GEOMETRIC ANALYSIS

PIOTR HAJŁASZ

1. The Rademacher theorem

1.1. **The Rademacher theorem.** Lipschitz functions defined on one dimensional intervals are differentiable a.e. This is a classical result that is covered in most of courses in measure theory. However, Rademacher proved a much deeper result that Lipschitz functions defined on open sets in $\mathbb{R}^n$ are also differentiable a.e. Let us first discuss differentiability of Lipschitz functions defined on one dimensional intervals. This result is a special case of differentiability of absolutely continuous functions.

Definition 1.1. We say that a function $f:[a,b]\to\mathbb{R}$ is *absolutely continuous* if for very $\varepsilon>0$ there is $\delta>0$ such that if $(x_1,x_1+h_1),\ldots,(x_k,x_k+h_k)$ are pairwise disjoint intervals in $[a,b]$ of total length less than δ, $\sum_{i=1}^k h_i<\delta$, then

$$\sum_{i=1}^k |f(x_i+h_i)-f(x_i)|<\varepsilon.$$

The definition of an absolutely continuous function reminds the definition of a uniformly continuous function. Indeed, we would obtain the definition of a uniformly continuous function if we would restrict just to a single interval $(x,x+h)$, i.e. if we would assume that $k=1$. Despite similarity, the class of absolutely continuous function is *much* smaller than the class of uniformly continuous function. For example Lipschitz function $f:[a,b]\to\mathbb{R}$ are absolutely continuous, but in general Hölder continuous functions are uniformly continuous, but not absolutely continuous.

Proposition 1.2. *If $f,g:[a,b]\to\mathbb{R}$ are absolutely continuous, then also the functions $f\pm g$ and fg are absolutely continuous. If in addition $g\geq c>0$ on $[a,b]$, then f/g is absolutely continuous.*

Exercise 1.3. *Prove it.*

It turns out that absolutely continuous functions are precisely the functions for which the fundamental theorem of calculus is satisfied.

Theorem 1.4. *If $f \in L^1([a,b])$, then the function*

$$F(x) = \int_a^x f(t)\, dt \tag{1.1}$$

is absolutely continuous. On the other hand, if $F : [a,b] \to \mathbb{R}$ is absolutely continuous, then F is differentiable a.e., $F' \in L^($[a,b]) and

$$F(x) = F(a) + \int_a^x F'(t)\, dt \quad \textit{for all } x \in [a,b].$$

Absolute continuity of the function F defined by (1.1) readily follows from the absolute continuity of the integral, but the second part of the theorem is difficult and we will not prove it here. In particular the theorem applies to Lipschitz functions.

Theorem 1.5 (Integration by parts)**.** *If $f, g : [a,b] \to \mathbb{R}$ are absolutely continuous, then*

$$\int_a^b f(t)g'(t)\, dt = fg|_a^b - \int_a^b f'(t)g'(t)\, dt.$$

Indeed, $fg' = (fg)' - f'g$. Integrating this identity and using absolute continuity of fg we obtain

$$\int_a^b f(t)g'(t)\, dt = \int_a^b (f(t)g(t))'\, dt - \int_a^b f'(t)g(t)\, dt = fg|_a^b - \int_a^b f'(t)g'(t)\, dt.$$

The aim of this section is to prove the following result of Rademacher and its generalizations - the Stepanov theorem and the Kirchheim therem.

Theorem 1.6 (Rademacher)**.** *If $f : \Omega \to \mathbb{R}$ is Lipschitz continuous, where $\Omega \subset \mathbb{R}^n$ is open, the f is differentiable a.e. That is the partial derivatives exist a.e. and*

$$\nabla f(x) = \left\langle \frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right\rangle$$

satisfies

$$\lim_{y \to x} \frac{f(y) - f(x) - \nabla f(x) \cdot (y - x)}{|y - x|} = 0 \quad \textit{for a.e. } x \in \Omega.$$

In the proof we will need the following lemma which is of independent interest.

Lemma 1.7. *If $f \in L^1_{\mathrm{loc}}(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is open, and $\int_\Omega f(x)\varphi(x)\, dx = 0$ for all $\varphi \in C_0^\infty(\Omega)$, then $f = 0$ a.e.*

Proof. Suppose to the contrary that $f \neq 0$ on a set of positive measure. Without loss of generality we may assume that $f > 0$ on a set of positive measure (otherwise we replace f by $-f$). Hence there is a compact set $K \subset \Omega$ and $\varepsilon > 0$ such that $f \geq \varepsilon$ on K. Let G_i be a decreasing sequence of open sets such that $K \subset G_i \Subset \Omega$ and let $\varphi \in C_0^\infty(G_i)$ be such that $0 \leq \varphi_i \leq 1$, $\varphi_i = 1$ on K. Then

$$0 = \int_\Omega f(x)\varphi_i(x)\, dx \geq \varepsilon|K| - \int_{G_i \setminus K} |f(x)|\, dx \to \varepsilon|K| \quad \text{as } \varepsilon \to 0$$

which is an obvious contradiction. We used here an absolute continuity of the integral: f is integrable on $G_1 \setminus K$ and measures of the sets $G_i \setminus K \subset G_1 \setminus K$ converge to zero. $\square$

Proof of Theorem 1.6. Let $\nu \in S^{n-1}$ and let

$$D_\nu f(x) = \left.\frac{d}{dt}\right|_{t=0} f(x + t\nu)$$

be the directional derivative. For each $\nu \in S^{n-1}$, $D_\nu f(x)$ exists a.e., because Lipschitz functions in dimension one are differentiable a.e.

If $\varphi \in C_0^\infty(\Omega)$, then for all sufficiently small $h > 0$

$$\int_\Omega \frac{f(x+h\nu) - f(x)}{h}\, \varphi(x)\, dx = -\int_\Omega \frac{\varphi(x-h\nu) - \varphi(x)}{-h}\, f(x)\, dx.$$

Although this can be regarded as a sort of integration by parts, it follows easily from a linear change of variables

$$\int_\Omega f(x+h\nu)\varphi(x)\, dx = \int_\Omega f(x)\varphi(x-h\nu)\, dx.$$

The dominated convergence theorem yields[1]

$$\int_\Omega D_\nu f(x)\varphi(x) = -\int_\Omega f(x) D_\nu \varphi(x)\, dx. \tag{1.2}$$

This is true for any $\nu \in S^{n-1}$. In particular

$$\int_\Omega \frac{\partial f}{\partial x_i}(x)\varphi(x)\, dx = -\int_\Omega f(x)\frac{\partial \varphi}{\partial x_i}(x)\, dx \quad \text{for } i = 1, 2, \ldots, n.$$

We want to prove that the directional derivative of f is linear in ν and that it equals $\nabla f(x) \cdot \nu$. The idea is to use the fact that $D_\nu \varphi(x) = \nabla \varphi(x) \cdot \nu$ in (1.2) and this should

[1] The difference quotients of f are bounded, because f is Lipschitz.

