Convex Optimization Theory

Chapter 1
Exercises and Solutions

Dimitri P. Bertsekas

Massachusetts Institute of Technology

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

CHAPTER 1: EXERCISES AND SOLUTIONS†

1.1

Let C be a nonempty subset of $\Re^n$, and let λ_1 and λ_2 be positive scalars. Show that if C is convex, then $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$ [cf. Prop. 1.1.1(c)]. Show by example that this need not be true when C is not convex.

Solution: We always have $(\lambda_1 + \lambda_2)C \subset \lambda_1 C + \lambda_2 C$, even if C is not convex. To show the reverse inclusion assuming C is convex, 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

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

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

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 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)

Show that:

(a) The intersection $\cap_{i \in I} C_i$ of a collection $\{C_i \mid i \in I\}$ of cones is a cone.

(b) The Cartesian product $C_1 \times C_2$ of two cones C_1 and C_2 is a cone.

(c) The vector sum $C_1 + C_2$ of two cones C_1 and C_2 is a cone.

(d) The image and the inverse image of a cone under a linear transformation is a cone.

† Many of the exercises and solutions given here were developed as part of my earlier convex optimization book [BNO03] (coauthored with Angelia Nedić and Asuman Ozdaglar), and are posted on the internet of that book's web site. The contribution of my coauthors in the development of these exercises and their solutions is gratefully acknowledged. Since some of the exercises and/or their solutions have been modified and also new exercises have been added, all errors are my sole responsibility.

(e) A subset C is a convex cone if and only if it is closed under addition and positive scalar multiplication, i.e., $C + C \subset C$, and $\gamma C \subset C$ for all $\gamma > 0$.

Solution: (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) 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 $A(\alpha x) \in C$, implying that $\alpha x \in A^{-1} \cdot C$, and showing that $A^{-1} \cdot C$ is a cone.

(e) Let C be a convex cone. Then $\gamma C \subset C$, for all $\gamma > 0$, by the definition of cone. Furthermore, by convexity of C, for all $x, y \in C$, we have $z \in C$, where

$$z = \frac{1}{2}(x + y).$$

Hence $(x + y) = 2z \in C$, since C is a cone, and it follows that $C + C \subset C$.

Conversely, assume that $C + C \subset C$, and $\gamma C \subset C$. Then C is a cone. Furthermore, if $x, y \in C$ and $\alpha \in (0, 1)$, we have $\alpha x \in C$ and $(1 - \alpha) y \in C$, and $\alpha x + (1 - \alpha) y \in C$ (since $C + C \subset C$). Hence C is convex.

1.3 (Convexity under Composition)

Let C be a nonempty convex subset of $\Re^n$. Let also $f = (f_1, \ldots, f_m)$, where $f_i : C \mapsto \Re$, $i = 1, \ldots, m$, are convex functions, and let $g : \Re^m \mapsto \Re$ be a function that is convex and monotonically nondecreasing over a convex set that contains the set $\big\{ f(x) \mid x \in C \big\}$, in the sense that for all $u, \overline{u}$ in this set such that $u \leq \overline{u}$, we have $g(u) \leq g(\overline{u})$. Show that the function h defined by $h(x) = g\big(f(x)\big)$ is convex over C. If in addition, $m = 1$, g is monotonically increasing and f is strictly convex, then h is strictly convex.

Solution: Let $x, y \in \Re^n$ and let $\alpha \in [0, 1]$. By the definitions of h and f, we have

$$h\big(\alpha x + (1 - \alpha) y\big) = g\Big(f\big(\alpha x + (1 - \alpha) y\big)\Big)$$

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

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

If $m = 1$, g is monotonically increasing, and f is strictly convex, then the first inequality is strict whenever $x \neq y$ and $\alpha \in (0, 1)$, showing that h is strictly convex.

1.4 (Examples of Convex Functions)

Show that the following functions from $\Re^n$ to $(-\infty, \infty]$ are convex:

(a)
$$
f_1(x_1, \ldots, x_n) = \begin{cases} -(x_1 x_2 \cdots x_n)^{\frac{1}{n}} & \text{if } x_1 > 0, \ldots, x_n > 0,\\ \infty & \text{otherwise.} \end{cases}
$$

(b) $f_2(x) = \ln\big(e^{x_1} + \cdots + e^{x_n}\big)$.

(c) $f_3(x) = \|x\|^p$ with $p \geq 1$.

(d) $f_4(x) = \frac{1}{f(x)}$, where f is concave and $0 < f(x) < \infty$ for all x.

(e) $f_5(x) = \alpha f(x) + \beta$, where $f : \Re^n \mapsto \Re$ is a convex function, and α and β are scalars, with $\alpha \geq 0$.

(f) $f_6(x) = e^{\beta x' A x}$, where A is a positive semidefinite symmetric $n \times n$ matrix and β is a positive scalar.

(g) $f_7(x) = f(Ax + b)$, where $f : \Re^m \mapsto \Re$ is a convex function, A is an $m \times n$ matrix, and b is a vector in $\Re^m$.

Solution: (a) Denote $X = \text{dom}(f_1)$. 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

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

in view of the fact that $2\alpha_j\alpha_k \le \alpha_j^2 + \alpha_k^2$. Hence, $\nabla^2 f_1(x)$ is positive semidefinite for all $x \in X$, and it follows from Prop. 1.1.10(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

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

Hence by Prop. 1.1.10(a), f_2 is convex.

(c) 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 $p \ge 1$ 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). Using Exercise 1.3, it follows that the function $f_3(x) = \|x\|^p$ is convex over $\Re^n$.

(d) 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 1.3, it follows that the function $f_4(x) = \frac{1}{f(x)}$ is convex over $\Re^n$.

(e) 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 $\alpha \ge 0$), while f is convex over $\Re^n$. Using Exercise 1.3, it follows that f_5 is convex over $\Re^n$.

(f) 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 1.3, it follows that f_6 is convex over $\Re^n$.

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

1.5 (Ascent/Descent Behavior of a Convex Function)

Let $f : \Re \mapsto \Re$ be a convex function.

(a) *(Monotropic Property)* Use the definition of convexity to show that f is "turning upwards" in the sense that if x_1, x_2, x_3 are three scalars such that $x_1 < x_2 < x_3$, then

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

(b) Use part (a) to show that there are four possibilities as x increases to ∞: (1) $f(x)$ decreases monotonically to $-\infty$, (2) $f(x)$ decreases monotonically to a finite value, (3) $f(x)$ reaches some value and stays at that value, (4) $f(x)$ increases monotonically to ∞ when $x \geq \overline{x}$ for some $\overline{x} \in \Re$.

Solution: (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

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

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

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

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

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

(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

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2} \leq \cdots \leq \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}. \tag{1.1}$$

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

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \to \gamma, \tag{1.2}$$

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

$$\frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} \leq \gamma, \qquad \forall\ k. \tag{1.3}$$

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. (1.2) is independent of the choice for $\{x_k\}$, and Eqs. (1.1) and (1.3) 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. (1.3), we obtain

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

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 $k \ge 3$. By using Eq. (1.3) with $k = 1$, we have

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

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. (1.3), we have $f(x_{k+1}) - f(x_k) \le 0$ 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 $k \ge 3$, where K is large enough so that $y < x_K$. By using Eqs. (1.1) and (1.3) with $\{y_k\}$, we have

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

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. (1.1) and (1.3) 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. (1.1) and (1.3) 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. (1.2) and (1.3)], there is a positive integer K and a positive scalar ϵ with $\epsilon < \gamma$ such that

$$\epsilon \leq \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}, \qquad \forall\ k \geq K. \tag{1.4}$$

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. (1.4) 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 $\overline{x} = x_K$.

1.6 (Posynomials)

A *posynomial* is a function of positive scalar variables $y_1, \ldots, y_n$ of the form

$$g(y_1, \ldots, y_n) = \sum_{i=1}^{m} \beta_i y_1^{a_{i1}} \cdots y_n^{a_{in}},$$

where a_{ij} and β_i are scalars, such that $\beta_i > 0$ for all i. Show the following:

(a) A posynomial need not be convex.

