Chapter Two

Generalized Gradients

What is now proved was once only imagined.

William Blake, *The Marriage of Heaven and Hell*

I was losing my distrust of generalizations.

Thornton Wilder, *Theophilus North*

In this chapter we gather a basic toolkit that will be used throughout the rest of the book. The theory and the calculus of generalized gradients are developed in detail, beginning with the case of a real-valued "locally Lipschitz" function defined on a Banach space. We also develop an associated geometric theory of normal and tangent cones, and explore the relationship between all of these concepts and their counterparts in smooth and in convex analysis. Later in the chapter special attention is paid to the case in which the underlying space is finite-dimensional, and also to the generalized gradients of certain important kinds of functionals. Examples are given, and extensions of the theory to non-Lipschitz and vector-valued functions are also carried out.

At the risk of losing some readers for one of our favorite chapters, we will concede that the reader who wishes only to have access to the statements of certain results of later chapters may find the introduction to generalized gradients given in Chapter 1 adequate for this purpose. Those who wish to follow all the details of the proofs (in this and later chapters) will require a background in real and functional analysis. In particular, certain standard constructs and results of the theory of Banach spaces are freely invoked. (Both the prerequisites and the difficulty of many proofs are greatly reduced if it is assumed that the Banach space is finite-dimensional.)

2.1 DEFINITION AND BASIC PROPERTIES

We shall be working in a Banach space X whose elements x we shall refer to as vectors or points, whose norm we shall denote $\|x\|$, and whose open unit ball is denoted B; the closed unit ball is denoted $\bar{B}$.

The Lipschitz Condition

Let Y be a subset of X. A function $f\colon Y \to R$ is said to satisfy a *Lipschitz condition* (on Y) provided that, for some nonnegative scalar K, one has

$$|f(y) - f(y')| \leqslant K\|y - y'\| \tag{1}$$

for all points y, y' in Y; this is also referred to as a Lipschitz condition of *rank* K. We shall say that f is Lipschitz (of rank K) *near* x if, for some $\varepsilon > 0$, f satisfies a Lipschitz condition (of rank K) on the set $x + \varepsilon B$ (i.e., within an ε-neighborhood of x).

For functions of a real variable, a Lipschitz condition is a requirement that the graph of f not be "too steep." It is easy to see that a function having this property near a point need not be differentiable there, nor need it admit directional derivatives in the classical sense.

The Generalized Directional Derivative

Let f be Lipschitz near a given point x, and let v be any other vector in X. The *generalized directional derivative* of f at x in the direction v, denoted $f^\circ(x; v)$, is defined as follows:

$$f^\circ(x; v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t},$$

where of course y is a vector in X and t is a positive scalar. Note that this definition does not presuppose the existence of any limit (since it involves an upper limit only), that it involves only the behavior of f near x, and that it differs from traditional definitions of directional derivatives in that the base point (y) of the difference quotient varies. The utility of f° stems from the following basic properties.

2.1.1 Proposition

Let f be Lipschitz of rank K near x. Then

(a) *The function $v \to f^\circ(x; v)$ is finite, positively homogeneous, and subadditive on X, and satisfies*

$$|f^\circ(x; v)| \leqslant K\|v\|.$$

(b) $f^\circ(x; v)$ *is upper semicontinuous as a function of* (x, v) *and, as a function of* v *alone, is Lipschitz of rank* K *on* X.
(c) $f^\circ(x; -v) = (-f)^\circ(x; v)$.

Proof. In view of the Lipschitz condition, the absolute value of the difference quotient in the definition of $f^\circ(x; v)$ is bounded by $K\|v\|$ when y is sufficiently near x and t sufficiently near 0. It follows that $|f^\circ(x; v)|$ admits the same upper bound. The fact that $f^\circ(x; \lambda v) = \lambda f^\circ(x; v)$ for any $\lambda > 0$ is immediate, so let us now turn to the subadditivity. With all the upper limits below understood to be taken as $y \to x$ and $t \downarrow 0$, we calculate:

$$f^\circ(x; v + w) = \limsup \frac{f(y + tv + tw) - f(y)}{t}$$

$$\leqslant \limsup \frac{f(y + tv + tw) - f(y + tw)}{t} + \limsup \frac{f(y + tw) - f(y)}{t}$$

(since the upper limit of a sum is bounded above by the sum of the upper limits). The first upper limit in this last expression is $f^\circ(x; v)$, since the term $y + tw$ represents in essence just a dummy variable converging to x. We conclude

$$f^\circ(x; v + w) \leqslant f^\circ(x; v) + f^\circ(x; w),$$

which establishes (a).

Now let $\{x_i\}$ and $\{v_i\}$ be arbitrary sequences converging to x and v, respectively. For each i, by definition of upper limit, there exist y_i in X and $t_i > 0$ such that

$$\|y_i - x_i\| + t_i < \frac{1}{i},$$

$$f^\circ(x_i; v_i) - \frac{1}{i} \leqslant \frac{f(y_i + t_i v_i) - f(y_i)}{t_i}$$

$$= \frac{f(y_i + t_i v) - f(y_i)}{t_i} + \frac{f(y_i + t_i v_i) - f(y_i + t_i v)}{t_i}.$$

Note that the last term is bounded in magnitude by $K\|v_i - v\|$ (in view of the Lipschitz condition). Upon taking upper limits (as $i \to \infty$), we derive

$$\limsup_{i \to \infty} f^\circ(x_i; v_i) \leqslant f^\circ(x; v),$$

which establishes the upper semicontinuity.

Finally, let any v and w in X be given. We have

$$f(y + tv) - f(y) \leqslant f(y + tw) - f(y) + K\|v - w\|t$$

for y near x, t near 0. Dividing by t and taking upper limits as $y \to x$, $t \downarrow 0$, gives

$$f^\circ(x; v) \leqslant f^\circ(x; w) + K\|v - w\|.$$

Since this also holds with v and w switched, (b) follows. To prove (c), we calculate:

$$f^\circ(x; -v) := \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x' - tv) - f(x')}{t}$$

$$= \limsup_{\substack{u \to x \\ t \downarrow 0}} \frac{(-f)(u + tv) - (-f)(u)}{t}, \quad \text{where } u := x' - tv$$

$$= (-f)^\circ(x; v), \quad \text{as stated.} \quad \square$$

The Generalized Gradient

The Hahn–Banach Theorem asserts that any positively homogeneous and subadditive functional on X majorizes some linear functional on X. Under the conditions of Proposition 2.1.1, therefore, there is at least one linear functional $\zeta: X \to R$ such that, for all v in X, one has $f^\circ(x; v) \geqslant \zeta(v)$. It follows also that ζ is bounded, and hence belongs to the dual space X^* of continuous linear functionals on X, for which we adopt the convention of using $\langle \zeta, v \rangle$ or $\langle v, \zeta \rangle$ for $\zeta(v)$. We are led to the following definition: the *generalized gradient* of f at x, denoted $\partial f(x)$, is the subset of X^* given by

$$\{\zeta \in X^* : f^\circ(x; v) \geqslant \langle \zeta, v \rangle \text{ for all } v \text{ in } X\}.$$

We denote by $\|\zeta\|_*$ the norm in X^*:

$$\|\zeta\|_* := \sup\{\langle \zeta, v \rangle : v \in X, \|v\| \leqslant 1\},$$

and B_* denotes the open unit ball in X^*. The following summarizes some basic properties of the generalized gradient.

2.1.2 Proposition

Let f be Lipschitz of rank K near x. Then

(a) *$\partial f(x)$ is a nonempty, convex, weak*-compact subset of X^* and $\|\zeta\|_* \leqslant K$ for every ζ in $\partial f(x)$.*
(b) *For every v in X, one has*

$$f^\circ(x; v) = \max\{\langle \zeta, v \rangle : \zeta \in \partial f(x)\}.$$

Proof. Assertion (a) is immediate from our preceding remarks and Proposition 2.1.1. (The weak*-compactness follows from Alaoglu's Theorem.) Assertion (b) is simply a restatement of the fact that $\partial f(x)$ is by definition the weak*-closed convex set whose support function (see Section 1.2 and below) is $f°(x; \cdot)$. To see this independently, suppose that for some v, $f°(x; v)$ exceeded the given maximum (it can't be less, by definition of $\partial f(x)$). According to a common version of the Hahn–Banach Theorem, there is a linear functional ζ majorized by $f°(x; \cdot)$ and agreeing with it at v. It follows that ζ belongs to $\partial f(x)$, whence $f°(x; v) > \langle \zeta, v \rangle = f°(x; v)$, a contradiction which establishes (b). □

2.1.3 Example—The Absolute-Value Function

As in the case of the classical calculus, one of our goals is to have to resort rarely to the definition in order to calculate generalized gradients in practice. For illustrative purposes, however, let us nonetheless so calculate the generalized gradient of the absolute-value function on the reals. In this case, $X = R$ and $f(x) = |x|$ (f is Lipschitz by the triangle inequality). If x is strictly positive, we calculate

$$f°(x; v) = \lim_{\substack{y \to x \\ t \downarrow 0}} \frac{y + tv - y}{t} = v,$$

so that $\partial f(x)$, the set of numbers ζ satisfying $v \geqslant \zeta v$ for all v, reduces to the singleton $\{1\}$. Similarly, $\partial f(x) = \{-1\}$ if $x < 0$. The remaining case is $x = 0$. We find

$$f°(0; v) = \begin{cases} v & \text{if } v \geqslant 0 \\ -v & \text{if } v < 0; \end{cases}$$

that is, $f°(0; v) = |v|$. Thus $\partial f(0)$ consists of those ζ satisfying $|v| \geqslant \zeta v$ for all v; that is, $\partial f(0) = [-1, 1]$.

Support Functions

As Proposition 2.1.2 makes clear, it is equivalent to know the set $\partial f(x)$ or the function $f°(x; \cdot)$; each is obtainable from the other. This is an instance of a general fact: closed convex sets are characterized by their support functions. Recall that the *support function* of a nonempty subset C of X is the function σ_C: $X^* \to R \cup \{+\infty\}$ defined by

$$\sigma_C(\zeta) := \sup\{\langle \zeta, x \rangle : x \in C\}.$$

If Σ is a subset of X^*, its support function is defined on X^{**}. If we view X as a subset of X^{**}, then for $x \in X$, one has

$$\sigma_\Sigma(x) = \sup\{\langle \zeta, x \rangle : \zeta \in \Sigma\}.$$

The following facts are known (see Hörmander, 1954).

2.1.4 Proposition

Let C, D be nonempty closed convex subsets of X, and let Σ, Δ be nonempty weak-closed convex subsets of X^*. Then*

(a) $C \subset D$ *iff* $\sigma_C(\zeta) \leqslant \sigma_D(\zeta)$ *for all* $\zeta \in X^*$.
(b) $\Sigma \subset \Delta$ *iff* $\sigma_\Sigma(x) \leqslant \sigma_\Delta(x)$ *for all* $x \in X$.
(c) Σ *is weak*-compact iff* $\sigma_\Sigma(\cdot)$ *is finite-valued on* X.
(d) *A given function* $\sigma\colon X \to R \cup \{+\infty\}$ *is positively homogeneous, subadditive, lower semicontinuous (strong or weak) and not identically* $+\infty$ *iff there is a nonempty weak*-closed convex subset* Σ *of* X^* *such that* $\sigma = \sigma_\Sigma$. *Any such* Σ *is unique.*

A *multifunction* $\Gamma\colon X \to Y$ is a mapping from X to the subsets of Y. Its *graph* is the set

$$\operatorname{Gr}\Gamma := \{(x, y) : x \in X,\ y \in \Gamma(x)\}.$$

Γ is *closed* if $\operatorname{Gr}\Gamma$ is closed in $X \times Y$ (relative to a given topology). When X and Y are Banach spaces, we define *upper semicontinuity* of Γ at x to be the following property: for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\Gamma(x') \subset \Gamma(x) + \varepsilon B_Y \quad \text{for all } x' \in x + \delta B_X.$$

We continue to assume that f is Lipschitz near x. The first assertion below reiterates that $f^\circ(x; \cdot)$ is the support function of $\partial f(x)$.

2.1.5 Proposition

(a) $\zeta \in \partial f(x)$ *iff* $f^\circ(x; v) \geqslant \langle \zeta, v\rangle$ *for all* v *in* X.
(b) *Let* x_i *and* ζ_i *be sequences in* X *and* X^* *such that* $\zeta_i \in \partial f(x_i)$. *Suppose that* x_i *converges to* x, *and that* ζ *is a cluster point of* ζ_i *in the weak* topology. Then one has* $\zeta \in \partial f(x)$. *(That is, the multifunction* ∂f *is weak*-closed.)*
(c)

$$\partial f(x) = \bigcap_{\delta > 0} \bigcup_{y \in x + \delta B} \partial f(y).$$

(d) *If* X *is finite-dimensional, then* ∂f *is upper semicontinuous at* x.

Proof. Assertion (a) is evident from Propositions 2.1.2 and 2.1.4. In order to prove (b), let any v in X be given. There is a subsequence of the numerical sequence $\langle \zeta_i, v\rangle$ which converges to $\langle \zeta, v\rangle$ (we do not relabel). One has $f^\circ(x_i; v) \geqslant \langle \zeta_i, v\rangle$ by property (a), which implies $f^\circ(x; v) \geqslant \langle \zeta, v\rangle$ by the upper semicontinuity of f° (see Proposition 2.1.1(b)). Since v is arbitrary, ζ belongs to $\partial f(x)$ by assertion (a).

Part (c) of the proposition is an immediate consequence of (b), so we turn now to (d). If ∂f fails to be upper semicontinuous at x, then we can construct a sequence x_i converging to x and a sequence ζ_i converging to ζ such that ζ_i belongs to $\partial f(x_i)$ for each i, yet ζ does not belong to $\partial f(x)$. This contradicts assertion (b). □

2.2 RELATION TO DERIVATIVES AND SUBDERIVATIVES

We pause to recall some terminology before pursuing the study of generalized gradients. The main results of this section will be that ∂f reduces to the derivative if f is C^1, and to the subdifferential of convex analysis if f is convex.

Classical Derivatives

Let F map X to another Banach space Y. The usual (one-sided) directional derivative of F at x in the direction v is

$$F'(x; v) := \lim_{t \downarrow 0} \frac{F(x + tv) - F(x)}{t}$$

when this limit exists. F is said to admit a *Gâteaux derivative* at x, an element in the space $\mathcal{L}(X, Y)$ of continuous linear functionals from X to Y denoted $DF(x)$, provided that for every v in X, $F'(x; v)$ exists and equals $\langle DF(x), v\rangle$. Let us note that this is equivalent to saying that the difference quotient converges for each v, that one has

$$\lim_{t \downarrow 0} \frac{F(x + tv) - F(x)}{t} = \langle DF(x), v\rangle,$$

and that the convergence is uniform with respect to v in finite sets (the last is automatically true). If the word "finite" in the preceding sentence is replaced by "compact," the derivative is known as *Hadamard*; for "bounded" we obtain the *Fréchet* derivative. In general, these are progressively more demanding requirements. When $X = R^n$, Hadamard and Fréchet differentiability are equivalent; when F is Lipschitz near x (i.e., when for some constant K one has $\|F(x') - F(x'')\|_Y \leqslant K\|x' - x''\|_X$ for all x', x'' near x) then Hadamard and Gâteaux differentiabilities coincide.

Strict Differentiability

It turns out that the differential concept most naturally linked to the theory of this chapter is that of *strict differentiability* (Bourbaki: "strictement dérivable"). We shall say that F admits a strict derivative at x, an element of $\mathcal{L}(X, Y)$ denoted $D_sF(x)$, provided that for each v, the following holds:

$$\lim_{\substack{x' \to x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = \langle D_sF(x), v\rangle,$$

and provided the convergence is uniform for v in compact sets. (This last condition is automatic if F is Lipschitz near x). Note that ours is a "Hadamard-type strict derivative."

2.2.1 Proposition

Let F map a neighborhood of x to Y, and let ζ be an element of $\mathfrak{L}(X, Y)$. The following are equivalent:

(a) *F is strictly differentiable at x and $D_s F(x) = \zeta$.*
(b) *F is Lipschitz near x, and for each v in X one has*

$$\lim_{\substack{x' \to x \\ t \downarrow 0}} \frac{F(x' + tv) - F(x')}{t} = \langle \zeta, v \rangle.$$

Proof. Assume (a). The equality in (b) holds by assumption, so to prove (b) we need only show that F is Lipschitz near x. If this is not the case, there exist sequences $\{x_i\}$ and $\{x_i'\}$ converging to x such that x_i, x_i' lie in $x + (1/i)B$ and

$$\|F(x_i') - F(x_i)\|_Y > i\|x_i' - x_i\|_X.$$

Let us define $t_i > 0$ and v_i via $x_i' = x_i + t_i v_i$ and $\|v_i\| = i^{-1/2}$. It follows that $t_i \to 0$.

Let V consist of the points in the sequence $\{v_i\}$ together with 0. Note that V is compact, so that by definition of $D_s F(x)$ for any $\varepsilon > 0$ there exists n_ε such that, for all $i \geqslant n_\varepsilon$, for all v in V, one has

$$\left\| \frac{F(x_i + t_i v) - F(x_i)}{t_i} - \langle DF(x), v \rangle \right\|_Y < \varepsilon.$$

But this is impossible since when $v = v_i$, the term $[F(x_i + t_i v) - F(x_i)]/t_i$ has norm exceeding $i^{1/2}$ by construction. Thus (b) holds.

We now posit (b). Let V be any compact subset of X and ε any positive number. In view of (b), there exists for each v in V a number $\delta(v) > 0$ such that

$$(1) \qquad \left\| \frac{F(x' + tv) - F(x')}{t} - \langle \zeta, v \rangle \right\|_Y < \varepsilon$$

for all $x' \in x + \delta B$ and $t \in (0, \delta)$. Since the norm of

$$\frac{F(x' + tv') - F(x')}{t} - \frac{F(x' + tv) - F(x')}{t}$$

is bounded above by $K\|v - v'\|$ (where K is a Lipschitz constant for F and where x', t are sufficiently near x and 0), we deduce from (1) that for a suitable redefinition of $\delta(v)$, one has

$$\left\| \frac{F(x' + tv') - F(x')}{t} - \langle \zeta, v' \rangle \right\|_Y < 2\varepsilon \tag{2}$$

for all x' in $x + \delta B$, v' in $v + \delta B$, and t in $(0, \delta)$. A finite number of the open sets $\{v + \delta(v)B : v \in V\}$ will cover V, say, those that correspond to $v_1, v_2, \ldots, v_n$. If we set $\delta' = \min_{1 \leqslant i \leqslant n} \delta(v_i)$, it follows then that (1) holds (with 2ε for ε) for any v in V, for all x' in $x + \delta' B$ and t in $(0, \delta')$. Thus ζ is the strict derivative of F at x, and the proof is complete. □

We shall call F *continuously* (*Gâteaux*) *differentiable* at x provided that on a neighborhood of x the Gâteaux derivative DF exists and is continuous as a mapping from X to $\mathcal{L}(X, Y)$ (with its operator norm topology).

Corollary

If F is continuously differentiable at x, then F is strictly differentiable at x and hence Lipschitz near x.

Proof. To prove that $DF(x)$ is actually the strict derivative, it suffices by the proposition to prove the equality in (b) for each v, for it follows easily from the vector mean value theorem (see McLeod, 1965) that F is Lipschitz near x. In turn, it suffices to show that, for any trio of sequences $\{x_i\}$, $\{v_i\}$, $\{t_i\}$ (with $t_i > 0$) converging to $x, v, 0$, respectively, one has

$$\lim_{i \to \infty} \sup_{\|\theta\|_{Y^*} \leqslant 1} \theta \left[\frac{F(x_i + t_i v_i) - F(x_i)}{t_i} - \langle DF(x), v_i \rangle \right] = 0.$$

By the mean-value theorem, there is x_i^* between x_i and $x_i + t_i v_i$ such that the last expression equals

$$\langle \theta\, DF(x_i^*), v_i \rangle - \langle \theta\, DF(x), v_i \rangle = \langle \theta [DF(x_i^*) - DF(x)], v_i \rangle.$$

This goes to 0 (uniformly in θ) as $i \to \infty$ by the continuity of DF. □

We are now ready to explore the relationship between the various derivatives defined above and the generalized gradient.

2.2.2 Proposition

Let f be Lipschitz near x and admit a Gâteaux (or Hadamard, or strict, or Fréchet) derivative $Df(x)$. Then $Df(x) \in \partial f(x)$.

Proof. By definition, $f'(x; v)$ exists for each v and equals $\langle Df(x), v \rangle$. Clearly one has $f' \leqslant f^\circ$ from the definition of the latter, so one has $f^\circ(x; v) \geqslant$

$\langle Df(x), v\rangle$ for all v in X. The required conclusion now follows from Proposition 2.1.5(a). □

2.2.3 Example

That $\partial f(x)$ can contain points other than $Df(x)$ is illustrated by the familiar example on R: $f(x) := x^2 \sin(1/x)$. This function is Lipschitz near 0, and it is easy to show that $f^\circ(0; v) = |v|$. It follows that $\partial f(0) = [-1, 1]$, a set which contains the (nonstrict) derivative $Df(0) = 0$.

2.2.4 Proposition

If f is strictly differentiable at x, then f is Lipschitz near x and $\partial f(x) = \{D_s f(x)\}$. Conversely, if f is Lipschitz near x and $\partial f(x)$ reduces to a singleton $\{\zeta\}$, then f is strictly differentiable at x and $D_s f(x) = \zeta$.

Proof. Suppose first that $D_s f(x)$ exists (so that f is Lipschitz near x by Proposition 2.2.1). Then, by definition of f°, one has $f^\circ(x; v) = \langle D_s f(x), v\rangle$ for all v, and it follows from Proposition 2.1.5(a) that $\partial f(x)$ reduces to $\{D_s f(x)\}$. To prove the converse, it suffices to show that the condition of Proposition 2.2.1(b) holds for each v in X. We begin by showing that $f^\circ(x; v) = \langle \zeta, v\rangle$ for each v. (Note that $f^\circ(x; v) \geqslant \langle \zeta, v\rangle$ by Proposition 2.1.2(b).) By the Hahn–Banach Theorem there exists $\zeta' \in X^*$ majorized by $f^\circ(x; \cdot)$ and agreeing with $f^\circ(x; \cdot)$ at v. It follows that $\zeta' \in \partial f(x)$, and we have $f^\circ(x; v) = \langle \zeta', v\rangle \geqslant \langle \zeta, v\rangle$. If $\langle \zeta, v\rangle$ were less than $f^\circ(x; v)$, then ζ, ζ' would be distinct elements of $\partial f(x)$, contrary to hypothesis. Thus $f^\circ(x; v) = \langle \zeta, v\rangle$ for all v.

We now calculate:

$$\begin{aligned}
\liminf_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x'+tv) - f(x')}{t} &= -\limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x') - f(x'+tv)}{t} \\
&= -\limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x'+tv-tv) - f(x'+tv)}{t} \\
&= -f^\circ(x; -v) = -\langle \zeta, -v\rangle = \langle \zeta, v\rangle \\
&= f^\circ(x; v) = \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x'+tv) - f(x')}{t}.
\end{aligned}$$

This establishes the limit condition of Proposition 2.2.1(b) and completes the proof. □

Corollary

If f is Lipschitz near x and X is finite-dimensional, then $\partial f(x')$ reduces to a singleton for every x' in $x + \varepsilon B$ iff f is continuously differentiable on $x + \varepsilon B$.

Proof. Invoke the corollary to Proposition 2.2.1 together with Proposition 2.1.5(d); note that for a point-valued map, continuity and upper semicontinuity coincide. □

2.2.5 *Example (Indefinite Integrals)*

Let ϕ: $[0,1] \to R$ belong to $L^\infty[0,1]$, and define a (Lipschitz) function f: $[0,1] \to R$ via $f(x) = \int_0^x \phi(t)\,dt$. Let us calculate $\partial f(x)$. We know that f is differentiable for almost all x, with $f'(x) = \phi(x)$; for any such x, we have $\phi(x) \in \partial f(x)$ by Proposition 2.2.2. It follows from this and from upper semicontinuity (see Proposition 2.1.5) that all essential cluster points of ϕ at x (i.e., those that persist upon the removal of any set of measure zero) belong to $\partial f(x)$.

Let $\phi^+(x)$ and $\phi^-(x)$ denote the essential supremum and essential infimum of ϕ at x. The remarks above and the fact that $\partial f(x)$ is convex imply that $\partial f(x)$ contains the interval $[\phi^-(x), \phi^+(x)]$. From the equality

$$f(y+t) - f(y) = \int_y^{y+t} \phi(s)\,ds$$

one easily deduces that $f^\circ(x;1)$ is bounded above by $\phi^+(x)$. Any ζ in $\partial f(x)$ satisfies $f^\circ(x;1) \geqslant 1\zeta$ (by Proposition 2.1.5(a)), so we derive $\zeta \leqslant \phi^+(x)$. Similarly, $\zeta \geqslant \phi^-(x)$. We arrive at the conclusion

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

Convex Functions

Let U be an open convex subset of X. Recall that a function f: $U \to R$ is said to be *convex* provided that, for all u, u' in U and λ in $[0,1]$, one has

$$f(\lambda u + (1-\lambda)u') \leqslant \lambda f(u) + (1-\lambda) f(u').$$

As we now see, convex functions are Lipschitz except in pathological cases.

2.2.6 *Proposition*

Let f above be bounded above on a neighborhood of some point of U. Then, for any x in U, f is Lipschitz near x.

