Convex Analysis and Optimization

Chapter 4 Solutions

Dimitri P. Bertsekas

with

Angelia Nedić and Asuman E. Ozdaglar

Massachusetts Institute of Technology

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

CHAPTER 4: SOLUTION MANUAL

4.1 (Directional Derivative of Extended Real-Valued Functions)

Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function, and let x be a vector in $\mathrm{dom}(f)$. Define

$$f'(x; y) = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x)}{\alpha}, \qquad y \in \Re^n.$$

Show the following:

(a) $f'(x; \lambda y) = \lambda f'(x; y)$ for all $\lambda \geq 0$ and $y \in \Re^n$.

(b) $f'(x; \cdot)$ is a convex function.

(c) $-f'(x; -y) \leq f'(x; y)$ for all $y \in \Re^n$.

(d) If $\mathrm{dom}(f) = \Re^n$, then the level set $\big\{y \mid f'(x; y) \leq 0\big\}$ is a closed convex cone and its polar is given by

$$\Big(\big\{y \mid f'(x; y) \leq 0\big\}\Big)^* = \mathrm{cl}\Big(\mathrm{cone}\big(\partial f(x)\big)\Big).$$

Solution: (a) Since $f'(x; 0) = 0$, the relation $f'(x; \lambda y) = \lambda f'(x; y)$ clearly holds for $\lambda = 0$ and all $y \in \Re^n$. Choose $\lambda > 0$ and $y \in \Re^n$. By the definition of directional derivative, we have

$$f'(x; \lambda y) = \inf_{\alpha > 0} \frac{f\big(x + \alpha(\lambda y)\big) - f(x)}{\alpha} = \lambda \inf_{\alpha > 0} \frac{f\big(x + (\alpha\lambda) y\big) - f(x)}{\alpha\lambda}.$$

By setting $\beta = \lambda\alpha$ in the preceding relation, we obtain

$$f'(x; \lambda y) = \lambda \inf_{\beta > 0} \frac{f(x + \beta y) - f(x)}{\beta} = \lambda f'(x; y).$$

(b) Let (y_1, w_1) and (y_2, w_2) be two points in $\mathrm{epi}\big(f'(x; \cdot)\big)$, and let γ be a scalar with $\gamma \in (0, 1)$. Consider a point (y_γ, w_γ) given by

$$y_\gamma = \gamma y_1 + (1 - \gamma) y_2, \qquad w_\gamma = \gamma w_1 + (1 - \gamma) w_2.$$

Since for all $y \in \Re^n$, the ratio

$$\frac{f(x + \alpha y) - f(x)}{\alpha}$$

is monotonically nonincreasing as $\alpha \downarrow 0$, we have

$$\frac{f(x + \alpha y_1) - f(x)}{\alpha} \leq \frac{f(x + \alpha_1 y_1) - f(x)}{\alpha_1}, \qquad \forall\ \alpha, \alpha_1,\ \text{with } 0 < \alpha \leq \alpha_1,$$

$$\frac{f(x+\alpha y_2)-f(x)}{\alpha} \le \frac{f(x+\alpha_2 y_2)-f(x)}{\alpha_2}, \qquad \forall\ \alpha,\alpha_2,\ \text{with } 0<\alpha\le\alpha_2.$$

Multiplying the first relation by γ and the second relation by $1-\gamma$, and adding, we have for all α with $0<\alpha\le\alpha_1$ and $0<\alpha\le\alpha_2$,

$$\frac{\gamma f(x+\alpha y_1)+(1-\gamma)f(x+\alpha y_2)-f(x)}{\alpha} \le \gamma\frac{f(x+\alpha_1 y_1)-f(x)}{\alpha_1} + (1-\gamma)\frac{f(x+\alpha_2 y_2)-f(x)}{\alpha_2}.$$

From the convexity of f and the definition of y_γ, it follows that

$$f(x+\alpha y_\gamma)\le \gamma f(x+\alpha\gamma y_1)+(1-\gamma)f(x+\alpha y_2).$$

Combining the preceding two relations, we see that for all $\alpha\le\alpha_1$ and $\alpha\le\alpha_2$,

$$\frac{f(x+\alpha y_\gamma)-f(x)}{\alpha} \le \gamma\frac{f(x+\alpha_1 y_1)-f(x)}{\alpha_1}+(1-\gamma)\frac{f(x+\alpha_2 y_2)-f(x)}{\alpha_2}.$$

By taking the infimum over α, and then over α_1 and α_2, we obtain

$$f'(x;y_\gamma)\le \gamma f'(x;y_1)+(1-\gamma)f'(x;y_2)\le \gamma w_1+(1-\gamma)w_2=w_\gamma,$$

where in the last inequality we use the fact $(y_1,w_1),(y_2,w_2)\in \mathrm{epi}\big(f'(x;\cdot)\big)$ Hence the point (y_γ,w_γ) belongs to $\mathrm{epi}\big(f'(x;\cdot)\big)$, implying that $f'(x;\cdot)$ is a convex function.

(c) Since $f'(x;0)=0$ and $(1/2)y+(1/2)(-y)=0$, it follows that

$$f'\big(x;(1/2)y+(1/2)(-y)\big)=0, \qquad \forall\ y\in\Re^n.$$

By part (b), the function $f'(x;\cdot)$ is convex, so that

$$0\le (1/2)f'(x;y)+(1/2)f'(x;-y),$$

and

$$-f'(x;-y)\le f'(x;y).$$

(d) Let a vector $\overline{y}$ be in the level set $\big\{y \mid f'(x;y)\le 0\big\}$, and let $\lambda>0$. By part (a),

$$f'(x;\lambda\overline{y})=\lambda f'(x;\overline{y})\le 0,$$

so that $\lambda\overline{y}$ also belongs to this level set, which is therefore a cone. By part (b), the function $f'(x;\cdot)$ is convex, implying that the level set $\big\{y \mid f'(x;y)\le 0\big\}$ is convex.

Since $\mathrm{dom}(f)=\Re^n$, $f'(x;\cdot)$ is a real-valued function, and since it is convex, by Prop. 1.4.6, it is also continuous over $\Re^n$. Therefore the level set $\big\{y \mid f'(x;y)\le 0\big\}$ is closed.

We now show that

$$\Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^* = \mathrm{cl}\Big(\mathrm{cone}\big(\partial f(x)\big)\Big).$$

By Prop. 4.2.2, we have

$$f'(x;y) = \max_{d \in \partial f(x)} y'd,$$

implying that $f'(x;y) \le 0$ if and only if $\max_{d \in \partial f(x)} y'd \le 0$. Equivalently, $f'(x;y) \le 0$ if and only if

$$y'd \le 0, \qquad \forall\ d \in \partial f(x).$$

Since

$$y'd \le 0, \qquad \forall\ d \in \partial f(x) \qquad \Longleftrightarrow \qquad y'd \le 0, \qquad \forall\ d \in \mathrm{cone}\big(\partial f(x)\big),$$

it follows from Prop. 3.1.1(a) that $f'(x;y) \le 0$ if and only if

$$y'd \le 0, \qquad \forall\ d \in \mathrm{cone}\big(\partial f(x)\big).$$

Therefore

$$\big\{y \mid f'(x;y) \le 0\big\} = \Big(\mathrm{cone}\big(\partial f(x)\big)\Big)^*,$$

and the desired relation follows by the Polar Cone Theorem [Prop. 3.1.1(b)].

4.2 (Chain Rule for Directional Derivatives)

Let $f : \Re^n \mapsto \Re^m$ and $g : \Re^m \mapsto \Re$ be some functions, and let x be a vector in $\Re^n$. Suppose that all the components of f and g are directionally differentiable at x, and that g is such that for all $w \in \Re^m$,

$$g'(y;w) = \lim_{\alpha \downarrow 0,\ z \to w} \frac{g\big(y + \alpha z\big) - g(y)}{\alpha}.$$

Then, the composite function $F(x) = g\big(f(x)\big)$ is directionally differentiable at x and the following chain rule holds:

$$F'(x;d) = g'\big(f(x); f'(x;d)\big), \qquad \forall\ d \in \Re^n.$$

Solution: For any $d \in \Re^n$, by using the directional differentiability of f at x, we have

$$\begin{aligned} F(x + \alpha d) - F(x) &= g\big(f(x + \alpha d)\big) - g\big(f(x)\big) \\ &= g\big(f(x) + \alpha f'(x;d) + o(\alpha)\big) - g\big(f(x)\big). \end{aligned}$$

Let $z_\alpha = f'(x;d) + o(\alpha)/\alpha$ and note that $z_\alpha \to f'(x;d)$ as $\alpha \downarrow 0$. By using this and the assumed property of g, we obtain

$$\lim_{\alpha \downarrow 0} \frac{F(x + \alpha d) - F(x)}{\alpha} = \lim_{\alpha \downarrow 0} \frac{g\big(f(x) + \alpha z_\alpha\big) - g\big(f(x)\big)}{\alpha} = g'\big(f(x); f'(x;d)\big),$$

showing that F is directionally differentiable at x and that the given chain rule holds.

4.3

Let $f : \Re^n \mapsto \Re$ be a convex function. Show that a vector $d \in \Re^n$ is a subgradient of f at x if and only if the function $d'y - f(y)$ attains its maximum at $y = x$.

Solution: By definition, a vector $d \in \Re^n$ is a subgradient of f at x if and only if

$$f(y) \geq f(x) + d'(y - x), \qquad \forall\ y \in \Re^n,$$

or equivalently

$$d'x - f(x) \geq d'y - f(y), \qquad \forall\ y \in \Re^n.$$

Therefore, $d \in \Re^n$ is a subgradient of f at x if and only if

$$d'x - f(x) = \max_y \big\{ d'y - f(y) \big\}.$$

4.4

Show that:

(a) For the function $f(x) = \|x\|$, we have

$$\partial f(x) = \begin{cases} \big\{ x/\|x\| \big\} & \text{if } x \neq 0, \\ \big\{ d \mid \|d\| \leq 1 \big\} & \text{if } x = 0. \end{cases}$$

(b) For a nonempty convex subset C of $\Re^n$ and the function $f : \Re^n \mapsto (-\infty, \infty]$ given by

$$f(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C, \end{cases}$$

we have

$$\partial f(x) = \begin{cases} N_C(x) & \text{if } x \in C, \\ \emptyset & \text{if } x \notin C. \end{cases}$$

Solution: (a) For $x \neq 0$, the function $f(x) = \|x\|$ is differentiable with $\nabla f(x) = x/\|x\|$, so that $\partial f(x) = \big\{ \nabla f(x) \big\} = \big\{ x/\|x\| \big\}$. Consider now the case $x = 0$. If a vector d is a subgradient of f at $x = 0$, then $f(z) \geq f(0) + d'z$ for all z, implying that

$$\|z\| \geq d'z, \qquad \forall\ z \in \Re^n.$$

By letting $z = d$ in this relation, we obtain $\|d\| \leq 1$, showing that $\partial f(0) \subset \big\{ d \mid \|d\| \leq 1 \big\}$.

On the other hand, for any $d \in \Re^n$ with $\|d\| \leq 1$, we have

$$d'z \leq \|d\| \cdot \|z\| \leq \|z\|, \qquad \forall\ z \in \Re^n,$$

which is equivalent to $f(0) + d'z \leq f(z)$ for all z, so that $d \in \partial f(0)$, and therefore $\big\{ d \mid \|d\| \leq 1 \big\} \subset \partial f(0)$.

Note that an alternative proof is obtained by writing

$$\|x\| = \max_{\|z\| \leq 1} x'z,$$

and by using Danskin's Theorem (Prop. 4.5.1).

(b) By convention $\partial f(x) = \emptyset$ when $x \notin \text{dom}(f)$, and since here $\text{dom}(f) = C$, we see that $\partial f(x) = \emptyset$ when $x \notin C$. Let now $x \in C$. A vector d is a subgradient of f at x if and only if

$$d'(z - x) \leq f(z), \qquad \forall\, z \in \Re^n.$$

Because $f(z) = \infty$ for all $z \notin C$, the preceding relation always holds when $z \notin C$, so the points $z \notin C$ can be ignored. Thus, $d \in \partial f(x)$ if and only if

$$d'(z - x) \leq 0, \qquad \forall\, z \in C.$$

Since C is convex, by Prop. 4.6.3, the preceding relation is equivalent to $d \in N_C(x)$, implying that $\partial f(x) = N_C(x)$ for all $x \in C$.

