Convex Analysis and Nonlinear Optimization
Second Edition

New Material

Fixed Points
8.1 The Brouwer Fixed Point Theorem ................. 179
8.2 Selection and the Kakutani–Fan Fixed Point Theorem ... 190
8.3 Variational Inequalities ................................... 200

8.3 now includes Fitzpatrick Function

More Nonsmooth Structure ................................. 213
9.1 Rademacher’s Theorem ................................. 213
9.2 Proximal Normals and Chebyshev Sets ............... 218
9.3 Amenable Sets and Prox-Regularity ................. 228
9.4 Partly Smooth Sets ................................. 233
"He was very big in Vienna."
Chapter 9

More Nonsmooth Structure

9.1 Rademacher’s Theorem

We mentioned Rademacher’s fundamental theorem on the differentiability of Lipschitz functions in the context of the Intrinsic Clarke subdifferential formula (Theorem 6.2.5):

$$\partial \circ f(x) = \text{conv} \{\lim_r \nabla f(x^r) \mid x^r \to x, \ x^r \notin Q\}, \quad (9.1.1)$$

valid whenever the function $f : E \to \mathbb{R}$ is locally Lipschitz around the point $x \in E$ and the set $Q \subset E$ has measure zero. We prove Rademacher’s theorem in this section, taking a slight diversion into some basic measure theory.

**Theorem 9.1.2 (Rademacher)** Any locally Lipschitz map between Euclidean spaces is Fréchet differentiable almost everywhere.

**Proof.** Without loss of generality (Exercise 1), we can consider a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$. In fact, we may as well further suppose that $f$ has Lipschitz constant $L$ throughout $\mathbb{R}^n$, by Exercise 2 in Section 7.1.

Fix a direction $h$ in $\mathbb{R}^n$. For any $t \neq 0$, the function $g_t$ defined on $\mathbb{R}^n$ by

$$g_t(x) = \frac{f(x + th) - f(x)}{t}$$

is continuous, and takes values in the interval $I = L||h||[-1, 1]$, by the Lipschitz property. Hence, for $k = 1, 2, \ldots$, the function $p_k : \mathbb{R}^n \to I$
defined by
\[ p_k(x) = \sup_{0 < |t| < 1/k} g_t(x) \]
is lower semicontinuous and therefore Borel measurable. Consequently, the upper Dini derivative \( D^+ h f : \mathbb{R}^n \to I \) defined by
\[ D^+ h f(x) = \limsup_{t \to 0} g_t(x) = \inf_{k \in \mathbb{N}} p_k(x) \]
is measurable, being the infimum of a sequence of measurable functions. Similarly, the lower Dini derivative \( D^- h f : \mathbb{R}^n \to I \) defined by
\[ D^- h f(x) = \liminf_{t \to 0} g_t(x) \]
is also measurable.

The subset of \( \mathbb{R}^n \) where \( f \) is not differentiable along the direction \( h \), namely
\[ A_h = \{ x \in \mathbb{R}^n \mid D^- h f(x) < D^+ h f(x) \} \]
is therefore also measurable. Given any point \( x \in \mathbb{R}^n \), the function \( t \mapsto f(x + th) \) is absolutely continuous (being Lipschitz), so the fundamental theorem of calculus implies this function is differentiable (or equivalently, \( x + th \notin A_h \)) almost everywhere on \( \mathbb{R} \).

Consider the nonnegative measurable function \( \phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \) defined by \( \phi(x, t) = \delta_{A_h}(x+th) \). By our observation above, for any fixed \( x \in \mathbb{R}^n \) we know \( \int_{\mathbb{R}} \phi(x, t) dt = 0 \). Denoting Lebesgue measure on \( \mathbb{R}^n \) by \( \mu \), Fubini’s theorem shows
\[ 0 = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}} \phi(x, t) dt \right) d\mu = \int_{\mathbb{R}} \left( \int_{\mathbb{R}^n} \phi(x, t) d\mu \right) dt = \int_{\mathbb{R}} \mu(A_h) dt \]
so the set \( A_h \) has measure zero. Consequently, we can define a measurable function \( D_h f : \mathbb{R}^n \to \mathbb{R} \) having the property \( D^- h f = D^+ h f = D^- h f \) almost everywhere.

Denote the standard basis vectors in \( \mathbb{R}^n \) by \( e_1, e_2, \ldots, e_n \). The function \( G : \mathbb{R}^n \to \mathbb{R}^n \) with components defined almost everywhere by
\[ G_i = D_{e_i} f = \frac{\partial f}{\partial x_i} \quad (9.1.3) \]
for each \( i = 1, 2, \ldots, n \) is the only possible candidate for the derivative of \( f \). Indeed, if \( f \) (or \(-f\)) is regular at \( x \), then it is easy to check that \( G(x) \) is the Fréchet derivative of \( f \) at \( x \) (Exercise 2). The general case needs a little more work.

Consider any continuously differentiable function \( \psi : \mathbb{R}^n \to \mathbb{R} \) that is zero except on a bounded set. For our fixed direction \( h \), if \( t \neq 0 \) we have
\[ \int_{\mathbb{R}^n} g_t(x) \psi(x) d\mu = \int_{\mathbb{R}^n} f(x) \frac{\psi(x - th) - \psi(x)}{t} d\mu. \]
defined by
\[ p_k(x) = \sup_{0 < |t| < 1/k} g_t(x) \]
is lower semicontinuous and therefore Borel measurable. Consequently, the upper Dini derivative \( D^+_h f : \mathbb{R}^n \to I \) defined by
\[ D^+_h f(x) = \limsup_{t \to 0} g_t(x) = \inf_{k \in \mathbb{N}} p_k(x) \]
is measurable, being the infimum of a sequence of measurable functions. Similarly, the lower Dini derivative \( D^-_h f : \mathbb{R}^n \to I \) defined by
\[ D^-_h f(x) = \liminf_{t \to 0} g_t(x) \]
is also measurable.

The subset of \( \mathbb{R}^n \) where \( f \) is not differentiable along the direction \( h \), namely
\[ A_h = \{ x \in \mathbb{R}^n \mid D^-_h f(x) < D^+_h f(x) \}, \]
is therefore also measurable. Given any point \( x \in \mathbb{R}^n \), the function \( t \mapsto f(x + th) \) is absolutely continuous (being Lipschitz), so the fundamental theorem of calculus implies this function is differentiable (or equivalently, \( x + th \notin A_h \)) almost everywhere on \( \mathbb{R} \).

Consider the nonnegative measurable function \( \phi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \) defined by \( \phi(x, t) = \delta_{A_h}(x + th) \). By our observation above, for any fixed \( x \in \mathbb{R}^n \) we know \( \int_{\mathbb{R}} \phi(x, t) \, dt = 0 \). Denoting Lebesgue measure on \( \mathbb{R}^n \) by \( \mu \), Fubini’s theorem shows
\[ 0 = \int_{\mathbb{R}^n} \left( \int_{\mathbb{R}} \phi(x, t) \, dt \right) \, d\mu = \int_{\mathbb{R}} \left( \int_{\mathbb{R}^n} \phi(x, t) \, d\mu \right) \, dt = \int_{\mathbb{R}} \mu(A_h) \, dt \]
so the set \( A_h \) has measure zero. Consequently, we can define a measurable function \( D_h f : \mathbb{R}^n \to \mathbb{R} \) having the property \( D_h f = D^+_h f = D^-_h f \) almost everywhere.

Denote the standard basis vectors in \( \mathbb{R}^n \) by \( e_1, e_2, \ldots, e_n \). The function \( G : \mathbb{R}^n \to \mathbb{R}^n \) with components defined almost everywhere by
\[ G_i = D_{e_i} f = \frac{\partial f}{\partial x_i} \tag{9.1.3} \]
for each \( i = 1, 2, \ldots, n \) is the only possible candidate for the derivative of \( f \). Indeed, if \( f \) (or \( -f \)) is regular at \( x \), then it is easy to check that \( G(x) \) is the Fréchet derivative of \( f \) at \( x \) (Exercise 2). The general case needs a little more work.

Consider any continuously differentiable function \( \psi : \mathbb{R}^n \to \mathbb{R} \) that is zero except on a bounded set. For our fixed direction \( h \), if \( t \neq 0 \) we have
\[ \int_{\mathbb{R}^n} g_t(x) \psi(x) \, d\mu = \int_{\mathbb{R}^n} f(x) \frac{\psi(x - th) - \psi(x)}{t} \, d\mu. \]
As \( t \to 0 \), the bounded convergence theorem applies, since both \( f \) and \( \psi \) are Lipschitz, so

\[
\int_{\mathbb{R}^n} D_h f(x) \psi(x) \, d\mu = -\int_{\mathbb{R}^n} f(x) \langle \nabla \psi(x), h \rangle \, d\mu.
\]

Setting \( h = e_i \) in the above equation, multiplying by \( h_i \), and adding over \( i = 1, 2, \ldots, n \), yields

\[
\int_{\mathbb{R}^n} \langle h, G(x) \rangle \psi(x) \, d\mu = -\int_{\mathbb{R}^n} f(x) \langle \nabla \psi(x), h \rangle \, d\mu = \int_{\mathbb{R}^n} D_h f(x) \psi(x) \, d\mu.
\]

Since \( \psi \) was arbitrary, we deduce \( D_h f = \langle h, G \rangle \) almost everywhere.

Now extend the basis \( e_1, e_2, \ldots, e_n \) to a dense sequence of unit vectors \( \{h_k\} \) in the unit sphere \( S_{n-1} \subset \mathbb{R}^n \). Define the set \( A \subset \mathbb{R}^n \) to consist of those points where each function \( D_{h_k} f \) is defined and equals \( \langle h_k, G \rangle \). Our argument above shows \( A^C \) has measure zero. We aim to show, at each point \( x \in A \), that \( f \) has Fréchet derivative \( G(x) \).

Fix any \( \epsilon > 0 \). For any \( t \neq 0 \), define a function \( r_t : \mathbb{R}^n \to \mathbb{R} \) by

\[
r_t(h) = \frac{f(x + th) - f(x) - \langle G(x), h \rangle}{t}.
\]

It is easy to check that \( r_t \) has Lipschitz constant \( 2L \). Furthermore, for each \( k = 1, 2, \ldots, M \), there exists \( \delta_k > 0 \) such that

\[
|r_t(h_k)| < \frac{\epsilon}{2} \quad \text{whenever } 0 < |t| < \delta_k.
\]

Since the sphere \( S_{n-1} \) is compact, there is an integer \( M \) such that

\[
S_{n-1} \subset \bigcup_{k=1}^M (h_k + \frac{\epsilon}{4L} B).
\]

If we define \( \delta = \min\{\delta_1, \delta_2, \ldots, \delta_M\} > 0 \), we then have

\[
|r_t(h_k)| < \frac{\epsilon}{2} \quad \text{whenever } 0 < |t| < \delta, \; k = 1, 2, \ldots, M.
\]

Finally, consider any unit vector \( h \). For some positive integer \( k \leq M \) we know \( \|h - h_k\| \leq \epsilon/4L \), so whenever \( 0 < |t| < \delta \) we have

\[
|r_t(h)| \leq |r_t(h) - r_t(h_k)| + |r_t(h_k)| \leq 2L \frac{\epsilon}{4L} + \frac{\epsilon}{2} = \epsilon.
\]

Hence \( G(x) \) is the Fréchet derivative of \( f \) at \( x \), as we claimed. 

