'z = a'\overline{x}$ for all $z \in \mathrm{dom}(f)$, contradicting Eq. (4.4). Hence, we must have $\gamma > 0$ and by dividing with γ in Eq. (4.3), we obtain

$$\overline{a}'z + \nu \geq \overline{a}'\big(x + \beta(\overline{x} - x)\big) + f(x) + \beta f'(x; \overline{x} - x), \qquad \forall\ (z, \nu) \in C, \quad \forall\ \beta \geq 0,$$

where $\overline{a} = a/\gamma$. By letting $\beta = 0$ and $\nu \downarrow f(x) + f'(x; z - x)$ in this relation, and by rearranging terms, we have

$$f'(x; z - x) \geq (-\overline{a})'(z - x), \qquad \forall\ z \in \mathrm{dom}(f).$$

Because

$$f(z) - f(x) = f\big(x + (z - x)\big) - f(x) \geq \inf_{\lambda > 0} \frac{f\big(x + \lambda(z - x)\big) - f(x)}{\lambda} = f'(x; z - x),$$

it follows that

$$f(z) - f(x) \geq (-\overline{a})'(z - x), \qquad \forall\ z \in \mathrm{dom}(f).$$

Finally, by using the fact $f(z) = \infty$ for all $z \notin \mathrm{dom}(f)$, we see that

$$f(z) - f(x) \geq (-\overline{a})'(z - x), \qquad \forall\ z \in \Re^n,$$

showing that $-\overline{a}$ is a subgradient of f at x and that $\partial f(x)$ is nonempty.

4.9 (Subdifferential of Sum of Extended Real-Valued Functions)

It will suffice to prove the result for the case where $f = f_1 + f_2$. If $d_1 \in \partial f_1(x)$ and $d_2 \in \partial f_2(x)$, then by the subgradient inequality, it follows that

$$f_1(z) \geq f_1(x) + (z - x)'d_1, \qquad \forall\ z,$$

$$f_2(z) \geq f_2(x) + (z - x)'d_2, \qquad \forall\ z,$$

so by adding these inequalities, we obtain

$$f(z) \geq f(x) + (z - x)'(d_1 + d_2), \qquad \forall\ z.$$

Hence, $d_1 + d_2 \in \partial f(x)$, implying that $\partial f_1(x) + \partial f_2(x) \subset \partial f(x)$.

Assuming that $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ and $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ have a point in common, we will prove the reverse inclusion. Let $d \in \partial f(x)$, and define the functions

$$g_1(y) = f_1(x + y) - f_1(x) - d'y, \qquad \forall\ y,$$

$$g_2(y) = f_2(x + y) - f_2(x), \qquad \forall\ y.$$

Then, for the function $g = g_1 + g_2$, we have $g(0) = 0$ and by using $d \in \partial f(x)$, we obtain

$$g(y) = f(x + y) - f(x) - d'y \geq 0, \qquad \forall\ y. \tag{4.5}$$

Consider the convex sets

$$C_1 = \big\{(y, \mu) \in \Re^{n+1} \mid y \in \mathrm{dom}(g_1),\ \mu \geq g_1(y)\big\},$$

$$C_2 = \big\{(u, \nu) \in \Re^{n+1} \mid u \in \mathrm{dom}(g_2),\ \nu \leq -g_2(y)\big\},$$

and note that

$$\mathrm{ri}(C_1) = \Big\{(y, \mu) \in \Re^{n+1} \mid y \in \mathrm{ri}\big(\mathrm{dom}(g_1)\big),\ \mu > g_1(y)\Big\},$$

$$\mathrm{ri}(C_2) = \Big\{(u, \nu) \in \Re^{n+1} \mid u \in \mathrm{ri}\big(\mathrm{dom}(g_2)\big),\ \nu < -g_2(y)\Big\}.$$

Suppose that there exists a vector $(\hat{y}, \hat{\mu}) \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$. Then,

$$g_1(\hat{y}) < \hat{\mu} < -g_2(\hat{y}),$$

yielding

$$g(\hat{y}) = g_1(\hat{y}) + g_2(\hat{y}) < 0,$$

which contradicts Eq. (4.5). Therefore, the sets $\mathrm{ri}(C_1)$ and $\mathrm{ri}(C_2)$ are disjoint, and by the Proper Separation (Prop. 2.4.6), the two convex sets C_1 and C_2 can be properly separated, i.e., there exists a vector $(w, \gamma) \in \Re^{n+1}$ such that

$$\inf_{(y,\mu)\in C_1} \{w'y + \gamma\mu\} \geq \sup_{(u,\nu)\in C_2} \{w'u + \gamma\nu\}, \tag{4.6}$$

$$\sup_{(y,\mu)\in C_1} \{w'y + \gamma\mu\} > \inf_{(u,\nu)\in C_2} \{w'u + \gamma\nu\}.$$

We cannot have $\gamma < 0$, because by letting $\mu \to \infty$ in Eq. (4.6), we will obtain a contradiction. Thus, we must have $\gamma \geq 0$. If $\gamma = 0$, then the preceding relations reduce to

$$\inf_{y\in \mathrm{dom}(g_1)} w'y \geq \sup_{u\in \mathrm{dom}(g_2)} w'u,$$

$$\sup_{y\in \mathrm{dom}(g_1)} w'y > \inf_{u\in \mathrm{dom}(g_2)} w'u,$$

which in view of the fact

$$\mathrm{dom}(g_1) = \mathrm{dom}(f_1) - x, \qquad \mathrm{dom}(g_2) = \mathrm{dom}(f_2) - x,$$

imply that $\mathrm{dom}(f_1)$ and $\mathrm{dom}(f_2)$ are properly separated. But this is impossible since $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ and $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ have a point in common. Hence $\gamma > 0$, and by dividing in Eq. (4.6) with γ and by setting $b = w/\gamma$, we obtain

$$\inf_{(y,\mu)\in C_1} \big\{b'y + \mu\big\} \geq \sup_{(u,\nu)\in C_2} \big\{b'u + \nu\big\}.$$

Since $g_1(0) = 0$ and $g_2(0) = 0$, we have $(0,0) \in C_1 \cap C_2$, implying that

$$b'y + \mu \geq 0 \geq b'u + \nu, \qquad \forall\ (y,\mu) \in C_1, \quad \forall\ (u,\nu) \in C_2.$$

Therefore, for $\mu = g_1(y)$ and $\nu = -g_2(u)$, we obtain

$$g_1(y) \geq -b'y, \qquad \forall\ y \in \mathrm{dom}(g_1),$$

$$g_2(u) \geq b'\nu, \qquad \forall\ u \in \mathrm{dom}(g_2),$$

and by using the definitions of g_1 and g_2, we see that

$$f_1(x+y) \geq f_1(x) + (d-b)'y, \qquad \text{for all } y \text{ with } x+y \in \mathrm{dom}(f_1),$$

$$f_2(x+u) \geq f_2(x) + b'u, \qquad \text{for all } u \text{ with } x+u \in \mathrm{dom}(f_2).$$

Hence,

$$f_1(z) \geq f_1(x) + (d-b)'(z-x), \qquad \forall\ z,$$

$$f_2(z) \geq f_2(x) + b'(z-x), \qquad \forall\ z,$$

so that $d - b \in \partial f_1(x)$ and $b \in \partial f_2(x)$, showing that $d \in \partial f_1(x) + \partial f_2(x)$ and $\partial f(x) \subset \partial f_1(x) + \partial f_2(x)$.

When some of the functions are polyhedral, we use a different separation argument for C_1 and C_2. In particular, since the sum of polyhedral functions is a polyhedral function (see Exercise 3.12), it will still suffice to consider the case $m = 2$. Thus, let f_1 be a convex function, and let f_2 be a polyhedral function such that

$$\mathrm{ri}\big(\mathrm{dom}(f_1)\big) \cap \mathrm{dom}(f_2) \neq \emptyset.$$

Then, in the preceding proof, g_2 is a polyhedral function and C_2 is a polyhedral set. Furthermore, $\mathrm{ri}(C_1)$ and C_2 are disjoint, for otherwise we would have for some $(\hat{y},\hat{\mu}) \in \mathrm{ri}(C_1) \cap C_2$,

$$g_1(\hat{y}) < \hat{\mu} \leq -g_2(\hat{y}),$$

implying that $g(\hat{y}) = g_1(\hat{y}) + g_2(\hat{y}) < 0$ and contradicting Eq. (4.5). Therefore, by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), the convex set C_1 and the polyhedral set C_2 can be properly separated by a hyperplane that does not contain C_1, i.e., there exists a vector $(w,\gamma) \in \Re^{n+1}$ such that

$$\inf_{(y,\mu)\in C_1} \big\{w'y + \gamma\mu\big\} \geq \sup_{(u,\nu)\in C_2} \big\{w'u + \gamma\nu\big\},$$

$$\inf_{(y,\mu)\in C_1} \big\{w'y + \gamma\mu\big\} < \sup_{(y,\mu)\in C_1} \big\{w'y + \gamma\mu\big\}.$$

We cannot have $\gamma < 0$, because by letting $\mu \to \infty$ in the first of the preceding relations, we will obtain a contradiction. Thus, we must have $\gamma \geq 0$. If $\gamma = 0$, then the preceding relations reduce to

$$\inf_{y\in \mathrm{dom}(g_1)} w'y \geq \sup_{u\in \mathrm{dom}(g_2)} w'u,$$

$$\inf_{y\in \text{dom}(g_1)} w'y < \sup_{y\in \text{dom}(g_1)} w'y.$$

In view of the fact

$$\text{dom}(g_1) = \text{dom}(f_1) - x, \qquad \text{dom}(g_2) = \text{dom}(f_2) - x,$$

it follows that $\text{dom}(f_1)$ and $\text{dom}(f_2)$ are properly separated by a hyperplane that does not contain $\text{dom}(f_1)$, while $\text{dom}(f_2)$ is polyhedral since f_2 is polyhedral [see Prop. 3.2.3). Therefore, by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), we have that $\text{ri}\big(\text{dom}(f_1)\big) \cap \text{dom}(f_2) = \emptyset$, which is a contradiction. Hence $\gamma > 0$, and the remainder of the proof is similar to the preceding one.

4.10 (Chain Rule for Extended Real-Valued Functions)

We note that $\text{dom}(F)$ is nonempty since in contains the inverse image under A of the common point of the range of A and the relative interior of $\text{dom}(f)$. In particular, F is proper. We fix an x in $\text{dom}(F)$. If $d \in A'\partial f(Ax)$, there exists a $g \in \partial f(Ax)$ such that $d = A'g$. We have for all $z \in \Re^m$,

$$\begin{aligned} F(z) - F(x) - (z-x)'d &= f(Az) - f(Ax) - (z-x)'A'g \\ &= f(Az) - f(Ax) - (Az - Ax)'g \\ &\geq 0, \end{aligned}$$

where the inequality follows from the fact $g \in \partial f(Ax)$. Hence, $d \in \partial F(x)$, and we have $A'\partial f(Ax) \subset \partial F(x)$.

We next show the reverse inclusion. By using a translation argument if necessary, we may assume that $x = 0$ and $F(0) = 0$. Let $d \in \partial F(0)$. Then we have

$$F(z) - z'd \geq 0, \qquad \forall\ z \in \Re^n,$$

or

$$f(Az) - z'd \geq 0, \qquad \forall\ z \in \Re^n,$$

or

$$f(y) - z'd \geq 0, \qquad \forall\ z \in \Re^n,\ y = Az,$$

or

$$H(y,z) \geq 0, \qquad \forall\ z \in \Re^n,\ y = Az,$$

where the function $H : \Re^m \times \Re^n \mapsto (-\infty, \infty]$ has the form

$$H(y,z) = f(y) - z'd.$$

Since the range of A contains a point in $\text{ri}\big(\text{dom}(f)\big)$, and $\text{dom}(H) = \text{dom}(f) \times \Re^n$, we see that the set $\big\{(y,z) \in \text{dom}(H) \mid y = Az\big\}$ contains a point in the relative interior of $\text{dom}(H)$. Hence, we can apply the Nonlinear Farkas' Lemma [part (b)] with the following identification:

$$x = (y,z), \qquad C = \text{dom}(H), \qquad g_1(y,z) = Az - y, \qquad g_2(y,z) = y - Az.$$

In this case, we have

$$\left\{x \in C \mid g_1(x) \leq 0,\ g_2(x) \leq 0\right\} = \left\{(y,z) \in \mathrm{dom}(H) \mid Az - y = 0\right\}.$$

As asserted earlier, this set contains a relative interior point of C, thus implying that the set

$$Q^* = \left\{\mu \geq 0 \mid H(y,z) + \mu_1' g_1(y,z) + \mu_2' g_2(y,z) \geq 0,\ \forall\ (y,z) \in \mathrm{dom}(H)\right\}$$

is nonempty. Hence, there exists (μ_1, μ_2) such that

$$f(y) - z'd + (\mu_1 - \mu_2)'(Az - y) \geq 0, \qquad \forall\ (y,z) \in \mathrm{dom}(H).$$

Since $\mathrm{dom}(H) = \Re^m \times \Re^n$, by letting $\lambda = \mu_1 - \mu_2$, we obtain

$$f(y) - z'd + \lambda'(Az - y) \geq 0, \qquad \forall\ y \in \Re^m,\ z \in \Re^n,$$

or equivalently

$$f(y) \geq \lambda' y + z'(d - A'\lambda), \qquad \forall\ y \in \Re^m,\ z \in \Re^n.$$

Because this relation holds for all z, we have $d = A'\lambda$ implying that

$$f(y) \geq \lambda' y, \qquad \forall\ y \in \Re^m,$$

which shows that $\lambda \in \partial f(0)$. Hence $d \in A'\partial f(0)$, thus completing the proof.

4.11

Suppose that the set $\cup_{x\in X}\partial_\epsilon f(x)$ is unbounded for some $\epsilon > 0$. Then, there exist a sequence $\{x_k\} \subset X$, and a sequence $\{d_k\}$ such that $d_k \in \partial_\epsilon f(x_k)$ for all k and $\|d_k\| \to \infty$. Without loss of generality, we may assume that $d_k \neq 0$ for all k, and we denote $y_k = d_k/\|d_k\|$. Since both $\{x_k\}$ and $\{y_k\}$ are bounded, they must contain convergent subsequences, and without loss of generality, we may assume that x_k converges to some x and y_k converges to some y with $\|y\| = 1$. Since $d_k \in \partial_\epsilon f(x_k)$ for all k, it follows that

$$f(x_k + y_k) \geq f(x_k) + d_k' y_k - \epsilon = f(x_k) + \|d_k\| - \epsilon.$$

By letting $k \to \infty$ and by using the continuity of f, we obtain $f(x + y) = \infty$, a contradiction. Hence, the set $\cup_{x\in X}\partial_\epsilon f(x)$ must be bounded for all $\epsilon > 0$.

4.12

Let $d \in \partial f(x)$. Then, by the definitions of subgradient and ϵ-subgradient, it follows that for any $\epsilon > 0$,

$$f(y) \geq f(x) + d'(y - x) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n,$$

implying that $d \in \partial_\epsilon f(x)$ for all $\epsilon > 0$. Therefore $d \in \cap_{\epsilon>0}\partial_\epsilon f(x)$, showing that $\partial f(x) \subset \cap_{\epsilon>0}\partial_\epsilon f(x)$.

Conversely, let $d \in \partial_\epsilon f(x)$ for all $\epsilon > 0$, so that

$$f(y) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n, \quad \forall\ \epsilon > 0.$$

By letting $\epsilon \downarrow 0$, we obtain

$$f(y) \geq f(x) + d'(y - x), \qquad \forall\ y \in \Re^n,$$

implying that $d \in \partial f(x)$, and showing that $\cap_{\epsilon>0}\partial_\epsilon f(x) \subset \partial f(x)$.

4.13 (Continuity Properties of ϵ-Subdifferential [Nur77])

(a) By the ϵ-subgradient definition, we have for all k,

$$f(y) \geq f(x) + d_k'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n.$$

Since the sequence $\{x_k\}$ is bounded, by Exercise 4.11, the sequence $\{d_k\}$ is also bounded and therefore, it has a limit point d. Taking the limit in the preceding relation along a subsequence of $\{d_k\}$ converging to d, we obtain

$$f(y) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n,$$

showing that $d \in \partial_\epsilon f(x)$.

(b) First we show that

$$\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big) = \partial_\epsilon f(x). \tag{4.7}$$

Let $d \in \partial_\delta f(x)$ for some scalar δ satisfying $0 < \delta < \epsilon$. Then, by the definition of ϵ-subgradient, we have

$$f(y) \geq f(x) - d'(y - x) - \delta \geq f(x) - d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n,$$

showing that $d \in \partial_\epsilon f(x)$. Therefore,

$$\partial_\delta f(x) \subset \partial_\epsilon f(x), \qquad \forall\ \delta \in (0, \epsilon), \tag{4.8}$$

implying that

$$\cup_{0<\delta<\epsilon} \partial_\delta f(x) \subset \partial_\epsilon f(x).$$

Since $\partial_\epsilon f(x)$ is closed, by taking the closures of both sides in the preceding relation, we obtain

$$\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big) \subset \partial_\epsilon f(x).$$

Conversely, assume to arrive at a contradiction that there is a vector $d \in \partial_\epsilon f(x)$ with $d \notin \mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big)$. Note that the set $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is bounded since it is contained in the compact set $\partial_\epsilon f(x)$. Furthermore, we claim that $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is convex. Indeed if d_1 and d_2 belong to this set, then $d_1 \in \partial_{\delta_1} f(x)$ and $d_2 \in \partial_{\delta_2} f(x)$ for some positive scalars δ_1 and δ_2. Without loss of generality, let $\delta_1 \leq \delta_2$. Then, by Eq. (4.8), it follows that $d_1, d_2 \in \partial_{\delta_2} f(x)$, which is a convex set by Prop. 4.3.1(a). Hence, $\lambda d_1 + (1 - \lambda) d_2 \in \partial_{\delta_2} f(x)$ for all $\lambda \in [0, 1]$, implying that $\lambda d_1 + (1 - \lambda) d_2 \in \cup_{0<\delta<\epsilon} \partial_\delta f(x)$ for all $\lambda \in [0, 1]$, and showing that the set $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is convex.

The vector d and the convex and compact set $\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big)$ can be strongly separated (see Exercise 2.17), i.e., there exists a vector $b \in \Re^n$ such that

$$b'd > \max_{g \in \mathrm{cl}\left(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\right)} b'g.$$

This relation implies that for some positive scalar β,

$$b'd > \max_{g \in \partial_\delta f(x)} b'g + 2\beta, \qquad \forall\ \delta \in (0, \epsilon).$$

By Prop. 4.3.1(a), we have

$$\inf_{\alpha>0} \frac{f(x+\alpha b) - f(x) + \delta}{\alpha} = \max_{g \in \partial_\delta f(x)} b'g,$$

so that

$$b'd > \inf_{\alpha>0} \frac{f(x+\alpha b) - f(x) + \delta}{\alpha} + 2\beta, \qquad \forall\ \delta,\ 0 < \delta < \epsilon.$$

Let $\{\delta_k\}$ be a positive scalar sequence converging to ϵ. In view of the preceding relation, for each δ_k, there exists a small enough $\alpha_k > 0$ such that

$$\alpha_k b'd \geq f(x+\alpha_k b) - f(x) + \delta_k + \beta. \tag{4.9}$$

Without loss of generality, we may assume that $\{\alpha_k\}$ is bounded, so that it has a limit point $\overline{\alpha} \geq 0$. By taking the limit in Eq. (4.9) along an appropriate subsequence, and by using $\delta_k \to \epsilon$, we obtain

$$\overline{\alpha} b'd \geq f(x+\overline{\alpha} b) - f(x) + \epsilon + \beta.$$

If $\overline{\alpha} = 0$, we would have $0 \geq \epsilon + \beta$, which is a contradiction. If $\overline{\alpha} > 0$, we would have

$$\overline{\alpha} b'd + f(x) - \epsilon > f(x+\overline{\alpha} b),$$

which cannot hold since $d \in \partial_\epsilon f(x)$. Hence, we must have

$$\partial_\epsilon f(x) \subset \mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big),$$

thus completing the proof of Eq. (4.7).

We now prove the statement of the exercise. Let $\{x_k\}$ be a sequence converging to x. By Prop. 4.3.1(a), the ϵ-subdifferential $\partial_\epsilon f(x)$ is bounded, so that there exists a constant $L > 0$ such that

$$\|g\| \leq L, \qquad \forall\ g \in \partial_\epsilon f(x).$$

Let

$$\gamma_k = \big|f(x_k) - f(x)\big| + L\ \|x_k - x\|, \qquad \forall\ k. \tag{4.10}$$

Since $x_k \to x$, by continuity of f, it follows that $\gamma_k \to 0$ as $k \to \infty$, so that $\epsilon_k = \epsilon - \gamma_k$ converges to ϵ. Let $\{k_i\} \subset \{0, 1, \ldots\}$ be an index sequence such that $\{\epsilon_{k_i}\}$ is positive and monotonically increasing to ϵ, i.e.,

$$\epsilon_{k_i} \uparrow \epsilon \qquad \text{with} \quad \epsilon_{k_i} = \epsilon - \gamma_{k_i} > 0, \quad \epsilon_{k_i} < \epsilon_{k_{i+1}}, \quad \forall\ i.$$

In view of relation (4.7), we have

$$\mathrm{cl}\big(\cup_{i\geq 0}\ \partial_{\epsilon_{k_i}} f(x)\big) = \partial_\epsilon f(x), \tag{4.11}$$

implying that for a given vector $d \in \partial_\epsilon f(x)$, there exists a sequence $\{d_{k_i}\}$ such that

$$d_{k_i} \to d \qquad \text{with} \quad d_{k_i} \in \partial_{\epsilon_{k_i}} f(x), \quad \forall\ i. \tag{4.12}$$

There remains to show that $d_{k_i} \in \partial_\epsilon f(x_{k_i})$ for all i. Since $d_{k_i} \in \partial_{\epsilon_{k_i}} f(x)$, it follows that for all i and $y \in \Re^n$,

$$\begin{aligned} f(y) &\geq f(x) + d_{k_i}'(y - x) - \epsilon_{k_i} \\ &= f(x_{k_i}) + \big(f(x) - f(x_{k_i})\big) + d_{k_i}'(y - x_{k_i}) + d_{k_i}'(x_{k_i} - x) - \epsilon_{k_i} \\ &\geq f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - \Big(\big|f(x) - f(x_{k_i})\big| + \big|d_{k_i}'(x_{k_i} - x)\big| + \epsilon_{k_i}\Big). \end{aligned} \tag{4.13}$$

Because $d_{k_i} \in \partial_\epsilon f(x)$ [cf. Eqs. (4.11) and (4.12)] and $\partial_\epsilon f(x)$ is bounded, there holds

$$\big|d_{k_i}'(x_{k_i} - x)\big| \leq L\, \|x_{k_i} - x\|.$$

Using this relation, the definition of γ_k [cf. Eq. (4.10)], and the fact $\epsilon_k = \epsilon - \gamma_k$ for all k, from Eq. (4.13) we obtain for all i and $y \in \Re^n$,

$$f(y) \geq f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - (\gamma_{k_i} + \epsilon_{k_i}) = f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - \epsilon.$$

Hence $d_{k_i} \in \partial_\epsilon f(x_{k_i})$ for all i, thus completing the proof.