4.5

Show that for a scalar convex function $f : \Re \mapsto \Re$, we have

$$\partial f(x) = \big\{ d \mid f^-(x) \leq d \leq f^+(x) \big\}, \qquad \forall\, x \in \Re.$$

Solution: When f is defined on the real line, by Prop. 4.2.1, $\partial f(x)$ is a compact interval of the form

$$\partial f(x) = [\alpha, \beta].$$

By Prop. 4.2.2, we have

$$f'(x; y) = \max_{d \in \partial f(x)} y'd, \qquad \forall\, y \in \Re^n,$$

from which we see that

$$f'(x; 1) = \alpha, \qquad f'(x; -1) = \beta.$$

Since

$$f'(x; 1) = f^+(x), \qquad f'(x; -1) = f^-(x),$$

we have

$$\partial f(x) = [f^-(x), f^+(x)].$$

4.6

Let $f : \Re^n \mapsto \Re$ be a convex function, and let x and y be given vectors in $\Re^n$. Consider the scalar function $\varphi : \Re \mapsto \Re$ defined by $\varphi(t) = f\big(tx + (1-t)y\big)$ for all $t \in \Re$, and show that

$$\partial\varphi(t) = \Big\{(x-y)'d \,\Big|\, d \in \partial f\big(tx + (1-t)y\big)\Big\}, \qquad \forall\, t \in \Re.$$

Hint: Apply the Chain Rule [Prop. 4.2.5(a)].

Solution: We can view the function

$$\varphi(t) = f\big(tx + (1-t)y\big), \qquad t \in \Re$$

as the composition of the form

$$\varphi(t) = f\big(g(t)\big), \qquad t \in \Re,$$

where $g(t) : \Re \mapsto \Re^n$ is an affine function given by

$$g(t) = y + t(x-y), \qquad t \in \Re.$$

By using the Chain Rule [Prop. 4.2.5(a)], where $A = (x-y)$, we obtain

$$\partial\varphi(t) = A'\,\partial f\big(g(t)\big), \qquad \forall\, t \in \Re,$$

or equivalently

$$\partial\varphi(t) = \Big\{(x-y)'d \mid d \in \partial f\big(tx + (1-t)y\big)\Big\}, \qquad \forall\, t \in \Re.$$

4.7

Let $f : \Re^n \mapsto \Re$ be a convex function, and let X be a nonempty bounded subset of $\Re^n$. Show that f is Lipschitz continuous over X, i.e., that there exists a scalar L such that

$$\big|f(x) - f(y)\big| \le L\,\|x-y\|, \qquad \forall\, x, y \in X.$$

Show also that

$$f'(x; y) \le L\,\|y\|, \qquad \forall\, x \in X, \quad \forall\, y \in \Re^n.$$

Hint: Use the boundedness property of the subdifferential (Prop. 4.2.3).

Solution: Let x and y be any two points in the set X. Since $\partial f(x)$ is nonempty, by using the subgradient inequality, it follows that

$$f(x) + d'(x-y) \le f(y), \qquad \forall\, d \in \partial f(x),$$

implying that

$$f(x) - f(y) \le \|d\| \cdot \|x - y\|, \qquad \forall\ d \in \partial f(x).$$

According to Prop. 4.2.3, the set $\cup_{x\in X} \partial f(x)$ is bounded, so that for some constant $L > 0$, we have

$$\|d\| \le L, \qquad \forall\ d \in \partial f(x), \quad \forall\ x \in X, \tag{4.1}$$

and therefore,

$$f(x) - f(y) \le L\ \|x - y\|.$$

By exchanging the roles of x and y, we similarly obtain

$$f(y) - f(x) \le L\ \|x - y\|,$$

and by combining the preceding two relations, we see that

$$\big|f(x) - f(y)\big| \le L\ \|x - y\|,$$

showing that f is Lipschitz continuous over X.

Also, by using Prop. 4.2.2 and the subgradient boundedness [Eq. (4.1)], we obtain

$$f'(x; y) = \max_{d\in\partial f(x)} d'y \le \max_{d\in\partial f(x)} \|d\| \cdot \|y\| \le L\ \|y\|, \qquad \forall\ x \in X, \quad \forall\ y \in \Re^n.$$

4.8 (Nonemptiness of Subdifferential)

Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function, and let x be a vector in $\mathrm{dom}(f)$. Show that $\partial f(x)$ is nonempty if and only if $f'(x; z - x)$ is finite for all $z \in \mathrm{dom}(f)$.

Solution: Suppose that $\partial f(x)$ is nonempty, and let $z \in \mathrm{dom}(f)$. By the definition of subgradient, for any $d \in \partial f(x)$, we have

$$\frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha} \ge d'(z - x), \qquad \forall\ \alpha > 0,$$

implying that

$$f'(x; z - x) = \inf_{\alpha>0} \frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha} \ge d'(z - x) > -\infty.$$

Furthermore, for all $\alpha \in (0, 1)$, and $x, z \in \mathrm{dom}(f)$, the vector $x+\alpha(z-x)$ belongs to $\mathrm{dom}(f)$. Therefore, for all $\alpha \in (0, 1)$,

$$\frac{f\big(x + \alpha(z - x)\big) - f(x)}{\alpha} < \infty,$$

implying that

$$f'(x; z - x) < \infty.$$

Hence, $f'(x; z - x)$ is finite.

Converseley, suppose that $f'(x; z - x)$ is finite for all $z \in \mathrm{dom}(f)$. Fix a vector $\overline{x}$ in the relative interior of $\mathrm{dom}(f)$. Consider the set

$$C = \big\{(z, \nu) \mid z \in \mathrm{dom}(f),\ f(x) + f'(x; z - x) < \nu\big\},$$

and the halfline

$$P = \big\{(u, \zeta) \mid u = x + \beta(\overline{x} - x),\ \zeta = f(x) + \beta f'(x; \overline{x} - x),\ \beta \geq 0\big\}.$$

By Exercise 4.1(b), the directional derivative function $f'(x; \cdot)$ is convex, implying that $f'(x; z - x)$ is convex in z. Therefore, the set C is convex. Furthermore, being a halfline, the set P is polyhedral.

Suppose that C and P have a point (z, ν) in common, so that we have

$$z \in \mathrm{dom}(f), \qquad f(x) + f'(x; z - x) < \nu, \tag{4.2}$$

$$z = x + \beta(\overline{x} - x), \qquad \nu = f(x) + \beta f'(x; \overline{x} - x),$$

for some scalar $\beta \geq 0$. Because $\beta f'(x; y) = f'(x; \beta y)$ for all $\beta \geq 0$ and $y \in \Re^n$ [see Exercise 4.1(a)], it follows that

$$\nu = f(x) + f'\big(x; \beta(\overline{x} - x)\big) = f(x) + f'\big(x; z - x),$$

contradicting Eq. (4.2), and thus showing that C and P do not have any common point. Hence, $\mathrm{ri}(C)$ and P do not have any common point, so by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), the polyhedral set P and the convex set C can be properly separated by a hyperplane that does not contain C, i.e., there exists a vector $(a, \gamma) \in \Re^{n+1}$ such that

$$a'z + \gamma\nu \geq a'\big(x + \beta(\overline{x} - x)\big) + \gamma\big(f(x) + \beta f'(x; \overline{x} - x)\big), \quad \forall\ (z, \nu) \in C, \quad \forall\ \beta \geq 0, \tag{4.3}$$

$$\inf_{(z,\nu)\in C} \big\{a'z + \gamma\nu\big\} < \sup_{(z,\nu)\in C} \big\{a'z + \gamma\nu\big\}, \tag{4.4}$$

We cannot have $\gamma < 0$ since then the left-hand side of Eq. (4.3) could be made arbitrarily small by choosing ν sufficiently large. Also if $\gamma = 0$, then for $\beta = 1$, from Eq. (4.3) we obtain

$$a'(z - \overline{x}) \geq 0, \qquad \forall\ z \in \mathrm{dom}(f).$$

Since $\overline{x} \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, we have that the linear function $a'z$ attains its minimum over $\mathrm{dom}(f)$ at a point in the relative interior of $\mathrm{dom}(f)$. By Prop. 1.4.2, it follows that $a'z$ is constant over $\mathrm{dom}(f)$, i.e., $a'z = a'\overline{x}$ for all $z \in \mathrm{dom}(f)$, contradicting Eq. (4.4). Hence, we must have $\gamma > 0$ and by dividing with γ in Eq. (4.3), we obtain

$$\overline{a}'z + \nu \geq \overline{a}'\big(x + \beta(\overline{x} - x)\big) + f(x) + \beta f'(x; \overline{x} - x), \qquad \forall\ (z, \nu) \in C, \quad \forall\ \beta \geq 0,$$

where $\overline{a} = a/\gamma$. By letting $\beta = 0$ and $\nu \downarrow f(x) + f'(x; z-x)$ in this relation, and by rearranging terms, we have

$$f'(x; z-x) \geq (-\overline{a})'(z-x), \qquad \forall\, z \in \mathrm{dom}(f).$$

Because

$$f(z) - f(x) = f\big(x + (z-x)\big) - f(x) \geq \inf_{\lambda > 0} \frac{f\big(x + \lambda(z-x)\big) - f(x)}{\lambda} = f'(x; z-x),$$

it follows that

$$f(z) - f(x) \geq (-\overline{a})'(z-x), \qquad \forall\, z \in \mathrm{dom}(f).$$

Finally, by using the fact $f(z) = \infty$ for all $z \notin \mathrm{dom}(f)$, we see that

$$f(z) - f(x) \geq (-\overline{a})'(z-x), \qquad \forall\, z \in \Re^n,$$

showing that $-\overline{a}$ is a subgradient of f at x and that $\partial f(x)$ is nonempty.

4.9 (Subdifferential of Sum of Extended Real-Valued Functions)

This exercise is a refinement of Prop. 4.2.4. Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be convex functions, and let $f = f_1 + \cdots + f_m$. Show that

$$\partial f_1(x) + \cdots + \partial f_m(x) \subset \partial f(x), \qquad \forall\, x \in \Re^n.$$

Furthermore, if

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

then

$$\partial f_1(x) + \cdots + \partial f_m(x) = \partial f(x), \qquad \forall\, x \in \Re^n.$$

In addition, if the functions f_i, $i = r+1, \ldots, m$, are polyhedral, the preceding relation holds under the weaker assumption that

$$\Big(\cap_{i=1}^r \mathrm{ri}\big(\mathrm{dom}(f_i)\big)\Big) \cap \big(\cap_{i=r+1}^m \mathrm{dom}(f_i)\big) \neq \varnothing, \qquad \forall\, x \in \Re^n.$$

Solution: It will suffice to prove the result for the case where $f = f_1 + f_2$. If $d_1 \in \partial f_1(x)$ and $d_2 \in \partial f_2(x)$, then by the subgradient inequality, it follows that

$$f_1(z) \geq f_1(x) + (z-x)'d_1, \qquad \forall\, z,$$

$$f_2(z) \geq f_2(x) + (z-x)'d_2, \qquad \forall\, z,$$

so by adding these inequalities, we obtain

$$f(z) \geq f(x) + (z-x)'(d_1 + d_2), \qquad \forall\, z.$$

Hence, $d_1 + d_2 \in \partial f(x)$, implying that $\partial f_1(x) + \partial f_2(x) \subset \partial f(x)$.

