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 every $\varepsilon > 0$ there is $\delta > 0$ such that if $(x_1, x_1+h_1), \dots, (x_k, x_k+h_k)$ 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., $F'\in L^1([a,b])$ 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 theorem.

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 $f\in L^1_{\rm loc}(\Omega)$, 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

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

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

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

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 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

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

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

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

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

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 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 L-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

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

is 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 $y\in A$, 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 $f\in L^1_{\rm loc}(\mathbb{R}^n)$, then for almost every x we have[2]

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

[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 lim sup 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

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

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

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

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]

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

[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 $x + y \notin E_{k,\ell}$.

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. □

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. □

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

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

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

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

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

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.

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 mapping $f : \mathbb{R}^n \supset \Omega \to X$ 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 mapping $f : \Omega \to \mathbb{R}^m$ 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 $v \in \mathbb{R}^n$

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

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

[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 $\mathbb{R}^n$.

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.

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, $|d(x, x_i) - d(y, x_i)| \le d(x, y)$ we conclude that $\|\kappa(x) - \kappa(y)\|_\infty \le d(x, y)$ 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 $\|\kappa(x) - \kappa(y)\|_\infty \ge d(x, y)$. $\square$

Remark 1.19. The embedding $\kappa : X \to \ell^\infty$ 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 $\ell^\infty = (\ell^1)^*$.

[7] What a disappointment.

[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. □

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.

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

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 $x \in \Omega$ and 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 $\mathbb{R}^n$. Indeed,

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

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 $g \mapsto \nabla f_g(x)$.

[10] Check that f_g is Lipschitz continuous.

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

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

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

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

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

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

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 $v\in S^{n-1}$

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

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

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

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

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

$$\limsup_{t\to 0}\left\| \frac{f(x+tv)-f(x)}{t}\right\|_Y \leq \limsup_{t\to 0} \fint_0^t \|Df(x+\tau v)v\|_Y\, d\tau = \|Df(x)v\|_Y$$

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.

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

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

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

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*

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

if and only if

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

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 how 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 $f : \Omega \to \mathbb{R}$ as 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 $f_i = \varphi_i f$. 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 $F(x) \neq f(x)$,

[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

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. $\square$

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 $E_x \subset E$ 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

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

as a function on $\mathbb{R}^n$ equals to the characteristic function of Ω which is not continuous at the boundary of Ω.

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 $r_{k+1} < r \le r_k$ 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

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

because

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

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 gives 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;*

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

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

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

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

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

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 $A \subset E$ be 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.*

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:

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

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

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 $|x - y| \ge 1/(3\ell)$, then $|f(x) - f(y)| \le 2k \le 6k\ell|x - y|$. 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

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

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

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

The proof is complete. $\square$

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

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

[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.

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.

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). $\square$

3. The area and the co-area formulas

3.1. **Hausdorff measure.** In this section we will recall the definition of the Hausdorff measure and we will state some of its basic properties. A more detailed discussion is postponed to Section **??**.

Let $\omega_s = \pi^{s/2}/\Gamma(1+\frac{s}{2})$, $s \geq 0$. If $s = n$ is a positive integer, then ω_n is volume of the unit ball in $\mathbb{R}^n$. Let X be a metric space. For $\varepsilon > 0$ and $E \subset X$ we define[20]

$$\mathcal{H}^s_\varepsilon(E) = \inf \frac{\omega_s}{2^s} \sum_{i=1}^{\infty} (\operatorname{diam} A_i)^s$$

where the infimum is taken over all possible coverings

$$E \subset \bigcup_{i=1}^{\infty} A_i \quad \text{with } \operatorname{diam} A_i \leq \varepsilon.$$

Since the function $\varepsilon \mapsto \mathcal{H}^s_\varepsilon(E)$ is nonincreasing, the limit

$$\mathcal{H}^s(E) = \lim_{\varepsilon \to 0} \mathcal{H}^s_\varepsilon(E)$$

exists. $\mathcal{H}^s$ is called the *Hausdorff measure*. It is easy to see that if $s = 0$, $\mathcal{H}^0$ is the counting measure.