4.14 (Subgradient Mean Value Theorem)

(a) **Scalar Case:** Define the scalar function $g : \Re \mapsto \Re$ by

$$g(t) = \varphi(t) - \varphi(a) - \frac{\varphi(b) - \varphi(a)}{b - a}(t - a),$$

and note that g is convex and $g(a) = g(b) = 0$. We first show that g attains its minimum over $\Re$ at some point $t^* \in [a, b]$. For $t < a$, we have

$$a = \frac{b - a}{b - t}t + \frac{a - t}{b - t}b,$$

and by using convexity of g and $g(a) = g(b) = 0$, we obtain

$$0 = g(a) \leq \frac{b - a}{b - t}g(t) + \frac{a - t}{b - t}g(b) = \frac{b - a}{b - t}g(t),$$

implying that $g(t) \geq 0$ for $t < a$. Similarly, for $t > b$, we have

$$b = \frac{b - a}{t - a}t + \frac{t - b}{t - a}a,$$

and by using convexity of g and $g(a) = g(b) = 0$, we obtain

$$0 = g(b) \leq \frac{b - a}{t - a}g(t) + \frac{t - b}{t - a}g(a) = \frac{t - b}{t - a}g(t),$$

implying that $g(t) \geq 0$ for $t > b$. Therefore $g(t) \geq 0$ for $t \notin (a, b)$, while $g(a) = g(b) = 0$. Hence

$$\min_{t \in \Re} g(t) = \min_{t \in [a,b]} g(t). \tag{4.14}$$

Because g is convex over $\Re$, it is also continuous over $\Re$, and since $[a,b]$ is compact, the set of minimizers of g over $[a,b]$ is nonempty. Thus, in view of Eq. (4.14), there exists a scalar $t^* \in [a,b]$ such that $g(t^*) = \min_{t\in\Re} g(t)$. If $t^* \in (a,b)$, then we are done. If $t^* = a$ or $t^* = b$, then since $g(a) = g(b) = 0$, it follows that every $t \in [a,b]$ attains the minimum of g over $\Re$, so that we can replace t^* by a point in the interval (a,b). Thus, in any case, there exists $t^* \in (a,b)$ such that $g(t^*) = \min_{t\in\Re} g(t)$.

We next show that

$$\frac{\varphi(b)-\varphi(a)}{b-a} \in \partial\varphi(t^*).$$

The function g is the sum of the convex function φ and the linear (and therefore smooth) function $-\frac{\varphi(b)-\varphi(a)}{b-a}(t-a)$. Thus the subdifferential of $\partial g(t^*)$ is the sum of the sudifferential of $\partial\varphi(t^*)$ and the gradient $-\frac{\varphi(b)-\varphi(a)}{b-a}$ (see Prop. 4.2.4),

$$\partial g(t^*) = \partial\varphi(t^*) - \frac{\varphi(b)-\varphi(a)}{b-a}.$$

Since t^* minimizes g over $\Re$, by the optimality condition, we have $0 \in \partial g(t^*)$. This and the preceding relation imply that

$$\frac{\varphi(b)-\varphi(a)}{b-a} \in \partial\varphi(t^*).$$

(b) **Vector Case:** Let x and y be any two vectors in $\Re^n$. If $x = y$, then $f(y) = f(x) + d'(y-x)$ trivially holds for any $d \in \partial f(x)$, and we are done. So assume that $x \neq y$, and consider the scalar function φ given by

$$\varphi(t) = f(x_t), \qquad x_t = tx + (1-t)y, \qquad t \in \Re.$$

By part (a), where $a = 0$ and $b = 1$, there exists $\alpha \in (0,1)$ such that

$$\varphi(1) - \varphi(0) \in \partial\varphi(\alpha),$$

while by Exercise 4.6, we have

$$\partial\varphi(\alpha) = \big\{ d'(x-y) \mid d \in \partial f(x_\alpha) \big\}.$$

Since $\varphi(1) = f(x)$ and $\varphi(0) = f(y)$, we see that

$$f(x) - f(y) \in \big\{ d'(x-y) \mid d \in \partial f(x_\alpha) \big\}.$$

Therefore, there exists $d \in \partial f(x_\alpha)$ such that $f(y) - f(x) = d'(y-x)$.

4.15 (Steepest Descent Direction of a Convex Function)

Note that the problem statement in the book contains a typo: $\bar{d}/\|\bar{d}\|$ should be replaced by $-\bar{d}/\|\bar{d}\|$.

The sets $\big\{d \mid \|d\| \le 1\big\}$ and $\partial f(x)$ are compact, and the function $\phi(d,g) = d'g$ is linear in each variable when the other variable is fixed, so that $\phi(\cdot,g)$ is convex and closed for all g, while the function $-\phi(d,\cdot)$ is convex and closed for all d. Thus, by Prop. 2.6.9, the order of min and max can be interchanged,

$$\min_{\|d\|\le 1} \max_{g\in\partial f(x)} d'g = \max_{g\in\partial f(x)} \min_{\|d\|\le 1} d'g,$$

and there exist associated saddle points.

By Prop. 4.2.2, we have $f'(x;d) = \max_{g\in\partial f(x)} d'g$, so

$$\min_{\|d\|\le 1} \max_{g\in\partial f(x)} d'g = \min_{\|d\|\le 1} f'(x;d). \tag{4.15}$$

We also have for all g,

$$\min_{\|d\|\le 1} d'g = -\|g\|,$$

and the minimum is attained for $d = -g/\|g\|$. Thus

$$\max_{g\in\partial f(x)} \min_{\|d\|\le 1} d'g = \max_{g\in\partial f(x)} \big(-\|g\|\big) = -\min_{g\in\partial f(x)} \|g\|. \tag{4.16}$$

From the generic characterization of a saddle point (cf. Prop. 2.6.1), it follows that the set of saddle points of $d'g$ is $D^* \times G^*$, where D^* is the set of minima of $f'(x;d)$ subject to $\|d\| \le 1$ [cf. Eq. (4.15)], and G^* is the set of minima of $\|g\|$ subject to $g \in \partial f(x)$ [cf. Eq. (4.16)], i.e., G^* consists of the unique vector g^* of minimum norm on $\partial f(x)$. Furthermore, again by Prop. 2.6.1, every $d^* \in D^*$ must minimize $d'g^*$ subject to $\|d\| \le 1$, so it must satisfy $d^* = -g^*/\|g^*\|$.

4.16 (Generating Descent Directions of Convex Functions)

Suppose that the process does not terminate in a finite number of steps, and let $\big\{(w_k, g_k)\big\}$ be the sequence generated by the algorithm. Since w_k is the projection of the origin on the set $\text{conv}\{g_1,\ldots,g_{k-1}\}$, by the Projection Theorem (Prop. 2.2.1), we have

$$(g - w_k)'w_k \ge 0, \qquad \forall\ g \in \text{conv}\{g_1,\ldots,g_{k-1}\},$$

implying that

$$g_i'w_k \ge \|w_k\|^2 \ge \|g^*\|^2 > 0, \qquad \forall\ i = 1,\ldots,k-1, \quad \forall\ k \ge 1, \tag{4.17}$$

where $g^* \in \partial f(x)$ is the vector with minimum norm in $\partial f(x)$. Note that $\|g^*\| > 0$ because x does not minimize f. The sequences $\{w_k\}$ and $\{g_k\}$ are contained in $\partial f(x)$, and since $\partial f(x)$ is compact, $\{w_k\}$ and $\{g_k\}$ have limit points in $\partial f(x)$.

Without loss of generality, we may assume that these sequences converge, so that for some $\hat{w}, \hat{g} \in \partial f(x)$, we have

$$\lim_{k\to\infty} w_k = \hat{w}, \qquad \lim_{k\to\infty} g_k = \hat{g},$$

which in view of Eq. (4.17) implies that $\hat{g}'\hat{w} > 0$. On the other hand, because none of the vectors $(-w_k)$ is a descent direction of f at x, we have $f'(x; -w_k) \geq 0$, so that

$$g_k'(-w_k) = \max_{g\in\partial f(x)} g'(-w_k) = f'(x; -w_k) \geq 0.$$

By letting $k \to \infty$, we obtain $\hat{g}'\hat{w} \leq 0$, thus contradicting $\hat{g}'\hat{w} > 0$. Therefore, the process must terminate in a finite number of steps with a descent direction.

4.17 (Generating ϵ-Descent Directions of Convex Functions [Lem74])

Suppose that the process does not terminate in a finite number of steps, and let $\big\{(w_k, g_k)\big\}$ be the sequence generated by the algorithm. Since w_k is the projection of the origin on the set $\text{conv}\{g_1, \ldots, g_{k-1}\}$, by the Projection Theorem (Prop. 2.2.1), we have

$$(g - w_k)'w_k \geq 0, \qquad \forall\ g \in \text{conv}\{g_1, \ldots, g_{k-1}\},$$

implying that

$$g_i'w_k \geq \|w_k\|^2 \geq \|g^*\|^2 > 0, \qquad \forall\ i = 1, \ldots, k-1, \quad \forall\ k \geq 1, \tag{4.18}$$

where $g^* \in \partial_\epsilon f(x)$ is the vector with minimum norm in $\partial_\epsilon f(x)$. Note that $\|g^*\| > 0$ because x is not an ϵ-optimal solution, i.e., $f(x) > \inf_{z\in\Re^n} f(z) + \epsilon$ [see Prop. 4.3.1(b)]. The sequences $\{w_k\}$ and $\{g_k\}$ are contained in $\partial_\epsilon f(x)$, and since $\partial_\epsilon f(x)$ is compact [Prop. 4.3.1(a)], $\{w_k\}$ and $\{g_k\}$ have limit points in $\partial_\epsilon f(x)$. Without loss of generality, we may assume that these sequences converge, so that for some $\hat{w}, \hat{g} \in \partial_\epsilon f(x)$, we have

$$\lim_{k\to\infty} w_k = \hat{w}, \qquad \lim_{k\to\infty} g_k = \hat{g},$$

which in view of Eq. (4.18) implies that $\hat{g}'\hat{w} > 0$. On the other hand, because none of the vectors $(-w_k)$ is an ϵ-descent direction of f at x, by Prop. 4.3.1(a), we have

$$g_k'(-w_k) = \max_{g\in\partial f(x)} g'(-w_k) = \inf_{\alpha>0} \frac{f(x - \alpha w_k) - f(x) + \epsilon}{\alpha} \geq 0.$$

By letting $k \to \infty$, we obtain $\hat{g}'\hat{w} \leq 0$, thus contradicting $\hat{g}'\hat{w} > 0$. Hence, the process must terminate in a finite number of steps with an ϵ-descent direction.

4.18

(a) For $x \in \mathrm{int}(C)$, we clearly have $F_C(x) = \Re^n$, implying that

$$T_C(x) = \Re^n.$$

Since C is convex, by Prop. 4.6.3, we have

$$N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \mathrm{int}(C)$, we have $\|x\| = 1$. By the definition of the set $F_C(x)$ of feasible directions at x, we have $y \in F_C(x)$ if and only if $x + \alpha y \in C$ for all sufficiently small positive scalars α. Thus, $y \in F_C(x)$ if and only if there exists $\overline{\alpha} > 0$ such that $\|x + \alpha y\|^2 \leq 1$ for all α with $0 < \alpha \leq \overline{\alpha}$, or equivalently

$$\|x\|^2 + 2\alpha x'y + \alpha^2 \|y\|^2 \leq 1, \qquad \forall\ \alpha,\ 0 < \alpha \leq \overline{\alpha}.$$

Since $\|x\| = 1$, the preceding relation reduces to

$$2x'y + \alpha \|y\|^2 \leq 0. \qquad \forall\ \alpha,\ 0 < \alpha \leq \overline{\alpha}.$$

This relation holds if and only if $y = 0$, or $x'y < 0$ and $\alpha \leq -2x'y/\|y\|^2$ (i.e., $\overline{\alpha} = -2x'y/\|y\|^2$). Therefore,

$$F_C(x) = \big\{y \mid x'y < 0\big\} \cup \{0\}.$$

Because C is convex, by Prop. 4.6.2(c), we have $T_C(x) = \mathrm{cl}\big(F_C(x)\big)$, implying that

$$T_C(x) = \big\{y \mid x'y \leq 0\big\}.$$

Furthermore, by Prop. 4.6.3, we have $N_C(x) = T_C(x)^*$, while by the Farkas' Lemma [Prop. 3.2.1(b)], $T_C(x)^* = \mathrm{cone}\big(\{x\}\big)$, implying that

$$N_C(x) = \mathrm{cone}\big(\{x\}\big).$$

(b) If C is a subspace, then clearly $F_C(x) = C$ for all $x \in C$. Because C is convex, by Props. 4.6.2(a) and 4.6.3, we have

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = C, \qquad N_C(x) = T_C(x)^* = C^\perp, \qquad \forall\ x \in C.$$

(c) Let C be a closed halfspace given by $C = \big\{x \mid a'x \leq b\big\}$ with a nonzero vector $a \in \Re^n$ and a scalar b. For $x \in \mathrm{int}(C)$, i.e., $a'x < b$, we have $F_C(x) = \Re^n$ and since C is convex, by Props. 4.6.2(a) and 4.6.3, we have

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = \Re^n, \qquad N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \mathrm{int}(C)$, we have $a'x = b$, so that $x + \alpha y \in C$ for some $y \in \Re^n$ and $\alpha > 0$ if and only if $a'y \leq 0$, implying that

$$F_C(x) = \big\{y \mid a'y \leq 0\big\}.$$

By Prop. 4.6.2(a), it follows that

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = \big\{y \mid a'y \le 0\big\},$$

while by Prop. 4.6.3 and the Farkas' Lemma [Prop. 3.2.1(b)], it follows that

$$N_C(x) = T_C(x)^* = \mathrm{cone}\big(\{a\}\big).$$

(d) For $x \in C$ with $x \in \mathrm{int}(C)$, i.e., $x_i > 0$ for all $i \in I$, we have $F_C(x) = \Re^n$. Then, by using Props. 4.6.2(a) and 4.6.3, we obtain

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = \Re^n, \qquad N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \mathrm{int}(C)$, the set $A_x = \{i \in I \mid x_i = 0\}$ is nonempty. Then, $x + \alpha y \in C$ for some $y \in \Re^n$ and $\alpha > 0$ if and only if $y_i \le 0$ for all $i \in A_x$, implying that

$$F_C(x) = \{y \mid y_i \le 0, \ \ \forall\, i \in A_x\}.$$

Because C is convex, by Prop. 4.6.2(a), we have that

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = \{y \mid y_i \le 0, \ \ \forall\, i \in A_x\},$$

or equivalently

$$T_C(x) = \big\{y \mid y'e_i \le 0, \ \ \forall\, i \in A_x\big\},$$

where $e_i \in \Re^n$ is the vector whose ith component is 1 and all other components are 0. By Prop. 4.6.3, we further have $N_C(x) = T_C(x)^*$, while by the Farkas' Lemma [Prop. propforea(b)], we see that $T_C(x)^* = \mathrm{cone}\big(\{e_i \mid i \in A_x\}\big)$, implying that

$$N_C(x) = \mathrm{cone}\big(\{e_i \mid i \in A_x\}\big).$$

4.19

(a) $\Rightarrow$ (b) Let $x \in \mathrm{ri}(C)$ and let S be the subspace that is parallel to $\mathrm{aff}(C)$. Then, for every $y \in S$, $x + \alpha y \in \mathrm{ri}(C)$ for all sufficiently small positive scalars α, implying that $y \in F_C(x)$ and showing that $S \subset F_C(x)$. Furthermore, by the definition of the set of feasible directions, it follows that if $y \in F_C(x)$, then there exists $\overline{\alpha} > 0$ such that $x + \alpha y \in C$ for all $\alpha \in (0, \overline{\alpha}]$. Hence $y \in S$, implying that $F_C(x) \subset S$. This and the relation $S \subset F_C(x)$ show that $F_C(x) = S$. Since C is convex, by Prop. 4.6.2(a), it follows that

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = S,$$

thus proving that $T_C(x)$ is a subspace.

(b) $\Rightarrow$ (c) Let $T_C(x)$ be a subspace. Then, because C is convex, from Prop. 4.6.3 it follows that

$$N_C(x) = T_C(x)^* = T_C(x)^\perp,$$

showing that $N_C(x)$ is a subspace.

(c) $\Rightarrow$ (a) Let $N_C(x)$ be a subspace, and to arrive at a contradiction suppose that x is not a point in the relative interior of C. Then, by the Proper Separation Theorem (Prop. 2.4.5), the point x and the relative interior of C can be properly separated, i.e., there exists a vector $a \in \Re^n$ such that

$$\sup_{y \in C} a'y \leq a'x, \tag{4.19}$$

$$\inf_{y \in C} a'y < \sup_{y \in C} a'y. \tag{4.20}$$

The relation (4.19) implies that

$$(-a)'(x - y) \leq 0, \qquad \forall\ y \in C. \tag{4.21}$$

Since C is convex, by Prop. 4.6.3, the preceding relation is equivalent to $-a \in T_C(x)^*$. By the same proposition, there holds $N_C(x) = T_C(x)^*$, implying that $-a \in N_C(x)$. Because $N_C(x)$ is a subspace, we must also have $a \in N_C(x)$, and by using

$$N_C(x) = T_C(x)^* = \big\{z \mid z'(x - y) \leq 0,\ \ \forall\ y \in C\big\}$$

(cf. Prop. 4.6.3), we see that

$$a'(x - y) \leq 0, \qquad \forall\ y \in C.$$

This relation and Eq. (4.21) yield

$$a'(x - y) = 0, \qquad \forall\ y \in C,$$

contradicting Eq. (4.20). Hence, x must be in the relative interior of C.

4.20 (Tangent and Normal Cones of Affine Sets)

Let $C = \{x \mid Ax = b\}$ and let $x \in C$ be arbitrary. We then have

$$F_C(x) = \{y \mid Ay = 0\} = N(A),$$

and by using Prop. 4.6.2(a), we obtain

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = N(A).$$

Since C is convex, by Prop. 4.6.3, it follows that

$$N_C(x) = T_C(x)^* = N(A)^\perp = R(A').$$

4.21 (Tangent and Normal Cones of Level Sets)

Let

$$C = \big\{z \mid f(z) \le f(x)\big\}.$$

We first show that

$$\mathrm{cl}\big(F_C(x)\big) = \big\{y \mid f'(x;y) \le 0\big\}.$$

Let $\overline{y} \in F_C(x)$ be arbitrary. Then, by the definition of $F_C(x)$, there exists a scalar $\overline{\alpha}$ such that $x + \alpha\overline{y} \in C$ for all $\alpha \in (0, \overline{\alpha}]$. By the definition of C, it follows that $f(x + \alpha\overline{y}) \le f(x)$ for all $\alpha \in (0, \overline{\alpha}]$, implying that

$$f'(x;\overline{y}) = \inf_{\alpha>0} \frac{f(x+\alpha\overline{y}) - f(x)}{\alpha} \le 0.$$

Therefore $\overline{y} \in \big\{y \mid f'(x;y) \le 0\big\}$, thus showing that

$$F_C(x) \subset \big\{y \mid f'(x;y) \le 0\big\}.$$

By Exercise 4.1(d), the set $\big\{y \mid f'(x;y) \le 0\big\}$ is closed, so that by taking closures in the preceding relation, we obtain

$$\mathrm{cl}\big(F_C(x)\big) \subset \big\{y \mid f'(x;y) \le 0\big\}.$$

To show the converse inclusion, let $\overline{y}$ be such that $f'(x;\overline{y}) < 0$, so that for all small enough $\alpha \ge 0$, we have

$$f(x + \alpha\overline{y}) - f(x) < 0.$$

Therefore $x + \alpha\overline{y} \in C$ for all small enough $\alpha \ge 0$, implying that $\overline{y} \in F_C(x)$ and showing that

$$\big\{y \mid f'(x;y) < 0\big\} \subset F_C(x).$$

By taking the closures of the sets in the preceding relation, we obtain

$$\big\{y \mid f'(x;y) \le 0\big\} = \mathrm{cl}\Big(\big\{y \mid f'(x;y) < 0\big\}\Big) \subset \mathrm{cl}\big(F_C(x)\big).$$

Hence

$$\mathrm{cl}\big(F_C(x)\big) = \big\{y \mid f'(x;y) \le 0\big\}.$$

Since C is convex, by Prop. 4.6.2(c), we have $\mathrm{cl}\big(F_C(x)\big) = T_C(x)$. This and the preceding relation imply that

$$T_C(x) = \big\{y \mid f'(x;y) \le 0\big\}.$$

Since by Prop. 4.6.3, $N_C(x) = T_C(x)^*$, it follows that

$$N_C(x) = \Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^*.$$

Furthermore, by Exercise 4.1(d), we have that

$$\Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^* = \mathrm{cl}\Big(\mathrm{cone}\big(\partial f(x)\big)\Big),$$

implying that

$$N_C(x) = \mathrm{cl}\Big(\mathrm{cone}\big(\partial f(x)\big)\Big).$$

If x does not minimize f over $\Re^n$, then the subdifferential $\partial f(x)$ does not contain the origin. Furthermore, by Prop. 4.2.1, $\partial f(x)$ is nonempty and compact, implying by Exercise 1.32(a) that the cone generated by $\partial f(x)$ is closed. Therefore, in this case, the closure operation in the preceding relation is unnecessary, i.e.,

$$N_C(x) = \mathrm{cone}\big(\partial f(x)\big).$$

4.22

It suffices to consider the case $m = 2$. From the definition of the cone of feasible directions, it can be seen that

$$F_{C_1 \times C_2}(x_1, x_2) = F_{C_1}(x_1) \times F_{C_2}(x_2).$$

By taking the closure of both sides in the preceding relation, and by using the fact that the closure of the Cartesian product of two sets coincides with the Cartesian product of their closures (see Exercise 1.37), we obtain

$$\mathrm{cl}\big(F_{C_1 \times C_2}(x_1, x_2)\big) = \mathrm{cl}\big(F_{C_1}(x_1)\big) \times \mathrm{cl}\big(F_{C_2}(x_2)\big).$$

Since C_1 and C_2 are convex, by Prop. 4.6.2(c), we have

$$T_{C_1}(x_1) = \mathrm{cl}\big(F_{C_1}(x_1)\big), \qquad T_{C_2}(x_2) = \mathrm{cl}\big(F_{C_2}(x_2)\big).$$

Furthermore, the Cartesian product $C_1 \times C_2$ is also convex, and by Prop. 4.6.2(c), we also have

$$T_{C_1 \times C_2}(x_1, x_2) = \mathrm{cl}\big(F_{C_1 \times C_2}(x_1, x_2)\big).$$

By combining the preceding three relations, we obtain

$$T_{C_1 \times C_2}(x_1, x_2) = T_{C_1}(x_1) \times T_{C_2}(x_2).$$

By taking polars in the preceding relation, we obtain

$$T_{C_1 \times C_2}(x_1, x_2)^* = \big(T_{C_1}(x_1) \times T_{C_2}(x_2)\big)^*,$$

and because the polar of the Cartesian product of two cones coincides with the Cartesian product of their polar cones (see Exercise 3.4), it follows that

$$T_{C_1 \times C_2}(x_1, x_2)^* = T_{C_1}(x_1)^* \times T_{C_2}(x_2)^*.$$

Since the sets C_1, C_2, and $C_1 \times C_2$ are convex, by Prop. 4.6.3, we have

$$T_{C_1 \times C_2}(x_1, x_2)^* = N_{C_1 \times C_2}(x_1, x_2),$$
$$T_{C_1}(x_1)^* = N_{C_1}(x_1), \qquad T_{C_2}(x_2)^* = N_{C_2}(x_2),$$

so that

$$N_{C_1 \times C_2}(x_1, x_2) = N_{C_1}(x_1) \times N_{C_2}(x_2).$$

4.23 (Tangent and Normal Cone Relations)

(a) We first show that

$$N_{C_1}(x) + N_{C_2}(x) \subset N_{C_1 \cap C_2}(x), \qquad \forall\ x \in C_1 \cap C_2.$$

For $i = 1, 2$, let $f_i(x) = 0$ when $x \in C$ and $f_i(x) = \infty$ otherwise, so that for $f = f_1 + f_2$, we have

$$f(x) = \begin{cases} 0 & \text{if } x \in C_1 \cap C_2, \\ \infty & \text{otherwise.} \end{cases}$$

By Exercise 4.4(d), we have

$$\partial f_1(x) = N_{C_1}(x), \qquad \forall\ x \in C_1,$$

$$\partial f_2(x) = N_{C_2}(x), \qquad \forall\ x \in C_2,$$

$$\partial f(x) = N_{C_1 \cap C_2}(x), \qquad \forall\ x \in C_1 \cap C_2,$$

while by Exercise 4.9, we have

$$\partial f_1(x) + \partial f_2(x) \subset \partial f(x), \qquad \forall\ x.$$

In particular, this relation holds for every $x \in \mathrm{dom}(f)$ and since $\mathrm{dom}(f) = C_1 \cap C_2$, we obtain

$$N_{C_1}(x) + N_{C_2}(x) \subset N_{C_1 \cap C_2}(x), \qquad \forall\ x \in C_1 \cap C_2. \tag{4.22}$$

If $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty, then by Exercise 4.9, we have

$$\partial f(x) = \partial f_1(x) + \partial f_2(x), \qquad \forall\ x,$$

implying that

$$N_{C_1 \cap C_2}(x) = N_{C_1}(x) + N_{C_2}(x), \qquad \forall\ x \in C_1 \cap C_2. \tag{4.23}$$

Furthermore, by Exercise 4.9, this relation also holds if C_2 is polyhedral and $\mathrm{ri}(C_1) \cap C_2$ is nonempty.