Assuming that $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ and $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ have a point in common, we will prove the reverse inclusion. Let $d \in \partial f(x)$, and define the functions

$$g_1(y) = f_1(x+y) - f_1(x) - d'y, \qquad \forall\, y,$$

$$g_2(y) = f_2(x+y) - f_2(x), \qquad \forall\, y.$$

Then, for the function $g = g_1 + g_2$, we have $g(0) = 0$ and by using $d \in \partial f(x)$, we obtain

$$g(y) = f(x+y) - f(x) - d'y \geq 0, \qquad \forall\, y. \tag{4.5}$$

Consider the convex sets

$$C_1 = \big\{(y,\mu) \in \Re^{n+1} \mid y \in \mathrm{dom}(g_1),\ \mu \geq g_1(y)\big\},$$

$$C_2 = \big\{(u,\nu) \in \Re^{n+1} \mid u \in \mathrm{dom}(g_2),\ \nu \leq -g_2(y)\big\},$$

and note that

$$\mathrm{ri}(C_1) = \Big\{(y,\mu) \in \Re^{n+1} \mid y \in \mathrm{ri}\big(\mathrm{dom}(g_1)\big),\ \mu > g_1(y)\Big\},$$

$$\mathrm{ri}(C_2) = \Big\{(u,\nu) \in \Re^{n+1} \mid u \in \mathrm{ri}\big(\mathrm{dom}(g_2)\big),\ \nu < -g_2(y)\Big\}.$$

Suppose that there exists a vector $(\hat{y},\hat{\mu}) \in \mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$. Then,

$$g_1(\hat{y}) < \hat{\mu} < -g_2(\hat{y}),$$

yielding

$$g(\hat{y}) = g_1(\hat{y}) + g_2(\hat{y}) < 0,$$

which contradicts Eq. (4.5). Therefore, the sets $\mathrm{ri}(C_1)$ and $\mathrm{ri}(C_2)$ are disjoint, and by the Proper Separation (Prop. 2.4.6), the two convex sets C_1 and C_2 can be properly separated, i.e., there exists a vector $(w,\gamma) \in \Re^{n+1}$ such that

$$\inf_{(y,\mu)\in C_1} \{w'y + \gamma\mu\} \geq \sup_{(u,\nu)\in C_2} \{w'u + \gamma\nu\}, \tag{4.6}$$

$$\sup_{(y,\mu)\in C_1} \{w'y + \gamma\mu\} > \inf_{(u,\nu)\in C_2} \{w'u + \gamma\nu\}.$$

We cannot have $\gamma < 0$, because by letting $\mu \to \infty$ in Eq. (4.6), we will obtain a contradiction. Thus, we must have $\gamma \geq 0$. If $\gamma = 0$, then the preceding relations reduce to

$$\inf_{y\in \mathrm{dom}(g_1)} w'y \geq \sup_{u\in \mathrm{dom}(g_2)} w'u,$$

$$\sup_{y\in \mathrm{dom}(g_1)} w'y > \inf_{u\in \mathrm{dom}(g_2)} w'u,$$

which in view of the fact

$$\mathrm{dom}(g_1) = \mathrm{dom}(f_1) - x, \qquad \mathrm{dom}(g_2) = \mathrm{dom}(f_2) - x,$$

imply that $\mathrm{dom}(f_1)$ and $\mathrm{dom}(f_2)$ are properly separated. But this is impossible since $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ and $\mathrm{ri}\big(\mathrm{dom}(f_1)\big)$ have a point in common. Hence $\gamma > 0$, and by dividing in Eq. (4.6) with γ and by setting $b = w/\gamma$, we obtain

$$\inf_{(y,\mu)\in C_1} \big\{b'y + \mu\big\} \geq \sup_{(u,\nu)\in C_2} \big\{b'u + \nu\big\}.$$

Since $g_1(0) = 0$ and $g_2(0) = 0$, we have $(0,0) \in C_1 \cap C_2$, implying that

$$b'y + \mu \geq 0 \geq b'u + \nu, \qquad \forall\ (y,\mu) \in C_1, \quad \forall\ (u,\nu) \in C_2.$$

Therefore, for $\mu = g_1(y)$ and $\nu = -g_2(u)$, we obtain

$$g_1(y) \geq -b'y, \qquad \forall\ y \in \mathrm{dom}(g_1),$$

$$g_2(u) \geq b'\nu, \qquad \forall\ u \in \mathrm{dom}(g_2),$$

and by using the definitions of g_1 and g_2, we see that

$$f_1(x+y) \geq f_1(x) + (d-b)'y, \qquad \text{for all } y \text{ with } x+y \in \mathrm{dom}(f_1),$$

$$f_2(x+u) \geq f_2(x) + b'u, \qquad \text{for all } u \text{ with } x+u \in \mathrm{dom}(f_2).$$

Hence,

$$f_1(z) \geq f_1(x) + (d-b)'(z-x), \qquad \forall\ z,$$

$$f_2(z) \geq f_2(x) + b'(z-x), \qquad \forall\ z,$$

so that $d - b \in \partial f_1(x)$ and $b \in \partial f_2(x)$, showing that $d \in \partial f_1(x) + \partial f_2(x)$ and $\partial f(x) \subset \partial f_1(x) + \partial f_2(x)$.

When some of the functions are polyhedral, we use a different separation argument for C_1 and C_2. In particular, since the sum of polyhedral functions is a polyhedral function (see Exercise 3.12), it will still suffice to consider the case $m = 2$. Thus, let f_1 be a convex function, and let f_2 be a polyhedral function such that

$$\mathrm{ri}\big(\mathrm{dom}(f_1)\big) \cap \mathrm{dom}(f_2) \neq \emptyset.$$

Then, in the preceding proof, g_2 is a polyhedral function and C_2 is a polyhedral set. Furthermore, $\mathrm{ri}(C_1)$ and C_2 are disjoint, for otherwise we would have for some $(\hat{y},\hat{\mu}) \in \mathrm{ri}(C_1) \cap C_2$,

$$g_1(\hat{y}) < \hat{\mu} \leq -g_2(\hat{y}),$$

implying that $g(\hat{y}) = g_1(\hat{y}) + g_2(\hat{y}) < 0$ and contradicting Eq. (4.5). Therefore, by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), the convex set C_1 and the polyhedral set C_2 can be properly separated by a hyperplane that does not contain C_1, i.e., there exists a vector $(w,\gamma) \in \Re^{n+1}$ such that

$$\inf_{(y,\mu)\in C_1} \big\{w'y + \gamma\mu\big\} \geq \sup_{(u,\nu)\in C_2} \big\{w'u + \gamma\nu\big\},$$

$$\inf_{(y,\mu)\in C_1}\big\{w'y+\gamma\mu\big\} < \sup_{(y,\mu)\in C_1}\big\{w'y+\gamma\mu\big\}.$$

We cannot have $\gamma < 0$, because by letting $\mu \to \infty$ in the first of the preceding relations, we will obtain a contradiction. Thus, we must have $\gamma \geq 0$. If $\gamma = 0$, then the preceding relations reduce to

$$\inf_{y\in \mathrm{dom}(g_1)} w'y \geq \sup_{u\in \mathrm{dom}(g_2)} w'u,$$

$$\inf_{y\in \mathrm{dom}(g_1)} w'y < \sup_{y\in \mathrm{dom}(g_1)} w'y.$$

In view of the fact

$$\mathrm{dom}(g_1) = \mathrm{dom}(f_1) - x, \qquad \mathrm{dom}(g_2) = \mathrm{dom}(f_2) - x,$$

it follows that $\mathrm{dom}(f_1)$ and $\mathrm{dom}(f_2)$ are properly separated by a hyperplane that does not contain $\mathrm{dom}(f_1)$, while $\mathrm{dom}(f_2)$ is polyhedral since f_2 is polyhedral [see Prop. 3.2.3). Therefore, by the Polyhedral Proper Separation Theorem (Prop. 3.5.1), we have that $\mathrm{ri}\big(\mathrm{dom}(f_1)\big) \cap \mathrm{dom}(f_2) = \emptyset$, which is a contradiction. Hence $\gamma > 0$, and the remainder of the proof is similar to the preceding one.

4.10 (Chain Rule for Extended Real-Valued Functions)

This exercise is a refinement of Prop. 4.2.5(a). Let $f : \Re^m \mapsto (-\infty, \infty]$ be a convex function, and let A be an $m \times n$ matrix. Assume that the range of A contains a point in the relative interior of $\mathrm{dom}(f)$. Then, the subdifferential of the function F, defined by

$$F(x) = f(Ax),$$

is given by

$$\partial F(x) = A'\partial f(Ax).$$

Solution: We note that $\mathrm{dom}(F)$ is nonempty since in contains the inverse image under A of the common point of the range of A and the relative interior of $\mathrm{dom}(f)$. In particular, F is proper. We fix an x in $\mathrm{dom}(F)$. If $d \in A'\partial f(Ax)$, there exists a $g \in \partial f(Ax)$ such that $d = A'g$. We have for all $z \in \Re^m$,

$$\begin{aligned} F(z) - F(x) - (z-x)'d &= f(Az) - f(Ax) - (z-x)'A'g \\ &= f(Az) - f(Ax) - (Az - Ax)'g \\ &\geq 0, \end{aligned}$$

where the inequality follows from the fact $g \in \partial f(Ax)$. Hence, $d \in \partial F(x)$, and we have $A'\partial f(Ax) \subset \partial F(x)$.

We next show the reverse inclusion. By using a translation argument if necessary, we may assume that $x = 0$ and $F(0) = 0$. Let $d \in \partial F(0)$. Then we have

$$F(z) - z'd \geq 0, \qquad \forall\ z \in \Re^n,$$

or

$$f(Az) - z'd \geq 0, \qquad \forall\, z \in \Re^n,$$

or

$$f(y) - z'd \geq 0, \qquad \forall\, z \in \Re^n,\ y = Az,$$

or

$$H(y,z) \geq 0, \qquad \forall\, z \in \Re^n,\ y = Az,$$

where the function $H : \Re^m \times \Re^n \mapsto (-\infty, \infty]$ has the form

$$H(y,z) = f(y) - z'd.$$

Since the range of A contains a point in $\text{ri}\big(\text{dom}(f)\big)$, and $\text{dom}(H) = \text{dom}(f) \times \Re^n$, we see that the set $\big\{(y,z) \in \text{dom}(H) \mid y = Az\big\}$ contains a point in the relative interior of $\text{dom}(H)$. Hence, we can apply the Nonlinear Farkas' Lemma [part (b)] with the following identification:

$$x = (y,z), \qquad C = \text{dom}(H), \qquad g_1(y,z) = Az - y, \qquad g_2(y,z) = y - Az.$$

In this case, we have

$$\big\{x \in C \mid g_1(x) \leq 0,\ g_2(x) \leq 0\big\} = \big\{(y,z) \in \text{dom}(H) \mid Az - y = 0\big\}.$$

As asserted earlier, this set contains a relative interior point of C, thus implying that the set

$$Q^* = \big\{\mu \geq 0 \mid H(y,z) + \mu_1' g_1(y,z) + \mu_2' g_2(y,z) \geq 0,\ \forall\, (y,z) \in \text{dom}(H)\big\}$$

is nonempty. Hence, there exists (μ_1, μ_2) such that

$$f(y) - z'd + (\mu_1 - \mu_2)'(Az - y) \geq 0, \qquad \forall\, (y,z) \in \text{dom}(H).$$

Since $\text{dom}(H) = \Re^m \times \Re^n$, by letting $\lambda = \mu_1 - \mu_2$, we obtain

$$f(y) - z'd + \lambda'(Az - y) \geq 0, \qquad \forall\, y \in \Re^m,\ z \in \Re^n,$$

or equivalently

$$f(y) \geq \lambda' y + z'(d - A'\lambda), \qquad \forall\, y \in \Re^m,\ z \in \Re^n.$$

Because this relation holds for all z, we have $d = A'\lambda$ implying that

$$f(y) \geq \lambda' y, \qquad \forall\, y \in \Re^m,$$

which shows that $\lambda \in \partial f(0)$. Hence $d \in A'\partial f(0)$, thus completing the proof.

4.11

Let $f : \Re^n \mapsto \Re$ be a convex function, and let X be a bounded subset of $\Re^n$. Show that for all $\epsilon > 0$, the set $\cup_{x \in X} \partial_\epsilon f(x)$ is bounded.

Solution: Suppose that the set $\cup_{x \in X} \partial_\epsilon f(x)$ is unbounded for some $\epsilon > 0$. Then, there exist a sequence $\{x_k\} \subset X$, and a sequence $\{d_k\}$ such that $d_k \in \partial_\epsilon f(x_k)$ for all k and $\|d_k\| \to \infty$. Without loss of generality, we may assume that $d_k \neq 0$ for all k, and we denote $y_k = d_k/\|d_k\|$. Since both $\{x_k\}$ and $\{y_k\}$ are bounded, they must contain convergent subsequences, and without loss of generality, we may assume that x_k converges to some x and y_k converges to some y with $\|y\| = 1$. Since $d_k \in \partial_\epsilon f(x_k)$ for all k, it follows that

$$f(x_k + y_k) \geq f(x_k) + d_k' y_k - \epsilon = f(x_k) + \|d_k\| - \epsilon.$$

By letting $k \to \infty$ and by using the continuity of f, we obtain $f(x + y) = \infty$, a contradiction. Hence, the set $\cup_{x \in X} \partial_\epsilon f(x)$ must be bounded for all $\epsilon > 0$.

4.12

Let $f : \Re^n \mapsto \Re$ be a convex function. Show that for all $x \in \Re^n$, we have

$$\cap_{\epsilon > 0} \partial_\epsilon f(x) = \partial f(x).$$

Solution: Let $d \in \partial f(x)$. Then, by the definitions of subgradient and ϵ-subgradient, it follows that for any $\epsilon > 0$,

$$f(y) \geq f(x) + d'(y - x) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\, y \in \Re^n,$$

implying that $d \in \partial_\epsilon f(x)$ for all $\epsilon > 0$. Therefore $d \in \cap_{\epsilon > 0} \partial_\epsilon f(x)$, showing that $\partial f(x) \subset \cap_{\epsilon > 0} \partial_\epsilon f(x)$.