An analogous argument using Fubini’s theorem now proves the subdifferential formula (9.1.1)—see Exercise 3.
It is easy to check that \( r_t \) has Lipschitz constant \( 2L \). Furthermore, for each \( k = 1, 2, \ldots \), there exists \( \delta_k > 0 \) such that

\[
| r_t(h_k) | < \frac{\varepsilon}{2} \quad \text{whenever } 0 < |t| < \delta_k.
\]

Since the sphere \( S_{n-1} \) is compact, there is an integer \( M \) such that

\[
S_{n-1} \subset \bigcup_{k=1}^{M} \left( h_k + \frac{\varepsilon}{4L} B \right).
\]

If we define \( \delta = \min\{\delta_1, \delta_2, \ldots, \delta_M\} > 0 \), we then have

\[
| r_t(h_k) | < \frac{\varepsilon}{2} \quad \text{whenever } 0 < |t| < \delta, \; k = 1, 2 \ldots, M.
\]

Finally, consider any unit vector \( h \). For some positive integer \( k \leq M \) we know \( \| h - h_k \| \leq \varepsilon/4L \), so whenever \( 0 < |t| < \delta \) we have

\[
| r_t(h) | \leq | r_t(h) - r_t(h_k) | + | r_t(h_k) | \leq 2L \frac{\varepsilon}{4L} + \frac{\varepsilon}{2} = \varepsilon.
\]

Hence \( G(x) \) is the Fréchet derivative of \( f \) at \( x \), as we claimed.
Exercises and Commentary

A basic reference for the measure theory and the version of the fundamental theorem of calculus we use in this section is [170]. Rademacher’s theorem is also proved in [71]. Various implications of the insensitivity of Clarke’s formula (9.1.1) to sets of measures zero are explored in [18]. In the same light, the generalized Jacobian of Exercise 4 is investigated in [72].

1. Assuming Rademacher’s theorem with range \( \mathbb{R} \), prove the general version.

2. \(^\ast\) (Rademacher’s theorem for regular functions) Suppose the function \( f : \mathbb{R}^n \to \mathbb{R} \) is locally Lipschitz around the point \( x \in \mathbb{R}^n \). Suppose the vector \( G(x) \) is well-defined by equation (9.1.3). By observing

\[
0 = f^-(x; e_i) + f^-(x; -e_i) = f^0(x; e_i) + f^0(x; -e_i)
\]

and using the sublinearity of \( f^0(x; \cdot) \), deduce \( G(x) \) is the Fréchet derivative of \( f \) at \( x \).

3. \(^\ast\ast\) (Intrinsic Clarke subdifferential formula) Derive formula (9.1.1) as follows.

(a) Using Rademacher’s theorem (9.1.2), show we can assume that the function \( f \) is differentiable everywhere outside the set \( Q \).

(b) Recall the one-sided inclusion following from the fact that the Clarke subdifferential is a closed multifunction (Exercise 12 in Section 6.2)

(c) For any vector \( v \in E \) and any point \( z \in E \), use Fubini’s theorem to show that the set \( \{ t \in \mathbb{R} \mid z + tv \in Q \} \) has measure zero, and deduce

\[
f(z + tv) - f(z) = \int_0^t \langle \nabla f(z + sv), v \rangle \, ds.
\]

(d) If formula (9.1.1) fails, show there exists \( v \in E \) such that

\[
f^0(x; v) > \limsup_{w \to x, w \notin Q} \langle \nabla f(w), v \rangle.
\]

Use part (c) to deduce a contradiction.

4. \(^\ast\ast\) (Generalized Jacobian) Consider a locally Lipschitz map between Euclidean spaces \( h : E \to Y \) and a set \( Q \subset E \) of measure zero
3. **(Intrinsic Clarke subdifferential formula)** Derive formula (9.1.1) as follows.

(a) Using Rademacher’s theorem (9.1.2), show we can assume that the function $f$ is differentiable everywhere outside the set $Q$.

(b) Recall the one-sided inclusion following from the fact that the Clarke subdifferential is a closed multifunction (Exercise 12 in Section 6.2)

(c) For any vector $v \in E$ and any point $z \in E$, use Fubini’s theorem to show that the set $\{t \in \mathbb{R} \mid z + tv \in Q\}$ has measure zero, and deduce

$$f(z + tv) - f(z) = \int_0^t \langle \nabla f(z + sv), v \rangle \, ds.$$ 

(d) If formula (9.1.1) fails, show there exists $v \in E$ such that

$$f^\circ(x; v) > \limsup_{w \to x, \, w \not\in Q} \langle \nabla f(w), v \rangle.$$ 

Use part (c) to deduce a contradiction.

4. **(Generalized Jacobian)** Consider a locally Lipschitz map between Euclidean spaces $h : E \to Y$ and a set $Q \subset E$ of measure zero
outside of which \( h \) is everywhere Gâteaux differentiable. By analogy with formula (9.1.1) for the Clarke subdifferential, we call

\[
\partial_Q h(x) = \text{conv} \{ \lim_{r \to 0} \nabla h(x^r) \mid x^r \to x, \ x^r \notin Q \},
\]

the Clarke generalized Jacobian of \( h \) at the point \( x \in E \).

(a) Prove that the set \( J_h(x) = \partial Q h(x) \) is independent of the choice of \( Q \).

(b) (Mean value theorem) For any points \( a, b \in E \), prove

\[
h(a) - h(b) \subset \text{conv} J_h[a, b](a - b).
\]

(c) (Chain rule) If the function \( g : Y \to R \) is locally Lipschitz, prove the formula

\[
\partial_h(g \circ h)(x) \subset J_h(x)^* \partial g(h(x)).
\]

(d) Propose a definition for the generalized Hessian of a continuously differentiable function \( f : E \to R \).
"Just a darn minute! — Yesterday you said that $X$ equals two!"
9.2 Proximal Normals and Chebyshev Sets

We introduced the Clarke normal cone in Section 6.3 (Tangent Cones), via the Clarke subdifferential. An appealing alternative approach begins with a more geometric notion of a normal vector. We call a vector \( y \in \mathbf{E} \) a **proximal normal** to a set \( S \subset \mathbf{E} \) at a point \( x \in S \) if, for some \( t > 0 \), the nearest point to \( x + ty \) in \( S \) is \( x \). The set of all such vectors is called the **proximal normal cone**, which we denote \( N_p^S(x) \).

The proximal normal cone, which may not be convex, is contained in the Clarke normal cone (Exercise 3). The containment may be strict, but we can reconstruct the Clarke normal cone from proximal normals using the following result.

**Theorem 9.2.1 (Proximal normal formula)** For any closed set \( S \subset \mathbf{E} \) and any point \( x \in S \), we have

\[
N_S(x) = \text{conv}\left\{ \lim_r y_r \mid y_r \in N_p^S(x_r), \ x_r \in S, \ x_r \to x \right\}.
\]

One route to this result uses Rademacher’s theorem (Exercise 7). In this section we take a more direct approach.

The Clarke normal cone to a set \( S \subset \mathbf{E} \) at a point \( x \in S \) is

\[
N_S(x) = \text{cl} \left( R^+ \partial d_S(x) \right),
\]

by Theorem 6.3.8, where

\[
ds(x) = \inf_{\bar{x} \in S} \| \bar{x} - x \|
\]

is the distance function. Notice the following elementary but important result that we use repeatedly in this section (Exercise 4(a) in Section 7.3).

**Proposition 9.2.2 (Projections)** If \( \bar{x} \) is a nearest point in the set \( S \subset \mathbf{E} \) to the point \( x \in \mathbf{E} \), then \( \bar{x} \) is the unique nearest point in \( S \) to each point on the half-open line segment \( [\bar{x}, x) \).

To derive the proximal normal formula from the subdifferential formula (9.1.1), we can make use of some striking differentiability properties of distance functions, summarized in the next result.

**Theorem 9.2.3 (Differentiability of distance functions)** Consider a nonempty closed set \( S \subset \mathbf{E} \) and a point \( x \notin S \). Then the following properties are equivalent:

(i) the Dini subdifferential \( \partial d_S(x) \) is nonempty;

(ii) \( x \) has a unique nearest point \( \bar{x} \) in \( S \).
(iii) the distance function $d_S$ is Fréchet differentiable at $x$.

In this case,

$$\nabla d_S(x) = \frac{x - \bar{x}}{\|x - \bar{x}\|} \in N_S(x) \subset N_S(\bar{x}).$$

The proof is outlined in Exercises 4 and 6.

For our alternate proof of the proximal normal formula without recourse to Rademacher’s theorem, we return to an idea we introduced in Section 8.2. A cusco is a USC multifunction with nonempty compact convex images. In particular, the Clarke subdifferential of a locally Lipschitz function on an open set is a cusco (Exercise 5 in Section 8.2).

Suppose $U \subset E$ is an open set, $Y$ is a Euclidean space, and $\Phi : U \to Y$ is a cusco. We call $\Phi$ minimal if its graph is minimal (with respect to set inclusion) among graphs of cuscors from $U$ to $Y$. For example, the subdifferential of a continuous convex function is a minimal cusco (Exercise 8). We next use this fact to prove that Clarke subdifferentials of distance functions are also minimal cuscors.

**Theorem 9.2.4 (Distance subdifferentials are minimal)** Outside a nonempty closed set $S \subset E$, the distance function $d_S$ can be expressed locally as the difference between a smooth convex function and a continuous convex function. Consequently, the Clarke subdifferential $\partial d_S : E \to E$ is a minimal cusco.

**Proof.** Consider any closed ball $T$ disjoint from $S$. For any point $y$ in $S$, it is easy to check that the Fréchet derivative of the function $x \mapsto \|x - y\|$ is Lipschitz on $T$. Suppose the Lipschitz constant is $2L$. It follows that the function $x \mapsto L\|x\|^2 - \|x - y\|$ is convex on $T$ (see Exercise 9). Since the function $h : T \to \mathbb{R}$ defined by

$$h(x) = L\|x\|^2 - d_S(x) = \sup_{y \in S} \{L\|x\|^2 - \|x - y\|\}$$

is convex, we obtain the desired expression $d_S = L\|\cdot\|^2 - h$.

To prove minimality, consider any cusco $\Phi : E \to E$ satisfying $\Phi(x) \subset \partial d_S(x)$ for all points $x$ in $E$. Notice that for any point $x \in \text{int } T$ we have

$$\partial d_S(x) = -\partial (-d_S)(x) = \partial h(x) - Lx.$$

Since $h$ is convex on int $T$, the subdifferential $\partial h$ is a minimal cusco on this set, and hence so is $\partial d_S$. Consequently, $\Phi$ must agree with $\partial d_S$ on int $T$, and hence throughout $S^c$, since $T$ was arbitrary.

On the set $\text{int } S$, the function $d_S$ is identically zero. Hence for all points $x$ in $\text{int } S$ we have $\partial d_S = \{0\}$ and therefore also $\Phi(x) = \{0\}$. We also deduce $0 \in \Phi(x)$ for all $x \in \text{cl (int } S)$. 


(iii) the distance function $d_S$ is Fréchet differentiable at $x$.

In this case,

$$\nabla d_S(x) = \frac{x - \bar{x}}{\|x - \bar{x}\|} \in N_S^p(\bar{x}) \subseteq N_S(\bar{x}).$$

The proof is outlined in Exercises 4 and 6.