By taking polars in Eq. (4.22), it follows that

$$N_{C_1 \cap C_2}(x)^* \subset \big(N_{C_1}(x) + N_{C_2}(x)\big)^*,$$

and since

$$\big(N_{C_1}(x) + N_{C_2}(x)\big)^* = N_{C_1}(x)^* \cap N_{C_2}(x)^*$$

(see Exercise 3.4), we obtain

$$N_{C_1 \cap C_2}(x)^* \subset N_{C_1}(x)^* \cap N_{C_2}(x)^*. \tag{4.24}$$

Because C_1 and C_2 are convex, their intersection $C_1 \cap C_2$ is also convex, and by Prop. 4.6.3, we have

$$N_{C_1 \cap C_2}(x)^* = T_{C_1 \cap C_2}(x),$$

$$N_{C_1}(x)^* = T_{C_1}(x), \qquad N_{C_2}(x)^* = T_{C_2}(x).$$

In view of Eq. (4.24), it follows that

$$T_{C_1 \cap C_2}(x) \subset T_{C_1}(x) \cap T_{C_2}(x).$$

When $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty, or when $\mathrm{ri}(C_1) \cap C_2$ is nonempty and C_2 is polyhedral, by taking the polars in both sides of Eq. (4.23), it can be similarly seen that

$$T_{C_1 \cap C_2}(x) = T_{C_1}(x) \cap T_{C_2}(x).$$

(b) Let $x_1 \in C_1$ and $x_2 \in C_2$ be arbitrary. Since C_1 and C_2 are convex, the sum $C_1 + C_2$ is also convex, so that by Prop. 4.6.3, we have

$$z \in N_{C_1+C_2}(x_1+x_2) \iff z'\big((y_1+y_2)-(x_1+x_2)\big) \le 0, \ \forall\, y_1 \in C_1, \ \forall\, y_2 \in C_2, \tag{4.25}$$

$$z_1 \in N_{C_1}(x_1) \iff z_1'(y_1 - x_1) \le 0, \quad \forall\, y_1 \in C_1, \tag{4.26}$$

$$z_2 \in N_{C_2}(x_2) \iff z_2'(y_2 - x_2) \le 0, \quad \forall\, y_2 \in C_2. \tag{4.27}$$

If $z \in N_{C_1+C_2}(x_1 + x_2)$, then by using $y_2 = x_2$ in Eq. (4.25), we obtain

$$z'(y_1 - x_1) \le 0, \quad \forall\, y_1 \in C_1,$$

implying that $z \in N_{C_1}(x_1)$. Similarly, by using $y_1 = x_1$ in Eq. (4.25), we see that $z \in N_{C_2}(x_2)$. Hence $z \in N_{C_1}(x_1) \cap N_{C_2}(x_2)$ implying that

$$N_{C_1+C_2}(x_1 + x_2) \subset N_{C_1}(x_1) \cap N_{C_2}(x_2).$$

Conversely, let $z \in N_{C_1}(x_1) \cap N_{C_2}(x_2)$, so that both Eqs. (4.26) and (4.27) hold, and by adding them, we obtain

$$z'\big((y_1 + y_2) - (x_1 + x_2)\big) \le 0, \qquad \forall\, y_1 \in C_1, \quad \forall\, y_2 \in C_2.$$

Therefore, in view of Eq. (4.25), we have $z \in N_{C_1+C_2}(x_1 + x_2)$, showing that

$$N_{C_1}(x_1) \cap N_{C_2}(x_2) \subset N_{C_1+C_2}(x_1 + x_2).$$

Hence

$$N_{C_1+C_2}(x_1 + x_2) = N_{C_1}(x_1) \cap N_{C_2}(x_2).$$

By taking polars in this relation, we obtain

$$N_{C_1+C_2}(x_1 + x_2)^* = \big(N_{C_1}(x_1) \cap N_{C_2}(x_2)\big)^*.$$

Since $N_{C_1}(x_1)$ and $N_{C_2}(x_2)$ are closed convex cones, by Exercise 3.4, it follows that

$$N_{C_1+C_2}(x_1 + x_2)^* = \mathrm{cl}\big(N_{C_1}(x_1)^* + N_{C_2}(x_2)^*\big).$$

The sets C_1, C_2, and $C_1 + C_2$ are convex, so that by Prop. 4.6.3, we have

$$N_{C_1}(x_1)^* = T_{C_1}(x_1), \qquad N_{C_2}(x_2)^* = T_{C_2}(x_2),$$

$$N_{C_1+C_2}(x_1+x_2)^* = T_{C_1+C_2}(x_1+x_2),$$

implying that

$$T_{C_1+C_2}(x_1+x_2) = \mathrm{cl}\big(T_{C_1}(x_1) + T_{C_2}(x_2)\big).$$

(c) Let $x \in C$ be arbitrary. Since C is convex, its image AC under the linear transformation A is also convex, so by Prop. 4.6.3, we have

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad z'(y - Ax) \leq 0, \quad \forall\, y \in AC,$$

which is equivalent to

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad z'(Av - Ax) \leq 0, \quad \forall\, v \in C.$$

Furthermore, the condition

$$z'(Av - Ax) \leq 0, \qquad \forall\, v \in C$$

is the same as

$$(A'z)'(v - x) \leq 0, \qquad \forall\, v \in C,$$

and since C is convex, by Prop. 4.6.3, this is equivalent to $A'z \in N_C(x)$. Thus,

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad A'z \in N_C(x),$$

which together with the fact $A'z \in N_C(x)$ if and only if $z \in (A')^{-1} \cdot N_C(x)$ yields

$$N_{AC}(Ax) = (A')^{-1} \cdot N_C(x).$$

By taking polars in the preceding relation, we obtain

$$N_{AC}(Ax)^* = \big((A')^{-1} \cdot N_C(x)\big)^*. \tag{4.28}$$

Because AC is convex, by Prop. 4.6.3, we have

$$N_{AC}(Ax)^* = T_{AC}(Ax). \tag{4.29}$$

Since C is convex, by the same proposition, we have $N_C(x) = T_C(x)^*$, so that $N_C(x)$ is a closed convex cone and by using Exercise 3.5, we obtain

$$\big((A')^{-1} \cdot N_C(x)\big)^* = \mathrm{cl}\big(A \cdot N_C(x)^*\big).$$

Furthermore, by convexity of C, we also have $N_C(x)^* = T_C(x)$, implying that

$$\big((A')^{-1} \cdot N_C(x)\big)^* = \mathrm{cl}\big(A \cdot T_C(x)\big). \tag{4.30}$$

Combining Eqs. (4.28), (4.29), and (4.30), we obtain

$$T_{AC}(Ax) = \mathrm{cl}\big(A \cdot T_C(x)\big).$$

4.24 [GoT71], [RoW98]

We assume for simplicity that all the constraints are inequalities. Consider the scalar function $\theta_0 : [0,\infty) \mapsto \Re$ defined by

$$\theta_0(r) = \sup_{x \in C,\, \|x - x^*\| \le r} y'(x - x^*), \qquad r \ge 0.$$

Clearly $\theta_0(r)$ is nondecreasing and satisfies

$$0 = \theta_0(0) \le \theta_0(r), \qquad \forall\, r \ge 0.$$

Furthermore, since $y \in T_C(x^*)^*$, we have $y'(x - x^*) \le o(\|x - x^*\|)$ for $x \in C$, so that $\theta_0(r) = o(r)$, which implies that θ_0 is differentiable at $r = 0$ with $\nabla\theta_0(0) = 0$. Thus, the function F_0 defined by

$$F_0(x) = \theta_0(\|x - x^*\|) - y'(x - x^*)$$

is differentiable at x^*, attains a global minimum over C at x^*, and satisfies

$$-\nabla F_0(x^*) = y.$$

If F_0 were smooth we would be done, but since it need not even be continuous, we will successively perturb it into a smooth function. We first define the function $\theta_1 : [0,\infty) \mapsto \Re$ by

$$\theta_1(r) = \begin{cases} \frac{1}{r}\int_r^{2r} \theta_0(s)ds & \text{if } r > 0, \\ 0 & \text{if } r = 0, \end{cases}$$

(the integral above is well-defined since the function θ_0 is nondecreasing). The function θ_1 is seen to be nondecreasing and continuous, and satisfies

$$0 \le \theta_0(r) \le \theta_1(r), \qquad \forall\, r \ge 0,$$

$\theta_1(0) = 0$, and $\nabla\theta_1(0) = 0$. Thus the function

$$F_1(x) = \theta_1(\|x - x^*\|) - y'(x - x^*)$$

has the same significant properties for our purposes as F_0 [attains a global minimum over C at x^*, and has $-\nabla F_1(x^*) = y$], and is in addition continuous.

We next define the function $\theta_2 : [0,\infty) \mapsto \Re$ by

$$\theta_2(r) = \begin{cases} \frac{1}{r}\int_r^{2r} \theta_1(s)ds & \text{if } r > 0, \\ 0 & \text{if } r = 0. \end{cases}$$

Again θ_2 is seen to be nondecreasing, and satisfies

$$0 \le \theta_1(r) \le \theta_2(r), \qquad \forall\, r \ge 0,$$

$\theta_2(0) = 0$, and $\nabla\theta_2(0) = 0$. Also, because θ_1 is continuous, θ_2 is smooth, and so is the function F_2 given by

$$F_2(x) = \theta_2(\|x - x^*\|) - y'(x - x^*).$$

The function F_2 fulfills all the requirements of the proposition, except that it may have global minima other than x^*. To ensure the uniqueness of x^* we modify F_2 as follows:

$$F(x) = F_2(x) + \|x - x^*\|^2.$$

The function F is smooth, attains a strict global minimum over C at x^*, and satisfies $-\nabla F(x^*) = y$.

4.25

We consider the problem

$$\begin{aligned} &\text{minimize} \quad \|x_1 - x_2\| + \|x_2 - x_3\| + \|x_3 - x_1\| \\ &\text{subject to} \quad x_i \in C_i, \quad i = 1, 2, 3, \end{aligned}$$

with the additional condition that x_1, x_2 and x_3 do not lie on the same line. Suppose that (x_1^*, x_2^*, x_3^*) defines an optimal triangle. Then, x_1^* solves the problem

$$\begin{aligned} &\text{minimize} \quad \|x_1 - x_2^*\| + \|x_2^* - x_3^*\| + \|x_3^* - x_1\| \\ &\text{subject to} \quad x_1 \in C_1, \end{aligned}$$

for which we have the following necessary optimality condition

$$d_1 = \frac{x_2^* - x_1^*}{\|x_2^* - x_1^*\|} + \frac{x_3^* - x_1^*}{\|x_3^* - x_1^*\|} \in T_{C_1}(x_1^*)^*.$$

The half-line $\{x \mid x = x_1^* + \alpha d_1, \ \alpha \geq 0\}$ is one of the bisectors of the optimal triangle. Similarly, there exist $d_2 \in T_{C_2}(x_2^*)^*$ and $d_3 \in T_{C_3}(x_3^*)^*$ which define the remaining bisectors of the optimal triangle. By elementary geometry, there exists a unique point z^* at which all three bisectors intersect (z^* is the center of the circle that is inscribed in the optimal triangle). From the necessary optimality conditions we have

$$z^* - x_i^* = \alpha_i d_i \in T_{C_i}(x_i^*)^*, \qquad i = 1, 2, 3.$$

4.26

Let us characterize the cone $T_X(x^*)^*$. Define

$$A(x^*) = \big\{ j \mid a_j' x^* = b_j \big\}.$$

Since X is convex, by Prop. 4.6.2, we have

$$T_X(x^*) = \mathrm{cl}\big(F_X(x^*)\big),$$

while from definition of X, we have

$$F_X(x^*) = \big\{ y \mid a_j' y \leq 0, \ \forall\, j \in A(x^*) \big\},$$

and since this set is closed, it follows that

$$T_X(x^*) = \big\{ y \mid a_j' y \leq 0, \ \forall\, j \in A(x^*) \big\}.$$

By taking polars in this relation and by using the Farkas' Lemma [Prop. 3.2.1(b)], we obtain

$$T_X(x^*)^* = \left\{ \sum_{j \in A(x^*)} \mu_j a_j \;\middle|\; \mu_j \geq 0,\ \forall\, j \in A(x^*) \right\},$$

and by letting $\mu_j = 0$ for all $j \notin A(x^*)$, we can write

$$T_X(x^*)^* = \left\{ \sum_{j=1}^{r} \mu_j a_j \;\middle|\; \mu_j \geq 0,\ \forall\, j,\quad \mu_j = 0,\ \forall\, j \notin A(x^*) \right\}. \tag{4.31}$$

By Prop. 4.7.2, the vector x^* minimizes f over X if and only if

$$0 \in \partial f(x^*) + T_X(x^*)^*.$$

In view of Eq. (4.31) and the definition of $A(x^*)$, it follows that x^* minimizes f over X if and only if there exist $\mu_1^*, \ldots, \mu_r^*$ such that

(i) $\mu_j^* \geq 0$ for all $j = 1, \ldots, r$, and $\mu_j^* = 0$ for all j such that $a_j' x^* < b_j$.

(ii) $0 \in \partial f(x^*) + \sum_{j=1}^{r} \mu_j^* a_j$.

4.27 (Quasiregularity)

(a) Let y be a nonzero tangent of X at x^*. Then there exists a sequence $\{\xi^k\}$ and a sequence $\{x^k\} \subset X$ such that $x^k \neq x^*$ for all k,

$$\xi^k \to 0, \qquad x^k \to x^*,$$

and

$$\frac{x^k - x^*}{\|x^k - x^*\|} = \frac{y}{\|y\|} + \xi^k. \tag{4.32}$$

By the mean value theorem, we have for all k

$$f(x^k) = f(x^*) + \nabla f(\tilde{x}^k)'(x^k - x^*),$$

where $\tilde{x}^k$ is a vector that lies on the line segment joining x^k and x^*. Using Eq. (4.32), the last relation can be written as

$$f(x^k) = f(x^*) + \frac{\|x^k - x^*\|}{\|y\|} \nabla f(\tilde{x}^k)' y^k, \tag{4.33}$$

where

$$y^k = y + \|y\| \xi^k.$$

If the tangent y satsifies $\nabla f(x^*)'y < 0$, then, since $\tilde{x}^k \to x^*$ and $y^k \to y$, we obtain for all sufficiently large k, $\nabla f(\tilde{x}^k)' y^k < 0$ and [from Eq. (4.33)] $f(x^k) < f(x^*)$. This contradicts the local optimality of x^*.

(b) Assume first that there are no equality constraints. Let $x \in X$ and let y be a nonzero tangent of X at x. Then there exists a sequence $\{\xi^k\}$ and a sequence $\{x^k\} \subset X$ such that $x^k \neq x$ for all k,

$$\xi^k \to 0, \qquad x^k \to x,$$

and

$$\frac{x^k - x}{\|x^k - x\|} = \frac{y}{\|y\|} + \xi^k.$$

By the mean value theorem, we have for all j and k

$$0 \geq g_j(x^k) = g_j(x) + \nabla g_j(\tilde{x}^k)'(x^k - x) = \nabla g_j(\tilde{x}^k)'(x^k - x),$$

where $\tilde{x}^k$ is a vector that lies on the line segment joining x^k and x. This relation can be written as

$$\frac{\|x^k - x\|}{\|y\|} \nabla g_j(\tilde{x}^k)' y^k \leq 0,$$

where $y^k = y + \xi^k \|y\|$, or equivalently

$$\nabla g_j(\tilde{x}^k)' y^k \leq 0, \qquad y^k = y + \xi^k \|y\|.$$

Taking the limit as $k \to \infty$, we obtain $\nabla g_j(x)'y \leq 0$ for all j, thus proving that $y \in V(x)$, and $T_X(x) \subset V(x)$. If there are some equality constraints $h_i(x) = 0$, they can be converted to the two inequality constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, and the result follows similarly.

(c) Assume first that there are no equality constraints. From part (a), we have $D(x^*) \cap V(x^*) = \emptyset$, which is equivalent to having $\nabla f(x^*)'y \geq 0$ for all y with $\nabla g_j(x^*)'y \leq 0$ for all $j \in A(x^*)$. By Farkas' Lemma, this is equivalent to the existence of Lagrange multipliers μ_j^* with the properties stated in the exercise. If there are some equality constraints $h_i(x) = 0$, they can be converted to the two inequality constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, and the result follows similarly.

Convex Analysis and Optimization

Chapter 5 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 5: SOLUTION MANUAL

5.1 (Second Order Sufficiency Conditions for Equality-Constrained Problems)

We first prove the following lemma.

Lemma 5.1: Let P and Q be two symmetric matrices. Assume that Q is positive semidefinite and P is positive definite on the nullspace of Q, that is, $x'Px > 0$ for all $x \neq 0$ with $x'Qx = 0$. Then there exists a scalar $\overline{c}$ such that

$$P + cQ : \text{positive definite}, \qquad \forall\ c > \overline{c}.$$

Proof: Assume the contrary. Then for every integer k, there exists a vector x^k with $\|x^k\| = 1$ such that

$$x^{k\prime} P x^k + k x^{k\prime} Q x^k \leq 0.$$

Since $\{x^k\}$ is bounded, there is a subsequence $\{x^k\}_{k\in K}$ converging to some $\overline{x}$, and since $\|x^k\| = 1$ for all k, we have $\|\overline{x}\| = 1$. Taking the limit superior in the above inequality, we obtain

$$\overline{x}' P \overline{x} + \limsup_{k\to\infty,\, k\in K} \left(k x^{k\prime} Q x^k\right) \leq 0. \tag{5.1}$$

Since, by the positive semidefiniteness of Q, $x^{k\prime} Q x^k \geq 0$, we see that $\{x^{k\prime} Q x^k\}_K$ must converge to zero, for otherwise the left-hand side of the above inequality would be ∞. Therefore, $\overline{x}' Q \overline{x} = 0$ and since P is positive definite, we obtain $\overline{x}' P \overline{x} > 0$. This contradicts Eq. (5.1). **Q.E.D.**

Let us introduce now the *augmented Lagrangian* function

$$L_c(x, \lambda) = f(x) + \lambda' h(x) + \frac{c}{2} \|h(x)\|^2,$$

where c is a scalar. This is the Lagrangian function for the problem

$$\begin{aligned} &\text{minimize} \ \ f(x) + \frac{c}{2}\|h(x)\|^2 \\ &\text{subject to} \ \ h(x) = 0, \end{aligned}$$

which has the same local minima as our original problem of minimizing $f(x)$ subject to $h(x) = 0$. The gradient and Hessian of L_c with respect to x are

$$\nabla_x L_c(x, \lambda) = \nabla f(x) + \nabla h(x)\big(\lambda + c h(x)\big),$$

$$\nabla^2_{xx} L_c(x,\lambda) = \nabla^2 f(x) + \sum_{i=1}^m \big(\lambda_i + c h_i(x)\big)\nabla^2 h_i(x) + c\nabla h(x)\nabla h(x)'.$$

In particular, if x^* and λ^* satisfy the given conditions, we have

$$\nabla_x L_c(x^*,\lambda^*) = \nabla f(x^*) + \nabla h(x^*)\big(\lambda^* + c h(x^*)\big) = \nabla_x L(x^*,\lambda^*) = 0, \qquad (5.2)$$

$$\begin{aligned}\nabla^2_{xx} L_c(x^*,\lambda^*) &= \nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + c\nabla h(x^*)\nabla h(x^*)' \\ &= \nabla^2_{xx} L(x^*,\lambda^*) + c\nabla h(x^*)\nabla h(x^*)'.\end{aligned}$$

By assumption, we have that $y'\nabla^2_{xx} L(x^*,\lambda^*)y > 0$ for all $y \neq 0$ such that $y'\nabla h(x^*)\nabla h(x^*)'y = 0$, so by applying Lemma 5.1 with $P = \nabla^2_{xx} L(x^*,\lambda^*)$ and $Q = \nabla h(x^*)\nabla h(x^*)'$, it follows that there exists a $\overline{c}$ such that

$$\nabla^2_{xx} L_c(x^*,\lambda^*) : \text{positive definite}, \qquad \forall\ c > \overline{c}. \qquad (5.3)$$

Using now the standard sufficient optimality condition for unconstrained optimization (see e.g., [Ber99a], Section 1.1), we conclude from Eqs. (5.2) and (5.3), that for $c > \overline{c}$, x^* is an unconstrained local minimum of $L_c(\cdot,\lambda^*)$. In particular, there exist $\gamma > 0$ and $\epsilon > 0$ such that

$$L_c(x,\lambda^*) \geq L_c(x^*,\lambda^*) + \frac{\gamma}{2}\|x - x^*\|^2, \qquad \forall\ x \text{ with } \|x - x^*\| < \epsilon.$$

Since for all x with $h(x) = 0$ we have $L_c(x,\lambda^*) = f(x)$, $\nabla_\lambda L(x^*,\lambda^*) = h(x^*) = 0$, it follows that

$$f(x) \geq f(x^*) + \frac{\gamma}{2}\|x - x^*\|^2, \qquad \forall\ x \text{ with } h(x) = 0, \text{ and } \|x - x^*\| < \epsilon.$$

Thus x^* is a strict local minimum of f over $h(x) = 0$.