Conversely, let $d \in \partial_\epsilon f(x)$ for all $\epsilon > 0$, so that

$$f(y) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\, y \in \Re^n, \quad \forall\, \epsilon > 0.$$

By letting $\epsilon \downarrow 0$, we obtain

$$f(y) \geq f(x) + d'(y - x), \qquad \forall\, y \in \Re^n,$$

implying that $d \in \partial f(x)$, and showing that $\cap_{\epsilon > 0} \partial_\epsilon f(x) \subset \partial f(x)$.

4.13 (Continuity Properties of ϵ-Subdifferential [Nur77])

Let $f : \Re^n \mapsto \Re$ be a convex function and let ϵ be a positive scalar. Show that for every $x \in \Re^n$, the following hold:

(a) If a sequence $\{x_k\}$ converges to x and $d_k \in \partial_\epsilon f(x_k)$ for all k, then the sequence $\{d_k\}$ is bounded and each of its limit points is an ϵ-subgradient of f at x.

(b) If $d \in \partial_\epsilon f(x)$, then for every sequence $\{x_k\}$ converging to x, there exists a subsequence $\{d_k\}_\mathcal{K}$ converging to d with $d_k \in \partial_\epsilon f(x_k)$ for all $k \in \mathcal{K}$.

Solution: (a) By the ϵ-subgradient definition, we have for all k,

$$f(y) \geq f(x) + d_k'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n.$$

Since the sequence $\{x_k\}$ is bounded, by Exercise 4.11, the sequence $\{d_k\}$ is also bounded and therefore, it has a limit point d. Taking the limit in the preceding relation along a subsequence of $\{d_k\}$ converging to d, we obtain

$$f(y) \geq f(x) + d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n,$$

showing that $d \in \partial_\epsilon f(x)$.

(b) First we show that

$$\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big) = \partial_\epsilon f(x). \tag{4.7}$$

Let $d \in \partial_\delta f(x)$ for some scalar δ satisfying $0 < \delta < \epsilon$. Then, by the definition of ϵ-subgradient, we have

$$f(y) \geq f(x) - d'(y - x) - \delta \geq f(x) - d'(y - x) - \epsilon, \qquad \forall\ y \in \Re^n,$$

showing that $d \in \partial_\epsilon f(x)$. Therefore,

$$\partial_\delta f(x) \subset \partial_\epsilon f(x), \qquad \forall\ \delta \in (0, \epsilon), \tag{4.8}$$

implying that

$$\cup_{0<\delta<\epsilon} \partial_\delta f(x) \subset \partial_\epsilon f(x).$$

Since $\partial_\epsilon f(x)$ is closed, by taking the closures of both sides in the preceding relation, we obtain

$$\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big) \subset \partial_\epsilon f(x).$$

Conversely, assume to arrive at a contradiction that there is a vector $d \in \partial_\epsilon f(x)$ with $d \notin \mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big)$. Note that the set $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is bounded since it is contained in the compact set $\partial_\epsilon f(x)$. Furthermore, we claim that $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is convex. Indeed if d_1 and d_2 belong to this set, then $d_1 \in \partial_{\delta_1} f(x)$ and $d_2 \in \partial_{\delta_2} f(x)$ for some positive scalars δ_1 and δ_2. Without loss of generality, let $\delta_1 \leq \delta_2$. Then, by Eq. (4.8), it follows that $d_1, d_2 \in \partial_{\delta_2} f(x)$, which is a convex set by Prop. 4.3.1(a). Hence, $\lambda d_1 + (1 - \lambda) d_2 \in \partial_{\delta_2} f(x)$ for all $\lambda \in [0, 1]$, implying that $\lambda d_1 + (1-\lambda)d_2 \in \cup_{0<\delta<\epsilon} \partial_\delta f(x)$ for all $\lambda \in [0,1]$, and showing that the set $\cup_{0<\delta<\epsilon} \partial_\delta f(x)$ is convex.

The vector d and the convex and compact set $\mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big)$ can be strongly separated (see Exercise 2.17), i.e., there exists a vector $b \in \Re^n$ such that

$$b'd > \max_{g \in \mathrm{cl}\left(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\right)} b'g.$$

This relation implies that for some positive scalar β,

$$b'd > \max_{g \in \partial_\delta f(x)} b'g + 2\beta, \qquad \forall\ \delta \in (0,\epsilon).$$

By Prop. 4.3.1(a), we have

$$\inf_{\alpha>0} \frac{f(x+\alpha b) - f(x) + \delta}{\alpha} = \max_{g \in \partial_\delta f(x)} b'g,$$

so that

$$b'd > \inf_{\alpha>0} \frac{f(x+\alpha b) - f(x) + \delta}{\alpha} + 2\beta, \qquad \forall\ \delta,\ 0<\delta<\epsilon.$$

Let $\{\delta_k\}$ be a positive scalar sequence converging to ϵ. In view of the preceding relation, for each δ_k, there exists a small enough $\alpha_k > 0$ such that

$$\alpha_k b'd \geq f(x+\alpha_k b) - f(x) + \delta_k + \beta. \tag{4.9}$$

Without loss of generality, we may assume that $\{\alpha_k\}$ is bounded, so that it has a limit point $\overline{\alpha} \geq 0$. By taking the limit in Eq. (4.9) along an appropriate subsequence, and by using $\delta_k \to \epsilon$, we obtain

$$\overline{\alpha} b'd \geq f(x+\overline{\alpha} b) - f(x) + \epsilon + \beta.$$

If $\overline{\alpha} = 0$, we would have $0 \geq \epsilon + \beta$, which is a contradiction. If $\overline{\alpha} > 0$, we would have

$$\overline{\alpha} b'd + f(x) - \epsilon > f(x+\overline{\alpha} b),$$

which cannot hold since $d \in \partial_\epsilon f(x)$. Hence, we must have

$$\partial_\epsilon f(x) \subset \mathrm{cl}\big(\cup_{0<\delta<\epsilon} \partial_\delta f(x)\big),$$

thus completing the proof of Eq. (4.7).

We now prove the statement of the exercise. Let $\{x_k\}$ be a sequence converging to x. By Prop. 4.3.1(a), the ϵ-subdifferential $\partial_\epsilon f(x)$ is bounded, so that there exists a constant $L > 0$ such that

$$\|g\| \leq L, \qquad \forall\ g \in \partial_\epsilon f(x).$$

Let

$$\gamma_k = \big|f(x_k) - f(x)\big| + L\ \|x_k - x\|, \qquad \forall\ k. \tag{4.10}$$

Since $x_k \to x$, by continuity of f, it follows that $\gamma_k \to 0$ as $k \to \infty$, so that $\epsilon_k = \epsilon - \gamma_k$ converges to ϵ. Let $\{k_i\} \subset \{0, 1, \ldots\}$ be an index sequence such that $\{\epsilon_{k_i}\}$ is positive and monotonically increasing to ϵ, i.e.,

$$\epsilon_{k_i} \uparrow \epsilon \qquad \text{with} \quad \epsilon_{k_i} = \epsilon - \gamma_{k_i} > 0, \quad \epsilon_{k_i} < \epsilon_{k_{i+1}}, \quad \forall\ i.$$

In view of relation (4.7), we have

$$\text{cl}\big(\cup_{i\geq 0}\ \partial_{\epsilon_{k_i}} f(x)\big) = \partial_\epsilon f(x), \tag{4.11}$$

implying that for a given vector $d \in \partial_\epsilon f(x)$, there exists a sequence $\{d_{k_i}\}$ such that

$$d_{k_i} \to d \qquad \text{with} \quad d_{k_i} \in \partial_{\epsilon_{k_i}} f(x), \quad \forall\ i. \tag{4.12}$$

There remains to show that $d_{k_i} \in \partial_\epsilon f(x_{k_i})$ for all i. Since $d_{k_i} \in \partial_{\epsilon_{k_i}} f(x)$, it follows that for all i and $y \in \Re^n$,

$$\begin{aligned} f(y) &\geq f(x) + d_{k_i}'(y - x) - \epsilon_{k_i} \\ &= f(x_{k_i}) + \big(f(x) - f(x_{k_i})\big) + d_{k_i}'(y - x_{k_i}) + d_{k_i}'(x_{k_i} - x) - \epsilon_{k_i} \\ &\geq f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - \Big(\big|f(x) - f(x_{k_i})\big| + \big|d_{k_i}'(x_{k_i} - x)\big| + \epsilon_{k_i}\Big). \end{aligned} \tag{4.13}$$

Because $d_{k_i} \in \partial_\epsilon f(x)$ [cf. Eqs. (4.11) and (4.12)] and $\partial_\epsilon f(x)$ is bounded, there holds

$$\big|d_{k_i}'(x_{k_i} - x)\big| \leq L\ \|x_{k_i} - x\|.$$

Using this relation, the definition of γ_k [cf. Eq. (4.10)], and the fact $\epsilon_k = \epsilon - \gamma_k$ for all k, from Eq. (4.13) we obtain for all i and $y \in \Re^n$,

$$f(y) \geq f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - (\gamma_{k_i} + \epsilon_{k_i}) = f(x_{k_i}) + d_{k_i}'(y - x_{k_i}) - \epsilon.$$

Hence $d_{k_i} \in \partial_\epsilon f(x_{k_i})$ for all i, thus completing the proof.

4.14 (Subgradient Mean Value Theorem)

(a) *Scalar Case*: Let $\varphi : \Re \mapsto \Re$ be a scalar convex function, and let a and b be scalars with $a < b$. Show that there exists a scalar $t^* \in (a, b)$ such that

$$\frac{\varphi(b) - \varphi(a)}{b - a} \in \partial\varphi(t^*).$$

Hint: Show that the scalar convex function

$$g(t) = \varphi(t) - \varphi(a) - \frac{\varphi(b) - \varphi(a)}{b - a}(t - a)$$

has a minimum $t^* \in (a, b)$, and use the optimality condition $0 \in \partial g(t^*)$.

(b) *Vector Case*: Let $f : \Re^n \mapsto \Re$ be a convex function, and let x and y be vectors in $\Re^n$. Show that there exist a scalar $\alpha \in (0,1)$ and a subgradient $d \in \partial f\big(\alpha x + (1-\alpha)y\big)$ such that

$$f(y) = f(x) + d'(y-x).$$

Hint: Apply part (a) to the scalar function $\varphi(t) = f\big(tx + (1-t)y\big)$, $t \in \Re$.

Solution: (a) **Scalar Case:** Define the scalar function $g : \Re \mapsto \Re$ by

$$g(t) = \varphi(t) - \varphi(a) - \frac{\varphi(b) - \varphi(a)}{b-a}(t-a),$$

and note that g is convex and $g(a) = g(b) = 0$. We first show that g attains its minimum over $\Re$ at some point $t^* \in [a,b]$. For $t < a$, we have

$$a = \frac{b-a}{b-t}t + \frac{a-t}{b-t}b,$$

and by using convexity of g and $g(a) = g(b) = 0$, we obtain

$$0 = g(a) \leq \frac{b-a}{b-t}g(t) + \frac{a-t}{b-t}g(b) = \frac{b-a}{b-t}g(t),$$

implying that $g(t) \geq 0$ for $t < a$. Similarly, for $t > b$, we have

$$b = \frac{b-a}{t-a}t + \frac{t-b}{t-a}a,$$

and by using convexity of g and $g(a) = g(b) = 0$, we obtain

$$0 = g(b) \leq \frac{b-a}{t-a}g(t) + \frac{t-b}{t-a}g(a) = \frac{t-b}{t-a}g(t),$$

implying that $g(t) \geq 0$ for $t > b$. Therefore $g(t) \geq 0$ for $t \notin (a,b)$, while $g(a) = g(b) = 0$. Hence

$$\min_{t \in \Re} g(t) = \min_{t \in [a,b]} g(t). \tag{4.14}$$

Because g is convex over $\Re$, it is also continuous over $\Re$, and since $[a,b]$ is compact, the set of minimizers of g over $[a,b]$ is nonempty. Thus, in view of Eq. (4.14), there exists a scalar $t^* \in [a,b]$ such that $g(t^*) = \min_{t \in \Re} g(t)$. If $t^* \in (a,b)$, then we are done. If $t^* = a$ or $t^* = b$, then since $g(a) = g(b) = 0$, it follows that every $t \in [a,b]$ attains the minimum of g over $\Re$, so that we can replace t^* by a point in the interval (a,b). Thus, in any case, there exists $t^* \in (a,b)$ such that $g(t^*) = \min_{t \in \Re} g(t)$.