Proof (Roberts and Varberg, 1974). We begin by proving that f is bounded on a neighborhood of x. Without loss of generality, let us suppose that f is bounded above by M on the set $\varepsilon B \subset U$. Choose $\rho > 1$ so that $y = \rho x$ is in U. If $\lambda = 1/\rho$, then the set

$$V = \{v : v = (1-\lambda)x' + \lambda y,\ x' \in \varepsilon B\}$$

is a neighborhood of $x = \lambda y$ with radius $(1 - \lambda)\varepsilon$. For all v in V, one has by convexity

$$f(v) \leqslant (1 - \lambda)f(x') + \lambda f(y) \leqslant M + \lambda f(y),$$

so that f is bounded above on a neighborhood of x. If z is any point in $x + (1 - \lambda)\varepsilon B$, there is another such point z' such that $x = (z + z')/2$, whence

$$f(x) \leqslant \tfrac{1}{2}f(z) + \tfrac{1}{2}f(z').$$

It follows that

$$f(z) \geqslant 2f(x) - f(z') \geqslant 2f(x) - M - \lambda f(y),$$

so that now f is also seen to be bounded below near x; we have established that f is bounded near x.

Let N be a bound on $|f|$ on the set $x + 2\delta B$, where $\delta > 0$. For distinct x_1, x_2 in $x + \delta B$, set $x_3 = x_2 + (\delta/\alpha)(x_2 - x_1)$ where $\alpha = \|x_2 - x_1\|$, and note that x_3 is in $x + 2\delta B$. Solving for x_2 gives

$$x_2 = \frac{\delta}{\alpha + \delta}x_1 + \frac{\alpha}{\alpha + \delta}x_3,$$

and so by convexity

$$f(x_2) \leqslant \frac{\delta}{\alpha + \delta}f(x_1) + \frac{\alpha}{\alpha + \delta}f(x_3).$$

Then

$$f(x_2) - f(x_1) \leqslant \frac{\alpha}{\alpha + \delta}[f(x_3) - f(x_1)] \leqslant \frac{\alpha}{\delta}|f(x_3) - f(x_1)|,$$

which combined with $|f| \leqslant N$ and $\alpha = \|x_2 - x_1\|$ yields

$$f(x_2) - f(x_1) \leqslant \frac{2N}{\delta}\|x_2 - x_1\|.$$

Since the roles of x_1 and x_2 may be interchanged, we conclude that f is Lipschitz near x. □

The proof showed:

Corollary

Let f be convex with $|f| \leqslant N$ on an open convex set U which contains a δ-neighborhood of a subset V. Then f satisfies a Lipschitz condition of rank $2N/\delta$ on V.

Recall that the *subdifferential* of the (convex) function f at x is defined to be the set of those ζ in X^* satisfying

$$f(x') - f(x) \geqslant \langle \zeta, x' - x \rangle \quad \text{for all } x' \text{ in } U.$$

Since ∂f is the established notation for the subdifferential, the following asserts that (fortunately) $\partial f = \partial f$.

2.2.7 Proposition

When f is convex on U and Lipschitz near x, then $\partial f(x)$ coincides with the subdifferential at x in the sense of convex analysis, and $f^\circ(x; v)$ coincides with the directional derivative $f'(x; v)$ for each v.

Proof. It is known from convex analysis that $f'(x; v)$ exists for each v and that $f'(x; \cdot)$ is the support function of the subdifferential at x. It suffices therefore to prove that for any v, $f^\circ(x; v) = f'(x; v)$. The former can be written as

$$\lim_{\varepsilon \downarrow 0} \sup_{\|x'-x\|<\varepsilon\delta} \sup_{0<t<\varepsilon} \frac{f(x'+tv)-f(x')}{t},$$

where δ is any fixed positive number. It follows readily from the definition of convex function that the function

$$t \to \frac{f(x'+tv)-f(x')}{t}$$

is nondecreasing, whence

$$f^\circ(x; v) = \lim_{\varepsilon \downarrow 0} \sup_{\|x'-x\|<\varepsilon\delta} \frac{f(x'+\varepsilon v)-f(x')}{\varepsilon}.$$

Now by the Lipschitz condition, for any x' in $x + \varepsilon\delta B$, one has

$$\left| \frac{f(x'+\varepsilon v)-f(x')}{\varepsilon} - \frac{f(x+\varepsilon v)-f(x)}{\varepsilon} \right| \leqslant 2\delta K,$$

so that

$$f^\circ(x; v) \leqslant \lim_{\varepsilon \downarrow 0} \frac{f(x+\varepsilon v)-f(x)}{\varepsilon} + 2\delta K = f'(x; v) + 2\delta K.$$

Since δ is arbitrary, we deduce $f^\circ(x; v) \leqslant f'(x; v)$. Of course equality follows, and the proof is complete. □

2.2.8 Example

Let us determine the generalized gradient of the function $f: R^n \to R$ defined by

$$f(x_1, x_2, \ldots, x_n) = \max\{x_i : i = 1, 2, \ldots, n\}.$$

Observe first that f, as a maximum of linear functions, is convex; let us calculate $f'(x; v)$. Let $I(x)$ denote the set of indices which $x_i = f(x)$ (i.e., the indices at which the maximum defining f is attained). We find

$$f'(x; v) := \lim_{t \downarrow 0} \max_i \frac{\{x_i + tv_i\} - f(x)}{t}$$

$$= \lim_{t \downarrow 0} \max_{i \in I(x)} \frac{\{x_i + tv_i\} - f(x)}{t}$$

(since for t small enough, any index i not in $I(x)$ can be ignored in the maximum)

$$= \lim_{t \downarrow 0} \max_{i \in I(x)} \frac{\{x_i + tv_i - x_i\}}{t}$$

$$= \max_{i \in I(x)} v_i.$$

Since f° and f' coincide (Proposition 2.2.7), $\partial f(x)$ consists of those vectors ζ in R^n satisfying

$$\max_{i \in I(x)} v_i \geqslant \zeta \cdot v \quad \text{for all } v \text{ in } R^n.$$

It follows that $\partial f(x)$ consists of all vectors $(\zeta_1, \zeta_2, \ldots, \zeta_n)$ such that $\zeta_i \geqslant 0$, $\Sigma\zeta_i = 1$, $\zeta_i = 0$ if $i \notin I(x)$.

We conclude the discussion by giving a criterion for convexity in terms of the generalized gradient. The proof, which can be based upon the mean value theorem, Theorem 2.3.7, is left as an exercise.

2.2.9 Proposition

Let f be Lipschitz near each point of an open convex subset U of X. Then f is convex on U iff the multifunction ∂f is monotone on U; that is, iff

$$\langle x - x', \zeta - \zeta' \rangle \geqslant 0 \quad \text{for all } x, x' \in U, \zeta \in \partial f(x), \zeta' \in \partial f(x').$$

2.3 BASIC CALCULUS

We now proceed to derive an assortment of formulas that facilitate greatly the calculation of ∂f when (as is often the case) f is "built up" from simple functionals through linear combination, maximization, composition, and so on. We assume that a function f is given which is Lipschitz near a given point x.

2.3.1 Proposition (Scalar Multiples)

For any scalar s, one has

$$\partial(sf)(x) = s\partial f(x).$$

Proof. Note that sf is also Lipschitz near x. When s is nonnegative, $(sf)^\circ = sf^\circ$, and it follows easily that $\partial(sf)(x) = s\partial f(x)$. It suffices now to prove the formula for $s = -1$. An element ζ of X^* belongs to $\partial(-f)(x)$ iff $(-f)^\circ(x; v) \geqslant \langle \zeta, v \rangle$ for all v. By Proposition 2.1.1(c), this is equivalent to: $f^\circ(x; -v) \geqslant \langle \zeta, v \rangle$ for all v, which is equivalent to $-\zeta$ belonging to $\partial f(x)$ (by Proposition 2.1.5(a)). Thus $\zeta \in \partial(-f)(x)$ iff $\zeta \in -\partial f(x)$, as claimed. □

2.3.2 Proposition (Local Extrema)

If f attains a local minimum or maximum at x, then $0 \in \partial f(x)$.

Proof. In view of the formula $\partial(-f) = -\partial f$, it suffices to prove the proposition when x is a local minimum. But in this case it is evident that for any v in X, one has $f^\circ(x; v) \geqslant 0$. Thus $\zeta = 0$ belongs to $\partial f(x)$ (by Proposition 2.1.5(a)). □

If f_i $(i = 1, 2, \dots, n)$ is a finite family of functions each of which is Lipschitz near x, it follows easily that their sum $f = \Sigma f_i$ is also Lipschitz near x.

2.3.3 Proposition (Finite Sums)

$$\partial\left(\sum f_i\right)(x) \subset \sum \partial f_i(x).$$

Proof. Note that the right-hand side denotes the (weak* compact) set of all points ζ obtainable as a sum $\Sigma\zeta_i$ (sum 1 to n), where each ζ_i belongs to $\partial f_i(x)$. It suffices to prove the formula for $n = 2$; the general case follows by induction.

The support functions (on X) of the left- and right-hand sides (evaluated at v) are, respectively, $(f_1 + f_2)^\circ(x; v)$ and $f_1^\circ(x; v) + f_2^\circ(x; v)$ (by Proposition 2.1.2(b)). In view of Proposition 2.1.4(b), it suffices therefore (in fact, it is equivalent) to prove the general inequality

$$(f_1 + f_2)^\circ(x; v) \leqslant f_1^\circ(x; v) + f_2^\circ(x; v).$$

This follows readily from the definition; we leave the verification to the reader. □

Corollary 1

Equality holds in Proposition 2.3.3 *if all but at most one of the functions* f_i *are strictly differentiable at* x.

Proof. By adding all the strictly differentiable functions together to get a single strictly differentiable function, we can reduce the proof to the case where we have two functions f_1 and f_2, with f_1 strictly differentiable. In this case, as is easily seen, we actually have

$$(f_1 + f_2)^\circ(x; v) = f_1'(x; v) + f_2^\circ(x; v) = f_1^\circ(x; v) + f_2^\circ(x; v),$$

so that the two sets in the statement of the proposition have support functions that coincide, and hence are equal. □

Corollary 2

For any scalars s_i, *one has*

$$\partial\left(\sum_{i=1}^{n} s_i f_i\right)(x) \subset \sum_{i=1}^{n} s_i \partial f_i(x),$$

and equality holds if all but at most one of the f_i *are strictly differentiable at* x.

Proof. Invoke Proposition 2.3.1. □

Regularity

It is often the case that calculus formulas for generalized gradients involve inclusions, such as in Proposition 2.3.3. The addition of further hypotheses can serve to sharpen such rules by turning the inclusions to equalities. For instance, equality certainly holds in Proposition 2.3.3 if all the functions in question are continuously differentiable, since then the generalized gradient is essentially the derivative, a linear operator. However, one would wish for a less extreme condition, one that would cover the convex (nondifferentiable) case, for example (in which equality does hold). A class of functions that proves useful in this connection is the following.

2.3.4 Definition

f is said to be *regular* at x provided

(i) For all v, the usual one-sided directional derivative $f'(x; v)$ exists.
(ii) For all v, $f'(x; v) = f^\circ(x; v)$.

To see how regularity contributes, let us note the following addendum to Proposition 2.3.3.

Corollary 3

If each f_i is regular at x, equality holds in Proposition 2.3.3. *Equality then holds in Corollary* 2 *as well, if in addition each s_i is nonnegative.*

Proof. Since, as we shall see below, a positive linear combination of regular functions is regular, it suffices again (in both cases) to consider the case $n = 2$. The proof made clear that equality would hold provided the two sets had the same support functions. The support function of the one on the left is $(f_1 + f_2)^\circ(x; \cdot) = (f_1 + f_2)'(x; \cdot) = f_1'(x; \cdot) + f_2'(x; \cdot) = f_1^\circ(x; \cdot) + f_2^\circ(x; \cdot)$, which is the support function of the one on the right. □

2.3.5 Remark

In view of the formula $\partial(-f) = -\partial f$, equality also holds in Proposition 2.3.3 if, for each i, $-f_i$ is regular at x. (In the future we shall often leave implicit such dual results.) To see that the inclusion can be strict in general, it suffices to take any function f such that $\partial f(x)$ is not a singleton, and set $f_1 = f$, $f_2 = -f$.

Here are some first observations about regular functions.

2.3.6 Proposition

Let f be Lipschitz near x.

(a) *If f is strictly differentiable at x, then f is regular at x.*
(b) *If f is convex, then f is regular at x.*
(c) *A finite linear combination (by nonnegative scalars) of functions regular at x is regular at x.*
(d) *If f admits a Gâteaux derivative $Df(x)$ and is regular at x, then $\partial f(x) = \{Df(x)\}$.*

Proof. (a) follows from Propositions 2.1.5(a) and 2.2.4, since $f^\circ(x; v)$ is the support function of $\partial f(x) = \{D_s f(x)\}$, and one has the equality $D_s f(x)(v) = f'(x; v)$. To prove (b) we invoke convex analysis, which asserts that $f'(x; \cdot)$ exists, and is the support function of the subdifferential; in view of Proposition 2.2.7, we derive $f'(x; \cdot) = f^\circ(x; \cdot)$ as required. As for (c), once again only the case of two functions f_1 and f_2 need be treated, and since sf is clearly regular when f is regular and $s \geqslant 0$, it suffices to prove $(f_1 + f_2)' = (f_1 + f_2)^\circ$ when f_1 and f_2 are regular at x (the existence of $(f_1 + f_2)'$ being evident). We have $(f_1 + f_2)' = f_1' + f_2' = f_1^\circ + f_2^\circ \geqslant (f_1 + f_2)^\circ$ (as in the proof of Proposition 2.3.3). Since we always have the opposite inequality $(f_1 + f_2)^\circ \geqslant (f_1 + f_2)'$, we arrive at (c). The assertion (d) is evident from Proposition 2.1.5(a). □

We shall extend (c) in Section 2.7 to sums which are not necessarily finite, as part of a more general study of integral functionals. Given x and y in X, the notation $[x, y]$ signifies the closed line segment consisting of all points $tx + (1 - t)y$ for $t \in [0, 1]$; (x, y) signifies the open line segment.

Mean-Value Theorem

2.3.7 Theorem (Lebourg)

Let x and y be points in X, and suppose that f is Lipschitz on an open set containing the line segment $[x, y]$. Then there exists a point u in (x, y) such that

$$f(y) - f(x) \in \langle \partial f(u), y - x \rangle. \tag{1}$$

We shall need the following special chain rule for the proof. We denote by x_t the point $x + t(y - x)$.

Lemma. The function $g: [0, 1] \to R$ defined by $g(t) = f(x_t)$ is Lipschitz on $(0, 1)$, and we have

$$\partial g(t) \subset \langle \partial f(x_t), y - x \rangle. \tag{2}$$

Proof. The fact that g is Lipschitz is plain. The two closed convex sets appearing in Eq. (2) are in fact intervals in R, so it suffices to prove that for $v = \pm 1$, we have

$$\max\{\partial g(t) v\} \leqslant \max\{\langle \partial f(x_t), y - x \rangle v\}.$$

Now the left-hand side is just $g^\circ(t; v)$; that is,

$$\limsup_{\substack{s \to t \\ \lambda \downarrow 0}} \frac{g(s + \lambda v) - g(s)}{\lambda}$$

$$= \limsup_{\substack{s \to t \\ \lambda \downarrow 0}} \frac{f(x + [s + \lambda v](y - x)) - f(x + s(y - x))}{\lambda}$$

$$\leqslant \limsup_{\substack{y' \to x_t \\ \lambda \downarrow 0}} \frac{f(y' + \lambda v(y - x)) - f(y')}{\lambda}$$

$$= f^\circ(x_t; v(y - x))$$

$$= \max\langle \partial f(x_t), v(y - x) \rangle.$$

Now to the proof of the theorem. Consider the function θ on $[0, 1]$ defined by

$$\theta(t) = f(x_t) + t[f(x) - f(y)].$$

Note $\theta(0) = \theta(1) = f(x)$, so that there is a point t in $(0, 1)$ at which θ attains a local minimum or maximum (by continuity). By Proposition 2.3.2, we have

$0 \in \partial\theta(t)$. We may calculate the latter by appealing to Propositions 2.3.1, 2.3.3, and the above lemma. We deduce

$$0 \in [f(x) - f(y)] + \langle \partial f(x_t), y - x \rangle,$$

which is the assertion of the theorem (take $u = x_t$). □

2.3.8 Example

We shall find the mean-value theorem very useful in what is to come. As an immediate illustration of its use, consider the function

$$f(x) = \int_0^x \phi(t)\, dt$$

of Example 2.2.5. If we combine the mean-value theorem with the expression for ∂f, we arrive at the following. For any $\varepsilon > 0$, there are points x and y in $(0, 1)$ such that $|x - y| < \varepsilon$ and such that

$$\phi(x) - \varepsilon < \int_0^1 \phi(t)\, dt < \phi(y) + \varepsilon.$$

Chain Rules

We now turn to a very useful chain rule that pertains to the following situation: $f = g \circ h$, where $h\colon X \to R^n$ and $g\colon R^n \to R$ are given functions. The component functions of h will be denoted h_i $(i = 1, 2, \dots, n)$. We assume that each h_i is Lipschitz near x and that g is Lipschitz near $h(x)$; this implies that f is Lipschitz near x (as usual). If we adopt the convention of identifying $(R^n)^*$ with R^n, then an element α of ∂g can be considered an n-dimensional vector: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$. (All sums below are from 1 to n.)

2.3.9 Theorem (Chain Rule I)

One has

$$\partial f(x) \subset \overline{\text{co}}\left\{\sum \alpha_i \zeta_i : \zeta_i \in \partial h_i(x), \alpha \in \partial g(h(x))\right\}$$

(where $\overline{\text{co}}$ denotes weak-closed convex hull), and equality holds under any one of the following additional hypotheses:*

(i) *g is regular at $h(x)$, each h_i is regular at x, and every element α of $\partial g(h(x))$ has nonnegative components. (In this case it follows that f is regular at x.)*

(ii) *g is strictly differentiable at $h(x)$ and $n = 1$. (In this case the $\overline{\text{co}}$ is superfluous.)*

(iii) *g is regular at $h(x)$ and h is strictly differentiable at x. (In this case it follows that f is regular at x, and the $\overline{\text{co}}$ is superfluous.)*

Proof. The set S whose convex hull is taken in the formula is weak*-compact, and therefore its convex hull has compact closure (and is in fact closed if X is finite-dimensional, so that $\overline{\text{co}}$ may be replaced by co in that case). The support function (of either S or $\overline{\text{co}}\, S$) evaluated at a point v in X is easily seen to be given by the quantity

$$(3) \qquad q_0 := \max\left\{\sum \alpha_i \langle \zeta_i, v\rangle : \zeta_i \in \partial h_i(x), \alpha \in \partial g(h(x))\right\}.$$

It suffices (by Proposition 2.1.4(b)) to prove that q_0 majorizes $f^\circ(x; v)$ for any v. To show this, it suffices in turn to show that for any $\varepsilon > 0$, the following quantity q_ε majorizes $f^\circ(x; v) - \varepsilon$:

$$\max\left\{\sum \alpha_i \langle \zeta_i, v\rangle : \zeta_i \in \partial h_i(x_i), \alpha \in \partial g(u), x_i \in x + \varepsilon B, u \in h(x) + \varepsilon B\right\}.$$

The reason for this is that, in light of a lemma that we shall provide later, q_ε decreases to q_0 as $\varepsilon \downarrow 0$.

From the definition of $f^\circ(x; v)$ it follows that we can find a point x' near x and a positive t near 0 such that

$$(4) \qquad f^\circ(x; v) \leqslant \frac{f(x' + tv) - f(x')}{t} + \varepsilon.$$

The degree of nearness is chosen to guarantee

$$x' \in x + \varepsilon B, \qquad x' + tv \in x + \varepsilon B, \qquad h(x') \in h(x) + \varepsilon B,$$
$$h(x' + tv) \in h(x) + \varepsilon B.$$

By the mean-value theorem 2.3.7 we may write

$$\begin{aligned} f(x' + tv) - f(x') &= g(h(x' + tv)) - g(h(x')) \\ &= \sum \alpha_i [h_i(x' + tv) - h_i(x')], \end{aligned}$$

where $\alpha \in \partial g(u)$ and u is a point in the segment $[h(x' + tv), h(x')]$ (and hence in $h(x) + \varepsilon B$). By another application of the mean-value theorem, the last term above can be expressed as

$$\sum \alpha_i \langle \zeta_i, tv\rangle,$$

where $\zeta_i \in \partial h_i(x_i)$ and x_i is a point in the segment $[x' + tv, x']$ (and hence in $x + \varepsilon B$). These observations combine with Eq. (4) to imply

$$f^\circ(x; v) \leqslant \sum \alpha_i \langle \zeta_i, v\rangle + \varepsilon,$$

which clearly yields $f^\circ(x; v) \leqslant q_\varepsilon + \varepsilon$ as required.

Here is the missing link we promised:

Lemma. $\lim_{\varepsilon\downarrow 0} q_\varepsilon = q_0$. To see this, let any $\delta > 0$ be given, and let K be a common local Lipschitz constant for the functions h_i, g. We shall show that q_ε is bounded above by $q_0 + n\delta(1 + K|v|)$ for all ε sufficiently small, which yields the desired result (since $q_\varepsilon \geqslant q_0$).

Choose ε so that each h_i is Lipschitz of rank K on $x + \varepsilon B$, and so that, for each index i, for any x_i in $x + \varepsilon B$, one has

$$h_i^\circ(x_i; \pm v) \leqslant h_i^\circ(x; \pm v) + \frac{\delta}{K}.$$

Multiplying across by $|\alpha_i|$, where α_i is the ith component of any element in $\partial g(h(x) + \varepsilon B)$

$$h_i^\circ(x_i; \zeta_i v) \leqslant h_i^\circ(x; \alpha_i v) + \delta$$

since $|\alpha_i| \leqslant K$. By Proposition 2.1.5(d), we can also choose ε small enough to guarantee that $\partial g(h(x) + \varepsilon B)$ is contained in $\partial g(h(x)) + \delta B$. We may now calculate as follows:

$$\begin{aligned} &q_\varepsilon \leqslant \\ &\max\left\{\sum_i \max[\alpha_i\langle\zeta_i, v\rangle : \zeta_i \in \delta h_i(x_i), x_i \in x + \varepsilon B] : \alpha \in \partial g(h(x)) + \delta B\right\} \\ &\leqslant \max\left\{\sum_i \left(h_i^\circ(x; \alpha_i v) + \delta\right) : \alpha \in \partial g(x) + \delta B\right\} \\ &\leqslant \max\left\{\sum_i \max[\alpha_i\langle\zeta_i, v\rangle : \zeta_i \in \partial h_i(x)] : \alpha \in \partial g(x) + \delta B\right\} + n\delta \\ &\leqslant q_0 + n\,\delta K|v| + n\delta, \end{aligned}$$

which completes the proof of the lemma.

Having completed the proof of the general formula, we turn now to the additional assertions, beginning with the situation depicted by (i).

Consider the quantity q_0 defined by Eq. (3). We have, since the α are all nonnegative,

$$\begin{aligned} q_0 &= \max\left\{\sum \alpha_i \max\{\langle\zeta_i, v\rangle : \zeta_i \in \partial h_i(x)\} : \alpha \in \partial g(h(x))\right\} \\ &= \max\left\{\sum \alpha_i h_i'(x; v) : \alpha \in \partial g(h(x))\right\} \\ &= g'(h(x); w), \quad \text{where } w_i := h_i'(x; v) \\ &= \lim_{t\downarrow 0} \frac{g(h(x) + tw) - g(h(x))}{t} \\ &= \lim_{t\downarrow 0} \left\{ \frac{g(h(x + tv)) - g(h(x))}{t} + \frac{g(h(x) + tw) - g(h(x + tv))}{t} \right\}. \end{aligned}$$

(note that the second term under the limit goes to 0 because g is Lipschitz near $h(x)$ and $w - [h(x+tv) - h(x)]/t$ goes to 0)

$$= \lim_{t \downarrow 0} \frac{g(h(x+tv)) - g(h(x))}{t} = f'(x; v).$$

Stepping back, let us note that we have just proven $q_0 \leqslant f'(x; v)$, and that earlier we proved $f^\circ(x; v) \leqslant q_0$. Since $f'(x; v) \leqslant f^\circ(x; v)$ always, we deduce that $q_0 = f'(x; v) = f^\circ(x; v)$ (thus f is regular). Recall that q_0 is the support function (evaluated at v) of the right-hand side of the general formula in the theorem, and that $f^\circ(x; \cdot)$ is that of $\partial f(x)$. The equality follows, and (i) is disposed of. The reader may verify that the preceding argument will adapt to case (iii) as well. This leaves case (ii), in which $D_s g(h(x))$ is a scalar α. We may assume $\alpha \geqslant 0$. We find, much as above,

$$\begin{aligned} q_0 &= \alpha h_1^\circ(x; v) \\ &= \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{\alpha[h(x'+tv) - h(x')]}{t} \\ &= \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{g(h(x'+tv)) - g(h(x'))}{t} \end{aligned}$$

(an immediate consequence of the strict differentiability of g at $h(x)$)

$$= f^\circ(x; v).$$

As before, this yields the equality in the theorem. The $\overline{\text{co}}$ is superfluous in cases (ii) and (iii), since when $\partial g(h(x))$ or else each $\partial h_i(x)$ consists of a single point, the set whose convex hull is being taken is already convex (it fails to be so in general) and closed. □

2.3.10 Theorem (Chain Rule II)

Let F be a map from X to another Banach space Y, and let g be a real-valued function on Y. Suppose that F is strictly differentiable at x and that g is Lipschitz near $F(x)$. Then $f = g \circ F$ is Lipschitz near x, and one has

$$\partial f(x) \subset \partial g(F(x)) \circ D_s F(x). \tag{5}$$

Equality holds if g (or $-g$) is regular at $F(x)$, in which case f (or $-f$) is also regular at x. Equality also holds if F maps every neighborhood of x to a set which is dense in a neighborhood of $F(x)$ (for example, if $D_s F(x)$ is onto).

