Now

$$\lim_{t\to 0} Q(x, v_k, t) = 0 \qquad (k = 1, \dots, N),$$

and thus there exists $\delta > 0$ so that

$$|Q(x, v_k, t)| < \frac{\epsilon}{2} \text{ for all } 0 < |t| < \delta,\ k = 1, \dots, N. \qquad (\star\star\star)$$

Consequently, for each $v \in \partial B(0,1)$, there exists $k \in \{1, \dots, N\}$ such that

$$|Q(x, v, t)| \le |Q(x, v_k, t)| + |Q(x, v, t) - Q(x, v_k, t)| < \epsilon$$

if $0 < |t| < \delta$, according to ($\star$) through ($\star\star\star$). Note the same $\delta > 0$ works for all $v \in \partial B(0,1)$.

Now choose any $y \in \mathbb{R}^n$, $y \ne x$. Write $v \equiv (y-x)/|y-x|$, so that $y = x+tv$, $t \equiv |x-y|$. Then

$$\begin{aligned} f(y) - f(x) - \text{grad } f(x) \cdot (y - x) &= f(x+tv) - f(x) - tv \cdot \text{grad } f(x) \\ &= o(t) \\ &= o(|x-y|), \text{ as } y \to x. \end{aligned}$$

Hence f is differentiable at x, with

$$Df(x) = \text{grad } f(x). \quad ∎$$

REMARK See Theorem 2 in Section 6.2 for another proof of Rademacher's Theorem and Theorem 1 in Section 6.2 for a generalization. In Section 6.4 we prove Aleksandrov's Theorem, stating that a convex function is twice differentiable a.e. ∎

We next record a technical lemma for use later.

COROLLARY 1

(i) *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz, and*

$$Z \equiv \{x \in \mathbb{R}^n \mid f(x) = 0\}.$$

Then $Df(x) = 0$ for $\mathcal{L}^n$ a.e. $x \in Z$.

(ii) *Let $f, g : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz, and*

$$Y \equiv \{x \in \mathbb{R}^n \mid g(f(x)) = x\}.$$

Then

$$Dg(f(x))Df(x) = I \quad \textit{for} \quad \mathcal{L}^n \textit{ a.e. } x \in Y.$$

3.1.2 Rademacher's Theorem

We next prove Rademacher's remarkable theorem that a Lipschitz function is differentiable $\mathcal{L}^n$ a.e. This is surprising since the inequality

$$|f(x) - f(y)| \le \text{Lip}(f)|x - y|$$

apparently says nothing about the possibility of locally approximating f by a linear map.

DEFINITION *The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** at $x \in \mathbb{R}^n$ if there exists a linear mapping*

$$L : \mathbb{R}^n \to \mathbb{R}^m$$

such that

$$\lim_{y \to x} \frac{|f(y) - f(x) - L(y - x)|}{|x - y|} = 0,$$

or, equivalently,

$$f(y) = f(x) + L(y - x) + o(|y - x|) \quad \textit{as} \quad y \to x.$$

NOTATION If such a linear mapping L exists, it is clearly unique, and we write

$$Df(x)$$

for L. We call $Df(x)$ the *derivative* of f at x.

THEOREM 2 RADEMACHER'S THEOREM
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function. Then f is differentiable $\mathcal{L}^n$ a.e.

PROOF

1. We may assume $m = 1$. Since differentiability is a local property, we may as well also suppose f is Lipschitz.

2. Fix any $v \in \mathbb{R}^n$ with $|v| = 1$, and define

$$D_v f(x) \equiv \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} \qquad (x \in \mathbb{R}^n),$$

provided this limit exists.

3. *Claim #1*: $D_v f(x)$ exists for $\mathcal{L}^n$ a.e. x.

Proof of Claim #1: Since f is continuous,

$$\overline{D}_v f(x) \equiv \limsup_{t\to 0} \frac{f(x+tv)-f(x)}{t} = \lim_{k\to\infty} \sup_{\substack{0<|t|<1/k \\ t \text{ rational}}} \frac{f(x+tv)-f(x)}{t}$$

is Borel measurable, as is

$$\underline{D}_v f(x) \equiv \liminf_{t\to 0} \frac{f(x+tv)-f(x)}{t}.$$

Thus

$$A_v \equiv \{x \in \mathbb{R}^n \mid D_v f(x) \text{ does not exist}\} = \{x \in \mathbb{R}^n \mid \underline{D}_v f(x) < \overline{D}_v f(x)\}$$

is Borel measurable.