5.2 (Second Order Sufficiency Conditions for Inequality-Constrained Problems)

We prove this result by using a transformation to an equality-constrained problem together with Exercise 5.1. Consider the equivalent equality-constrained problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && h_1(x) = 0, \ldots, h_m(x) = 0, \\ &&& g_1(x) + z_1^2 = 0, \ldots, g_r(x) + z_r^2 = 0, \end{aligned} \qquad (5.4),$$

which is an optimization problem in variables x and $z = (z_1,\ldots,z_r)$. Consider the vector (x^*,z^*), where $z^* = (z_1^*,\ldots,z_r^*)$,

$$z_j^* = \big(-g_j(x^*)\big)^{1/2}, \qquad j = 1,\ldots,r.$$

We will show that (x^*, z^*) and (λ^*, μ^*) satisfy the sufficiency conditions of Exercise 5.1, thus showing that (x^*, z^*) is a strict local minimum of problem (5.4), proving that x^* is a strict local minimum of the original inequality-constrained problem.

Let $\overline{L}(x, z, \lambda, \mu)$ be the Lagrangian function for this problem, i.e.,

$$\overline{L}(x, z, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j \big(g_j(x) + z_j^2\big).$$

We have

$$\begin{aligned}\nabla_{(x,z)}\overline{L}(x^*, z^*, \lambda^*, \mu^*)' &= \big[\nabla_x \overline{L}(x^*, z^*, \lambda^*, \mu^*)',\ \nabla_z \overline{L}(x^*, z^*, \lambda^*, \mu^*)'\big] \\ &= \big[\nabla_x L(x^*, \lambda^*, \mu^*)',\ 2\mu_1^* z_1^*, \ldots, 2\mu_r^* z_r^*\big] \\ &= [0,\ 0],\end{aligned}$$

where the last equality follows since, by assumption, we have $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$, and $\mu_j^* = 0$ for all $j \notin A(x^*)$, whereas $z_j^* = \big(-g_j(x^*)\big)^{1/2} = 0$ for all $j \in A(x^*)$. We also have

$$\begin{aligned}\nabla_{(\lambda,\mu)}\overline{L}(x^*, z^*, \lambda^*, \mu^*)' = \Big[&h_1(x^*), \ldots, h_m(x^*), \\ &\big(g_1(x^*) + (z_1^*)^2\big), \ldots, \big(g_r(x^*) + (z_r^*)^2\big)\Big] \\ = [0, 0]&.\end{aligned}$$

Hence the first order conditions of the sufficiency conditions for equality-constrained problems, given in Exercise 5.1, are satisfied.

We next show that for all $(y, w) \neq (0, 0)$ satisfying

$$\nabla h(x^*)' y = 0, \qquad \nabla g_j(x^*)' y + 2z_j^* w_j = 0, \qquad j = 1, \ldots, r, \tag{5.5}$$

we have

$$(\, y' \quad w' \,) \begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) & 0 \\ 0 & \begin{matrix} 2\mu_1^* & 0 & \ldots & 0 \\ 0 & 2\mu_2^* & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 2\mu_r^* \end{matrix} \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} > 0. \tag{5.6}$$

The left-hand side of the preceding expression can also be written as

$$y' \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y + 2 \sum_{j=1}^{r} \mu_j^* w_j^2. \tag{5.7}$$

Let $(y, w) \neq (0, 0)$ be a vector satisfying Eq. (5.5). We have that $z_j^* = 0$ for all $j \in A(x^*)$, so it follows from Eq. (5.5) that

$$\nabla h_i(x^*)' y = 0, \quad \forall\ i = 1, \ldots, m, \qquad \nabla g_j(x^*)' y = 0, \quad \forall\ j \in A(x^*).$$

Hence, if $y \neq 0$, it follows by assumption that

$$y' \nabla^2_{xx} L(x^*, \lambda^*, \mu^*) y > 0,$$

which implies, by Eq. (5.7) and the assumption $\mu_j^* \geq 0$ for all j, that (y, w) satisfies Eq. (5.6), proving our claim.

If $y = 0$, it follows that $w_k \neq 0$ for some $k = 1, \ldots, r$. In this case, by using Eq. (5.5), we have

$$2 z_j^* w_j = 0, \qquad j = 1, \ldots, r,$$

from which we obtain that z_k^* must be equal to 0, and hence $k \in A(x^*)$. By assumption, we have that

$$\mu_j^* > 0, \qquad \forall\, j \in A(x^*).$$

This implies that $\mu_k^* w_k^2 > 0$, and therefore

$$2 \sum_{j=1}^{r} \mu_j^* w_j^2 > 0,$$

showing that (y, w) satisfies Eq. (5.6), completing the proof.

5.3 (Sensitivity Under Second Order Conditions)

We first prove the result for the special case of equality-constrained problems.

Proposition 5.3: Let x^* and λ^* be a local minimum and Lagrange multiplier, respectively, satisfying the second order sufficiency conditions of Exercise 5.1, and assume that the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, are linearly independent. Consider the family of problems

$$\begin{aligned} &\text{minimize} \ \ f(x) \\ &\text{subject to} \ \ h(x) = u, \end{aligned} \tag{5.8}$$

parameterized by the vector $u \in \Re^m$. Then there exists an open sphere S centered at $u = 0$ such that for every $u \in S$, there is an $x(u) \in \Re^n$ and a $\lambda(u) \in \Re^m$, which are a local minimum-Lagrange multiplier pair of problem (5.8). Furthermore, $x(\cdot)$ and $\lambda(\cdot)$ are continuously differentiable functions within S and we have $x(0) = x^*$, $\lambda(0) = \lambda^*$. In addition, for all $u \in S$ we have

$$\nabla p(u) = -\lambda(u),$$

where $p(u)$ is the optimal cost parameterized by u, that is,

$$p(u) = f\big(x(u)\big).$$

Proof: Consider the system of equations

$$\nabla f(x) + \nabla h(x) \lambda = 0, \qquad h(x) = u. \tag{5.9}$$

For each fixed u, this system represents $n+m$ equations with $n+m$ unknowns – the vectors x and λ. For $u=0$ the system has the solution (x^*,λ^*). The corresponding $(n+m)\times(n+m)$ Jacobian matrix with respect to (x,λ) is given by

$$J = \begin{pmatrix} \nabla^2_{xx}L(x^*,\lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)' & 0 \end{pmatrix}.$$

Let us show that J is nonsingular. If it were not, some nonzero vector $(y',z')'$ would belong to the nullspace of J, that is,

$$\nabla^2_{xx}L(x^*,\lambda^*)y + \nabla h(x^*)z = 0, \tag{5.10}$$

$$\nabla h(x^*)'y = 0. \tag{5.11}$$

Premultiplying Eq. (5.10) by y' and using Eq. (5.11), we obtain

$$y'\nabla^2_{xx}L(x^*,\lambda^*)y = 0.$$

In view of Eq. (5.11), it follows that $y=0$, for otherwise our second order sufficiency assumption would be violated. Since $y=0$, Eq. (5.10) yields $\nabla h(x^*)z=0$, which in view of the linear independence of the columns $\nabla h_i(x^*)$, $i=1,\ldots,m$, of $\nabla h(x^*)$, yields $z=0$. Thus, we obtain $y=0$, $z=0$, which is a contradiction. Hence, J is nonsingular.

Returning now to the system (5.9), it follows from the nonsingularity of J and the Implicit Function Theorem that for all u in some open sphere S centered at $u=0$, there exist $x(u)$ and $\lambda(u)$ such that $x(0)=x^*$, $\lambda(0)=\lambda^*$, the functions $x(\cdot)$ and $\lambda(\cdot)$ are continuously differentiable, and

$$\nabla f\big(x(u)\big) + \nabla h\big(x(u)\big)\lambda(u) = 0, \tag{5.12}$$

$$h\big(x(u)\big) = u.$$

For u sufficiently close to 0, the vectors $x(u)$ and $\lambda(u)$ satisfy the second order sufficiency conditions for problem (5.8), since they satisfy them by assumption for $u=0$. This is straightforward to verify by using our continuity assumptions. [If it were not true, there would exist a sequence $\{u^k\}$ with $u^k\to 0$, and a sequence $\{y^k\}$ with $\|y^k\|=1$ and $\nabla h\big(x(u^k)\big)'y^k=0$ for all k, such that

$$y^{k\prime}\nabla^2_{xx}L\big(x(u^k),\lambda(u^k)\big)y^k \le 0, \qquad \forall\ k.$$

By taking the limit along a convergent subsequence of $\{y^k\}$, we would obtain a contradiction of the second order sufficiency condition at (x^*,λ^*).] Hence, $x(u)$ and $\lambda(u)$ are a local minimum-Lagrange multiplier pair for problem (5.8).

There remains to show that $\nabla p(u) = \nabla_u\left\{f\big(x(u)\big)\right\} = -\lambda(u)$. By multiplying Eq. (5.12) by $\nabla x(u)$, we obtain

$$\nabla x(u)\nabla f\big(x(u)\big) + \nabla x(u)\nabla h\big(x(u)\big)\lambda(u) = 0.$$

By differentiating the relation $h\big(x(u)\big) = u$, it follows that

$$I = \nabla_u \Big\{h\big(x(u)\big)\Big\} = \nabla x(u)\nabla h\big(x(u)\big), \tag{5.13}$$

where I is the $m \times m$ identity matrix. Finally, by using the chain rule, we have

$$\nabla p(u) = \nabla_u \Big\{f\big(x(u)\big)\Big\} = \nabla x(u)\nabla f\big(x(u)\big).$$

Combining the above three relations, we obtain

$$\nabla p(u) + \lambda(u) = 0, \tag{5.14}$$

and the proof is complete. **Q.E.D.**

We next use the preceding result to show the corresponding result for inequality-constrained problems. We assume that x^* and (λ^*, μ^*) are a local minimum and Lagrange multiplier, respectively, of the problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \ h_1(x) = 0, \ldots, h_m(x) = 0, \\ &\qquad\qquad\quad g_1(x) \le 0, \ldots, g_r(x) \le 0, \end{aligned} \tag{5.15}$$

and they satisfy the second order sufficiency conditions of Exercise 5.2. We also assume that the gradients $\nabla h_i(x^*)$, $i = 1, \ldots, m$, $\nabla g_j(x^*)$, $j \in A(x^*)$ are linearly independent, i.e., x^* is regular. We consider the equality-constrained problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_1(x) = 0, \ldots, h_m(x) = 0, \\ & g_1(x) + z_1^2 = 0, \ldots, g_r(x) + z_r^2 = 0, \end{aligned} \tag{5.16}$$

which is an optimization problem in variables x and $z = (z_1, \ldots, z_r)$. Let z^* be a vector with

$$z_j^* = \big(-g_j(x^*)\big)^{1/2}, \qquad j = 1, \ldots, r.$$

It can be seen that, since x^* and (λ^*, μ^*) satisfy the second order assumptions of Exercise 5.2, (x^*, z^*) and (λ^*, μ^*) satisfy the second order assumptions of Exercise 5.1, thus showing that (x^*, z^*) is a strict local minimum of problem (5.16)(cf. proof of Exercise 5.2). It is also straightforward to see that since x^* is regular for problem (5.15), (x^*, z^*) is regular for problem (5.16). We consider the family of problems

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = u_i, \qquad i = 1, \ldots, m, \\ & g_j(x) + z_j^2 = v_j, \qquad j = 1, \ldots, r, \end{aligned} \tag{5.17}$$

parametrized by u and v.

Using Prop. 5.3, given in the beginning of this exercise, we have that there exists an open sphere S centered at $(u,v) = (0,0)$ such that for every $(u,v) \in S$ there is an $x(u,v) \in \Re^n$, $z(u,v) \in \Re^r$ and $\lambda(u,v) \in \Re^m$, $\mu(u,v) \in \Re^r$, which are a local minimum and associated Lagrange multiplier vectors of problem (5.17).

We claim that the vectors $x(u,v)$ and $\lambda(u,v) \in \Re^m$, $\mu(u,v) \in \Re^r$ are a local minimum and Lagrange multiplier vector for the problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && h_i(x) = u_i, \qquad \forall\ i = 1,\ldots,m, \\ & && g_j(x) \le v_j, \qquad \forall\ j = 1,\ldots,r. \end{aligned} \tag{5.18}$$

It is straightforward to see that $x(u,v)$ is a local minimum of the preceding problem. To see that $\lambda(u,v)$ and $\mu(u,v)$ are the corresponding Lagrange multipliers, we use the first order necessary optimality conditions for problem (5.17) to write

$$\nabla f\big(x(u,v)\big) + \sum_{i=1}^m \lambda_i(u,v)\nabla h_i\big(x(u,v)\big) + \sum_{j=1}^r \mu_j(u,v)\nabla g_j\big(x(u,v)\big) = 0,$$

$$2\mu_j(u,v)z_j(u,v) = 0, \qquad j = 1,\ldots,r.$$

Since $z_j(u,v) = \Big(v_j - g_j\big(x(u,v)\big)\Big)^{1/2} > 0$ for $j \notin A\big(x(u,v)\big)$, where

$$A\big(x(u,v)\big) = \Big\{ j \mid g_j\big(x(u,v)\big) = v_j \Big\},$$

the last equation can also be written as

$$\mu_j(u,v) = 0, \qquad \forall\ j \notin A\big(x(u,v)\big). \tag{5.19}$$

Thus, to show $\lambda(u,v)$ and $\mu(u,v)$ are Lagrange multipliers for problem (5.18), there remains to show the nonnegativity of $\mu(u,v)$. For this purpose we use the second order necessary condition for the equivalent equality constrained problem (5.17). It yields

$$\begin{pmatrix} y' & w' \end{pmatrix} \begin{pmatrix} \nabla^2_{xx} L\big(x(u,v),\lambda(u,v),\mu(u,v)\big) & 0 \\ 0 & \begin{matrix} 2\mu_1(u,v) & 0 & \ldots & 0 \\ 0 & 2\mu_2(u,v) & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 2\mu_r(u,v) \end{matrix} \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} \ge 0, \tag{5.20}$$

for all $y \in \Re^n$ and $w \in \Re^r$ satisfying

$$\nabla h\big(x(u,v)\big)' y = 0, \qquad \nabla g_j\big(x(u,v)\big)' y + 2z_j(u,v)w_j = 0, \qquad j \in A\big(x(u,v)\big). \tag{5.21}$$

Next let us select, for every $j \in A\big(x(u,v)\big)$, a vector (y,w) with $y = 0$, $w_j \neq 0$, $w_k = 0$ for all $k \neq j$. Such a vector satisfies the condition of Eq. (5.21). By using such a vector in Eq. (5.20), we obtain $2\mu_j(u,v)w_j^2 \ge 0$, and

$$\mu_j(u,v) \ge 0, \qquad \forall\ j \in A\big(x(u,v)\big).$$

Furthermore, by Prop. 5.3 given in the beginning of this exercise, it follows that $x(\cdot,\cdot)$, $\lambda(\cdot,\cdot)$, and $\mu(\cdot,\cdot)$ are continuously differentiable in S and we have $x(0,0) = x^*$, $\lambda(0,0) = \lambda^*$, $\mu(0,0) = \mu^*$. In addition, for all $(u,v) \in S$, there holds

$$\nabla_u p(u,v) = -\lambda(u,v),$$

$$\nabla_v p(u,v) = -\mu(u,v),$$

where $p(u,v)$ is the optimal cost of problem (5.17), parameterized by (u,v), which is the same as the optimal cost of problem (5.18), completing our proof.

5.4 (General Sufficiency Condition)

We have

$$\begin{aligned} f(x^*) &= f(x^*) + \mu^{*\prime} g(x^*) \\ &= \min_{x\in X}\big\{f(x) + \mu^{*\prime} g(x)\big\} \\ &\le \min_{x\in X,\, g(x)\le 0}\big\{f(x) + \mu^{*\prime} g(x)\big\} \\ &\le \min_{x\in X,\, g(x)\le 0} f(x) \\ &\le f(x^*), \end{aligned}$$

where the first equality follows from the hypothesis, which implies that $\mu^{*\prime} g(x^*) = 0$, the next-to-last inequality follows from the nonnegativity of μ^*, and the last inequality follows from the feasibility of x^*. It folows that equality holds throughout, and x^* is a optimal solution.

5.5

(a) We note that

$$\inf_{d\in\Re^n} L_0(d,\mu) = \begin{cases} -\frac{1}{2}\|\mu\|^2 & \text{if } \mu \in M, \\ -\infty & \text{otherwise,} \end{cases}$$

so since μ^* is the vector of minimum norm in M, we obtain for all $\gamma > 0$,

$$\begin{aligned} -\frac{1}{2}\|\mu^*\|^2 &= \sup_{\mu\ge 0}\inf_{d\in\Re^n} L_0(d,\mu) \\ &\le \inf_{d\in\Re^n}\sup_{\mu\ge 0} L_0(d,\mu), \end{aligned}$$

where the inequality follows from the minimax inequality (cf. Chapter 2). For any $d \in \Re^n$, the supremum of $L_0(d,\mu)$ over $\mu \ge 0$ is attained at

$$\mu_j = (a_j'd)^+, \qquad j = 1,\ldots,r.$$

[to maximize $\mu_j a_j'd - (1/2)\mu_j^2$ subject to the constraint $\mu_j \ge 0$, we calculate the unconstrained maximum, which is $a_j'd$, and if it is negative we set it to 0, so that the maximum subject to $\mu_j \geq 0$ is attained for $\mu_j = (a_j'd)^+]$. Hence, it follows that, for any $d \in \Re^n$,

$$\sup_{\mu \geq 0} L_0(d, \mu) = a_0'd + \frac{1}{2} \sum_{j=1}^{r} \big((a_j'd)^+ \big)^2,$$

which yields the desired relations.

(b) Since the infimum of the quadratic cost function $a_0'd + \frac{1}{2} \sum_{j=1}^{r} \big((a_j'd)^+ \big)^2$ is bounded below, as given in part (a), it follows from the results of Section 2.3 that the infimum of this function is attained at some $d^* \in \Re^n$.

(c) From the Saddle Point Theorem, for all $\gamma > 0$, the coercive convex/concave quadratic function L_γ has a saddle point, denoted (d^γ, μ^γ), over $d \in \Re^n$ and $\mu \geq 0$. This saddle point is unique and can be easily characterized, taking advantage of the quadratic nature of L_γ. In particular, similar to part (a), the maximization over $\mu \geq 0$ when $d = d^\gamma$ yields

$$\mu_j^\gamma = (a_j'd^\gamma)^+, \qquad j = 1, \ldots, r. \tag{5.22}$$

Moreover, we can find $L_\gamma(d^\gamma, \mu^\gamma)$ by minimizing $L_\gamma(d, \mu^\gamma)$ over $d \in \Re^n$. To find the unconstrained minimum d^γ, we take the gradient of $L_\gamma(d, \mu^\gamma)$ and set it equal to 0. This yields

$$d^\gamma = -\frac{a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j}{\gamma}.$$

Hence,

$$L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\left\| a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j \right\|^2}{2\gamma} - \frac{1}{2} \|\mu^\gamma\|^2.$$

We also have

$$\begin{aligned} -\frac{1}{2} \|\mu^*\|^2 &= \sup_{\mu \geq 0} \inf_{d \in \Re^n} L_0(d, \mu) \\ &\leq \inf_{d \in \Re^n} \sup_{\mu \geq 0} L_0(d, \mu) \\ &\leq \inf_{d \in \Re^n} \sup_{\mu \geq 0} L_\gamma(d, \mu) \\ &= L_\gamma(d^\gamma, \mu^\gamma), \end{aligned} \tag{5.23}$$

where the first two relations follow from part (a), thus yielding the desired relation.

(d) From part (c), we have

$$-\frac{1}{2} \|\mu^\gamma\|^2 \geq L_\gamma(d^\gamma, \mu^\gamma) = -\frac{\left\| a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j \right\|^2}{2\gamma} - \frac{1}{2} \|\mu^\gamma\|^2 \geq -\frac{1}{2} \|\mu^*\|^2. \tag{5.24}$$

From this, we see that $\|\mu^\gamma\| \leq \|\mu^*\|$, so that μ^γ remains bounded as $\gamma \to 0$. By taking the limit above as $\gamma \to 0$, we see that

$$\lim_{\gamma \to 0} \left(a_0 + \sum_{j=1}^{r} \mu_j^\gamma a_j \right) = 0,$$

so any limit point of μ^γ, call it $\overline{\mu}$, satisfies $-\left(a_0 + \sum_{j=1}^r \overline{\mu}_j a_j\right) = 0$. Since $\mu^\gamma \geq 0$, it follows that $\overline{\mu} \geq 0$, so $\overline{\mu} \in M$. We also have $\|\overline{\mu}\| \leq \|\mu^*\|$ (since $\|\mu^\gamma\| \leq \|\mu^*\|$), so by using the minimum norm property of μ^*, we conclude that any limit point $\overline{\mu}$ of μ^γ must be equal to μ^*. Thus, $\mu^\gamma \to \mu^*$. From Eq. (5.24), we then obtain

$$L_\gamma(d^\gamma, \mu^\gamma) \to -\frac{1}{2}\|\mu^*\|^2. \tag{5.25}$$

(e) Equations (5.23) and (5.25), together with part (b), show that

$$L_0(d^*, \mu^*) = \inf_{d \in \Re^n} \sup_{\mu \geq 0} L_0(d, \mu) = \sup_{\mu \geq 0} \inf_{d \in \Re^n} L_0(d, \mu),$$

[thus proving that (d^*, μ^*) is a saddle point of $L_0(d, \mu)$], and that

$$a_0' d^* = -\|\mu^*\|^2, \qquad (a_j' d^*)^+ = \mu_j^*, \quad j = 1, \ldots, r.$$

5.6 (Strict Complementarity)

Consider the following example

$$\begin{aligned} &\text{minimize } \ x_1 + x_2 \\ &\text{subject to } x_1 \leq 0,\ x_2 \leq 0,\ -x_1 - x_2 \leq 0. \end{aligned}$$

The only feasible vector is $x^* = (0, 0)$, which is therefore also the optimal solution of this problem. The vector $(1, 1, 2)'$ is a Lagrange multiplier vector which satisfies strict complementarity. However, it is not possible to find a vector that violates simultaneously all the constraints, showing that this Lagrange multiplier vector is not informative.