2.3.11 Remark

The meaning of (5), of course, is that every element z of $\partial f(x)$ can be represented as a composition of a map ζ in $\partial g(F(x))$ and $D_s F(x)$: $\langle z, v\rangle =$

$\langle \zeta, D_s F(x)(v)\rangle$ for all v in X. If * denotes adjoint, we may write (5) in the equivalent form

$$\partial f(x) \subset [D_s F(x)]^* \partial g(F(x)).$$

Proof. The fact that f is Lipschitz at x is straightforward. Once again (5) is an inclusion between convex, weak*-compact sets. In terms of support functions, therefore, (5) is equivalent to the following inequality being true for any v (we let $A = D_s F(x)$):

$$\begin{aligned}(6)\qquad f^\circ(x; v) &\leqslant \max\{\langle z, Av\rangle : z \in \partial g(F(x))\}\\ &= g^\circ(F(x); Av).\end{aligned}$$

The proof of this is quite analogous to that of Theorem 2.3.9, and we omit it.

Now suppose that g is regular. (The case in which $-g$ is regular is then handled by looking at $-f$ and invoking $\partial(-f) = -\partial f$.) Then the last term in (6) coincides with $g'(F(x); Av)$, which is

$$\lim_{t\downarrow 0} \frac{g(F(x) + tAv) - g(F(x))}{t} = \lim_{t\downarrow 0} \frac{g(F(x + tv)) - g(F(x))}{t}$$

(since $[F(x + tv) - F(x) - tAv]/t$ goes to 0 with t and g is Lipschitz)

$$= \lim_{t\downarrow 0} \frac{f(x + tv) - f(x)}{t}$$

$$= f'(x; v) \leqslant f^\circ(x; v).$$

This shows that f' exists and establishes the opposite inequality to that of (6). Thus equality actually holds above (so f is regular) and in (5).

Finally, suppose that F maps every neighborhood of x to a set dense in a neighborhood of $F(x)$. This permits us to write the final term in (6) as follows:

$$\begin{aligned} g^\circ(F(x); Av) &= \limsup_{\substack{y\to F(x)\\ t\downarrow 0}} \frac{g(y + tAv) - g(y)}{t}\\ &= \limsup_{\substack{x'\to x\\ t\downarrow 0}} \frac{g(F(x') + tAv) - g(F(x'))}{t}\\ &= \limsup_{\substack{x'\to x\\ t\downarrow 0}} \frac{g(F(x' + tv)) - g(F(x'))}{t}\end{aligned}$$

(since $[F(x' + tv) - F(x') - tAv]/t$ goes to zero as $x' \to x$ and $t \downarrow 0$, and g is

Lipschitz)

$$= f^\circ(x; v).$$

As before, this establishes the reverse inequality to (6), and equality in (5) ensues (but not, this time, that f is regular). □

Corollary

Let $g: Y \to R$ be Lipschitz near x, and suppose the space X is continuously imbedded in Y, is dense in Y, and contains the point x. Then the restriction f of g to X is Lipschitz near x, and $\partial f(x) = \partial g(x)$, in the sense that every element z of $\partial f(x)$ admits a unique extension to an element of $\partial g(x)$.

Proof. Apply the theorem with F being the imbedding map from X to Y. □

Suppose now that $\{f_i\}$ is a finite collection of functions ($i = 1, 2, \ldots, n$) each of which is Lipschitz near x. The function f defined by

$$f(x') = \max\{f_i(x') : i = 1, 2, \ldots, n\}$$

is easily seen to be Lipschitz near x as well. For any x' we let $I(x')$ denote the set of indices i for which $f_i(x') = f(x')$ (i.e., the indices at which the maximum defining f is attained).

2.3.12 Proposition (Pointwise Maxima)

$$\partial f(x) \subset \text{co}\{\partial f_i(x) : i \in I(x)\},$$

and if f_i is regular at x for each i in $I(x)$, then equality holds and f is regular at x.

Proof. Define $g: R^n \to R$ via

$$g(u_1, u_2, \ldots, u_n) = \max\{u_i : i = 1, 2, \ldots, n\},$$

and define $h: X \to R^n$ via

$$h(x) = [f_1(x), f_2(x), \ldots, f_n(x)].$$

Observe that $f = g \circ h$. It suffices now to apply Theorem 2.3.9 and the characterization of ∂g from Example 2.2.8. Because g is convex, it is regular at $h(x)$ (see Proposition 2.3.6). The assertions regarding equality and regularity follow from Theorem 2.3.9(i). (We may assume that $I(x) = \{1, 2, \ldots, n\}$, for dropping any f_i for which $i \notin I(x)$ has no bearing on f locally; the closure operation is superfluous here.) □

2.3.13 Proposition (Products)

Let f_1, f_2 be Lipschitz near x. Then $f_1 f_2$ is Lipschitz near x, and one has

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

If in addition $f_1(x) \geqslant 0$, $f_2(x) \geqslant 0$ and if f_1, f_2 are both regular at x, then equality holds and $f_1 f_2$ is regular at x.

Proof. Let g: $R^2 \to R$ be the function $g(u_1, u_2) = u_1 \cdot u_2$, and let h: $X \to R^2$ be the function

$$h(x) = [f_1(x), f_2(x)].$$

Note that $f_1 f_2 = g \circ h$. It suffices now to apply Theorem 2.3.9; condition (i) of that result applies to yield the additional assertions. □

The next result is proven very similarly:

2.3.14 Proposition (Quotients)

Let f_1, f_2 be Lipschitz near x, and suppose $f_2(x) \neq 0$. Then f_1/f_2 is Lipschitz near x, and one has

$$\partial\left(\frac{f_1}{f_2}\right)(x) \subset \frac{f_2(x)\,\partial f_1(x) - f_1(x)\,\partial f_2(x)}{f_2^2(x)}.$$

If in addition $f_1(x) \geqslant 0$, $f_2(x) > 0$ and if f_1 and $-f_2$ are regular at x, then equality holds and f_1/f_2 is regular at x.

Partial Generalized Gradients

Let $X = X_1 \times X_2$, where X_1, X_2 are Banach spaces, and let $f(x_1, x_2)$ on X be Lipschitz near (x_1, x_2). We denote by $\partial_1 f(x_1, x_2)$ the (partial) generalized gradient of $f(\cdot, x_2)$ at x_1, and by $\partial_2 f(x_1, x_2)$ that of $f(x_1, \cdot)$ at x_2. The notation $f_1^\circ(x_1, x_2; v)$ will represent the generalized directional derivative at x_1 in the direction $v \in X_1$ of the function $f(\cdot, x_2)$. It is a fact that in general neither of the sets $\partial f(x_1, x_2)$ and $\partial_1 f(x_1, x_2) \times \partial_2 f(x_1, x_2)$ need be contained in the other; an example will be given in Section 2.5. For regular functions, however, a general relationship does hold between these sets.

2.3.15 Proposition

If f is regular at $x = (x_1, x_2)$, then

$$\partial f(x_1, x_2) \subset \partial_1 f(x_1, x_2) \times \partial_2 f(x_1, x_2).$$

Proof. Let $z = (z_1, z_2)$ belong to $\partial f(x_1, x_2)$. It suffices to prove that z_1 belongs to $\partial_1 f(x_1, x_2)$, which in turn is equivalent to proving that for any v in

X_1, one has $\langle z_1, v\rangle \leqslant f_1^\circ(x_1, x_2; v)$. But the latter coincides with $f_1'(x_1, x_2; v) = f'(x_1, x_2; v, 0) = f^\circ(x_1, x_2; v, 0)$, which majorizes $\langle z_1, v\rangle$ (by Proposition 2.1.2(b)). □

To obtain a relationship in the nonregular case, define the projection $\pi_1 \partial f(x_1, x_2)$ as the set

$$\{z_1 \in X_1^* : \text{for some } z_2 \in X_2^*, (z_1, z_2) \in \partial f(x_1, x_2)\},$$

and analogously for $\pi_2 \partial f(x_1, x_2)$.

2.3.16 Proposition

$$\partial_1 f(x_1, x_2) \subset \pi_1 \partial f(x_1, x_2).$$

Proof. Fix x_2, and let f_1 be the function $f_1(x) = f(x, x_2)$ on X_1. Let F: $X_1 \to X_1 \times X_2$ be defined by $F(x) = (x, x_2)$, and note that $D_s F(x)$ is given by $\langle D_s F(x), v\rangle = (v, 0)$. Since $f_1 = f \circ F$, the result now follows by applying Theorem 2.3.10, with $x = x_1$. □

Corollary

$$\partial_1 f(x_1, x_2) \times \partial_2 f(x_1, x_2) \subset \pi_1 \partial f(x_1, x_2) \times \pi_2 \partial f(x_1, x_2).$$

2.3.17 Example

Let us prove the assertion made in Section 1.1 (see Figure 1.1) about the best L^1-approximating straight line. Specifically, let the $N + 1$ data points be $(0,0), (1,1), \ldots, (N-1, N-1)$ together with $(N, 0)$. Recall that the problem becomes that of minimizing the following function on R^2:

$$f(\alpha, \beta) = |\alpha N + \beta| + \sum_{i=1}^{N-1} |\alpha i + \beta - i|.$$

Note first that the function $f_{c,k}$ defined by

$$f_{c,k}(\alpha, \beta) = |\alpha c + \beta - k|$$

is the composition of g and F, where

$$g(y) = |y|, \qquad F(\alpha, \beta) = \alpha c + \beta - k.$$

Note that F is strictly differentiable, with $D_s F(\alpha, \beta) = [c, 1]$. Since g is regular (in fact, convex), it follows from Theorem 2.3.10 that one has

$$\partial f_{c,k}(\alpha, \beta) = \begin{cases} \{[c, 1]\} & \text{if } \alpha c + \beta - k > 0 \\ \{[-c, -1]\} & \text{if } \alpha c + \beta - k < 0 \\ \{\lambda[c, 1] : |\lambda| \leqslant 1\} & \text{if } \alpha c + \beta - k = 0. \end{cases}$$

There is a point (α, β) at which f is minimized, and it must satisfy $0 \in \partial f(\alpha, \beta)$ by Proposition 2.3.2. The preceding remarks and Proposition 2.3.3 allow us to interpret this as follows:

$$0 = \lambda_N [N, 1] + \sum_{i=0}^{N-1} \lambda_i [i, 1]. \tag{7}$$

If this necessary condition holds for the line $y = x$ (i.e., for $\alpha = 1$, $\beta = 0$), then $\lambda_N = 1$, and $|\lambda_i| \leqslant 1$ ($i = 0, \ldots, N-1$). It is easy to see that if $N \geqslant 3$, then there exist such λ_i's satisfying Eq. (7). Since f is convex, the condition $0 \in \partial f(1, 0)$ is also sufficient to imply that the line $y = x$ is a solution to the problem. Any other candidate would pass through at most two of the data points; the reader is invited to show that Eq. (7) cannot hold in such a case if N exceeds 4.

2.4 ASSOCIATED GEOMETRIC CONCEPTS

The Distance Function

Let C be a nonempty subset of X, and consider its *distance function*; that is, the function $d_C(\cdot): X \to R$ defined by

$$d_C(x) = \inf\{\|x - c\| : c \in C\}.$$

If C happens to be closed (which we do not assume), then $x \in C$ iff $d_C(x) = 0$. The function d_C is certainly not differentiable in any of the standard senses; it is, however, globally Lipschitz, as we are about to prove. We shall use the generalized gradient of d_C to lead us to new concepts of tangents and normals to an arbitrary set C. Subsequently, we shall characterize these normals and tangents topologically, thus making it clear that they do not depend on the particular norm (or distance function) that we are using. We shall prove that the new tangents and normals defined here reduce to the known ones in the smooth or convex cases. Finally, we shall indicate how these geometrical ideas lead to an extended definition of the generalized gradient ∂f of a function f which is not necessarily locally Lipschitz, and possibly extended-valued.

2.4.1 Proposition

The function d_C satisfies the following global Lipschitz condition on X:

$$|d_C(x) - d_C(y)| \leqslant \|x - y\|.$$

Proof. Let any $\varepsilon > 0$ be given. By definition, there is a point c in C such that $d_C(y) \geqslant \|y - c\| - \varepsilon$. We have

$$d_C(x) \leqslant \|x - c\| \leqslant \|x - y\| + \|y - c\|$$
$$\leqslant \|x - y\| + d_C(y) + \varepsilon.$$

Since ε is arbitrary, and since the argument can be repeated with x and y switched, the result follows. □

Tangents

Suppose now that x is a point in C. A vector v in X is *tangent* to C at x provided $d_C^\circ(x; v) = 0$. The set of all tangents to C at x is denoted $T_C(x)$. Of course, only the local nature of C near x is involved in this definition.

It is an immediate consequence of Proposition 2.1.1 that $T_C(x)$ is a closed convex cone in X (in particular, $T_C(x)$ always contains 0).

Normals

We define the *normal cone* to C at x by polarity with $T_C(x)$:

$$N_C(x) = \{\zeta \in X^* : \langle \zeta, v\rangle \leqslant 0 \text{ for all } v \text{ in } T_C(x)\}.$$

We have the following alternate characterization of $N_C(x)$ in terms of generalized gradients:

2.4.2 Proposition

$$N_C(x) = \operatorname{cl}\left\{ \bigcup_{\lambda \geqslant 0} \lambda\, \partial d_C(x)\right\},$$

where cl denotes weak closure.*

Proof. The definition of $T_C(x)$, along with Proposition 2.1.2(b), implies that v belongs to $T_C(x)$ iff $\langle v, \zeta\rangle \leqslant 0$ for every ζ in $\partial d_C(x)$. It follows that the cone polar to $T_C(x)$ is the weak*-closed, convex cone generated by $\partial d_C(x)$, which is the statement of the proposition. □

The distance function figures in later chapters on optimization largely because of its role in "exact penalization." Here is a first such result:

2.4.3 Proposition

Let f be Lipschitz of rank K on a set S. Let x belong to a set $C \subset S$ and suppose that f attains a minimum over C at x. Then for any $\hat{K} \geqslant K$, the function

$g(y) = f(y) + \hat{K}d_C(y)$ attains a minimum over S at x. If $\hat{K} > K$ and C is closed, then any other point minimizing g over S must also lie in C.

Proof. Let us prove the first assertion by supposing the contrary. Then there is a point y in S and $\varepsilon > 0$ such that $f(y) + \hat{K}d_C(y) < f(x) - \hat{K}\varepsilon$. Let c be a point in C such that $\|y - c\| \leqslant d_C(y) + \varepsilon$. Then

$$f(c) \leqslant f(y) + \hat{K}\|y - c\| \leqslant f(y) + \hat{K}(d_C(y) + \varepsilon) < f(x),$$

which contradicts the fact that x minimizes f over C. Now let $\hat{K} > K$, and let y also minimize g over S. Then

$$f(y) + \hat{K}d_C(y) = f(x) \leqslant f(y) + (K + \hat{K})d_C(y)/2$$

(by the first assertion applied to $(\hat{K} + K)/2$), which implies $d_C(y) = 0$, and hence that y belongs to C. □

Corollary

Suppose that f is Lipschitz near x and attains a minimum over C at x. Then $0 \in \partial f(x) + N_C(x)$.

Proof. Let S be a neighborhood of x upon which f is Lipschitz (of rank K). We may suppose $C \subset S$ (since C and $C \cap S$ have the same normal cones at x), so by the proposition we deduce that x minimizes $f(y) + Kd_C(y)$ locally. Thus

$$0 \in \partial(f + Kd_C)(x) \subset \partial f(x) + K\partial d_C(x).$$

The result now follows from Proposition 2.4.2. □

When C is convex, there is a well-known concept of normal vector: $\zeta \in X^*$ is said to be normal to C at x (in the sense of convex analysis) provided $\langle \zeta, x - c \rangle \geqslant 0$ for all $c \in C$.

2.4.4 Proposition

If C is convex, $N_C(x)$ coincides with the cone of normals in the sense of convex analysis.

Proof. Let ζ be normal to C at x as per convex analysis. Then the point $c = x$ minimizes $f(c) = \langle \zeta, x - c \rangle$ over C, so that by the corollary to Proposition 2.4.3 (and since f is continuously, or strictly, differentiable) we obtain

$$0 \in -\zeta + N_C(x),$$

which shows that ζ belongs to $N_C(x)$.

To complete the proof, it will suffice, in view of Proposition 2.4.2, to prove that any element ζ of $\partial d_C(x)$ is normal to C at x in the sense of convex analysis (since the set of such normals is a weak*-closed convex cone).

Lemma. $d_C(\cdot)$ is convex.

Let x, y in X and λ in $(0,1)$ be given. For any $\varepsilon > 0$, pick c_x, c_y in C such that

$$\|c_x - x\| \leqslant d_C(x) + \varepsilon, \qquad \|c_y - y\| \leqslant d_C(y) + \varepsilon,$$

and let c in C be given by $c = \lambda c_x + (1-\lambda)c_y$. Then

$$\begin{aligned} d_C(\lambda x + (1-\lambda)y) &\leqslant \|c - \lambda x - (1-\lambda)y\| \\ &\leqslant \lambda\|c_x - x\| + (1-\lambda)\|c_y - y\| \\ &\leqslant \lambda d_C(x) + (1-\lambda)d_C(y) + \varepsilon. \end{aligned}$$

Since ε is arbitrary, the lemma is proved.

We now resume the proof of the proposition. By Proposition 2.2.7, ζ is a subgradient of d_C at x; that is,

$$d_C(y) - d_C(x) \geqslant \langle \zeta, y - x\rangle \qquad \text{for all } y \text{ in } X.$$

Thus $\langle \zeta, c - x\rangle \leqslant 0$ for any c in C. □

Corollary

If C is convex, then $v \in T_C(x)$ iff $d_C^0(x; v) = d_C'(x; v) = 0$.

Proof. The proof showed that d_C is convex, and therefore regular by Proposition 2.3.6. Thus d_C^0 and d_C' coincide, and the result follows. □

An Intrinsic Characterization of $T_C(x)$

We now show that the tangency concept defined above is actually independent of the norm (and hence distance function) used on X. Knowing this allows us to choose in particular circumstances a distance function which makes calculating tangents (or normals) more convenient.

2.4.5 Theorem

An element v of X is tangent to C at x iff, for every sequence x_i in C converging to x and sequence t_i in $(0, \infty)$ decreasing to 0, there is a sequence v_i in X converging to v such that $x_i + t_i v_i \in C$ for all i.

Proof. Suppose first that $v \in T_C(x)$, and that sequences $x_i \to x$ (with $x_i \in C$), $t_i \downarrow 0$ are given. We must produce the sequence v_i alluded to in the statement of the theorem. Since $d_C^0(x; v) = 0$ by assumption, we have

$$(1) \qquad \lim_{i\to\infty} \frac{d_C(x_i + t_i v) - d_C(x_i)}{t_i} = \lim_{i\to\infty} \frac{d_C(x_i + t_i v)}{t_i} = 0.$$

Let c_i be a point in C which satisfies

$$(2) \qquad \|x_i + t_i v - c_i\| \leqslant d_C(x_i + t_i v) + \frac{t_i}{i}$$

and let us set

$$v_i = \frac{c_i - x_i}{t_i}.$$

Then (1) and (2) imply that $\|v - v_i\| \to 0$; that is, that v_i converges to v. Furthermore, $x_i + t_i v_i = c_i \in C$, as required.

Now for the converse. Let v have the stated property concerning sequences, and choose a sequence y_i converging to x and t_i decreasing to 0 such that

$$(3) \qquad \lim_{i\to\infty} \frac{d_C(y_i + t_i v) - d_C(y_i)}{t_i} = d_C^\circ(x; v).$$

Our purpose is to prove this quantity nonpositive, for then $v \in T_C(x)$ by definition. Let c_i in C satisfy

$$(4) \qquad \|c_i - y_i\| \leqslant d_C(y_i) + \frac{t_i}{i}.$$

It follows that c_i converges to x. Thus there is a sequence v_i converging to v such that $c_i + t_i v_i \in C$. But then, since d_C is Lipschitz,

$$\begin{aligned} d_C(y_i + t_i v) &\leqslant d_C(c_i + t_i v_i) + \|y_i - c_i\| + t_i\|v - v_i\| \\ &\leqslant d_C(y_i) + t_i\left(\|v - v_i\| + \frac{1}{i}\right) \qquad \text{(by (4))}. \end{aligned}$$

We deduce that the limit (3) is nonpositive, which completes the proof. □

Corollary

Let $X = X_1 \times X_2$, where X_1, X_2 are Banach spaces, and let $x = (x_1, x_2) \in C_1 \times C_2$, where C_1, C_2 are subsets of X_1, X_2, respectively. Then

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

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

Proof. The first equality follows readily from the characterization of the tangent cone given in the theorem, and the second then follows by polarity. □

Regularity of Sets

In order to establish the relationship between the geometric concepts defined above and the previously known ones in smooth contexts, we require a notion of regularity for sets which will play the role that regularity for functions played in Section 2.3. We recall first the *contingent cone* $K_C(x)$ of tangents to a set C at a point x. A vector v in X belongs to $K_C(x)$ iff, for all $\varepsilon > 0$, there exist t in $(0, \varepsilon)$ and a point w in $v + \varepsilon B$ such that $x + tw \in C$ (thus $x \in \operatorname{cl} C$ necessarily). It follows immediately from Theorem 2.4.5 that $T_C(x)$ is always contained in $K_C(x)$. The latter may not be convex, however.

2.4.6 Definition

The set C is *regular* at x provided $T_C(x) = K_C(x)$.

Any convex set is regular at each of its points by the corollary to Proposition 2.4.4. The second corollary to the following theorem will confirm the fact that N_C and T_C reduce to the classical notions when C is a "smooth" set.

2.4.7 Theorem

Let f be Lipschitz near x, and suppose $0 \notin \partial f(x)$. If C is defined as $\{y \in X\colon f(y) \leqslant f(x)\}$, then one has

$$(5) \qquad \{v \in X\colon f^\circ(x; v) \leqslant 0\} \subset T_C(x).$$

If f is regular at x, then equality holds, and C is regular at x.

Proof. We observe first that there is a point $\hat{v}$ in X such that $f^\circ(x; \hat{v}) < 0$, since $f^\circ(x; \cdot)$ is the support function of a set (i.e., $\partial f(x)$) not containing zero. If v belongs to the left-hand side of (5), then for any $\varepsilon > 0$, $f^\circ(x; v + \varepsilon\hat{v}) < 0$, since $f^\circ(x; \cdot)$ is subadditive (Proposition 2.1.1(a)). In consequence, it suffices to prove that any v for which $f^\circ(x; v) < 0$ belongs to $T_C(x)$, as we now proceed to do.

It follows from the definition of $f^\circ(x; v)$ that there are ε and $\delta > 0$ such that, for all y within ε of x and t in $(0,\varepsilon)$, we have

$$f(y + tv) - f(y) \leqslant -\delta t.$$

Now let x_i be any sequence in C converging to x, and t_i any sequence decreasing to 0. By definition of C, we have $f(x_i) \leqslant f(x)$, and for all i sufficiently large,

$$\begin{aligned} f(x_i + t_i v) &\leqslant f(x_i) - \delta t_i \\ &\leqslant f(x) - \delta t_i. \end{aligned}$$

It follows that $x_i + t_i v$ (for i large) belongs to C, and this establishes that $v \in T_C(x)$ (Theorem 2.4.5).

Now suppose that f is regular at x. In order to derive the extra assertions in this case, it will suffice to prove that any member v of $K_C(x)$ belongs to the left-hand side of (5). Since we always have $T_C(x) \subset K_C(x)$, it will then follow that the three sets coincide.

So let v belong to $K_C(x)$. Then by definition

$$\liminf_{t \downarrow 0} \frac{d_C(x + tv)}{t} = 0.$$

For any $\varepsilon > 0$, we may therefore choose a sequence t_i decreasing to 0 such that, for all i sufficiently large, we have

$$d_C(x + t_i v) \leqslant \varepsilon t_i.$$

Consequently there is a point x_i in C satisfying

$$\|x + t_i v - x_i\| \leqslant 2\varepsilon t_i,$$

and of course $f(x_i) \leqslant f(x)$. We deduce

$$\frac{f(x + t_i v) - f(x)}{t_i} \leqslant \frac{f(x_i) + 2\varepsilon K t_i - f(x)}{t_i} \leqslant 2\varepsilon K,$$

where K is the Lipschitz constant for f near x. Taking limits, and recalling that ε is arbitrary, we arrive at $f'(x; v) = f^\circ(x; v) \leqslant 0$, as required. □

Corollary 1

$$N_C(x) \subset \bigcup_{\lambda \geqslant 0} \lambda\, \partial f(x).$$

If f is regular at x, then equality holds.

Proof. The cone polar to the left-hand side of (5) is precisely the weak*-closed convex cone generated by $\partial f(x)$; that is, the right-hand side of the above inclusion (which is already closed because $\partial f(x)$ is a weak*-compact set not containing 0). Since taking polars reverses inclusions, the result follows immediately from (5). □

Corollary 2

Let C be given as follows:

$$\{y \in X : f_1(y) \leqslant 0, f_2(y) \leqslant 0, \ldots, f_n(y) \leqslant 0\},$$

and let x be such that $f_i(x) = 0$ $(i = 1, 2, \ldots, n)$. Then, if each f_i is strictly differentiable at x, and if $D_s f_i(x)$, $i = 1, 2, \ldots, n$, are positively linearly independent, it follows that C is regular at x, and one has

$$N_C(x) = \left\{ \sum_{i=1}^{n} \lambda_i D_s f_i(x) : \lambda_i \geqslant 0, i = 1, 2, \ldots, n \right\}.$$

Proof. Define $f(y)$ via

$$f(y) = \max\{f_i(y) : i = 1, 2, \ldots, n\}.$$

Then (by Propositions 2.3.6(a) and 2.3.12) f is Lipschitz near x and regular at x. C is the set $\{y : f(y) \leqslant 0\}$, and $f(x) = 0$, so that C is regular and the assertion of Corollary 1 holds with equality. By Proposition 2.3.12, $\partial f(x)$ is the convex hull of the n points $D_s f_i(x)$, which leads to the desired formula for $N_C(x)$. □

Hypertangents

A vector v in X is said to be *hypertangent* to the set C at the point x in C if, for some $\varepsilon > 0$,

$$y + tw \in C \qquad \text{for all } y \in (x + \varepsilon B) \cap C, w \in v + \varepsilon B, t \in (0, \varepsilon).$$

It follows easily that any vector v hypertangent to C at x belongs to $T_C(x)$. It is possible, however, for there to be no hypertangents at all.