(b) By a logarithmic change of variables, where we set

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

we obtain a convex function

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

where $\exp(z) = e^{z_1} + \cdots + e^{z_m}$ for all $z \in \Re^m$, A is an $m \times n$ matrix with components a_{ij}, and $b \in \Re^m$ is a vector with components b_i.

(c) Every 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, can be transformed by a logarithmic change of variables into a convex function f given by

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

with the matrix A_k and the vector b_k being associated with the posynomial g_k for each k.

Solution: (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

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

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 $\ln \exp(z) = \ln\bigl(e^{z_1} + \cdots + e^{z_m}\bigr)$ 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 1.4(b). With this identification, $f(x)$ can be viewed as the composition $f(x) = f_2(Ax + b)$, which is convex by Exercise 1.4(g).

(c) Consider a 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 the weighted sum of convex functions with nonnegative coefficients [part (b)], it follows that $f(x)$ is convex.

1.7 (Arithmetic-Geometric Mean Inequality)

Show that if $\alpha_1, \ldots, \alpha_n$ are positive scalars with $\sum_{i=1}^n \alpha_i = 1$, then for every set of positive scalars $x_1, \ldots, x_n$, we have

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + a_2 x_2 + \cdots + \alpha_n x_n,$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$. *Hint*: Show that $(-\ln x)$ is a strictly convex function on $(0, \infty)$.

Solution: 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 $x_1, \ldots, x_n$ 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 the scalars $x_1, \ldots, x_n$ are all equal.

1.8 (Young and Holder Inequalities)

Use the result of Exercise 1.7 to verify Young's inequality

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

where $p > 0$, $q > 0$, and

$$1/p + 1/q = 1.$$

Then, use Young's inequality to verify Holder's inequality

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} \left(\sum_{i=1}^n |y_i|^q \right)^{1/q}.$$

Solution: According to Exercise 1.7, 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.

1.9 (Characterization of Differentiable Convex Functions)

Let $f : \Re^n \mapsto \Re$ be a differentiable function. Show that f is convex over a nonempty convex set C if and only if

$$\big(\nabla f(x) - \nabla f(y)\big)'(x - y) \geq 0, \qquad \forall\ x, y \in C.$$

Note: The condition above says that the function f, restricted to the line segment connecting x and y, has monotonically nondecreasing gradient.

Solution: If f is convex, then by Prop. 1.1.7(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

$$\big(\nabla f(y) - \nabla f(x)\big)'(x - y) \geq 0. \tag{1.5}$$

Conversely, let Eq. (1.5) 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. (1.5), 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.

1.10 (Strong Convexity)

Let $f : \Re^n \mapsto \Re$ be a function that is continuous over a closed convex set $C \subset \text{dom}(f)$, and let $\sigma > 0$. We say that f *is strongly convex over* C *with coefficient* σ if for all $x, y \in C$ and all $\alpha \in [0,1]$, we have

$$f\big(\alpha x + (1-\alpha)y\big) + \frac{\sigma}{2}\alpha(1-\alpha)\|x-y\|^2 \leq \alpha f(x) + (1-\alpha)f(y).$$

(a) Show that if f is strongly convex over C with coefficient σ, then f is strictly convex over C. Furthermore, there exists a unique $x^* \in C$ that minimizes f over C, and we have

$$f(x) \geq f(x^*) + \frac{\sigma}{2}\|x-x^*\|^2, \qquad \forall\, x \in C.$$

(b) Assume that $\text{int}(C)$, the interior of C, is nonempty, and that f is continuously differentiable over $\text{int}(C)$. Show that the following are equivalent:

(i) f is strongly convex with coefficient σ over C.

(ii) We have

$$\big(\nabla f(x) - \nabla f(y)\big)'(x-y) \geq \sigma\|x-y\|^2, \qquad \forall\, x, y \in \text{int}(C).$$

Furthermore, if f is twice continuously differentiable over $\mathrm{int}(C)$, the above two properties are equivalent to:

(iii) The matrix $\nabla^2 f(x) - \sigma I$ is positive semidefinite for every $x \in \mathrm{int}(C)$, where I is the identity matrix.

Solution: (a) The strict convexity of f over C is evident from the definition of strong convexity and the hypothesis. Strict convexity also implies that there can be at most one minimum of f over C.

To show existence of a vector x^* that minimizes f over C, we show that every level set $\{x \in C \mid f(x) \leq \gamma\}$ is bounded and hence compact (since C is closed and f is continuous over C), and then use Weierstrass' Theorem. Assume to arrive at a contradiction that a level set $L = \{x \in C \mid f(x) \leq \gamma\}$ is unbounded, and let $\{x_k\} \subset L$ be an unbounded sequence. We assume with no loss of generality that $\|x_k - x_0\| \geq 1$ for all k. Let $\alpha_k = 1/\|x_k - x_0\|$, and note that $\alpha_k \to 0$. Define

$$y_k = \alpha_k x_k + (1 - \alpha_k)x_0 = \frac{x_k - x_0}{\|x_k - x_0\|} + x_0,$$

and note that $\|y_k - x_0\| = 1$ for all $k \geq 1$. By strong convexity of f, we have

$$\begin{aligned} f(y_k) &\leq \alpha_k f(x_k) + (1 - \alpha_k) f(x_0) - \frac{1}{2}\sigma\alpha_k(1 - \alpha_k)\|x_k - x_0\|^2 \\ &= \alpha_k f(x_k) + (1 - \alpha_k) f(x_0) - \frac{1}{2}\sigma(1 - \alpha_k)\|x_k - x_0\| \\ &\leq \gamma - \frac{1}{2}\sigma(1 - \alpha_k)\|x_k - x_0\|. \end{aligned}$$

Hence $f(y_k) \to -\infty$, which contradicts the boundedness of $\{y_k\}$ and the continuity of f.

To show the inequality $f(x) \geq f(x^*) + (\sigma/2)\|x - x^*\|^2$, we write for any $x \in C$ and $\alpha \in (0, 1)$,

$$\begin{aligned} \alpha f(x) + (1 - \alpha) f(x^*) &\geq f\big(\alpha x + (1 - \alpha)x^*\big) + \frac{1}{2}\sigma\alpha(1 - \alpha)\|x - x^*\|^2 \\ &\geq f(x^*) + \frac{1}{2}\sigma\alpha(1 - \alpha)\|x - x^*\|^2. \end{aligned}$$

It follows that $f(x) \geq f(x^*) + (\sigma/2)(1 - \alpha)\|x - x^*\|^2$, and by taking the limit as $\alpha \to 0$, we obtain the desired inequality.

(b) We first show that (i) implies (ii). We have, using the definition of strong convexity,

$$f(y) + \alpha\nabla f(y)'(x - y) \leq f\big(y + \alpha(x - y)\big) \leq \alpha f(x) + (1 - \alpha) f(y) - \frac{\sigma}{2}\alpha(1 - \alpha)\|x - y\|^2,$$

for all $x, y \in \mathrm{int}(C)$ and $\alpha \in (0, 1)$, from which

$$f(y) + \nabla f(y)'(x - y) \leq f(x) - \frac{\sigma}{2}(1 - \alpha)\|x - y\|^2.$$

Similarly,

$$f(x) + \nabla f(x)'(y - x) \leq f(y) - \frac{\sigma}{2}(1 - \alpha)\|x - y\|^2,$$

and adding these two inequalities:

$$\big(\nabla f(y) - \nabla f(x)\big)'(x-y) \le -\sigma(1-\alpha)\|x-y\|^2,$$

or

$$\big(\nabla f(y) - \nabla f(x)\big)'(y-x) \ge \sigma(1-\alpha)\|x-y\|^2.$$

Taking the limit as $\alpha \to 0$, we obtain

$$\big(\nabla f(y) - \nabla f(x)\big)'(y-x) \ge \sigma\|x-y\|^2.$$

Next we show that (ii) implies (i). For any $\alpha \in (0,1)$ and $x_1, x_2 \in \mathrm{int}(C)$ with $x_1 \neq x_2$, let

$$x_\alpha = \alpha x_1 + (1-\alpha)x_2.$$

We have

$$f(x_a) = f(x_1) + \int_0^1 \nabla f\big(x_1 + t(x_\alpha - x_1)\big)'(x_a - x_1)dt,$$

$$f(x_a) = f(x_2) + \int_0^1 \nabla f\big(x_2 + t(x_a - x_2)\big)'(x_a - x_2)dt.$$