We next show that

$$\frac{\varphi(b) - \varphi(a)}{b-a} \in \partial\varphi(t^*).$$

The function g is the sum of the convex function φ and the linear (and therefore smooth) function $-\frac{\varphi(b)-\varphi(a)}{b-a}(t-a)$. Thus the subdifferential of $\partial g(t^*)$ is the sum of the sudifferential of $\partial\varphi(t^*)$ and the gradient $-\frac{\varphi(b)-\varphi(a)}{b-a}$ (see Prop. 4.2.4),

$$\partial g(t^*) = \partial\varphi(t^*) - \frac{\varphi(b)-\varphi(a)}{b-a}.$$

Since t^* minimizes g over $\Re$, by the optimality condition, we have $0 \in \partial g(t^*)$. This and the preceding relation imply that

$$\frac{\varphi(b)-\varphi(a)}{b-a} \in \partial\varphi(t^*).$$

(b) **Vector Case:** Let x and y be any two vectors in $\Re^n$. If $x = y$, then $f(y) = f(x) + d'(y-x)$ trivially holds for any $d \in \partial f(x)$, and we are done. So assume that $x \neq y$, and consider the scalar function φ given by

$$\varphi(t) = f(x_t), \qquad x_t = tx + (1-t)y, \qquad t \in \Re.$$

By part (a), where $a = 0$ and $b = 1$, there exists $\alpha \in (0,1)$ such that

$$\varphi(1) - \varphi(0) \in \partial\varphi(\alpha),$$

while by Exercise 4.6, we have

$$\partial\varphi(\alpha) = \big\{d'(x-y) \mid d \in \partial f(x_\alpha)\big\}.$$

Since $\varphi(1) = f(x)$ and $\varphi(0) = f(y)$, we see that

$$f(x) - f(y) \in \big\{d'(x-y) \mid d \in \partial f(x_\alpha)\big\}.$$

Therefore, there exists $d \in \partial f(x_\alpha)$ such that $f(y) - f(x) = d'(y-x)$.

4.15 (Steepest Descent Direction of a Convex Function)

Let $f : \Re^n \mapsto \Re$ be a convex function and let x be a vector in $\Re^n$. Show that a vector $\bar{d}$ is the vector of minimum norm in $\partial f(x)$ if and only if either $\bar{d} = 0$ or else $\bar{d}/\|\bar{d}\|$ minimizes $f'(x;d)$ over all d with $\|d\| \leq 1$.

Solution: Note that the problem statement in the book contains a typo: $\bar{d}/\|\bar{d}\|$ should be replaced by $-\bar{d}/\|\bar{d}\|$.

The sets $\big\{d \mid \|d\| \leq 1\big\}$ and $\partial f(x)$ are compact, and the function $\phi(d,g) = d'g$ is linear in each variable when the other variable is fixed, so that $\phi(\cdot,g)$ is convex and closed for all g, while the function $-\phi(d,\cdot)$ is convex and closed for all d. Thus, by Prop. 2.6.9, the order of min and max can be interchanged,

$$\min_{\|d\|\leq 1} \max_{g\in\partial f(x)} d'g = \max_{g\in\partial f(x)} \min_{\|d\|\leq 1} d'g,$$

and there exist associated saddle points.

By Prop. 4.2.2, we have $f'(x;d) = \max_{g\in\partial f(x)} d'g$, so

$$\min_{\|d\|\le 1} \max_{g\in\partial f(x)} d'g = \min_{\|d\|\le 1} f'(x;d). \tag{4.15}$$

We also have for all g,

$$\min_{\|d\|\le 1} d'g = -\|g\|,$$

and the minimum is attained for $d = -g/\|g\|$. Thus

$$\max_{g\in\partial f(x)} \min_{\|d\|\le 1} d'g = \max_{g\in\partial f(x)} \big(-\|g\|\big) = -\min_{g\in\partial f(x)} \|g\|. \tag{4.16}$$

From the generic characterization of a saddle point (cf. Prop. 2.6.1), it follows that the set of saddle points of $d'g$ is $D^* \times G^*$, where D^* is the set of minima of $f'(x;d)$ subject to $\|d\| \le 1$ [cf. Eq. (4.15)], and G^* is the set of minima of $\|g\|$ subject to $g \in \partial f(x)$ [cf. Eq. (4.16)], i.e., G^* consists of the unique vector g^* of minimum norm on $\partial f(x)$. Furthermore, again by Prop. 2.6.1, every $d^* \in D^*$ must minimize $d'g^*$ subject to $\|d\| \le 1$, so it must satisfy $d^* = -g^*/\|g^*\|$.

4.16 (Generating Descent Directions of Convex Functions)

This exercise provides a method for generating a descent direction in circumstances where obtaining a single subgradient is relatively easy.

Let $f : \Re^n \mapsto \Re$ be a convex function, and let x be a fixed vector in $\Re^n$. A vector $d \in \Re^n$ is said to be a *descent direction* of f at x if the corresponding directional derivative of f satisfies

$$f'(x;d) < 0.$$

Assume that x does not minimize f, and let g_1 be a subgradient of f at x. For $k = 2, 3, \ldots$, let w_k be the vector of minimum norm in the convex hull of $g_1, \ldots, g_{k-1}$,

$$w_k = \arg \min_{g\in \mathrm{conv}\{g_1,\ldots,g_{k-1}\}} \|g\|.$$

If $-w_k$ is a descent direction of f at x, then stop; else let g_k be a vector in $\partial f(x)$ such that

$$g_k{}'w_k = \min_{g\in\partial f(x)} g'w_k.$$

Show that this process terminates in a finite number of steps with a descent direction of f at x. *Hint*: If $-w_k$ is not a descent direction, then $g_i{}'w_k \ge \|w_k\|^2 \ge \|g^*\|^2 > 0$ for all $i = 1, \ldots, k-1$, where g^* is the subgradient of f at x with minimum norm, while at the same time $g_k{}'w_k \le 0$. Consider a limit point of $\big\{(w_k, g_k)\big\}$.

Solution: Suppose that the process does not terminate in a finite number of steps, and let $\big\{(w_k, g_k)\big\}$ be the sequence generated by the algorithm. Since w_k

is the projection of the origin on the set $\text{conv}\{g_1, \ldots, g_{k-1}\}$, by the Projection Theorem (Prop. 2.2.1), we have

$$(g - w_k)' w_k \geq 0, \qquad \forall\, g \in \text{conv}\{g_1, \ldots, g_{k-1}\},$$

implying that

$$g_i' w_k \geq \|w_k\|^2 \geq \|g^*\|^2 > 0, \qquad \forall\, i = 1, \ldots, k-1, \quad \forall\, k \geq 1, \tag{4.17}$$

where $g^* \in \partial f(x)$ is the vector with minimum norm in $\partial f(x)$. Note that $\|g^*\| > 0$ because x does not minimize f. The sequences $\{w_k\}$ and $\{g_k\}$ are contained in $\partial f(x)$, and since $\partial f(x)$ is compact, $\{w_k\}$ and $\{g_k\}$ have limit points in $\partial f(x)$. Without loss of generality, we may assume that these sequences converge, so that for some $\hat{w}, \hat{g} \in \partial f(x)$, we have

$$\lim_{k \to \infty} w_k = \hat{w}, \qquad \lim_{k \to \infty} g_k = \hat{g},$$

which in view of Eq. (4.17) implies that $\hat{g}' \hat{w} > 0$. On the other hand, because none of the vectors $(-w_k)$ is a descent direction of f at x, we have $f'(x; -w_k) \geq 0$, so that

$$g_k'(-w_k) = \max_{g \in \partial f(x)} g'(-w_k) = f'(x; -w_k) \geq 0.$$

By letting $k \to \infty$, we obtain $\hat{g}' \hat{w} \leq 0$, thus contradicting $\hat{g}' \hat{w} > 0$. Therefore, the process must terminate in a finite number of steps with a descent direction.

4.17 (Generating ϵ-Descent Directions of Convex Functions [Lem74])

This exercise shows how the procedure of Exercise 4.16 can be modified so that it generates an ϵ-descent direction.

Let $f : \Re^n \mapsto \Re$ be a convex function, and let x be a fixed vector in $\Re^n$. Assume that x is not an ϵ-minimizer of the function f, i.e., $f(x) > \inf_{z \in \Re^n} f(z) + \epsilon$, and let g_1 be an ϵ-subgradient of f at x. For $k = 2, 3, \ldots$, let w_k be the vector of minimum norm in the convex hull of $g_1, \ldots, g_{k-1}$,

$$w_k = \arg \min_{g \in \text{conv}\{g_1, \ldots, g_{k-1}\}} \|g\|.$$

By a search along the direction $-w_k$, determine whether there exists a scalar $\overline{\alpha}$ such that $f(x - \overline{\alpha} w_k) < f(x) - \epsilon$. If such an $\overline{\alpha}$ exists, then stop ($-w_k$ is an ϵ-descent direction of f at x); otherwise, let g_k be a vector in $\partial_\epsilon f(x)$ such that

$${g_k}' w_k = \min_{g \in \partial_\epsilon f(x)} g' w_k.$$

Show that this process will terminate in a finite number of steps with an ϵ-descent direction of f at x.

Solution: Suppose that the process does not terminate in a finite number of steps, and let $\big\{(w_k, g_k)\big\}$ be the sequence generated by the algorithm. Since w_k

is the projection of the origin on the set $\text{conv}\{g_1, \ldots, g_{k-1}\}$, by the Projection Theorem (Prop. 2.2.1), we have

$$(g - w_k)' w_k \geq 0, \qquad \forall\, g \in \text{conv}\{g_1, \ldots, g_{k-1}\},$$

implying that

$$g_i' w_k \geq \|w_k\|^2 \geq \|g^*\|^2 > 0, \qquad \forall\, i = 1, \ldots, k-1, \quad \forall\, k \geq 1, \tag{4.18}$$

where $g^* \in \partial_\epsilon f(x)$ is the vector with minimum norm in $\partial_\epsilon f(x)$. Note that $\|g^*\| > 0$ because x is not an ϵ-optimal solution, i.e., $f(x) > \inf_{z \in \Re^n} f(z) + \epsilon$ [see Prop. 4.3.1(b)]. The sequences $\{w_k\}$ and $\{g_k\}$ are contained in $\partial_\epsilon f(x)$, and since $\partial_\epsilon f(x)$ is compact [Prop. 4.3.1(a)], $\{w_k\}$ and $\{g_k\}$ have limit points in $\partial_\epsilon f(x)$. Without loss of generality, we may assume that these sequences converge, so that for some $\hat{w}, \hat{g} \in \partial_\epsilon f(x)$, we have

$$\lim_{k \to \infty} w_k = \hat{w}, \qquad \lim_{k \to \infty} g_k = \hat{g},$$

which in view of Eq. (4.18) implies that $\hat{g}'\hat{w} > 0$. On the other hand, because none of the vectors $(-w_k)$ is an ϵ-descent direction of f at x, by Prop. 4.3.1(a), we have

$$g_k'(-w_k) = \max_{g \in \partial f(x)} g'(-w_k) = \inf_{\alpha > 0} \frac{f(x - \alpha w_k) - f(x) + \epsilon}{\alpha} \geq 0.$$

By letting $k \to \infty$, we obtain $\hat{g}'\hat{w} \leq 0$, thus contradicting $\hat{g}'\hat{w} > 0$. Hence, the process must terminate in a finite number of steps with an ϵ-descent direction.

4.18

For the following subsets C of $\Re^n$, specify the tangent cone and the normal cone at every point of C.

(a) C is the unit ball.

(b) C is a subspace.

(c) C is a closed halfspace, i.e., $C = \{x \mid a'x \leq b\}$ for a nonzero vector $a \in \Re^n$ and a scalar b.

(d) $C = \{x \mid x_i \geq 0,\ i \in I\}$ with $I \subset \{1, \ldots, n\}$.