2.4.8 Theorem (Rockafellar)

Suppose there is at least one vector v hypertangent to C at x. Then the set of all hypertangents to C at x coincides with int $T_C(x)$.

Proof (Rockafellar, 1980). Let K denote the set of all hypertangents to C at x; it follows easily that K is an open set containing all positive multiples of its elements, and that $K \subset T_C(x)$. It will therefore suffice to prove int $T_C(x) \subset K$. This inclusion can be established by verifying

$$K + T_C(x) \subset K, \tag{6}$$

for if this holds and $v \in$ int $T_C(x)$, then for arbitrary w in K there exists $\lambda > 0$ with $v - \lambda w \in$ int $T_C(x)$. Since also $\lambda w \in K$ and $\lambda w + (v - \lambda w) = v$, it follows from (6) that $v \in K$. We therefore establish (6).

Let $v_1 \in K$ and $v_2 \in T_C(x)$. In order to verify (6), we must demonstrate that $v_1 + v_2 \in K$; that is, that there exists an $\varepsilon > 0$ such that

$$(7) \qquad C \cap (x + \varepsilon B) + t(v_1 + v_2 + \varepsilon B) \subset C \qquad \text{for all } t \text{ in } (0, \varepsilon).$$

Since $v_1 \in K$, we know there exists an $\varepsilon_1 > 0$ such that

$$(8) \qquad C \cap (x + \varepsilon_1 B) + t(v_1 + \varepsilon_1 B) \subset C \qquad \text{for all } t \text{ in } (0, \varepsilon_1).$$

Next, because $v_2 \in T_C(x)$, there exists $\varepsilon_2 > 0$ such that

$$(9) \qquad C \cap (x + \varepsilon_2 B) + tv_2 \subset C + t\left(\frac{\varepsilon_1}{2}\right)B \qquad \text{for } t \in (0, \varepsilon_2),$$

a consequence of the characterization of T_C given in Theorem 2.4.5. Let $\varepsilon < \min(\varepsilon_2, \varepsilon_1/2, \varepsilon_1/[1 + \varepsilon_1 + \|v_2\|])$. We claim that (7) is valid for this ε.

To see this, let v be any element of the left-hand side of (7). Then $v = y + t(v_1 + v_2 + \varepsilon u)$ where $y \in C \cap (x + \varepsilon B)$ and $u \in B$. Since $\varepsilon \leqslant \varepsilon_2$, we derive from (9) the fact that $y + tv_2 - t\varepsilon_1 w/2 \in C$ for some vector w of norm at most 1. The vector $y + t(v_2 - \varepsilon_1 w/2)$ also belongs to $x + \varepsilon_1 B$:

$$\left|x - y - t\left(v_2 - \frac{\varepsilon_1 w}{2}\right)\right| \leqslant \varepsilon + \varepsilon(\|v_2\| + \varepsilon_1) < \varepsilon_1.$$

Consequently $y + t(v_2 - \varepsilon_1 w/2) \in C \cap (x + \varepsilon_1 B)$, and by (8) we conclude that

$$(10) \qquad y + t\left(v_2 - \frac{\varepsilon_1 w}{2}\right) + tv_1 + t\varepsilon_1 B \subset C.$$

If we now write v as follows:

$$v = y + t\left(v_2 - \frac{\varepsilon_1 w}{2}\right) + tv_1 + t\left(\varepsilon u + \frac{\varepsilon_1 w}{2}\right)$$

and observe $|\varepsilon u + \varepsilon_1 w/2| \leqslant \varepsilon + \varepsilon_1/2 < \varepsilon_1$, we deduce from (10) that v belongs to C. □

Corollary

Let $x \in C$, and suppose that a hypertangent to C at x exists. Then the multifunction N_C is closed at x; that is, if $\zeta_i \in N_C(x_i)$ and $\zeta_i \to \zeta$ (weak), $x_i \to x$, then it follows that $\zeta \in N_C(x)$.*

Proof. Consider first any v hypertangent to C at x. It follows easily that $v \in T_C(y)$ for all y in C near x, whence $0 \geqslant \lim_{i \to \infty} \langle \zeta_i, v \rangle = \langle \zeta, v \rangle$. Since int $T_C(x)$ consists of hypertangents by the theorem, we have proved

$$\langle \zeta, v\rangle \leqslant 0 \quad \text{for all } v \in \operatorname{int} T_C(x).$$

But $T_C(x)$ is convex with nonempty interior, so $T_C(x)$ is the closure of its interior, hence $\langle \zeta, v\rangle \leqslant 0$ for all $v \in T_C(x)$; that is, $\zeta \in N_C(x)$. □

Epigraphs and Non-Lipschitz Functions

As we have seen, the distance function d_C serves as a bridge between the analytical theory of generalized gradients and the geometric theory of the preceding section. A different link can be forged through the notion of *epigraph*. The epigraph of a real-valued function f defined on X is the following subset of $X \times R$:

$$\operatorname{epi} f := \{(x, r) \in X \times R : f(x) \leqslant r\}.$$

Clearly epi f captures all information about f. The following confirms that tangency is consistent with the generalized directional derivative. (Note that only the local nature of epi f near the point $(x, f(x))$ is involved.)

2.4.9 Theorem

Let f be Lipschitz near x. Then

(i) *The epigraph of $f^\circ(x; \cdot)$ is $T_{\operatorname{epi} f}(x, f(x))$; that is, (v, r) lies in $T_{\operatorname{epi} f}(x, f(x))$ iff $r \geqslant f^\circ(x; v)$.*
(ii) *f is regular at x iff epi f is regular at $(x, f(x))$.*

Proof. Suppose first that (v, r) lies in $T_{\operatorname{epi} f}(x, f(x))$. Choose sequences $y_i \to x$, $t_i \downarrow 0$ such that

$$\lim_{i \to \infty} \frac{f(y_i + t_i v) - f(y_i)}{t_i} = f^\circ(x; v). \tag{11}$$

Note that $(y_i, f(y_i))$ is a sequence in epi f converging to $(x, f(x))$. By Theorem 2.4.5, therefore, there exists a sequence (v_i, r_i) converging to (v, r) such that $(y_i, f(y_i)) + t_i(v_i, r_i) \in \operatorname{epi} f$. Thus

$$f(y_i) + t_i r_i \geqslant f(y_i + t_i v_i).$$

We rewrite this as

$$\frac{f(y_i + t_i v_i) - f(y_i)}{t_i} \leqslant r_i.$$

Taking limits, and recalling Eq. (11), we obtain $f^\circ(x; v) \leqslant r$ as desired.

It suffices now to prove that for any v, for any $\delta \geqslant 0$, the point $(v, f^\circ(x; v) + \delta)$ lies in $T_{\text{epi} f}(x, f(x))$. Accordingly, let (x_i, r_i) be any sequence in epi f converging to $(x, f(x))$, and let $t_i \downarrow 0$. We are to produce a sequence (v_i, s_i) converging to $(v, f^\circ(x; v) + \delta)$ with the property that $(x_i, r_i) + t_i(v_i, s_i)$ lies in epi f for each i; that is, such that

$$r_i + t_i s_i \geqslant f(x_i + t_i v_i). \tag{12}$$

Let us define $v_i = v$ and

$$s_i = \max\left\{ f^\circ(x; v) + \delta, \frac{f(x_i + t_i v) - f(x_i)}{t_i} \right\}.$$

Observe first that $s_i \to f^\circ(x; v) + \delta$, since

$$\limsup_{i \to \infty} \frac{f(x_i + t_i v) - f(x_i)}{t_i} \leqslant f^\circ(x; v).$$

We need only verify (12) to complete the proof of (i). We have

$$r_i + t_i s_i \geqslant r_i + [f(x_i + t_i v) - f(x_i)]$$

and $r_i \geqslant f(x_i)$ (since $(x_i, r_i) \in$ epi f); (12) is the result. We now turn to (ii).

Suppose first that f is regular at x. Let the function g: $X \times R \to R$ be defined via

$$g(x', r) = f(x') - r,$$

and note that g is regular at $(x, f(x))$, and that epi f is the set $\{g \leqslant 0\}$. Since $0 \notin \partial g(x, f(x)) = \partial f(x) \times \{-1\}$, Theorem 2.4.7 applies to yield the regularity of epi f at $(x, f(x))$.

To finish the proof we require the following result. Let f'_+ signify the following Dini derivate:

$$f'_+(x; v) = \liminf_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

Lemma

$$K_{\text{epi} f}(x, f(x)) = \text{epi } f'_+(x; \cdot).$$

The proof is an exercise along the lines of the proof of (i) above; we omit it.

Now suppose that epi f is regular at $(x, f(x))$. Then, in view of the lemma and (i) of the theorem, one has

$$\text{epi } f'_+(x;\cdot) = \text{epi } f^\circ(x;\cdot),$$

which implies that $f'_+(x; v) = f^\circ(x; v)$ for all v. It follows that $f'(x; v)$ exists and coincides with $f^\circ(x; v)$; that is, that f is regular at x. □

Corollary

An element ζ of X^* belongs to $\partial f(x)$ iff $(\zeta, -1)$ belongs to $N_{\text{epi} f}(x, f(x))$.

Proof. We know that ζ belongs to $\partial f(x)$ iff, for any v, we have $f^\circ(x; v) \geqslant \langle \zeta, v\rangle$; that is, iff for any v and for any $r \geqslant f^\circ(x; v)$, we have

$$\langle(\zeta, -1), (v, r)\rangle \leqslant 0.$$

By the theorem, this is equivalent to the condition that this last inequality be valid for all elements (v, r) in $T_{\text{epi} f}(x, f(x))$; that is, that $(\zeta, -1)$ lie in $N_{\text{epi} f}(x, f(x))$. □

An Extended Definition of ∂f

The above corollary, which of course is analogous to the fact that $[f'(x), -1]$ is normal to the graph of the smooth function $f: R \to R$, provides a strong temptation: to define, for any f, locally Lipschitz or not, $\partial f(x)$ as the set of ζ for which $(\zeta, -1)$ lies in $N_{\text{epi} f}(x, f(x))$. (The corollary would then guarantee that this new definition is consistent with the previous one for the locally Lipschitz case.) It would be feasible then to define $\partial f(x)$ even for extended-valued functions f on X (i.e., those taking values in $R \cup \{\pm\infty\}$), as long as f is finite at x. In similar fashion, the concept of regularity could be extended to such f. We proceed to succumb to temptation:

2.4.10 Definition

Let $f: X \to R \cup \{\pm\infty\}$ be finite at a point x. We define $\partial f(x)$ to be the set of all ζ in X^* (if any) for which $(\zeta, -1) \in N_{\text{epi} f}(x, f(x))$. We say that f is regular at x provided epi f is regular at $(x, f(x))$.

It follows that $\partial f(x)$ is a weak*-closed subset of X^*, which may no longer be compact and, as the definition implies, may be empty (but not, of course, if f is locally Lipschitz). It turns out that $\partial f(x)$ is never empty if f attains a local minimum at x:

2.4.11 Proposition

Let $f: X \to R \cup \{\pm\infty\}$ be finite at x, and suppose that for all x' in a neighborhood of x, $f(x') \geqslant f(x)$. Then $0 \in \partial f(x)$.

Proof. We wish to prove that $(0, -1) \in N_{\text{epi} f}(x, f(x))$, or equivalently, that $\langle (0, -1), (v, r) \rangle \leqslant 0$ for all (v, r) in $T_{\text{epi} f}(x, f(x))$, or again that $r \geqslant 0$ for all such (v, r). But for any (v, r) in $T_{\text{epi} f}(x, f(x))$, for any sequence $t_i \downarrow 0$, there is a sequence $(v_i, r_i) \to (v, r)$ such that $(x, f(x)) + t_i(v_i, r_i) \in \text{epi} f$. But then

$$f(x) + t_i r_i \geqslant f(x + t_i v_i) \geqslant f(x)$$

for all i sufficiently large, and so $t_i r_i \geqslant 0$. It follows that one has $\lim r_i = r \geqslant 0$. □

Indicators

The *indicator* of a set C in X is the extended-valued function ψ_C: $X \to R \cup \{\infty\}$ defined as follows:

$$\psi_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Here is more evidence confirming the compatibility between the analytical and the geometrical concepts we have introduced.

2.4.12 Proposition

Let the point x belong to C. Then

$$\partial \psi_C(x) = N_C(x),$$

and ψ_C is regular at x iff C is regular at x.

Proof. ζ belongs to $\partial \psi_C(x)$ iff $(\zeta, -1) \in N_{\text{epi} \psi_C}(x, 0)$. But clearly

$$\text{epi}\, \psi_C = C \times [0, \infty),$$

so this is equivalent, by the corollary to Theorem 2.4.5, to

$$\zeta \in N_C(x), \qquad -1 \in N_{[0,\infty)}(0).$$

The second of these is always true (Proposition 2.4.4); the formula follows. The regularity assertion follows easily from the definition as well. □

2.5 *THE CASE IN WHICH X IS FINITE-DIMENSIONAL*

We explore in this section the additional properties of generalized gradients, normals, and tangents that follow when $X = R^n$. As usual, we identify X^* with X in this case, so that $\partial f(x)$ is viewed as a subset of R^n. Our most important result is the characterization given by Theorem 2.5.1. It facilitates greatly the calculation of ∂f in finite dimensions. We recall Rademacher's Theorem, which states that a function which is Lipschitz on an open subset of R^n is differentiable almost everywhere (a.e.) (in the sense of Lebesgue measure) on that subset. The set of points at which a given function f fails to be differentiable is denoted Ω_f.

2.5.1 Theorem

Let f be Lipschitz near x, and suppose S is any set of Lebesgue measure 0 *in R^n. Then*

$$\partial f(x) = \operatorname{co}\{\lim \nabla f(x_i) : x_i \to x,\ x_i \notin S,\ x_i \notin \Omega_f\}. \tag{1}$$

(The meaning of Eq. (1) is the following: consider any sequence x_i converging to x while avoiding both S and points at which f is not differentiable, and such that the sequence $\nabla f(x_i)$ converges; then the convex hull of all such limit points is $\partial f(x)$.)

Proof. Let us note to begin with that there are "plenty" of sequences x_i which converge to x and avoid $S \cup \Omega_f$, since the latter has measure 0 near x. Further, because ∂f is locally bounded near x (Proposition 2.1.2(a)) and $\nabla f(x_i)$ belongs to $\partial f(x_i)$ for each i (Proposition 2.2.2), the sequence $\{\nabla f(x_i)\}$ admits a convergent subsequence by the Bolzano–Weierstrass Theorem. The limit of any such sequence must belong to $\partial f(x)$ by the closure property of ∂f proved in Proposition 2.1.5(b). It follows that the set

$$\{\lim \nabla f(x_i) : x_i \to x,\ x_i \notin S \cup \Omega_f\}$$

is contained in $\partial f(x)$ and is nonempty and bounded, and in fact compact, since it is rather obviously closed. Since $\partial f(x)$ is convex, we deduce that the left-hand side of (1) contains the right. Now, the convex hull of a compact set in R^n is compact, so to complete the proof we need only show that the support function of the left-hand side of (1) (i.e., $f^\circ(x;\cdot)$) never exceeds that of the right. This is what the following lemma does:

Lemma. For any $v \neq 0$ in R^n, for any $\varepsilon > 0$, we have

$$f^\circ(x; v) - \varepsilon \leqslant \limsup\{\nabla f(y)\cdot v : y \to x,\ y \notin S \cup \Omega_f\}.$$

To prove this, let the right-hand side be α. Then by definition, there is a $\delta > 0$ such that the conditions

$$y \in x + \delta B,\ y \notin S \cup \Omega_f$$

imply $\nabla f(y) \cdot v \leqslant \alpha + \varepsilon$. We also choose δ small enough so that $S \cup \Omega_f$ has measure 0 in $x + \delta B$. Now consider the line segments $L_y = \{y + tv : 0 < t < \delta/(2|v|)\}$. Since $S \cup \Omega_f$ has measure 0 in $x + \delta B$, it follows from Fubini's Theorem that for almost every y in $x + (\delta/2)B$, the line segment L_y meets $S \cup \Omega_f$ in a set of 0 one-dimensional measure. Let y be any point in $x + (\delta/2)B$ having this property, and let t lie in $(0, \delta/(2|v|))$. Then

$$f(y + tv) - f(y) = \int_0^t \nabla f(y + sv) \cdot v \, ds,$$

since Df exists a.e. on L_y. Since one has $|y + sv - x| < \delta$ for $0 < s < t$, it follows that $\nabla f(y + sv) \cdot v \leqslant \alpha + \varepsilon$, whence

$$f(y + tv) - f(y) \leqslant t(\alpha + \varepsilon).$$

Since this is true for all y within $\delta/2$ of x except those in a set of measure 0, and for all t in $(0, \delta/(2|v|))$, and since f is continuous, it is in fact true for all such y and t. We deduce

$$f^\circ(x; v) \leqslant \alpha + \varepsilon,$$

which completes the proof. □

Corollary

$$f^\circ(x; v) = \limsup_{y \to x} \{\nabla f(y) \cdot v : y \notin S \cup \Omega_f\}.$$

2.5.2 *Example*

Let us use the theorem to calculate $\partial f(0,0)$ where f: $R^2 \to R$ is given by

$$f(x, y) = \max\{\min[x, -y], y - x\}.$$

Define

$$C_1 = \{(x, y) : y \leqslant 2x \text{ and } y \leqslant -x\}$$

$$C_2 = \{(x, y) : y \leqslant x/2 \text{ and } y \geqslant -x\}$$

$$C_3 = \{(x, y) : y \geqslant 2x \text{ or } y \geqslant x/2\}.$$

Then $C_1 \cup C_2 \cup C_3 = R^2$, and we have

$$f(x, y) = \begin{cases} x & \text{for } (x, y) \in C_1 \\ -y & \text{for } (x, y) \in C_2 \\ y - x & \text{for } (x, y) \in C_3. \end{cases}$$

Note that the boundaries of these three sets form a set S of measure 0, and that if (x, y) does not lie in S, then f is differentiable and $\nabla f(x, y)$ is one of the points $(1,0)$, $(0, -1)$, or $(-1, 1)$. It follows from the theorem that $\partial f(0,0)$ is the convex hull of these three points.

Note that $f(0, y) = \max[0, y]$, so that $\partial_y f(0,0)$ is the interval $[0, 1]$. Similarly, $\partial_x f(0,0) = [-1, 0]$. For this example therefore we have

$$\partial_x f(0,0) \times \partial_y f(0,0) \not\subset \partial f(0,0) \not\subset \partial_x f(0,0) \times \partial_y f(0,0).$$

We saw in Proposition 2.3.15 that the second of these potential inclusions holds if f is regular; here is a result in a similar vein.

2.5.3 Proposition

Let $f\colon R^n \times R^m \to R$ be Lipschitz on a convex neighborhood $U \times V$ of a point $x = (\alpha, \beta)$, and suppose that for each α' near α, the function $f(\alpha', \cdot)$ is convex on V. Then whenever (z, w) belongs to $\partial f(x)$ we also have $w \in \partial_2 f(\alpha, \beta)$.

Proof. In view of Theorem 2.5.1, it suffices to prove the implication for points (z, w) of the form

$$\lim_{i \to \infty} \nabla f(x_i), \qquad \text{where } x_i = (\alpha_i, \beta_i) \to x,\ x_i \notin \Omega_f,$$

since it will follow for convex combinations of such points, and hence for $\partial f(x)$. Let $\nabla f(x_i) = [z_i, w_i]$. We have $w_i \in \partial_2 f(\alpha_i, \beta_i)$ by Proposition 2.2.2 so that (by Proposition 2.2.7) w_i is a subgradient of the convex function $f(\alpha_i, \cdot)$ on V. Thus for all v in R^m sufficiently small, and for all i sufficiently large, we have

$$f(\alpha_i, \beta_i + v) - f(\alpha_i, \beta_i) \geqslant v \cdot w_i.$$

Taking limits gives us

$$f(\alpha, \beta + v) - f(\alpha, \beta) \geqslant v \cdot w.$$

This implies $f_2^\circ(\alpha, \beta; v) \geqslant v \cdot w$ for all v, so w belongs to $\partial_2 f(\alpha, \beta)$. □

The Euclidean Distance Function

Let C be an arbitrary nonempty subset of R^n. We can also use Theorem 2.5.1 to characterize the generalized gradient of the distance function d_C relative to the usual Euclidean norm:

$$d_C(x) = \inf\{|x - c| : c \in C\},$$

where $|x - c| = \{\sum_{i=1}^n |x_i - c_i|^2\}^{1/2}$. (The symbol $|\cdot|$ is reserved for this norm on R^n.) Recall that d_C was shown to be (globally) Lipschitz of rank 1 in Proposition 2.4.1.

2.5.4 Proposition

Let $\nabla d_C(x)$ exist and be different from zero. Then $x \notin \mathrm{cl}\, C$, there is a unique closest point c_0 in $\mathrm{cl}\, C$ to x, and $\nabla d_C(x) = (x - c_0)/|x - c_0|$.

Proof. If x lies in $\mathrm{cl}\, C$, then for any v in R^n we calculate

$$v \cdot \nabla d_C(x) = \lim_{t \downarrow 0} \frac{d_C(x + tv) - d_C(x)}{t}$$

$$= \lim_{t \downarrow 0} \frac{d_C(x + tv)}{t} \geqslant 0,$$

whence $\nabla d_C(x) = 0$, contrary to assumption. Thus x does not lie in $\mathrm{cl}\, C$, and admits at least one closest point c_0 in $\mathrm{cl}\, C$ (i.e., a point c_0 in $\mathrm{cl}\, C$ such that $d_C(x) = |x - c_0|$). We shall now show that $\nabla d_C(x) = (x - c_0)/|x - c_0|$, from which it follows that there are no other closest points.

For t in $(0, 1)$, the closest point in $\mathrm{cl}\, C$ to $x + t(c_0 - x)$ is still c_0, whence

$$d_C(x + t(c_0 - x)) = (1 - t)|x - c_0|.$$

Subtracting $d_C(x) = |x - c_0|$ from both sides, dividing by t, and taking the limit as $t \downarrow 0$, produces

$$d_C'(x; c_0 - x) = \nabla d_C(x) \cdot (c_0 - x) = -|c_0 - x|.$$

Now $|\nabla d_C(x)| \leqslant 1$, since d_C is Lipschitz of rank 1, so this last equation yields $\nabla d_C(x) = (x - c_0)/|x - c_0|$. (This is where the special nature of the Euclidean norm $|\cdot|$ is used.) This completes the proof. □

Let us define a nonzero vector v to be *perpendicular* to C at $x \in \mathrm{cl}\, C$ (symbolically, $v \perp C$ at x), provided $v = x' - x$, where x' has unique closest point x in $\mathrm{cl}\, C$. The following property of perpendiculars will prove useful:

2.5.5 Proposition

Let v be perpendicular to C at x, where $x \in \mathrm{cl}\, C$. Then, for all c in $\mathrm{cl}\, C$, one has

$$\langle v, c - x \rangle \leqslant \tfrac{1}{2}|x - c|^2.$$

Proof. We have by definition, for all c in $\mathrm{cl}\, C$,

$$|x' - c| \geqslant |x' - x|,$$

where $v = x' - x$. This is equivalent to

$$\langle x' - c, x' - c\rangle \geqslant \langle x' - x, x' - x\rangle.$$

To obtain the conclusion of the proposition, it suffices to replace $x' - c$ by $v + (x - c)$ on the left in this last inequality, and to expand the inner product. □

2.5.6 Theorem

Let x belong to $\mathrm{cl}\, C$. Then $\partial d_C(x)$ equals the convex hull of the origin and the set

$$\left\{ v = \lim \frac{v_i}{|v_i|} : v_i \perp C \text{ at } x_i \to x, v_i \to 0\right\}.$$

Proof. The gradient of d_C, whenever it exists, is either 0 or a unit vector of the type described in Proposition 2.5.4 (i.e., a normalized perpendicular). We immediately derive from Theorem 2.5.1 the conclusion that $\partial d_C(x)$ is contained in the given set. The opposite inclusion requires first that we show that $0 \in \partial d_C(x)$. This follows from Proposition 2.3.2, since d_C attains a (global) minimum at x. Finally we need to show that any $v = \lim v_i/|v_i|$ as described belongs to $\partial d_C(x)$. Let $v_i = y_i - x_i$ as in the definition of perpendicular. By the mean-value theorem,

$$d_C(y_i) - d_C(x_i) \in \langle \partial d_C(x_i^*), y_i - x_i\rangle$$

for some point x_i^* in (x_i, y_i). But $d_C(x_i) = 0$ and $d_C(y_i) = |y_i - x_i|$, so that

$$1 \in \left\langle \partial d_C(x_i^*), \frac{(y_i - x_i)}{|y_i - x_i|}\right\rangle.$$

Since $|\partial d_C| \leqslant 1$, we deduce $v_i/|v_i| \in \partial d_C(x_i^*)$. Taking limits, and observing that x_i^* converges to x, we obtain $v \in \partial d_C(x)$ by Proposition 2.1.5(d). □

Corollary 1

If x lies on the boundary of $\mathrm{cl}\, C$, then $\partial d_C(x)$ contains nonzero points.