somehow translate to a similar property of the derivative of f. We have

$$\begin{aligned}\int_\Omega D_\nu f(x)\varphi(x)\,dx &= -\int_\Omega f(x)D_\nu\varphi(x)\,dx = -\int_\Omega f(x)\left(\nabla\varphi(x)\cdot\nu\right)\,dx \\ &= -\sum_{i=1}^n \int_\Omega f(x)\frac{\partial\varphi_i}{\partial x_i}\nu_i\,dx = \sum_{i=1}^n \int_\Omega \varphi(x)\frac{\partial f}{\partial x_i}(x)\nu_i\,dx \\ &= \int_\Omega \varphi(x)\left(\nabla f(x)\cdot\nu\right)\,dx\end{aligned}$$

for all $\varphi\in C_0^\infty(\Omega)$. This and Lemma 1.7 implies that for every $\nu\in S^{n-1}$

$$D_\nu f(x) = \nabla f(x)\cdot\nu \quad \text{a.e.}$$

Let $\nu_1,\nu_2,\ldots$ be a countable and dense subset of S^{n-1} and let

$$A_k = \{x\in\Omega:\ \nabla f(x) \text{ exists, } D_{\nu_k}f(x) \text{ exists, and } D_{\nu_k}f(x) = \nabla f(x)\cdot\nu_k.\}$$

Each of the sets $\Omega\setminus A_k$ has measure zero and hence

$$A = \bigcap_{k=1}^\infty A_k \quad \text{satisfies} \quad |\Omega\setminus A| = 0.$$

Clearly

$$D_{\nu_k}f(x) = \nabla f(x)\cdot\nu_k \quad \text{for all } x\in A \text{ and all } k=1,2,\ldots$$

We will prove that f is differentiable at every point of the set A. For each $x\in A$, $\nu\in S^{n-1}$ and $h>0$ we define

$$Q(x,\nu,h) = \frac{f(x+h\nu)-f(x)}{h} - \nabla f(x)\cdot\nu.$$

It suffices to prove that if $x\in A$, then for every $\varepsilon>0$ there is $\delta>0$ such that

$$|Q(x,\nu,h)|<\varepsilon \quad \text{whenever } 0<h<\delta,\ \nu\in S^{n-1}.$$

Assume that f is L-Lipschitz. Since difference quotients for f are bounded by L we have

$$\left|\frac{\partial f}{\partial x_i}\right| \leq L \quad \text{and hence} \quad |\nabla f(x)|\leq\sqrt{n}L \text{ a.e.}$$

Thus for any $\nu,\nu'\in S^{n-1}$ and $h>0$

$$|Q(x,\nu,h)-Q(x,\nu',h)| \leq (\sqrt{n}+1)L|\nu-\nu'|.$$

Given $\varepsilon>0$ let p be so large that for each $\nu\in S^{n-1}$

$$|\nu-\nu_k| \leq \frac{\varepsilon}{2(\sqrt{n}+1)L} \quad \text{for some } k=1,2,\ldots,p.$$

Since $\nabla f(x) \cdot \nu_i = D_{\nu_i} f(x)$ for $x \in A$, the definition of the directional derivative yields

$$\lim_{h\to 0^+} Q(x,\nu_i,h) = 0 \quad \text{for all } x \in A \text{ and } i = 1,2,\ldots$$

Thus given $x \in A$, there is $\delta > 0$ such that

$$|Q(x,\nu_i,h)| < \varepsilon/2 \quad \text{whenever } 0 < h < \delta \text{ and } i = 1,2,\ldots,p.$$

Now for $0 < h < \delta$ and $\nu \in S^{n-1}$ we have

$$|Q(x,\nu,h)| \le |Q(x,\nu_k,h)| + |Q(x,\nu_k,h) - Q(x,\nu,h)| \le \frac{\varepsilon}{2} + (\sqrt{n}+1)L|\nu_k - \nu| < \varepsilon.$$

The proof is complete. □

1.2. **The Stepanov theorem.** There is a very elegant characterization of functions that are differentiable a.e. due to Stepanov.

Theorem 1.8 (Stepanov)**.** *Let $\Omega \subset \mathbb{R}^n$ be open. Then a any measurable function $f : \Omega \to \mathbb{R}$ is differentiable at almost every point of the set*

$$A_f = \left\{ x \in \Omega : \limsup_{y\to x} \frac{|f(y)-f(x)|}{|y-x|} < \infty \right\}.$$

As an immediate consequence we obtain

Corollary 1.9. *A measurable function $f : \Omega \to \mathbb{R}$ is differentiable a.e. if and only if*

$$\limsup_{y\to x} \frac{|f(y)-f(x)|}{|y-x|} < \infty \text{ a.e.}$$

We will provide two proofs of the Stepanov theorem. The first one is natural and intuitive, while the second one is tricky and very elegant. We will need some basic properties of Lipschitz functions.

Lemma 1.10. *If $\{u_\alpha\}_{\alpha\in I}$ is a family of Lipschitz functions on a metric space X, then*

$$U(x) = \sup_{\alpha\in I} u_\alpha(x)$$

is L-Lipschitz provided it is finite at one point. Also

$$u(x) = \inf_{\alpha\in I} u_\alpha(x)$$

if L-Lipschitz provided it is finite at one point.

Proof. We will only prove the first part of the lemma. The proof of the second part is very similar. For $x, y \in X$ we have $u_\alpha(y) \leq u_\alpha(x) + Ld(x,y)$. Taking supremum with respect to α (firs on the right hand side and then on the left hand side) we obtain $U(y) \leq U(x) + Ld(x,y)$. If $U(x) < \infty$, then $U(y) < \infty$ for all $y \in X$ and hence $U(y) - U(x) \leq Ld(x,y)$ for all $x, y \in X$. Changing the role of x and y gives $U(x) - U(y) \leq Ld(x,y)$ and thus $|U(y) - U(x)| \leq Ld(x,y)$. □

Theorem 1.11 (McShane)**.** *If $f : A \to \mathbb{R}$ is an L-Lipschitz function defined on a subset $A \subset X$ of a metric space, then there is an L-Lipschitz function $\tilde{f} : X \to \mathbb{R}$ such that $\tilde{f}|_A = f$.*

In other words a Lipschitz function defined on a subset of a metric space can be extended to a Lipschitz function defined on the whole space with the same Lipschitz constant.

Proof. For $x \in X$ we simply define

$$\tilde{f}(x) = \inf_{y \in A} (f(y) + Ld(x,y)).$$

For each $y \in X$, the function $x \mapsto f(y) + Ld(x,y)$ is L-Lipschitz, so $\tilde{f}$ is L Lipschitz by Lemma 1.10. Clearly $\tilde{f}(x) = f(x)$ for $x \in A$, because $f(x) \leq f(y) + Ld(x,y)$ for any $y \in A$ by the L-Lipschitz continuity of f and $f(x) = f(x) + Ld(x,x)$. □

We also need to recall the notion of a density point of a measurable set. We say that x is a density point of a measurable set E if

$$\lim_{r \to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 1.$$

That condition simply means that in a very small ball centered at x the set E fills most of the ball. More than 99.999999% of the ball. Note that we do not require that x belongs to E, but the most interesting question is which points of the set are its density points.

Theorem 1.12. *Almost every point of a measurable set $E \subset \mathbb{R}^n$ is a density point of E.*

Proof. Recall that if $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then for almost every x we have[2]

$$⨍_{B(x,r)} f(y)\, dy := \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)\, dy \to f(x) \quad \text{as } r \to 0.$$

[2] Here and in what follows the barred integral will denote the integral average, i.e. the integral divided by the measure of the set over which we integrate the function.