For the converse statement, consider the example of Fig. 5.1.3. The Lagrange multiplier vectors, that involve three nonzero components out of four, are informative, but they do not satisfy strict complementarity.

5.7

Let x^* be a local minimum of the problem

$$\begin{aligned} &\text{minimize } \ \sum_{i=1}^n f_i(x_i) \\ &\text{subject to } \ x \in S, \qquad x_i \in X_i, \quad i = 1, \ldots, n, \end{aligned}$$

where $f_i : \Re \mapsto \Re$ are smooth functions, X_i are closed intervals of real numbers of $\Re^n$, and S is a subspace of $\Re^n$. We introduce artificial optimization variables

$z_1, \ldots, z_n$ and the linear constraints $x_i = z_i$, $i = 1, \ldots, n$, while replacing the constraint $x \in S$ with $z \in S$, so that the problem becomes

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{n} f_i(x_i) \\ &\text{subject to } \ z \in S, \qquad x_i \in X_i, \qquad x_i = z_i, \quad i = 1, \ldots, n. \end{aligned} \tag{5.26}$$

Let $a_1, \ldots, a_m$ be a basis for $S^\perp$, the orthogonal complement of S. Then, we can represent S as

$$S = \{y \mid a_j' y = 0, \ \forall\, j = 1, \ldots, m\}.$$

We also represent the closed intervals X_i as

$$X_i = \{y \mid c_i \le y \le d_i\}.$$

With the previous identifications, the constraint set of problem (5.26) can be described alternatively as

$$\begin{aligned} \text{minimize } & \sum_{i=1}^{n} f_i(x_i) \\ \text{subject to } & a_j' z = 0, \quad j = 1, \ldots, m, \\ & c_i \le x_i \le d_i, \quad i = 1, \ldots, n, \\ & x_i = z_i, \quad i = 1, \ldots, n \end{aligned}$$

[cf. extended representation of the constraint set of problem (5.26)]. This is a problem with linear constraints, so by Prop. 5.4.1, it admits Lagrange multipliers. But, by Prop. 5.6.1, this implies that the problem admits Lagrange multipliers in the original representation as well. We associate a Lagrange multiplier λ_i^* with each equality constraint $x_i = z_i$ in problem (5.26). By taking the gradient with respect to the variable x, and using the definition of Lagrange multipliers, we get

$$\big(\nabla f_i(x_i^*) + \lambda_i^*\big)(x_i - x_i^*) \ge 0, \qquad \forall\, x_i \in X_i, \ i = 1, \ldots, n,$$

whereas, by taking the gradient with respect to the variable z, we obtain $\lambda^* \in S^\perp$, thus completing the proof.

5.8

We first show that CQ5a implies CQ6. Assume CQ5a holds:

(a) There does not exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

$$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*).$$

(b) There exists a $d \in N_X(x^*)^* = T_X(x^*)$ (since X is regular at x^*) such that

$$\nabla h_i(x^*)'d = 0, \quad i = 1, \ldots, m, \qquad \nabla g_j(x^*)'d < 0, \quad \forall\, j \in A(x^*).$$

To arrive at a contradiction, assume that CQ6 does not hold, i.e., there are scalars $\lambda_1, \ldots, \lambda_m$, $\mu_1, \ldots, \mu_r$, not all of them equal to zero, such that

(i)
$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*).$$

(ii) $\mu_j \geq 0$ for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

In view of our assumption that X is regular at x^*, condition (i) can be written as

$$-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in T_X(x^*)^*,$$

or equivalently,

$$\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*). \tag{5.27}$$

Since not all the λ_i and μ_j are equal to 0, we conclude that $\mu_j > 0$ for at least one $j \in A(x^*)$; otherwise condition (a) of CQ5a would be violated. Since $\mu_j^* \geq 0$ for all j, with $\mu_j^* = 0$ for $j \notin A(x^*)$ and $\mu_j^* > 0$ for at least one j, we obtain

$$\sum_{i=1}^m \lambda_i \nabla h_i(x^*)'d + \sum_{j=1}^r \mu_j \nabla g_j(x^*)'d < 0,$$

where $d \in T_X(x^*)$ is the vector in condition (b) of CQ5a. But this contradicts Eq. (5.27), showing that CQ6 holds.

Conversely, assume that CQ6 holds. It can be seen that this implies condition (a) of CQ5a. Let H denote the subspace spanned by the vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$, and let G denote the cone generated by the vectors $\nabla g_j(x^*)$, $j \in A(x^*)$. Then, the orthogonal complement of H is given by

$$H^\perp = \big\{y \mid \nabla h_i(x^*)'y = 0,\ \forall\, i = 1, \ldots, m\big\},$$

whereas the polar of G is given by

$$G^* = \big\{y \mid \nabla g_j(x^*)'y \leq 0,\ \forall\, j \in A(x^*)\big\},$$

(cf. the results of Section 3.1). The interior of G^* is the set

$$\mathrm{int}(G^*) = \big\{y \mid \nabla g_j(x^*)'y < 0,\ \forall\, j \in A(x^*)\big\}.$$

Under CQ6, we have $\text{int}(G^*) \neq \emptyset$, since otherwise the vectors $\nabla g_j(x^*)$, $j \in A(x^*)$ would be linearly dependent, contradicting CQ6. Similarly, under CQ6, we have

$$H^\perp \cap \text{int}(G^*) \neq \emptyset. \tag{5.28}$$

To see this, assume the contrary, i.e., $H^\perp$ and $\text{int}(G^*)$ are disjoint. The sets $H^\perp$ and $\text{int}(G^*)$ are convex, therefore by the Separating Hyperplane Theorem, there exists some nonzero vector ν such that

$$\nu' x \leq \nu' y, \qquad \forall\ x \in H^\perp,\ \forall\ y \in \text{int}(G^*),$$

or equivalently,

$$\nu'(x - y) \leq 0, \qquad \forall\ x \in H^\perp,\ \forall\ y \in G^*,$$

which implies, using also Exercise 3.4., that

$$\nu \in (H^\perp - G^*)^* = H \cap (-G).$$

But this contradicts CQ6, and proves Eq. (5.28).

Finally, we show that CQ6 implies condition (b) of CQ5a. Assume, to arrive at a contradiction, that condition (b) of CQ5a does not hold. This implies that

$$N_X(x^*)^* \cap H^\perp \cap \text{int}\big(G^*\big) = \emptyset.$$

Since X is regular at x^*, the preceding is equivalent to

$$T_X(x^*) \cap H^\perp \cap \text{int}\big(G^*\big) = \emptyset.$$

The regularity of X at x^* implies that $T_X(x^*)$ is convex. Similarly, since the interior of a convex set is convex and the intersection of two convex sets is convex, it follows that the set $H^\perp \cap \text{int}\big(G^*\big)$ is convex. It is also nonempty by Eq. (5.28). Thus, by the Separating Hyperplane Theorem, there exists some vector $a \neq 0$ such that

$$a'x \leq a'y, \qquad \forall\ x \in T_X(x^*),\ \forall\ y \in H^\perp \cap \text{int}\big(G^*\big),$$

or equivalently,

$$a'(x - y) \leq 0, \qquad \forall\ x \in T_X(x^*),\ \forall\ y \in H^\perp \cap G^*,$$

which implies that

$$a \in \big(T_X(x^*) - (H^\perp \cap G^*)\big)^*.$$

We have

$$\begin{aligned}
\big(T_X(x^*) - (H^\perp \cap G^*)\big)^* &= T_X(x^*)^* \cap \big(-(H^\perp \cap G^*)^*\big) \\
&= T_X(x^*)^* \cap \Big(-\big(\text{cl}(H + G)\big)\Big) \\
&= T_X(x^*)^* \cap \big(-(H + G)\big) \\
&= N_X(x^*) \cap \big(-(H + G)\big),
\end{aligned}$$

where the second equality follows since $H^\perp$ and G^* are closed and convex, and the third equality follows since H and G are both polyhedral cones (cf. Chapter 3). Combining the preceding relations, it follows that there exists a nonzero vector a that belongs to the set

$$N_X(x^*) \cap \big(-(H + G)\big).$$

But this contradicts CQ6, thus completing our proof.

5.9 (Minimax Problems)

Let x^* be a local minimum of the minimax problem,

$$\begin{aligned} &\text{minimize} \quad \max\big\{f_1(x),\ldots,f_p(x)\big\} \\ &\text{subject to} \quad x \in X. \end{aligned}$$

We introduce an additional scalar variable z and convert the preceding problem to the smooth problem

$$\begin{aligned} &\text{minimize} \quad z \\ &\text{subject to} \quad x \in X, \quad f_i(x) \le z, \quad i = 1,\ldots,p, \end{aligned}$$

which is an optimization problem in the variables x and z and with an abstract set constraint $(x,z) \in X \times \Re$. Let

$$z^* = \max\big\{f_1(x^*),\ldots,f_p(x^*)\big\}.$$

It can be seen that (x^*, z^*) is a local minimum of the above problem.

It is straightforward to show that

$$N_{X\times\Re}(x^*, z^*) = N_X(x^*) \times \{0\}, \tag{5.29}$$

and

$$N_{X\times\Re}(x^*, z^*)^* = N_X(x^*)^* \times \Re. \tag{5.30}$$

Let $d = (0,1)$. By Eq. (5.30), this vector belongs to the set $N_{X\times\Re}(x^*, z^*)^*$, and also

$$[\nabla f_i(x^*)',\ -1] \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1 < 0, \qquad \forall\ i = 1,\ldots,p.$$

Hence, CQ5a is satisfied, which together with Eq. (5.29) implies that there exists a nonnegative vector $\mu^* = (\mu_1^*,\ldots,\mu_p^*)$ such that

(i) $-\left(\sum_{j=1}^p \mu_j^* \nabla f_i(x^*)\right) \in N_X(x^*)$.

(ii) $\sum_{j=1}^p \mu_j^* = 1$.

(iii) For all $j = 1,\ldots,p$, if $\mu_j^* > 0$, then

$$f_j(x^*) = \max\big\{f_1(x^*),\ldots,f_p(x^*)\big\}.$$

5.10 (Exact Penalty Functions)

We consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{5.31}$$

where

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \le 0, \ldots, g_r(x) \le 0\big\},$$

and the exact penalty function

$$F_c(x) = f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right),$$

where c is a positive scalar.

(a) In view of our assumption that, for some given $c > 0$, x^* is also a local minimum of F_c over X, we have, by Prop. 5.5.1, that there exist $\lambda_1, \ldots, \lambda_m$ and $\mu_1, \ldots, \mu_r$ such that

$$-\left(\nabla f(x^*) + c\left(\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right)\right) \in N_X(x^*),$$

$$\lambda_i = 1 \quad \text{if } h_i(x^*) > 0, \qquad \lambda_i = -1 \quad \text{if } h_i(x^*) < 0,$$

$$\lambda_i \in [-1, 1] \quad \text{if } h_i(x^*) = 0,$$

$$\mu_j = 1 \quad \text{if } g_j(x^*) > 0, \qquad \mu_j = 0 \quad \text{if } g_j(x^*) < 0,$$

$$\mu_j \in [0, 1] \quad \text{if } g_j(x^*) = 0.$$

By the definition of R-multipliers, the preceding relations imply that the vector $(\lambda^*, \mu^*) = c(\lambda, \mu)$ is an R-multiplier for problem (5.31) such that

$$|\lambda_i^*| \le c, \quad i = 1, \ldots, m, \qquad \mu_j^* \in [0, c], \quad j = 1, \ldots, r. \tag{5.32}$$

(b) Assume that the functions f and the g_j are convex, the functions h_i are linear, and the set X is convex. Since x^* is a local minimum of problem (5.31), and (λ^*, μ^*) is a corresponding Lagrange multiplier vector, we have by definition that

$$\left(\nabla f(x^*) + \left(\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)\right)'(x - x^*) \ge 0, \qquad \forall\, x \in X.$$

In view of the convexity assumptions, this is a sufficient condition for x^* to be a local minimum of the function $f(x) + \sum_{i=1}^{m} \lambda_i^* h_i(x) + \sum_{j=1}^{r} \mu_j^* g_j(x)$ over $x \in X$.

Since x^* is feasible for the original problem, and (λ^*, μ^*) satisfy Eq. (5.32), we have for all $x \in X$,

$$\begin{aligned} F_C(x^*) &= f(x^*) \\ &\le f(x) + \sum_{i=1}^{m} \lambda_i^* h_i(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \\ &\le f(x) + c\left(\sum_{i=1}^{m} h_i(x) + \sum_{j=1}^{r} g_j(x)\right) \\ &\le f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right) \\ &= F_c(x), \end{aligned}$$

implying that x^* is a local minimum of F_c over X.

5.11 (Extended Representations)

(a) The hypothesis implies that for every smooth cost function f for which x^* is a local minimum there exist scalars $\lambda_1^*, \ldots, \lambda_{\overline{m}}^*$ and $\mu_1^*, \ldots, \mu_{\overline{r}}^*$ satisfying

$$\left(\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*)\right)' y \ge 0, \qquad \forall\ y \in T_{\overline{X}}(x^*), \quad (5.33)$$

$$\mu_j^* \ge 0, \qquad \forall\ j = 1, \ldots, \overline{r},$$
$$\mu_j^* = 0, \qquad \forall\ j \notin \overline{A}(x^*),$$

where

$$\overline{A}(x^*) = \big\{ j \mid g_j(x^*) = 0,\ j = 1, \ldots, \overline{r} \big\}.$$

Since $X \subset \overline{X}$, we have $T_X(x^*) \subset T_{\overline{X}}(x^*)$, so Eq. (5.33) implies that

$$\left(\nabla f(x^*) + \sum_{i=1}^{\overline{m}} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*)\right)' y \ge 0, \qquad \forall\ y \in T_X(x^*). \quad (5.34)$$

Let $V(x^*)$ denote the set

$$\begin{aligned} V(x^*) = \big\{ y \mid{}& \nabla h_i(x^*)' y = 0,\ i = m+1, \ldots, \overline{m}, \\ & \nabla g_j(x^*)' y \le 0,\ j = r+1, \ldots, \overline{r} \text{ with } j \in \overline{A}(x^*) \big\}. \end{aligned}$$

We claim that $T_X(x^*) \subset V(x^*)$. To see this, let y be a nonzero vector that belongs to $T_X(x^*)$. Then, there exists a sequence $\{x_k\} \subset X$ such that $x_k \ne x^*$ for all k and

$$\frac{x_k - x^*}{\|x_k - x^*\|} \to \frac{y}{\|y\|}.$$

Since $x_k \in X$, for all $i = m+1, \ldots, \overline{m}$ and k, we have

$$0 = h_i(x_k) = h_i(x^*) + \nabla h_i(x^*)'(x_k - x^*) + o(\|x_k - x^*\|),$$

which can be written as

$$\nabla h_i(x^*)' \frac{(x_k - x^*)}{\|x_k - x^*\|} + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|} = 0.$$

Taking the limit as $k \to \infty$, we obtain

$$\nabla h_i(x^*)'y = 0, \qquad \forall\ i = m+1, \ldots, \overline{m}. \tag{5.35}$$

Similarly, we have for all $j = r+1, \ldots, \overline{r}$ with $j \in \overline{A}(x^*)$ and for all k

$$0 \geq g_j(x_k) = g_j(x^*) + \nabla g_j(x^*)'(x_k - x^*) + o(\|x_k - x^*\|),$$

which can be written as

$$\nabla g_j(x^*)' \frac{(x_k - x^*)}{\|x_k - x^*\|} + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|} \leq 0.$$

By taking the limit as $k \to \infty$, we obtain

$$\nabla g_j(x^*)'y \leq 0, \qquad \forall\ j = r+1, \ldots, \overline{r} \text{ with } j \in \overline{A}(x^*).$$

Equation (5.35) and the preceding relation imply that $y \in V(x^*)$, showing that $T_X(x^*) \subset V(x^*)$.

Hence Eq. (5.34) implies that

$$\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\ y \in T_X(x^*),$$

and it follows that λ_i^*, $i = 1, \ldots, m$, and μ_j^*, $j = 1, \ldots, r$, are Lagrange multipliers for the original representation.

(b) Consider the exact penalty function for the extended representation:

$$\overline{F}_c(x) = f(x) + c\left(\sum_{i=1}^{\overline{m}} |h_i(x)| + \sum_{j=1}^{\overline{r}} g_j^+(x)\right).$$

We have $F_c(x) = \overline{F}_c(x)$ for all $x \in X$. Hence if $x^* \in C$ is a local minimum of $\overline{F}_c(x)$ over $x \in \overline{X}$, it is also a local minimum of $F_c(x)$ over $x \in X$. Thus, for a given $c > 0$, if x^* is both a strict local minimum of f over C and a local minimum of $\overline{F}_c(x)$ over $x \in \overline{X}$, it is also a local minimum of $F_c(x)$ over $x \in X$.

Convex Analysis and Optimization

Chapter 6 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 6: SOLUTION MANUAL

6.1

We consider the dual function

$$\begin{aligned} q(\mu_1,\mu_2) &= \inf_{x_1\ge 0,\ x_2\ge 0} \big\{x_1 - x_2 + \mu_1(x_1+1) + \mu_2(1-x_1-x_2)\big\} \\ &= \inf_{x_1\ge 0,\ x_2\ge 0} \big\{x_1(1+\mu_1-\mu_2) + x_2(-1-\mu_2) + \mu_1 + \mu_2\big\}. \end{aligned}$$

It can be seen that if $-\mu_1+\mu_2-1\le 0$ and $\mu_2+1\le 0$, then the infimum above is attained at $x_1=0$ and $x_2=0$. In this case, the dual function is given by $q(\mu_1,\mu_2)=\mu_1+\mu_2$. On the other hand, if $1+\mu_1-\mu_2<0$ or $-1-\mu_2<0$, then we have $q(\mu_1,\mu_2)=-\infty$. Thus, the dual problem is

$$\begin{aligned} &\text{maximize} \quad \mu_1+\mu_2 \\ &\text{subject to} \quad \mu_1\ge 0,\ \mu_2\ge 0,\ -\mu_1+\mu_2-1\le 0,\ \mu_2+1\le 0. \end{aligned}$$

6.2 (Extended Representation)

Assume that there exists a geometric multiplier in the extended representation. This implies that there exist nonnegative scalars $\lambda_1^*,\ldots,\lambda_m^*,\lambda_{m+1}^*,\ldots,\lambda_{\overline{m}}^*$ and $\mu_1^*,\ldots,\mu_r^*,\mu_{r+1}^*,\ldots,\mu_{\overline{r}}^*$ such that

$$f^* = \inf_{x\in\Re^n}\left\{f(x)+\sum_{i=1}^{\overline{m}}\lambda_i^* h_i(x) + \sum_{j=1}^{\overline{r}}\mu_j^* g_j(x)\right\},$$

implying that

$$f^* \le f(x)+\sum_{i=1}^{\overline{m}}\lambda_i^* h_i(x) + \sum_{j=1}^{\overline{r}}\mu_j^* g_j(x), \qquad \forall\ x\in\Re^n.$$

For any $x\in X$, we have $h_i(x)=0$ for all $i=m+1,\ldots,\overline{m}$, and $g_j(x)\le 0$ for all $j=r+1,\ldots,\overline{r}$, so that $\mu_j^* g_j(x)\le 0$ for all $j=r+1,\ldots,\overline{r}$. Therefore, it follows from the preceding relation that

$$f^* \le f(x)+\sum_{i=1}^{m}\lambda_i^* h_i(x) + \sum_{j=1}^{r}\mu_j^* g_j(x), \qquad \forall\ x\in X.$$

Taking the infimum over all $x \in X$, it follows that

$$\begin{aligned} f^* &\leq \inf_{x\in X} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* h_i(x) + \sum_{j=1}^r \mu_j^* g_j(x) \right\} \\ &\leq \inf_{\substack{x\in X,\ h_i(x)=0,\ i=1,\ldots,m \\ g_j(x)\leq 0,\ j=1,\ldots,r}} \left\{ f(x) + \sum_{i=1}^m \lambda_i^* h_i(x) + \sum_{j=1}^r \mu_j^* g_j(x) \right\} \\ &\leq \inf_{\substack{x\in X,\ h_i(x)=0,\ i=1,\ldots,m \\ g_j(x)\leq 0,\ j=1,\ldots,r}} f(x) \\ &= f^*. \end{aligned}$$

Hence, equality holds throughout above, showing that the scalars $\lambda_1^*, \ldots, \lambda_m^*$, $\mu_1^*, \ldots, \mu_r^*$ constitute a geometric multiplier for the original representation.

6.3 (Quadratic Programming Duality)

Consider the extended representation of the problem in which the linear inequalities that represent the polyhedral part are lumped with the remaining linear inequality constraints. From Prop. 6.3.1, finiteness of the optimal value implies that there exists an optimal solution and a geometric multiplier. From Exercise 6.2, it follows that there exists a geometric multiplier for the original representation of the problem.

6.4 (Sensitivity)

We have

$$\overline{f} = \inf_{x\in X} \big\{ f(x) + \overline{\mu}'\big(g(x) - \overline{u}\big)\big\},$$

$$\tilde{f} = \inf_{x\in X} \big\{ f(x) + \tilde{\mu}'\big(g(x) - \tilde{u}\big)\big\}.$$

Let $\overline{q}(\mu)$ denote the dual function of the problem corresponding to $\overline{u}$:

$$\overline{q}(\mu) = \inf_{x\in X} \big\{ f(x) + \mu'\big(g(x) - \overline{u}\big)\big\}.$$

We have

$$\begin{aligned} \overline{f} - \tilde{f} &= \inf_{x\in X} \big\{ f(x) + \overline{\mu}'\big(g(x) - \overline{u}\big)\big\} - \inf_{x\in X} \big\{ f(x) + \tilde{\mu}'\big(g(x) - \tilde{u}\big)\big\} \\ &= \inf_{x\in X} \big\{ f(x) + \overline{\mu}'\big(g(x) - \overline{u}\big)\big\} - \inf_{x\in X} \big\{ f(x) + \tilde{\mu}'\big(g(x) - \overline{u}\big)\big\} + \tilde{\mu}'(\tilde{u} - \overline{u}) \\ &= \overline{q}(\overline{\mu}) - \overline{q}(\tilde{\mu}) + \tilde{\mu}'(\tilde{u} - \overline{u}) \\ &\geq \tilde{\mu}'(\tilde{u} - \overline{u}), \end{aligned}$$

where the last inequality holds because $\overline{\mu}$ maximizes $\overline{q}$.

This proves the left-hand side of the desired inequality. Interchanging the roles of $\overline{f}$, $\overline{u}$, $\overline{\mu}$, and $\tilde{f}$, $\tilde{u}$, $\tilde{\mu}$, shows the desired right-hand side.