Solution: (a) For $x \in \text{int}(C)$, we clearly have $F_C(x) = \Re^n$, implying that

$$T_C(x) = \Re^n.$$

Since C is convex, by Prop. 4.6.3, we have

$$N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \text{int}(C)$, we have $\|x\| = 1$. By the definition of the set $F_C(x)$ of feasible directions at x, we have $y \in F_C(x)$ if and only if $x + \alpha y \in C$ for all sufficiently small positive scalars α. Thus, $y \in F_C(x)$ if and only if there exists $\overline{\alpha} > 0$ such that $\|x + \alpha y\|^2 \leq 1$ for all α with $0 < \alpha \leq \overline{\alpha}$, or equivalently

$$\|x\|^2 + 2\alpha x'y + \alpha^2 \|y\|^2 \leq 1, \qquad \forall\ \alpha,\ 0 < \alpha \leq \overline{\alpha}.$$

Since $\|x\| = 1$, the preceding relation reduces to

$$2x'y + \alpha\|y\|^2 \leq 0. \qquad \forall\ \alpha,\ 0 < \alpha \leq \overline{\alpha}.$$

This relation holds if and only if $y = 0$, or $x'y < 0$ and $\alpha \leq -2x'y/\|y\|^2$ (i.e., $\overline{\alpha} = -2x'y/\|y\|^2$). Therefore,

$$F_C(x) = \big\{y \mid x'y < 0\big\} \cup \{0\}.$$

Because C is convex, by Prop. 4.6.2(c), we have $T_C(x) = \text{cl}\big(F_C(x)\big)$, implying that

$$T_C(x) = \big\{y \mid x'y \leq 0\big\}.$$

Furthermore, by Prop. 4.6.3, we have $N_C(x) = T_C(x)^*$, while by the Farkas' Lemma [Prop. 3.2.1(b)], $T_C(x)^* = \text{cone}\big(\{x\}\big)$, implying that

$$N_C(x) = \text{cone}\big(\{x\}\big).$$

(b) If C is a subspace, then clearly $F_C(x) = C$ for all $x \in C$. Because C is convex, by Props. 4.6.2(a) and 4.6.3, we have

$$T_C(x) = \text{cl}\big(F_C(x)\big) = C, \qquad N_C(x) = T_C(x)^* = C^\perp, \qquad \forall\ x \in C.$$

(c) Let C be a closed halfspace given by $C = \big\{x \mid a'x \leq b\big\}$ with a nonzero vector $a \in \Re^n$ and a scalar b. For $x \in \text{int}(C)$, i.e., $a'x < b$, we have $F_C(x) = \Re^n$ and since C is convex, by Props. 4.6.2(a) and 4.6.3, we have

$$T_C(x) = \text{cl}\big(F_C(x)\big) = \Re^n, \qquad N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \text{int}(C)$, we have $a'x = b$, so that $x + \alpha y \in C$ for some $y \in \Re^n$ and $\alpha > 0$ if and only if $a'y \leq 0$, implying that

$$F_C(x) = \big\{y \mid a'y \leq 0\big\}.$$

By Prop. 4.6.2(a), it follows that

$$T_C(x) = \text{cl}\big(F_C(x)\big) = \big\{y \mid a'y \leq 0\big\},$$

while by Prop. 4.6.3 and the Farkas' Lemma [Prop. 3.2.1(b)], it follows that

$$N_C(x) = T_C(x)^* = \text{cone}\big(\{a\}\big).$$

(d) For $x \in C$ with $x \in \text{int}(C)$, i.e., $x_i > 0$ for all $i \in I$, we have $F_C(x) = \Re^n$. Then, by using Props. 4.6.2(a) and 4.6.3, we obtain

$$T_C(x) = \text{cl}\big(F_C(x)\big) = \Re^n, \qquad N_C(x) = T_C(x)^* = \{0\}.$$

For $x \in C$ with $x \notin \text{int}(C)$, the set $A_x = \{i \in I \mid x_i = 0\}$ is nonempty. Then, $x + \alpha y \in C$ for some $y \in \Re^n$ and $\alpha > 0$ if and only if $y_i \leq 0$ for all $i \in A_x$, implying that

$$F_C(x) = \{y \mid y_i \leq 0, \ \ \forall\, i \in A_x\}.$$

Because C is convex, by Prop. 4.6.2(a), we have that

$$T_C(x) = \text{cl}\big(F_C(x)\big) = \{y \mid y_i \leq 0, \ \ \forall\, i \in A_x\},$$

or equivalently

$$T_C(x) = \big\{y \mid y'e_i \leq 0, \ \ \forall\, i \in A_x\big\},$$

where $e_i \in \Re^n$ is the vector whose ith component is 1 and all other components are 0. By Prop. 4.6.3, we further have $N_C(x) = T_C(x)^*$, while by the Farkas' Lemma [Prop. propforea(b)], we see that $T_C(x)^* = \text{cone}\big(\{e_i \mid i \in A_x\}\big)$, implying that

$$N_C(x) = \text{cone}\big(\{e_i \mid i \in A_x\}\big).$$

4.19

Let C be a convex subset of $\Re^n$, and let x be a vector in C. Show that the following properties are equivalent:

(a) x lies in the relative interior of C.

(b) $T_C(x)$ is a subspace.

(c) $N_C(x)$ is a subspace.

Solution: (a) $\Rightarrow$ (b) Let $x \in \text{ri}(C)$ and let S be the subspace that is parallel to $\text{aff}(C)$. Then, for every $y \in S$, $x + \alpha y \in \text{ri}(C)$ for all sufficiently small positive scalars α, implying that $y \in F_C(x)$ and showing that $S \subset F_C(x)$. Furthermore, by the definition of the set of feasible directions, it follows that if $y \in F_C(x)$, then there exists $\overline{\alpha} > 0$ such that $x + \alpha y \in C$ for all $\alpha \in (0, \overline{\alpha}]$. Hence $y \in S$, implying that $F_C(x) \subset S$. This and the relation $S \subset F_C(x)$ show that $F_C(x) = S$. Since C is convex, by Prop. 4.6.2(a), it follows that

$$T_C(x) = \text{cl}\big(F_C(x)\big) = S,$$

thus proving that $T_C(x)$ is a subspace.

(b) $\Rightarrow$ (c) Let $T_C(x)$ be a subspace. Then, because C is convex, from Prop. 4.6.3 it follows that

$$N_C(x) = T_C(x)^* = T_C(x)^{\perp},$$

showing that $N_C(x)$ is a subspace.

(c) $\Rightarrow$ (a) Let $N_C(x)$ be a subspace, and to arrive at a contradiction suppose that x is not a point in the relative interior of C. Then, by the Proper Separation Theorem (Prop. 2.4.5), the point x and the relative interior of C can be properly separated, i.e., there exists a vector $a \in \Re^n$ such that

$$\sup_{y \in C} a'y \leq a'x, \tag{4.19}$$

$$\inf_{y \in C} a'y < \sup_{y \in C} a'y. \tag{4.20}$$

The relation (4.19) implies that

$$(-a)'(x - y) \leq 0, \qquad \forall\, y \in C. \tag{4.21}$$

Since C is convex, by Prop. 4.6.3, the preceding relation is equivalent to $-a \in T_C(x)^*$. By the same proposition, there holds $N_C(x) = T_C(x)^*$, implying that $-a \in N_C(x)$. Because $N_C(x)$ is a subspace, we must also have $a \in N_C(x)$, and by using

$$N_C(x) = T_C(x)^* = \big\{z \mid z'(x - y) \leq 0, \ \ \forall\, y \in C\big\}$$

(cf. Prop. 4.6.3), we see that

$$a'(x - y) \leq 0, \qquad \forall\, y \in C.$$

This relation and Eq. (4.21) yield

$$a'(x - y) = 0, \qquad \forall\, y \in C,$$

contradicting Eq. (4.20). Hence, x must be in the relative interior of C.

4.20 (Tangent and Normal Cones of Affine Sets)

Let A be an $m \times n$ matrix and b be a vector in $\Re^m$. Show that the tangent cone and the normal cone of the set $\{x \mid Ax = b\}$ at any of its points are the null space of A and the range space of A', respectively.

Solution: Let $C = \{x \mid Ax = b\}$ and let $x \in C$ be arbitrary. We then have

$$F_C(x) = \{y \mid Ay = 0\} = N(A),$$

and by using Prop. 4.6.2(a), we obtain

$$T_C(x) = \mathrm{cl}\big(F_C(x)\big) = N(A).$$

Since C is convex, by Prop. 4.6.3, it follows that

$$N_C(x) = T_C(x)^* = N(A)^\perp = R(A').$$

4.21 (Tangent and Normal Cones of Level Sets)

Let $f : \Re^n \mapsto \Re$ be a convex function, and let x be a vector in $\Re^n$ such that the level set $\big\{z \mid f(z) \leq f(x)\big\}$ is nonempty. Show that the tangent cone and the normal cone of the level set $\big\{z \mid f(z) \leq f(x)\big\}$ at the point x coincide with $\big\{y \mid f'(x;y) \leq 0\big\}$ and $\mathrm{cl}\big(\mathrm{cone}\big(\partial f(x)\big)\big)$, respectively. Furthermore, if x does not minimize f over $\Re^n$, the closure operation is unnecessary.

Solution: Let

$$C = \big\{z \mid f(z) \leq f(x)\big\},$$

and consider $F_C(x)$, the cone of feasible directions of C at x. We first show that

$$\mathrm{cl}\big(F_C(x)\big) = \big\{y \mid f'(x;y) \leq 0\big\}.$$

Let $\overline{y} \in F_C(x)$ be arbitrary. Then, by the definition of $F_C(x)$, there exists a scalar $\overline{\alpha}$ such that $x + \alpha\overline{y} \in C$ for all $\alpha \in (0, \overline{\alpha}]$. By the definition of C, it follows that $f(x + \alpha\overline{y}) \leq f(x)$ for all $\alpha \in (0, \overline{\alpha}]$, implying that

$$f'(x;\overline{y}) = \inf_{\alpha>0} \frac{f(x+\alpha\overline{y}) - f(x)}{\alpha} \leq 0.$$

Therefore $\overline{y} \in \big\{y \mid f'(x;y) \leq 0\big\}$, thus showing that

$$F_C(x) \subset \big\{y \mid f'(x;y) \leq 0\big\}.$$

By Exercise 4.1(d), the set $\big\{y \mid f'(x;y) \leq 0\big\}$ is closed, so that by taking closures in the preceding relation, we obtain

$$\mathrm{cl}\big(F_C(x)\big) \subset \big\{y \mid f'(x;y) \leq 0\big\}.$$

To show the converse inclusion, let $\overline{y}$ be such that $f'(x;\overline{y}) < 0$, so that for all small enough $\alpha \geq 0$, we have

$$f(x + \alpha\overline{y}) - f(x) < 0.$$

Therefore $x + \alpha\overline{y} \in C$ for all small enough $\alpha \geq 0$, implying that $\overline{y} \in F_C(x)$ and showing that

$$\big\{y \mid f'(x;y) < 0\big\} \subset F_C(x).$$

By taking the closures of the sets in the preceding relation, we obtain

$$\big\{y \mid f'(x;y) \leq 0\big\} = \mathrm{cl}\Big(\big\{y \mid f'(x;y) < 0\big\}\Big) \subset \mathrm{cl}\big(F_C(x)\big).$$

Hence

$$\mathrm{cl}\big(F_C(x)\big) = \big\{y \mid f'(x;y) \leq 0\big\}.$$

Since C is convex, by Prop. 4.6.2(c), we have $\mathrm{cl}\big(F_C(x)\big) = T_C(x)$. This and the preceding relation imply that

$$T_C(x) = \big\{y \mid f'(x;y) \leq 0\big\}.$$

Since by Prop. 4.6.3, $N_C(x) = T_C(x)^*$, it follows that

$$N_C(x) = \Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^*.$$

Furthermore, by Exercise 4.1(d), we have that

$$\Big(\big\{y \mid f'(x;y) \le 0\big\}\Big)^* = \text{cl}\Big(\text{cone}\big(\partial f(x)\big)\Big),$$

implying that

$$N_C(x) = \text{cl}\Big(\text{cone}\big(\partial f(x)\big)\Big).$$

If x does not minimize f over $\Re^n$, then the subdifferential $\partial f(x)$ does not contain the origin. Furthermore, by Prop. 4.2.1, $\partial f(x)$ is nonempty and compact, implying by Exercise 1.32(a) that the cone generated by $\partial f(x)$ is closed. Therefore, in this case, the closure operation in the preceding relation is unnecessary, i.e.,

$$N_C(x) = \text{cone}\big(\partial f(x)\big).$$

4.22

Let $C_i \subset \Re^{n_i}$, $i = 1, \ldots, m$, be convex sets and let $x_i \in C_i$ for all i. Show that

$$T_{C_1 \times \cdots \times C_m}(x_1, \ldots, x_m) = T_{C_1}(x_1) \times \cdots \times T_{C_m}(x_m),$$

$$N_{C_1 \times \cdots \times C_m}(x_1, \ldots, x_m) = N_{C_1}(x_1) \times \cdots \times N_{C_m}(x_m).$$