Multiplying these relations with α and $1-\alpha$, respectively, adding, and collecting terms using the relations $x_\alpha - x_1 = (1-\alpha)(x_2 - x_1)$, $x_\alpha - x_2 = \alpha(x_1 - x_2)$, and

$$\big(x_1 + t(x_a - x_1)\big) - \big(x_2 + t(x_a - x_2)\big) = (1-t)(x_1 - x_2),$$

we obtain

$$\begin{aligned}
&\alpha f(x_1) + (1-\alpha)f(x_2) - f(x_a)\\
&\quad = \alpha(1-\alpha)\int_0^1 \Big(\nabla f\big(x_1 + t(x_a - x_1)\big) - \nabla f\big(x_2 + t(x_a - x_2)\big)\Big)'(x_1 - x_2)dt\\
&\quad \ge \sigma\alpha(1-\alpha)\|x_1 - x_2\|^2 \int_0^1 (1-t)dt\\
&\quad = \frac{1}{2}\sigma\alpha(1-\alpha)\|x_1 - x_2\|^2,
\end{aligned}$$

verifying the strong convexity inequality for x_1, x_2 in the interior of C [and using the continuity of f, for x_1, x_2 in the boundary of C as well].

Assume now that f is twice continuously differentiable over $\mathrm{int}(C)$.

First we show that (iii) implies (ii). Let $x, y \in \mathrm{int}(C)$ and 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, 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,

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

where the last inequality holds because $\nabla^2 f\big(tx+(1-t)y\big) - \sigma I$ is positive semidefinite. Combining the last two relations, we obtain the desired inequality.

We finally show that (i) implies (iii). For any $\alpha \in (0, 1)$ and $x_1, x_2 \in \text{int}(C)$ with $x_1 \neq x_2$, let

$$x_\alpha = \alpha x_1 + (1 - \alpha)x_2.$$

Using the 2nd order Mean Value Theorem, we have

$$f(x_1) = f(x_\alpha) + \nabla f(x_\alpha)'(x_1 - x_\alpha) + \frac{1}{2}(x_1 - x_\alpha)'\nabla^2 f(\tilde{x}_\alpha)(x_1 - x_\alpha),$$

$$f(x_2) = f(x_\alpha) + \nabla f(x_\alpha)'(x_2 - x_\alpha) + \frac{1}{2}(x_2 - x_\alpha)'\nabla^2 f(\hat{x}_\alpha)(x_2 - x_\alpha),$$

where $\tilde{x}_\alpha$ and $\hat{x}_\alpha$ are vectors that lie in the intervals connecting x_α with x_1 and x_2, respectively. Multiplying these relations with α and $1 - \alpha$, respectively, adding, canceling the terms involving $\nabla f(x_\alpha)$, and using the relations $x_\alpha - x_1 = (1 - \alpha)(x_2 - x_1)$ and $x_\alpha - x_2 = \alpha(x_1 - x_2)$ and the definition of strong convexity, we obtain

$$\begin{aligned} f(x_\alpha) + \frac{1}{2}\sigma\alpha(1 - \alpha)\|x_1 - x_2\|^2 &\leq \alpha f(x_1) + (1 - \alpha)f(x_2) \\ &= f(x_\alpha) + \frac{1}{2}\alpha(1 - \alpha)^2(x_1 - x_2)'\nabla^2 f(\tilde{x}_\alpha)(x_1 - x_2) \\ &\quad + \frac{1}{2}\alpha^2(1 - \alpha)(x_1 - x_2)'\nabla^2 f(\hat{x}_\alpha)(x_1 - x_2) \end{aligned}$$

and finally,

$$\sigma\|x_1 - x_2\|^2 \leq (x_1 - x_2)'\big((1 - \alpha)\nabla^2 f(\tilde{x}_\alpha) + \alpha\nabla^2 f(\hat{x}_\alpha)\big)(x_1 - x_2).$$

Dividing by $\|x_1 - x_2\|^2$ and letting x_2 approach x_1, we obtain

$$\sigma \leq d'\nabla^2 f(x_1)d,$$

where $d = (x_1 - x_2)/\|x_1 - x_2\|$. Since x_1 and x_2 were chosen arbitrarily within $\text{int}(C)$, it follows that the matrix $\nabla^2 f(x) - \sigma I$ is positive semidefinite for every $x \in \text{int}(C)$. Since by the convexity of C, every point in the boundary of C can be approached from the interior and $\nabla^2 f$ is continuous, $\nabla^2 f(x) - \sigma I$ is also positive semidefinite for every x in the boundary of C.

1.11 (Generated Cones and Convex Hulls)

Show that:

(a) For a nonempty convex subset C of $\Re^n$, we have

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

(b) A cone C is convex if and only if $C + C \subset C$.

(c) For any two convex cones C_1 and C_2 containing the origin, we have

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

Solution: (a) Let $y \in \text{cone}(C)$. If $y = 0$, then $y \in \cup_{x\in C}\{\gamma x \mid \gamma \geq 0\}$. If $y \neq 0$, then by definition of 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

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

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

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

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

(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 $C_1 + C_2 \subset \text{conv}(C_1 \cup C_2)$. 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 $x \in \text{conv}(C_1 \cup C_2)$.

Next, we show that $\text{conv}(C_1 \cup C_2) \subset C_1 + C_2$. 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

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

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.

1.12 (Extension of Caratheodory's Theorem)

Let X_1 and X_2 be nonempty subsets of $\Re^n$, and let $X = \text{conv}(X_1) + \text{cone}(X_2)$. Show that every vector x in X can be represented in the form

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

where m is a positive integer with $m \leq n+1$, 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 $\alpha_1, \ldots, \alpha_m$ are nonnegative with $\alpha_1+\cdots+\alpha_k=1$. Furthermore, the vectors $x_2-x_1,\ldots,x_k-x_1,y_{k+1},\ldots,y_m$ are linearly independent.

Solution: The proof will be an application of Caratheodory's Theorem [Prop. 1.2.1(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, $(x,1)\in\text{cone}(Y)$. 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 $m\le n+1$) 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

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

1.13

Let X be a nonempty bounded subset of $\Re^n$. Show that

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

In particular, if X is compact, then $\mathrm{conv}(X)$ is compact (cf. Prop. 1.2.2).

Solution: The set $\mathrm{cl}(X)$ is compact since X is bounded by assumption. Hence, by Prop. 1.2.2, 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 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. 1.1.1(d), the closure of a convex set is convex. Hence, the result follows.

1.14 (Convex Hulls and Generated Cones of Cartesian Products)

Given nonempty sets $X_i \subset \Re^{n_i}$, $i = 1, \ldots, m$, let $X = X_1 \times \cdots \times X_m$ be their Cartesian product. Show that:

(a) The convex hull (closure, affine hull) of X is equal to the Cartesian product of the convex hulls (closures, affine hulls, respectively) of the X_i.

(b) If all the sets $X_1, \ldots, X_m$ contain the origin, then

$$\mathrm{cone}(X) = \mathrm{cone}(X_1) \times \cdots \times \mathrm{cone}(X_m).$$

Furthermore, the result fails if one of the sets does not contain the origin.

Solution: (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 $y \in \text{conv}(X_1) \times \cdots \times \text{conv}(X_m)$.

To prove the reverse inclusion, assume that y is a vector in $\text{conv}(X_1)\times\cdots\times \text{conv}(X_m)$. Then, we can represent y as $y = (y_1, \ldots, y_m)$ with $y_i \in \text{conv}(X_i)$, 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 $\text{conv}(X)$, 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 $\text{conv}(X)$, thus proving our claim.

Next, we show the corresponding result for the closure of X. Assume that $y = (x_1, \ldots, x_m) \in \text{cl}(X)$. 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 $x_i \in \text{cl}(X_i)$ for each i, and hence $y \in \text{cl}(X_1) \times \cdots \times \text{cl}(X_m)$. Conversely, suppose that $y = (x_1, \ldots, x_m) \in$

$\text{cl}(X_1) \times \cdots \times \text{cl}(X_m)$. 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 $y \in \text{cl}(X)$.