6.5

We first consider the relation

$$(P) \qquad \min_{A'x \geq b} c'x \Longleftrightarrow \max_{A\mu = c, \mu \geq 0} b'\mu. \quad (D)$$

The dual problem to (P) is

$$\max_{\mu \geq 0} q(\mu) = \max_{\mu \geq 0} \inf_{x \in \Re^n} \left\{ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \mu_i a_{ij} \right) x_j + \sum_{i=1}^{m} \mu_i b_i \right\}.$$

If $c_j - \sum_{i=1}^{m} \mu_i a_{ij} \neq 0$, then $q(\mu) = -\infty$. Thus the dual problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} \mu_i b_i \\ &\text{subject to} \quad \sum_{i=1}^{m} \mu_i a_{ij} = c_j, \quad j = 1, \ldots, n, \quad \mu \geq 0. \end{aligned}$$

To determine the dual of (D), note that (D) is equivalent to

$$\min_{A\mu = c, \mu \geq 0} -b'\mu,$$

and so its dual problem is

$$\max_{x \in \Re^n} p(x) = \max_x \inf_{\mu \geq 0} \big\{ (Ax - b)'\mu - c'x \big\}.$$

If $a_i'x - b_i < 0$ for any i, then $p(x) = -\infty$. Thus the dual of (D) is

$$\begin{aligned} &\text{maximize} \quad -c'x \\ &\text{subject to} \quad A'x \geq b, \end{aligned}$$

or

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad A'x \geq b. \end{aligned}$$

The Lagrangian optimality condition for (P) is

$$x^* = \arg\min_x \left\{ \left(c - \sum_{i=1}^{m} \mu_i^* a_i \right)' x + \sum_{i=1}^{m} \mu_i^* b_i \right\},$$

from which we obtain the complementary slackness conditions for (P):

$$A\mu = c.$$

The Lagrangian optimality condition for (D) is

$$\mu^* = \arg\min_{\mu\geq 0}\{(Ax^* - b)'\mu - c'x^*\},$$

from which we obtain the complementary slackness conditions for (D):

$$Ax^* - b \geq 0, \qquad (Ax^* - b)_i\mu_i^* = 0, \quad \forall\ i.$$

Next, consider

$$(P) \qquad \min_{A'x\geq b, x\geq 0} c'x \Longleftrightarrow \max_{A\mu\leq c, \mu\geq 0} b'\mu. \quad (D)$$

The dual problem to (P) is

$$\max_{\mu\geq 0} q(\mu) = \max_{\mu\geq 0}\inf_{x\geq 0}\left\{\sum_{j=1}^{n}\left(c_j - \sum_{i=1}^{m}\mu_i a_{ij}\right)x_j + \sum_{i=1}^{m}\mu_i b_i\right\}.$$

If $c_j - \sum_{i=1}^m \mu_i a_{ij} < 0$, then $q(\mu) = -\infty$. Thus the dual problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m}\mu_i b_i \\ &\text{subject to} \quad \sum_{i=1}^{m}\mu_i a_{ij} \leq c_j, \quad j = 1, \ldots, n, \quad \mu \geq 0. \end{aligned}$$

To determine the dual of (D), note that (D) is equivalent to

$$\min_{A\mu\leq c, \mu\geq 0} -b'\mu,$$

and so its dual problem is

$$\max_{x\geq 0} p(x) = \max_{x\geq 0}\inf_{\mu\geq 0}\big\{(Ax - b)'\mu - c'x\big\}.$$

If $a_i'x - b_i < 0$ for any i, then $p(x) = -\infty$. Thus the dual of (D) is

$$\begin{aligned} &\text{maximize} \quad -c'x \\ &\text{subject to} \quad A'x \geq b, \quad x \geq 0 \end{aligned}$$

or

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad A'x \geq b, \quad x \geq 0. \end{aligned}$$

The Lagrangian optimality condition for (P) is

$$x^* = \arg\min_{x\geq 0}\left\{\left(c - \sum_{i=1}^{m}\mu_i^* a_i\right)'x + \sum_{i=1}^{m}\mu_i^* b_i\right\},$$

from which we obtain the complementary slackness conditions for (P):

$$\left(c_j - \sum_{i=1}^{m} \mu_i^* a_{ij} \right) x_j^* = 0, \quad x_j^* \geq 0, \quad \forall\, j = 1, \ldots, n,$$

$$c - \sum_{i=1}^{m} \mu_i^* a_i \geq 0.$$

The Lagrangian optimality condition for (D) is

$$\mu^* = \arg\min_{\mu \geq 0} \big\{ (Ax^* - b)'\mu - c'x^* \big\},$$

from which we obtain the complementary slackness conditions for (D):

$$Ax^* - b \geq 0, \qquad (Ax^* - b)_i \mu_i^* = 0, \quad \forall\, i.$$

6.6 (Duality and Zero Sum Games)

Consider the linear program

$$\min_{\substack{\zeta e \geq A'x \\ \sum_{i=1}^{n} x_i = 1,\, x_i \geq 0}} \zeta,$$

whose optimal value is equal to $\min_{x \in X} \max_{z \in Z} x'Az$. Introduce dual variables $z \in \Re^m$ and $\xi \in \Re$, corresponding to the constraints $A'x - \zeta e \leq 0$ and $\sum_{i=1}^{n} x_i = 1$, respectively. The dual function is

$$\begin{aligned} q(z, \xi) &= \inf_{x_i \geq 0,\, i=1,\ldots,n} \left\{ \zeta + z'(A'x - \zeta e) + \xi \left(1 - \sum_{i=1}^{n} x_i \right) \right\} \\ &= \inf_{x_i \geq 0,\, i=1,\ldots,n} \left\{ \zeta \left(1 - \sum_{j=1}^{m} z_j \right) + x'(Az - \xi e) + \xi \right\} \\ &= \begin{cases} \xi & \text{if } \sum_{j=1}^{m} z_j = 1,\ \xi e - Az \leq 0, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus the dual problem, which is to maximize $q(z, \xi)$ subject to $z \geq 0$ and $\xi \in \Re$, is equivalent to the linear program

$$\max_{\xi e \leq Az,\, z \in Z} \xi,$$

whose optimal value is equal to $\max_{z \in Z} \min_{x \in X} x'Az$.

6.7 (Goldman-Tucker Complementarity Theorem [GoT56])

Consider the subspace

$$S = \big\{(x, w) \mid bw - Ax = 0,\ c'x = wv,\ x \in \Re^n,\ w \in \Re\big\},$$

where v is the optimal value of (LP). Its orthogonal complement is the range of the matrix

$$\begin{bmatrix} -A' & c \\ b & -v \end{bmatrix},$$

so it has the form

$$S^\perp = \big\{(c\zeta - A'\lambda, b'\lambda - v\zeta) \mid \lambda \in \Re^m,\ \zeta \in \Re\big\}.$$

Applying the Tucker Complementarity Theorem (Exercise 3.32) for this choice of S, we obtain a partition of the index set $\{1, \ldots, n+1\}$ in two subsets. There are two possible cases: (1) the index $n+1$ belongs to the first subset, or (2) the index $n+1$ belongs to the second subset. Since the vectors $(x, 1)$ such that $x \in X^*$ satisfy $Ax - bw = 0$ and $c'x = wv$, we see that case (1) holds, i.e., the index $n+1$ belongs to the first index subset. In particular, we have that there exist disjoint index sets I and $\overline{I}$ such that $I \cup \overline{I} = \{1, \ldots, n\}$ and the following properties hold:

(a) There exist vectors $(x, w) \in S$ and $(\lambda, \zeta) \in \Re^{m+1}$ with the property

$$x_i > 0, \quad \forall\, i \in I, \qquad x_i = 0, \quad \forall\, i \in \overline{I}, \qquad w > 0, \tag{6.1}$$

$$c_i\zeta - (A'\lambda)_i = 0, \quad \forall\, i \in I, \quad c_i\zeta - (A'\lambda)_i > 0, \quad \forall\, i \in \overline{I}, \qquad b'\lambda = v\zeta. \tag{6.2}$$

(b) For all $(x, w) \in S$ with $x \geq 0$, and $(\lambda, \zeta) \in \Re^{m+1}$ with $c\zeta - A'\lambda \geq 0$, $v\zeta - b'\lambda \geq 0$, we have

$$x_i = 0, \qquad \forall\, i \in \overline{I},$$

$$c_i\zeta - (A'\lambda)_i = 0, \qquad \forall\, i \in I, \qquad b'\lambda = v\zeta.$$

By dividing (x, w) by w, we obtain [cf. Eq. (6.1)] an optimal primal solution $x^* = x/w$ such that

$$x_i^* > 0, \quad \forall\, i \in I, \qquad x_i^* = 0, \quad \forall\, i \in \overline{I}.$$

Similarly, if the scalar ζ in Eq. (6.2) is positive, by dividing with ζ in Eq. (6.2), we obtain an optimal dual solution $\lambda^* = \lambda/\zeta$, which satisfies the desired property

$$c_i - (A'\lambda^*)_i = 0, \quad \forall\, i \in I, \qquad c_i - (A'\lambda^*)_i > 0, \quad \forall\, i \in \overline{I}.$$

If the scalar ζ in Eq. (6.2) is nonpositive, we choose any optimal dual solution λ^*, and we note, using also property (b), that we have

$$c_i - (A'\lambda^*)_i = 0, \qquad \forall\, i \in I, \qquad c_i - (A'\lambda^*)_i \geq 0, \quad \forall\, i \in \overline{I}, \qquad b'\lambda^* = v. \tag{6.3}$$

Consider the vector

$$\tilde{\lambda} = (1 - \zeta)\lambda^* + \lambda.$$

By multiplying Eq. (6.3) with the positive number $1 - \zeta$, and by combining it with Eq. (6.2), we see that

$$c_i - (A'\tilde{\lambda})_i = 0, \qquad \forall\, i \in I, \qquad c_i - (A'\tilde{\lambda})_i > 0, \quad \forall\, i \in \overline{I}, \qquad b'\tilde{\lambda} = v.$$

Thus, $\tilde{\lambda}$ is an optimal dual solution that satisfies the desired property.

6.8

The problem of finding the minimum distance from the origin to a line is written as

$$\min \tfrac{1}{2}\|x\|^2$$

$$\text{subject to } Ax = b,$$

where A is a 2×3 matrix with full rank, and $b \in \Re^2$. Let f^* be the optimal value and consider the dual function

$$q(\lambda) = \min_x \left\{ \tfrac{1}{2}\|x\|^2 + \lambda'(Ax - b) \right\}.$$

By Prop. 6.3.1, since the optimal value is finite, it follows that this problem has no duality gap.

Let V^* be the supremum over all distances of the origin from planes that contain the line $\{x \mid Ax = b\}$. Clearly, we have $V^* \leq f^*$, since the distance to the line $\{x \mid Ax = b\}$ cannot be smaller than the distance to the plane that contains the line.

We now note that any plane of the form $\{x \mid p'Ax = p'b\}$, where $p \in \Re^2$, contains the line $\{x \mid Ax = b\}$, so we have for all $p \in \Re^2$,

$$V(p) \equiv \min_{p'Ax = p'x} \tfrac{1}{2}\|x\|^2 \leq V^*.$$

On the other hand, by duality in the minimization of the preceding equation, we have

$$U(p, \gamma) \equiv \min_x \left\{ \tfrac{1}{2}\|x\|^2 + \gamma(p'Ax - p'x) \right\} \leq V(p), \qquad \forall\, p \in \Re^2,\ \gamma \in \Re.$$

Combining the preceding relations, it follows that

$$\sup_\lambda q(\lambda) = \sup_{p,\gamma} U(p, \gamma) \leq \sup_p U(p, 1) \leq \sup_p V(p) \leq V^* \leq f^*.$$

Since there is no duality gap for the original problem, we have $\sup_\lambda q(\lambda) = f^*$, it follows that equality holds throughout above. Hence $V^* = f^*$, which was to be proved.

6.9

We introduce artificial variables $x_0, x_1, \ldots, x_m$, and we write the problem in the equivalent form

$$\begin{aligned} &\text{minimize } \sum_{i=0}^{m} f_i(x_i) \\ &\text{subject to } x_i \in X_i, \quad i = 0, \ldots, m \qquad x_i = x_0, \quad i = 1, \ldots, m. \end{aligned} \tag{6.4}$$

By relaxing the equality constraints, we obtain the dual function

$$
\begin{aligned}
q(\lambda_1, \ldots, \lambda_m) &= \inf_{x_i \in X_i,\ i=0,\ldots,m} \left\{ \sum_{i=0}^{m} f_i(x_i) + \lambda_i'(x_i - x_0) \right\} \\
&= \inf_{x \in X_0} \big\{ f_0(x) - (\lambda_1 + \cdots \lambda_m)'x \big\} + \sum_{i=1}^{m} \inf_{x \in X_i} \big\{ f_i(x) + \lambda_i' x \big\},
\end{aligned}
$$

which is of the form given in the exercise. Note that the infima above are attained since f_i are continuous (being convex functions over $\Re^n$) and X_i are compact polyhedra.

Because the primal problem involves minimization of the continuous function $\sum_{i=0}^{m} f_i(x)$ over the compact set $\cap_{i=0}^{m} X_i$, a primal optimal solution exists. Applying Prop. 6.4.2 to problem (6.4), we see that there is no duality gap and there exists at least one geometric multiplier, which is a dual optimal solution.

6.10

Let M denote the set of geometric multipliers, i.e.,

$$
M = \left\{ \mu \geq 0 \;\Big|\; f^* = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\} \right\}.
$$

We will show that if the set M is nonempty and compact, then the Slater condition holds. Indeed, if this were not so, then 0 would not be an interior point of the set

$$
D = \big\{ u \mid \text{there exists some } x \in X \text{ such that } g(x) \leq u \big\}.
$$

By a similar argument as in the proof of Prop. 6.6.1, it can be seen that D is convex. Therefore, we can use the Supporting Hyperplane Theorem to assert the existence of a hyperplane that passes through 0 and contains D in its positive halfspace, i.e., there is a nonzero vector $\overline{\mu}$ such that $\overline{\mu}' u \geq 0$ for all $u \in D$. This implies that $\overline{\mu} \geq 0$, since for each $u \in D$, we have that $(u_1, \ldots, u_j + \gamma, \ldots, u_r) \in D$ for all $\gamma > 0$ and j. Since $g(x) \in D$ for all $x \in X$, it follows that

$$
\overline{\mu}' g(x) \geq 0, \qquad \forall\, x \in X.
$$

Thus, for any $\mu \in M$, we have

$$
f(x) + (\mu + \gamma \overline{\mu})' g(x) \geq f^*, \qquad \forall\, x \in X,\ \forall\, \gamma \geq 0.
$$

Hence, it follows that $(\mu + \gamma\overline{\mu}) \in M$ for all $\gamma \geq 0$, which contradicts the boundedness of M.

6.11 (Inconsistent Convex Systems of Inequalities)

The dual function for the problem in the hint is

$$
\begin{aligned}
q(\mu) &= \inf_{y\in\Re,\, x\in X} \left\{ y + \sum_{j=1}^{r} \mu_j \big(g_j(x) - y\big) \right\} \\
&= \begin{cases} \inf_{x\in X} \sum_{j=1}^{r} \mu_j g_j(x) & \text{if } \sum_{j=1}^{r} \mu_j = 1, \\ -\infty & \text{if } \sum_{j=1}^{r} \mu_j \neq 1. \end{cases}
\end{aligned}
$$

The problem in the hint satisfies Assumption 6.4.2, so by Prop. 6.4.3, the dual problem has an optimal solution μ^* and there is no duality gap.

Clearly the problem in the hint has an optimal value that is greater or equal to 0 if and only if the system of inequalities

$$
g_j(x) < 0, \qquad j = 1, \ldots, r,
$$

has no solution within X. Since there is no duality gap, we have

$$
\max_{\mu\geq 0,\, \sum_{j=1}^{r} \mu_j = 1} q(\mu) \geq 0
$$

if and only if the system of inequalities $g_j(x) < 0$, $j = 1, \ldots, r$, has no solution within X. This is equivalent to the statement we want to prove.

6.12

Since $c \in \mathrm{cone}\{a_1, \ldots, a_r\} + \mathrm{ri}(N)$, there exists a vector $\overline{\mu} \geq 0$ such that

$$
-\left(-c + \sum_{j=1}^{r} \overline{\mu}_j a_j \right) \in \mathrm{ri}(N).
$$

By Example 6.4.2, this implies that the problem

$$
\begin{aligned}
&\text{minimize} \quad -c'd + \tfrac{1}{2} \textstyle\sum_{j=1}^{r} \big((a_j'd)^+\big)^2 \\
&\text{subject to} \quad d \in N^*,
\end{aligned}
$$

has an optimal solution, which we denote by d^*. Consider the set

$$
M = \left\{ \mu \geq 0 \;\middle|\; -\left(-c + \sum_{j=1}^{r} \mu_j a_j \right) \in N \right\},
$$

which is nonempty by assumption. Let μ^* be the vector of minimum norm in M and let the index set J be defined by

$$
J = \{ j \mid \mu_j^* > 0 \}.
$$

Then, it follows from Lemma 5.3.1 that

$$
a_j'd^* > 0, \qquad \forall\, j \in J,
$$

and

$$
a_j'd^* \leq 0, \qquad \forall\, j \notin J,
$$

thus proving that the properties (1) and (2) of the exercise hold.

6.13 (Pareto Optimality)

(a) Assume that x^* is not a Pareto optimal solution. Then there is a vector $\overline{x} \in X$ such that either

$$f_1(\bar{x}) \leq f_1(x^*), \quad f_2(\bar{x}) < f_2(x^*),$$

or

$$f_1(\bar{x}) < f_1(x^*), \quad f_2(\bar{x}) \leq f_2(x^*).$$

Multiplying the left equation by λ_1^*, the right equation by λ_2^*, and adding the two in either case yields

$$\lambda_1^* f_1(\bar{x}) + \lambda_2^* f_2(\bar{x}) < \lambda_1^* f_1(x^*) + \lambda_2^* f_2(x^*),$$

yielding a contradiction. Therefore x^* is a Pareto optimal solution.

(b) Let

$$A = \big\{(z_1, z_2) | \text{ there exists } x \in X \text{ such that } f_1(x) \leq z_1, f_2(x) \leq z_2\big\}.$$

We first show that A is convex. Indeed, let (a_1, a_2), and (b_1, b_2) be elements of A, and let $(c_1, c_2) = \alpha(a_1, a_2) + (1 - \alpha)(b_1, b_2)$ for any $\alpha \in [0, 1]$. Then for some $x_a \in X$, $x_b \in X$, we have $f_1(x_a) \leq a_1$, $f_2(x_a) \leq a_2$, $f_1(x_b) \leq b_1$, and $f_2(x_b) \leq b_2$. Let $x_c = \alpha x_a + (1 - \alpha)x_b$. Since X is convex, $x_c \in X$. Since f is convex, we also have

$$f_1(x_c) \leq c_1, \quad \text{and} \quad f_2(x_c) \leq c_2.$$

Hence, $(c_1, c_2) \in A$ and it follows that A is a convex set.

For any $x \in X$, we have $\big(f_1(x), f_2(x)\big) \in A$. In addition, $\big(f_1(x^*), f_2(x^*)\big)$ is in the boundary of A. [If this were not the case, then either (1) or (2) would hold and x^* would not be Pareto optimal.] Then by the Supporting Hyperplane Theorem, there exists λ_1^* and λ_2^*, not both equal to 0, such that

$$\lambda_1^* z_1 + \lambda_2^* z_2 \geq \lambda_1^* f_1(x^*) + \lambda_2^* f_2(x^*), \qquad \forall\ (z_1, z_2) \in A.$$

Since z_1 and z_2 can be made arbitrarily large, we must have $\lambda_1^*, \lambda_2^* \geq 0$. Since $\big(f_1(x), f_2(x)\big) \in A$, the above equation yields

$$\lambda_1^* f_1(x) + \lambda_2^* f_2(x) \geq \lambda_1^* f_1(x^*) + \lambda_2^* f_2(x^*), \quad \forall x \in X,$$

or, equivalently,

$$\min_{x \in X} \big\{\lambda_1^* f_1(x) + \lambda_2^* f_2(x)\big\} \geq \lambda_1^* f_1(x^*) + \lambda_2^* f_2(x^*).$$

Combining this with the fact that

$$\min_{x \in X} \big\{\lambda_1^* f_1(x) + \lambda_2^* f_2(x)\big\} \leq \lambda_1^* f_1(x^*) + \lambda_2^* f_2(x^*)$$

yields the desired result.

(c) Generalization of (a): If x^* is a vector in X, and $\lambda_1^*, \ldots, \lambda_m^*$ are positive scalars such that

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = \min_{x \in X} \left\{ \sum_{i=1}^{m} \lambda_i^* f_i(x) \right\},$$

then x^* is a Pareto optimal solution.
Generalization of (b): Assume that X is convex and $f_1, \ldots, f_m$ are convex over X. If x^* is a Pareto optimal solution, then there exist non-negative scalars $\lambda_1^*, \ldots, \lambda_m^*$, not all zero, such that

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = \min_{x \in X} \left\{ \sum_{i=1}^{m} \lambda_i^* f_i(x) \right\}.$$

6.14 (Polyhedral Programming)

Using Prop. 3.2.3, it follows that

$$g_j(x) = \max_{i=1,\ldots,m} \left\{ a_{ij}' x + b_{ij} \right\},$$

where a_{ij} are vectors in $\Re^n$ and b_{ij} are scalars. Hence the constraint functions can equivalently be represented as

$$a_{ij}' x + b_{ij} \leq 0, \qquad \forall\ i = 1, \ldots, m,\ \forall\ j = 1, \ldots, r.$$

By assumption, the set X is a polyhedral set, and the cost function f is a polyhedral function, hence convex over $\Re^n$. Therefore, we can use the Strong Duality Theorem for linear constraints (cf. Prop. 6.4.2) to conclude that there is no duality gap and there exists at least one geometric multiplier, i.e., there exists a nonnegative vector μ such that

$$f^* = \inf_{x \in X} \left\{ f(x) + \mu' g(x) \right\}.$$

Let $p(u)$ denote the primal function for this problem. The preceding relation implies that

$$\begin{aligned} p(0) - \mu' u &= \inf_{x \in X} \Big\{ f(x) + \mu' \big(g(x) - u \big) \Big\} \\ &\leq \inf_{x \in X,\ g(x) \leq u} \Big\{ f(x) + \mu' \big(g(x) - u \big) \Big\} \\ &\leq \inf_{x \in X,\ g(x) \leq u} f(x) \\ &= p(u), \end{aligned}$$

which, in view of the assumption that $p(0)$ is finite, shows that $p(u) > -\infty$ for all $u \in \Re^r$.

The primal function can be obtained by partial minimization as

$$p(u) = \inf_{x \in \Re^n} F(x, u),$$

where

$$F(x, u) = \begin{cases} f(x) & \text{if } g_j(x) \leq u_j \ \forall\ j,\ x \in X, \\ \infty & \text{otherwise.} \end{cases}$$

Since, by assumption f is polyhedral, the g_j are polyhedral (which implies that the level sets of the g_j are polyhedral), and X is polyhedral, it follows that $F(x, u)$ is a polyhedral function. Since we have also shown that $p(u) > -\infty$ for all $u \in \Re^r$, we can use Exercise 3.13 to conclude that the primal function p is polyhedral, and therefore also closed. Since $p(0)$, the optimal value, is assumed finite, it follows that p is proper.