The Hausdorff content $\mathcal{H}^s_\infty(E)$ is defined as the infimum of $\sum_{i=1}^{\infty} r_i^s$ over all coverings

$$E \subset \bigcup_{i=1}^{\infty} B(x_i, r_i)$$

of E by balls of radii r_i. It is an easy exercise to show that $\mathcal{H}^s(E) = 0$ if and only if $\mathcal{H}^s_\infty(E) = 0$. Often it is easier to use the Hausdorff content to show that the Hausdorff measure of a set is zero, because one does not have to worry about the diameters of the sets in the covering. The Hausdorff content is an outer measure, but very few sets are actually measurable, and it is not countably additive on Borel sets. This is why $\mathcal{H}^s_\infty$ is called content, but not measure.

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

[20] If $B \subset \mathbb{R}^n$ is a ball, then $\frac{\omega_n}{2^n}(\operatorname{diam} B)^n = |B|$. This explains the choice of the coefficient $\omega_s/2^s$ in the definition of the Hausdorff measure.

Theorem 3.1. *$\mathcal{H}^s$ is a metric outer measure i.e. $\mathcal{H}^s(E \cup F) = \mathcal{H}^s(E) + \mathcal{H}^s(F)$ whenever E and F are arbitrary sets with* $\operatorname{dist}(E, F) > 0$. *Hence all Borel sets are $\mathcal{H}^s$ measurable.*

It is an easy exercise to prove that $\mathcal{H}^s$ is an outer measure. The fact that it is a metric outer measure follows from the observation that if $\varepsilon < \operatorname{dist}(E, F)/2$, we can assume that sets of diameter less than ε that cover E are disjoint from the sets of diameter less than ε that cover F. We leave details as an exercise. Finally measurability of Borel sets is a general property of metric outer measures.

The next result is very important and difficult. We will prove it in Section **??**

Theorem 3.2. *$\mathcal{H}^n$ on $\mathbb{R}^n$ coincides with the outer Lebesgue measure $\mathcal{L}^n$. Hence a set is $\mathcal{H}^n$ measurable if and only if it is Lebesgue measurable and both measures are equal on the class of measurable sets.*

This result generalizes to the case of the Lebesgue measure on submanifolds of $\mathbb{R}^n$. We will discuss it in the Subsection 3.3.

In what follows we will often use the Hausdorff measure notation to denote the Lebesgue measure.

Proposition 3.3. *If $f : X \supset E \to Y$ is a Lipschitz mapping between metric spaces, then $\mathcal{H}^s(f(E)) \leq L^s \mathcal{H}^s(E)$. In particular if $\mathcal{H}^s(E) = 0$, then $\mathcal{H}^s(f(E)) = 0$.*

This is very easy. Indeed if $A \subset E$, then $f(A)$ has diameter less than or equal to $L \operatorname{diam} A$, where L is the Lipschitz constant of f. This observation and the definition of the Hausdorff measure easily yields the result.

In particular, if $f : \mathbb{R}^n \supset E \to \mathbb{R}^m$ is a Lipschitz mapping and $|E| = 0$, then $\mathcal{H}^n(f(E)) = 0$. We will prove a stronger result which is known as the Sard theorem. A more general version of the Sard theorem will be discussed in Section **??**.

Theorem 3.4 (Sard)**.** *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^m$ be Lipschitz continuous and let*

$$\operatorname{Crit}(f) = \{x \in E : \operatorname{rank} \operatorname{ap}Df(x) < n\},$$

then $\mathcal{H}^n(f(\operatorname{Crit}(f))) = 0$.

In the proof we will need the so called $5r$*-covering lemma*. It is also called a *Vitali type covering lemma*. Here and in what follows by σB we denote a ball concentric with the ball B and σ times the radius.