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 $\text{aff}(X_1), \ldots, \text{aff}(X_m)$ as well as $\text{aff}(X)$ are all subspaces.

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

1.15 (Characterization of Twice Continuously Differentiable Convex Functions)

Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over $\Re^n$. Let S be the subspace that is parallel to the affine hull of C. Show that f is convex over C if and only if $y' \nabla^2 f(x) y \geq 0$ for all $x \in C$ and $y \in S$. [In particular, when C has nonempty interior, f is convex over C if and only if $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.]

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

$$y' \nabla^2 f(x) y < 0, \qquad \forall\, x \in B(\overline{x}, \epsilon). \tag{1.6}$$

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. (1.6), 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 $x \in \text{ri}(C)$.

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 $\{x_k\} \subset \text{ri}(C)$ 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 $y \in S$. 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. 1.1.7(a), we conclude that f is convex over C.

1.16

Construct an example of a point in a nonconvex set X that has the prolongation property of Prop. 1.3.3 but is not a relative interior point of X.

Solution: Take two intersecting lines in the plane, and consider the point of intersection.

1.17 (Characterizations of Relative Interior)

Let C be a nonempty convex set.

(a) Show the following refinement of the Prolongation Lemma (Prop. 1.3.3): $x \in \text{ri}(C)$ if and only if for every $\overline{x} \in \text{aff}(C)$, there exists a $\gamma > 0$ such that $x + \gamma(x - \overline{x}) \in C$.

(b) Show that $\text{cone}(C) = \text{aff}(C)$ if and only if $0 \in \text{ri}(C)$.

Solution: (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. 1.3.1(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. 1.3.1(c).

(b) Assume that $0 \in \text{ri}(C)$. Then, the inclusion $\text{cone}(C) \subset \text{aff}(C)$ is evident. For the reverse inclusion, note that if $\overline{x} \in \text{aff}(C)$, then $-\overline{x} \in \text{aff}(C)$, so applying part (a) with $x = 0$, we have that $\gamma\overline{x} \in C$ for some $\gamma > 0$. Hence $\overline{x} \in \text{cone}(C)$ and $\text{aff}(C) \subset \text{cone}(C)$.

Conversely, assume that $\text{aff}(C) = \text{cone}(C)$. We will show that $0 \in \text{ri}(C)$. Indeed if this is not so, by applying part (a) with $x = 0$, it follows that there exists $\overline{x} \in \text{aff}(C)$ such that $\gamma(-\overline{x}) \notin C$ for all $\gamma > 0$. Hence $-\overline{x} \notin \text{cone}(C)$, a contradiction.

1.18

Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function, let γ be a scalar, and let C be a nonempty convex subset of $\Re^n$.

(a) Show that if $f(x) < \gamma$ for some x, then $f(x) < \gamma$ for some $x \in \text{ri}\big(\text{dom}(f)\big)$.

(b) Show that if $C \subset \text{ri}\big(\text{dom}(f)\big)$ and $f(x) < \gamma$ for some $x \in \text{cl}(C)$, then $f(x) < \gamma$ for some $x \in \text{ri}(C)$.

(c) Show that if $C \subset \text{dom}(f)$ and $f(x) \geq \gamma$ for all $x \in C$, then $f(x) \geq \gamma$ for all $x \in \text{cl}(C)$.

Solution: (a) Assume the contrary, i.e., that $f(x) \geq \gamma$ for all $x \in \text{ri}\big(\text{dom}(f)\big)$. Let $\overline{x}$ be such that $f(\overline{x}) < \gamma$ and let $\tilde{x}$ be any vector in $\text{ri}\big(\text{dom}(f)\big)$. By the Line Segment Principle, all the points on the line segment connecting $\overline{x}$ and $\tilde{x}$, except possibly $\overline{x}$, belong to $\text{ri}\big(\text{dom}(f)\big)$ and therefore,

$$f\big(\alpha\tilde{x} + (1 - \alpha)\overline{x}\big) \geq \gamma, \qquad \forall\ \alpha \in (0, 1].$$

Thus, we have

$$\alpha f(\tilde{x}) + (1-\alpha) f(\overline{x}) \geq f\big(\alpha\tilde{x} + (1-\alpha)\overline{x}\big) \geq \gamma, \qquad \forall\ \alpha \in (0,1].$$

By letting $\alpha \to 0$, it follows that $f(\overline{x}) \geq \gamma$, a contradiction.

(b) Define

$$g(x) = \begin{cases} f(x) & \text{if } x \in \mathrm{cl}(C), \\ \infty & \text{otherwise.} \end{cases}$$

Then

$$\mathrm{ri}(C) \subset \mathrm{dom}(g) \subset \mathrm{cl}(C),$$

so that $\mathrm{ri}\big(\mathrm{dom}(g)\big) = \mathrm{ri}(C)$, by Prop. 1.3.5. By hypothesis, there is an $\overline{x}$ with $g(\overline{x}) < \gamma$, so by part (a), there exists an $\tilde{x} \in \mathrm{ri}\big(\mathrm{dom}(g)\big)$ with $g(\tilde{x}) < \gamma$. This vector belongs to $\mathrm{ri}(C)$ and satisfies $f(\tilde{x}) < \alpha$.

(c) Assume the contrary, i.e., that $f(x) < \gamma$ for some $x \in \mathrm{cl}(C)$. Then, by part (b), we have $f(x) < \gamma$ for some $x \in \mathrm{ri}(C)$, which contradicts the hypothesis.

1.19 (Closure and Relative Interior of Cones)

(a) Let C be a nonempty convex cone. Show that $\mathrm{cl}(C)$ and $\mathrm{ri}(C)$ is also a convex cone.

(b) Let $C = \mathrm{cone}\big(\{x_1, \ldots, x_m\}\big)$. Show that

$$\mathrm{ri}(C) = \left\{ \sum_{i=1}^{m} \alpha_i x_i \;\middle|\; \alpha_i > 0,\ i = 1, \ldots, m \right\}.$$

Solution: (a) 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, and it also convex since the closure of a convex set if convex.

By Prop. 1.3.2(a), the relative interior of a convex set is convex. To show that $\mathrm{ri}(C)$ is a cone, let $x \in \mathrm{ri}(C)$. Then, $x \in C$ and since C is a cone, $\alpha x \in C$ for all $\alpha > 0$. By the Line Segment Principle, all the points on the line segment connecting x and αx, except possibly αx, belong to $\mathrm{ri}(C)$. Since this is true for every $\alpha > 0$, it follows that $\alpha x \in \mathrm{ri}(C)$ for all $\alpha > 0$, showing that $\mathrm{ri}(C)$ 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 A. Therefore, by using Prop. 1.3.6, we have

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

1.20 (Closure and Relative Interior of Level Sets)

Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function, and let $\gamma > \inf_{x \in \Re^n} f(x)$.

(a) Show that

$$\mathrm{ri}\big(\{x \mid f(x) \le \gamma\}\big) = \mathrm{ri}\big(\{x \mid f(x) < \gamma\}\big) = \big\{x \in \mathrm{ri}\big(\mathrm{dom}(f)\big) \mid f(x) < \gamma\big\},$$

$$\mathrm{cl}\big(\{x \mid f(x) \le \gamma\}\big) = \mathrm{cl}\big(\{x \mid f(x) < \gamma\}\big) = \big\{x \mid (\mathrm{cl}\, f)(x) \le \gamma\big\}.$$

(b) The sets $\{x \mid f(x) \le \gamma\}$ and $\{x \mid f(x) < \gamma\}$ have the same dimension as $\mathrm{dom}(f)$.

(c) If f is real-valued, $\{x \mid f(x) \le \gamma\}$ has nonempty interior. Furthermore, for all $\gamma, \overline{\gamma}$ with $\inf_{x \in \Re^n} f(x) < \gamma < \overline{\gamma}$, the interior of $\{x \mid f(x) \le \gamma\}$ is contained in the interior of $\{x \mid f(x) \le \overline{\gamma}\}$.