For our alternate proof of the proximal normal formula without recourse to Rademacher’s theorem, we return to an idea we introduced in Section 8.2. A cusco is a USC multifunction with nonempty compact convex images. In particular, the Clarke subdifferential of a locally Lipschitz function on an open set is a cusco (Exercise 5 in Section 8.2).

Suppose $U \subset E$ is an open set, $Y$ is a Euclidean space, and $\Phi : U \to Y$ is a cusco. We call $\Phi$ minimal if its graph is minimal (with respect to set inclusion) among graphs of cuscos from $U$ to $Y$. For example, the subdifferential of a continuous convex function is a minimal cusco (Exercise 8). We next use this fact to prove that Clarke subdifferentials of distance functions are also minimal cuscos.

**Theorem 9.2.4 (Distance subdifferentials are minimal)** Outside a nonempty closed set $S \subset E$, the distance function $d_S$ can be expressed locally as the difference between a smooth convex function and a continuous convex function. Consequently, the Clarke subdifferential $\partial d_S : E \to E$ is a minimal cusco.

**Proof.** Consider any closed ball $T$ disjoint from $S$. For any point $y$ in $S$, it is easy to check that the Fréchet derivative of the function $x \mapsto \|x - y\|$ is Lipschitz on $T$. Suppose the Lipschitz constant is $2L$. It follows that the function $x \mapsto L\|x\|^2 - \|x - y\|$ is convex on $T$ (see Exercise 9). Since the function $h : T \to \mathbb{R}$ defined by

$$h(x) = L\|x\|^2 - d_S(x) = \sup_{y \in S} \{L\|x\|^2 - \|x - y\|\}$$

is convex, we obtain the desired expression $d_S = L\|\cdot\|^2 - h$.

To prove minimality, consider any cusco $\Phi : E \to E$ satisfying $\Phi(x) \subseteq \partial d_S(x)$ for all points $x$ in $E$. Notice that for any point $x \in \text{int } T$ we have

$$\partial d_S(x) = -\partial(-d_S)(x) = \partial h(x) - Lx.$$

Since $h$ is convex on int $T$, the subdifferential $\partial h$ is a minimal cusco on this set, and hence so is $\partial d_S$. Consequently, $\Phi$ must agree with $\partial d_S$ on int $T$, and hence throughout $S^c$, since $T$ was arbitrary.

On the set int $S$, the function $d_S$ is identically zero. Hence for all points $x$ in int $S$ we have $\partial d_S = \{0\}$ and therefore also $\Phi(x) = \{0\}$. We also deduce $0 \in \Phi(x)$ for all $x \in \text{cl } (\text{int } S)$. 

Now consider a point \( x \in \text{bd} S \). The Mean value theorem (Exercise 9 in Section 6.1) shows
\[
\partial_d d_S(x) = \text{conv} \left\{ 0, \lim_{y^r \rightarrow x} y^r \in \partial_d d_S(x^r), \ x^r \rightarrow x, \ x^r \notin S \right\}
\]
where 0 can be omitted from the convex hull unless \( x \in \text{cl} (\text{int} S) \) (see Exercise 10). But the final set is contained in \( \Phi(x) \), so the result now follows.

The Proximal normal formula (Theorem 9.2.1) follows rather quickly from this result (and indeed can be strengthened), using the fact that Clarke subgradients of the distance function are proximal normals (Exercise 11).

We end this section with another elegant illustration of the geometry of nearest points. We call a set \( S \subset \mathbb{E} \) a Chebyshev set if every point in \( \mathbb{E} \) has a unique nearest point \( \text{P}_S(x) \in S \). Any nonempty closed convex set is a Chebyshev set (Exercise 8 in Section 2.1). Much less obvious is the converse, stated in the following result.

**Theorem 9.2.5 (Convexity of Chebyshev sets)** A subset of a Euclidean space is a Chebyshev set if and only if it is nonempty, closed and convex.

**Proof.** Consider a Chebyshev set \( S \subset \mathbb{E} \). Clearly \( S \) is nonempty and closed, and it is easy to verify that the projection \( P_S : \mathbb{E} \rightarrow \mathbb{E} \) is continuous. To prove \( S \) is convex, we first introduce another new notion. We call \( S \) a sun if, for each point \( x \in \mathbb{E} \), every point on the ray \( P_S(x) + \mathbb{R}^+ (x - P_S(x)) \) has nearest point \( P_S(x) \). We begin by proving that the following properties are equivalent (see Exercise 13):

(i) \( S \) is convex;

(ii) \( S \) is a sun;

(iii) \( P_S \) is nonexpansive.

So, we need to show that \( S \) is a sun.

Suppose \( S \) is not a sun, so there is a point \( x \notin S \) with nearest point \( P_S(x) = \bar{x} \) such that the ray \( L = \bar{x} + \mathbb{R}^+ (x - \bar{x}) \) strictly contains
\[
\{ z \in L \mid P_S(z) = \bar{x} \}.
\]
Hence by Proposition 9.2.2 (Projections) and the continuity of \( P_S \), the above set is nontrivial closed line segment \([\bar{x}, x_0]\) containing \( x \).

Choose a radius \( \epsilon > 0 \) so that the ball \( x_0 + \epsilon B \) is disjoint from \( S \). The continuous self map of this ball
\[
z \mapsto x_0 + \epsilon \frac{x_0 - P_S(z)}{\|x_0 - P_S(z)\|}
\]
Now consider a point $x \in \partial S$. The Mean value theorem (Exercise 9 in Section 6.1) shows
\[
\partial S(x) = \text{conv}\left\{0, \lim_{r \to 0} y^r \mid y^r \in \partial S(x^r), \ x^r \to x, \ x^r \notin S \right\}
\]
where 0 can be omitted from the convex hull unless $x \in \text{cl}(\text{int} S)$ (see Exercise 10). But the final set is contained in $\Phi(x)$, so the result now follows.

The Proximal normal formula (Theorem 9.2.1), follows rather quickly from this result (and indeed can be strengthened), using the fact that Clarke subgradients of the distance function are proximal normals (Exercise 11).

We end this section with another elegant illustration of the geometry of nearest points. We call a set $S \subset E$ a Chebyshev set if every point in $E$ has a unique nearest point $P_S(x)$ in $S$. Any nonempty closed convex set is a Chebyshev set (Exercise 8 in Section 2.1). Much less obvious is the converse, stated in the following result.

**Theorem 9.2.5 (Convexity of Chebyshev sets)** A subset of a Euclidean space is a Chebyshev set if and only if it is nonempty, closed and convex.

**Proof.** Consider a Chebyshev set $S \subset E$. Clearly $S$ is nonempty and closed, and it is easy to verify that the projection $P_S : E \to E$ is continuous. To prove $S$ is convex, we first introduce another new notion. We call $S$ a sun if, for each point $x \in E$, every point on the ray $P_S(x) + \mathbb{R}_+(x - P_S(x))$ has nearest point $P_S(x)$. We begin by proving that the following properties are equivalent (see Exercise 13):

(i) $S$ is convex;
(ii) $S$ is a sun;
(iii) $P_S$ is nonexpansive.

So, we need to show that $S$ is a sun.

Suppose $S$ is not a sun, so there is a point $x \notin S$ with nearest point $P_S(x) = \bar{x}$ such that the ray $L = \bar{x} + \mathbb{R}_+(x - \bar{x})$ strictly contains
\[
\{z \in L \mid P_S(z) = \bar{x}\}.
\]
Hence by Proposition 9.2.2 (Projections) and the continuity of $P_S$, the above set is nontrivial closed line segment $[\bar{x}, x_0]$ containing $x$.

Choose a radius $\epsilon > 0$ so that the ball $x_0 + \epsilon B$ is disjoint from $S$. The continuous self map of this ball
\[
z \mapsto x_0 + \epsilon \frac{x_0 - P_S(z)}{\|x_0 - P_S(z)\|}
\]
Approximately convex sets
and Suns
9.2 Proximal Normals and Chebyshev Sets

We introduced the Clarke normal cone in Section 6.3 (Tangent Cones), via the Clarke subdifferential. An appealing alternative approach begins with a more geometric notion of a normal vector. We call a vector \( y \in \mathbb{E} \) a \textit{proximal normal} to a set \( S \subset \mathbb{E} \) at a point \( x \in S \) if, for some \( t > 0 \), the nearest point to \( x + ty \) in \( S \) is \( x \). The set of all such vectors is called the \textit{proximal normal cone}, which we denote \( N_p^S(x) \).

The proximal normal cone, which may not be convex, is contained in the Clarke normal cone (Exercise 3). The containment may be strict, but we can reconstruct the Clarke normal cone from proximal normals using the following result.

\[ \text{Theorem 9.2.1 (Proximal normal formula)} \quad \text{For any closed set } S \subset \mathbb{E} \text{ and any point } x \in S, \text{ we have} \]
\[ N_S(x) = \text{conv} \left\{ \lim_{r \to 0} y_r \mid y_r \in N_p^S(x_r), \ x_r \in S, \ x_r \to x \right\}. \]

One route to this result uses Rademacher’s theorem (Exercise 7). In this section we take a more direct approach.

The Clarke normal cone to a set \( S \subset \mathbb{E} \) at a point \( x \in S \) is
\[ N_S(x) = \text{cl} (\mathbb{R}_+ \partial d_S(x)), \]
by Theorem 6.3.8, where
\[ d_S(x) = \inf_{z \in S} \|z - x\| \]
is the distance function. Notice the following elementary but important result that we use repeatedly in this section (Exercise 4(a) in Section 7.3).

\[ \text{Proposition 9.2.2 (Projections)} \quad \text{If } \bar{x} \text{ is a nearest point in the set } S \subset \mathbb{E} \text{ to the point } x \in \mathbb{E}, \text{ then } \bar{x} \text{ is the unique nearest point in } S \text{ to each point on the half-open line segment } [\bar{x}, x). \]

To derive the proximal normal formula from the subdifferential formula (9.1.1), we can make use of some striking differentiability properties of distance functions, summarized in the next result.

\[ \text{Theorem 9.2.3 (Differentiability of distance functions)} \quad \text{Consider a nonempty closed set } S \subset \mathbb{E} \text{ and a point } x \notin S. \text{ Then the following properties are equivalent:} \]

(i) the Dini subdifferential \( \partial_- d_S(x) \) is nonempty;

(ii) \( x \) has a unique nearest point \( \bar{x} \) in \( S \);
So, we need to show that $S$ is a sun.

Suppose $S$ is not a sun, so there is a point $x \notin S$ with nearest point
$P_S(x) = \bar{x}$ such that the ray $L = \bar{x} + \mathbb{R}_+(x - \bar{x})$ strictly contains

$$\{ z \in L \mid P_S(z) = \bar{x} \}.$$

Hence by Proposition 9.2.2 (Projections) and the continuity of $P_S$, the
above set is nontrivial closed line segment $[\bar{x}, x_0]$ containing $x$.

Choose a radius $\epsilon > 0$ so that the ball $x_0 + \epsilon B$ is disjoint from $S$. The
continuous self map of this ball

$$z \mapsto x_0 + \epsilon \frac{x_0 - P_S(z)}{\|x_0 - P_S(z)\|}$$
Not a Sun

P(x)

x

z?

P(z)

x0

P(x)
has a fixed point by Brouwer’s theorem (8.1.3). We then quickly derive a contradiction to the definition of the point \( x_0 \).