Convex Analysis and Optimization

Chapter 7 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 7: SOLUTION MANUAL

7.1 (Fenchel's Inequality)

(a) From the definition of g,

$$g(\lambda) = \sup_{x\in\Re^n} \big\{x'\lambda - f(x)\big\},$$

we have the inequality $x'\lambda \le f(x) + g(\lambda)$. In view of this inequality, the equality $x'\lambda = f(x) + g(\lambda)$ of (i) is equivalent to the inequality

$$x'\lambda - f(x) \ge g(\lambda) = \sup_{z\in\Re^n} \big\{z'\lambda - f(z)\big\},$$

or

$$x'\lambda - f(x) \ge z'\lambda - f(z), \qquad \forall\ z \in \Re^n,$$

or

$$f(z) \ge f(x) + \lambda'(z - x), \qquad \forall\ z \in \Re^n,$$

which is equivalent to (ii). Since f is closed, f is equal to the conjugate of g, so by using the equivalence of (i) and (ii) with the roles of f and g reversed, we obtain the equivalence of (i) and (iii).

(b) A vector x^* minimizes f if and only if $0 \in \partial f(x^*)$, which by part (a), is true if and only if $x^* \in \partial g(0)$.

(c) The result follows by combining part (b) and Prop. 4.4.2.

7.2

Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function, and let g be its conjugate. Show that the lineality space of g is equal to the orthogonal complement of the subspace parallel to $\text{aff}\big(\text{dom}(f)\big)$.

7.3

Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be proper convex functions, and let $f = f_1 + \cdots + f_m$. Show that if $\cap_{i=1}^m \text{ri}\big(\text{dom}(f_i)\big)$ is nonempty, then we have

$$g(\lambda) = \inf_{\substack{\lambda_1+\cdots+\lambda_m=\lambda \\ \lambda_i\in\Re^n,\, i=1,\ldots,m}} \big\{g_1(\lambda_1) + \cdots + g_m(\lambda_m)\big\}, \qquad \forall\ \lambda \in \Re^n,$$

where $g, g_1, \ldots, g_m$ are the conjugates of $f, f_1, \ldots, f_m$, respectively.

7.4 (Finiteness of the Optimal Dual Value)

Consider the function $\tilde{q}$ given by

$$\tilde{q}(\mu) = \begin{cases} q(\mu) & \text{if } \mu \geq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

and note that $-\tilde{q}$ is closed and convex, and that by the calculation of Example 7.1.6, we have

$$\tilde{q}(\mu) = \inf_{u \in \Re^r} \big\{ p(u) + \mu' u \big\}, \qquad \forall\, \mu \in \Re^r. \tag{1}$$

Since $\tilde{q}(\mu) \leq p(0)$ for all $\mu \in \Re^r$, given the feasibility of the problem [i.e., $p(0) < \infty$], we see that q^* is finite if and only if $-\tilde{q}$ is proper. From Eq. (1), $-\tilde{q}$ is the conjugate of $p(-u)$, and by the Conjugacy Theorem [Prop. 7.1.1(b)], $-\tilde{q}$ is proper if and only if p is proper. Hence, (i) is equivalent to (ii).

We note that the epigraph of p is the closure of M. Hence, given the feasibility of the problem, (ii) is equivalent to the closure of M not containing a vertical line. Since M is convex, its closure does not contain a line if and only if M does not contain a line (since the closure and the relative interior of M have the same recession cone). Hence (ii) is equivalent to (iii).

7.5 (General Perturbations and Min Common/Max Crossing Duality)

(a) We have

$$\begin{aligned} h(\lambda) &= \sup_u \big\{ \lambda' u - p(u) \big\} \\ &= \sup_u \big\{ \lambda' u - \inf_x F(x,u) \big\} \\ &= \sup_{x,u} \big\{ \lambda' u - F(x,u) \big\} \\ &= G(0, \lambda). \end{aligned}$$

Also

$$\begin{aligned} q(\lambda) &= \inf_{(u,w) \in M} \{ w + \lambda' u \} \\ &= \inf_{x,u} \big\{ F(x,u) + \lambda' u \big\} \\ &= -\sup_{x,u} \big\{ -\lambda' u - F(x,u) \big\} \\ &= -G(0, -\lambda). \end{aligned}$$

Consider the constrained minimization propblem of Example 7.1.6:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad g(x) \leq 0, \end{aligned}$$

and define

$$F(x,u) = \begin{cases} f(x) & \text{if } x \in X \text{ and } g(x) \leq u, \\ \infty & \text{otherwise.} \end{cases}$$

Then p is the primal function of the constrained minimization problem. Consider now $q(\lambda)$, the cost function of the max crossing problem corresponding to M. For $\lambda \geq 0$, $q(\lambda)$ is equal to the dual function value of the constrained optimization problem, and otherwise $q(\lambda)$ is equal to $-\infty$. Thus, the relations $h(\lambda) = G(0, \lambda)$ and $q(\lambda) = -G(0, -\lambda)$ proved earlier, show the relation proved in Example 7.1.6, i.e., that $q(\lambda) = -h(-\lambda)$.

(b) Let

$$M = \{(u, w) \mid \text{there is an } x \text{ such that } F(x, u) \leq w\}.$$

Then the corresponding min common value is

$$\inf_{\{(x,w) \ \mid \ F(x,0) \leq w\}} w = \inf_x F(x, 0) = p(0).$$

Since $p(0)$ is the min common value corresponding to $\text{epi}(p)$, the min common values corresponding to the two choises for M are equal. Similarly, we show that the cost functions of the max crossing problem corresponding to the two choises for M are equal.

(c) If $F(x, u) = f_1(x) - f_2(Qx + u)$, we have

$$p(u) = \inf_x \{f_1(x) - f_2(Qx + u)\},$$

so $p(0)$, the min common value, is equal to the primal optimal value in the Fenchel duality framework. By part (a), the max crossing value is

$$q^* = \sup_\lambda \{-h(-\lambda)\},$$

where h is the conjugate of p. By using the change of variables $z = Qx + u$ in the following calculation, we have

$$\begin{aligned}
-h(-\lambda) &= -\sup_u \{-\lambda' u - \inf_x \{f_1(x) - f_2(Qx + u)\}\} \\
&= -\sup_{z,x} \{-\lambda'(z - Qx) - f_1(x) + f_2(z)\} \\
&= g_2(\lambda) - g_1(Q\lambda),
\end{aligned}$$

where g_1 and g_2 are the conjugate convex and conjugate concave functions of f_1 and f_2, respectively:

$$g_1(\lambda) = \sup_x \{x'\lambda - f_1(x)\}, \qquad g_2(\lambda) = \inf_z \{z'\lambda - f_2(z)\}.$$

Thus, no duality gap in the min common/max crossing framework [i.e., $p(0) = q^* = \sup_\lambda \{-h(-\lambda)\}$] is equivalent to no duality gap in the Fenchel duality framework.

The minimax framework of Section 2.6.1 (using the notation of that section) is obtained for

$$F(x, u) = \sup_{z \in Z} \{\phi(x, z) - u'z\}.$$

The constrained optimization framework of Section 6.1 (using the notation of that section) is obtained for the function

$$F(x, u) = \begin{cases} f(x) & \text{if } x \in X, \ h(x) = u_1, \ g(x) \leq u_2, \\ \infty & \text{otherwise,} \end{cases}$$

where $u = (u_1, u_2)$.

7.6

By Exercise 1.35,

$$\mathrm{cl}\, f_1 + \mathrm{cl}\,(-f_2) = \mathrm{cl}\,(f_1 - f_2).$$

Furthermore,

$$\inf_{x\in\Re^n} \mathrm{cl}\,(f_1 - f_2)(x) = inf_{x\in\Re^n}\big(f_1(x) - f_2(x)\big).$$

Thus, we may replace f_1 and $-f_2$ with their closures, and the result follows by applying Minimax Theorem III.

7.7 (Monotropic Programming Duality)

We apply Fenchel duality with

$$f_1(x) = \begin{cases} \sum_{i=1}^n f_i(x_i) & \text{if } x \in X_1 \times \cdots \times X_n, \\ \infty & \text{otherwise,} \end{cases}$$

and

$$f_2(x) = \begin{cases} 0 & \text{if } x \in S, \\ -\infty & \text{otherwise.} \end{cases}$$

The corresponding conjugate concave and convex functions g_2 and g_1 are

$$\inf_{x\in S} \lambda' x = \begin{cases} 0 & \text{if } \lambda \in S^\perp, \\ -\infty & \text{if } \lambda \notin S^\perp, \end{cases}$$

where $S^\perp$ is the orthogonal subspace of S, and

$$\sup_{x_i\in X_i} \left\{ \sum_{i=1}^n \big(x_i\lambda_i - f_i(x_i)\big) \right\} = \sum_{i=1}^n g_i(\lambda_i),$$

where for each i,

$$g_i(\lambda_i) = \sup_{x_i\in X_i} \big\{x_i\lambda_i - f_i(x_i)\big\}.$$

By the Primal Fenchel Duality Theorem (Prop. 7.2.1), the dual problem has an optimal solution and there is no duality gap if the functions f_i are convex over X_i and one of the following two conditions holds:

(1) The subspace S contains a point in the relative interior of $X_1 \times \cdots \times X_n$.

(2) The intervals X_i are closed (so that the Cartesian product $X_1 \times \cdots \times X_n$ is a polyhedral set) and the functions f_i are convex over the entire real line.

These conditions correspond to the two conditions for no duality gap given following Prop. 7.2.1.

7.8 (Network Optimization and Kirchhoff's Laws)

This problem is a monotropic programming problem, as considered in Exercise 7.7. For each $(i,j) \in \mathcal{A}$, the function $f_{ij}(x_{ij}) = \frac{1}{2}R_{ij}x_{ij}^2 - t_{ij}x_{ij}$ is continuously differentiable and convex over $\Re$. The dual problem is

$$\begin{aligned} &\text{maximize } q(v) \\ &\text{subject to no constraints on } p, \end{aligned}$$

with the dual function q given by

$$q(v) = \sum_{(i,j)\in\mathcal{A}} q_{ij}(v_i - v_j),$$

where

$$q_{ij}(v_i - v_j) = \min_{x_{ij}\in\Re} \left\{ \frac{1}{2}R_{ij}x_{ij}^2 - (v_i - v_j + t_{ij})x_{ij} \right\}.$$

Since the primal cost functions f_{ij} are real-valued and convex over the entire real line, there is no duality gap. The necessary and sufficient conditions for a set of variables $\{x_{ij} \mid (i,j) \in \mathcal{A}\}$ and $\{v_i \mid i \in \mathcal{N}\}$ to be an optimal solution-Lagrange multiplier pair are:

(1) The set of variables $\{x_{ij} \mid (i,j) \in \mathcal{A}\}$ must be primal feasible, i.e., Kirchhoff's current law must be satisfied.

(2)
$$x_{ij} \in \arg\min_{y_{ij}\in\Re} \left\{ \frac{1}{2}R_{ij}y_{ij}^2 - (v_i - v_j + t_{ij})y_{ij} \right\}, \quad \forall\ (i,j) \in \mathcal{A},$$

which is equivalent to Ohm's law:

$$R_{ij}x_{ij} - (v_i - v_j + t_{ij}) = 0, \quad \forall\ (i,j) \in \mathcal{A}.$$

Hence a set of variables $\{x_{ij} \mid (i,j) \in \mathcal{A}\}$ and $\{v_i \mid i \in \mathcal{N}\}$ are an optimal solution-Lagrange multiplier pair if and only if they satisfy Kirchhoff's current law and Ohm's law.

7.9 (Symmetry of Duality)

(a) We have $f^* = p(0)$. Since $p(u)$ is monotonically nonincreasing, its minimal value over $u \in P$ and $u \leq 0$ is attained for $u = 0$. Hence, $f^* = p^*$, where $p^* = \inf_{u\in P,\, u\leq 0} p(u)$. For $\mu \geq 0$, we have

$$\begin{aligned} \inf_{x\in X} \big\{ f(x) + \mu' g(x) \big\} &= \inf_{u\in P} \inf_{x\in X,\, g(x)\leq u} \big\{ f(x) + \mu' g(x) \big\} \\ &= \inf_{u\in P} \big\{ p(u) + \mu' u \big\}. \end{aligned}$$

Since $f^* = p^*$, we see that $f^* = \inf_{x \in X}\big\{f(x) + \mu' g(x)\big\}$ if and only if $p^* = \inf_{u \in P}\big\{p(u) + \mu' u\big\}$. In other words, the two problems have the same geometric multipliers.

(b) This part was proved by the preceding argument.

(c) From Example 7.1.6, we have that $-q(-\mu)$ is the conjugate convex function of p. Let us view the dual problem as the minimization problem

$$\begin{aligned} &\text{minimize} \quad -q(-\mu) \\ &\text{subject to} \ \ \mu \leq 0. \end{aligned} \tag{1}$$

Its dual problem is obtained by forming the conjugate convex function of its primal function, which is p, based on the analysis of Example 7.1.6, and the closedness and convexity of p. Hence the dual of the dual problem (1) is

$$\begin{aligned} &\text{maximize} \quad -p(u) \\ &\text{subject to} \ \ u \leq 0 \end{aligned}$$

and the optimal solutions to this problem are the geometric multipliers to problem (1).

7.10 (Second-Order Cone Programming)

(a) Define

$$X = \Big\{(x,u,t) \mid x \in \Re^n,\, u_j = A_j x + b_j,\ t_j = e_j' x + d_j,\, j = 1,\ldots,r\Big\},$$

$$C = \Big\{(x,u,t) \mid x \in \Re^n,\ \|u_j\| \leq t_j,\, j = 1,\ldots,r\Big\}.$$

It can be seen that X is convex and C is a cone. Therefore the modified problem can be written as

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ x \in X \cap C, \end{aligned}$$

and is a cone programming problem of the type described in Section 7.2.2.

(b) Let $(\lambda, z, w) \in \hat{C}$, where $\hat{C}$ is the dual cone ($\hat{C} = -C^*$, where C^* is the polar cone). Then we have

$$\lambda' x + \sum_{j=1}^{r} z_j' u_j + \sum_{j=1}^{r} w_j t_j \geq 0, \qquad \forall\ (x,u,t) \in C.$$

Since x is unconstrained, we must have $\lambda = 0$ for otherwise the above inequality will be violated. Furthermore, it can be seen that

$$\hat{C} = \Big\{(0,z,w) \mid \|z_j\| \leq w_j,\, j = 1,\ldots,r\Big\}.$$

By the conic duality theory of Section 7.2.2, the dual problem is given by

$$\begin{aligned}&\text{minimize } \sum_{j=1}^{r}(z_j'b_j + w_j d_j)\\&\text{subject to } \sum_{j=1}^{r}(A_j'z_j + w_j e_j) = c, \quad \|z_j\| \le w_j,\ j = 1,\ldots,r.\end{aligned}$$

If there exists a feasible solution of the modified primal problem satisfying strictly all the inequality constraints, then the relative interior condition $\mathrm{ri}(X)\cap\mathrm{ri}(C) \neq \emptyset$ is satisfied, and there is no duality gap. Similarly, if there exists a feasible solution of the dual problem satisfying strictly all the inequality constraints, there is no duality gap.

7.11 (Quadratically Constrained Quadratic Problems [LVB98])

Since each P_i is symmetric and positive definite, we have

$$\begin{aligned}x'P_i x + 2q_i'x + r_i &= \left(P_i^{1/2}x\right)' P_i^{1/2}x + 2\left(P_i^{-1/2}q_i\right)' P_i^{1/2}x + r_i\\&= \|P_i^{1/2}x + P_i^{-1/2}q_i\|^2 + r_i - q_i'P_i^{-1}q_i,\end{aligned}$$

for $i = 0, 1, \ldots, p$. This allows us to write the original problem as

$$\begin{aligned}&\text{minimize } \|P_0^{1/2}x + P_0^{-1/2}q_0\|^2 + r_0 - q_0'P_0^{-1}q_0\\&\text{subject to } \|P_i^{1/2}x + P_i^{-1/2}q_i\|^2 + r_i - q_i'P_i^{-1}q_i \le 0,\ i = 1,\ldots,p.\end{aligned}$$

By introducing a new variable x_{n+1}, this problem can be formulated in $\Re^{n+1}$ as

$$\begin{aligned}&\text{minimize } x_{n+1}\\&\text{subject to } \|P_0^{1/2}x + P_0^{-1/2}q_0\| \le x_{n+1}\\&\|P_i^{1/2}x + P_i^{-1/2}q_i\| \le \left(q_i'P_i^{-1}q_i - r_i\right)^{1/2}, \quad i = 1,\ldots,p.\end{aligned}$$

The optimal values of this problem and the original problem are equal up to a constant and a square root. The above problem is of the type described in Exercise 7.10. To see this, define $A_i = \left(P_i^{1/2} \mid 0\right)$, $b_i = P_i^{-1/2}q_i$, $e_i = 0$, $d_i = \left(q_i'P_i^{-1}q_i - r_i\right)^{1/2}$ for $i = 1,\ldots,p$, $A_0 = \left(P_0^{1/2} \mid 0\right)$, $b_0 = P_0^{-1/2}q_0$, $e_0 = (0,\ldots,0,1)$, $d_0 = 0$, and $c = (0,\ldots,0,1)$. Its dual is given by

$$\begin{aligned}&\text{maximize } -\sum_{i=1}^{p}\left(q_i'P_i^{-1/2}z_i + \left(q_i'P_i^{-1}q_i - r_i\right)^{1/2} w_i\right) - q_0'P_0^{-1/2}z_0\\&\text{subject to } \sum_{i=0}^{p}P_i^{1/2}z_i = 0, \quad \|z_0\| \le 1, \quad \|z_i\| \le w_i, \quad i = 1,\ldots,p.\end{aligned}$$

7.12 (Minimizing the Sum or the Maximum of Norms [LVB98])

Consider the problem

$$\begin{aligned} &\text{minimize} \ \sum_{i=1}^{p} \|F_i x + g_i\| \\ &\text{subject to } x \in \Re^n. \end{aligned}$$

By introducing variables $t_1, \ldots, t_p$, this problem can be expressed as a second-order cone programming problem (see Exercise 7.10):

$$\begin{aligned} &\text{minimize} \ \sum_{i=1}^{p} t_i \\ &\text{subject to } \|F_i x + g_i\| \le t_i, \quad i = 1, \ldots, p. \end{aligned}$$

Define

$$X = \{(x, u, t) \mid x \in \Re^n, \ u_i = F_i x + g_i, \ t_i \in \Re, \ i = 1, \ldots, p\},$$

$$C = \{(x, u, t) \mid x \in \Re^n, \ \|u_i\| \le t_i, \quad i = 1, \ldots, p\}.$$

Then, similar to Exercise 7.10, we have

$$-C^* = \{(0, z, w) \mid \|z_i\| \le w_i, \quad i = 1, \ldots, p\},$$

and

$$\begin{aligned} g(0, z, w) &= \sup_{(x,u,t) \in X} \left\{ \sum_{i=1}^{p} z_i' u_i + \sum_{i=1}^{p} w_i t_i - \sum_{i=1}^{p} t_i \right\} \\ &= \sup_{x \in \Re^n, t \in \Re^p} \left\{ \sum_{i=1}^{p} z_i'(F_i x + g_i) + \sum_{i=1}^{p} (w_i - 1) t_i \right\} \\ &= \sup_{x \in \Re^n} \left\{ \left(\sum_{i=1}^{p} F_i' z_i \right)' x \right\} + \sup_{t \in \Re^p} \left\{ \sum_{i=1}^{p} (w_i - 1) t_i \right\} + \sum_{i=1}^{p} g_i' z_i \\ &= \begin{cases} \sum_{i=1}^{p} g_i' z_i & \text{if } \sum_{i=1}^{p} F_i' z_i = 0, \ w_i = 1, \ i = 1, \ldots, p \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Hence the dual problem is given by

$$\begin{aligned} &\text{maximize} \ -\sum_{i=1}^{p} g_i' z_i \\ &\text{subject to} \ \sum_{i=1}^{p} F_i' z_i = 0, \quad \|z_i\| \le 1, \ i = 1, \ldots, p. \end{aligned}$$

Now, consider the problem

$$\begin{aligned} &\text{minimize } \max_{1 \le i \le p} ||F_i x + g_i|| \\ &\text{subject to } x \in \Re^n. \end{aligned}$$

By introducing a new variable x_{n+1}, we obtain

$$\begin{aligned} &\text{minimize } x_{n+1} \\ &\text{subject to } ||F_i x + g_i|| \le x_{n+1}, \ i = 1, \ldots, p, \end{aligned}$$

or equivalently

$$\begin{aligned} &\text{minimize } e_{n+1}' x \\ &\text{subject to } ||A_i x + g_i|| \le e_{n+1}' x, \ i = 1, \ldots, p, \end{aligned}$$

where $x \in \Re^{n+1}$, $A_i = (F_i, 0)$, and $e_{n+1} = (0, \ldots, 0, 1)' \in \Re^{n+1}$. Evidently, this is a second-order cone programming problem. From Exercise 7.10 we have that its dual problem is given by

$$\begin{aligned} &\text{maximize } -\sum_{i=1}^{p} g_i' z_i \\ &\text{subject to } \sum_{i=1}^{p} \left(\begin{pmatrix} F_i' \\ 0 \end{pmatrix} z_i + e_{n+1} w_i \right) = e_{n+1}, \quad ||z_i|| \le w_i, \quad i = 1, \ldots, p, \end{aligned}$$

or equivalently

$$\begin{aligned} &\text{maximize } -\sum_{i=1}^{p} g_i' z_i \\ &\text{subject to } \sum_{i=1}^{p} F_i' z_i = 0, \quad \sum_{i=1}^{p} w_i = 1, \quad ||z_i|| \le w_i, \quad i = 1, \ldots, p. \end{aligned}$$

7.13 (Complex l_1 and l_∞ Approximation [LVB98])

For $v \in \mathcal{C}^p$ we have

$$\|v\|_1 = \sum_{i=1}^{p} |v_i| = \sum_{i=1}^{p} \left|\left| \begin{pmatrix} \mathcal{R}e(v_i) \\ \mathcal{I}m(v_i) \end{pmatrix} \right|\right|,$$

where $\mathcal{R}e(v_i)$ and $\mathcal{I}m(v_i)$ denote the real and the imaginary parts of v_i, respectively. Then the complex l_1 approximation problem is equivalent to

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{p} \left|\left| \begin{pmatrix} \mathcal{R}e(a_i' x - b_i) \\ \mathcal{I}m(a_i' x - b_i) \end{pmatrix} \right|\right| \\ &\text{subject to } x \in \mathcal{C}^n, \end{aligned} \tag{1}$$

where a_i' is the i-th row of A (A is a $p \times n$ matrix). Note that

$$\begin{pmatrix} \mathcal{R}e(a_i'x - b_i) \\ \mathcal{I}m(a_i'x - b_i) \end{pmatrix} = \begin{pmatrix} \mathcal{R}e(a_i') & -\mathcal{I}m(a_i') \\ \mathcal{I}m(a_i') & \mathcal{R}e(a_i') \end{pmatrix} \begin{pmatrix} \mathcal{R}e(x) \\ \mathcal{I}m(x) \end{pmatrix} - \begin{pmatrix} \mathcal{R}e(b_i) \\ \mathcal{I}m(b_i) \end{pmatrix}.$$

By introducing new variables $y = (\mathcal{R}e(x'), \mathcal{I}m(x'))'$, problem (1) can be rewritten as

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{p} \|F_i y + g_i\| \\ &\text{subject to } y \in \Re^{2n}, \end{aligned}$$

where

$$F_i = \begin{pmatrix} \mathcal{R}e(a_i') & -\mathcal{I}m(a_i') \\ \mathcal{I}m(a_i') & \mathcal{R}e(a_i') \end{pmatrix}, \qquad g_i = -\begin{pmatrix} \mathcal{R}e(b_i) \\ \mathcal{I}m(b_i) \end{pmatrix}. \tag{2}$$

According to Exercise 7.12, the dual problem is given by

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{p} \big(\mathcal{R}e(b_i), \mathcal{I}m(b_i)\big) z_i \\ &\text{subject to } \sum_{i=1}^{p} \begin{pmatrix} \mathcal{R}e(a_i') & \mathcal{I}m(a_i') \\ -\mathcal{I}m(a_i') & \mathcal{R}e(a_i') \end{pmatrix} z_i = 0, \qquad \|z_i\| \le 1, \quad i = 1, \ldots, p, \end{aligned}$$

where $z_i \in \Re^{2n}$ for all i.