Solution: It suffices to consider the case $m = 2$. From the definition of the cone of feasible directions, it can be seen that

$$F_{C_1 \times C_2}(x_1, x_2) = F_{C_1}(x_1) \times F_{C_2}(x_2).$$

By taking the closure of both sides in the preceding relation, and by using the fact that the closure of the Cartesian product of two sets coincides with the Cartesian product of their closures (see Exercise 1.37), we obtain

$$\text{cl}\big(F_{C_1 \times C_2}(x_1, x_2)\big) = \text{cl}\big(F_{C_1}(x_1)\big) \times \text{cl}\big(F_{C_2}(x_2)\big).$$

Since C_1 and C_2 are convex, by Prop. 4.6.2(c), we have

$$T_{C_1}(x_1) = \text{cl}\big(F_{C_1}(x_1)\big), \qquad T_{C_2}(x_2) = \text{cl}\big(F_{C_2}(x_2)\big).$$

Furthermore, the Cartesian product $C_1 \times C_2$ is also convex, and by Prop. 4.6.2(c), we also have

$$T_{C_1 \times C_2}(x_1, x_2) = \text{cl}\big(F_{C_1 \times C_2}(x_1, x_2)\big).$$

By combining the preceding three relations, we obtain

$$T_{C_1\times C_2}(x_1,x_2) = T_{C_1}(x_1)\times T_{C_2}(x_2).$$

By taking polars in the preceding relation, we obtain

$$T_{C_1\times C_2}(x_1,x_2)^* = \big(T_{C_1}(x_1)\times T_{C_2}(x_2)\big)^*,$$

and because the polar of the Cartesian product of two cones coincides with the Cartesian product of their polar cones (see Exercise 3.4), it follows that

$$T_{C_1\times C_2}(x_1,x_2)^* = T_{C_1}(x_1)^*\times T_{C_2}(x_2)^*.$$

Since the sets C_1, C_2, and $C_1\times C_2$ are convex, by Prop. 4.6.3, we have

$$T_{C_1\times C_2}(x_1,x_2)^* = N_{C_1\times C_2}(x_1,x_2),$$

$$T_{C_1}(x_1)^* = N_{C_1}(x_1), \qquad T_{C_2}(x_2)^* = N_{C_2}(x_2),$$

so that

$$N_{C_1\times C_2}(x_1,x_2) = N_{C_1}(x_1)\times N_{C_2}(x_2).$$

4.23 (Tangent and Normal Cone Relations)

Let C_1, C_2, and C be nonempty convex subsets of $\Re^n$. Show the following properties:

(a) We have

$$N_{C_1\cap C_2}(x) \supset N_{C_1}(x) + N_{C_2}(x), \qquad \forall\ x\in C_1\cap C_2,$$

$$T_{C_1\cap C_2}(x) \subset T_{C_1}(x)\cap T_{C_2}(x), \qquad \forall\ x\in C_1\cap C_2.$$

Furthermore, if $\mathrm{ri}(C_1)\cap \mathrm{ri}(C_2)$ is nonempty, the preceding relations hold with equality. This is also true if $\mathrm{ri}(C_1)\cap C_2$ is nonempty and the set C_2 is polyhedral.

(b) For $x_1\in C_1$ and $x_2\in C_2$, we have

$$N_{C_1+C_2}(x_1+x_2) = N_{C_1}(x_1)\cap N_{C_2}(x_2),$$

$$T_{C_1+C_2}(x_1+x_2) = \mathrm{cl}\big(T_{C_1}(x_1)+T_{C_2}(x_2)\big).$$

(c) For an $m\times n$ matrix A and any $x\in C$, we have

$$N_{AC}(Ax) = (A')^{-1}\cdot N_C(x), \qquad T_{AC}(Ax) = \mathrm{cl}\big(A\cdot T_C(x)\big).$$

Solution: (a) We first show that

$$N_{C_1}(x) + N_{C_2}(x) \subset N_{C_1\cap C_2}(x), \qquad \forall\ x\in C_1\cap C_2.$$

For $i = 1, 2$, let $f_i(x) = 0$ when $x \in C$ and $f_i(x) = \infty$ otherwise, so that for $f = f_1 + f_2$, we have

$$f(x) = \begin{cases} 0 & \text{if } x \in C_1 \cap C_2, \\ \infty & \text{otherwise.} \end{cases}$$

By Exercise 4.4(d), we have

$$\partial f_1(x) = N_{C_1}(x), \qquad \forall\, x \in C_1,$$

$$\partial f_2(x) = N_{C_2}(x), \qquad \forall\, x \in C_2,$$

$$\partial f(x) = N_{C_1 \cap C_2}(x), \qquad \forall\, x \in C_1 \cap C_2,$$

while by Exercise 4.9, we have

$$\partial f_1(x) + \partial f_2(x) \subset \partial f(x), \qquad \forall\, x.$$

In particular, this relation holds for every $x \in \mathrm{dom}(f)$ and since $\mathrm{dom}(f) = C_1 \cap C_2$, we obtain

$$N_{C_1}(x) + N_{C_2}(x) \subset N_{C_1 \cap C_2}(x), \qquad \forall\, x \in C_1 \cap C_2. \tag{4.22}$$

If $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty, then by Exercise 4.9, we have

$$\partial f(x) = \partial f_1(x) + \partial f_2(x), \qquad \forall\, x,$$

implying that

$$N_{C_1 \cap C_2}(x) = N_{C_1}(x) + N_{C_2}(x), \qquad \forall\, x \in C_1 \cap C_2. \tag{4.23}$$

Furthermore, by Exercise 4.9, this relation also holds if C_2 is polyhedral and $\mathrm{ri}(C_1) \cap C_2$ is nonempty.

By taking polars in Eq. (4.22), it follows that

$$N_{C_1 \cap C_2}(x)^* \subset \big(N_{C_1}(x) + N_{C_2}(x)\big)^*,$$

and since

$$\big(N_{C_1}(x) + N_{C_2}(x)\big)^* = N_{C_1}(x)^* \cap N_{C_2}(x)^*$$

(see Exercise 3.4), we obtain

$$N_{C_1 \cap C_2}(x)^* \subset N_{C_1}(x)^* \cap N_{C_2}(x)^*. \tag{4.24}$$

Because C_1 and C_2 are convex, their intersection $C_1 \cap C_2$ is also convex, and by Prop. 4.6.3, we have

$$N_{C_1 \cap C_2}(x)^* = T_{C_1 \cap C_2}(x),$$

$$N_{C_1}(x)^* = T_{C_1}(x), \qquad N_{C_2}(x)^* = T_{C_2}(x).$$

In view of Eq. (4.24), it follows that

$$T_{C_1 \cap C_2}(x) \subset T_{C_1}(x) \cap T_{C_2}(x).$$

When $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2)$ is nonempty, or when $\mathrm{ri}(C_1) \cap C_2$ is nonempty and C_2 is polyhedral, by taking the polars in both sides of Eq. (4.23), it can be similarly seen that

$$T_{C_1 \cap C_2}(x) = T_{C_1}(x) \cap T_{C_2}(x).$$

(b) Let $x_1 \in C_1$ and $x_2 \in C_2$ be arbitrary. Since C_1 and C_2 are convex, the sum $C_1 + C_2$ is also convex, so that by Prop. 4.6.3, we have

$$z \in N_{C_1+C_2}(x_1+x_2) \iff z'\big((y_1+y_2)-(x_1+x_2)\big) \le 0, \ \forall\, y_1 \in C_1, \ \forall\, y_2 \in C_2, \tag{4.25}$$

$$z_1 \in N_{C_1}(x_1) \iff z_1'(y_1 - x_1) \le 0, \quad \forall\, y_1 \in C_1, \tag{4.26}$$

$$z_2 \in N_{C_2}(x_2) \iff z_2'(y_2 - x_2) \le 0, \quad \forall\, y_2 \in C_2. \tag{4.27}$$

If $z \in N_{C_1+C_2}(x_1 + x_2)$, then by using $y_2 = x_2$ in Eq. (4.25), we obtain

$$z'(y_1 - x_1) \le 0, \quad \forall\, y_1 \in C_1,$$

implying that $z \in N_{C_1}(x_1)$. Similarly, by using $y_1 = x_1$ in Eq. (4.25), we see that $z \in N_{C_2}(x_2)$. Hence $z \in N_{C_1}(x_1) \cap N_{C_2}(x_2)$ implying that

$$N_{C_1+C_2}(x_1 + x_2) \subset N_{C_1}(x_1) \cap N_{C_2}(x_2).$$

Conversely, let $z \in N_{C_1}(x_1) \cap N_{C_2}(x_2)$, so that both Eqs. (4.26) and (4.27) hold, and by adding them, we obtain

$$z'\big((y_1 + y_2) - (x_1 + x_2)\big) \le 0, \qquad \forall\, y_1 \in C_1, \quad \forall\, y_2 \in C_2.$$

Therefore, in view of Eq. (4.25), we have $z \in N_{C_1+C_2}(x_1 + x_2)$, showing that

$$N_{C_1}(x_1) \cap N_{C_2}(x_2) \subset N_{C_1+C_2}(x_1 + x_2).$$

Hence

$$N_{C_1+C_2}(x_1 + x_2) = N_{C_1}(x_1) \cap N_{C_2}(x_2).$$

By taking polars in this relation, we obtain

$$N_{C_1+C_2}(x_1 + x_2)^* = \big(N_{C_1}(x_1) \cap N_{C_2}(x_2)\big)^*.$$

Since $N_{C_1}(x_1)$ and $N_{C_2}(x_2)$ are closed convex cones, by Exercise 3.4, it follows that

$$N_{C_1+C_2}(x_1 + x_2)^* = \mathrm{cl}\big(N_{C_1}(x_1)^* + N_{C_2}(x_2)^*\big).$$

The sets C_1, C_2, and $C_1 + C_2$ are convex, so that by Prop. 4.6.3, we have

$$N_{C_1}(x_1)^* = T_{C_1}(x_1), \qquad N_{C_2}(x_2)^* = T_{C_2}(x_2),$$

$$N_{C_1+C_2}(x_1 + x_2)^* = T_{C_1+C_2}(x_1 + x_2),$$

implying that

$$T_{C_1+C_2}(x_1 + x_2) = \mathrm{cl}\big(T_{C_1}(x_1) + T_{C_2}(x_2)\big).$$

(c) Let $x \in C$ be arbitrary. Since C is convex, its image AC under the linear transformation A is also convex, so by Prop. 4.6.3, we have

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad z'(y - Ax) \leq 0, \quad \forall\, y \in AC,$$

which is equivalent to

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad z'(Av - Ax) \leq 0, \quad \forall\, v \in C.$$

Furthermore, the condition

$$z'(Av - Ax) \leq 0, \qquad \forall\, v \in C$$

is the same as

$$(A'z)'(v - x) \leq 0, \qquad \forall\, v \in C,$$

and since C is convex, by Prop. 4.6.3, this is equivalent to $A'z \in N_C(x)$. Thus,

$$z \in N_{AC}(Ax) \quad \Longleftrightarrow \quad A'z \in N_C(x),$$

which together with the fact $A'z \in N_C(x)$ if and only if $z \in (A')^{-1} \cdot N_C(x)$ yields

$$N_{AC}(Ax) = (A')^{-1} \cdot N_C(x).$$

By taking polars in the preceding relation, we obtain

$$N_{AC}(Ax)^* = \big((A')^{-1} \cdot N_C(x)\big)^*. \tag{4.28}$$

Because AC is convex, by Prop. 4.6.3, we have

$$N_{AC}(Ax)^* = T_{AC}(Ax). \tag{4.29}$$

Since C is convex, by the same proposition, we have $N_C(x) = T_C(x)^*$, so that $N_C(x)$ is a closed convex cone and by using Exercise 3.5, we obtain

$$\big((A')^{-1} \cdot N_C(x)\big)^* = \mathrm{cl}\big(A \cdot N_C(x)^*\big).$$

Furthermore, by convexity of C, we also have $N_C(x)^* = T_C(x)$, implying that

$$\big((A')^{-1} \cdot N_C(x)\big)^* = \mathrm{cl}\big(A \cdot T_C(x)\big). \tag{4.30}$$

Combining Eqs. (4.28), (4.29), and (4.30), we obtain

$$T_{AC}(Ax) = \mathrm{cl}\big(A \cdot T_C(x)\big).$$

4.24 [GoT71], [RoW98]

Let C be a subset of $\Re^n$ and let $x^* \in C$. Show that for every $y \in T_C(x^*)^*$ there is a smooth function f with $-\nabla f(x^*) = y$, and such that x^* is the unique global minimum of f over C.