Solution: We have for every $\gamma \in \Re$

$$\big\{(x, \gamma) \mid f(x) \le \gamma\big\} = \mathrm{epi}(f) \cap M, \tag{1.7}$$

where M is the set

$$M = \big\{(x, \gamma) \mid x \in \Re^n\big\}.$$

By Prop. 1.3.10,

$$\mathrm{ri}\big(\mathrm{epi}(f)\big) = \big\{(x, w) \mid x \in \mathrm{ri}\big(\mathrm{dom}(f)\big),\ f(x) < w\big\}, \tag{1.8}$$

so for $\gamma > \inf_{x \in \Re^n} f(x)$, using also Exercise 1.18(a), we have $\mathrm{ri}\big(\mathrm{epi}(f)\big) \cap M \neq \emptyset$. It follows from Eq. (1.7) and Prop. 1.3.8 that

$$\mathrm{ri}\Big(\big\{(x, \gamma) \mid f(x) \le \gamma\big\}\Big) = \mathrm{ri}\big(\mathrm{epi}(f)\big) \cap M,$$

$$\mathrm{cl}\Big(\big\{(x, \gamma) \mid f(x) \le \gamma\big\}\Big) = \mathrm{cl}\big(\mathrm{epi}(f)\big) \cap M.$$

The last two equations together with Eq. (1.8) show that for every $\gamma > \inf_{x \in \Re^n} f(x)$, we have

$$\mathrm{ri}\Big(\big\{x \mid f(x) \le \gamma\big\}\Big) = \big\{x \in \mathrm{ri}(\mathrm{dom}(f)) \mid f(x) < \gamma\big\},$$

$$\mathrm{cl}\Big(\big\{x \mid f(x) \le \gamma\big\}\Big) = \big\{x \mid (\mathrm{cl}\, f)(x) \le \gamma\big\}.$$

Next, we show that

$$\mathrm{cl}\big(\{x \mid f(x) \le \gamma\}\big) = \mathrm{cl}\big(\{x \mid f(x) < \gamma\}\big). \tag{1.9}$$

Clearly,

$$\mathrm{cl}\big(\{x \mid f(x) \le \gamma\}\big) \supset \mathrm{cl}\big(\{x \mid f(x) < \gamma\}\big).$$

To show the reverse inclusion, let $\overline{x} \in \mathrm{cl}\big(\{x \mid f(x) \leq \gamma\}\big)$, or equivalently, $(\mathrm{cl}\, f)(\overline{x}) \leq \gamma$. Also, choose $\tilde{x}$ such that $\tilde{x} \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$ and $f(\tilde{x}) < \gamma$ [such a vector exists by Exercise 1.18(a), in view of the assumption $\gamma > \inf_{x\in\Re^n} f(x)$]. Then, by Prop. 1.3.15, along the line segment connecting $\tilde{x}$ and $\overline{x}$, there is a sequence $\{x_k\} \subset \mathrm{ri}\big(\mathrm{dom}(f)\big)$ that converges to $\overline{x}$ and satisfies $f(x_k) < \gamma$ for all k. It follows that $\overline{x} \in \mathrm{cl}\big(\{x \mid f(x) < \gamma\}\big)$, showing that

$$\mathrm{cl}\big(\{x \mid f(x) \leq \gamma\}\big) \subset \mathrm{cl}\big(\{x \mid f(x) < \gamma\}\big),$$

and thereby proving Eq. (1.9).

Next note that since the sets $\{x \mid f(x) \leq \gamma\}$ and $\{x \mid f(x) < \gamma\}$ have the same closure, by Prop. 1.3.5(c), they have the same relative interior, i.e.,

$$\mathrm{ri}\big(\{x \mid f(x) \leq \gamma\}\big) = \mathrm{ri}\big(\{x \mid f(x) < \gamma\}\big).$$

Finally, since the sets $\{x \mid f(x) \leq \gamma\}$ and $\{x \mid f(x) < \gamma\}$ have the same closure and relative interior, they also have the same affine hull, and hence the same dimension.

1.21 (Relative Interior Intersection Lemma)

Let C_1 and C_2 be convex sets. Show that

$$C_1 \cap \mathrm{ri}(C_2) \neq \emptyset \qquad \text{if and only if} \qquad \mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big) \cap \mathrm{ri}(C_2) \neq \emptyset.$$

Hint: Choose $\overline{x} \in \mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big)$ and $x \in C_1 \cap \mathrm{ri}(C_2)$ [which belongs to $C_1 \cap \mathrm{aff}(C_2)$], consider the line segment connecting x and $\overline{x}$, and use the Line Segment Principle to conclude that points close to x belong to $\mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big) \cap \mathrm{ri}(C_2)$.

Solution: Let $x \in C_1 \cap \mathrm{ri}(C_2)$ and $\overline{x} \in \mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big)$. Let L be the line segment connecting x and $\overline{x}$. Then L belongs to $C_1 \cap \mathrm{aff}(C_2)$ since both of its endpoints belong to $C_1 \cap \mathrm{aff}(C_2)$. Hence, by the Line Segment Principle, all points of L except possibly x, belong to $\mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big)$. On the other hand, by the definition of relative interior, all points of L that are sufficiently close to x belong to $\mathrm{ri}(C_2)$, and these points, except possibly for x belong to $\mathrm{ri}\big(C_1 \cap \mathrm{aff}(C_2)\big) \cap \mathrm{ri}(C_2)$.

1.22 (Retractiveness of Convex Cones)

(a) Show that a nonpolyhedral closed convex cone need not be retractive, by using as an example the cone $C = \{(u, v, w) \mid \|(u, v)\| \leq w\}$, the recession direction $d = (1, 0, 1)$, and the corresponding asymptotic sequence $\big\{(k, \sqrt{k}, \sqrt{k^2 + k})\big\}$.

(b) Verify that the cone C of part (a) can be written as the intersection of an infinite number of closed halfspaces, thereby showing that a nested set sequence obtained by intersection of an infinite number of retractive nested set sequences need not be retractive.

Solution: (a) Clearly, $d = (1, 0, 1)$ is the recession direction associated with the asymptotic sequence $\{x_k\}$, where $x_k = (k, \sqrt{k}, \sqrt{k^2 + k})$. On the other hand, it can be verified by straightforward calculation that the vector

$$x_k - d = (k - 1, \sqrt{k}, \sqrt{k^2 + k} - 1)$$

does not belong to C. Indeed, denoting

$$u_k = k - 1, \qquad v_k = \sqrt{k}, \qquad w_k = \sqrt{k^2 + k} - 1,$$

we have

$$\|(u_k, v_k)\|^2 = (k - 1)^2 + k = k^2 - k + 1,$$

while

$$w_k^2 = \left(\sqrt{k^2 + k} - 1\right)^2 = k^2 + k + 1 - 2\sqrt{k^2 + k},$$

and it can be seen that

$$\|(u_k, v_k)\|^2 > w_k^2, \qquad \forall\ k \geq 1.$$

(b) Since by the Schwarz inequality, we have

$$\max_{\|(w,y)\|=1} (uw + vy) = \|(u, v)\|,$$

it follows that the cone

$$C = \{(u, v, w) \mid \|(u, v)\| \leq w\}$$

can be written as

$$C = \cap_{\|(w,y)\|=1} \big\{(u, v, w) \mid uw + vy \leq w\big\}.$$

Hence C is the intersection of an infinite number of closed halfspaces.

1.23 (Frank-Wolfe Theorem – Existence of Solutions of General Quadratic Programs)

Let Q be a symmetric $n \times n$ matrix, let c and $a_1, \ldots, a_r$ be vectors in $\Re^n$, and let $b_1, \ldots, b_r$ be scalars. Assume that the optimal value of the problem

$$\begin{aligned} &\text{minimize} \ \ x'Qx + c'x \\ &\text{subject to} \ \ a_j'x \leq b_j, \quad j = 1, \ldots, r, \end{aligned}$$

is finite. Show that the problem has at least one optimal solution.

Solution: Let f^* be the optimal value and let X be the feasible region

$$X = \big\{x \mid a_j'x \leq b_j,\ j = 1, \ldots, r\big\}.$$

Let $\{\gamma_k\}$ be a scalar sequence with $\gamma_k \downarrow f^*$, and denote

$$S_k = \{x \in X \mid x'Qx + c'x \leq \gamma_k\}.$$

Then the set of optimal solutions of the problem is $\cap_{k=0}^{\infty} S_k$, so by Prop. 1.4.10, it will suffice to show that all asymptotic sequences of $\{S_k\}$ are retractive.