Exercises and Commentary

Proximal normals provide an alternative comprehensive approach to non-smooth analysis: a good reference is [56]. Our use of the minimality of distance subdifferentials here is modelled on [38]. Theorem 9.2.5 (Convexity of Chebyshev sets) is sometimes called the “Motzkin-Bunt theorem”. Our discussion closely follows [62]. In the exercises, we outline three nonsmooth proofs. The first (Exercises 14, 15, 16) is a variational proof following [82]. The second (Exercises 17, 18, 19) follows [96], and uses Fenchel conjugacy. The third argument (Exercises 20 and 21) is due to Asplund [2]. It is the most purely geometric, first deriving an interesting dual result on furthest points, and then proceeding via inversion in the unit sphere. Asplund extended the argument to Hilbert space, where it remains unknown whether a norm-closed Chebyshev set must be convex. Asplund showed that, in seeking a nonconvex Chebyshev set, we can restrict attention to “Klee caverns”: complements of closed bounded convex sets.

1. Consider a closed set \( S \subset E \) and a point \( x \in S \).
   (a) Show that the proximal normal cone \( N^p_S(x) \) may not be convex.
   (b) Prove \( x \in \text{int} \, S \implies N^p_S(x) = \{0\} \).
   (c) Is the converse to part (b) true?
   (d) Prove the set \( \{z \in S \mid N^p_S(z) \neq \{0\} \} \) is dense in the boundary of \( S \).

2. (Projections) Prove Proposition 9.2.2.

3. (Proximal normals are normals) Consider a set \( S \subset E \). Suppose the unit vector \( y \in E \) is a proximal normal to \( S \) at the point \( x \in S \).
   (a) Use Proposition 9.2.2 (Projections) to prove \( d'_S(x; y) = 1 \).
   (b) Use the Lipschitz property of the distance function to prove \( \partial_S d_S(x) \subset B \).
   (c) Deduce \( y \in \partial_S d_S(x) \).
   (d) Deduce that any proximal normal lies in the Clarke normal cone.

4. * (Unique nearest points) Consider a closed set \( S \subset E \) and a point \( x \) outside \( S \) with unique nearest point \( \bar{x} \) in \( S \). Complete the following steps to prove
   \[
   \frac{x - \bar{x}}{\|x - \bar{x}\|} \in \partial_{-} d_S(x).
   \]
has a fixed point by Brouwer’s theorem (8.1.3). We then quickly derive a contradiction to the definition of the point \( x_0 \). 

\[ \square \]

**Exercises and Commentary**

Proximal normals provide an alternative comprehensive approach to non-smooth analysis: a good reference is [56]. Our use of the minimality of distance subdifferentials here is modelled on [38]. Theorem 9.2.5 (Convexity of Chebyshev sets) is sometimes called the “Motzkin-Bunt theorem”. Our discussion closely follows [62]. In the exercises, we outline three nonsmooth proofs. The first (Exercises 14, 15, 16) is a variational proof following [82]. The second (Exercises 17, 18, 19) follows [96], and uses Fenchel conjugacy. The third argument (Exercises 20 and 21) is due to Asplund [2]. It is the most purely geometric, first deriving an interesting dual result on furthest points, and then proceeding via inversion in the unit sphere. Asplund extended the argument to Hilbert space, where it remains unknown whether a norm-closed Chebyshev set must be convex. Asplund showed that, in seeking a nonconvex Chebyshev set, we can restrict attention to “Klee caverns”: complements of closed bounded convex sets.

1. Consider a closed set \( S \subset E \) and a point \( x \in S \).
   
   (a) Show that the proximal normal cone \( N_p^S(x) \) may not be convex.
   
   (b) Prove \( x \in \text{int } S \Rightarrow N_p^S(x) = \{0\} \).
   
   (c) Is the converse to part (b) true?
   
   (d) Prove the set \( \{ z \in S \mid N_p^S(z) \neq \{0\} \} \) is dense in the boundary of \( S \).

2. **(Projections) Prove Proposition 9.2.2.**

3. **(Proximal normals are normals)** Consider a set \( S \subset E \). Suppose the unit vector \( y \in E \) is a proximal normal to \( S \) at the point \( x \in S \).
   
   (a) Use Proposition 9.2.2 (Projections) to prove \( d'_S(x; y) = 1 \).
   
   (b) Use the Lipschitz property of the distance function to prove \( \partial_0 d_S(x) \subset B \).
   
   (c) Deduce \( y \in \partial_0 d_S(x) \).
   
   (d) Deduce that any proximal normal lies in the Clarke normal cone.

4. "**(Unique nearest points)** Consider a closed set \( S \subset E \) and a point \( x \) outside \( S \) with unique nearest point \( \bar{x} \) in \( S \). Complete the following steps to prove
   
   \[ \frac{x - \bar{x}}{\|x - \bar{x}\|} \in \partial d_S(x). \]
(a) Assuming the result fails, prove there exists a direction $h \in \mathbf{E}$ such that
\[ d_S^-(x; h) < \langle \|x - \bar{x}\|^{-1}(x - \bar{x}), h \rangle. \]

(b) Consider a sequence $t_r \downarrow 0$ such that
\[ \frac{d_S(x + t_r h) - d_S(x)}{t_r} \rightarrow d_S^-(x; h) \]
and suppose each point $x + t_r h$ has a nearest point $s_r$ in $S$. Prove $s_r \rightarrow \bar{x}$.

(c) Use the fact that the gradient of the norm at the point $x - s_r$ is a subgradient to deduce a contradiction.

5. **(Nearest points and Clarke subgradients)** Consider a closed set $S \subset \mathbf{E}$ and a point $x$ outside $S$ with a nearest point $\bar{x}$ in $S$. Use Exercise 4 to prove $x - \bar{x} / \|x - \bar{x}\| \in \partial d_S(x)$.

6. **(Differentiability of distance functions)** Consider a nonempty closed set $S \subset \mathbf{E}$.

(a) For any points $x, z \in \mathbf{E}$, observe the identity
\[ d_S^2(z) - d_S^2(x) = 2d_S(x)(d_S(z) - d_S(x)) + (d_S(z) - d_S(x))^2. \]

(b) Use the Lipschitz property of the distance function to deduce
\[ 2d_S(x)\partial_+ d_S(x) \subset \partial_+ d_S^2(x). \]

Now suppose $y \in \partial_+ d_S(x)$.

(c) If $\bar{x}$ is any nearest point to $x$ in $S$, use part (b) to prove $\bar{x} = x - d_S(x)y$, so $\bar{x}$ is in fact the unique nearest point.

(d) Prove $-2d_S(x)y \in \partial_-(d_S^2)(x)$.

(e) Deduce $d_S^2$ is Fréchet differentiable at $x$.

Assume $x \notin S$.

(f) Deduce $d_S$ is Fréchet differentiable at $x$.

(g) Use Exercises 3 and 4 to complete the proof of Theorem 9.2.3.

7. **(Proximal normal formula via Rademacher)** Prove Theorem 9.2.1 using the subdifferential formula (9.1.1) and Theorem 9.2.3 (Differentiability of distance functions).
(a) Assuming the result fails, prove there exists a direction \( h \in E \) such that
\[
d_S^-(x; h) < \langle \|x - \bar{x}\|^{-1}(x - \bar{x}), h \rangle.
\]

(b) Consider a sequence \( t_r \downarrow 0 \) such that
\[
\frac{d_S(x + t_r h) - d_S(x)}{t_r} \rightarrow d_S^-(x; h)
\]
and suppose each point \( x + t_r h \) has a nearest point \( s_r \) in \( S \). Prove \( s_r \rightarrow \bar{x} \).

(c) Use the fact that the gradient of the norm at the point \( x - s_r \) is a subgradient to deduce a contradiction.

5. **(Nearest points and Clarke subgradients)** Consider a closed set \( S \subset E \) and a point \( x \) outside \( S \) with a nearest point \( \bar{x} \) in \( S \). Use Exercise 4 to prove
\[
\frac{x - \bar{x}}{\|x - \bar{x}\|} \in \partial_S d_S(x).
\]

6. **(Differentiability of distance functions)** Consider a nonempty closed set \( S \subset E \).

   (a) For any points \( x, z \in E \), observe the identity
   \[
d_S^2(z) - d_S^2(x) = 2d_S(x)(d_S(z) - d_S(x)) + (d_S(z) - d_S(x))^2.
   \]

   (b) Use the Lipschitz property of the distance function to deduce
   \[
   2d_S(x)\partial_- d_S(x) \subset \partial_- d_S^2(x).
   \]

   Now suppose \( y \in \partial_- d_S(x) \).

   (c) If \( \bar{x} \) is any nearest point to \( x \) in \( S \), use part (b) to prove \( \bar{x} = x - d_S(x)y \), so \( \bar{x} \) is in fact the unique nearest point.

   (d) Prove \( -2d_S(x)y \in \partial_- (-d_S^2)(x) \).

   (e) Deduce \( d_S^2 \) is Fréchet differentiable at \( x \).

Assume \( x \not\in S \).

(f) Deduce \( d_S \) is Fréchet differentiable at \( x \).

(g) Use Exercises 3 and 4 to complete the proof of Theorem 9.2.3.

7. **(Proximal normal formula via Rademacher)** Prove Theorem 9.2.1 using the subdifferential formula (9.1.1) and Theorem 9.2.3 (Differentiability of distance functions).
8. **(Minimality of convex subdifferentials)** If the open set $U \subset E$ is convex and the function $f : U \to \mathbb{R}$ is convex, use the Max formula (Theorem 3.1.8) to prove that the subdifferential $\partial f$ is a minimal cusco.

9. **(Smoothness and DC functions)** Suppose the set $C \subset E$ is open and convex, and the Fréchet derivative of the function $g : C \to \mathbb{R}$ has Lipschitz constant $2L$ on $C$. Deduce that the function $L\|\cdot\|^2 - g$ is convex on $C$.

10. **(Subdifferentials at minimizers)** Consider a locally Lipschitz function $f : E \to \mathbb{R}_+$, a point $x$ in $f^{-1}(0)$. Prove

$$
\partial f(x) = \text{conv}\left\{0, \lim_{r \to 0^+} y^r \bigg| y^r \in \partial f(x^r), \ x^r \to x, \ f(x^r) > 0\right\},
$$

where 0 can be omitted from the convex hull if int $f^{-1}(0) = \emptyset$.

11. **(Proximal normals and the Clarke subdifferential)** Consider a closed set $S \subset E$ and a point $x$ in $S$ Use Exercises 3 and 5 and the minimality of the subdifferential $\partial d_S : E \to E$ to prove

$$
\partial d_S(x) = \text{conv}\left\{0, \lim_{r \to 0^+} y^r \bigg| y^r \in N^p_S(x^r), \ \|y^r\| = 1, \ x^r \to x, \ x^r \in S\right\}.
$$

Deduce the Proximal normal formula (Theorem 9.2.1). Assuming $x \in \text{bd} S$, prove the following stronger version. Consider any dense subset $Q$ of $S^c$, and suppose $P : Q \to S$ maps each point in $Q$ to a nearest point in $S$. Prove

$$
\partial d_S(x) = \text{conv}\left\{0, \lim_{r \to 0^+} \frac{x^r - P(x^r)}{\|x^r - P(x^r)\|} \bigg| x^r \to x, \ x^r \in Q\right\},
$$

and derive a stronger version of the Proximal normal formula.