Proof. If x lies on the boundary of $\mathrm{cl}\, C$, then for every $\varepsilon > 0$, the set of points y which lie within ε of x but not in $\mathrm{cl}\, C$ is of positive measure. In consequence, any such set contains points y at which $\nabla d_C(y)$ exists, and we know the latter is of the form $v/|v|$ for some perpendicular v. Letting y converge to x (i.e., $\varepsilon \downarrow 0$) leads to a sequence of unit vectors, any of whose limit points must belong to $\partial d_C(x)$. □

Corollary 2

If x lies on the boundary of $\mathrm{cl}\, C$, then $N_C(x)$ contains nonzero points.

Proof. Invoke Proposition 2.4.2 and Corollary 1. □

A Characterization of Normal Vectors

The alternate characterization of normal vectors to sets in R^n given below is useful in many particular calculations; it is a geometric analogue of Theorem 2.5.1.

2.5.7 Proposition

Let x belong to $\operatorname{cl} C$. Then $N_C(x)$ is the closed convex cone generated by the origin and the set

$$\left\{ v = \lim \frac{v_i}{|v_i|} : v_i \perp C \text{ at } x_i \to x, v_i \to 0 \right\}.$$

Proof. Proposition 2.4.2 asserts that $N_C(x)$ is the closed convex cone generated by $\partial d_C(x)$. The result follows from Theorem 2.5.6. □

Interior of the Tangent Cone

2.5.8 Theorem

Let x belong to a closed subset C of R^n. Then one has $v \in \operatorname{int} T_C(x)$ iff there is an $\varepsilon > 0$ such that

$$\text{(2)} \quad d_C(y + tw) \leqslant d_C(y) \quad \text{for all } y \in x + \varepsilon B, w \in v + \varepsilon B, t \in [0, \varepsilon).$$

Proof. Suppose first that v satisfies (2) for some $\varepsilon > 0$. We shall show that any w in $v + \varepsilon B$ lies in $T_C(x)$ (and hence that $v \in \operatorname{int} T_C(x)$). To see this we shall use the characterization of T_C given by Theorem 2.4.5. Let x_i be a sequence in C converging to x, t_i a sequence decreasing to 0. Then, in light of (2), $d_C(x_i + t_i w) \leqslant 0$ for all i sufficiently large. But then $x_i + t_i w$ lies in C for all such i. This confirms $w \in T_C(x)$.

To prove the necessity of (2), let $v \in \operatorname{int} T_C(x)$ be given. It follows that $v \cdot z < 0$ for every nonzero $z \in N_C(x)$, for if $v \cdot z = 0$, then every neighborhood of v would contain points w for which $w \cdot z > 0$, which cannot be since $T_C(x) \cdot z \leqslant 0$. Equivalently, there exists $k > 0$ such that

$$v \cdot z \leqslant -k|z| \qquad \text{for all } z \in N_C(x).$$

In view of Proposition 2.4.2, this transmutes to

$$\text{(3)} \qquad v \cdot z \leqslant -k|z| \qquad \text{for all } z \in \partial d_C(x).$$

Recall that whenever y is such that $\nabla d_C(y)$ exists and is nonzero, it is a unit vector (by Proposition 2.5.4) and that $\partial d_C(x)$ contains all limits of such gradients as $y \to x$ (any such limit is of course a unit vector). In light of this, we derive from (3) the conclusion that for some $\delta > 0$,

$$(4) \qquad v \cdot \nabla d_C(y) \leqslant 0 \qquad \text{for all } y \in x + \delta B,$$

if $\nabla d_C(y)$ exists and is nonzero; of course the inequality is also valid if $\nabla d_C(y) = 0$.

Now choose $\varepsilon > 0$ such that $y + tw$ lies in $x + \delta B$ whenever $y \in x + \varepsilon B$, $w \in v + \varepsilon B$, $t \in [0, \varepsilon)$. Fix any such w, and note that for almost all y in $x + \varepsilon B$, the ray $t \to y + tw$ meets in a set of 0 one-dimensional measure the set where d_C is not differentiable (since this latter set has 0 measure in R^n). For any such y, and for any t in $[0, \varepsilon)$,

$$d_C(y + tw) = d_C(y) + \int_0^t \nabla d_C(y + sw) \cdot w \, ds.$$

Since the integrand is negative or zero by (4), we derive $d_C(y + tw) \leqslant d_C(y)$, for almost all y in $x + \varepsilon B$. Since d_C is continuous, this must in fact be the case for all y in $x + \varepsilon B$; that is, (2) holds. □

Corollary 1 (Rockafellar)

When C is a closed subset of R^n, the set of hypertangents to C at a point $x \in C$ coincides with int $T_C(x)$.

Proof. It follows from the theorem, by taking y in (2) to lie in C, that v is a hypertangent whenever v lies in int $T_C(x)$. The converse results from Theorem 2.4.8. □

Corollary 2 (Rockafellar)

Let C be a closed subset of R^n containing x, and suppose int $T_C(x) \neq \phi$. Then the multifunction N_C is closed at x; that is,

$$x_i \to x, \qquad \zeta_i \in N_C(x_i), \qquad \zeta_i \to \zeta \text{ implies } \zeta \in N_C(x).$$

Proof. Invoke Corollary 1 and the corollary to Theorem 2.4.8. □

2.6 GENERALIZED JACOBIANS

Let us now consider a vector-valued function $F: R^n \to R^m$, written in terms of component functions as $F(x) = [f^1(x), f^2(x), \ldots, f^m(x)]$. We assume that each f^i (and hence F) is Lipschitz near a given point x of interest. As before, Rademacher's Theorem asserts that F is differentiable (i.e., each f^i is differentiable) a.e. on any neighborhood of x in which F is Lipschitz. We shall continue to denote the set of points at which F fails to be differentiable by Ω_F.

We shall write $JF(y)$ for the usual $m \times n$ Jacobian matrix of partial derivatives whenever y is a point at which the necessary partial derivatives exist.

2.6.1 Definition

The *generalized Jacobian* of F at x, denoted $\partial F(x)$, is the convex hull of all $m \times n$ matrices Z obtained as the limit of a sequence of the form $JF(x_i)$, where $x_i \to x$ and $x_i \notin \Omega_F$.

Symbolically, then, one has

$$\partial F(x) = \operatorname{co}\{\lim JF(x_i) : x_i \to x,\, x_i \notin \Omega_F\}.$$

Of course, there is no problem in taking convex combinations in the space $R^{m \times n}$ of $m \times n$ matrices. We shall endow this space with the norm

$$\|M\|_{m \times n} = \left\{ \sum_{i=1}^{m} |r_i|^2 : r_i \text{ is the } i\text{th row of } M \right\}^{1/2},$$

and we denote by $B_{m \times n}$ the open unit ball in $R^{m \times n}$. We proceed to summarize some properties of ∂F.

2.6.2 Proposition

(a) *$\partial F(x)$ is a nonempty convex compact subset of $R^{m \times n}$.*

(b) *∂F is closed at x; that is, if $x_i \to x$, $Z_i \in \partial F(x_i)$, $Z_i \to Z$, then $Z \in \partial F(x)$.*

(c) *∂F is upper semicontinuous at x: for any $\varepsilon > 0$ there is $\delta > 0$ such that, for all y in $x + \delta B$,*

$$\partial F(y) \subset \partial F(x) + \varepsilon B_{m \times n}.$$

(d) *If each component function f^i is Lipschitz of rank K_i at x, then F is Lipschitz at x of rank $K = |(K_1, K_2, \ldots, K_m)|$, and $\partial F(x) \subset K\bar{B}_{m \times n}$.*

(e) *$\partial F(x) \subset \partial f^1(x) \times \partial f^2(x) \times \cdots \times \partial f^m(x)$, where the latter denotes the set of all matrices whose i th row belongs to $\partial f^i(x)$ for each i. If $m = 1$, then $\partial F(x) = \partial f^1(x)$ (i.e., the generalized gradient and the generalized Jacobian coincide).*

2.6.3 Remark

We must now confess that $\partial f(x)$ is not really a subset of R^n (when $f\colon R^n \to R$) as we have pretended up to now. To be consistent with the generalized Jacobian, ∂f should consist of $1 \times n$ matrices (i.e., row vectors). On the other hand, the usual convention that linear operators from R^n to R^m, such as $F(x) = Ax$, are represented by matrix multiplication by an $m \times n$ matrix on the left, forces us to view R^n as $n \times 1$ matrices (i.e., column vectors), and we get $\partial F(x) = \{A\}$ as we would wish. This distinction is irrelevant as long as we remain in the case $m = 1$, but must be adhered to in interpreting, for example, certain chain rules to come.

Proof. Assertions (a) and (d) together follow easily, the former as in the first part of the proof of Theorem 2.5.1. It is clear that (c) subsumes (b). To prove (c), suppose, for a given ε, that no such δ existed. Then (since the right-hand side is convex), for each i there would necessarily be an element $JF(y_i)$ such that $y_i \in x + (1/i)B$ yet $JF(y_i) \notin \partial F(x) + \varepsilon B_{m \times n}$. We may suppose (by local boundedness of ∂F) that $JF(y_i)$ converges to an element Z. By definition, $Z \in \partial F(x)$, a contradiction.

Assertion (e) follows immediately from the definition of ∂F and Theorem 2.5.1. □

In comparing the statement of Theorem 2.5.1 and the definition of the generalized Jacobian, we may well ask: what happened to the set S (of the former)? In fact, although the generalized gradient is "blind" to sets of measure 0, as proven in Theorem 2.5.1, it remains unknown whether (for $m > 1$) this is true of the generalized Jacobian. Conceivably (although we doubt it), an altered definition in which the points x_i are also restricted to lie outside a null set S could lead to a different generalized Jacobian $\partial_S F(x)$. The possibility that the generalized Jacobian is nonintrinsic in this sense is unresolved.†

In most applications, it is the *images* of vectors under the generalized Jacobian that enter into the picture. In this sense, we now show that the putative ambiguity of the generalized Jacobian is irrelevant.

2.6.4 Proposition

For any v in R^n, w in R^m,

$$\partial F(x)v = \partial_S F(x)v$$

$$\partial F(x)^* w = \partial_S F(x)^* w,$$

*where * denotes transpose.*

Proof. We shall prove the first of these, the other being analogous. Since both sets in question are compact and convex, and since ∂F includes $\partial_S F$, we need only show that for any point u in R^m, the value σ_1 of the support function of $\partial F(x)v$ evaluated at u does not exceed the corresponding value σ_2 for $\partial_S F(x)v$. One has

$$\begin{aligned}\sigma_1 &= \limsup\{\langle u, JF(y)v\rangle : y \to x,\ y \notin \Omega_F\}\\ &= \limsup\{\langle JF(y)^* u, v\rangle : y \to x,\ y \notin \Omega_F\}.\end{aligned}$$

† In fact the intrinsic nature of the generalized Jacobian has been confirmed by J. Warga; see J. Math. Anal. Appl. 81 (1981) 545–560.

Let g be the function $g(y) = u \cdot F(y)$. Recall that ∇g, when it exists, is a row vector (see Remark 2.6.3). For $y \notin \Omega_F$, one has $\nabla g(y)^* = JF(y)^* u$, so we may write

$$\begin{aligned}\sigma_1 &= \limsup\{\nabla g(y) \cdot v : y \to x, y \notin \Omega_F\}\\ &= g^\circ(x; v) \qquad \text{(by corollary, Theorem 2.5.1)}\\ &= \limsup\{\nabla g(y) \cdot v : y \to x, y \notin S \cup \Omega_F\}\\ &= \limsup\{\langle u, JF(y)v\rangle : y \to x, y \notin S \cup \Omega_F\},\end{aligned}$$

which is precisely σ_2. □

The following is a straightforward extension of the vector mean-value theorem. Note that as a consequence of Proposition 2.6.4, the right-hand side below is blind to sets of measure 0 in calculating ∂F.

2.6.5 Proposition

Let F be Lipschitz on an open convex set U in R^n, and let x and y be points in U. Then one has

$$F(y) - F(x) \in \text{co}\, \partial F([x, y])(y - x).$$

Proof. (The right-hand side above denotes the convex hull of all points of the form $Z(y - x)$, where $Z \in \partial F(u)$ for some point u in $[x, y]$. Since $[\text{co}\, \partial F([x, y])](y - x) = \text{co}[\partial F([x, y])(y - x)]$, there is no ambiguity.) Let us fix x. It suffices to prove this inclusion for points y having the property that the line segment $[x, y]$ meets in a set of 0 one-dimensional measure the set Ω_F, for, by a now familiar argument, almost all y have this property, and the general case will follow by a limiting argument based on the continuity of F and the upper semicontinuity of ∂F. For such y, however, we may write

$$F(y) - F(x) = \int_0^1 JF(x + t(y - x))(y - x)\, dt,$$

which directly expresses $F(y) - F(x)$ as a (continuous) convex combination of points from $\partial F([x, y])(y - x)$. This completes the proof. □

A Jacobian Chain Rule

2.6.6 Theorem

Let $f = g \circ F$, where $F: R^n \to R^m$ is Lipschitz near x and where $g: R^m \to R$ is Lipschitz near $F(x)$. Then f is Lipschitz near x and one has

(1) $$\partial f(x) \subset \text{co}\{\partial g(F(x))\, \partial F(x)\}.$$

If in addition g is strictly differentiable at $F(x)$, then equality holds (and co is superfluous).

Proof. It is easy to verify that f is Lipschitz near x. The right-hand side of (1) means the convex hull of all points of the form ζZ where $\zeta \in \partial g(F(x))$ and $Z \in \partial F(x)$. (Recall that ζ is $1 \times m$ and Z $m \times n$.) We shall establish the inclusion by showing that for any v in R^n, there is such an element $\zeta_0 Z_0$ for which $(\zeta_0 Z_0)v \geqslant f^\circ(x; v)$. Since the latter is the support function of $\partial f(x)$, (1) will result. (Note that the right-hand side of (1) is closed.)

For y near x, any expression of the form $[f(y + tv) - f(y)]/t$ can, by the mean-value theorem 2.3.7, be written as $\zeta(F(y + tv) - F(y))/t$, where ζ belongs to $\partial g(u)$ for some point u in $[F(y), F(y + tv)]$. In turn, $[F(y + tv) - F(y)]/t$ equals some element w of co $\partial F([y, y + tv])v$ (by Proposition 2.6.5), so there exists an element Z of $\partial F([y, y + tv])$ such that $\zeta w \leqslant \zeta(Zv)$. Gathering the threads, we have deduced

$$(2) \qquad \frac{f(y + tv) - f(y)}{t} \leqslant (\zeta Z)v,$$

where $\zeta \in \partial g(u)$, $Z \in \partial F([y, y + tv])$. Now choose sequences $y_i \to x$, $t_i \downarrow 0$, for which the corresponding terms on the left-hand side of (2) converge to $f^\circ(x; v)$. Note that u_i must converge to $F(x)$, and that the line segment $[y, y + tv]$ shrinks to x. By extracting subsequences, we may suppose that $\zeta_i \to \zeta_0$, which must belong to $\partial g(F(x))$, and that $Z_i \to Z_0$, which must belong to $\partial F(x)$ (by Proposition 2.6.2(b)). We derive from (2):

$$f^\circ(x; v) \leqslant \zeta_0 Z_0 v,$$

as required.

Now suppose that g is strictly differentiable at $F(x)$, and let $D_s g(F(x)) = \zeta$. We first prove a technical result.

Lemma. For any $\varepsilon > 0$ so that F is Lipschitz on $x + \varepsilon B$, there exists $\delta > 0$ such that for all $y \notin \Omega_f \cup \Omega_F$, $y \in x + \delta B$, one has

$$\nabla f(y) \in \zeta \partial F(y) + \varepsilon B.$$

To see this, pick δ in $(0, \varepsilon)$ so small that for all y in $x + \delta B$, F is Lipschitz near y, g is Lipschitz near $F(y)$, and finally so that $\partial g(F(y)) \subset \zeta + (\varepsilon/K)B$ (where K is the Lipschitz rank of F on $x + \varepsilon B$); this is possible because $\partial g(F(x)) = \{\zeta\}$. We claim first that $\nabla f(y) \in \partial g(F(y))JF(y)$ if $y \in x +$

δB, $y \notin \Omega_f \cup \Omega_F$. To see this, take any vector v and observe

$$\nabla f(y)v = \lim_{t\downarrow 0} \frac{g(F(y+tv)) - g(F(y))}{t}$$

$$= \lim_{t\downarrow 0} \zeta_t \frac{F(y+tv) - F(y)}{t}$$

(where $\zeta_t \in \partial g(u_t)$ and u_t lies in $[F(y), F(y+tv)]$, by the mean-value theorem)

$$\in \partial g(F(y))JF(y)v,$$

by taking a convergent subsequence of ζ_t, and since F is differentiable at y. Since v is arbitrary and $\partial g(F(y))JF(y)$ is convex, this establishes that $\nabla f(y)$ belongs to $\partial g(F(y))JF(y)$.

Now for any $y \in x + \delta B$, $y \notin \Omega_f \cup \Omega_F$, one has

$$\nabla f(y) \in \partial g(F(y))JF(y)$$

$$\subset \left(\zeta + \frac{\varepsilon}{K}B\right)JF(y) \quad \text{(by choice of } \delta)$$

$$\subset \zeta JF(y) + \varepsilon B.$$

This establishes the lemma.

Consider now the quantity

$$q = \max \zeta \partial F(x)v,$$

where v is any vector in R^n. Then by definition

$$q = \limsup\{\zeta JF(y)v : y \to x,\ y \notin \Omega_F\}$$

$$\geqslant \limsup\{\zeta JF(y)v : y \to x,\ y \notin \Omega_f \cup \Omega_F\}$$

$$= \limsup\{\nabla f(y)v : y \to x,\ y \notin \Omega_f \cup \Omega_F\},$$

in view of the lemma. This last quantity is precisely $f°(x; v)$, by the corollary to Theorem 2.5.1. Thus the support functions of the two sides of (1) are equal, and the proof is complete. □

Corollary

With F as in the theorem, let G: $R^m \to R^k$ be Lipschitz near $F(x)$. Then, for any v in R^n, one has

$$\partial(G \circ F)(x)v \subset \mathrm{co}\{\partial G(F(x))\,\partial F(x)v\}.$$

If G is continuously differentiable near $F(x)$, then equality holds (and co is superfluous).

Proof. Let α be any vector in R^k. Then

$$\alpha^*\partial(G \circ F)(x)v = \partial(\alpha^*[G \circ F])(x)v$$

(by the theorem applied to $F = G \circ F$, $g(y) = \alpha \cdot y$)

$$\subset \mathrm{co}\{\partial(\alpha^* G)(F(x))\,\partial F(x)v\}$$

(by the theorem applied to $F = F$, $g = \alpha^* G$)

$$= \mathrm{co}\{\alpha^*\partial G(F(x))\,\partial F(x)v\}$$

$$= \alpha^*\,\mathrm{co}\,\{\partial G(F(x))\,F(x)v\}.$$

Since α is arbitrary, the result follows. We turn now to the last assertion.

When G is C^1, then it is clear from the definition that $\partial(G \circ F)(x) \supset DG(F(x))\partial F(x)$, since $\Omega_{G \circ F} \subset \Omega_F$. Consequently equality must hold in the corollary. □

2.7 *GENERALIZED GRADIENTS OF INTEGRAL FUNCTIONALS*

In Section 3, we studied generalized gradients of finite sums. We now extend that study to integrals, which will be taken over a positive measure space $(T, \mathcal{T}, \mu)$. Suppose that U is an open subset of a Banach space X, and that we are given a family of functions f_t: $U \to R$ satisfying the following conditions:

2.7.1 *Hypotheses*

(i) For each x in U, the map $t \to f_t(x)$ is measurable.

(ii) For some $k(\cdot) \in L^1(T, R)$ (the space of integrable functions from T to R), for all x and y in U and t in T, one has

$$|f_t(x) - f_t(y)| \leqslant k(t)\|x - y\|.$$

We shall consider the integral functional f on X given by

$$f(x) = \int_T f_t(x)\,\mu(dt)$$

whenever this integral is defined. Our goal is to assert the appealing formula

$$(1) \qquad \partial f(x) = \partial \int_T f_t(x)\,\mu(dt) \subset \int_T \partial f_t(x)\,\mu(dt),$$

where ∂f_t denotes the generalized gradient of the (locally Lipschitz) function $f_t(\cdot)$. The interpretation of (1) is as follows: To every ζ in $\partial f(x)$ there corresponds a mapping $t \to \zeta_t$ from T to X^* with

$$\zeta_t \in \partial f_t(x) \qquad \mu\text{-a.e.},$$

(that is, almost everywhere relative to the measure μ) and having the property that for every v in X, the function $t \to \langle \zeta_t, v\rangle$ belongs to $L^1(X, R)$ and one has

$$\langle \zeta, v\rangle = \int_T \langle \zeta_t, v\rangle\,\mu(dt).$$

Thus, every ζ in the left-hand side of (1) is an element of X^* that can be written $\zeta(\cdot) = \int_T \langle \zeta_t, \cdot\rangle\,\mu(dt)$ where $t \to \zeta_t$ is a (measurable) selection of $\partial f_t(x)$. The theorem below is valid in any of the three following cases:

(a) T is countable.
(b) X is separable.
(c) T is a separable metric space, μ is a regular measure, and the mapping $t \to \partial f_t(y)$ is upper semicontinuous (weak*) for each y in U.

2.7.2 Theorem

Suppose that f is defined at some point x in U. Then f is defined and Lipschitz in U. If at least one of (a), (b), or (c) is satisfied, then formula (1) *holds. If in addition each $f_t(\cdot)$ is regular at x, then f is regular at x and equality holds in expression* (1).

Proof. Let y be any point in U. The map $t \to f_t(y)$ is measurable by hypothesis, and one has

$$\begin{aligned} |f(y) - f(x)| &\leqslant \int_T |f_t(y) - f_t(x)|\,\mu(dt) \\ &\leqslant \int_T k(t)\|y - x\|\,\mu(dt) \quad \text{(by 2.7.1(ii))} \\ &= K\|y - x\|, \end{aligned}$$

where K is given by $\int_T k(t)\,\mu(dt)$. We conclude that f is defined and Lipschitz on U. Now let ζ belong to $\partial f(x)$, and let v be any element of X. Then, by definition,

$$f^{\circ}(x; v) = \limsup_{\substack{y \to x \\ \lambda \downarrow 0}} \int_T \frac{f_t(y + \lambda v) - f_t(y)}{\lambda} \mu(dt).$$

Conditions (i) and (ii) of 2.7.1 allow us to invoke Fatou's Lemma to bring the lim sup under the integral and deduce

$$\int_T f_t^{\circ}(x; v)\, \mu(dt) \geqslant f^{\circ}(x; v) \geqslant \langle \zeta, v \rangle, \tag{2}$$

where the second inequality results from the definition of $\partial f(x)$. Let us define $\hat{f}_t, \hat{f}$ as follows:

$$\hat{f}_t(v) = f_t^{\circ}(x; v), \qquad \hat{f}(v) = \int_T \hat{f}_t(v)\, \mu(dt).$$

Then each $\hat{f}_t(\cdot)$ is convex (Proposition 2.1.1(a)) and thus, so is $\hat{f}$. If we observe that $\hat{f}_t(0) = \hat{f}(0) = 0$, we may rewrite (2) in the form

$$\hat{f}(v) - \hat{f}(0) \geqslant \langle \zeta, v \rangle,$$

and this holds for all v. In other words, ζ belongs to the subdifferential $\partial \hat{f}(0)$ of the integral functional $\hat{f}(v) = \int_T \hat{f}_t(v)\, \mu(dt)$. The subdifferentials of such convex functionals have been characterized, for example, by Ioffe and Levin (1972); the only requirement to be able to apply their results (which amount to saying that formula (1) is true in the convex case) to our situation is to verify that the map $t \to \hat{f}_t(v)$ is measurable for each v (in each of the possible cases (a), (b), or (c)). Let us postpone this to the lemma below and proceed now to show how the known result in the convex case yields the required conclusion.

Since (1) is true for $\hat{f}$ (where the role of x is played by zero) (by p. 8 of Ioffe and Levin (1972) if (a) or (b), by p. 13 if (c)), then there is a map $t \to \zeta_t$ with $\zeta_t \in \partial \hat{f}_t(0)$, μ-a.e., such that for every v in X,

$$\langle \zeta, v \rangle = \int_T \langle \zeta_t, v \rangle\, \mu(dt).$$

However, $\partial \hat{f}_t(0) = \partial f_t(x)$ by construction, so the result would follow.

An Alternate Proof

The possibility of reducing to the convex case is one of the useful features of generalized gradients. However, one might wish for a self-contained derivation of (1). We sketch one here in a brief digression, picking up the proof above at

(2). If Z denotes the set of all measurable maps $t \to \zeta_t$ with $\zeta_t \in \partial f_t(x)$ μ-a.e., then we have

$$\int_T f_t^\circ(x; v)\,\mu(dt) = \int_T \max\langle \partial f_t(x), v\rangle\,\mu(dt)$$

$$= \max_Z \int_T \langle \zeta_t, v\rangle\,\mu(dt)$$

(by a measurable selection theorem). This allows us to write (2) as follows:

$$\min_{v\in X}\max_Z\left\{\int_T \langle \zeta_t, v\rangle\,\mu(dt) - \langle \zeta, v\rangle\right\} = 0.$$

Next we apply the "lop-sided" minimax theorem of Aubin (1978a) to deduce the existence of an element $\{\zeta_t\}$ of Z such that

$$\min_{v\in X}\left\{\int_T \langle \zeta_t, v\rangle\,\mu(dt) - \langle \zeta, v\rangle\right\} = 0.$$

It follows that for each v, the quantity in braces vanishes, and so once again we have expressed ζ in the required form.