Solution: We assume for simplicity that all the constraints are inequalities. Consider the scalar function $\theta_0 : [0, \infty) \mapsto \Re$ defined by

$$\theta_0(r) = \sup_{x \in C,\, \|x - x^*\| \leq r} y'(x - x^*), \qquad r \geq 0.$$

Clearly $\theta_0(r)$ is nondecreasing and satisfies

$$0 = \theta_0(0) \leq \theta_0(r), \qquad \forall\ r \geq 0.$$

Furthermore, since $y \in T_C(x^*)^*$, we have $y'(x - x^*) \leq o(\|x - x^*\|)$ for $x \in C$, so that $\theta_0(r) = o(r)$, which implies that θ_0 is differentiable at $r = 0$ with $\nabla\theta_0(0) = 0$. Thus, the function F_0 defined by

$$F_0(x) = \theta_0(\|x - x^*\|) - y'(x - x^*)$$

is differentiable at x^*, attains a global minimum over C at x^*, and satisfies

$$-\nabla F_0(x^*) = y.$$

If F_0 were smooth we would be done, but since it need not even be continuous, we will successively perturb it into a smooth function. We first define the function $\theta_1 : [0, \infty) \mapsto \Re$ by

$$\theta_1(r) = \begin{cases} \frac{1}{r}\int_r^{2r} \theta_0(s)ds & \text{if } r > 0, \\ 0 & \text{if } r = 0, \end{cases}$$

(the integral above is well-defined since the function θ_0 is nondecreasing). The function θ_1 is seen to be nondecreasing and continuous, and satisfies

$$0 \leq \theta_0(r) \leq \theta_1(r), \qquad \forall\ r \geq 0,$$

$\theta_1(0) = 0$, and $\nabla\theta_1(0) = 0$. Thus the function

$$F_1(x) = \theta_1(\|x - x^*\|) - y'(x - x^*)$$

has the same significant properties for our purposes as F_0 [attains a global minimum over C at x^*, and has $-\nabla F_1(x^*) = y$], and is in addition continuous.

We next define the function $\theta_2 : [0, \infty) \mapsto \Re$ by

$$\theta_2(r) = \begin{cases} \frac{1}{r}\int_r^{2r} \theta_1(s)ds & \text{if } r > 0, \\ 0 & \text{if } r = 0. \end{cases}$$

Again θ_2 is seen to be nondecreasing, and satisfies

$$0 \leq \theta_1(r) \leq \theta_2(r), \qquad \forall\, r \geq 0,$$

$\theta_2(0) = 0$, and $\nabla\theta_2(0) = 0$. Also, because θ_1 is continuous, θ_2 is smooth, and so is the function F_2 given by

$$F_2(x) = \theta_2(\|x - x^*\|) - y'(x - x^*).$$

The function F_2 fulfills all the requirements of the proposition, except that it may have global minima other than x^*. To ensure the uniqueness of x^* we modify F_2 as follows:

$$F(x) = F_2(x) + \|x - x^*\|^2.$$

The function F is smooth, attains a strict global minimum over C at x^*, and satisfies $-\nabla F(x^*) = y$.

4.25

Let C_1, C_2, and C_3 be nonempty closed subsets of $\Re^n$. Consider the problem of finding a triangle with minimum perimeter that has one vertex on each of the sets, i.e., the problem of minimizing $\|x_1 - x_2\| + \|x_2 - x_3\| + \|x_3 - x_1\|$ subject to $x_i \in C_i$, $i = 1, 2, 3$, and the additional condition that x_1, x_2, and x_3 do not lie on the same line. Show that if (x_1^*, x_2^*, x_3^*) defines an optimal triangle, there exists a vector z^* in the triangle such that

$$(z^* - x_i^*) \in T_{C_i}(x_i^*)^*, \qquad i = 1, 2, 3.$$

Solution: We consider the problem

$$\begin{aligned} &\text{minimize} \quad \|x_1 - x_2\| + \|x_2 - x_3\| + \|x_3 - x_1\| \\ &\text{subject to} \quad x_i \in C_i, \quad i = 1, 2, 3, \end{aligned}$$

with the additional condition that x_1, x_2 and x_3 do not lie on the same line. Suppose that (x_1^*, x_2^*, x_3^*) defines an optimal triangle. Then, x_1^* solves the problem

$$\begin{aligned} &\text{minimize} \quad \|x_1 - x_2^*\| + \|x_2^* - x_3^*\| + \|x_3^* - x_1\| \\ &\text{subject to} \quad x_1 \in C_1, \end{aligned}$$

for which we have the following necessary optimality condition

$$d_1 = \frac{x_2^* - x_1^*}{\|x_2^* - x_1^*\|} + \frac{x_3^* - x_1^*}{\|x_3^* - x_1^*\|} \in T_{C_1}(x_1^*)^*.$$

The half-line $\{x \mid x = x_1^* + \alpha d_1,\ \alpha \geq 0\}$ is one of the bisectors of the optimal triangle. Similarly, there exist $d_2 \in T_{C_2}(x_2^*)^*$ and $d_3 \in T_{C_3}(x_3^*)^*$ which define the remaining bisectors of the optimal triangle. By elementary geometry, there exists a unique point z^* at which all three bisectors intersect (z^* is the center of the circle that is inscribed in the optimal triangle). From the necessary optimality conditions we have

$$z^* - x_i^* = \alpha_i d_i \in T_{C_i}(x_i^*)^*, \qquad i = 1, 2, 3.$$

4.26

Consider the problem of minimizing a convex function $f : \Re^n \mapsto \Re$ over the polyhedral set

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

Show that x^* is an optimal solution if and only if there exist scalars $\mu_1^*, \ldots, \mu_r^*$ such that

(i) $\mu_j^* \ge 0$ for all j, and $\mu_j^* = 0$ for all j such that $a_j'x^* < b_j$.

(ii) $0 \in \partial f(x^*) + \sum_{j=1}^r \mu_j^* a_j$.

Hint: Characterize the cone $T_X(x^*)^*$, and use Prop. 4.7.2 and Farkas' Lemma.

Solution: Let us characterize the cone $T_X(x^*)^*$. Define

$$A(x^*) = \big\{ j \mid a_j'x^* = b_j \big\}.$$

Since X is convex, by Prop. 4.6.2, we have

$$T_X(x^*) = \mathrm{cl}\big(F_X(x^*)\big),$$

while from definition of X, we have

$$F_X(x^*) = \big\{ y \mid a_j'y \le 0,\ \forall\, j \in A(x^*) \big\},$$

and since this set is closed, it follows that

$$T_X(x^*) = \big\{ y \mid a_j'y \le 0,\ \forall\, j \in A(x^*) \big\}.$$

By taking polars in this relation and by using the Farkas' Lemma [Prop. 3.2.1(b)], we obtain

$$T_X(x^*)^* = \left\{ \sum_{j \in A(x^*)} \mu_j a_j \ \middle|\ \mu_j \ge 0,\ \forall\, j \in A(x^*) \right\},$$

and by letting $\mu_j = 0$ for all $j \notin A(x^*)$, we can write

$$T_X(x^*)^* = \left\{ \sum_{j=1}^r \mu_j a_j \ \middle|\ \mu_j \ge 0,\ \forall\, j,\quad \mu_j = 0,\ \forall\, j \notin A(x^*) \right\}. \tag{4.31}$$

By Prop. 4.7.2, the vector x^* minimizes f over X if and only if

$$0 \in \partial f(x^*) + T_X(x^*)^*.$$

In view of Eq. (4.31) and the definition of $A(x^*)$, it follows that x^* minimizes f over X if and only if there exist $\mu_1^*, \ldots, \mu_r^*$ such that

(i) $\mu_j^* \ge 0$ for all $j = 1, \ldots, r$, and $\mu_j^* = 0$ for all j such that $a_j'x^* < b_j$.

(ii) $0 \in \partial f(x^*) + \sum_{j=1}^r \mu_j^* a_j$.

4.27 (Quasiregularity)

Let $f : \Re^n \mapsto \Re$ be a smooth function, let X be a subset of $\Re^n$, and let x^* be a local minimum of f over X. Denote $D(x) = \big\{y \mid \nabla f(x)'y < 0\big\}$.

(a) Show that $D(x^*) \cap T_X(x^*) = \emptyset$.

(b) Suppose that X has the form

$$X = \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0,\ g_1(x) \le 0, \ldots, g_r(x) \le 0\big\},$$

where the functions $h_i : \Re^n \mapsto \Re$, $i = 1, \ldots, m$, and $g_j : \Re^n \mapsto \Re$, $j = 1, \ldots, r$, are smooth. For any $x \in X$ consider the cone

$$V(x) = \big\{y \mid \nabla h_i(x)'y = 0,\ i = 1, \ldots, m,\ \nabla g_j(x)'y \le 0,\ j \in A(x)\big\},$$

where $A(x) = \big\{j \mid g_j(x) = 0\big\}$. Show that $T_X(x) \subset V(x)$.

(c) Use Farkas' Lemma and part (a) to show that if $T_X(x^*) = V(x^*)$, then there exist scalars $\lambda_1^*, \ldots, \lambda_m^*$ and $\mu_1^*, \ldots, \mu_r^*$, satisfying

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* \ge 0, \quad \forall\ j = 1, \ldots, r, \qquad \mu_j^* = 0, \quad \forall\ j \in A(x^*).$$

Note: The condition $T_X(x^*) = V(x^*)$ is called *quasiregularity* at x^*, and will be discussed further in Chapter 5.

Solution: (a) Let y be a nonzero tangent of X at x^*. Then there exists a sequence $\{\xi^k\}$ and a sequence $\{x^k\} \subset X$ such that $x^k \ne x^*$ for all k,

$$\xi^k \to 0, \qquad x^k \to x^*,$$

and

$$\frac{x^k - x^*}{\|x^k - x^*\|} = \frac{y}{\|y\|} + \xi^k. \tag{4.32}$$

By the mean value theorem, we have for all k

$$f(x^k) = f(x^*) + \nabla f(\tilde{x}^k)'(x^k - x^*),$$

where $\tilde{x}^k$ is a vector that lies on the line segment joining x^k and x^*. Using Eq. (4.32), the last relation can be written as

$$f(x^k) = f(x^*) + \frac{\|x^k - x^*\|}{\|y\|} \nabla f(\tilde{x}^k)'y^k, \tag{4.33}$$

where

$$y^k = y + \|y\|\xi^k.$$

If the tangent y satsifies $\nabla f(x^*)'y < 0$, then, since $\tilde{x}^k \to x^*$ and $y^k \to y$, we obtain for all sufficiently large k, $\nabla f(\tilde{x}^k)'y^k < 0$ and [from Eq. (4.33)] $f(x^k) < f(x^*)$. This contradicts the local optimality of x^*.

(b) Assume first that there are no equality constraints. Let $x \in X$ and let y be a nonzero tangent of X at x. Then there exists a sequence $\{\xi^k\}$ and a sequence $\{x^k\} \subset X$ such that $x^k \neq x$ for all k,

$$\xi^k \to 0, \qquad x^k \to x,$$

and

$$\frac{x^k - x}{\|x^k - x\|} = \frac{y}{\|y\|} + \xi^k.$$

By the mean value theorem, we have for all j and k

$$0 \geq g_j(x^k) = g_j(x) + \nabla g_j(\tilde{x}^k)'(x^k - x) = \nabla g_j(\tilde{x}^k)'(x^k - x),$$

where $\tilde{x}^k$ is a vector that lies on the line segment joining x^k and x. This relation can be written as

$$\frac{\|x^k - x\|}{\|y\|} \nabla g_j(\tilde{x}^k)'y^k \leq 0,$$

where $y^k = y + \xi^k \|y\|$, or equivalently

$$\nabla g_j(\tilde{x}^k)'y^k \leq 0, \qquad y^k = y + \xi^k \|y\|.$$

Taking the limit as $k \to \infty$, we obtain $\nabla g_j(x)'y \leq 0$ for all j, thus proving that $y \in V(x)$, and $T_X(x) \subset V(x)$. If there are some equality constraints $h_i(x) = 0$, they can be converted to the two inequality constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, and the result follows similarly.

(c) Assume first that there are no equality constraints. From part (a), we have $D(x^*) \cap V(x^*) = \emptyset$, which is equivalent to having $\nabla f(x^*)'y \geq 0$ for all y with $\nabla g_j(x^*)'y \leq 0$ for all $j \in A(x^*)$. By Farkas' Lemma, this is equivalent to the existence of Lagrange multipliers μ_j^* with the properties stated in the exercise. If there are some equality constraints $h_i(x) = 0$, they can be converted to the two inequality constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, and the result follows similarly.