12. **(Continuity of the projection)** Consider a Chebyshev set $S$. Prove directly from the definition that the projection $P_S$ is continuous.

13. * **(Suns)** Complete the details in the proof of Theorem 9.2.5 (Convexity of Chebyshev sets) as follows.

   (a) Prove (iii) $\Rightarrow$ (i).

   (b) Prove (i) $\Rightarrow$ (ii).

   (c) Denoting the line segment between points $y, z \in E$ by $[y, z]$, prove property (ii) implies

   $$
P_S(x) = P_{[z, P_S(z)]}(x) \text{ for all } x \in E, \ z \in S. \quad (9.2.6)
$$
8. (Minimality of convex subdifferentials) If the open set $U \subset E$ is convex and the function $f : U \rightarrow \mathbb{R}$ is convex, use the Max formula (Theorem 3.1.8) to prove that the subdifferential $\partial f$ is a minimal cusco.

9. (Smoothness and DC functions) Suppose the set $C \subset E$ is open and convex, and the Fréchet derivative of the function $g : C \rightarrow \mathbb{R}$ has Lipschitz constant $2L$ on $C$. Deduce that the function $L\|\cdot\|^2 - g$ is convex on $C$.

10. ** (Subdifferentials at minimizers) Consider a locally Lipschitz function $f : E \rightarrow \mathbb{R}_+$, and a point $x$ in $f^{-1}(0)$. Prove

$$\partial_y f(x) = \text{conv}\left\{0, \lim_{r \to 0} y^r \left| y^r \in \partial_y f(x^r), x^r \to x, f(x^r) > 0\right\},$$

where 0 can be omitted from the convex hull if int $f^{-1}(0) = \emptyset$.

11. ** (Proximal normals and the Clarke subdifferential) Consider a closed set $S \subset E$ and a point $x$ in $S$. Use Exercises 3 and 5 and the minimality of the subdifferential $\partial_y d_S : E \rightarrow E$ to prove

$$\partial_y d_S(x) = \text{conv}\left\{0, \lim_{r \to 0} y^r \left| y^r \in N^y_S(x^r), \|y^r\| = 1, x^r \to x, x^r \in S\right\}.$$ 

Deduce the Proximal normal formula (Theorem 9.2.1). Assuming $x \in \text{bd} S$, prove the following stronger version. Consider any dense subset $Q$ of $S^c$, and suppose $P : Q \rightarrow S$ maps each point in $Q$ to a nearest point in $S$. Prove

$$\partial_y d_S(x) = \text{conv}\left\{0, \lim_{r \to 0} \frac{x^r - P(x^r)}{\|x^r - P(x^r)\|} \left| x^r \to x, x^r \in Q\right\},$$

and derive a stronger version of the Proximal normal formula.

12. (Continuity of the projection) Consider a Chebyshev set $S$. Prove directly from the definition that the projection $P_S$ is continuous.

13. * (Suns) Complete the details in the proof of Theorem 9.2.5 (Convexity of Chebyshev sets) as follows.

(a) Prove (iii) $\Rightarrow$ (i).

(b) Prove (i) $\Rightarrow$ (ii).

(c) Denoting the line segment between points $y, z \in E$ by $[y, z]$, prove property (ii) implies

$$P_S(x) = P_{[z, P_S(x)]}(x) \text{ for all } x \in E, \ z \in S. \quad (9.2.6)$$
12. **(Continuity of the projection)** Consider a Chebyshev set $S$. Prove directly from the definition that the projection $P_S$ is continuous.

13. ***(Suns)*** Complete the details in the proof of Theorem 9.2.5 (Convexity of Chebyshev sets) as follows.

(a) Prove (iii) $\Rightarrow$ (i).

(b) Prove (i) $\Rightarrow$ (ii).

(c) Denoting the line segment between points $y, z \in E$ by $[y, z]$, prove property (ii) implies

$$P_S(x) = P_{[z, P_S(x)]}(x) \text{ for all } x \in E, z \in S.$$  

(9.2.6)
(d) Prove (9.2.6) \(\Rightarrow\) (iii).

(e) Fill in the remaining details of the proof.

**14. (Basic Ekeland variational principle [43])** Prove the following version of the Ekeland variation principle (Theorem 7.1.2). Suppose the function \(f : E \to (-\infty, +\infty]\) is closed and the point \(x \in E\) satisfies \(f(x) < \inf f + \epsilon\) for some real \(\epsilon > 0\). Then for any real \(\lambda > 0\) there is a point \(v \in E\) satisfying the conditions

\[
\begin{align*}
(a) \quad & \|x - v\| \leq \lambda, \\
(b) \quad & f(v) + (\epsilon/\lambda)\|x - v\| \leq f(x), \text{ and} \\
(c) \quad & v \text{ minimizes the function } f(\cdot) + (\epsilon/\lambda)\|\cdot - v\|.
\end{align*}
\]

**15. (Approximately convex sets)** Consider a closed set \(C \subset E\). We call \(C\) approximately convex if, for any closed ball \(D \subset E\) disjoint from \(C\), there exists a closed ball \(D' \supset D\) disjoint from \(C\) with arbitrarily large radius.

(a) If \(C\) is convex, prove it is approximately convex.

(b) Suppose \(C\) is approximately convex but not convex.

(i) Prove there exist points \(a, b \in C\) and a closed ball \(D\) centered at the point \(c = (a + b)/2\) and disjoint from \(C\).

(ii) Prove there exists a sequence of points \(x_1, x_2, \ldots \in E\) such that the balls \(B_r = x_r + rB\) are disjoint from \(C\) and satisfy \(D \subset B_r \subset B_{r+1}\) for all \(r = 1, 2, \ldots\).

(iii) Prove the set \(H = \operatorname{cl} \cup_r B_r\) is closed and convex, and its interior is disjoint from \(C\) but contains \(c\).

(iv) Suppose the unit vector \(u\) lies in the polar set \(H^\circ\). By considering the quantity \(\langle u, \|x_r - x\|^{-1}(x_r - x) \rangle\) as \(r \to \infty\), prove \(H^\circ\) must be a ray.

(v) Deduce a contradiction.

(c) Conclude that a closed set is convex if and only if it is approximately convex.

**16. (Chebyshev sets and approximate convexity)** Consider a Chebyshev set \(C \subset E\), and a ball \(x + \beta B\) disjoint from \(C\).

(a) Use Theorem 9.2.3 (Differentiability of distance functions) to prove

\[ \limsup_{v \to x} \frac{d_C(v) - d_C(x)}{\|v - x\|} = 1. \]
Ekeland replaces Brouwer

A Chebyshev Sun is Approximately Convex

• replaces exact Hilbert proximal analysis by $\varepsilon$-variations
• works for weakly closed sets in smooth rotund space
(d) Prove (9.2.6) $\Rightarrow$ (iii).
(e) Fill in the remaining details of the proof.

14. **(Basic Ekeland variational principle [43])** Prove the following version of the Ekeland variation principle (Theorem 7.1.2). Suppose the function $f : E \to (\infty, +\infty]$ is closed and the point $x \in E$ satisfies $f(x) < \inf f + \epsilon$ for some real $\epsilon > 0$. Then for any real $\lambda > 0$ there is a point $v \in E$ satisfying the conditions

(a) $\|x - v\| \leq \lambda$,
(b) $f(v) + (\epsilon/\lambda)\|x - v\| \leq f(x)$, and
(c) $v$ minimizes the function $f(\cdot) + (\epsilon/\lambda)\|\cdot - v\|$.

15. * (Approximately convex sets) Consider a closed set $C \subset E$. We call $C$ *approximately convex* if, for any closed ball $D \subset E$ disjoint from $C$, there exists a closed ball $D' \supset D$ disjoint from $C$ with arbitrarily large radius.

(a) If $C$ is convex, prove it is approximately convex.
(b) Suppose $C$ is approximately convex but not convex.

(i) Prove there exist points $a, b \in C$ and a closed ball $D$ centered at the point $c = (a + b)/2$ and disjoint from $C$.
(ii) Prove there exists a sequence of points $x_1, x_2, \ldots \in E$ such that the balls $B_r = x_r + rB$ are disjoint from $C$ and satisfy $D \subset B_r \subset B_{r+1}$ for all $r = 1, 2, \ldots$
(iii) Prove the set $H = \text{cl} \cup_r B_r$ is closed and convex, and its interior is disjoint from $C$ but contains $c$.
(iv) Suppose the unit vector $u$ lies in the polar set $H^\circ$. By considering the quantity $(u, \|x_r - x\|^{-1}(x_r - x))$ as $r \to \infty$, prove $H^\circ$ must be a ray.
(v) Deduce a contradiction.
(c) Conclude that a closed set is convex if and only if it is approximately convex.

16. **(Chebyshev sets and approximate convexity)** Consider a Chebyshev set $C \subset E$, and a ball $x + \beta B$ disjoint from $C$.

(a) Use Theorem 9.2.3 (Differentiability of distance functions) to prove
\[
\limsup_{v \to x} \frac{d_C(v) - d_C(x)}{\|v - x\|} = 1.
\]
(b) Consider any real $\alpha > d_C(x)$. Fix reals $\sigma \in (0, 1)$ and $\rho$ satisfying
\[
\frac{\alpha - d_C(x)}{\sigma} < \rho < \alpha - \beta.
\]
By applying the Basic Ekeland variational principle (Exercise 14) to the function $-d_C + \delta_{x+\rho B}$, prove there exists a point $v \in E$ satisfying the conditions
\[
\begin{align*}
d_C(x) + \sigma \|x-v\| &\leq d_C(v) \\
d_C(z) - \sigma \|z-v\| &\leq d_C(v)
\end{align*}
\]
for all $z \in x + \rho B$.

Use part (a) to deduce $\|x-v\| = \rho$, and hence $x + \beta B \subset v + \alpha B$.

(c) Conclude that $C$ is approximately convex, and hence convex by Exercise 15.

(d) Extend this argument to an arbitrary norm on $E$.

17. ** (Smoothness and biconjugacy) Consider a function $f : E \to (\infty, +\infty]$ that is closed and bounded below and satisfies the condition
\[
\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = +\infty.
\]
Consider also a point $x \in \text{dom} f$.

(a) Using Carathéodory’s theorem (Section 2.2, Exercise 5), prove there exist points $x_1, x_2, \ldots, x_m \in E$ and real $\lambda_1, \lambda_2, \ldots, \lambda_m > 0$ satisfying
\[
\sum_i \lambda_i = 1, \quad \sum_i \lambda_i x_i = x, \quad \sum_i \lambda_i f(x_i) = f^{**}(x).
\]

(b) Use the Fenchel-Young inequality (Proposition 3.3.4) to prove
\[
\partial(f^{**})(x) = \bigcap_i \partial f(x_i).
\]

Suppose furthermore that the conjugate $f^*$ is everywhere differentiable.

(c) If $x \in \text{ri}(\text{dom}(f^{**}))$, prove $x_i = x$ for each $i$.

(d) Deduce $\text{ri}(\text{epi}(f^{**})) \subset \text{epi}(f)$.

(e) Use the fact that $f$ is closed to deduce $f = f^{**}$, so $f$ is convex.