Back to the Proof. As in our first proof, this one imposes measurability (selection) requirements. Here is the missing link we promised earlier:

Lemma. In either of cases (a), (b), or (c), the map $t \to \hat{f}_t(v) = f_t^\circ(x; v)$ is measurable for each v.

In case (a) this is automatic; so consider (b). Since $f_t(\cdot)$ is continuous, we may express $f_t^\circ(x; v)$ as the upper limit of

$$\frac{f_t(y+\lambda v) - f_t(y)}{\lambda}, \tag{3}$$

where $\lambda \downarrow 0$ taking rational values, and $y \to x$ taking values in a countable dense subset $\{x_i\}$ of X. But expression (3) defines a measurable function of t by hypothesis, so $f_t^\circ(x; v)$, as the "countable lim sup" of measurable functions of t, is measurable in t.

To complete the proof of the lemma, we consider now case (c). Consider the formula

$$f_t^\circ(x; v) = \max\{\langle \zeta, v\rangle : \zeta \in \partial f_t(x)\}.$$

Since now the multifunction $t \to \partial f_t(x)$ is upper semicontinuous, and weak*-

compact for each t, it follows routinely that $f_t^\circ(x; v)$ is upper semicontinuous as a function of t, and therefore measurable. The lemma is proved.

The Final Steps. There remain the regularity and equality assertions of the theorem to verify. When f_t is regular at x for each t, the existence of $f'(x; v)$ for any v in X and the equality

$$f'(x; v) = \int_T f_t'(x; v)\,\mu(dt) \tag{4}$$

follow from the dominated convergence theorem. The left-hand side of (4) is bounded above by $f^\circ(x; v)$, and the right-hand side coincides with $\int_T f_t^\circ(x; v)\,\mu(dt)$, which is no less than $f^\circ(x; v)$ by (2). It follows that the terms in (4) coincide with $f^\circ(x; v)$, and hence that f is regular.

Finally, let $\zeta = \int_T \langle \zeta_t, \cdot \rangle\,\mu(dt)$ belong to the right-hand side of relation (1). Then, since $\zeta_t \in \partial f_t(x)$ μ-a.e., one has

$$\begin{aligned} f^\circ(x; v) &= \int_T f_t^\circ(x; v)\,\mu(dt) \\ &\geqslant \int_T \langle \zeta_t, v \rangle\,\mu(dt) \\ &= \langle \zeta, v \rangle, \end{aligned}$$

which shows that ζ belongs to $\partial f(x)$, and completes the proof. □

A type of integral functional which occurs frequently in applications is that in which X is a space of functions on T and $f_t(x) = g(t, x(t))$. When the "evaluation map" $x \to x(t)$ is continuous (e.g., when $X = C[0, 1]$), Theorem 2.7.2 will apply. However, there are important instances when it is not, notably when X is an L^p space. We now treat this situation for $p = \infty$, and subsequently for $1 \leqslant p < \infty$.

Integrals on Subspaces of L^∞

We suppose now that $(T, \mathfrak{T}, \mu)$ is a σ-finite positive measure space and Y a separable Banach space. $L^\infty(T, Y)$ denotes the space of measurable essentially bounded functions mapping T to Y, equipped with the usual supremum norm. We are also given a closed subspace X of $L^\infty(T, Y)$, and a family of functions $f_t\colon Y \to R$ $(t \in T)$. We define a function f on X by the formula

$$f(x) = \int_T f_t(x(t))\,\mu(dt).$$

Suppose that f is defined (finitely) at a point x. We wish to characterize $\partial f(x)$. To this end, we suppose that there exist $\varepsilon > 0$ and a function k in $L^1(T, R)$ such that, for all $t \in T$, for all v_1, v_2 in $x(t) + \varepsilon B_Y$,

$$|f_t(v_1) - f_t(v_2)| \leqslant k(t)\|v_1 - v_2\|_Y.$$

We also suppose that the mapping $t \to f_t(v)$ is measurable for each v in Y.

2.7.3 Theorem

Under the hypotheses described above, f is Lipschitz in a neighborhood of x and one has

$$\partial f(x) \subset \int_T \partial f_t(x(t))\, \mu(dt). \tag{5}$$

Further, if f_t is regular at $x(t)$ for each t, then f is regular at x and equality holds.

Proof. Let us note first that the right-hand side of expression (5) is interpreted analogously to that of (1) in Theorem 2.7.2. Thus it consists of those maps $\zeta(\cdot) = \int_T \langle \zeta_t, \cdot \rangle\, \mu(dt)$ such that $\zeta_t \in \partial f_t(x(t)) \subset Y^*$ μ-a.e., and such that for any v in X, $\int_T \langle \zeta_t, v(t) \rangle\, \mu(dt)$ is defined and equal to $\langle \zeta, v \rangle$.

The fact that f is defined and Lipschitz near x follows easily, as in Theorem 2.7.2. To prove (5), suppose that ζ belongs to $\partial f(x)$, and let v be any element of X.

Let $\hat{f}_t(\cdot)$ signify the function $f_t^\circ(x(t); \cdot)$. The measurability of $t \to \hat{f}_t(v)$ follows as in the lemma in Theorem 2.7.2, case (b), and of course $\hat{f}_t(\cdot)$ is continuous. It follows that the map $t \to \hat{f}_t(v(t))$ is measurable, and arguing precisely as in the proof of Theorem 2.7.2, we deduce

$$\int_T f_t^\circ(x(t); v(t))\, \mu(dt) \geqslant f^\circ(x; v) \geqslant \langle \zeta, v \rangle. \tag{6}$$

This can be interpreted as saying that ζ belongs to the subdifferential at 0 of the integral functional on X defined by

$$\hat{f}(v) = \int_T \hat{f}_t(v(t))\, \mu(dt).$$

The requisites of Ioffe and Levin (1972, p. 22) are met and, as in Theorem 2.7.2, the truth of relation (5) for $\hat{f}$, $\hat{f}_t$ at 0 immediately gives (5) for f, f_t.

We remark that a direct argument avoiding recourse to convex analysis is available, one quite analogous to the one sketched in the proof of Theorem 2.7.2.

Regularity. Now suppose that each f_t is regular at $x(t)$. Fix any v in X and set

$$\theta = \liminf_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda}.$$

Invoking Fatou's Lemma, we derive

$$f^\circ(x; v) \geqslant \theta \geqslant \int_T f_t'(x(t); v(t))\, \mu(dt)$$

$$= \int_T f_t^\circ(x(t); v(t))\, \mu(dt) \geqslant f^\circ(x; v) \quad \text{(by (6))}.$$

Consequently all are equal and f is regular at x.

Finally, if $\zeta = \int_T \langle \zeta_t, \cdot \rangle\, \mu(dt)$ belongs to the right-hand side of expression (5), then

$$\langle \zeta, v \rangle = \int_T \langle \zeta_t, v(t) \rangle\, \mu(dt)$$

$$\leqslant \int_T f_t^\circ(x(t); v(t))\, \mu(dt)$$

$$= \int_T f_t'(x(t); v(t))\, \mu(dt) = f'(x; v),$$

by the preceding calculation. Thus ζ belongs to $\partial f(x)$, and the proof is complete. □

2.7.4 *Example (Variational Functionals)*

A common type of functional in the calculus of variations is the following:

$$J(y) = \int_a^b L(t, y(t), \dot{y}(t))\, dt,$$

where y is an absolutely continuous function from $[a, b]$ to R^n. We can calculate ∂J by means of Theorem 2.7.3 with the following cast of characters:

$(T, \mathcal{T}, \mu) = [a, b]$ with Lebesgue measure, $Y = R^n \times R^n$

$$X = \left\{ (s, v) \in L^\infty(T, Y) : \text{for some } c \text{ in } R^n,\ s(t) = c + \int_a^t v(\tau)\, d\tau \right\}$$

$f_t(s, v) = L(t, s, v)$.

Note that for any (s, v) in X, we have $f(s, v) = J(s)$. With $(\hat{s}, \hat{v})$ a given element of X (so that $\hat{v} = (d/dt)\hat{s}$), we assume that the integrand L is measurable in t, and that for some $\varepsilon > 0$ and some function $k(\cdot)$ in $L^1[a, b]$, one has

$$|L(t, s_1, v_1) - L(t, s_2, v_2)| \leqslant k(t)|(s_1 - s_2, v_1 - v_2)|$$

for all (s_i, v_i) in $(\hat{s}(t), \hat{v}(t)) + \varepsilon B$. It is easy to see that the hypotheses of the theorem are satisfied, so that if ζ belongs to $\partial f(\hat{s}, \hat{v})$, we deduce the existence of a measurable function $(q(t), p(t))$ such that

$$(q(t), p(t)) \in \partial L(t, \hat{s}(t), \hat{v}(t)) \quad \text{a.e.}$$

(where ∂L denotes generalized gradient with respect to (s, v)), and where, for any (s, v) in X, one has

$$\langle \zeta, (s, v) \rangle = \int_a^b \{q(t) \cdot s(t) + p(t) \cdot v(t)\}\, dt.$$

In particular, if $\zeta = 0$ (as when J attains a local minimum at $\hat{s}$), it then follows easily that $p(\cdot)$ is absolutely continuous and that $q = \dot{p}$ a.e.

In this case then we have

$$(\dot{p}, p) \in \partial L(t, \hat{s}, \dot{\hat{s}}) \quad \text{a.e.},$$

which implies the classical Euler–Lagrange equation if L is C^1:

$$\frac{d}{dt} L_v(t, \hat{s}, \dot{\hat{s}}) = L_s(t, \hat{s}, \dot{\hat{s}}) \quad \text{a.e.}$$

Integral Functionals on L^p

We assume now that $(T, \mathfrak{T}, \mu)$ is a positive complete measure space with $\mu(T) < \infty$, and that Y is a separable Banach space. Let X be a closed subspace of $L^p(T, Y)$ (for some p in $[1, \infty)$), the space of p-integrable functions from T to Y. We define a functional f on X via

$$f(x) = \int_T f_t(x(t))\, \mu(dt),$$

where $f_t \colon Y \to R$ $(t \in T)$ is a given family of functions. We shall again suppose that for each y in Y the function $t \to f_t(y)$ is measurable, and that x is a point at which $f(x)$ is defined (finitely). We shall characterize $\partial f(x)$ under either of two additional hypotheses. Let q solve $1/p + 1/q = 1$ $(q = \infty$ if $p = 1)$.

Hypothesis A. There is a function k in $L^q(T, R)$ such that, for all $t \in T$,

$$|f_t(y_1) - f_t(y_2)| \leqslant k(t)\|y_1 - y_2\|_Y \quad \text{for all } y_1, y_2 \text{ in } Y.$$

Hypothesis B. Each function f_t is Lipschitz (of some rank) near each point of Y, and for some constant c, for all $t \in T$, $y \in Y$, one has

$$\zeta \in \partial f_t(y) \quad \text{implies} \quad \|\zeta\|_{Y^*} \leqslant c\{1 + \|y\|_Y^{p-1}\}.$$

2.7.5 *Theorem*

Under the conditions described above, under either of Hypotheses A or B, f is uniformly Lipschitz on bounded subsets of X, and one has

$$(7) \qquad \partial f(x) \subset \int_T \partial f_t(x(t))\, \mu(dt).$$

Further, if each f_t is regular at $x(t)$ then f is regular at x and equality holds.

2.7.6 *Remark*

The interpretation of expression (7) is in every way analogous to expressions (1) and (5) of Theorems 2.7.2 and 2.7.3. When $X = L^p(T, Y)$, then any element ζ of $\partial f(x)$ must belong to $L^q(T, Y^*) = X^*$.

Proof (f Lipschitz). That f is defined and globally Lipschitz on X when Hypothesis A holds is immediate, for one has

$$\begin{aligned} |f(x_1) - f(x_2)| &\leqslant \int_T |f_t(x_1(t)) - f_t(x_2(t))|\, \mu(dt) \\ &\leqslant \int_T k(t)\|x_1(t) - x_2(t)\|_Y \mu(dt) \\ &\leqslant K\|x_1 - x_2\|_X, \end{aligned}$$

the last by Hölder's inequality, where $K = \|k\|_q$. We now prove under Hypothesis B that f is uniformly Lipschitz on bounded subsets of X. Let any $m > \|x\|_X$ be given, and let u be any element of X satisfying $\|u\|_p \leqslant m$. By the mean-value theorem we write

$$(8) \qquad f_t(u(t)) - f_t(x(t)) = \langle \zeta_t, u(t) - x(t) \rangle,$$

where $\zeta_t \in \partial f_t(x_t^*)$ and x_t^* lies in the interval $[u(t), x(t)]$. It follows from

Hypothesis B that

$$\begin{aligned}\|\zeta_t\|_{Y^*} &\leqslant c\{1 + \|x_t^*\|^{p-1}\}\\ &\leqslant c\{1 + \|u(t)\|^{p-1} + \|x(t)\|^{p-1}\}.\end{aligned}$$

Let us call this last quantity $\theta(t)$; note $\theta \in L^q(T, R)$. If we substitute this into Eq. (8), take absolute values, and integrate, we obtain

$$\begin{aligned}\int_T |f_t(u(t)) - f_t(x(t))|\, \mu(dt) &\leqslant \int_T \theta(t)\|u(t) - x(t)\|\, \mu(dt)\\ &\leqslant \|\theta\|_q \|u - x\|_p \quad \text{(by Hölder's inequality)}.\end{aligned}$$

It is easy to see that $\|\theta\|_q$ is bounded above by a number K depending only on c and m (but not on u!). We deduce then that $f(u)$ is not only finite, but satisfies

$$|f(u) - f(x)| \leqslant K\|u - x\|_p = K\|u - x\|_X.$$

The process by which we got this inequality can now be repeated for x replaced by any v satisfying $\|v\|_p \leqslant m$, so f is indeed Lipschitz on bounded subsets of X.

Deriving (7). In Theorem 2.7.2 we began by proving that for any v in X, for any ζ in $\partial f(x)$, one has

$$\text{(9)} \qquad \int_T f_t^\circ(x(t); v(t))\, \mu(dt) \geqslant f^\circ(x; v) \geqslant \langle \zeta, v \rangle.$$

Under Hypothesis A this is done just as in Theorem 2.7.2. Under Hypothesis B the process is essentially the same, except that the use of Fatou's Lemma must be justified by appealing to Eq. (8) (for $u = x + \lambda v$). Again we let $\hat{f}_t(v) = f_t^\circ(x(t); v)$, and we interpret (9) as saying that ζ belongs to the subgradient at 0 of the convex functional $\hat{f}(v) = \int_T \hat{f}_t(v(t))\, \mu(dt)$. This remains true if we restrict v to $X \cap L^\infty(T, Y)$, which is a closed subspace of $L^\infty(T, Y)$. Consequently we can apply Theorem 2.7.3 to $\hat{f}$, provided we can verify its hypotheses. Postponing that to the lemma below, we conclude that ζ is expressible in the form $\zeta(\cdot) = \int_T \langle \zeta_t, \cdot \rangle\, \mu(dt)$, where $\zeta_t \in \partial \hat{f}_t(0) = \partial f_t(x(t))$. This proves relation (7).

Lemma. The hypotheses of Theorem 2.7.3 are satisfied when $x = 0$, $f_t = \hat{f}_t$.

The required measurability of $t \to \hat{f}_t(v)$ follows exactly as in case (b) of the lemma in the proof of Theorem 2.7.2. Of course, $\int_T \hat{f}_t(0)\, \mu(dt) = \hat{f}(0)$ is defined, so we need only verify the Lipschitz hypothesis. Now, the set $\partial f_t(x(t))$

is bounded by an integrable function $r(t)$: if Hypothesis A holds, take $r = k$ (in view of Proposition 2.1.2(a)); if Hypothesis B holds, take $r = c\{1 + \|x(t)\|_Y^{p-1}\}$. Thus $\hat{f}_t(\cdot)$, as the support function of $\partial f_t(x(t))$, is globally Lipschitz with constant $r(t)$, and the lemma is proved.

Regularity. We now turn to verifying the last conclusion of the theorem. For any v in X, one derives

$$\liminf_{\lambda \downarrow 0} \frac{f(x + \lambda v) - f(x)}{\lambda} \geqslant \int_T f_t'(x(t); v(t))\, \mu(dt)$$

$$= \int_T f_t^\circ(x(t)); v(t))\, \mu(dt) \geqslant f^\circ(x; v).$$

It follows that $f'(x; v)$ exists and equals $f^\circ(x; v)$; that is, that f is regular at x.

Given any ζ in the right-hand side of expression (7), one has

$$\langle \zeta, v \rangle = \int_T \langle \zeta_t, v(t) \rangle\, \mu(dt)$$

$$\leqslant \int_T f_t'(x(t); v(t))\, \mu(dt) = f^\circ(x; v),$$

and so $\zeta \in \partial f(x)$. Thus relation (7) holds with equality. □

2.8 POINTWISE MAXIMA

A General Formula

Functionals which are explicitly expressed as pointwise maxima of some indexed family of functions will play an important role in later chapters. We have already studied the case in which the index set is finite, in Proposition 2.3.12. This section is devoted to the much more complex situation in which the family is an infinite one.

Let f_t be a family of functions on X parametrized by $t \in T$, where T is a topological space. Suppose that for some point x in X, each function f_t is Lipschitz near x. We shall find it convenient to define a new kind of partial generalized gradient, one which takes account of variations in parameters other than the primary one.

We denote by $\partial_{[T]} f_t(x)$ the set ($\overline{\text{co}}$ denotes weak*-closed convex hull)

$$\overline{\text{co}}\{\zeta \in X^* : \zeta_i \in \partial f_{t_i}(x_i),$$

$$x_i \to x,\ t_i \to t,\ t_i \in T,\ \zeta \text{ is a weak* cluster point of } \zeta_i\}.$$

2.8.1 Definition

The multifunction $(\tau, y) \to \partial f_\tau(y)$ is said to be (weak*) closed at (t, x) provided $\partial_{[T]} f_t(x) = \partial f_t(x)$.

(Note that this condition certainly holds if t is isolated in T (in view of Proposition 2.1.5(b)).)

We make the following hypotheses:

(i) T is a sequentially compact space.
(ii) For some neighborhood U of x, the map $t \to f_t(y)$ is upper semicontinuous for each y in U.
(iii) Each f_t, $t \in T$, is Lipschitz of given rank K on U, and $\{f_t(x) : t \in T\}$ is bounded.

We define a function f: $X \to R$ via

$$f(y) = \max\{f_t(y) : t \in T\},$$

and we observe that our hypotheses imply that f is defined and finite (with the maximum defining f attained) on U. It also follows readily that f is Lipschitz on U (of rank K), since each f_t is.

We denote by $M(y)$ the set $\{t \in T: f_t(y) = f(y)\}$. It is easy to see that $M(y)$ is nonempty and closed for each y in U. Finally, for any subset S of T, $P[S]$ signifies the collection of probability Radon measures supported on S.

2.8.2 Theorem

In addition to the hypotheses given above, suppose that either

(iv) *X is separable, or*
(iv)′ *T is metrizable (which is true in particular if T is separable).*

Then one has

$$\partial f(x) \subset \left\{ \int_T \partial_{[T]} f_t(x)\, \mu(dt) : \mu \in P[M(x)] \right\}. \tag{1}$$

Further, if the multifunction $(\tau, y) \to \partial f_\tau(y)$ is closed at (t, x) for each $t \in M(x)$, and if f_t is regular at x for each t in $M(x)$, then f is regular at x and equality holds in expression (1) (with $\partial_{[T]} f_t(x) = \partial f_t(x)$).

2.8.3 Remark

The interpretation of the set occurring on the right-hand side of expression (1) is completely analogous to that of (1) in Theorem 2.7.2. Specifically, an element ζ of that set is an element of X^* to which there corresponds a mapping $t \to \zeta_t \in \partial_{[T]} f_t(x)$ from T to X^* and an element μ of $P[M(x)]$ such that, for every v in X, $t \to \langle \zeta_t, v \rangle$ is μ-integrable, and

$$\langle \zeta, v \rangle = \int_T \langle \zeta_t, v \rangle \, \mu(dt).$$

Let us note two important special cases before proving the theorem.

Corollary 1

In addition to the basic hypotheses (i)–(iii) and (iv) or (iv)′, assume that U is convex, and that $f_t(x)$ is continuous as a function of t and convex as a function of x. Then f is convex on U, and, for each x in U,

$$\partial f(x) = \left\{ \int_T \partial f_t(x) \, \mu(dt) : \mu \in P[M(x)] \right\}.$$

Proof. A pointwise maximum of convex functions is convex, so f is convex in this case. Because convex functions are regular (Proposition 2.3.6), the corollary follows immediately from the theorem once one establishes that the multifunction $(t, x) \to \partial f_t(x)$ is closed. To see this, let ζ_i belong to $\partial f_{t_i}(x_i)$, where $x_i \to x$, $t_i \to t$, and let ζ be a cluster point of $\{\zeta_i\}$. We wish to show that ζ belongs to $\partial f_t(x)$. Now for any v sufficiently small,

$$f_{t_i}(x_i + v) - f_{t_i}(x_i) \geqslant \langle \zeta_i, v \rangle.$$

We deduce (from hypothesis (iii))

$$f_{t_i}(x + v) - f_{t_i}(x) \geqslant \langle \zeta_i, v \rangle - 2K\|x_i - x\|.$$

There is a subsequence $\{\zeta_i'\}$ such that $\lim \langle \zeta_i', v \rangle = \langle \zeta, v \rangle$. Taking limits along this subsequence yields

$$f_t(x + v) - f_t(x) \geqslant \langle \zeta, v \rangle.$$

This implies

$$\zeta \in \partial f_t(x). \quad \square$$

Corollary 2

In addition to the basic hypotheses (i)–(iii) and (iv) or (iv)′, assume that each f_t admits a strict derivative $D_s f_t(x)$ on U, and that $D_s f_t(x)$ is continuous as a function of (t, x). Then for each $x \in U$, f is regular at x and one has

$$\partial f(x) = \left\{ \int_T D_s f_t(x) \, \mu(dt) : \mu \in P[M(x)] \right\}.$$

This follows immediately from the theorem, since $\partial f_t(x) = \{D_s f_t(x)\}$ by Proposition 2.2.4.

2.8.4 Example (Supremum Norm)

Let us apply Corollary 1 to obtain a result originally due to Banach. Let X be the space of continuous functions x mapping the interval $[a, b]$ to R^n, and define

$$\|x\| = \max_{a \leqslant t \leqslant b} |x(t)|.$$

At which points x in X is this norm differentiable?

We appeal to Corollary 1 in the case:

$$T = [a, b], \qquad f_t(x) = |x(t)|.$$

Since $f(x) = \|x\|$ is convex, f is regular by Proposition 2.3.6 and therefore differentiable iff $\partial f(x)$ reduces to a singleton. In view of the corollary, this happens only when $M(x)$ consists of just one point. We have proved: $\|x\|$ is differentiable at those x which admit a unique t maximizing $|x(t)|$.

Proof of Theorem 2.8.2

Step 1. Let us define the function $g: U \times X \to R$ as follows:

$$g(x; v) = \max\{\langle \zeta, v \rangle : \zeta \in \partial_{[T]} f_t(x), t \in M(x)\}. \tag{2}$$

The use of max is justified by the fact that $\partial_{[T]} f_t$ is weak*-compact (by hypothesis (iii)) and $M(x)$ is closed. We wish to establish, for any v in X,

Lemma. $f^\circ(x; v) \leqslant g(x; v)$.

To see this, let $y_i \to x$ and $\lambda_i \downarrow 0$ be sequences such that the terms

$$\Delta_i = \frac{f(y_i + \lambda_i v) - f(y_i)}{\lambda_i}$$

converge to $f^\circ(x; v)$. Pick any $t_i \in M(y_i + \lambda_i v)$. Then it follows that

$$\Delta_i \leqslant \frac{f_{t_i}(y_i + \lambda_i v) - f_{t_i}(y_i)}{\lambda_i}. \tag{3}$$

By the mean-value theorem, the expression on the right-hand side of (3) may be expressed $\langle \zeta_i, v \rangle$, where ζ_i belongs to $\partial f_{t_i}(y_i^*)$ for some y_i^* between y_i and $y_i + \lambda_i v_i$. By extracting a subsequence (without relabeling) we may suppose that $t_i \to t \in T$, and that $\{\zeta_i\}$ admits a weak* cluster point with $\lim_{i\to\infty}\langle \zeta_i, v\rangle = \langle \zeta, v\rangle$. Note that y_i^* has no choice but to converge to x, so that $\zeta \in \partial_{[T]} f_t(x)$ by definition. We derive from all this, in view of (3), the following:

$$f^\circ(x; v) = \lim_{i\to\infty} \Delta_i \leqslant \langle \zeta, v\rangle.$$

Now the lemma will follow if $t \in M(x)$, for then $\langle \zeta, v\rangle \leqslant g(x; v)$ by definition. To see that this is the case, observe that for any τ in T, one has

$$f_{t_i}(y_i + \lambda_i v) \geqslant f_\tau(y_i + \lambda_i v),$$

(since $t_i \in M(y_i + \lambda_i v)$), whence (by hypothesis (iii)):

$$f_{t_i}(x) \geqslant f_\tau(x) - 2K\|y_i + \lambda_i v - x\|.$$

From hypothesis (ii) we now conclude, upon taking limits:

$$f_t(x) \geqslant f_\tau(x).$$

Since τ is arbitrary, it follows that $t \in M(x)$, and the lemma is proved.