Theorem 3.5 ($5r$-covering lemma)**.** *Let $\mathcal{B}$ be a family of balls in a metric space such that $\sup\{\operatorname{diam} B : B \in \mathcal{B}\} < \infty$. Then there is a subfamily of pairwise disjoint balls $\mathcal{B}' \subset \mathcal{B}$ such that*

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{B \in \mathcal{B}'} 5B \, .$$

If the metric space is separable, then the family $\mathcal{B}'$ is countable and we can arrange it as a sequence $\mathcal{B}' = \{B_i\}_{i=1}^\infty$, so

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{i=1}^\infty 5B_i \, .$$

Remark 3.6. Here $\mathcal{B}$ can be either a family of open balls or closed balls. In both cases the proof is the same.

Proof. Let $\sup\{\operatorname{diam} B:\ B\in\mathcal{B}\}=R<\infty$. Divide the family $\mathcal{B}$ according to the diameter of the balls
$$\mathcal{F}_j=\{B\in\mathcal{B}:\frac{R}{2^j}<\operatorname{diam} B\leq\frac{R}{2^{j-1}}\}\,.$$
Clearly $\mathcal{B}=\bigcup_{j=1}^\infty\mathcal{F}_j$. Define $\mathcal{B}_1\subset\mathcal{F}_1$ to be the maximal family of pairwise disjoint balls. Suppose the families $\mathcal{B}_1,\ldots,\mathcal{B}_{j-1}$ are already defined. Then we define $\mathcal{B}_j$ to be the maximal family of pairwise disjoint balls in
$$\mathcal{F}_j\cap\{B:\ B\cap B'=\emptyset\text{ for all }B'\in\bigcup_{i=1}^{j-1}\mathcal{B}_i\}\,.$$
Next we define $\mathcal{B}'=\bigcup_{j=1}^\infty\mathcal{B}_j$. Observe that every ball $B\in\mathcal{F}_j$ intersects with a ball in $\bigcup_{i=1}^j\mathcal{B}_j$. Suppose that $B\cap B_1\neq\emptyset$, $B_1\in\bigcup_{i=1}^j\mathcal{B}_i$. Then
$$\operatorname{diam} B\leq\frac{R}{2^{j-1}}=2\cdot\frac{R}{2^j}\leq 2\ \operatorname{diam} B_1$$
and hence $B\subset 5B_1$. The proof is complete. □

Proof of the Sard theorem. Using the McShane extension (Theorem 1.11) we can assume that f is defined on all of $\mathbb{R}^n$ and replace the approximate derivative by the classical one. Indeed, the set of points in E where the approximate derivative exists, but the extension to $\mathbb{R}^n$ is not differentiable at these points has measure zero and this set is mapped onto a set of $\mathcal{H}^n$ measure zero.

Let Z be the set of points in $\mathbb{R}^n$ such that $Df(x)$ exists and $\operatorname{rank} Df(x)<n$. We need to show that $\mathcal{H}^n(f(Z))=0$. By splitting Z into bounded pieces we may assume that Z is contained in the interior of the unit cube Q.[21] For $L>\varepsilon>0$ and $x\in Z$ there is $r_x>0$ such that $B(x,r_x)\subset Q$ and
$$|f(y)-f(x)-Df(x)(y-x)|<\varepsilon r_x\quad\text{if }y\in B(x,5r_x).$$
Hence
$$\operatorname{dist}(f(y),W_x)\leq\varepsilon r_x\quad\text{for }y\in B(x,5r_x),$$
where $W_x=f(x)+Df(x)(\mathbb{R}^n)$ is an affine space through $f(x)$. Clearly $\dim W_x\leq n-1$. Thus
$$f(B(x,5r_x))\subset B(f(x),5Lr_x)\cap\{z:\ \operatorname{dist}(z,W_x)\leq\varepsilon r_x\}.\tag{3.1}$$
Since $\dim W_x=k\leq n-1$ we have that
$$\mathcal{H}^n_\infty(f(B(x,5r_x))\leq C\varepsilon L^{n-1}r_x^n,$$
where the constant C depends on n only. Indeed, the k dimensional ball $B(f(x),5Lr_x)\cap W_x$ can be covered by
$$C\left(\frac{Lr_x}{\varepsilon r_x}\right)^k\leq C\left(\frac{L}{\varepsilon}\right)^{n-1}$$