This is the classical Lebesgue differentiation theorem. Now our result is an immediate consequence of this result applied to the characteristic function of the set E, $f = \chi_E$. $\square$

First proof of Theorem 1.8. The idea is as follows. First we show that the set A_f can be written as the union of countably many sets $A_f = \bigcup_{i=1}^\infty E_i$ such that $f|_{E_i}$ is Lipschitz. The function $f|_{E_i}$ can be extended to a Lipschitz function $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ (McShane) which, by the Rademacher theorem, is differentiable a.e. Then it follows from the triangle inequality that if $x \in E_i$ is a density point of E_i and a point of differentiability of $\tilde{f}$, then f is differentiable at x with $\nabla f(x) = \nabla \tilde{f}(x)$.

Let

$$E_{k,\ell} = \left\{ x \in A_f : \, |f(x)| \leq k, \text{ and } \frac{|f(x) - f(y)|}{|x-y|} \leq k \text{ if } |x-y| \leq \frac{1}{\ell} \right\}.$$

It follows from the definition of $\limsup$ and the definition of the set A_f that

$$A_f = \bigcup_{k,\ell} E_{k,\ell}.$$

Hence it suffices to prove that f is differentiable a.e. in each set $E_{k,\ell}$. First observe that the function $f|_{E_{k,\ell}}$ is Lipschitz continuous. Indeed, if $|x-y| < 1/\ell$, then $|f(x)-f(y)| \leq k|x-y|$ and if $|x-y| \geq 1/\ell$, then $|f(x) - f(y)| \leq 2k \leq 2k\ell|x-y|$. Let $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz extension of $f|_{E_{k,\ell}}$. We will prove that f is differentiable at all density points of $E_{k,\ell}$ which are points of differentiability of $\tilde{f}$ and that

$$\nabla f(x) = \nabla \tilde{f}(x) \quad \text{at such points.}$$

Let $x \in E_{k,\ell}$ be a density point such that $\tilde{f}$ is differentiable at x. We need to show that

$$\frac{|f(x+y) - f(x) - \nabla \tilde{f}(x) \cdot y|}{|y|} \to 0 \quad \text{as } y \to 0.$$

This is obvious if $x + y \in E_{k,\ell}$ because $f(x+y) = \tilde{f}(x+y)$, $f(x) = \tilde{f}(x)$ and $\tilde{f}$ is differentiable at x. If $x + y \notin E_{k,\ell}$, then the fact that x is a density point of $E_{k,\ell}$ implies that there is $\tilde{y}$ such that $x + \tilde{y} \in E_{k,\ell}$ and $|y - \tilde{y}| = o(|y|)$ as $|y| \to 0$. We have

$$\begin{aligned} \frac{|f(x+y) - f(x) - \nabla \tilde{f}(x) \cdot y|}{|y|} &\leq \frac{|f(x+\tilde{y}) - f(x) - \nabla \tilde{f}(x) \cdot \tilde{y}|}{|\tilde{y}|} \frac{|\tilde{y}|}{|y|} \\ &+ \frac{|f(x+\tilde{y}) - f(x+y)|}{|y|} + \frac{|\nabla \tilde{f}(x) \cdot (y - \tilde{y})|}{|y|} \to 0 \quad \text{as } y \to 0. \end{aligned}$$

The convergence to zero of the first and the third expression on the right hand side is obvious. For the middle term observe that $x+\tilde{y} \in E_{k,\ell}$ and we can assume that $|y-\tilde{y}| < 1/\ell$.

Thus the definition of the set $E_{k\ell}$ yields[3]

$$\frac{|f(x+\tilde{y})-f(x+y)|}{|y-\tilde{y}|}\leq k\,.$$

Hence

$$\frac{|f(x+\tilde{y})-f(x+y)|}{|y|}\leq k\frac{|y-\tilde{y}|}{|y|}\to 0 \quad \text{as } y\to 0.$$

The proof is complete. $\square$

Second proof of Theorem 1.8. We need the following elementary fact.

Lemma 1.13. *If $g\leq f\leq h$, $g(x_0)=f(x_0)=h(x_0)$ and the functions g and h are differentiable at x_0, then f is differentiable at x_0 and*

$$\nabla f(x_0)=\nabla g(x_0)=\nabla h(x_0).$$

Proof. Since $h-g\geq 0$ and $(h-g)(x_0)=0$, we have $\nabla(h-g)(x_0)=0$ and hence $\nabla h(x_0)=\nabla g(x_0)$. Let $L=\nabla g(x_0)=\nabla h(x_0)$. Then

$$\begin{aligned}\frac{g(y)-g(x_0)-\nabla g(x_0)(y-x_0)}{|y-x_0|} &\leq \frac{f(y)-f(x_0)-L(y-x_0)}{|y-x_0|}\\ &\leq \frac{h(y)-h(x_0)-\nabla h(x_0)(y-x_0)}{|y-x_0|}\end{aligned}$$

Clearly the left and the right term converge to zero when $y\to x_0$ and so does the middle one. $\square$

Now we can complete the proof of the Stepanov theorem. Let $\{U_i\}_{i=1}^\infty$ be the family of *all* balls with rational radius and center[4] in Ω such that $f|_{U_i}$ is bounded. Clearly $A_f\subset\bigcup_i U_i$. It is important here that we consider all such balls and not only the largest ones, so every point in A_f is covered by arbitrarily small balls U_i with arbitrarily large indexes i.

Let $a_i:U_i\to\mathbb{R}$ be the supremun of all i-Lipschitz functions $\leq f|_{U_i}$ and let $b_i:U_i\to\mathbb{R}$ be the infimum of all i-Lipschitz functions $\geq f|_{U_i}$. According to Lemma 1.10 the functions a_i and b_i are i-Lipschitz. Clearly

$$a_i\leq f|_{U_i}\leq b_i \quad \text{on } U_i. \tag{1.3}$$

[3]Note that the Lipschitz type condition in the definition of the set $E_{k,\ell}$ requires that only one of the points is in the set $E_{k,\ell}$ while the other point can be arbitrary, but close, so this is a stronger condition than being Lipschitz on the set $E_{k,\ell}$ and this is very important here, because $x+y\notin E_{k,\ell}$.

[4]We need to take rational radius and center as otherwise the family of balls would be uncountable.

Let

$$E_i = \{x \in U_i : \text{ both } a_i \text{ and } b_i \text{ are differentiable at } x.\}$$

By the Rademacher theorem the set

$$Z = \bigcup_{i=1}^{\infty} U_i \setminus E_i$$

has measure zero. It remains to show that f is differentiable at all points of $A_f \setminus Z$. Let $x \in A_f \setminus Z$. It suffices to prove that there is an index i such that $x \in E_i$ and $a_i(x) = b_i(x)$. Indeed, differentiability of f at x will follow from lemma 1.13 and inequality (1.3). Since $x \in A_f$, there are $r > 0$ and $\lambda > 0$ such that

$$|f(y) - f(x)| \leq \lambda |y - x| \quad \text{for } y \in B(x, r).$$

There are infinitely many balls U_i of arbitrarily small radii that contain x, so we can find $i \geq \lambda$ such that $x \in U_i \subset B(x, r)$. Since $x \notin Z$ we have $x \in E_i$.[5] Now for $y \in U_i \subset B(x, r)$ we ave

$$f(y) \leq f(x) + \lambda |y - x| \leq f(x) + i|y - x|.$$

The function $y \mapsto f(x) + i|y - x|$ is i-Lipschitz and larger than f on U_i, so the definition of b_i yields

$$f(y) \leq b_i(y) \leq f(x) + i|y - x| \quad \text{on } U_i.$$

Taking $y = x$ yields $f(x) = b_i(x)$. Similarly

$$f(y) \geq a_i(y) \geq f(x) - i|y - x|$$

and hence $f(x) = a_i(x)$. In particular $a_i(x) = b_i(x)$ and the proof is complete. □

1.3. **The Kirchheim theorem.** Since differentiability of a mapping

$$f = (f_1, \ldots, f_m) : \mathbb{R}^n \supset \Omega \to \mathbb{R}^m$$

is equivalent to differentiability of components $f_1, \ldots, f_m$, Lipschitz mappings $f : \Omega \to \mathbb{R}^m$ are differentiable a.e.