Let $\{x_k\}$ be an asymptotic sequence and let d be the corresponding common nonzero direction of recession of S_k. We claim that

$$d'Qd \leq 0, \qquad a_j'd \leq 0, \quad j = 1, \ldots, r. \tag{1.10}$$

Indeed, since $x_k \in S_k$, we have

$$x_k'Qx_k + c'x_k \leq \gamma_k,$$

and by denoting

$$d_k = \frac{x_k}{\|x_k\|}\|d\|,$$

it follows that

$$d_k{}'Qd_k + \frac{c'd_k}{\|x_k\|}\|d\| \leq \frac{\gamma_k}{\|x_k\|^2}\|d\|^2.$$

Taking the limit as $k \to \infty$, and using the fact $d_k \to d$ and $\|x_k\| \to \infty$, we see that $d'Qd \leq 0$. Also, for all j, we have $a_j'x_k \leq b_j$, so that

$$a_j'd_k \leq \frac{b_j}{\|x_k\|}\|d\|,$$

and by taking the limit as $k \to \infty$, we obtain $a_j'd \leq 0$, for all j.

For any $x \in X$, consider the vectors $\tilde{x}_k = x + kd$. We will show that for each j and all k sufficiently large, we have $a_j'(x_k - d) \leq b_j$. Indeed, for a given j, there are two possibilities:

(1) $\{a_j'x_k\}$ has a bounded subsequence. Then, along this subsequence, we have $(a_j'x_k/\|x_k\|) \to 0$ and hence $a_j'd = 0$, since $x_k/\|x_k\| \to d/\|d\|$. It follows that $a_j'(x_k - d) = a_j'x_k \leq b_j$ for all k.

(2) $a_j'x_k \to -\infty$, in which case $a_j'(x_k - d) \leq b_j$ for all sufficiently large k.

Thus, for k sufficiently large, we have $x_k - d \in X$.

Also, from Eq. (1.10) it can be seen that $\tilde{x}_k \in X$. Thus, the cost function value corresponding to $\tilde{x}_k$ satisfies

$$\begin{aligned} f^* &\leq (x + kd)'Q(x + kd) + c'(x + kd) \\ &= x'Qx + c'x + k^2 d'Qd + k(c + 2Qx)'d \\ &\leq x'Qx + c'x + k(c + 2Qx)'d, \end{aligned}$$

where the last inequality follows from the fact $d'Qd \leq 0$. From the finiteness of f^*, it follows that

$$(c + 2Qx)'d \geq 0, \qquad \forall\ x \in X.$$

We now show that $\{x_k\}$ is retractive, so that we can use Prop. 1.4.10. Indeed since $\|x_k\| \to \infty$, it follows that for k sufficiently large, we have $x_k - d \in X$. Furthermore, we have

$$\begin{aligned} f(x_k - d) &= (x_k - d)'Q(x_k - d) + c'(x_k - d) \\ &= x_k{}'Qx_k + c'x_k - (c + 2Qx_k)'d + d'Qd \\ &\le x_k{}'Qx_k + c'x_k \\ &\le \gamma_k, \end{aligned}$$

where the first inequality follows from the facts $d'Qd \le 0$ and $(c + 2Qx_k)'d \ge 0$ shown earlier. Thus for sufficiently large k, we have $x_k - d \in S_k$, so that $\{x_k\}$ is retractive. The existence of an optimal solution now follows from Prop. 1.4.10.

1.24 (Radon's Theorem)

Let $x_1, \ldots, x_m$ be vectors in $\Re^n$, where $m \ge n + 2$. Show that there exists a partition of the index set $\{1, \ldots, m\}$ into two disjoint sets I and J such that

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

As an illustration, show that given four points in the plane, either the (possibly degenerate) triangle formed by three of the points contains the fourth, or else the four points define a (possibly degenerate) quadrilateral. *Hint*: The system of $n + 1$ equations in the m unknowns $\lambda_1, \ldots, \lambda_m$,

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

has a nonzero solution λ^*. Let $I = \{i \mid \lambda_i^* \ge 0\}$ and $J = \{j \mid \lambda_j^* < 0\}$.

Solution: 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

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

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 scalars 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

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

Given four distinct points in the plane (i.e., $m = 4$ and $n = 2$), Radon's Theorem guarantees the existence of a partition into two subsets, the convex hulls of which intersect. Assuming, there is no subset of three points lying on the same line, there are two possibilities:

(1) Each set in the partition consists of two points, in which case the convex hulls intesect and define the diagonals of a quadrilateral.

(2) One set in the partition consists of three points and the other consists of one point, in which case the triangle formed by the three points must contain the fourth.

In the case where three of the points define a line segment on which they lie, and the fourth does not, the triangle formed by the two ends of the line segment and the point outside the line segment form a triangle that contains the fourth point. In the case where all four of the points lie on a line segment, the degenerate triangle formed by three of the points, including the two ends of the line segment, contains the fourth point.

1.25 (Helly's Theorem [Hel21])

Consider a collection $\mathcal{S}$ of convex sets in $\Re^n$, with at least $n + 1$ members, and assume that the intersection of every subcollection of $n + 1$ sets has nonempty intersection.

(a) Assuming that $\mathcal{S}$ is a finite collection, show that the entire collection has nonempty intersection. *Hint*: Use induction. Assume that the conclusion holds for every collection of M sets, where $M \geq n + 1$, and show that the conclusion holds for every collection of $M + 1$ sets. In particular, let $C_1, \ldots, C_{M+1}$ be a collection of $M+1$ convex sets, and consider the collection of $M+1$ sets $B_1, \ldots, B_{M+1}$, where

$$B_j = \cap_{\substack{i=1,\ldots,M+1,\\ i\neq j}} C_i, \qquad j = 1, \ldots, M+1.$$

Note that, by the induction hypothesis, each set B_j is the intersection of a collection of M sets that have the property that every subcollection of $n+1$ (or fewer) sets has nonempty intersection. Hence each set B_j is nonempty. Let x_j be a vector in B_j. Apply Radon's Theorem (Exercise 1.24) to the vectors $x_1, \ldots, x_{M+1}$. Show that any vector in the intersection of the corresponding convex hulls belongs to the intersection of $C_1, \ldots, C_{M+1}$.

(b) Assuming that the members of $\mathcal{S}$ are compact, show that the entire collection has nonempty intersection. *Hint*: Use part (a) and the fact that if a collection of compact sets has empty intersection, so does one of its finite subcollections [cf. Prop. A.2.4(i)].

(c) Use part (a) to show that given a finite family of vertical intervals on the plane every three of which can be intersected by a line, the entire family can be intersected by the same line.

Solution: (a) Consider the induction argument of the hint, 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 $I \cup J = \{1, \ldots, M+1\}$, 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 $C_1, \ldots, C_{M+1}$.

(b) Evident from the hint.

(c) Consider a finite family $\{S_i \mid i = 1, \ldots, m\}$ of vertical line segments on the plane:

$$S_i = \big\{(x, y) \mid x = x_i,\ \underline{y}_i \leq y \leq \overline{y}_i\big\}, \qquad i = 1, \ldots, m,$$

where $m \geq 3$, and $x_i, \underline{y}_i, \overline{y}_i$, $i = 1, \ldots, m$, are given scalars. For each i, consider the set of lines that intersect S_i:

$$C_i = \big\{(a, b) \mid \underline{y}_i \leq a x_i + b \leq \overline{y}_i\big\}.$$

The sets C_i are convex, and every three of them have a common point. By Helly's Theorem, it follows that all the sets C_i have a common point, which is a line that intersects all the intervals S_i.