[21] Indeed, if each bounded piece of Z is mapped into a set of $\mathcal{H}^n$ measure zero, then Z is mapped into a set of measure zero.

balls of radius εr_x. Then balls with radii[22] $2\varepsilon r_x$ and the same centers cover the right hand side of (3.1). Thus

$$\mathcal{H}^n_\infty(f(B(x,5r_x)) \leq C\left(\frac{L}{\varepsilon}\right)^{n-1}(4\varepsilon r_x)^n = C'\varepsilon r_x^n L^{n-1}.$$

From the covering $Z \subset \bigcup_{x\in Z} B(x,r_x)$ we can select a family of pairwise disjoint balls $B(x_i,r_{x_i})$, $i=1,2,\ldots$ such that $Z\subset\bigcup_i B(x_i,5r_{x_i})$. We have

$$\mathcal{H}^n_\infty(f(Z)) \leq \sum_{i=1}^\infty \mathcal{H}^n_\infty(f(B(x_i,5r_{x_i})) \leq C\varepsilon L^{n-1}\sum_{i=1}^\infty r^n_{x_i} \leq C'\varepsilon L^{n-1},$$

because the balls $B(x_i,r_{x_i})$ are disjoint and contained in the unit cube; hence the sum of their volumes is less than one. Since ε can be arbitrarily small we conclude that $\mathcal{H}^n_\infty(f(Z))=0$ and thus $\mathcal{H}^n(f(Z))=0$. □

Exercise 3.7. *Show that if*

- $\mathcal{H}^s(E)<\infty$, *then* $\mathcal{H}^t(E)=0$ *for all* $t>s\geq 0$;
- $\mathcal{H}^s(E)>0$, *then* $\mathcal{H}^t(E)=\infty$ *for all* $0\leq t<s$.

Definition 3.8. The *Hausdorff dimension* is defined as follows. If $\mathcal{H}^s(E)>0$ for all $s\geq 0$, then $\dim_H(E)=\infty$. Otherwise we define

$$\dim_H(E)=\inf\{s\geq 0:\ \mathcal{H}^s(E)=0\}.$$

It follows from the exercise that there is $s\in[0,\infty]$ such that $\mathcal{H}^t(E)=0$ for $t>s$ and $\mathcal{H}^t(E)=\infty$ for $0<t<s$. Hausdorff dimension of E equals s. It also easily follows from Proposition 3.3 that Lipschitz mappings do not increase the Hausdorff dimension.

3.2. **Countably rectifiable sets.**

Definition 3.9. We say that a metric space X is *countably n-rectifiable* if there is a family of Lipschitz mappings $f_i:\mathbb{R}^n\supset E_i\to X$ defined on measurable sets such that

$$\mathcal{H}^n\left(X\setminus\bigcup_{i=1}^\infty f(E_i)\right)=0.$$

In particular we can talk about sets $X\subset\mathbb{R}^m$ that are countably n-rectifiable.

Clearly any Borel subset of a countably n-rectifiable set is countably n-rectifiable.

In other words X is countably n-rectifiable if it can be covered by countably many Lipschitz images of subsets of $\mathbb{R}^n$ up to a set of $\mathcal{H}^n$ measure zero. Since Lipschitz mappings map sets of finite $\mathcal{H}^n$ measure onto sets of finite $\mathcal{H}^n$ measure, the $\mathcal{H}^n$ measure on X is σ-finite and hence $\dim_H X\leq n$. We do not require the mappings f_i to be one-to-one and one can imagine that X can be very complicated. However as we will see, if X is a subset of $\mathbb{R}^m$ its structure is relatively simple.