We will show now that Lipschitz mappings $f : \Omega \to X$ from an open subset of $\mathbb{R}^n$ to an *arbitrary* metric space X are differentiable a.e. in some generalized sense. This will be a really surprising generalization of the Rademacher theorem.

[5]Indeed, $x \in U_i$, so assuming $x \notin E_i$ would imply that $x \in U_i \setminus E_i \subset Z$.

Suppose that $f:\Omega\to\mathbb{R}^m$ (not necessarily Lipschitz) is differentiable at $x\in\Omega$. Then

$$\left|\frac{|f(y)-f(x)|-|Df(x)(y-x)|}{|y-x|}\right|\leq\frac{|f(y)-f(x)-Df(x)(y-x)|}{|y-x|}\xrightarrow{y\to x}0.$$

Observe that $\|z\|_x:=|Df(x)z|$ is a seminorm.[6]

Definition 1.14. Let X be a metric space. We say that a mappings $f:\mathbb{R}^n\supset\Omega\to X$ is *metrically differentiable at* $x\in\Omega$ if there is a seminorm $\|\cdot\|_x$ on $\mathbb{R}^n$ such that

$$\frac{d(f(y),f(x))-\|y-x\|_x}{|y-x|}\to 0\quad\text{as } y\to x.$$

The seminorm $\|\cdot\|_x$ is called the *metric derivative* of f and it will be denoted by

$$mDf(x)(z)=\|z\|_x$$

Exercise 1.15. *Show that if the seminorm $\|\cdot\|_x$ exists, it is unique.*

Clearly a mappings $f:\Omega\to\mathbb{R}^m$ differentiable at $x\in\Omega$ is metrically differentiable with $mDf(x)(z)=\|z\|_x=|Df(x)z|$. In particular Lipschitz mappings into $\mathbb{R}^m$ are metrically differentiable a.e.

The metric derivative however, does not see directions in the target, because there is no linear structure in the space X. It follows directly from the definition of the metric derivative that for $v\in\mathbb{R}^n$

$$\frac{d(f(x+tv),f(x))-mDf(x)(tv)}{|t|}\to 0\quad\text{as } t\to 0.$$

Since $mDf(x)(tv)=|t|mDf(x)(v)$ (seminorm property) we conclude that

$$mDf(x)(v)=\lim_{t\to 0}\frac{d(f(x+tv),f(x))}{|t|}\quad\text{for all } v\in\mathbb{R}^n. \tag{1.4}$$

Thus if f is metrically differentiable at x, the metric derivative $mDf(x)(v)$ equals the "speed" of the curve $t\mapsto f(x+tv)$ at $t=0$. The metric derivative gives us also a control (in some convex way) on how the speed may change if we change the direction v.

Exercise 1.16. *Show that the function $f:\mathbb{R}\to\mathbb{R}$, $f(x)=|x|$ is metrically differentiable at every point.*[7]

The aim of this subsection is to prove the following surprising result.

[6]That mens $\|z_1+z_2\|_x\leq\|z_1\|_x+\|z_2\|_x$, $\|tz\|_x=|t|\|z\|_x$, but $\|\cdot\|_x$ may vanish on a subspace of $\mathbb{R}^n$.

[7]What a disappointment.

Theorem 1.17 (Kirchheim). *Lipschitz mappings $f : \mathbb{R}^n \supset \Omega \to X$ into an arbitrary metric space are metrically differentiable a.e.*

The main idea is to embed X isometrically into the Banach space ℓ^∞ of bounded sequences. The linear structure of ℓ^∞ will allow us to prove some kind of weak differentiability of Lipschitz mappings $f : \Omega \to \ell^\infty$ in a functional analysis sense. Thus we need

Theorem 1.18 (Kuratowski). *Any separable metric space X admits an isometric embedding into ℓ^∞.*

Proof. Fix $x_0 \in X$ and let $\{x_i\}_{i=1}^\infty \subset X$ be a countable and dense subset of X. Then the mapping

$$X \ni x \mapsto \kappa(x) = \{d(x, x_i) - d(x_i, x_0)\}_{i=1}^\infty \in \ell^\infty$$

is an isometric embedding of X into ℓ^∞. To prove this it suffices to show that

$$\|\kappa(x) - \kappa(y)\|_\infty = d(x, y) \quad \text{for all } x, y \in X.$$

We have

$$\|\kappa(x) - \kappa(y)\|_\infty = \sup_i |d(x, x_i) - d(y, x_i)|.$$

Since for any i, $|d(x, x_i) - d(y, x_i)| \leq d(x, y)$ we conclude that $\|\kappa(x) - \kappa(y)\|_\infty \leq d(x, y)$ and it remains to prove the opposite inequality. Choose a sequence[8] $x_{i_k} \to y$. Then $|d(x, x_{i_k}) - d(y, x_{i_k})| \to d(x, y)$ and hence $\|\kappa(x) - \kappa(y)\|_\infty \geq d(x, y)$. □

Remark 1.19. The embdding $\kappa : X \to \infty$ is called the *Kuratowski embedding.*

In Kirchheim's theorem we *do not* assume that the space X is separable. However, the subspace $\tilde{X} = f(\Omega) \subset X$ is separable and the metric differentiability condition refers only to the points of the space $\tilde{X}$. Thus after all we can assume in the Kirchheim theorem that X is separable by restricting the space to $\tilde{X}$ if necessary. Hence we can assume that $X \subset \ell^\infty$ and it remains to prove that Lipschitz mappings $f : \Omega \to \ell^\infty$ are metrically differentiable a.e. This allows us to use functional analysis and we need another notion of differentiability.

Although the space ℓ^∞ is known to be ugly, it has a nice and useful property of being dual to a separable Banach space $\ell^\infty = (\ell^1)^*$.

[8]The sequence exists, because the set $\{x_i\}_{i=1}^\infty \subset X$ is dense.

Definition 1.20. Let $Y = G^*$ be dual to a separable real Banach space G. We say that a mapping $f : \mathbb{R}^n \supset \Omega \to Y$ is *w^*-differentiable* at $x \in \Omega$ if there is a bounded linear mapping $L : \mathbb{R}^n \to Y$ such that[9]

$$\left\langle \frac{f(y) - f(x) - L(y-x)}{|y-x|}, g \right\rangle \to 0 \quad \text{as } y \to x.$$

for every $g \in G$. The mapping L is called the *w^*-derivative* of f and it will be denoted by

$$wDf(x) : \mathbb{R}^n \to Y.$$

In other words we assume that the expression

$$\frac{f(y) - f(x) - L(y-x)}{|y-x|}$$

converges to zero as $y \to x$ in the weak-$*$ sense.

The next lemma shows a basic comparison between the metric derivative and the w^*-derivative.