1.26 (Kirchberger's Theorem [Kir1903])

Let S be a finite subset of $\Re^n$ with at least $n+2$ points, and let $S = B \cup R$ be a partition of S in two disjoint subsets B (the "blue" points) and R (the "red" points). Suppose that every subset $\overline{S}$ of $n+2$ points of S can be linearly separated, in the sense that there is a vector $\overline{a}$ and a scalar $\overline{c}$ such that $\overline{a}'b+\overline{c} < 0$ for all $b \in \overline{S} \cap B$ and $\overline{a}'r + \overline{c} > 0$ for all $r \in \overline{S} \cap R$. Use Helly's Theorem (Exercise 1.25) to show that the entire set S can be linearly separated, i.e., that there is a vector a and a scalar c such that $a'b + c < 0$ for all $b \in B$ and $a'r + c > 0$ for all $r \in R$. *Hint*: For each $b \in B$ consider the set $G(b)$ of vectors $(x_1, \ldots, x_{n+1})$ such that

$$\sum_{i=1}^{n} x_i b_i + x_{n+1} < 0,$$

and for each $r \in R$, consider the set $H(r)$ of vectors $(x_1, \ldots, x_{n+1})$ such that

$$\sum_{i=1}^{n} x_i r_i + x_{n+1} > 0.$$

Let $\mathcal{C}$ be the collection of the sets $G(b)$ and $H(r)$ as b and r ranges over B and R, respectively. Use Helly's Theorem (Exercise 1.25) to show that $\mathcal{C}$ has nonempty intersection.

Solution: For each $b \in B$ consider the set $G(b)$ of vectors $(x_1, \ldots, x_{n+1})$ such that

$$\sum_{i=1}^{n} x_i b_i + x_{n+1} < 0,$$

and for each $r \in R$, consider the set $H(r)$ of vectors $(x_1, \ldots, x_{n+1})$ such that

$$\sum_{i=1}^{n} x_i r_i + x_{n+1} > 0.$$

Let $\mathcal{C}$ be the collection of the convex sets $G(b)$ and $H(r)$ as b and r ranges over B and R, respectively. By assumption, for any subset $C \subset B \cup R$, consisting of $n+2$ points, the sets $B \cap C$ and $R \cap C$ can be linearly separated, so there exist $\overline{a} \in \Re^n$ and $\overline{c} \in \Re$ such that

$$\overline{a}'b + \overline{c} < 0, \qquad \forall\ b \in B \cap C,$$

$$\overline{a}'r + \overline{c} > 0, \qquad \forall\ r \in R \cap C.$$

Thus, $(\overline{a}, \overline{c}) \in L(b)$ for all $b \in B \cap C$, and $(\overline{a}, \overline{c}) \in G(r)$ for all $r \in R \cap C$. It follows that $\mathcal{C}$ is a finite family of convex sets in $\Re^{n+1}$, which contains at least $n+2$ members and every collection of $n+2$ of these members has nonempty intersection. By Helly's Theorem, there is a vector (a, c) that belongs to all members of $\mathcal{C}$, and for which we have $a'x + c < 0$ for all $x \in B$ and $a'x + c > 0$ for all $x \in R$. (Proof given in Webster [Web02], and credited to H. Rademacher and I. J. Shoenberg, "Helly's Theorem on Convex Domains and Tchebycheff's Approximation Problem," Canadian J. of Math., Vol. 2, 1950, pp. 245-256.)

1.27 (Krasnosselsky's Theorem [Kra46])

Let S be a nonempty compact subset of $\Re^n$. For any two points x and y of S, we say that x *is visible from* y if the line segment connecting x and y belongs to S. Assume that S has the property that for any subset of $n+1$ points of S, there is a point of S from which all $n+1$ points are visible. Show that there is a point in S from which all points of S are visible. *Hint*: For each $y \in S$, let S_y be the set of points of S that are visible from y. Show that the set $C_y = \text{conv}(S_y)$ is compact, and consider the family of sets $\{C_y \mid y \in S\}$. Use Helly's Theorem (Exercise 1.25) to show that there is a vector $a \in S$ that belongs to $\cap_{y \in S} \text{conv}(S_y)$. Show that $a \in \cap_{y \in S} S_y$ (this last part is not simple).

Solution: For each $y \in S$, let S_y be the set of points of S that are visible from y. The set S_y is easily seen to be closed, and hence its convex hull, $C_y = \text{conv}(S_y)$, is compact by Prop. 1.2.2. Consider the family of sets $\{C_y \mid y \in S\}$. Let $y_0, \ldots, y_n$ be points in S. By the hypothesis, there is a vector $x \in S$ from which $y_0, \ldots, y_n$ are visible. Thus $x \in S_{y_0} \cap \cdots \cap S_{y_n}$, and hence also $x \in C_{y_0} \cap \cdots \cap C_{y_n}$. It follows that any subcollection of $n+1$ sets from the family $\{C_y \mid y \in S\}$ is nonempty. By Helly's Theorem [Exercise 1.25(b)], the entire family is nonempty. Thus, there exists a vector a such that

$$a \in \text{conv}(S_y), \qquad \forall\, y \in S.$$

We claim now that every $y \in S$ is visible from a. Assume the contrary, so there exists a vector $b \in S$ and a vector c in the line segment connecting a and b such that $c \notin S$. Let C be a closed ball of nonzero radius, which is centered at c and does not intersect S. Let

$$\alpha = \inf\big\{\lambda \geq 0 \mid S \cap \big(C + \lambda(b-c)\big) \neq \varnothing\big\},$$

denote the closed ball $C + \alpha(b-c)$ by D and denote its center by d. Then by construction, S meets the boundary of D but not its interior. Let e a vector in $S \cap D$. We will show that $a \notin \text{conv}(S_e)$, thus arriving at a contradiction.

Indeed, consider the halfspaces

$$H^- = \big\{z \mid (z-e)'(e-d) < 0\big\}, \qquad H^+ = \big\{z \mid (z-e)'(e-d) \geq 0\big\}.$$

Then, by elementary geometry, it follows that $a \in H^-$, while $S_e \subset H^+$ and hence also $\text{conv}(S_e) \subset H^+$. Since $H^- \cap H^+ = \varnothing$, it follows that $a \notin \text{conv}(S_e)$, a contradiction. (Proof given in Webster [Web02].)

1.28

(a) Let C_1 be a convex set with nonempty interior and C_2 be a nonempty convex set that does not intersect the interior of C_1. Show that there exists a hyperplane such that one of the associated closed halfspaces contains C_2, and does not intersect the interior of C_1.

(b) Show by an example that we cannot replace interior with relative interior in the statement of part (a).

Solution: (a) In view of the assumption that $\mathrm{int}(C_1)$ and C_2 are disjoint and convex [cf Prop. 1.1.1(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

$$a'x \leq b, \qquad \forall\, x \in \mathrm{int}(C_1). \tag{1.11}$$

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. (1.11). 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 $\mathrm{ri}(C_1)$.

1.29

Let C be a nonempty convex set in $\Re^n$, and let M be a nonempty affine set in $\Re^n$. Show that $M \cap \mathrm{ri}(C) = \emptyset$ is a necessary and sufficient condition for the existence of a hyperplane H containing M, and such that $\mathrm{ri}(C)$ is contained in one of the open halfspaces associated with H.

Solution: 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. 1.5.5. 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. 1.3.4. This contradicts the proper separation property, thus showing that $\mathrm{ri}(C)$ is contained in one of the open halfspaces.

1.30

Let C_1 and C_2 be nonempty convex subsets of $\Re^n$ such that C_2 is a cone.

(a) Suppose that there exists a hyperplane that separates C_1 and C_2 properly. Show that there exists a hyperplane which separates C_1 and C_2 properly and passes through the origin.

(b) Suppose that there exists a hyperplane that separates C_1 and C_2 strictly. Show that there exists a hyperplane that passes through the origin such that one of the associated closed halfspaces contains the cone C_2 and does not intersect C_1.

Solution: (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

$$\inf_{x \in C_1} a'x \geq \sup_{x \in C_2} a'x, \tag{1.12}$$

$$\sup_{x \in C_1} a'x > \inf_{x \in C_2} a'x. \tag{1.13}$$

Let

$$b = \sup_{x \in C_2} a'x. \tag{1.14}$$

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. (1.14) 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

$$a'x_2 < \beta < a'x_1, \qquad \forall\, x_1 \in C_1, \quad \forall\, x_2 \in C_2. \tag{1.15}$$

We choose b to be

$$b = \sup_{x \in C_2} a'x, \tag{1.16}$$

and consider the closed halfspace

$$K = \{x \mid a'x \leq b\},$$

which contains C_2. By Eq. (1.15), 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.

1.31 (Separation Properties of Cones)

Define a *homogeneous halfspace* to be a closed halfspace associated with a hyperplane that passes through the origin. Show that:

(a) A nonempty closed convex cone is the intersection of the homogeneous halfspaces that contain it.