Now, for each $x, v \in \mathbb{R}^n$, with $|v| = 1$, define $\varphi : \mathbb{R} \to \mathbb{R}$ by

$$\varphi(t) \equiv f(x+tv) \qquad (t \in \mathbb{R}).$$

Then φ is Lipschitz, thus absolutely continuous, and thus differentiable $\mathcal{L}^1$ a.e. Hence

$$\mathcal{H}^1(A_v \cap L) = 0$$

for each line L parallel to v. Fubini's Theorem then implies

$$\mathcal{L}^n(A_v) = 0.$$

4. As a consequence of Claim #1, we see

$$\text{grad } f(x) \equiv \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right)$$

exists for $\mathcal{L}^n$ a.e. x.

5. *Claim #2*: $D_v f(x) = v \cdot \text{grad } f(x)$ for $\mathcal{L}^n$ a.e. x.

Proof of Claim #2: Let $\zeta \in C_c^\infty(\mathbb{R}^n)$. Then

$$\int_{\mathbb{R}^n} \left[\frac{f(x+tv)-f(x)}{t}\right] \zeta(x)\, dx = -\int_{\mathbb{R}^n} f(x) \left[\frac{\zeta(x)-\zeta(x-tv)}{t}\right] dx.$$

Let $t = 1/k$ for $k = 1, \ldots$ in the above equality and note

$$\left|\frac{f(x+\frac{1}{k}v)-f(x)}{\frac{1}{k}}\right| \le \text{Lip}\,(f)|v| = \text{Lip}\,(f).$$

Thus the Dominated Convergence Theorem implies

$$\begin{aligned}\int_{\mathbb{R}^n} D_v f(x)\zeta(x)\, dx &= -\int_{\mathbb{R}^n} f(x) D_v\zeta(x)\, dx \\ &= -\sum_{i=1}^n v_i \int_{\mathbb{R}^n} f(x) \frac{\partial \zeta}{\partial x_i}(x)\, dx \\ &= \sum_{i=1}^n v_i \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(x)\zeta(x)\, dx \\ &= \int_{\mathbb{R}^n} (v \cdot \text{grad } f(x))\zeta(x)\, dx,\end{aligned}$$

where we used Fubini's Theorem and the absolute continuity of f on lines. The above equality holding for each $\zeta \in C_c(\mathbb{R}^n)$ implies $D_v f = v \cdot \text{grad } f$ $\mathcal{L}^n$ a.e.

6. Now choose $\{v_k\}_{k=1}^\infty$ to be a countable, dense subset of $\partial B(0,1)$. Set

$$A_k \equiv \{x \in \mathbb{R}^n \mid D_{v_k} f(x),\ \text{grad } f(x) \text{ exist and } D_{v_k} f(x) = v_k \cdot \text{grad } f(x)\}$$

for $k = 1, 2, \ldots$, and define

$$A \equiv \bigcap_{k=1}^\infty A_k.$$

Observe

$$\mathcal{L}^n(\mathbb{R}^n - A) = 0.$$

7. *Claim #3*: f is differentiable at each point $x \in A$.

Proof of Claim #3: Fix any $x \in A$. Choose $v \in \partial B(0,1)$, $t \in \mathbb{R}$, $t \neq 0$, and write

$$Q(x,v,t) \equiv \frac{f(x+tv)-f(x)}{t} - v \cdot \text{grad } f(x).$$

Then if $v' \in \partial B(0,1)$, we have

$$\begin{aligned}|Q(x,v,t) - Q(x,v',t)| &\le \left|\frac{f(x+tv)-f(x+tv')}{t}\right| + |(v-v') \cdot \text{grad } f(x)| \\ &\le \text{Lip}\,(f)|v-v'| + |\text{grad } f(x)|\,|v-v'| \\ &\le (\sqrt{n}+1)\text{Lip}\,(f)|v-v'|. \qquad (\star)\end{aligned}$$

Now fix $\epsilon > 0$, and choose N so large that if $v \in \partial B(0,1)$, then

$$|v - v_k| \le \frac{\epsilon}{2(\sqrt{n}+1)\text{Lip}\,(f)} \quad \text{for some} \quad k \in \{1, \ldots, N\}. \qquad (\star\star)$$

THEOREM 1 ALEKSANDROV'S THEOREM
Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex. Then f has a second derivative $\mathcal{L}^n$ a.e. More precisely, for $\mathcal{L}^n$ a.e. x,

$$\left| f(y) - f(x) - Df(x)\cdot(y-x) - \frac{1}{2}(y-x)^T \cdot D^2 f(x)\cdot(y-x)\right| = o(|y-x|^2) \quad \text{as } y \to x. \qquad (\star)$$

PROOF
1. $\mathcal{L}^n$ a.e. point x satisfies these conditions:

(a) $Df(x)$ exists and $\lim_{r\to 0} ⨍_{B(x,r)} |Df(y) - Df(x)|\, dy = 0.$

(b) $\lim_{r\to 0} ⨍_{B(x,r)} |D^2 f(y) - D^2 f(x)|\, dy = 0.$ $(\star\star)$

(c) $\lim_{r\to 0} |[D^2 f]_s|\,(B(x,r))/r^n = 0.$

2. Fix such a point x; we may as well assume $x = 0$. Choose $r > 0$ and let $f^\epsilon \equiv \eta_\epsilon * f$. Fix $y \in B(r)$. By Taylor's Theorem,

$$f^\epsilon(y) = f^\epsilon(0) + Df^\epsilon(0)\cdot y + \int_0^1 (1-s) y^T \cdot D^2 f^\epsilon(sy)\cdot y\, ds.$$

Add and subtract $(1/2)y^T \cdot D^2 f(0)\cdot y$:

$$f^\epsilon(y) = f^\epsilon(0) + Df^\epsilon(0)\cdot y + \frac{1}{2} y^T\cdot D^2 f(0)\cdot y + \int_0^1 (1-s)y^T\cdot [D^2 f^\epsilon(sy) - D^2 f(0)]\cdot y\, ds.$$

3. Fix any function $\varphi \in C_c^2(B(r))$ with $|\varphi| \le 1$, multiply the equation above by φ, and average over $B(r)$:

$$⨍_{B(r)} \varphi(y)(f^\epsilon(y) - f^\epsilon(0) - Df^\epsilon(0)\cdot y - \frac{1}{2}y^T\cdot D^2 f(0)\cdot y)\, dy$$
$$= \int_0^1 (1-s)\left(⨍_{B(r)} \varphi(y) y^T\cdot[D^2 f^\epsilon(sy) - D^2 f(0)]\cdot y\, dy\right) ds$$
$$= \int_0^1 \frac{(1-s)}{s^2}\left(⨍_{B(rs)} \varphi(\frac{z}{s}) z^T\cdot [D^2 f^\epsilon(z) - D^2 f(0)]\cdot z\, dz\right) ds. \quad (\star\star\star)$$

Now

$$g_\epsilon(s) \equiv \int_{B(rs)} \varphi\left(\frac{z}{s}\right) z^T\cdot D^2 f^\epsilon(z)\cdot z\, dz$$
$$= \int_{B(rs)} f^\epsilon(z) \sum_{i,j=1}^n \frac{\partial^2}{\partial z_i \partial z_j}\left(\varphi\left(\frac{z}{s}\right) z_i z_j\right) dz$$
$$\to \int_{B(rs)} f(z) \sum_{i,j=1}^n \frac{\partial^2}{\partial z_i \partial z_j}\left(\varphi\left(\frac{z}{s}\right) z_i z_j\right) dz \qquad \text{as } \epsilon \to 0$$
$$= \sum_{i,j=1}^n \int_{B(rs)} \varphi\left(\frac{z}{s}\right) z_i z_j\, d\mu^{ij}$$
$$= \int_{B(rs)} \varphi\left(\frac{z}{s}\right) z^T\cdot D^2 f(z)\cdot z\, dz + \sum_{i,j=1}^n \int_{B(rs)} \varphi\left(\frac{z}{s}\right) z_i z_j\, d\mu_s^{ij}.$$

Furthermore, as in Section 6.1.1, we may calculate

$$\frac{|g_\epsilon(s)|}{s^{n+2}} \le \frac{r^2}{s^n}\int_{B(rs)} |D^2 f^\epsilon(z)|\, dz$$
$$= \frac{r^2}{s^n}\int_{B(rs)} \left|\int_{\mathbb{R}^n} D^2\eta_\epsilon(z-y) f(y)\, dy\right| dz$$
$$\le \frac{r^2}{s^n}\int_{B(rs)} \left|\int_{\mathbb{R}^n} \eta_\epsilon(z-y)\, d[D^2 f]\right| dz$$
$$\le \frac{C}{s^n\epsilon^n}\int_{B(rs+\epsilon)} \left(\int_{B(rs)\cap B(y,\epsilon)} dz\right) d\|D^2 f\|$$
$$\le C\frac{\min((rs)^n, \epsilon^n)}{s^n \epsilon^n}\|D^2 f\|(B(rs+\epsilon))$$
$$\le C\frac{\min((rs)^n, \epsilon^n)(rs+\epsilon)^n}{s^n\epsilon^n}$$
$$\le C \qquad \text{for } 0 < \epsilon, s \le 1 \text{ by } (\star\star).$$

4. Hence we may apply the Dominated Convergence Theorem to let $\epsilon \to 0$ in $(\star\star\star)$:

$$⨍_{B(r)} \varphi(y)\left[f(y) - f(0) - Df(0)\cdot y - \frac{1}{2}y^T\cdot D^2 f(0)\cdot y\right] dy$$
$$\le Cr^2 \int_0^1 ⨍_{B(rs)} |D^2 f(z) - D^2 f(0)|\, dz\, ds + Cr^2\int_0^1 \frac{|[D^2 f]_s|(B(rs))}{(sr)^n}\, ds$$
$$= o(r^2) \text{ as } r \to 0, \qquad \text{by } (\star\star) \text{ with } x = 0.$$