Lemma 1.21. *If $f : \Omega \to Y = G^*$ is both w^*-differentiable and metrically differentiable at $x \in \Omega$, then*

$$\|wDf(x)(v)\|_Y \leq mDf(x)(v) \quad \textit{for all } v \in \mathbb{R}^n.$$

Proof. It follows from (1.4) that

$$mDf(x)(v) = \lim_{t \to 0} \left\| \frac{f(x+tv) - f(x)}{t} \right\|_Y.$$

On the other hand w^*-differentiability yields

$$\left\langle \frac{f(x+tv) - f(x) - wDf(x)(tv)}{t}, g \right\rangle \to 0 \quad \text{as } t \to 0$$

and hence

$$\lim_{t \to 0} \left\langle \frac{f(x+tv) - f(x)}{t}, g \right\rangle = \langle wDf(x)(v), g \rangle.$$

If $\|g\|_G = 1$, then

$$\left\langle \frac{f(x+tv) - f(x)}{t}, g \right\rangle \leq \left\| \frac{f(x+tv) - f(x)}{t} \right\|_Y$$

and passing to the limit as $t \to 0$ gives

$$\langle wDf(x)(v), g \rangle \leq mDf(x)(v).$$

Taking the supremum over all $g \in G$ with $\|g\|_G = 1$ completes the proof. □

[9]Here $\langle z, g \rangle$ denotes the evaluation of the functional $z \in Y = G^*$ on the element $g \in G$.

Since we reduced the Kirchheim theorem to the case of Lipschitz mappings into ℓ^∞, the Kirchheim's result is a direct consequence of the following slightly stronger result.

Theorem 1.22. *Let $Y = G^*$ be dual to a separable real Banach space G. Then any Lipschitz mapping $f : \mathbb{R}^n \supset \Omega \to Y$ is w^*-differentiable a.e., metrically differentiable a.e. and*

$$mDf(x)(v) = \|wDf(x)(v)\|_Y$$

for almost all $x \in \Omega$ an all $v \in \mathbb{R}^n$.

Proof. Let $D \subset G$ be a countable and a dense subset. According to the Rademacher theorem and the fact that D is countable, there is a set $N \subset \Omega$ of Lebesgue measure zero, $|N| = 0$, such that for every $g \in D$, the real valued Lipschitz function[10]

$$x \mapsto f_g(x) = \langle f(x), g\rangle$$

is differentiable at every point of the set $\Omega \setminus N$, i.e. there is a vector $\nabla f_g(x) \in \mathbb{R}^n$ such that

$$\frac{f_g(y) - f_g(x) - \nabla f_g(x)\cdot(y-x)}{|y-x|} \to 0 \quad \text{as } y \to x \tag{1.5}$$

for every $x \in \Omega \setminus N$ and every $g \in D$. Observe that (1.5) implies that for every $v \in \mathbb{R}^n$, $g \in D$ and $x \in \Omega \setminus N$

$$\left\langle \frac{f(x+tv) - f(x)}{t}, g \right\rangle \to \nabla f_g(x)\cdot v \quad \text{as } t \to 0. \tag{1.6}$$

This shows that the mapping $g \mapsto \nabla f_g(x)$ is linear. There is however, a problem, because g belongs to a countable set D which has no linear structure and it really does not make sense to talk about linearity of the mapping $g \mapsto \nabla f_g(x)$. To overcome this difficulty we can assume that D is a linear space over the field $\mathbb{Q}$ of rational numbers, i.e.

$$\text{if } g_1, g_2 \in D \text{ and } a_1, a_2 \in \mathbb{Q}, \text{ then } a_1g_1 + a_2g_2 \in D.$$

If not, we simply replace D by its linear span over $\mathbb{Q}$

$$\{a_1g_1 + \ldots + a_kg_k :\ k \geq 1,\ a_i \in \mathbb{Q},\ g_i \in D\}.$$

This set is still countable. Thus assuming a $\mathbb{Q}$-linear structure in D we see that the mapping $g \mapsto \nabla f_g(x)$ is $\mathbb{Q}$-linear on the space D. It is also bounded as a mapping from D to $\mathbb{R}^n$.

[10]Check that f_g is Lipschitz continuous.

Indeed,

$$
\begin{aligned}
(1.7)\qquad |\nabla f_g(x)| &= \sup_{|v|=1} |\nabla f_g(x)\cdot v| = \sup_{|v|=1}\lim_{t\to 0}\left|\left\langle \frac{f(x+tv)-f(x)}{t}\, , g\right\rangle\right| \\
&\leq \sup_{|v|=1}\liminf_{t\to 0}\left\|\frac{f(x+tv)-f(x)}{t}\right\|_Y \|g\|_G \leq L\|g\|_G,
\end{aligned}
$$

where L is the Lipschitz constant of f. That means the mapping

$$D\ni g\mapsto \nabla f_g(x)\in\mathbb{R}^n$$

is linear and bounded with the norm bounded by L. Since the set $D\subset G$ is dense, it uniquely extends to a linear and bounded mapping $G\to\mathbb{R}^n$ which still will be denoted by $g\mapsto\nabla f_g(x)$.

Let $Df(x):\mathbb{R}^n\to Y$ be defined as follows. For $v\in\mathbb{R}^n$, $Df(x)v\in Y=G^*$ is a functional on G defined by the formula

$$(1.8)\qquad \langle Df(x)v, g\rangle := \nabla f_g(x)\cdot v.$$

Clearly $Df(x)$ is a linear mapping. Moreover the operator norm of $Df(x)$ is bounded by L. Indeed,

$$
\begin{aligned}
\|Df(x)\| &= \sup_{|v|=1}\|Df(x)v\|_Y = \sup_{|v|=1}\sup_{\|g\|_G=1}|\langle Df(x)v,g\rangle| \\
&= \sup_{\|g\|_G=1}\sup_{|v|=1}|\nabla f_g(x)\cdot v|\leq L
\end{aligned}
$$

by (1.7). Now (1.5) can be rewritten as

$$(1.9)\qquad \left\langle\frac{f(y)-f(x)-Df(x)(y-x)}{|y-x|}\, , g\right\rangle \stackrel{y\to x}{\longrightarrow} 0 \quad\text{for every } g\in D.$$

Since

$$\left\|\frac{f(y)-f(x)-Df(x)(y-x)}{|y-x|}\right\|_Y\leq 2L,$$

density of $D\subset G$ implies that (1.9) is true for every $g\in G$ (why?). Hence $f:\Omega\to Y$ is w^*-differentiable in all points of $\Omega\setminus N$ and

$$wDf(x)=Df(x).$$

In order to show metric differentiability of f with $mDf(x)(v)=\|wDf(x)v\|_Y$ it suffices to show that

$$(1.10)\qquad \sup_{|v|=1}\left|\left\|\frac{f(x+tv)-f(x)}{t}\right\|_Y - \|Df(x)v\|_Y\right|\to 0\quad\text{as } t\to 0.$$

To this end it suffices to show that for every $v \in S^{n-1}$

$$\left\| \frac{f(x+tv)-f(x)}{t} \right\|_Y \to \|Df(x)v\|_Y \quad \text{as } t \to 0. \tag{1.11}$$

Indeed, (1.10) can be concluded from (1.11) by the following argument. Let $\{v_i\}_{i=1}^\infty$ be a dense subset of S^{n-1}. For any $v \in S^{n-1}$ and any v_k we have[11]

$$\begin{aligned}
&\left| \left\| \frac{f(x+tv)-f(x)}{t} \right\|_Y - \|Df(x)v\|_Y \right| \\
\leq\ & \left| \left\| \frac{f(x+tv_k)-f(x)}{t} \right\|_Y - \|Df(x)v_k\|_Y \right| \\
+\ & \left\| \frac{f(x+tv_k)-f(x+tv)}{t} \right\|_Y + \|Df(x)(v-v_k)\|_Y \\
\leq\ & \left| \left\| \frac{f(x+tv_k)-f(x)}{t} \right\|_Y - \|Df(x)v_k\|_Y \right| + 2L|v-v_k|.
\end{aligned} \tag{1.12}$$