18. * (Chebyshev sets and differentiability) Use Theorem 9.2.3 (Differentiability of distance functions) to prove that a closed set $S \subset E$ is a Chebyshev set if and only if the function $d_S^2$ is Fréchet differentiable throughout $E$. 
(b) Consider any real \( \alpha > d_C(x) \). Fix reals \( \sigma \in (0, 1) \) and \( \rho \) satisfying
\[
\alpha - d_C(x) < \frac{\alpha - d_C(x)}{\sigma} < \rho < \alpha - \beta.
\]
By applying the Basic Ekeland variational principle (Exercise 14) to the function \( -d_C + \delta_{x + \rho B} \), prove there exists a point \( v \in E \) satisfying the conditions
\[
d_C(x) + \sigma \|x - v\| \leq d_C(v)
\]
\[
d_C(z) - \sigma \|z - v\| \leq d_C(v) \text{ for all } z \in x + \rho B.
\]
Use part (a) to deduce \( \|x - v\| = \rho \), and hence \( x + \beta B \subset v + \alpha B \).

(c) Conclude that \( C \) is approximately convex, and hence convex by Exercise 15.

(d) Extend this argument to an arbitrary norm on \( E \).

17. ** (Smoothness and biconjugacy) Consider a function \( f : E \to (\infty, +\infty] \) that is closed and bounded below and satisfies the condition
\[
\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = +\infty.
\]
Consider also a point \( x \in \text{dom} \ f \).

(a) Using Carathéodory’s theorem (Section 2.2, Exercise 5), prove there exist points \( x_1, x_2, \ldots, x_m \in E \) and real \( \lambda_1, \lambda_2, \ldots, \lambda_m > 0 \) satisfying
\[
\sum_i \lambda_i = 1, \quad \sum_i \lambda_i x_i = x, \quad \sum_i \lambda_i f(x_i) = f^{**}(x).
\]

(b) Use the Fenchel-Young inequality (Proposition 3.3.4) to prove
\[
\partial(f^{**})(x) = \bigcap_i \partial f(x_i).
\]
Suppose furthermore that the conjugate \( f^* \) is everywhere differentiable.

(c) If \( x \in \text{ri(dom}(f^{**})) \), prove \( x_i = x \) for each \( i \).

(d) Deduce \( \text{ri(epi}(f^{**})) \subset \text{epi}(f) \).

(e) Use the fact that \( f \) is closed to deduce \( f = f^{**} \), so \( f \) is convex.

18. * (Chebyshev sets and differentiability) Use Theorem 9.2.3 (Differentiability of distance functions) to prove that a closed set \( S \subset E \) is a Chebyshev set if and only if the function \( d_S^2 \) is Fréchet differentiable throughout \( E \).
$W$ is convex if and only if $W^*$ is Frechet

$$W(x) = \left(1 - x^2\right)^2$$
19. **(Chebyshev convexity via conjugacy)** For any nonempty closed set \( S \subset E \), prove

\[
\left( \frac{\| \cdot \|^2 + \delta_S}{2} \right)^* = \frac{\| \cdot \|^2 - d_S^2}{2}
\]

Deduce, using Exercises 17 and 18, that Chebyshev sets are convex.

20. **(Unique furthest points)** Consider a set \( S \subset E \), and define a function \( r_S : E \to [-\infty, +\infty] \) by

\[
r_S(x) = \sup_{y \in S} \| x - y \|
\]

Any point \( y \) attaining the above supremum is called a furthest point in \( S \) to the point \( x \in E \).

(a) Prove that the function \( (r_S^2 - \| \cdot \|^2)/2 \) is the conjugate of the function

\[
g_S = \frac{\delta_{-S} - \| \cdot \|^2}{2}
\]

(b) Prove that the function \( r_S^2 \) is strictly convex on its domain.

Now suppose each point \( x \in E \) has a unique nearest point \( q_S(x) \) in \( S \).

(c) Prove that the function \( q_S \) is continuous.

We consider two alternative proofs that a set has the unique furthest point property if and only if it is a singleton.

(d) (i) Use Section 6.1, Exercise 10 (Max-functions) to show that the function \( r_S^2/2 \) has Clarke subdifferential the singleton \( \{ x - q_S(x) \} \) at any point \( x \in E \), and hence is everywhere differentiable.

(ii) Use Exercise 17 (Smoothness and biconjugacy) to deduce that the function \( g_S \) is convex, and hence that \( S \) is a singleton.

(e) Alternatively, suppose \( S \) is not a singleton. Denote the unique minimizer of the function \( r_S \) by \( y \). By investigating the continuity of the function \( q_S \) on the line segment \([y, q_S(y)]\), derive a contradiction without using part (d).

21. **(Chebyshev convexity via inversion)** The map \( \iota : E \setminus \{0\} \to E \) defined by \( \iota(x) = \| x \|^{-2} x \) is called the inversion in the unit sphere.

(a) If \( D \subset E \) is a ball with \( 0 \in \text{bd} D \), prove \( \iota(D \setminus \{0\}) \) is a halfspace disjoint from 0.
19. **(Chebyshev convexity via conjugacy)** For any nonempty closed set $S \subset E$, prove

\[
\left( \frac{\| \cdot \|^2 + \delta_S}{2} \right)^* = \frac{\| \cdot \|^2 - d_S^2}{2}
\]

Deduce, using Exercises 17 and 18, that Chebyshev sets are convex.

20. **(Unique furthest points)** Consider a set $S \subset E$, and define a function $r_S : E \to [-\infty, +\infty]$ by

\[
r_S(x) = \sup_{y \in S} \| x - y \|.
\]

Any point $y$ attaining the above supremum is called a furthest point in $S$ to the point $x \in E$.

(a) Prove that the function $r_2^2 / 2$ is the conjugate of the function $g_S = \delta_{-S} - \| \cdot \|^2 / 2$.

(b) Prove that the function $r_2^2$ is strictly convex on its domain.

Now suppose each point $x \in E$ has a unique nearest point $q_S(x)$ in $S$.

(c) Prove that the function $q_S$ is continuous.

We consider two alternative proofs that a set has the unique furthest point property if and only if it is a singleton.

(d) (i) Use Section 6.1, Exercise 10 (Max-functions) to show that the function $r_2^2 / 2$ has Clarke subdifferential the singleton $\{ x - q_S(x) \}$ at any point $x \in E$, and hence is everywhere differentiable.

(ii) Use Exercise 17 (Smoothness and biconjugacy) to deduce that the function $g_S$ is convex, and hence that $S$ is a singleton.

(e) Alternatively, suppose $S$ is not a singleton. Denote the unique minimizer of the function $r_S$ by $y$. By investigating the continuity of the function $q_S$ on the line segment $[y, q_S(y)]$, derive a contradiction without using part (d).

21. **(Chebyshev convexity via inversion)** The map $\iota : E \setminus \{0\} \to E$ defined by $\iota(x) = \| x \|^{-2} x$ is called the inversion in the unit sphere.

(a) If $D \subset E$ is a ball with $0 \in \text{bd} D$, prove $\iota(D \setminus \{0\})$ is a halfspace disjoint from $0$.  

Inverse Geometry for Hunters

Preserves circles (spheres, lines …)
(b) For any point $x \in \mathbf{E}$ and radius $\delta > \|x\|$, prove

$$\iota((x + \delta B) \setminus \{0\}) = \frac{1}{\delta^2 - \|x\|^2} \{y \in \mathbf{E} : \|y + x\| \geq \delta\}.$$ 

Prove that any Chebyshev set $C \subset \mathbf{E}$ must be convex as follows.

Without loss of generality, suppose $0 \notin C$ but $0 \in \text{cl}(\text{conv } C)$. Consider any point $x \in \mathbf{E}$.

(c) Prove the quantity 

$$\rho = \inf \{\delta > 0 | \iota C \subset x + \delta B\}$$

satisfies $\rho > \|x\|$.

(d) Let $z$ denote the unique nearest point in $C$ to the point 

$$\frac{-x}{\rho^2 - \|x\|^2}.$$ 

Use part (b) to prove that $\iota z$ is the unique furthest point in $\iota C$ to $x$.

(e) Use Exercise 20 to derive a contradiction.
The Chebyshev Problem in Infinite Dimensions is OPEN

In any Banach space (JMB & JV, CUP in press):

Corollary 3.14.2. Suppose \( f : X \to (-\infty, \infty] \) is such that \( f^{**} \) is proper.

(a) If \( f^* \) is Fréchet differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is lower semicontinuous, then \( f \) is convex.

(a) If \( f^* \) is Gâteaux differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is sequentially weakly lower semicontinuous, then \( f \) is convex.

Theorem 3.14.7. Let \( X \) be a Hilbert space and suppose \( C \) is a nonempty weakly closed subset of \( X \). Then the following are equivalent.

(i) \( C \) is convex.
(ii) \( C \) is a Chebyshev set.
(iii) \( d(\cdot, C)^2 \) is Fréchet differentiable.
(iv) \( d(\cdot, C)^2 \) is Gâteaux differentiable.
3.14 Chebyshev Sets

Given a bornology $\beta$ on $X^*$, again we use $\tau_\beta$ to denote the topology on $X^{**}$ of uniform convergence on $\beta$-sets. We will say $f : X \to (-\infty, \infty]$ is sequentially $\tau_\beta$-lower semicontinuous if for every sequence $(x_n) \subset X \subset X^{**}$, and $x \in X$, $\liminf f(x_n) \geq f(x)$ whenever $x_n \rightharpoonup x$. Notice that the $\tau_\beta$-topology restricted to $X$ is at least as strong as the weak topology on $X$.

**Theorem 3.14.1.** Suppose $f : X \to (-\infty, \infty]$ is such that $f^{**}$ is proper. If $f^*$ is $\beta$-differentiable at all $x^* \in \text{dom}(\partial f^*)$ and $f$ is sequentially $\tau_\beta$-lower semicontinuous, then $f$ is convex.

**Proof.** First, $f^{**} \leq f$, so it suffices to show that $f(x) \leq f^{**}(x)$ for all $x \in X$. For notational purposes, let $\hat{f} = f^{**}|_X$. If $x \not\in \text{dom} \hat{f}$ then $\hat{f}(x) = \infty$, so $\hat{f}(x) \geq f(x)$. So let $x \in \text{dom} \hat{f}$. We first handle the case $x \in \text{dom}(\partial \hat{f})$. Indeed, for such $x$, let $\phi \in \partial \hat{f}(x)$. Then $x \in \partial f^*(\phi)$. Also, $f^*$ is lower semicontinuous everywhere and $\beta$-differentiable at $\phi$ and so $f^*$ is continuous at $\phi$. Now choose $x_n \in X$ such that

$$\phi(x_n) - f(x_n) \geq f^*(\phi) - \epsilon_n \quad \text{where } \epsilon_n \to 0^+.$$ 

Then $\phi(x_n) - \hat{f}(x_n) \geq f^*(\phi) - \epsilon_n$ and so $x_n \in \partial f^*(\phi)$ for all $n$. According to Šmulyan’s criterion $x_n \rightharpoonup_{\tau_\beta} x$. In particular, $\phi(x_n) \to \phi(x)$. Therefore,

$$\hat{f}(x) = \phi(x) - f^*(\phi) = \lim_{n \to \infty} \phi(x) - [\phi(x_n) - f(x_n)] = \lim_{n \to \infty} f(x_n).$$

Now $f$ is sequentially $\tau_\beta$-lower semicontinuous, and so $\liminf_{n \to \infty} f(x_n) \geq f(x)$. Therefore, $\hat{f}(x) \geq f(x)$ when $x \in \text{dom}(\partial \hat{f})$.