Step 2. Now let ζ be an element of $\partial f(x)$. We deduce from the lemma that, for all v in X,

$$g(x; v) \geqslant \langle \zeta, v\rangle. \tag{4}$$

Since $g(x; 0) = 0$, this means that ζ belongs to the subdifferential of the convex function $g(x; \cdot)$ at 0, which has been characterized in convex analysis, for example by Ioffe and Levin (1972, p. 33 if (iv) holds, or p. 34 if (iv)$'$ holds). It follows that ζ can be expressed in the form $\int_T \langle \zeta_t, \cdot\rangle\, \mu(dt)$ for some μ in $P[M(x)]$ and measurable mapping $t \to \zeta_t$ assuming values in $\partial_{[T]} f_t(x)$ μ-a.e.. This establishes expression (1).

It is possible to proceed directly rather than appeal to convex analysis. We sketch the alternate approach before proceeding with the proof of the theorem. Let us call the right-hand side of expression (1) Z. Note that Z is the weak*-closed convex hull of all points in $\partial_{[T]} f_t(x)$ for t in $M(x)$. Then (4) and the lemma assert

$$\min_{v\in X} \max_{w\in Z} \{\langle w, v\rangle - \langle \zeta, v\rangle\} = 0.$$

We apply Aubin's lop-sided Minimax Theorem to deduce the existence of a point w in Z, satisfying, for all v in X,

$$\langle w - \zeta, v\rangle \geqslant 0.$$

It follows that $\zeta = w$, and we obtain expression (1) as before.

Step 3. We now assume the additional hypotheses. Let

$$\alpha = \liminf_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda}.$$

To conclude that f is regular at x, it suffices to show that $f^\circ(x;v) \leqslant \alpha$. Pick any t in $M(x)$. One has

$$\frac{f(x+\lambda v) - f(x)}{\lambda} \geqslant \frac{f_t(x+\lambda v) - f_t(x)}{\lambda},$$

and taking the lower limit of each side yields $\alpha \geqslant f_t'(x;v) = f_t^\circ(x;v)$, the last equality resulting from the fact that f_t is regular at x. We conclude

$$\alpha \geqslant \max\{f_t^\circ(x;v) : t \in M(x)\}. \tag{5}$$

By Definition 2.8.1, we have $\partial_{[T]} f_t(x) = \partial f_t(x)$, so the function $g(x;v)$ defined at the beginning of the proof reduces to $\max\{\langle \zeta, v\rangle : \zeta \in \partial f_t(x), t \in M(x)\}$, which is precisely the right-hand side of (5). The lemma of Step 1, together with (5), thus combine to imply $f^\circ(x;v) \leqslant \alpha$. As noted, this proves that f is regular at x.

In order to show that expression (1) holds with equality, let $\zeta = \int_T \langle \zeta_t, \cdot\rangle\, \mu(dt)$ be any element of the right-hand side of (1). Then, for any v in X,

$$\begin{aligned}
\langle \zeta, v\rangle &= \int_T \langle \zeta_t, v\rangle\, \mu(dt) \\
&\leqslant \int_T f_t^\circ(x;v)\, \mu(dt) = \int_T f_t'(x;v)\, \mu(dt) \\
&= \lim_{\lambda \downarrow 0} \frac{\int f_t(x+\lambda v)\, \mu(dt) - \int f_t(x)\, \mu(dt)}{\lambda}
\end{aligned}$$

(by dominated convergence)

$$\leqslant \limsup_{\lambda \downarrow 0} \frac{\int f(x+\lambda v)\, \mu(dt) - \int f(x)\, \mu(dt)}{\lambda}$$

(since $f_t(x) = f(x)$ for $t \in M(x)$ and μ is supported on $M(x)$)

$$\begin{aligned}
&= \limsup_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda} \\
&= f'(x;v) = f^\circ(x;v) \quad \text{(as just proven above).}
\end{aligned}$$

We have shown $\langle \zeta, v \rangle \leqslant f^\circ(x; v)$. Thus $\zeta \in \partial f(x)$, and the proof of the theorem is complete. □

When there is Lipschitz behavior of f_t in t (as well as in x), it has been demonstrated by Hiriart-Urruty (1978) that a different sort of estimate for ∂f is possible. We use Theorem 2.8.2 to extend his result to the case in which X is infinite-dimensional.

We now write $\phi(t, x)$ for $f_t(x)$, and we make a single hypothesis which subsumes all the hypotheses (i)–(iii) and (iv)′:

(v) T is a compact subset of a finite-dimensional Banach space Y, and for some $\varepsilon > 0$, for all t_1, t_2 in $T + 2\varepsilon B_Y$, and for all x_1, x_2 in U, one has

$$|\phi(t_1, x_1) - \phi(t_2, x_2)| \leqslant K(\|t_1 - t_2\|_Y + \|x_1 - x_2\|_X).$$

As usual, d_T denotes the distance function of T.

Corollary 3

If z belongs to $\partial f(x)$, then $(0, z)$ belongs to the set

$$\overline{\mathrm{co}}\{\partial\phi(t, x) - \hat{K}\,\partial d_T(t) \times \{0\} : t \in M(x)\}, \tag{6}$$

where $\hat{K}$ is any number greater than K.

Proof. We are going to place ourselves in the context of Theorem 2.8.2, but with altered data. We define

$$\hat{T} = T + \varepsilon B_Y, \qquad \hat{U} = \varepsilon B_Y \times U \subset Y \times X$$

$$\hat{f}_t(y, x) = \phi(t + y, x) - \hat{K} d_T(t + y)$$

$$\hat{f}(y, x) = \max\{\hat{f}_t(y, x) : t \in \hat{T}\}.$$

The mapping $(t, y, x) \to \partial\hat{f}_t(y, x)$ is closed because so are $\partial\phi$ and ∂d_T, so that $\partial_{[\hat{T}]}\hat{f}_t$ and $\partial\hat{f}_t$ coincide (Definition 2.8.1). The basic hypotheses of the theorem are satisfied. We conclude that $\partial\hat{f}(0, x)$ is contained in the set

$$\overline{\mathrm{co}}\{\partial\phi(t, x) - \hat{K}\,\partial d_T(t) \times \{0\} : t \in \hat{M}(0, x)\}, \tag{7}$$

where we have replaced $\partial\hat{f}_t(0, x)$ by the (possibly larger) set $\partial\phi(t, x) - \hat{K}\,\partial d_T(t) \times \{0\}$ (Proposition 2.3.3).

Lemma. For all (y, x) in $\hat{U}$, $\hat{M}(y, x) = M(x) - y$ and $\hat{f}(y, x) = f(x)$.

Note that the second assertion follows from the first by substitution. Once the lemma is proven, the resulting formula $\partial\hat{f}(0, x) = \{0\} \times \partial f(x)$, together with the fact that expression (6) and (7) coincide, immediately yield the corollary. To prove the lemma's first assertion, note that t lies in $\hat{M}(y, x)$ iff $\tau = t + y$ maximizes $\phi(\cdot, x) - \hat{K}d_T(\cdot)$ over $\hat{T} + y$. By Proposition 2.4.3 (with $S = \hat{T} + y$, $C = T$), this is equivalent to the statement that τ lies in T and maximizes $\phi(\cdot, x)$ over T; that is, that τ belongs to $M(x)$. □

2.8.5 *Remark*

When T is a subset of a finite-dimensional Banach space Y, as in Corollary 3 above, and when $M(x)$ admits a locally bounded selection $m(x)$, we may replace $M(x)$ in the preceding results by the set $\tilde{M}(x)$ = all limit points of $m(y)$ as y approaches x. (This observation is also due to Hiriart-Urruty.) To see this, simply apply the theorem to the modified problem in which T is replaced by $(\tilde{M}(x) + \varepsilon B_Y) \cap T$. For arbitrarily small $\varepsilon > 0$, this has no effect on f locally.

When X is Finite-Dimensional

When $X = R^n$ we can exploit the characterization of ∂f given by Theorem 2.5.1 to obtain a result along the lines of Theorem 2.8.2, but requiring far less structure on the data. As far as T is concerned, we now let it be simply a given abstract set. We define

$$f(x) = \sup\{f_t(x) : t \in T\}$$

as before, but our hypotheses will no longer imply that this supremum is attained. Now we posit only hypothesis (iii), namely,

$$|f_t(x_1) - f_t(x_2)| \leqslant K\|x_1 - x_2\|$$

for $t \in T$, x_1, x_2 in U. We assume that f is finite at some point in U, from which it follows easily that f is finite and Lipschitz on all of U.

2.8.6 *Theorem*

Let x be any point in U, and let S be a subset of measure 0 in U. Then $\partial f(x)$ is contained in the set C defined by

$$\operatorname{co}\Big\{\lim_{i\to\infty} \nabla f_{t_i}(x_i) : x_i \to x, x_i \notin S, t_i \in T, f_{t_i}(x) \to f(x)\Big\}.$$

Proof. The definition of C expresses it as the convex hull of all points z of the form $\lim_{i\to\infty}\nabla f_{t_i}(x_i)$, where $x_i \notin S$, $\nabla f_{t_i}(x_i)$ exists, $x_i \to x$, and t_i is a maximizing sequence for $f_t(x)$ in T. It is easy to see that C is nonempty and compact. Let v be any point in R^n, and define

$$m = \max\{\langle \zeta, v\rangle : \zeta \in C\}.$$

It suffices to prove that for any $\varepsilon > 0$, one has

$$f^\circ(x; v) \leqslant m + \varepsilon, \tag{8}$$

since $f^\circ(x; \cdot)$ is the support function of $\partial f(x)$. We may assume $|v| = 1$.

The definition of C implies the existence of some δ in $(0, 1)$ such that whenever $y \in x + 2\delta B$ and $t \in T$ satisfy

$$y \notin S, \qquad \nabla f_t(y) \text{ exists}, \qquad f_t(x) \geqslant f(x) - \delta,$$

then one has

$$\nabla f_t(y) \cdot v \leqslant m + \varepsilon. \tag{9}$$

Let λ be any point in $(0, \delta)$ and t any point in T satisfying $f_t(x) \geqslant f(x) - \delta$. Let Ω be the set of points in U at which $Df_t(\cdot)$ fails to exist. Now let y be any point in $x + \delta B$ such that the line segment $[y, y + \delta v]$ meets $S \cup \Omega$ in a set of 0 one-dimensional measure. (Note that almost all y in $x + \delta B$ have this property, since $S \cup \Omega$ has measure zero.) With the help of (9) we obtain

$$f_t(y + \lambda v) - f_t(y) = \int_0^\lambda \nabla f_t(y + \tau v) \cdot v \, d\tau \leqslant \lambda(m + \varepsilon). \tag{10}$$

Since f_t is continuous, this inequality must in fact hold for all y in $x + \delta B$.

Pick r in $(0, \delta)$ such that $r^2 + 4rK < \delta$.

Lemma. For all y in $x + rB$ and λ in $(0, r)$, one has

$$f(y + \lambda v) - f(y) \leqslant \lambda(m + \varepsilon) + \lambda^2.$$

Clearly this will imply (8), and thus complete the proof.

To prove the lemma, given y and λ as stated, choose any t such that $f_t(y + \lambda v) \geqslant f(y + \lambda v) - \lambda^2$. Then

$$\begin{aligned}
f_t(x) &\geqslant f_t(y + \lambda v) - K\|x - y - \lambda v\| \\
&\geqslant f(y + \lambda v) - \lambda^2 - K\|x - y - \lambda v\| \quad \text{(by choice of } t) \\
&\geqslant f(x) - \lambda^2 - 2K\|x - y - \lambda v\| \\
&\geqslant f(x) - r^2 - 4rK \geqslant f(x) - \delta.
\end{aligned}$$

This shows that (10) is applicable to this t and λ, whence

$$f(y+\lambda v)-f(y) \leqslant f_t(y+\lambda v)-f_t(y)+\lambda^2$$
$$\leqslant \lambda(m+\varepsilon)+\lambda^2,$$

which proves the lemma. □

2.8.7 *Example (Greatest Eigenvalue)*

Let $A(x)$ be an $n \times n$ matrix-valued function of a variable x in R^m. A function which arises naturally as a criterion in many engineering design problems, and which is known to be nondifferentiable in general, is

$$f(x) = \text{greatest eigenvalue of } A(x).$$

We shall assume that $A(\cdot)$ is C^1, and that for each x, $A(x)$ is positive-definite and symmetric. We wish to calculate $\partial f(x)$.

Some notation is in order. We know that there exists for each x a unitary matrix U (depending on x) such that

$$U^* A(x) U = D,$$

where D is the diagonal matrix of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $A(x)$. We shall assume without loss of generality that (for the particular x of interest) $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. For any integer j between 1 and n, let S_j denote the unit vectors w of R^n whose components beyond the jth are zero. (Thus, for example, S_1 consists of the two vectors $[\pm 1, 0, \ldots, 0]$.) Finally, let r be the multiplicity of the eigenvalue λ_1, and let $A_i'(x)$ denote the matrix $A(x)$ differentiated with respect to x_i.

2.8.8 **Proposition**

The function f is Lipschitz and regular at x, and one has

$$\partial f(x) = \mathrm{co}\{[\langle U^* A_1' U w, w\rangle, \ldots, \langle U^* A_m' U w, w\rangle] : w \in S_r\}.$$

f is differentiable at x if λ_1 has multiplicity 1, *and in that case one has*

$$\frac{\partial}{\partial x_i} f(x) = \text{the } (1,1)\text{-entry in } U^*\left[\frac{\partial}{\partial x_i} A(x)\right] U.$$

Proof. From the characterization

$$(11) \qquad f(x) = \max\{\langle A(x) w, v\rangle : w, v \in \bar{B}\}$$

(where B is the unit ball in R^n) combined with Theorem 2.8.2, it ensues that f

is Lipschitz and regular at x. From Corollary 2 we have

$$(12)\quad \partial f(x) = \mathrm{co}\left\{ \frac{d}{dx}\langle A(x)w, v\rangle : (w, v) \in M(x)\right\}$$

$$= \mathrm{co}\{[\langle A_1'(x)w, v\rangle, \ldots, \langle A_m'(x)w, v\rangle] : (w, v) \in M(x)\},$$

where $M(x)$ is the set of maximizing (w, v) in Eq. (11). Let U be the unitary matrix defined earlier. Since $U\overline{B} = \overline{B}$, one has

$$f(x) = \max\{\langle A(x)U\tilde{w}, U\tilde{v}\rangle : \tilde{w}, \tilde{v} \in \overline{B}\}$$

$$= \max\{\langle U^*A(x)U\tilde{w}, \tilde{v}\rangle : \tilde{w}, \tilde{v} \in \overline{B}\}$$

$$= \max\{\langle D\tilde{w}, \tilde{v}\rangle : \tilde{w}, \tilde{v} \in \overline{B}\}.$$

It is a simple exercise to verify that the vectors $(\tilde{w}, \tilde{v})$ at which this last maximum is attained are precisely those of the form $(\tilde{w}, \tilde{w})$, where $\tilde{w}$ belongs to S_r. It follows that the elements (w, v) of $M(x)$ are those of the form $(U\tilde{w}, U\tilde{w})$, where $\tilde{w}$ belongs to S_r. Substitution into Eq. (12) now yields the formula as stated in the proposition. The last assertion is an immediate consequence of this formula. □

2.9 EXTENDED CALCULUS

In Section 2.4 we defined $\partial f(x)$ for any extended-valued function f provided only that it be finite at x. With few exceptions, however, the calculus we have developed has been for locally Lipschitz functions; it is natural to ask to what extent it would be possible to treat more general functions. Recently, significant progress on this issue has been made by Rockafellar, Hiriart-Urruty, and Aubin.

We must expect somewhat sparser results and greater technical detail in such an endeavor. A major contribution of Rockafellar's has been to suitably extend the definition of the generalized directional derivative $f°$, and to identify a subclass of extended-valued functions (those that are "directionally Lipschitz") which is general enough for most applications yet well-behaved enough to lead to several useful extensions of the Lipschitz calculus. These results have been obtained in the broader context of locally convex spaces rather than in Banach spaces, with new limiting concepts. For our purposes, however, it shall suffice to remain in a Banach space X. The results and proofs of this section are due to Rockafellar.

First on the agenda then, is the study of what becomes of $f°$ in the non-Lipschitz case. In view of Theorem 2.4.9 we would expect $f°$ to be such that the epigraph of $f°(x; \cdot)$ is the set $T_{\mathrm{epi}\, f}(x, f(x))$ (assuming for the moment that the latter is an epigraph). But what is the direct characterization of $f°$? The answer below involves some complicated limits for which some additional notation is in order. Following Rockafellar, the expression

$$(y, \alpha) \downarrow_f x$$

shall mean that $(y, \alpha) \in \operatorname{epi} f$, $y \to x$, $\alpha \to f(x)$.

2.9.1 Theorem (Rockafellar)

Let $f\colon X \to R \cup \{\infty\}$ be an extended real-valued function and x a point where f is finite. Then the tangent cone $T_{epi\,f}(x, f(x))$ is the epigraph of the function $f°(x; \cdot)\colon X \to R \cup \{\pm\infty\}$ defined as follows:

$$f°(x; v) = \lim_{\varepsilon \downarrow 0} \limsup_{\substack{(y,\alpha)\downarrow_f x \\ t \downarrow 0}} \inf_{w \in v + \varepsilon B} \frac{f(y + tw) - \alpha}{t}.$$

Proof. We must show that a point (v, r) of $X \times R$ belongs to $T_{\operatorname{epi} f}(x, f(x))$ iff $f°(x; v) \leqslant r$.

Necessity. Let $(v, r) \in T_{\operatorname{epi} f}(x, f(x))$, and let any $\varepsilon > 0$ be given. It suffices to show

$$\limsup_{\substack{(y,\alpha)\downarrow_f x \\ t \downarrow 0}} \inf_{w \in v + \varepsilon B} \frac{f(y + tw) - \alpha}{t} \leqslant r. \tag{1}$$

To see this, let $(y_i, \alpha_i) \downarrow_f x$ and $t_i \downarrow 0$. Since $(v, r) \in T_{\operatorname{epi} f}(x, f(x))$, there exists (by Theorem 2.4.5) a sequence (v_i, r_i) converging to (v, r) such that $(y_i, \alpha_i) + t_i(v_i, r_i) \in \operatorname{epi} f$, that is, such that

$$\alpha_i + t_i r_i \geqslant f(y_i + t_i v_i).$$

Consequently, for i sufficiently large,

$$\inf_{w \in v + \varepsilon B} \frac{f(y_i + t_i w) - \alpha_i}{t_i} \leqslant \frac{f(y_i + t_i v_i) - \alpha_i}{t_i} \leqslant r_i,$$

and (1) follows.

Sufficiency. Suppose $f°(x; v) \leqslant r$ and let (y_i, α_i) be any sequence in $\operatorname{epi} f$ converging to $(x, f(x))$ and t_i any sequence decreasing to 0. To prove that (v, r) belongs to $T_{\operatorname{epi} f}(x, f(x))$, it suffices to produce a sequence (v_i, r_i) converging to (v, r) such that

$$(y_i, \alpha_i) + t_i(v_i, r_i) \in \operatorname{epi} f \qquad \text{infinitely often;}$$

that is,

(2) $$\alpha_i + t_i r_i \geqslant f(y_i + t_i v_i) \quad \text{infinitely often}$$

(by an evident "subsequence version" of the characterization given in Theorem 2.4.5). For each positive integer n, there is an $\varepsilon_n < 1/n$ such that

$$\limsup_{\substack{(y,\alpha)\downarrow_f x\\ t\downarrow 0}} \inf_{w\in v+\varepsilon_n B} \frac{f(y+tw)-\alpha}{t} \leqslant r + \frac{1}{n}$$

(since $f^\circ(x; v) \leqslant r$). Consequently, since $(y_i, \alpha_i)\downarrow_f x$, there must be an index $i = i(n) > n$ such that

$$\inf_{w\in v+\varepsilon_n B} \frac{f(y_i+t_i w)-\alpha_i}{t_i} \leqslant r + \frac{2}{n},$$

and therefore a point v_i in $v + \varepsilon_n B$ such that

(3) $$\frac{f(y_i+t_i v_i)-\alpha_i}{t_i} \leqslant r + \frac{3}{n}.$$

Let us define, for indices i in the subsequence $i(1), i(2),\dots,$

(4) $$r_i = \max\left\{r, \frac{f(y_i+t_i v_i)-\alpha_i}{t_i}\right\}.$$

Note that the inequality in (2) is satisfied, so we shall be through provided we verify that r_i converges to r. But this is evident from (3) and (4), and so the proof is complete. □

It is not hard to see that if f is lower semicontinuous at x, then $f^\circ(x; v)$ is given by the slightly simpler expression

$$\lim_{\varepsilon\downarrow 0} \limsup_{\substack{y\downarrow_f x\\ t\downarrow 0}} \inf_{w\in v+\varepsilon B} \frac{f(y+tw)-f(y)}{t},$$

where $y\downarrow_f x$ signifies that y and $f(y)$ converge to x and $f(x)$, respectively. Note also (in all cases) that the limit over $\varepsilon > 0$ is equivalent to a supremum over $\varepsilon > 0$. As a consequence of the theorem, we see that the extended f° plays the same role vis-à-vis ∂f as it did in the Lipschitz case:

Corollary

One has $\partial f(x) = \emptyset$ iff $f^\circ(x; 0) = -\infty$. Otherwise, one has

$$\partial f(x) = \{\zeta \in X^* : f^\circ(x; v) \geqslant \langle \zeta, v\rangle \text{ for all } v \in X\},$$

and

$$f^\circ(x;v) = \sup\{\langle \zeta, v\rangle : \zeta \in \partial f(x)\}.$$

Proof. ζ belongs to $\partial f(x)$, by definition, iff $\langle(\zeta,-1),(v,r)\rangle \leqslant 0$ for all $(v,r) \in T_{\text{epi} f}(x, f(x)) = \text{epi } f^\circ(x;\cdot)$ (by the theorem). Equivalently, $\zeta \in \partial f(x)$ iff $\langle \zeta, v\rangle \leqslant r$ for all v in X and r in R such that $r \geqslant f^\circ(x;v)$. If $f^\circ(x;v) = -\infty$ for some v, there can be no such ζ; that is, $\partial f(x) = \varnothing$. On the other hand, we have $f^\circ(x;\lambda v) = \lambda f^\circ(x;v)$ for all v and $\lambda > 0$, since epi $f^\circ(x;\cdot)$ is a cone (we use the convention $\lambda(-\infty) = -\infty$). Thus $f^\circ(x;\cdot)$ is $-\infty$ somewhere iff it is $-\infty$ at 0, and the first assertion follows.

Assuming then that $f^\circ(x;\cdot) > -\infty$, we have $\zeta \in \partial f(x)$ iff $\langle\zeta, v\rangle \leqslant f^\circ(x;v)$ for all v, as we have seen above. This is the second assertion, which is equivalent to the statement that $f^\circ(x;\cdot)$ is the support function of $\partial f(x)$; that is, the third assertion (see Proposition 2.1.4). □

In view of Theorem 2.4.9, the expression for f° given in the theorem must reduce, when f is locally Lipschitz, to the one now so familiar from the preceding sections. This is in fact quite easy to show but, again following Rockafellar, we pursue the more ambitious goal of identifying a useful intermediate class of functions f for which this is "essentially" the case.

2.9.2 Definition

f is *directionally Lipschitz at x with respect to $v \in X$* if (f is finite at x and) the quantity

$$f^+(x;v) := \limsup_{\substack{(y,\alpha)\downarrow_f x,\, w\to v \\ t\downarrow 0}} \frac{f(y+tw)-\alpha}{t}$$

is not $+\infty$. We say f is *directionally Lipschitz at x* if f is directionally Lipschitz at x with respect to at least one v in X.

We can relate this concept to a geometrical one introduced in Section 2.4.

2.9.3 Proposition

f is directionally Lipschitz at x with respect to v iff for some $\beta \in R$, (v,β) is hypertangent to epi f at $(x, f(x))$.

Proof. Suppose first that (v,β) is hypertangent to epi f at $(x, f(x))$. Then, for some $\varepsilon > 0$, one has by definition

$$(5) \qquad [(x, f(x)) + \varepsilon B] \cap C + t[(v,\beta) + \varepsilon B] \subset C \quad \text{for } t \in (0,\varepsilon),$$

where $B \doteq B_{X\times R}$ and $C = \text{epi } f$. Consequently, whenever $(y,\alpha) \in C$ is close to $(x, f(x))$, and w is close to v, one has $(y, \alpha) + t(w, \beta) \in C$; that is,

$$\alpha + t\beta \geqslant f(y + tw).$$

It follows that the expression $f^+(x; v)$ in Definition 2.9.2 is bounded above by β, so that f is directionally Lipschitz at x with respect to v.

Now let us assume this last conclusion, and let β be any number greater than $f^+(x; v)$. Then, for all $(y, \alpha) \in \operatorname{epi} f$ near $(x, f(x))$, for all w near v and $t > 0$ near 0, one has

$$\frac{f(y + tw) - \alpha}{t} < \beta.$$

This clearly implies (5) for some ε; that is, that (v, β) is hypertangent to epi f at $(x, f(x))$. □

The proof actually showed:

Corollary

If $f^+(x; v)$ is less than $+\infty$, then it equals

$$\inf\{\beta : (v, \beta) \text{ is hypertangent to epi } f \text{ at } (x, f(x))\}.$$

Here are some criteria assuring that f is directionally Lipschitz.

2.9.4 Theorem

Let f be extended real-valued, and let x be a point at which f is finite. If any of the following hold, then f is directionally Lipschitz at x:

(i) *f is Lipschitz near x.*
(ii) *$f = \psi_C$, where C admits a hypertangent at x.*
(iii) *f is convex, and bounded in a neighborhood of some point (not necessarily x).*
(iv) *f is nondecreasing with respect to the partial ordering induced by some closed convex cone K with nonempty interior.*
(v) *$X = R^n$, f is lower semicontinuous on a neighborhood of x, and $\partial f(x)$ is nonempty and does not include an entire line.*

Proof. (i) has already been noted, and (ii) is immediate from Proposition 2.9.3. Let us turn to (iii).