The remaining argument is pretty standard. Given $\varepsilon > 0$ there is p such that for every $v \in S^{n-1}$

$$|v - v_k| < \frac{\varepsilon}{4L} \quad \text{for some } k = 1, 2, \ldots, p.$$

It follows from (1.11) that there is $\delta > 0$ such that for any $-\delta < t < \delta$

$$\sup_{i \in \{1,2,\ldots,p\}} \left| \left\| \frac{f(x+tv_i)-f(x)}{t} \right\|_Y - \|Df(x)v_i\|_Y \right| < \frac{\varepsilon}{2}\,.$$

Hence (1.12) yields

$$\sup_{|v|=1} \left| \left\| \frac{f(x+tv)-f(x)}{t} \right\|_Y - \|Df(x)v\|_Y \right| \leq \frac{\varepsilon}{2} + 2L \cdot \frac{\varepsilon}{4L} = \varepsilon.$$

This proves (1.10). Therefore it remains to prove (1.11). For $g \in G$, $\|g\|_G = 1$ we have[12]

$$\langle Df(x)v, g\rangle = \lim_{t\to 0} \left\langle \frac{f(x+tv)-f(x)}{t}, g \right\rangle \leq \liminf_{t\to 0} \left\| \frac{f(x+tv)-f(x)}{t} \right\|_Y.$$

Taking the supremum over all g with $\|g\|_G = 1$ we obtain

$$\|Df(x)v\|_Y \leq \liminf_{t\to 0} \left\| \frac{f(x+tv)-f(x)}{t} \right\|_Y. \tag{1.13}$$

Observe that

$$\nabla f_g(x) \cdot v = \langle Df(x)v, g\rangle$$

---

[11] We are using here an elementary inequality $\big|\|a\| - \|b\|\big| \leq \big|\|c\| - \|d\|\big| + \|a-c\| + \|b-d\|$

[12] See (1.9) and a comment that follows it.

is the directional derivative of the function f_g in the direction v. Thus the directional derivative exists for all[13] $x \in \Omega \setminus N$ and all $g \in G$, $v \in \mathbb{R}^n$. The Fubini theorem and Theorem 1.4 imply that for almost all $x \in \Omega$ and all $g \in G$, $v \in \mathbb{R}^n$

$$\langle f(x+tv) - f(x), g\rangle = \int_0^t \langle Df(x+\tau v)v, g\rangle \, d\tau.$$

Taking the supremum over[14] $g \in G$ with $\|g\|_G = 1$ we get

$$\|f(x+tv) - f(x)\|_Y \leq \int_0^t \|Df(x+\tau v)v\|_Y \, d\tau.$$

Since the function $\tau \mapsto \|Df(x+\tau v)v\|_Y$ is bounded by $L|v|$, the Lebesgue differentiation theorem implies that

$$\limsup_{t\to 0} \left\| \frac{f(x+tv) - f(x)}{t} \right\|_Y \leq \limsup_{t\to 0} ⨍_0^t \|Df(x+\tau v)v\|_Y \, d\tau = \|Df(x)v\|_Y$$

almost everywhere. This together with (1.13) implies (1.11). The proof of Theorem 1.22 and hence that of Kirchheim's theorem are complete. □

2. Whitney extension and approximately differentiable functions

2.1. **The Whitney extension theorem.** Whitney provided a complete answer to the following important problem. Given a continuous function f on a compact set $K \subset \mathbb{R}^n$, and a positive integer m, find a necessary and sufficient condition for the existence of a function $F \in C^m(\mathbb{R}^n)$ such that $F|_K = f$. We will formulate and prove this result in Section **??**, but now we will state a special case of this result when $m = 1$ and we will show some applications.

Theorem 2.1 (Whitney extension theorem)**.** *Let $K \subset \mathbb{R}^n$ be a compact set and let $f : K \to \mathbb{R}$, $L : K \to \mathbb{R}^n$ be continuous functions. Then there is a function $F \in C^1(\mathbb{R}^n)$ such that*

$$F|_K = f \qquad \text{and} \qquad DF|_K = L$$

if and only if

$$\lim_{\substack{x,y\in K,\, x\neq y \\ |x-y|\to 0}} \frac{|f(y) - f(x) - L(x)(y-x)|}{|y-x|} = 0.$$

[13]See (1.6) and a comment after (1.9).

[14]First on the right hand side and then on the left hand side.

Necessity of the condition easily follows from the Taylor formula of the first order. The sufficiency is difficult since we need to construct the extension explicitly and that requires a lot of work. Note that in the limit we require the uniform convergence to zero as $|x-y| \to 0$. It is not enough to assume that the limit equals zero as $y \to x$ for every $x \in K$.

2.2. **A surprising example.** Whitney constructed a surprising example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ of class C^1 which is not constant on a certain arc, but whose gradient equals zero on that arc. We will see later in Section **??** that if a function $f : \mathbb{R}^2 \to \mathbb{R}$ is of class C^2 and its gradient equals zero on an arc, then f is constant on that arc. Now we will show who a Whitney type example can be constructed using the Whitney theorem. The famous van Koch snowflake K is homeomorphic to a unit circle. It follows from the construction of the curve that there is a homeomorphism $\Phi : S^1 \to K$ such that

$$C_1|x-y|^{\frac{\log 3}{\log 4}} \leq |\Phi(x) - \Phi(y)| \leq C_2|x-y|^{\frac{\log 3}{\log 4}}$$

for some positive constants C_1 and C_2. We will not prove this fact here. In particular $\Phi^{-1}(y) - \Phi^{-1}(x)| \leq C|y-x|^{\log 4/\log 3}$. Let $f = \Phi^{-1} : K \to \mathbb{R}^2$ and let $L = 0$ on K. We have

$$\frac{|f(y) - f(x) - L(x)(y-x)|}{|y-x|} \leq C|y-x|^{(\log 4/\log 3)-1} \to 0 \quad \text{when } x, y \in K,\ |y-x| \to 0.$$

Now the Whitney extension theorem implies that the function f extends to a C^1 mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ whose derivative restricted to K equals $L = 0$. In Section **??** we will prove a more general result of this type.

2.3. **The C^1-Lusin property.** A measurable function coincides with a continuous function outside a set of an arbitrarily small measure. This is the Lusin property of measurable functions. The following result, our first application of the Whitney theorem, shows a similar C^1-Lusin property of differentiable functions.