Now suppose $x \in \text{dom} \hat{f} \setminus \text{dom}(\partial \hat{f})$. According to Borwein’s generalization of Brøsted-Rockafellar there exists a sequence $x_n \to x$ such that $x_n \in \text{dom}(\partial \hat{f})$ and $|\hat{f}(x_n) - f(x)| \to 0$. Consequently,

$$f(x) \leq \liminf f(x_n) = \liminf \hat{f}(x_n) = \hat{f}(x).$$

Thus, for any $x \in X$, $f(x) \leq \hat{f}(x)$, and so $f(x)$ is convex. \qed

The important special cases of this are recorded as follows.

**Corollary 3.14.2.** Suppose $f : X \to (-\infty, \infty]$ is such that $f^{**}$ is proper.
3.14 Chebyshev Sets

Given a bornology $\beta$ on $X^*$, again we use $\tau_\beta$ to denote the topology on $X^{**}$ of uniform convergence on $\beta$-sets. We will say $f : X \to (-\infty, \infty]$ is sequentially $\tau_\beta$-lower semicontinuous if for every sequence $(x_n) \subset X \subset X^{**}$, and $x \in X$, $\liminf f(x_n) \geq f(x)$ whenever $x_n \to_{\tau_\beta} x$. Notice, that the $\tau_\beta$-topology restricted to $X$ is at least as strong as the weak topology on $X$.

**Theorem 3.14.1.** Suppose $f : X \to (-\infty, \infty]$ is such that $f^{**}$ is proper. If $f^*$ is $\beta$-differentiable at all $x^* \in \text{dom}(\partial f^*)$ and $f$ is sequentially $\tau_\beta$-lower semicontinuous, then $f$ is convex.

**Proof.** First, $f^{**} \leq f$, so it suffices to show that $f(x) \leq f^{**}(x)$ for all $x \in X$. For notational purposes, let $\hat{f} = f^{**}|_X$. If $x \notin \text{dom} \hat{f}$ then $\hat{f}(x) = \infty$, so $\hat{f}(x) \geq f(x)$. So let $x \in \text{dom} \hat{f}$. We first handle the case $x \in \text{dom}(\partial \hat{f})$. Indeed, for such $x$, let $\phi \in \partial \hat{f}(x)$. Then $x \in \partial f^*(\phi)$. Also, $f^*$ is lower semicontinuous everywhere and $\beta$-differentiable at $\phi$ and so $f^*$ is continuous at $\phi$. Now choose $x_n \in X$ such that

$$\phi(x_n) - f(x_n) \geq f^*(\phi) - \varepsilon_n \text{ where } \varepsilon_n \to 0^+.$$

Then $\phi(x_n) - \hat{f}(x_n) \geq f^*(\phi) - \varepsilon_n$ and so $x_n \in \partial f^*(\phi)$ for all $n$. According to Smulyan's criterion $x_n \to_{\tau_\beta} x$. In particular, $\phi(x_n) \to \phi(x)$. Therefore,

$$\hat{f}(x) = \phi(x) - f^*(\phi) = \lim_{n \to \infty} \phi(x) - [\phi(x_n) - f(x_n)] = \lim_{n \to \infty} f(x_n).$$

Now $f$ is sequentially $\tau_\beta$-lower semicontinuous, and so $\liminf f(x_n) \geq f(x)$. Therefore, $\hat{f}(x) \geq f(x)$ when $x \in \text{dom}(\partial \hat{f})$.

Now suppose $x \in \text{dom} \hat{f} \setminus \text{dom}(\partial \hat{f})$. According to Borwein's generalization of Brøsted-Rockafellar there exists a sequence $x_n \to x$ such that $x_n \in \text{dom}(\partial \hat{f})$ and $|\hat{f}(x_n) - \hat{f}(x)| \to 0$. Consequently,

$$f(x) \leq \liminf f(x_n) = \liminf \hat{f}(x_n) = \hat{f}(x).$$

Thus, for any $x \in X$, $f(x) \leq \hat{f}(x)$, and so $f(x)$ is convex. \qed

The important special cases of this are recorded as follows.

**Corollary 3.14.2.** Suppose $f : X \to (-\infty, \infty]$ is such that $f^{***}$ is proper.
3.14. CHEBYSHEV SETS

(a) If \( f^* \) is Fréchet differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is lower semicontinuous, then \( f \) is convex.

(a) If \( f^* \) is Gâteaux differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is sequentially weakly lower semicontinuous, then \( f \) is convex.

Fact 3.14.3. Let \( C \) be a closed nonempty subset of the Hilbert space \( X \). Let \( f = \frac{1}{2} \| \cdot \|^2 + \delta_C \). Then \( d^2_C = \| \cdot \|^2 - 2f^* \).

Proof. Given \( f \) as defined, we compute

\[
f^*(y) = \sup \left\{ \langle x, y \rangle - \frac{1}{2} \| x \|^2 : x \in C \right\}
\]

\[
= \sup \left\{ \langle x, y \rangle - \frac{1}{2} \langle x, x \rangle : x \in C \right\}
\]

\[
= \sup \left\{ \frac{1}{2} \langle y, y \rangle + \langle x, y \rangle - \frac{1}{2} \langle y, y \rangle - \frac{1}{2} \langle x, x \rangle : x \in C \right\}
\]

\[
= \frac{1}{2} \langle y, y \rangle + \sup \left\{ -\frac{1}{2} \langle x, x \rangle + \langle x, y \rangle - \frac{1}{2} \langle y, y \rangle : x \in C \right\}
\]

\[
= \frac{1}{2} \langle y, y \rangle - \frac{1}{2} \inf \{ \langle x, x \rangle - 2 \langle x, y \rangle + \langle y, y \rangle : x \in C \}
\]

\[
= \frac{1}{2} \langle y, y \rangle - \frac{1}{2} d^2_C(y).
\]

Therefore,

\[
d^2_C = \| \cdot \|^2 - 2f^* \tag{3.14.26}
\]

as desired.

Theorem 3.14.4. Let \( X \) be a Hilbert space and let \( C \) be a nonempty closed subset of \( X \). Then the following are equivalent.

(a) \( C \) is convex.

(b) \( d^2_C \) is Fréchet differentiable.

(c) \( d^2_C \) is Gâteaux differentiable.

Proof. (a) \( \Rightarrow \) (c): If \( C \) is convex, then \( d^2_C \) is even Fréchet differentiable by Exercise 3.11.5.

(c) \( \Rightarrow \) (b): Let \( f = \frac{1}{2} \| \cdot \|^2 + \delta_C \). Then \( f^* = (\| \cdot \|^2 - d^2_C) / 2 \) and so \( f^* \) is Gâteaux differentiable. Thus the derivatives of \( f^* \) and \( \| \cdot \|^2 \) are both norm to weak continuous. Consequently the derivative of \( d^2_C \) is norm to weak continuous. Moreover, the derivative of \( d^2_C \) is maximal norm, so by
(a) If \( f^* \) is Fréchet differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is lower semicontinuous, then \( f \) is convex.

(a) If \( f^* \) is Gâteaux differentiable at all \( x^* \in \text{dom}(\partial f^*) \) and \( f \) is sequentially weakly lower semicontinuous, then \( f \) is convex.

**Fact 3.14.3.** Let \( C \) be a closed nonempty subset of the Hilbert space \( X \). Let \( f = \frac{1}{2} \| \cdot \|^2 + \delta_C \). Then \( d_C^2 = \| \cdot \|^2 - 2f^* \).

**Proof.** Given \( f \) as defined, we compute

\[
    f^*(y) = \sup \left\{ \langle x, y \rangle - \frac{1}{2} \| x \|^2 : x \in C \right\}
    = \sup \left\{ \langle x, y \rangle - \frac{1}{2} \langle x, x \rangle : x \in C \right\}
    = \sup \left\{ \frac{1}{2} \langle y, y \rangle + \langle x, y \rangle - \frac{1}{2} \langle y, y \rangle - \frac{1}{2} \langle x, x \rangle : x \in C \right\}
    = \frac{1}{2} \langle y, y \rangle + \sup \left\{ -\frac{1}{2} \langle x, x \rangle + \langle x, y \rangle - \frac{1}{2} \langle y, y \rangle : x \in C \right\}
    = \frac{1}{2} \langle y, y \rangle - \frac{1}{2} \inf \left\{ \langle x, x \rangle - 2\langle x, y \rangle + \langle y, y \rangle : x \in C \right\}
    = \frac{1}{2} \langle y, y \rangle - \frac{1}{2} d_C^2(y).
\]

Therefore,

\[
d_C^2 = \| \cdot \|^2 - 2f^* \tag{3.14.26}
\]

as desired. \( \Box \)

**Theorem 3.14.4.** Let \( X \) be a Hilbert space and let \( C \) be a nonempty closed subset of \( X \). Then the following are equivalent.

(a) \( C \) is convex.

(b) \( d_C^2 \) is Fréchet differentiable.

(c) \( d_C^2 \) is Gâteaux differentiable.

**Proof.** (a) \( \Rightarrow \) (c): If \( C \) is convex, then \( d_C^2 \) is even Fréchet differentiable by Exercise 3.11.5.

(c) \( \Rightarrow \) (b): Let \( f = \frac{1}{2} \| \cdot \|^2 + \delta_C \). Then \( f^* = (\| \cdot \|^2 - d_C^2)/2 \) and so \( f^* \) is Gâteaux differentiable. Thus the derivatives of \( f^* \) and \( \| \cdot \|^2 \) are both norm to weak continuous. Consequently the derivative of \( d_C^2 \) is norm to weak continuous. Moreover, the derivative of \( d_C^2 \) is maximal norm, so by
the Kadec-Klee property of the Hilbert norm, the derivative of $d_C^2$ is norm to norm continuous. Consequently, so is the derivative of $f^*$, and so $f^*$ is Fréchet differentiable. Therefore, $d_C^2$ is also Fréchet differentiable.

(b) ⇒ (a): Suppose $d_C^2$ is Fréchet differentiable. With $f$ as in the previous part, we conclude that $f^*$ is Fréchet differentiable. By Corollary 3.14.2(a), we know $f$ is convex. Therefore, $C = \text{dom } f$ is convex.  

Application: Chebyshev sets
Let $C$ be a nonempty subset of a normed space. We define the nearest point mapping by

$$P_C(x) = \{v \in C : \|v - x\| = d(x, C)\}.$$ 

A set $C$ is said to be a Chebyshev set if $P_C(x)$ is a singleton for every $x \in X$.

**Proposition 3.14.5.** Let $C$ be a nonempty closed subset of $X$.

(i) If $X$ is reflexive, then $P_C(x) \neq \emptyset$ for each $x \in X$.

(ii) If $X$ is strictly convex, then $P_C(x)$ is either empty or a singleton for each $x \in X$.

Proof. (i) Let $x \in X$, and define $f$ by $f(u) = \|u - x\| + \delta_C$ for $u \in C$. Then $f$ is a lower semicontinuous coercive convex function so it attains its minimum on $C$ (add general exercise!).

(ii) Another exercise.

**Proposition 3.14.6.** Suppose $X$ is a reflexive Banach space, and let $C$ be a weakly closed Chebyshev subset of $X$. Then $x \rightarrow P_C(x)$ is norm to weak continuous. If, moreover, the norm on $X$ has the Kadec-Klee property, then $x \rightarrow P_C(x)$ is norm to norm continuous.