Take the supremum over all φ as above to obtain

$$\fint_{B(r)} |h(y)|\, dy = o(r^2) \text{ as } r \to 0 \qquad (\star\star\star\star)$$

for

$$h(y) \equiv f(y) - f(0) - Df(0)\cdot y - \frac{1}{2} y^T \cdot D^2 f(0) \cdot y.$$

5. Claim #1: There exists a constant C such that

$$\sup_{B(r/2)} |Dh| \le \frac{C}{r} \fint_{B(r)} |h|\, dy + Cr \qquad (r > 0).$$

Proof of Claim #1: Let $\Lambda \equiv |D^2 f(0)|$. Then $g \equiv h + (\Lambda/2)|y|^2$ is convex: apply Theorem 1 from Section 6.3.

6. Claim #2: $\sup_{B(r/2)} |h| = o(r^2)$ as $r \to 0$.

Proof of Claim #2: Fix $0 < \epsilon, \eta < 1$, $\eta^{1/n} \le 1/2$. Then

$$\begin{aligned} \mathcal{L}^n\{z \in B(r) \mid |h(z)| \ge \epsilon r^2\} &\le \frac{1}{\epsilon r^2} \int_{B(r)} |h|\, dz \\ &= o(r^n) \text{ as } r \to 0, \text{ by } (\star\star\star\star) \\ &< \eta \mathcal{L}^n(B(r)) \qquad \text{for } 0 < r < r_0 \equiv r_0(\epsilon, \eta). \end{aligned}$$

Thus for each $y \in B(r/2)$ there exists $z \in B(r)$ such that

$$|h(z)| \le \epsilon r^2$$

and

$$|y - z| \le \sigma \equiv \eta^{1/n} r,$$

for if not,

$$\mathcal{L}^n\{z \in B(r) \mid |h(z)| \ge \epsilon r^2\} \ge \mathcal{L}^n(B(y,\sigma)) = \alpha(n)\eta r^n = \eta \mathcal{L}^n(B(r)).$$

Consequently,

$$\begin{aligned} |h(y)| &\le |h(z)| + |h(y) - h(z)| \\ &\le \epsilon r^2 + \sigma \sup_{B(r)} |Dh| \\ &\le \epsilon r^2 + C\eta^{1/n} r^2 \qquad \text{by Claim \#1 and } (\star\star\star\star) \\ &= 2\epsilon r^2, \end{aligned}$$

provided we fix η such that $C\eta^{1/n} = \epsilon$ and then choose $0 < r < r_0$.

7. According to Claim #2,

$$\sup_{B(r/2)} \left| f(y) - f(0) - Df(0)\cdot y - \frac{1}{2} y^T \cdot D^2 f(0) \cdot y \right| = o(r^2) \qquad \text{as } r \to 0.$$

This proves $(\star)$ for $x = 0$. ∎

6.5 Whitney's Extension Theorem

We next identify conditions ensuring the existence of a C^1 extension $\bar{f}$ of a given function f defined on a closed subset C of $\mathbb{R}^n$.

Let $C \subset \mathbb{R}^n$ be a closed set and assume $f : C \to \mathbb{R}$, $d : C \to \mathbb{R}^n$ are given functions.

NOTATION

(i) $R(y,x) \equiv \frac{f(y) - f(x) - d(x)\cdot(y-x)}{|x-y|}$ $\qquad (x, y \in C, x \ne y)$.

(ii) Let $K \subset C$ be compact, and set

$$\rho_K(\delta) \equiv \sup\{|R(y,x)| \mid 0 < |x - y| \le \delta,\ x, y \in K\}.$$

THEOREM 1 WHITNEY'S EXTENSION THEOREM

Assume f, d are continuous, and for each compact set $K \subset C$,

$$\rho_K(\delta) \to 0 \text{ as } \delta \to 0. \qquad (\star)$$

Then there exists a function $\bar{f} : \mathbb{R}^n \to \mathbb{R}$ such that

(i) *$\bar{f}$ is C^1.*

(ii) *$\bar{f} = f$, $D\bar{f} = d$ on C.*

PROOF

1. The proof will be a kind of "C^1-version" of the proof of the Extension Theorem presented in Section 1.2.

Let $U \equiv \mathbb{R}^n - C$; U is open. Define

$$r(x) \equiv \frac{1}{20} \min\{1, \text{dist}(x, C)\}.$$

By Vitali's Covering Theorem, there exists a countable set $\{x_j\}_{j=1}^\infty \subset U$ such that

$$U = \bigcup_{j=1}^\infty B(x_j, 5r(x_j))$$