Theorem 2.2 (Federer)**.** *If $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}$ is differentiable a.e., then for any $\varepsilon > 0$ there is a function $g \in C^1(\mathbb{R}^n)$ such that*

$$|\{x \in \Omega :\ f(x) \neq g(x)\}| < \varepsilon.$$

Proof. Let $L(x) = Df(x)$. L is a measurable function. Assume for a moment that $f : \mathbb{R}^n \to \mathbb{R}$ is defined on $\mathbb{R}^n$ and that f vanished outside an open ball B. According to the Lusin theorem there is a compact set $K' \subset B$ such that f is differentiable on K', $f|_{K'}$, $L|_{K'}$ are continuous, and $|B \setminus K'| < \varepsilon/2$. Hence

$$\lim_{\substack{K' \ni y \to x \\ y \neq x}} \frac{|f(y) - f(x) - L(x)(y-x)|}{|y-x|} = 0 \quad \text{for all } x \in K'. \tag{2.1}$$

This condition is however, weaker than the one required in the Whitney theorem – we need uniform convergence over a compact set as $|x-y| \to 0$. Let

$$R(x,y) = \frac{|f(y) - f(x) - L(x)(y-x)|}{|y-x|}$$

and let

$$\eta_k(x) = \sup\{R(x,y) :\ K' \ni y \neq x,\ |x-y| < 1/k\} \quad k = 1, 2, \ldots$$

The condition (2.1) means that for very $x \in K'$, $\eta_k(x) \to 0$ as $k \to \infty$. According to Egorov's theorem there is another compact set $K \subset K'$ such that $|K' \setminus K| < \varepsilon/2$ and

$$\eta_k \rightrightarrows 0 \quad \text{uniformly on } K \text{ as } k \to \infty.$$

Hence

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

and according to the Whitney theorem there is $F \in C^1(\mathbb{R}^n)$ such that $F|_K = f|_K$, $DF|_K = Df|_K$, $|B \setminus K| < \varepsilon$. Multiplying F by a function $\varphi \in C_0^\infty$ that is equal 1 on K we may further assume that F has compact support in B. Since both functions f and F and their derivatives vanish outside B we have that $F = f$ and $DF = Df$ on $\mathbb{R}^n$ except for a set of measure less than ε.

To prove the result in the general case we need to represent the function $f : \Omega \to \mathbb{R}$ a a sum of functions as in the case described above. This can be done with the help of a partition of unity.

Lemma 2.3 (Partition of unity). *Let $\Omega \subset \mathbb{R}^n$ be open. Then there is a family of balls $B(x_i, r_i) \subset \Omega$, $i = 1, 2, \ldots$ and a family of functions $\varphi_i \in C_0^\infty(B(x_i, 2r_i))$ such that*

(1) $\bigcup_{i=1}^\infty B(x_i, r_i) = \Omega$;
(2) $B(x_i, 2r_i) \subset \Omega$;
(3) *No point of Ω belongs to more than 40^n balls $B(x, 2r_i)$;*
(4) $\sum_{i=1}^\infty \varphi_i(x) = 1$ *for every $x \in \Omega$.*

We will not prove this lemma.

Let $f_i = \varphi f$. Then $f = \sum_{i=1}^{\infty} f_i$. Note that in a neighborhood of any point in Ω only a finite number of terms is different than zero, i.e. the sum is locally finite. Let $F_i \in C_0^1(B(x_i, 2r_i))$ be such that $F_i = f_i$ on $\mathbb{R}^n$ except for a set of measure less than $\varepsilon/2^i$ and let $F = \sum_{i=1}^{\infty} F_i$. Since the sum is locally finite in Ω, $F \in C^1(\Omega)$.[15] If $F(x) \neq f(x)$, then there is i such that $F_i(x) \neq f_i(x)$ and hence the set of such points has measure less than $\sum_{i=1}^{\infty} \varepsilon/2^i = \varepsilon$.

We constructed $F \in C^1(\Omega)$. In order to obtain a function F of class $C^1(\mathbb{R}^n)$ we simply need to multiply it by a suitable cut-off function which vanishes near the boundary of Ω. We leave details as an exercise. □

2.4. **Approximately differentiable functions.**

Definition 2.4. Let $f : E \to \mathbb{R}$ be a measurable function defined on a measurable set $E \subset \mathbb{R}^n$. We say that f is approximately differentiable at $x \in E$ if there is a linear function $L : \mathbb{R}^n \to \mathbb{R}$ such that for any $\varepsilon > 0$ the set

$$\left\{ y \in E : \frac{|f(y) - f(x) - L(y-x)|}{|y-x|} < \varepsilon \right\} \tag{2.2}$$

has x as a density point. L is called the *approximate derivative of f* and it is often denoted by $\mathrm{ap}Df(x)$.

Exercise 2.5. *Prove that* $\mathrm{ap}Df(x)$ *is uniquely determined.*

The next result provides a useful characterization of approximately differentiable functions. In what follows we will rather refer to the condition given in this characterization than to the original definition.

Proposition 2.6. *A measurable function $f : E \to \mathbb{R}$ defined in a measurable set $E \subset \mathbb{R}^n$ is approximately differentiable at $x \in E$ if and only if there is a measurable set $E_x \subset E$*

[15] The functions F_i have compact support and they are defined on $\mathbb{R}^n$. Moreover the series that defines F is finite at every point of $\mathbb{R}^n$. Does it mean that $F \in C^1(\mathbb{R}^n)$? Not necessarily. The sum is locally finite in Ω, but not in $\mathbb{R}^n$. Any neighborhood of a boundary point of Ω contains infinitely many balls $B(x_i, 2r_i)$. For example the functions φ_i have compact support and hence they are defined on $\mathbb{R}^n$. However, their sum as a function on $\mathbb{R}^n$ equals to the characteristic function of Ω which is not continuous at the boundary of Ω.

and a linear function $L:\mathbb{R}^n\to\mathbb{R}$ such that x is a density point of E_x and

$$\lim_{E_x\ni y\to x}\frac{f(y)-f(x)-L(y-x)}{|y-x|}=0. \tag{2.3}$$

Proof. The implication from right to left is obvious, because the set (2.2) contains $E_x\cap B(x,r)$ for some small r and clearly x is a density point of this set. To prove the opposite implication we need to define the set E_x. Let r_k be a sequence strictly decreasing to 0 such that

$$r_{k+1}\leq\frac{r_k}{2^{k/n}}$$

and

$$\left|\left\{y\in B(x,r)\cap E:\ \frac{|f(y)-f(x)-L(y-x)|}{|y-x|}<\frac{1}{k}\right\}\right|\geq\omega_n r^n\left(1-\frac{1}{2^k}\right) \tag{2.4}$$

whenever $0<r\leq r_k$. Here and in what follows ω_n stands for the volume of the unit ball in $\mathbb{R}^n$. Let

$$E_k=\left\{y\in B(x,r_k)\cap E:\ \frac{|f(y)-f(x)-L(y-x)|}{|y-x|}<\frac{1}{k}\right\}.$$

It follows from (2.4) that

$$|E_k|\geq\omega_n r_k^n\left(1-\frac{1}{2^k}\right).$$

Finally we define

$$E_x=\bigcup_{k=1}^{\infty}(E_k\setminus B(x,r_{k+1})).$$

The set E_x is the union of the parts of the sets E_k that are contained in the annuli $B(x,r_k)\setminus B(x,r_{k+1})$. Clearly the condition (2.3) is satisfied and we only need to prove that x is a density point of E_x. If r is small, then $r_{k+1}<r\leq r_k$ for some large k and we need to show that

$$\frac{|B(x,r)\cap E_x|}{\omega_n r^n}\to 1.\quad\text{as } k\to\infty. \tag{2.5}$$

We have

$$\begin{aligned}
|B(x,r)\cap E_x| &\geq |(B(x,r)\cap E_k)\setminus B(x,r_{k+1})|+|E_{k+1}\setminus B(x,r_{k+2})|\\
&\geq \left(\omega_n r^n\left(1-\frac{1}{2^k}\right)-\omega_n r_{k+1}^n\right)+\left(\omega_n r_{k+1}^n\left(1-\frac{1}{2^{k+1}}\right)-\omega_n r_{k+2}^n\right)\\
&= \omega_n r^n\left(1-\frac{1}{2^k}\right)-\frac{\omega_n r_{k+1}^n}{2^{k+1}}-\omega_n r_{k+2}^n>\omega_n r^n\left(1-\frac{1}{2^{k-1}}\right),
\end{aligned}$$

because

$$r_{k+2}^n\leq\frac{r_{k+1}^n}{2^{k+1}}\qquad\text{and}\qquad r_{k+1}<r.$$

Now (2.5) follows easily. □

If the restriction of f to the line[16] $t \mapsto x + te_i$ is approximately differentiable at x we say that f has *approximate partial derivative* at x.