[22]and hence diameter $4\varepsilon r_x$

Theorem 3.10. *A Borel set $E \subset \mathbb{R}^m$ is countably n-rectifiable, $m \geq n$, if and only if there is a sequence of n-dimensional C^1-submanifolds $\{\mathcal{M}_i\}_{i=1}^\infty$ of $\mathbb{R}^m$ such that*

$$\mathcal{H}^n\left(E \setminus \bigcup_{i=1}^\infty \mathcal{M}_i\right) = 0. \tag{3.2}$$

Proof. Clearly the condition (3.2) is sufficient for the countable n-rectifiability and we need to prove its necessity. Each mapping $f_i : E_i \to \mathbb{R}^m$ can be approximated by C^1-mappings in the sense of Theorem 2.7(d). Using a sequence of such C^1 maps we can approximate f_i up to a set of measure zero. Since sets of measure zero are mapped by Lipschitz maps to sets of measure zero, we can simply assume that the mappings $f_i : \mathbb{R}^n \to \mathbb{R}^m$ are C^1 and

$$\mathcal{H}^n\left(E \setminus \bigcup_{i=1}^\infty f_i(\mathbb{R}^n)\right) = 0.$$

A neighborhood of any point in $\mathbb{R}^n$ where $\operatorname{rank} Df_i = n$ is mapped to a C^1-submanifold of $\mathbb{R}^m$ and the remaining set of points where $\operatorname{rank} Df_i < n$ in mapped to a set of $\mathcal{H}^n$ measure zero by Theorem 3.4. $\square$

Definition 3.11. We say that a measurable mappings $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^n$, has the *Lusin property N* if for any measurable set $A \subset \Omega$ we have

$$|A| = 0 \quad \Rightarrow \quad |f(A)| = 0.$$

More generally we say that a measurable mapping $f : \mathbb{R}^n \supset A \to X$ to a metric space has the *Lusin property N* if for any measurable set $E \subset A$ we have

$$|A| = 0 \quad \Rightarrow \quad \mathcal{H}^n(f(E)) = 0.$$

Exercise 3.12. *Prove that a measurable mapping $f : \mathbb{R}^n \supset \Omega \to \mathbb{R}^n$ maps Lebesgue measurable sets onto Lebesgue measurable sets if and only if it has the Lusin property N.*

For example Lipschitz mappings have the Lusin property, Proposition 3.3.

Theorem 3.13. *Let $f : \mathbb{R}^n \supset E \to \mathbb{R}^m$ be an a.e. approximately differentiable mapping with the Lusin property N, and let*

$$\mathrm{Crit}\,(f) = \{x \in E : \operatorname{rank} \mathrm{ap}Df(x) < n\},$$

then $\mathcal{H}^n(f(\mathrm{Crit}\,(f))) = 0$.

Indeed, this result is a direct consequence of Lemma 2.8, the Sard theorem, and the Lusin property of f. A similar argument yields

Proposition 3.14. *$X \subset \mathbb{R}^m$, $m \geq n$ is countably n-rectifiable if and only if there are a.e. approximately differentiable mappings $f_i : \mathbb{R}^n \supset E_i \to \mathbb{R}^m$ with the Lusin property N such that*

$$\mathcal{H}^n\left(X \setminus \bigcup_{i=1}^\infty f(E_i)\right) = 0.$$

Proposition 3.15. *$E \subset \mathbb{R}^m$, $m \geq n$ is countably n-rectifiable if and only if there is a locally Lipschitz map $f : \mathbb{R}^n \to \mathbb{R}^m$ such that $\mathcal{H}^n(E \setminus f(\mathbb{R}^n)) = 0$.*

Indeed, we can assume in the definition of a countably n-rectifiable set that the sets E_i are contained in a unit cube. We can place such sets in disjoint unit cubes in $\mathbb{R}^n$ that are separated by a positive distance. On each cube we apply the McShane extension to the mapping $f : E_i \to \mathbb{R}^m$. Then we glue the mappings to form a locally Lipschitz mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ by multiplying the extension of f_i by a cut-off function[23] that equals 1 on the unit cube that contains E_i. Note that the result is not true for countably rectifiable subsets of metric spaces, because in such a general setting the McShane theorem is not available.

3.3. **The area formula.** Recall the classical change of variable s formula.