In view of Proposition 2.9.3, it suffices to prove that $C = \operatorname{epi} f$ admits a hypertangent at $(x, f(x))$. If f is bounded above by $\beta - 1$ on a neighborhood of $x + \lambda v$ for some $\lambda > 0$, it follows that (v, α) satisfies $(x, f(x)) + \lambda(v, \alpha) \in \operatorname{int} C$, where $\alpha = (\beta - f(x))/\lambda$. Thus the following lemma is all we need:

Lemma. If a set C is convex and contains, along with x, a neighborhood of $x + \lambda v$ for some vector v and $\lambda > 0$, then v is hypertangent to C at x.

To see this, suppose that $x + \lambda v + \varepsilon B \subset C$ for $\varepsilon > 0$. Choose $\delta > 0$ such that $(x + \delta B) + \lambda(v + \delta B) \subset C$. Then, for any t in $(0, \lambda)$, $y \in x + \delta B$, $w \in v + \delta B$, one has

$$y + tw = \left(1 - \frac{t}{\lambda}\right)y + \frac{t}{\lambda}(y + \lambda w).$$

If also $y \in C$, then this last expression belongs to C by convexity. That is, we have shown $(x + \delta B) \cap C + t(v + \delta B) \subset C$ for $t \in (0, \lambda)$, which establishes that v is hypertangent to C at x.

Next on the agenda is the proof of (iv). The condition is that $f(x_1) \leqslant f(x_2)$ whenever $x_1 \leqslant_K x_2$; that is, $x_2 - x_1 \in K$. Suppose $-v \in \operatorname{int} K$, so that $-v + \varepsilon B \subset K$ for some $\varepsilon > 0$. Then for all points w such that $-w \in -v + \varepsilon B$, and for all $t \geqslant 0$, one has $-tw \in K$, so that $y + tw \leqslant_K y$ for all y. Consequently,

$$\frac{f(y + tw) - f(y)}{t} \leqslant 0$$

for y near x, w near v, and we derive the result that $f^+(x; v)$ is finite, so that f is directionally Lipschitz at x (with respect to v).

Finally we turn to (v). The statement that $\partial f(x)$ does not include a line is equivalent, in light of the corollary to Theorem 2.9.1, to the condition that the set $D = \{v : f^\circ(x; v) < \infty\}$ is not contained in any subspace of dimension less than n, which is equivalent to saying that D has nonempty interior; suppose $v \in \operatorname{int} D$. Now $f^\circ(x; \cdot)$ is locally bounded in int D, because it is convex and X is finite-dimensional. It follows that for some $\beta \in R, (v, \beta) \in \operatorname{int} \operatorname{epi} f^\circ(x; \cdot) = \operatorname{int} T_{\operatorname{epi} f}(x, f(x))$. But when $X = R^n$, the existence of hypertangents to a closed set at a point is equivalent to the tangent cone at the point having nonempty interior (Corollary 1, Theorem 2.5.8). It now follows from Proposition 2.9.3 that f is directionally Lipschitz at x. (Note that the lower semicontinuity hypothesis of (v) could be weakened to: epi f is locally closed near $(x, f(x))$.) □

We denote by $D_f(x)$ the set of vectors v, if any, with respect to which f is directionally Lipschitz at x.

2.9.5 Theorem

Let f be finite at x, and suppose $D_f(x) \neq \varnothing$. Then

$$D_f(x) = \operatorname{int}\{v : f^\circ(x; v) < \infty\}.$$

Furthermore, $f^\circ(x; \cdot)$ is continuous at each $v \in D_f(x)$ and agrees there with $f^+(x; \cdot)$.

Proof. We shall derive this as a consequence of Theorem 2.4.8 (for $C = \operatorname{epi} f$). Let K denote the set of all hypertangents to C at $(x, f(x))$; K is nonempty by assumption. Then $D_f(x) = \pi_X K$ (projection of K on X) by Proposition 2.9.3, and $K = \operatorname{int} T_C(x, f(x))$ by Theorem 2.4.8. So in order to prove the first assertion of the theorem, it suffices to establish

$$\pi_X \operatorname{int} T_C(x, f(x)) = \operatorname{int}\{v : f^\circ(x; v) < \infty\}, \tag{6}$$

which we now proceed to do.

If v lies in the left-hand side of Eq. (6), then $(v, \beta) \in \operatorname{int} T_C(x, f(x))$ for some β. By Theorem 2.9.1 then, $(v, \beta) \in \operatorname{int} \operatorname{epi} f^\circ(x; \cdot)$. Consequently, $f^\circ(x; w) < \infty$ for all w near v, and v lies in the right-hand side of Eq. (6).

If now v lies in the right-hand side of Eq. (6), then for some $\varepsilon > 0$ and β in R, $f^\circ(x; w) < \beta - 1$ for all w in $v + \varepsilon B$. It follows that (w, α) lies in $\operatorname{epi} f^\circ(x; \cdot) = T_C(x, f(x))$ for all w near v and α near β; that is, $(v, \beta) \in \operatorname{int} T_C(x, f(x))$, so $v \in \pi_X \operatorname{int} T_C(x, f(x))$.

A convex function such as $f^\circ(x; \cdot)$ is always continuous on the interior of the set $\{v : f^\circ(x; v) < \infty\}$, provided it is bounded above in a neighborhood of one point; this is the case precisely because $D_f(x) \neq \varnothing$.

Finally let us examine the expressions

$$f^\circ(x; v) = \inf\{\beta : (v, \beta) \in T_C(x, f(x))\}$$

$$f^+(x; v) = \inf\{\beta : (v, \beta) \in K\},$$

where v belongs to $D_f(x) = \pi_X K$. The first equality is a result of the fact that $T_C(x, f(x))$ is the epigraph of $f^\circ(x; \cdot)$ (and $f^\circ(x; v) < \infty$); the second is the corollary to Proposition 2.9.3. Since, as we have seen, $K = \operatorname{int} T_C(x, f(x))$, it follows that $f^+(x; v) = f^\circ(x; v)$, and the theorem is proved. □

In view of Corollary 1 to Theorem 2.5.8, the proof above can be modified to yield the following:

Corollary

Let X be finite-dimensional, and suppose epi f is locally closed near $(x, f(x))$. Then one has

$$D_f(x) = \operatorname{int}\{v : f^\circ(x; v) < \infty\},$$

and so f is directionally Lipschitz at x precisely when the set $\{v : f^\circ(x; v) < \infty\}$ has nonempty interior.

The Asymptotic Generalized Gradient

A convenient measure of the degree to which a given function f fails to be Lipschitz near a point x (at which f is finite) is provided by the *asymptotic* *generalized gradient* of f at x, denoted $\partial^\infty f(x)$, which is defined to be the set

$$\{\zeta \in X^* : (\zeta, 0) \in N_{\text{epi}\, f}(x, f(x))\}$$

(compare with Definition 2.4.10). Note that $\partial^\infty f(x)$ always contains 0. The first two assertions below follow from the definition and the theory of recession cones (see Rockafellar, 1979b, p. 350); the third is a consequence of Theorem 2.5.6, Corollary 2.

2.9.6 Proposition

$\partial^\infty f(x)$ is a closed convex cone. If $\partial f(x)$ is nonempty, then one has

$$N_{\text{epi}\, f}(x, f(x)) = \bigcup_{\lambda > 0} \lambda[\partial f(x), -1] \cup [\partial^\infty f(x), 0],$$

and in this case $\partial^\infty f(x)$ reduces to $\{0\}$ if $\partial f(x)$ is bounded. If X is finite-dimensional, and if epi f is locally closed near $(x, f(x))$ (e.g., if f is lower semicontinuous) then $\partial f(x) \cup (\partial^\infty f(x) \setminus \{0\}) \neq \varnothing$.

When f is Lipschitz near x, then of course $\partial f(x)$ is nonempty and bounded (and so $\partial^\infty f(x) = \{0\}$). The converse holds for a large class of functions:

2.9.7 Proposition

Suppose that X is finite-dimensional, that f is finite at x, and that epi f is locally closed near $(x, f(x))$. Then the following are equivalent:

(a) *$\partial f(x)$ is nonempty and bounded.*
(b) *f is Lipschitz near x.*
(c) $\partial^\infty f(x) = \{0\}$.

Proof. We shall show that (a) implies (b) implies (c) implies (a). If (a) holds, then f is directionally Lipschitz at x by Theorem 2.9.4. It then follows from the corollary to Theorem 2.9.1 together with Theorem 2.9.5 that $f^+(x; 0)$ is finite. This is readily seen to imply (b). We have already observed that (b) implies (c), and the fact that (c) implies (a) is immediate from Proposition 2.9.6. □

Sum of Two Functions

We are now in a position to prove the extended formula for the generalized gradient of a sum of two functions. Recall that regularity in the extended setting was defined in Definition 2.4.10.

2.9.8 Theorem (Rockafellar)

Let f_1 and f_2 be finite at x, and let f_2 be directionally Lipschitz at x. Suppose that

$$\{v : f_1^\circ(x; v) < \infty\} \cap \text{int}\{v : f_2^\circ(x; v) < \infty\} \neq \varnothing. \tag{7}$$

Then

$$\partial(f_1 + f_2)(x) \subset \partial f_1(x) + \partial f_2(x), \tag{8}$$

where the set on the right is weak-closed. If in addition f_1 and f_2 are regular at x, then equality holds, and if $\partial f_1(x)$ and $\partial f_2(x)$ are nonempty, then $f_1 + f_2$ is regular at x.*

Proof. As in the proof of Proposition 2.3.3 in the locally Lipschitz case, we begin by proving:

Lemma.

$$(f_1 + f_2)^\circ(x; v) \leqslant f_1^\circ(x; v) + f_2^\circ(x; v).$$

Let $f_0 = f_1 + f_2$, $l_i(v) = f_i^\circ(x; v)$, for $i = 0, 1, 2$. We wish to prove

$$l_0(v) \leqslant l_1(v) + l_2(v), \tag{9}$$

where the convention $\infty - \infty = \infty$ is in force, and we start with the case in which v lies in the left-hand side of (7).

Let $\beta > l_2(v) = f_2^\circ(x; v)$. By Theorem 2.9.5, we have $f_2^\circ(x; v) = f_2^+(x; v) < \beta$. Hence, for some $\delta > 0$, one has

$$\frac{f_2(y + tw) - \alpha_2}{t} < \beta \tag{10}$$

whenever $t \in (0, \delta)$, $y \in x + \delta B$, $w \in v + \delta B$, $\alpha_2 \geqslant f_2(y)$, $|\alpha_2 - f(y)| < \delta$. Recall that by definition

$$l_0(y) = \lim_{\substack{\varepsilon \downarrow 0}} \limsup_{\substack{(y,\alpha)\downarrow_f x \\ t \downarrow 0}} \inf_{w \in v + \varepsilon B} \frac{f_0(y + tw) - \alpha}{t}.$$

We can write this difference quotient as

$$\frac{f_1(y + tw) - \alpha_1}{t} + \frac{f_2(y + tw) - \alpha_2}{t}, \tag{11}$$

where $\alpha_1 + \alpha_2 = \alpha$, and the lim sup can be taken equivalently as $t \downarrow 0$, $y \to x$, $\alpha_i \to f_i(x)$ with $\alpha_i \geqslant f_i(y)$. Invoking (10) in (11), we obtain

$$l_0(y) \leqslant \lim_{\substack{\varepsilon \downarrow 0}} \limsup_{\substack{(y,\alpha)\downarrow_f x \\ t \downarrow 0}} \inf_{w \in v + \varepsilon B} \left\{ \frac{f_1(y + tw) - \alpha_1}{t} + \beta \right\}$$

$$= f_1^\circ(x; v) + \beta = l_1(v) + \beta.$$

Since this is true for all $\beta > l_2(v)$, we derive (9), which we still need to establish for general v. If either $l_1(v)$ or $l_2(v) = +\infty$, then (9) is trivial, so suppose that v belongs to $D_1 \cap D_2$, where D_i is the convex set $\{w : l_i(w) < \infty\}$. By (7) there is a point $\tilde{v}$ in $D_1 \cap \operatorname{int} D_2$; by convexity, we have $v_\varepsilon := (1-\varepsilon)v + \varepsilon\tilde{v} \in D_1 \cap \operatorname{int} D_2$ for ε in $(0,1)$. By the case already treated, then, we know

$$l_0(v_\varepsilon) \leqslant l_1(v_\varepsilon) + l_2(v_\varepsilon).$$

Because the functions l_i are convex and lower semicontinuous, we have $\lim_{\varepsilon \to 0} l_i(v_\varepsilon) = l_i(v)$, so (9) results, and hence the lemma is proved.

Let us now prove (8). If either $l_1(0)$ or $l_2(0)$ is $-\infty$, then, as the proof of the lemma actually showed, $l_0(0) = -\infty$ also. Then (corollary, Theorem 2.9.1) both sides of (8) are empty and there is nothing to prove. Assume therefore that $l_1(0) = l_2(0) = 0$ (the only other possibility, since l_i is sublinear). Then one has

$$\begin{aligned} \partial f_0(x) &= \{z : l_0(v) \geqslant \langle z, v\rangle \text{ for all } v\} \\ &\subset \{z : (l_1 + l_2)(v) \geqslant \langle z, v\rangle \text{ for all } v\} \\ &= \partial(l_1 + l_2)(0), \end{aligned} \tag{12}$$

where the final equality is valid because $l_1 + l_2$ is a convex function which is 0 at 0. The assumption (7) is precisely what is needed to make the following formula from convex analysis valid:

$$\partial(l_1 + l_2)(0) = \partial l_1(0) + \partial l_2(0) \qquad (= \partial f_1(x) + \partial f_2(x)).$$

This together with (12) yields (8). As the subdifferential of $l_1 + l_2$ at 0, $\partial f_1(x) + \partial f_2(x)$ is automatically weak*-closed.

The final component to be proved concerns the equality in expression (8). If either $l_1(0)$ or $l_2(0)$ is $-\infty$, then, as we have seen, $l_0(0)$ is also, and both sides of (8) are empty (and hence, equal). Assuming therefore that this is not the case, and that f_1, f_2 are regular at x, we have (as an easy consequence of regularity) that, for any v,

$$l_i(v) = \liminf_{\substack{w \to v \\ t \downarrow 0}} \frac{f_i(x+tw) - f_i(x)}{t} \qquad (i = 1, 2).$$

Taking lower limits in the expression

$$\frac{f_1(x+tw) - f_1(x)}{t} + \frac{f_2(x+tw) - f_2(x)}{t} = \frac{f_0(x+tw) - f_0(x)}{t},$$

we derive (since the first two terms tend separately to $l_1(v), l_2(v) > -\infty$) the inequality

$$(13) \qquad l_1(v) + l_2(v) \leqslant \liminf_{\substack{w \to v \\ t \downarrow 0}} \frac{[f_0(x + tw) - f_0(x)]}{t} \leqslant l_0(v).$$

In view of the lemma, we have therefore

$$l_1(v) + l_2(v) = l_0(v) \quad \text{for all } v.$$

This permits the reversal of the inclusion in (12), so (8) holds with equality. Further, $l_0(\cdot)$ is now seen to be the support function of $\partial f_0(x)$, so $l_0(v) = f_0^\circ(x; v)$ and (13) implies that f_0 is regular at x. □

Corollary 1

Suppose that f_1 is finite at x and f_2 is Lipschitz near x. Then one has $\partial(f_1 + f_2)(x) \subset \partial f_1(x) + \partial f_2(x)$, and there is equality if f_1 and f_2 are also regular at x.

Proof. The theorem may be applied directly, since 0 belongs to the set in (7). □

Corollary 2

Let C_1 and C_2 be subsets of X and let $x \in C_1 \cap C_2$. Suppose that

$$T_{C_1}(x) \cap \operatorname{int} T_{C_2}(x) \neq \varnothing,$$

and that C_2 admits at least one hypertangent vector v at x. Then one has

$$(14) \qquad T_{C_1 \cap C_2}(x) \supset T_{C_1}(x) \cap T_{C_2}(x)$$

$$(15) \qquad N_{C_1 \cap C_2}(x) \subset N_{C_1}(x) + N_{C_2}(x),$$

where the set on the right in expression (15) *is weak*-closed. Equality holds in* (14) *and* (15) *if C_1 and C_2 are regular at x, in which case $C_1 \cap C_2$ is also regular at x.*

Proof. Apply the theorem to $f_1 = \psi_{C_1}$, $f_2 = \psi_{C_2}$. The existence of v as above is equivalent to f_2 being directionally Lipschitz (by Proposition 2.9.3). □

Corollary 3

Suppose $X = R^n$. Then the hypothesis in Theorem 2.9.8 that f_2 be directionally Lipschitz can be replaced by the assumption that f_2 is lower semicontinuous in a *neighborhood of x. Likewise, the existence of v in Corollary 2 can be replaced by the assumption that C_2 is locally closed near x.*

Proof. Invoke the corollary to Theorem 2.9.5 in the first case, Corollary 1 to Theorem 2.5.8 in the second. □

Composition with a Differentiable Map

The following is Rockafellar's extended analogue of Theorem 2.3.10:

2.9.9 Theorem

Let $f = g \circ F$, where F is a strictly differentiable mapping from X to another Banach space Y and g is an extended-valued function on Y. Suppose that g is finite and directionally Lipschitz at $F(x)$ with

$$(\text{range } D_s F(x)) \cap \operatorname{int}\{v : g^\circ(F(x); v) < \infty\} \neq \varnothing .$$

Then one has

$$\partial f(x) \subset D_s F(x)^* \circ \partial g(F(x)), \tag{16}$$

and equality holds if g is regular at x.

Proof. Let $y = F(x)$, and define h on $X \times Y$ by

$$h(x', y') = \begin{cases} f(x') & \text{if } y' = F(x') \\ +\infty & \text{otherwise.} \end{cases}$$

Note that $h = f_1 + f_2$, where f_1 is the indicator of the graph of F and where $f_2(x', y') = g(y')$. We let A stand for $D_s F(x)$.

Lemma.

$$h^\circ(x, y; v, w) = \begin{cases} f^\circ(x; v) & \text{if } w = A(v) \\ +\infty & \text{otherwise.} \end{cases}$$

To see this, recall that h° above is given by

$$\lim_{\varepsilon \to 0} \limsup_{\substack{(x', y', \alpha) \downarrow_h (x, y) \\ t \downarrow 0}} \inf_{(v', w') \in (v, w) + \varepsilon B} \frac{h(x' + tv', y' + tw') - \alpha}{t},$$

where the condition $\alpha \geqslant h(x', y')$ implicit above is equivalent to $\alpha \geqslant f(x')$ and $y' = F(x')$, and where

$$h(x' + tv', y' + tw') = \begin{cases} f(x' + tv') & \text{if } w' = \dfrac{F(x' + tv') - F(x')}{t} \\ +\infty & \text{otherwise.} \end{cases}$$

Since F is strictly differentiable at x, we have

$$\limsup_{\substack{x' \to x,\, v' \to v \\ t \downarrow 0}} \frac{F(x' + tv') - F(x')}{t} = A(v),$$

as an easy consequence of Proposition 2.2.1. It follows that the expression for h° above is $+\infty$ if $w \neq A(v)$, while otherwise it is

$$\lim_{\varepsilon \to 0} \limsup_{\substack{(x', \alpha) \downarrow_f x \\ t \downarrow 0}} \inf_{v' \in v + \varepsilon B} \frac{f(x' + tv') - \alpha}{t} = f^\circ(x; v),$$

as stated in the lemma. We now calculate as follows:

$$\begin{aligned}
\partial h(x, y) &= \{(\zeta, \phi) : h^\circ(x, y; v, w) \geqslant \langle \zeta, v \rangle + \langle \phi, w \rangle \text{ for all } v, w\} \\
&= \{(\zeta, \phi) : f^\circ(x, v) \geqslant \langle \zeta, v \rangle + \langle \phi, A(v) \rangle \text{ for all } v\} \\
&= \{(\zeta, \phi) : f^\circ(x; v) \geqslant \langle \zeta + A^*(\phi), v \rangle \text{ for all } v\} \\
&= \{(\zeta, \phi) : \zeta + A^*(\phi) \in \partial f(x)\},
\end{aligned}$$

and consequently,

$$\partial f(x) = \{\zeta : (\zeta, 0) \in \partial h(x, y)\}. \tag{17}$$

The next step is to apply Theorem 2.9.8 to the representation $h = f_1 + f_2$ noted above. The strict differentiability implies for the set $G = \text{graph } F$ that

$$T_G(x, y) = K_G(x, y) = \text{graph } A.$$

Thus, f_1 is regular at (x, y) and

$$f_1^\circ(x, y; v, w) = \begin{cases} 0 & \text{if } w = A(v) \\ +\infty & \text{otherwise,} \end{cases}$$

$$\begin{aligned}
\partial f_1(x, y) &= N_G(x, y) = \text{polar of graph } A \qquad (18) \\
&= \{(\zeta, \phi) : \zeta = -A^*(\phi)\}.
\end{aligned}$$

Clearly f_2 inherits the directionally Lipschitz property from g, and one has

$$\begin{aligned}
f_2^\circ(x, y; v, w) &= g^\circ(y; w), \\
\partial f_2(x, y) &= \{(0, \phi) : \phi \in \partial g(y)\}. \qquad (19)
\end{aligned}$$

In particular, the set

$$\{(v,w) : f_1^\circ(x,y;v,w) < \infty\} \cap \operatorname{int}\{(v,w) : f_2^\circ(x,y;v,w) < \infty\}$$
$$= \{(v, A(v)) : A(v) \in \operatorname{int}\{w : g^\circ(y;w) < \infty\}\}$$

is nonempty by assumption. Therefore the hypotheses of Theorem 2.9.8 are satisfied, and we have

$$\partial h(x,y) \subset \partial f_1(x,y) + \partial f_2(x,y).$$

Combining this with (17) (18) (19) gives (16). The remaining assertions about equality are also direct translations from Theorem 2.9.8. □

Corollary 1

Let $x \in C := F^{-1}(\Omega)$, where $\Omega \subset Y$ and where F is strictly differentiable at x. Suppose that Ω admits a hypertangent at $F(x)$, and that

$$(\text{range } D_s F(x)) \cap \operatorname{int} T_\Omega(F(x)) \neq \varnothing .$$

Then one has

$$T_C(x) \supset D_s F(x)^{-1}[T_\Omega(F(x))]$$
$$N_C(x) \subset D_s F(x)^* \circ N_\Omega(F(x)),$$

and equality holds if Ω is also regular at $F(x)$.

Proof. Apply Theorem 2.9.9 to $g = \psi_\Omega$. □

Corollary 2

Suppose $Y = R^n$. Then the hypothesis in Theorem 2.9.9 that g is directionally Lipschitz can be replaced with the condition that g is lower semicontinuous on a neighborhood of $F(x)$. The existence of a hypertangent in Corollary 1 can be replaced with the condition that Ω is closed in a neighborhood of $F(x)$.

Proof. Invoke the corollary to Theorem 2.9.5 in the first case, Corollary 1 to Theorem 2.5.8 in the second. □

Sets Defined by Inequalities

2.9.10 Theorem

Let $C = \{x' : f(x') \leqslant 0\}$, and let x be a point satisfying $f(x) = 0$. Suppose that f is directionally Lipschitz at x with $0 \notin \partial f(x) \neq \varnothing$, and let $D = \{v : f^\circ(x;v) <$

$\infty\}$. *Then C admits a hypertangent at x and one has*

$$T_C(x) \supset \{v : f^\circ(x; v) \leqslant 0\} \tag{20}$$

$$\operatorname{int} T_C(x) \supset \{v \in \operatorname{int} D : f^\circ(x; v) < 0\} \neq \emptyset \tag{21}$$

$$N_C(x) \subset \bigcup_{\lambda > 0} \lambda\, \partial f(x) \cup \partial^\infty f(x), \tag{22}$$

where the set on the right of (22) is weak-closed. If in addition f is regular at x, then equality holds in each of the above and C is regular at x.*

Proof. The result will be derived by invoking Corollary 2 of Theorem 2.9.8, for the sets $C_1 = \{(z, \mu) : \mu = 0\}$ and $C_2 = \operatorname{epi} f$ at the point $(x, 0)$ (note that $C \times \{0\} = C_1 \cap C_2$). One has

$$T_{C_1}(x,0) = \{(z, \mu) : \mu = 0\}, \qquad T_{C_2}(x,0) = \operatorname{epi} f^\circ(x; \cdot), \tag{23}$$

and by polarity

$$N_{C_1}(x,0) = \{(\zeta, \mu) : \zeta = 0\} \tag{24}$$

$$N_{C_2}(x,0) = \text{the cone in the statement of Proposition 2.9.6.}$$

From Theorem 2.9.5, the function $f^\circ(x; \cdot)$ is continuous on int D, so one also has

$$\operatorname{int} \operatorname{epi} f^\circ(x; \cdot) = \{(v, \beta) : v \in \operatorname{int} D, f^\circ(x; v) < \beta\}, \tag{25}$$

and this set is nonempty. Therefore by Eq. (23),

$$T_{C_1}(x,0) \cap \operatorname{int} T_{C_2}(x,0) = \{(v,0) : v \in \operatorname{int} D, f^\circ(x; v) < 0\}.$$

If this latter set were empty, then the set (25) would be contained in the half-space $\{(v, \beta) : \beta \geqslant 0\}$, and the same would then be true of its closure which includes epi $f^\circ(x; \cdot)$. This would imply $f^\circ(x; v) \geqslant 0$ for all v in contradiction with the hypothesis $0 \notin \partial f(x)$. The hypotheses of Corollary 2, Theorem 2.9.8, are therefore satisfied. (C_2 admits a hypertangent at $(x, f(x))$ since f is directionally Lipschitz at x.) The conclusions of the theorem follow directly from those of the corollary together with Eqs. (23), (24), and (25). □