The following result given an important characterization of functions that are approximately differentiable a.e.

Theorem 2.7 (Stepanov-Whitney). *Let $f : E \to \mathbb{R}$ be a measurable function defined on a measurable set $E \subset \mathbb{R}^n$. Then the following conditions are equivalent.*

(a) *The function f has approximate partial derivatives a.e. in E;*
(b) *The function f is approximately differentiable a.e. in E;*
(c) *For every $\varepsilon > 0$ there is a locally Lipschitz function $g : \mathbb{R}^n \to \mathbb{R}$ such that;*

$$|\{x \in E : \; f(x) \neq g(x)\}| < \varepsilon;$$

(d) *For every $\varepsilon > 0$ there is a function $g \in C^1(\mathbb{R}^n)$ such that*

$$|\{x \in E : \; f(x) \neq g(x)\}| < \varepsilon;$$

If in addition the set E has finite measure, we can take g in (c) to be globally Lipschitz on $\mathbb{R}^n$.

Proof. The implications (d)⇒(c) and (b)⇒(a) are obvious. For the implication (c)⇒(b) just observe that the function f is approximately differentiable a.e. in the set $\{x \in E : f(x) = g(y)\}$, namely at the density point of the set that are points of differentiability of g.[17] Now (b) follows since we can exhaust the set E with such sets up to a set of measure zero. The implication from (a) to (b) is a result of Stepanov and the proof can be found in [**?**]. We will not present it here. The implication from (c) to (d) is a direct consequence of Theorem 2.2. Hence we only need to prove the implication from (b) to (c). This will follow from the next lemma which is of independent interest. This lemma is somewhat similar to an argument used in the proof of the Stepanov theorem (Theorem 1.8).

Lemma 2.8. *Let $f : E \to \mathbb{R}$ be a measurable function defined on a measurable set $E \subset \mathbb{R}^n$. Let $A \subset E$ the set of all points of approximate differentiability of f. Then A is the union of countably many sets E_i, $A = \bigcup_{i=1}^{\infty} E_i$ such that f restricted to E_i is Lipschitz continuous.*

[16] e_i is the direction of the ith coordinate.

[17] Note also that $\mathrm{ap}Df(x) = Dg(x)$ at such points.

Proof. For positive integers k, ℓ let $E_{k,\ell}$ be the set of all points $x \in A$ such that the following two conditions are satisfied:

$$|f(x)| \leq k \quad \text{and} \quad \frac{|f(y)-f(x)|}{|y-x|} \leq k \quad \text{when } |x-y| < 1/\ell \text{ and } y \in E_x,$$

$$|B(x,r) \cap E_x| > \frac{2^n}{2^n+1}|B(x,r)| \quad \text{if } 0 < r < 1/\ell.$$

It follows from (2.3) that $A = \bigcup_{k,\ell} E_{k,\ell}$. We will prove that $f|_{E_{k,\ell}}$ is Lipschitz continuous. Let $x, y \in E_{k,\ell}$. If $|x-y| \geq 1/(3\ell)$, then $|f(x)-f(y)| \leq 2k \leq 6k\ell|x-y|$. Thus we may assume that $|x-y| < 1/(3\ell)$. Let $r = |x-y|$. Then $B(x,r) \subset B(y,2r)$. Since $2r < 1/\ell$ we have

$$|B(y,2r) \cap E_x| \geq |B(x,r) \cap E_x| > \frac{2^n}{2^n+1}\omega_n r^n \quad \text{and} \quad |B(y,2r) \cap E_y| > \frac{2^n}{2^n+1}\omega_n (2r)^n.$$

Since

$$|B(y,2r) \cap E_x| + |B(y,2r) \cap E_y| > \frac{2^n}{2^n+1}\left(\omega_n r^n + \omega_n (2r)^n\right) = \omega_n (2r)^n = |B(y,2r)|,$$

there is $z \in B(y,2r) \cap E_x \cap E_y$. Clearly $|y-z| < 2r = 2|x-y| < 1/\ell$ and $|x-z| \leq |x-y| + |y-z| < 3|x-y| < 1/\ell$. Since $y \in E_{k,\ell}$, $z \in E_y$ and $|y-z| < 1/\ell$ the definition of the set $E_{k,\ell}$ yields $|f(z)-f(y)| \leq k|y-z| < 2k|x-y|$. Similarly $x \in E_{k,\ell}$, $z \in E_x$ and $|x-z| < 1/\ell$ gives $|f(z)-f(x)| \leq k|z-x| < 3k|x-y|$. Thus

$$|f(x)-f(y)| \leq |f(x)-f(z)| + |f(z)-f(y)| \leq 5k|x-y|.$$

The proof is complete. □

Now we can return to the proof of the implication from (b) to (c). That proof will give also global Lipschitz continuity of g in the case in which E has finite measure. Actually we can assume that E has finite measure. The general case will follow from this one. Indeed, if the measure is infinite we divide the set E into bounded sets contained in the unit cubes $\{Q_i\}_{i=1}^\infty$ with integer vertices. The function $f|_{Q_i \cap E}$ can be approximated by a Lipschitz function g_i up to a set of measure $\varepsilon/2^i$. By multiplying the function g_i by a smooth function compactly supported in Q_i that equals 1 on the substantial part of the cube we can further assume that the function g_i is supported in cube Q_i.[18] Since the functions g_i have disjoint supports, the function $g = \sum_{i=1}^\infty g_i$ is locally Lipschitz and coincides with f outside a set of measure less than ε.

[18] That will perhaps increase the Lipschitz constant of g_i enormously, but it does not matter. We are looking only for a locally Lipschitz function.

Thus assume that the set E has finite measure. Let $A = \bigcup_{i=1}^{\infty} E_i$ be a decomposition of the set of points of approximate differentiability as described in the lemma. By removing unnecessary parts of the sets E_i we may assume that the sets E_i are pairwise disjoint. Let $K_i \subset E_i$ be compact and such that $|E_i \setminus K_i| < \varepsilon/2^{i+1}$. Then $|A \setminus \bigcup_i K_i| < \varepsilon/2$ and hence there is N such that $|A \setminus \bigcup_{i=1}^{N} K_i| < \varepsilon$.[19] Each of the functions $f|_{K_i}$ is Lipschitz continuous. The sets K_i are disjoint and hence the distance between K_i and K_j, $i \neq j$ is always positive. This easily implies that the function f restricted to the set $\bigcup_{i=1}^{N} K_i$ is Lipschitz continuous. Now (c) follows from the McShane theorem (Theorem 1.11). $\square$

Department of Mathematics, University of Pittsburgh, 301 Thackeray Hall, Pittsburgh, PA 15260, USA, hajlasz@pitt.edu

[19] This is where we use the assumption that $|A| = |E| < \infty$.