Proof. Let $x \in X$ and suppose $x_n \rightarrow x$. Then $\|x_n - P_C(x_n)\| \rightarrow \|x - P_C(x)\|$. According to the Eberlein-Smulian theorem, there is a subsequence $(P_C(x_{n_k}))$ that converges weakly to some $\bar{x} \in X$. Then $\bar{x} \in C$ because $C$ is weakly closed. Now

$$\|x - P_C(x)\| = \lim_{k \rightarrow \infty} \|x_{n_k} - P_C(x_{n_k})\| \geq \|x - \bar{x}\|.$$ 

Because $\bar{x} \in C$, we conclude that $\bar{x} = P_C(x)$. Thus $P_C(x_{n_k}) \rightharpoonup P_C(x)$. Standard arguments now imply $x \rightarrow P_C(x)$ is norm to weak continuous.

If the norm on $X$ has the Kadec-Klee property, then in the above, we have $x_{n_k} - P_C(x_{n_k}) \rightarrow_w x - P_C(x)$, and hence $x_{n_k} - P_C(x_{n_k}) \rightarrow x - P_C(x)$ which implies $P_C(x_{n_k}) \rightarrow P_C(x)$ from which we deduce the norm to norm continuity.
the Kadec-Klee property of the Hilbert norm, the derivative of $d_C^2$ is norm to norm continuous. Consequently, so is the derivative of $f^*$, and so $f^*$ is Fréchet differentiable. Therefore, $d_C^2$ is also Fréchet differentiable.

(b) $\Rightarrow$ (a): Suppose $d_C^2$ is Fréchet differentiable. With $f$ as in the previous part, we conclude that $f^*$ is Fréchet differentiable. By Corollary 3.14.2(a), we know $f$ is convex. Therefore, $C = \text{dom } f$ is convex. \hfill $\square$

Application: Chebyshev sets

Let $C$ be a nonempty subset of a normed space. We define the nearest point mapping by

$$P_C(x) = \{ v \in C : \| v - x \| = d(x, C) \}.$$  

A set $C$ is said to be a Chebyshev set if $P_C(x)$ is a singleton for every $x \in X$.

**Proposition 3.14.5.** Let $C$ be a nonempty closed subset of $X$.

(i) If $X$ is reflexive, then $P_C(x) \neq \emptyset$ for each $x \in X$.

(ii) If $X$ is strictly convex, then $P_C(x)$ is either empty or a singleton for each $x \in X$.

**Proof.** (i) Let $x \in X$, and define $f$ by $f(u) = \| u - x \| + \delta_C$ for $u \in C$. Then $f$ is a lower semicontinuous coercive convex function so it attains its minimum on $C$ (add general exercise!).

(ii) Another exercise. \hfill $\square$

**Proposition 3.14.6.** Suppose $X$ is a reflexive Banach space, and let $C$ be a weakly closed Chebyshev subset of $X$. Then $x \to P_C(x)$ is norm to weak continuous. If, moreover, the norm on $X$ has the Kadec-Klee property, then $x \to P_C(x)$ is norm to norm continuous.

**Proof.** Let $x \in X$ and suppose $x_n \to x$. Then $\| x_n - P_C(x_n) \| \to \| x - P_C(x) \|$. According to the Eberlein-Smulian theorem, there is a subsequence $(P_C(x_{n_k}))$ that converges weakly to some $\bar{x} \in X$. Then $\bar{x} \in C$ because $C$ is weakly closed. Now

$$\| x - P_C(x) \| = \lim_{k \to \infty} \| x_{n_k} - P_C(x_{n_k}) \| \geq \| x - \bar{x} \|.$$  

Because $\bar{x} \in C$, we conclude that $\bar{x} = P_C(x)$. Thus $P_C(x_{n_k}) \to_w P_C(x)$. Standard arguments now imply $x \to P_C(x)$ is norm to weak continuous.

If the norm on $X$ has the Kadec-Klee property, then in the above, we have $x_{n_k} - P_C(x_{n_k}) \to_w x - P_C(x)$, and hence $x_{n_k} - P_C(x_{n_k}) \to x - P_C(x)$ which implies $P_C(x_{n_k}) \to P_C(x)$ from which we deduce the norm to norm continuous.
Theorem 3.14.7. Let $X$ be a Hilbert space and suppose $C$ is a nonempty weakly closed subset of $X$. Then the following are equivalent.

(i) $C$ is convex.

(ii) $C$ is a Chebyshev set.

(iii) $d(\cdot, C)^2$ is Fréchet differentiable.

(iv) $d(\cdot, C)^2$ is Gâteaux differentiable.

Proof. The implication (i) $\Rightarrow$ (ii) follows from Proposition 3.14.5(b). We next prove (ii) $\Rightarrow$ (iii). For this, again consider $f = \frac{1}{2} \| \cdot \|^2 + \delta_C$. We first show that

$$
\partial f^*(x) = \{ PC(x) \}, \text{ for all } x \in X. \tag{3.14.27}
$$

Indeed, for $x \in X$, (3.14.26) implies

$$
f^*(x) = \frac{1}{2} \| x \|^2 - \frac{1}{2} \| x - PC(x) \|^2 = \langle x, PC(x) \rangle - \frac{1}{2} \| PC(x) \|^2
$$

$$
= \langle x, PC(x) \rangle - f(PC(x)).
$$

Consequently, $PC(x) \in \partial f^*(x)$ for $x \in X$. Now suppose $y \in \partial f^*(x)$, and define $x_n = x + \frac{1}{n}(y - PC(x))$. Then $x_n \to x$, and hence $PC(x_n) \to PC(x)$ by Proposition 3.14.6. Using the subdifferential inequality, we have

$$
0 \leq \langle x_n - x, PC(x_n) - y \rangle = \frac{1}{n} \langle y - PC(x), PC(x_n) - y \rangle.
$$

This now implies:

$$
0 \leq \lim_{n \to \infty} \langle y - PC(x), PC(x_n) - y \rangle = -\| y - PC(x) \|^2.
$$

Consequently, $y = PC(x)$ and so (3.14.27) is established. Now $f^*$ is continuous, and Proposition 3.14.6 the ensures that the mapping $x \to PC(x)$ is norm to norm continuous. Consequently, Šmulyan's criterion implies that $f^*$ and $d(\cdot, C)^2$ are Fréchet differentiable.

The implication (iii) $\Rightarrow$ (iv) is trivial. Finally, (iv) $\Rightarrow$ (i) follows from Corollary 3.14.2(b) because $f$ is weakly lower semicontinuous (use the fact that $C$ is weakly closed). $\square$
Theorem 3.14.7. Let $X$ be a Hilbert space and suppose $C$ is a nonempty weakly closed subset of $X$. Then the following are equivalent.

(i) $C$ is convex.

(ii) $C$ is a Chebyshev set.

(iii) $d(\cdot, C)^2$ is Fréchet differentiable.

(iv) $d(\cdot, C)^2$ is Gâteaux differentiable.

Proof. The implication (i) $\Rightarrow$ (ii) follows from Proposition 3.14.5(b). We next prove (ii) $\Rightarrow$ (iii). For this, again consider $f = \frac{1}{2} \| \cdot \|^2 + \delta_C$. We first show that

$$\partial f^*(x) = \{ P_C(x) \}, \text{ for all } x \in X. \quad (3.14.27)$$

Indeed, for $x \in X$, (3.14.26) implies

$$f^*(x) = \frac{1}{2} \| x \|^2 - \frac{1}{2} \| x - P_C(x) \|^2 = \langle x, P_C(x) \rangle - \frac{1}{2} \| P_C(x) \|^2$$

$$= \langle x, P_C(x) \rangle - f(P_C(x)).$$

Consequently, $P_C(x) \in \partial f^*(x)$ for $x \in X$. Now suppose $y \in \partial f^*(x)$, and define $x_n = x + \frac{1}{n} (y - P_C(x))$. Then $x_n \to x$, and hence $P_C(x_n) \to P_C(x)$ by Proposition 3.14.6. Using the subdifferential inequality, we have

$$0 \leq \langle x_n - x, P_C(x_n) - y \rangle = \frac{1}{n} \langle y - P_C(x), P_C(x_n) - y \rangle.$$

This now implies:

$$0 \leq \lim_{n \to \infty} \langle y - P_C(x), P_C(x_n) - y \rangle = - \| y - P_C(x) \|^2.$$

Consequently, $y = P_C(x)$ and so (3.14.27) is established. Now $f^*$ is continuous, and Proposition 3.14.6 the ensures that the mapping $x \to P_C(x)$ is norm to norm continuous. Consequently, Šmulyan's criterion implies that $f^*$ and $d(\cdot, C)^2$ are Fréchet differentiable.

The implication (iii) $\Rightarrow$ (iv) is trivial. Finally, (iv) $\Rightarrow$ (i) follows from Corollary 3.14.2(b) because $f$ is weakly lower semicontinuous (use the fact that $C$ is weakly closed). \qed
"We have reason to believe Bingleman is an irrational number himself."
A norm is Kadec-Klee (sequentially) if the weak and norm topologies coincide (sequentially) on the boundary of the unit ball, as in Hilbert space.

**Theorem 10** Let $C$ be a closed subset of a reflexive Banach space $X$ with a Kadec-Klee norm.

(a) (Density) The set of points in $X$ at which every minimizing sequence clusters to a best approximation is dense in $X$.

(b) (Projection) If in addition, the original norm is Fréchet then

$$
\partial_{Fd} C(x) \subset \partial_{Fd} C(P_C(x))
$$

where $P_C(x)$ is the (set of) best approximations of $x$ on $C$.

(c) In particular, in any Fréchet LUR norm on a reflexive space, this holds for all sets in the Fréchet sense with a single-valued metric projection.
Proof. (a) We may assume $x_n \to_w p$ and at any of the dense set of points with

$$\phi \in \partial_F d_C(x) \neq \emptyset$$

all minimizing sequences actually converge in norm to $p$ since

$$\phi(x_n - x) \to d_C(x) \Rightarrow \|x_n - x\| \to \|p - x\|,$$

and by Kadec-Klee $x_n \to p$, and $p = P_C(x)$.

The Fréchet slice forces
the approximating sequence to line up

The corresponding subgradient is a proximal normal to $C$ at $p$.  

35
Finally, when the norm is $F$-smooth, simple derivative estimates show that any member of $\partial_F d_C(x)$ must lie in

$$\partial_F d_C(P_C(x)).$$

✓ This used to be hard.

- **(Lau-Konjagin (1976-86))** $X$ is reflexive and Kadec-Klee iff best approximations always exist densely (or generically).

- Theorem 10 easily shows the normal cone defined in terms of distance functions is always contained in the normal cone defined in terms of indicator functions.

- In Hilbert space we may conclude

$$\partial_F d_C(x) \subset \partial_\pi d_C(P_C(x)),$$

where $\partial_\pi$ denotes the set of proximal subgradients.
Two Open Questions

• Every closed set in every reflexive space (every renorm of Hilbert space) admits at least one best approximation.

(Stronger variant.) For every closed set of every reflexive space the proximal normal points are norm dense in the norm boundary.

✓ Any counter-example is necessarily unbounded (and fractal-like)

• Every norm closed set in a reflexive Banach space with unique best approximations for every point in A (a Chebyshev set) is convex.

[True in weak topology, and so in $R^N$.]
"...and, as you go out into the world, I predict that you will, gradually and imperceptibly, forget all you ever learned at this university."