(b) The closure of the convex cone generated by a nonempty set X is the intersection of all the homogeneous halfspaces containing X.

Solution: (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 1.30(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 $\mathrm{cl}\big(\mathrm{cone}(X)\big)$.

1.32 (Strong Separation)

Let C_1 and C_2 be nonempty convex subsets of $\Re^n$, and let B denote the unit ball in $\Re^n$, $B = \{x \mid \|x\| \leq 1\}$. A hyperplane H is said to *separate strongly* C_1 and C_2 if there exists an $\epsilon > 0$ such that $C_1 + \epsilon B$ is contained in one of the open halfspaces associated with H and $C_2 + \epsilon B$ is contained in the other. Show that:

(a) The following three conditions are equivalent.

  (i) There exists a hyperplane separating strongly C_1 and C_2.

  (ii) There exists a vector $a \in \Re^n$ such that $\inf_{x \in C_1} a'x > \sup_{x \in C_2} a'x$.

  (iii) $\inf_{x_1 \in C_1,\, x_2 \in C_2} \|x_1 - x_2\| > 0$, i.e., $0 \notin \mathrm{cl}(C_2 - C_1)$.

(b) If C_1 and C_2 are disjoint, any one of the five conditions for strict separation, given in Prop. 1.5.3, implies that C_1 and C_2 can be strongly separated.

Solution: (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

$$\inf_{x \in C_1} a'x > \sup_{x \in C_2} a'x, \tag{1.17}$$

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. 1.5.3 imply condition (1) of that proposition (see the discussion in the proof of Prop. 1.5.3), 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.

1.33

Let C be a nonempty closed convex subset of $\Re^{n+1}$ that contains no vertical lines. Show that C is equal to the intersection of the closed halfspaces that contain it and correspond to nonvertical hyperplanes.

Solution: 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. 1.5.8, 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.

1.34 (Logarithmic/Exponential Conjugacy)

Let $f : \Re \mapsto \Re$ be the exponential function

$$f(x) = e^x.$$

Show that the conjugate is

$$f^\star(y) = \begin{cases} y \ln y - y & \text{if } y > 0, \\ 0 & \text{if } y = 0, \\ \infty & \text{if } y < 0. \end{cases}$$

Solution: The conjugate is

$$f^\star(y) = \sup_{x \in \Re} \{xy - e^x\},$$

and can be calculated in closed form. For $y < 0$, by taking $x \to -\infty$, we see that $xy - e^x$ can be made arbitrarily large, so $f^\star(y) = \infty$. For $y = 0$, we have

$$f^\star(0) = \sup_{x \in \Re} \{-e^x\} = -\inf_{x \in \Re} e^x = 0.$$

Finally, for $y > 0$, by setting the derivative of $xy - e^x$ to zero, we see that the supremum of $xy - e^x$ is obtained for $x = \ln y$, and by substitution, we obtain $f^\star(y) = y \ln y - y$.

1.35 (Conjugates of p-Norms)

Let $f : \Re \mapsto \Re$ be the function

$$f(x) = \frac{1}{p}|x|^p,$$

where $1 < p$. Show that the conjugate is

$$f^\star(y) = \frac{1}{q}|y|^q, \tag{1.18}$$

where q is defined by the relation

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Solution: We set the derivative of $xy - (1/p)|x|^p$ to zero, and we see that the supremum over x is attained when $\operatorname{sgn}(x)|x|^{p-1} = y$, which implies that $xy = |x|^p$ and $|x|^{p-1} = |y|$. By substitution in the formula for the conjugate, we obtain

$$f^\star(y) = |x|^p - \frac{1}{p}|x|^p = \left(1 - \frac{1}{p}\right)|x|^p = \frac{1}{q}|y|^{\frac{p}{p-1}} = \frac{1}{q}|y|^q.$$

We now note that for any function $f : \Re^n \mapsto (-\infty, \infty]$ that has the form

$$f(x) = f_1(x_1) + \cdots + f_n(x_n),$$

where $x = (x_1, \ldots, x_n)$ and $f_i : \Re \mapsto (-\infty, \infty]$, $i = 1, \ldots, n$, the conjugate is given by

$$f^\star(y) = f_1^\star(y_1) + \cdots + f_n^\star(y_n),$$

where $f_i^\star : \Re \mapsto (-\infty, \infty]$ is the conjugate of f_i, $i = 1, \ldots, n$. By combining this fact with the formula (1.18), we see that the conjugate of the function

$$f(x) = \frac{1}{p}\sum_{i=1}^n |x_i|^p$$

is the function

$$f^\star(y) = \frac{1}{q}\sum_{i=1}^n |y_i|^q.$$

1.36 (Conjugate of a Quadratic)

Let

$$f(x) = \tfrac{1}{2}x'Qx + a'x + b,$$

where Q is a symmetric positive semidefinite $n \times n$ matrix, a is a vector in $\Re^n$, and b is a scalar. Derive the conjugate of f.

Solution: Let us assume first that Q is nonsingular. Then the maximum of $x'y - f(x)$ over x is attained when $Qx + a = \nabla f(x) = y$. By substitution, we obtain

$$f^\star(y) = x'y - f(x) = x'(Qx + a) - \tfrac{1}{2}x'Qx - a'x - b = \tfrac{1}{2}x'Qx - b,$$

and finally, using $x = Q^{-1}(y - a)$,

$$f^\star(y) = \tfrac{1}{2}(y - a)'Q^{-1}(y - a) - b.$$

Consider now the general case where Q may be singular. Then if the equation $y = Qx+a$ has no solution, i.e., $y-a$ does not belong to the range $R(Q)$ of Q, we have $f^\star(y) = \infty$. Otherwise, let x be the solution of the equation $y = Qx+a$ that has minimum Euclidean norm. Then it is known (see e.g., Luenberger, Optimization by Vector Space Methods, 1969, p. 165) that x is linearly related to $y - a$ and can be written as $x = Q^\dagger(y - a)$ where $Q^\dagger$ is a symmetric positive semidefinite matrix, called the *pseudoinverse* of Q, which satisfies $Q^\dagger Q Q^\dagger = Q^\dagger$. Similar to the case where Q is invertible, we have $f^\star(y) = (1/2)x'Qx - b$, and it follows, using $x = Q^\dagger(y - a)$, that

$$f^\star(y) = \begin{cases} \tfrac{1}{2}(y - a)'Q^\dagger(y - a) - b & \text{if } y - a \in R(Q), \\ \infty & \text{otherwise.} \end{cases}$$

1.37 (Support Function of a Bounded Ellipsoid)

Let X be an ellipsoid of the form

$$X = \big\{x \mid (x - \overline{x})'Q(x - \overline{x}) \le b\big\},$$

where Q is a symmetric positive definite matrix, $\overline{x}$ is a vector, and b is a positive scalar. Calculate the support function of X.

Solution: To calculate $\sigma_X(y)$, we write

$$X = \{\overline{x}\} + \overline{X},$$

where

$$\overline{X} = \big\{x \mid x'Qx \le b\big\},$$

we calculate the support function $\sigma_{\overline{X}}(y)$, and we use the equation

$$\sigma_X(y) = y'\overline{x} + \sigma_{\overline{X}}(y). \tag{1.19}$$

To calculate

$$\sigma_{\overline{X}}(y) = \sup_{x'Qx \le b} y'x,$$

we introduce the transformation $z = Q^{1/2}x$ and we write

$$\sigma_{\overline{X}}(y) = \sup_{\|z\| \le b^{1/2}} y'Q^{-1/2}z.$$

It can be seen that for $y \neq 0$, the supremum over z above is attained at

$$z(y) = b^{1/2} \frac{Q^{-1/2}y}{\|Q^{-1/2}y\|},$$

and by substitution in the expression for $\sigma_{\overline{X}}(y)$, we have

$$\sigma_{\overline{X}}(y) = \left(b\, y'Q^{-1}y\right)^{1/2}.$$

Thus, using Eq. (1.19), we finally obtain

$$\sigma_X(y) = y'\overline{x} + \left(b\, y'Q^{-1}y\right)^{1/2}, \qquad \forall\ y \in \Re^n.$$