For $v \in \mathcal{C}^p$ we have

$$\|v\|_\infty = \max_{1 \le i \le p} |v_i| = \max_{1 \le i \le p} \left\| \begin{pmatrix} \mathcal{R}e(v_i) \\ \mathcal{I}m(v_i) \end{pmatrix} \right\|.$$

Therefore the complex l_∞ approximation problem is equivalent to

$$\begin{aligned} &\text{minimize } \max_{1 \le i \le p} \left\| \begin{pmatrix} \mathcal{R}e(a_i'x - b_i) \\ \mathcal{I}m(a_i'x - b_i) \end{pmatrix} \right\| \\ &\text{subject to } x \in \mathcal{C}^n. \end{aligned}$$

By introducing new variables $y = (\mathcal{R}e(x'), \mathcal{I}m(x'))'$, this problem can be rewritten as

$$\begin{aligned} &\text{minimize } \max_{1 \le i \le p} \|F_i y + g_i\| \\ &\text{subject to } y \in \Re^{2n}, \end{aligned}$$

where F_i and g_i are given by Eq. (2). From Exercise 7.12, it follows that the dual problem is

$$\begin{aligned} &\text{maximize } \sum_{i=1}^{p} \big(\mathcal{R}e(b_i), \mathcal{I}m(b_i)\big) z_i \\ &\text{subject to } \sum_{i=1}^{p} \begin{pmatrix} \mathcal{R}e(a_i') & -\mathcal{I}m(a_i') \\ \mathcal{I}m(a_i') & \mathcal{R}e(a_i') \end{pmatrix} z_i = 0, \quad \sum_{i=1}^{p} w_i = 1, \quad \|z_i\| \le w_i, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1, \ldots, p, \end{aligned}$$

where $z_i \in \Re^2$ for all i.

7.14

The condition $u'\mu^* \le P(u)$ for all $u \in \Re^r$ can be written as

$$\sum_{j=1}^{r} u_j \mu_j^* \le \sum_{j=1}^{r} P_j(u_j), \qquad \forall\, u = (u_1, \ldots, u_r),$$

and is equivalent to

$$u_j \mu_j^* \le P_j(u_j), \qquad \forall\, u_j \in \Re,\ \forall\, j = 1, \ldots, r.$$

In view of the requirement that P_j is convex with $P_j(u_j) = 0$ for $u_j \le 0$, and $P_j(u_j) > 0$ for all $u_j > 0$, it follows that the condition $u_j\mu_j^* \le P_j(u_j)$ for all $u_j \in \Re$, is equivalent to $\mu_j^* \le \lim_{z_j \downarrow 0}\big(P_j(z_j)/z_j\big)$. Similarly, the condition $u_j\mu_j^* < P_j(u_j)$ for all $u_j \in \Re$, is equivalent to $\mu_j^* < \lim_{z_j \downarrow 0}\big(P_j(z_j)/z_j\big)$.

7.15 [Ber99b]

Following [Ber99b], we address the problem by embedding it in a broader class of problems. Let Y be a subset of $\Re^n$, let y be a parameter vector taking values in Y, and consider the parametric program

$$\begin{aligned} &\text{minimize} \ \ f(x, y) \\ &\text{subject to} \ \ x \in X, \qquad g_j(x, y) \le 0, \quad j = 1, \ldots, r, \end{aligned} \tag{1}$$

where X is a convex subset of $\Re^n$, and for each $y \in Y$, $f(\cdot, y)$ and $g_j(\cdot, y)$ are real-valued functions that are convex over X. We assume that for each $y \in Y$, this program has a finite optimal value, denoted by $f^*(y)$. Let $c > 0$ denote a penalty parameter and assume that the penalized problem

$$\begin{aligned} &\text{minimize} \ \ f(x, y) + c\big\|g^+(x, y)\big\| \\ &\text{subject to} \ \ x \in X \end{aligned} \tag{2}$$

has a finite optimal value, thereby coming under the framework of Section 7.3. By Prop. 7.3.1, we have

$$f^*(y) = \inf_{x \in X} \Big\{ f(x, y) + c\big\|g^+(x, y)\big\| \Big\}, \qquad \forall\, y \in Y, \tag{3}$$

if and only if

$$u'\mu^*(y) \le c\|u^+\|, \qquad \forall\, u \in \Re^r,\ \forall\, y \in Y,$$

for some geometric multiplier $\mu^*(y)$.

It is seen that Eq. (3) is equivalent to the bound

$$f^*(y) \le f(x, y) + c\big\|g^+(x, y)\big\|, \qquad \forall\, x \in X,\ \forall\, y \in Y, \tag{4}$$

so this bound holds if and only if there exists a uniform bounding constant $c > 0$ such that

$$u'\mu^*(y) \le c\|u^+\|, \qquad \forall\, u \in \Re^r,\ \forall\, y \in Y. \tag{5}$$

Thus the bound (4), holds if and only if for every $y \in Y$, it is possible to select a geometric multiplier $\mu^*(y)$ of the parametric problem (1) such that the set $\big\{\mu^*(y) \mid y \in Y\big\}$ is bounded.

Let us now specialize the preceding discussion to the parametric program

$$\begin{aligned} &\text{minimize} \quad f(x, y) = \|y - x\| \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r, \end{aligned} \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm, X is a convex subset of $\Re^n$, and g_j are convex over X. This is the projection problem of the exercise. Let us take $Y = X$. If c satisfies Eq. (5), the bound (4) becomes

$$d(y) \le \|y - x\| + c\Big\|\big(g(x)\big)^+\Big\|, \qquad \forall\, x \in X,\ \forall\, y \in X,$$

and (by taking $x = y$) implies the bound

$$d(y) \le c\Big\|\big(g(y)\big)^+\Big\|, \qquad \forall\, y \in X. \tag{7}$$

This bound holds if a geometric multiplier $\mu^*(y)$ of the projection problem (6) can be found such that Eq. (5) holds. We will now show the reverse assertion.

Indeed, assume that for some c, Eq. (7) holds, and to arrive at a contradiction, assume that there exist $x \in X$ and $y \in Y$ such that

$$d(y) > \|y - x\| + c\Big\|\big(g(x)\big)^+\Big\|.$$

Then, using Eq. (7), we obtain

$$d(y) > \|y - x\| + d(x).$$

From this relation and the triangle inequality, it follows that

$$\begin{aligned} \inf_{z \in X,\, g(z) \le 0} \|y - z\| &> \|y - x\| + \inf_{z \in X,\, g(z) \le 0} \|x - z\| \\ &= \inf_{z \in X,\, g(z) \le 0} \big\{\|y - x\| + \|x - z\|\big\} \\ &\ge \inf_{z \in X,\, g(z) \le 0} \|y - z\|, \end{aligned}$$

which is a contradiction. Thus Eq. (7) implies that we have

$$d(y) \le \|y - x\| + c\Big\|\big(g(x)\big)^+\Big\|, \qquad \forall\, x \in X,\ \forall\, y \in X.$$

Using Prop. 7.3.1, this implies that there exists a geometric multiplier $\mu^*(y)$ such that

$$u'\mu^*(y) \le c\|u^+\|, \qquad \forall\, u \in \Re^r,\ \forall\, y \in X.$$

This in turn implies the boundedness of the set $\big\{\mu^*(y) \mid y \in X\big\}$.

Convex Analysis and Optimization

Chapter 8 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts
http://www.athenasc.com

CHAPTER 8: SOLUTION MANUAL

8.1

To obtain a contradiction, assume that q is differentiable at some dual optimal solution $\mu^* \in M$, where $M = \{\mu \in \Re^r \mid \mu \geq 0\}$. Then by the optimality theory of Section 4.7 (cf. Prop. 4.7.2, concave function q), we have

$$\nabla q(\mu^*)(\mu^* - \mu) \geq 0, \qquad \forall\ \mu \geq 0.$$

If $\mu_j^* = 0$, then by letting $\mu = \mu^* + \gamma e_j$ for a scalar $\gamma \geq 0$, and the vector e_j whose jth component is 1 and the other components are 0, from the preceding relation we obtain $\partial q(\mu^*)/\partial \mu_j \leq 0$. Similarly, if $\mu_j^* > 0$, then by letting $\mu = \mu^* + \gamma e_j$ for a sufficiently small scalar γ (small enough so that $\mu^* + \gamma e_j \in M$), from the preceding relation we obtain $\partial q(\mu^*)/\partial \mu_j = 0$. Hence

$$\partial q(\mu^*)/\partial \mu_j \leq 0, \qquad \forall\ j = 1, \ldots, r,$$

$$\mu_j^* \partial q(\mu^*)/\partial \mu_j = 0, \qquad \forall\ j = 1, \ldots, r.$$

Since q is differentiable at μ^*, we have that

$$\nabla q(\mu^*) = g(x^*),$$

for some vector $x^* \in X$ such that $q(\mu^*) = L(x^*, \mu^*)$. This and the preceding two relations imply that x^* and μ^* satisfy the necessary and sufficient optimality conditions for an optimal solution-geometric multiplier pair (cf. Prop. 6.2.5). It follows that there is no duality gap, a contradiction.

8.2 (Sharpness of the Error Tolerance Estimate)

Consider the incremental subgradient method with the stepsize α and the starting point $\overline{x} = (\alpha M C_0, \alpha M C_0)$, and the following component processing order:

M components of the form $|x_1|$ [endpoint is $(0, \alpha M C_0)$],

M components of the form $|x_1 + 1|$ [endpoint is $(-\alpha M C_0, \alpha M C_0)$],

M components of the form $|x_2|$ [endpoint is $(-\alpha M C_0, 0)$],

M components of the form $|x_2 + 1|$ [endpoint is $(-\alpha M C_0, -\alpha M C_0)$],

M components of the form $|x_1|$ [endpoint is $(0, -\alpha M C_0)$],

M components of the form $|x_1 - 1|$ [endpoint is $(\alpha M C_0, -\alpha M C_0)$],

M components of the form $|x_2|$ [endpoint is $(\alpha MC_0, 0)$], and

M components of the form $|x_2 - 1|$ [endpoint is $(\alpha MC_0, \alpha MC_0)$].

With this processing order, the method returns to $\overline{x}$ at the end of a cycle. Furthermore, the smallest function value within the cycle is attained at points $(\pm\alpha MC_0, 0)$ and $(0, \pm\alpha MC_0)$, and is equal to $4MC_0 + 2\alpha M^2 C_0^2$. The optimal function value is $f^* = 4MC_0$, so that

$$\liminf_{k\to\infty} f(\psi_{i,k}) \geq f^* + 2\alpha M^2 C_0^2.$$

Since $m = 8M$ and $mC_0 = C$, we have $M^2 C_0^2 = C^2/64$, implying that

$$2\alpha M^2 C_0^2 = \frac{1}{16}\frac{\alpha C^2}{2},$$

and therefore

$$\liminf_{k\to\infty} f(\psi_{i,k}) \geq f^* + \frac{\beta\alpha C^2}{2},$$

with $\beta = 1/16$.

8.3 (A Variation of the Subgradient Method [CFM75])

At first, by induction, we show that

$$(\mu^* - \mu^k)' d^k \geq (\mu^* - \mu^k)' g^k. \tag{8.0}$$

Since $d^0 = g^0$, the preceding relation obviously holds for $k = 0$. Assume now that this relation holds for $k-1$. By using the definition of d^k,

$$d^k = g^k + \beta^k d^{k-1},$$

we obtain

$$(\mu^* - \mu^k)' d^k = (\mu^* - \mu^k)' g^k + \beta^k (\mu^* - \mu^k)' d^{k-1}. \tag{8.1}$$

We further have

$$\begin{aligned}(\mu^* - \mu^k)' d^{k-1} &= (\mu^* - \mu^{k-1})' d^{k-1} + (\mu^{k-1} - \mu^k)' d^{k-1} \\ &\geq (\mu^* - \mu^{k-1})' d^{k-1} - \|\mu^{k-1} - \mu^k\| \|d^{k-1}\|.\end{aligned}$$

By the induction hypothesis, we have that

$$(\mu^* - \mu^{k-1})' d^{k-1} \geq (\mu^* - \mu^{k-1})' g^{k-1},$$

while by the subgradient inequality, we have that

$$(\mu^* - \mu^{k-1})' g^{k-1} \geq q(\mu^*) - q(\mu^{k-1}).$$

Combining the preceding three relations, we obtain

$$(\mu^* - \mu^k)' d^{k-1} \geq q(\mu^*) - q(\mu^{k-1}) - \|\mu^{k-1} - \mu^k\| \|d^{k-1}\|.$$

Since

$$\|\mu^{k-1} - \mu^k\| \le s^{k-1}\|d^{k-1}\|,$$

it follows that

$$(\mu^* - \mu^k)'d^{k-1} \ge q(\mu^*) - q(\mu^{k-1}) - s^{k-1}\|d^{k-1}\|^2.$$

Finally, because $0 < s^{k-1} \le \big(q(\mu^*) - q(\mu^{k-1})\big)/\|d^{k-1}\|^2$, we see that

$$(\mu^* - \mu^k)'d^{k-1} \ge 0. \tag{8.2}$$

Since $\beta^k \ge 0$, the preceding relation and equation (8.1) imply that

$$(\mu^* - \mu^k)'d^k \ge (\mu^* - \mu^k)'g^k.$$

Assuming $\mu^k \ne \mu^*$, we next show that

$$\|\mu^* - \mu^{k+1}\| < \|\mu^* - \mu^k\|, \qquad \forall\ k.$$

Similar to the proof of Prop. 8.2.1, it can be seen that this relation holds for $k = 0$. For $k > 0$, by using the nonexpansive property of the projection operation, we obtain

$$\begin{aligned}\|\mu^* - \mu^{k+1}\|^2 &\le \|\mu^* - \mu^k - s^k d^k\|^2 \\ &= \|\mu^* - \mu^k\|^2 - 2s^k(\mu^* - \mu^k)'d^k + (s^k)^2\|d^k\|^2.\end{aligned}$$

By using equation (8.1) and the subgradient inequality,

$$(\mu^* - \mu^k)'g^k \ge q(\mu^*) - q(\mu^k),$$

we further obtain

$$\begin{aligned}\|\mu^* - \mu^{k+1}\|^2 &\le \|\mu^k - \mu^*\|^2 - 2s^k(\mu^* - \mu^k)'g^k + (s^k)^2\|d^k\|^2 \\ &\le \|\mu^k - \mu^*\|^2 - 2s^k\big(q(\mu^*) - q(\mu^k)\big) + (s^k)^2\|d^k\|^2.\end{aligned}$$

Since $0 < s^k \le \big(q(\mu^*) - q(\mu^k)\big)/\|d^k\|^2$, it follows that

$$-2s^k\big(q(\mu^*) - q(\mu^k)\big) + (s^k)^2\|d^k\|^2 \le -s^k\big(q(\mu^*) - q(\mu^k)\big) < 0,$$

implying that

$$\|\mu^* - \mu^{k+1}\|^2 < \|\mu^k - \mu^*\|^2.$$

We next prove that

$$\frac{(\mu^* - \mu^k)'d^k}{\|d^k\|} \ge \frac{(\mu^* - \mu^k)'g^k}{\|g^k\|}.$$

It suffices to show that

$$\|d^k\| \le \|g^k\|,$$

since this inequality and Eq. (8.0) imply the desired relation. If $g^{k'}d^{k-1} \geq 0$, then by the definition of d^k and β^k, we have that $d^k = g^k$, and we are done, so assume that $g^{k'}d^{k-1} < 0$. We then have

$$\|d^k\|^2 = \|g^k\|^2 + 2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2.$$

Since $\beta^k = -\gamma g^{k'}d^{k-1}/\|d^{k-1}\|^2$, it follows that

$$2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2 = 2\beta^k g^{k'}d^{k-1} - \gamma\beta^k g^{k'}d^{k-1} = (2-\gamma)\beta^k g^{k'}d^{k-1}.$$

Furthermore, since $g^{k'}d^{k-1} < 0$, $\beta^k \geq 0$, and $\gamma \in [0,2]$, we see that

$$2\beta^k g^{k'}d^{k-1} + (\beta^k)^2\|d^{k-1}\|^2 \leq 0,$$

implying that

$$\|d^k\|^2 \leq \|g^k\|^2.$$

8.4 (Subgradient Randomization for Stochastic Programming)

The stochastic programming problem of Example 8.2.2 can be written in the following form

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} \pi_i\big(f_0(x) + f_i(x)\big) \\ &\text{subject to} \quad x \in X, \end{aligned}$$

where

$$f_i(x) = \max_{B_i'\lambda_i \leq d_i} (b_i - Ax)'\lambda_i, \qquad i = 1, \ldots, m,$$

and the outcome i occurs with probability π_i. Assume that for each outcome $i \in \{1, ..., m\}$ and each vector $x \in \Re^n$, the maximum in the expression for $f_i(x)$ is attained at some $\lambda_i(x)$. Then, the vector $A'\lambda_i(x)$ is a subgradient of f_i at x. One possible form of the randomized incremental subgradient method is

$$x_{k+1} = P_X\big(x_k - \alpha_k(g_k + A'\lambda^k_{\omega_k})\big),$$

where g_k is a subgradient of f_0 at x_k, $\lambda^k_{\omega_k} = \lambda_{\omega_k}(x_k)$, and the random variable ω_k takes value i from the set $\{1, \ldots, m\}$ with probability π_i. The convergence analysis of Section 8.2.2 goes through in its entirety for this method, with only some adjustments in various bounding constants.

In an alternative method, we could use as components the $m+1$ functions $f_0, f_1, \ldots, f_m$, with f_0 chosen with probability 1/2 and each component f_i, $1, \ldots, m$, chosen with probability $\pi_i/2$.

8.5

Consider the cutting plane method applied to the following one-dimensional problem

$$\text{maximize } q(\mu) = -\mu^2,$$
$$\text{subject to } \mu \in [0, 1].$$

Suppose that the method is started at $\mu^0 = 0$, so that the initial polyhedral approximation is $Q^1(\mu) = 0$ for all μ. Suppose also that in all subsequent iterations, when maximizing $Q^k(\mu)$, $k = 0, 1, ...,$ over $[0, 1]$, we choose μ^k to be the largest of all maximizers of $Q^k(\mu)$ over $[0, 1]$. We will show by induction that in this case, we have $\mu^k = 1/2^{k-1}$ for $k = 1, 2, ...$

Since $Q^1(\mu) = 0$ for all μ, the set of maximizers of $Q^1(\mu) = 0$ over $[0, 1]$ is the entire interval $[0, 1]$, so that the largest maximizer is $\mu^1 = 1$. Suppose now that $\mu^i = 1/2^{i-1}$ for $i = 1, ..., k$. Then

$$Q^{k+1}(\mu) = \min\{l_0(\mu), l_1(\mu), ..., l_k(\mu)\},$$

where $l_0(\mu) = 0$ and

$$l_i(\mu) = q(\mu^i) + \nabla q(\mu^i)'(\mu - \mu^i) = -2\mu^i\mu + (\mu^i)^2, \qquad i = 1, ..., k.$$

The maximum value of $Q^{k+1}(\mu)$ over $[0, 1]$ is 0 and it is attained at any point in the interval $[0, \mu^k/2]$. By the induction hypothesis, we have $\mu^k = 1/2^{k-1}$, implying that the largest maximizer of $Q^{k+1}(\mu)$ over $[0, 1]$ is $\mu^{k+1} = 1/2^k$.

Hence, in this case, the cutting plane method generates an infinite sequence $\{\mu^k\}$ converging to the optimal solution $\mu^* = 0$, thus showing that the method need not terminate finitely even if it starts at an optimal solution.

8.6 (Approximate Subgradient Method)

(a) We have for all $\mu \in \Re^r$

$$\begin{aligned} q(\mu) &= \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\} \\ &\leq f(x^k) + \mu' g(x^k) \\ &= f(x^k) + \mu^{k\prime} g(x^k) + g(x^k)'(\mu - \mu^k) \\ &= q(\mu^k) + \epsilon + g(x^k)'(\mu - \mu^k), \end{aligned}$$

where the last inequality follows from the equation

$$L(x^k, \mu^k) \leq \inf_{x \in X} L(x, \mu^k) + \epsilon.$$

Thus $g(x^k)$ is an ϵ-subgradient of q at μ^k.

(b) For all $\mu \in M$, by using the nonexpansive property of the projection, we have

$$\begin{aligned} \|\mu^{k+1} - \mu\|^2 &\leq \|\mu^k + s^k g^k - \mu\|^2 \\ &\leq \|\mu^k - \mu\|^2 - 2s^k g^{k\prime}(\mu - \mu^k) + (s^k)^2 \|g^k\|^2, \end{aligned}$$

where

$$s^k = \frac{q(\mu^*) - q(\mu^k)}{\|g^k\|^2},$$

and $g^k \in \partial_\epsilon q(\mu^k)$. From this relation and the definition of an ϵ-subgradient we obtain

$$\|\mu^{k+1} - \mu\|^2 \leq \|\mu^k - \mu\|^2 - 2s^k\big(q(\mu) - q(\mu^k) - \epsilon\big) + (s^k)^2\|g^k\|^2, \qquad \forall\ \mu \in M.$$

Let μ^* be an optimal solution. Substituting the expression for s^k and taking $\mu = \mu^*$ in the above inequality, we have

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - \frac{q(\mu^*) - q(\mu^k)}{\|g^k\|^2}\big(q(\mu^*) - q(\mu^k) - 2\epsilon\big).$$

Thus, if $q(\mu^k) < q(\mu^*) - 2\epsilon$, we obtain

$$\|\mu^{k+1} - \mu^*\| \leq \|\mu^k - \mu^*\|.$$