Theorem 3.16. *Let $\Phi : \Omega \to \mathbb{R}^n$ be a C^1 diffeomorphism between domains $\Omega \subset \mathbb{R}^n$ and $\Phi(\Omega) \subset \mathbb{R}^n$. If $f : \Omega \to [0, \infty]$ is a nonnegative measurable function or if $f|J_\Phi| \in L^1(\Omega)$ is integrable, then*

$$\int_{\Phi(\Omega)} f(\Phi^{-1}(y))\, dy = \int_\Omega f(x)|J_\Phi(x)|\, dx,$$

where $J_\Phi(x) = \det D\Phi(x)$ is the Jacobian of the diffeomorphism Φ.

In the case in which the function f is defined on $\Phi(\Omega)$ we have

$$\int_{\Phi(\Omega)} f(y)\, dy = \int_\Omega (f \circ \Phi)(x)|J_\Phi(x)|\, dx,$$

where we assume that $f \geq 0$ or that $f \in L^1(\Omega)$. Theorem 3.16 generalizes to the case of integration over an n-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^m$, $m \geq n$. A neighborhood of any point in $\mathcal{M}$ can be represented as the image of a parametrization. Recall that a *parametrization* of $\mathcal{M}$ is a one-to-one mapping

$$\Phi : \mathbb{R}^n \supset \Omega \to \mathbb{R}^m, \quad \Phi(\Omega) \subset \mathcal{M}$$

of class C^1 such that $\operatorname{rank} Dg(x) = n$ for all $x \in \Omega$.

Observe that $\det(D\Phi)^T D\Phi$ is the Gramm determinant of vectors $\partial\Phi/\partial x_i$ and hence $\sqrt{\det(D\Phi)^T D\Phi(x)}$ is the n-dimensional volume of the parallelepiped with edges $\partial\Phi(x)/\partial x_i$. Thus it is natural to define

$$|J_\Phi(x)| = \sqrt{\det(D\Phi)^T D\Phi(x)},$$

even if $m > n$. Note that this definition is consistent with the standard definition of the absolute value of the Jacobian when $m = n$.

In the case when $m = n$ and Φ is a diffeomerphism, the change of variables formula implies that the Lebesgue measure $|E|$ of a set $E \subset \Phi(\Omega)$ equals

$$|E| = \int_{\Phi^{-1}(E)} |J_\Phi(x)|\, dx.$$

[23] We multiply each component of the function f_i by a cut-off function.

If $\Phi : \mathbb{R}^n \supset \Omega \to \mathcal{M} \subset \mathbb{R}^m$ is a parametrization, we define the Lebesgue measure (surface measure) $\sigma(E)$ of a set $E \subset \Phi(\Omega)$ by the formula (3.3), i.e.

$$\sigma(E) = \int_{\Phi^{-1}(E)} |J_\Phi(x)|\, dx. \tag{3.3}$$

If $E \subset \mathcal{M}$ is not necessarily contained in the image of a single parametrization, we divide it into small pieces that are contained in the images of parametrizations and we add the measures. One only needs to observe that the measure of a set does not depend on the choice of a parametrization. Indeed, suppose that $E \subset \Phi_1(\Omega_1) \cap \Phi_2(\Omega_2)$. By taking smaller domains we can assume that $\Phi_1(\Omega_1) = \Phi_2(\Omega_2)$. Then $\Phi_1^{-1} \circ \Phi_2 : \Omega_2 \to \Omega_1$ is a diffeomorphism and the change of variables formula easily implies that

$$\int_{\Phi_1^{-1}(E)} |J_{\Phi_1}(x)|\, dx = \int_{\Phi_2^{-1}(E)} |J_{\Phi_2}(x)|\, dx.$$

Note that the formula (3.3) can be written as

$$\int_{\Phi(\Omega)} f(y)\, d\sigma(y) = \int_\Omega (f \circ \Phi)(x) |J_\Phi(x)|\, dx,$$

where f is the characteristic function of the set E. Since measurable functions can be approximated by simple functions which are linear combinations of characteristic functions, standard limiting procedure yields

Theorem 3.17. *Let $\Phi : \mathbb{R}^n \supset \Omega \to \mathcal{M} \subset \mathbb{R}^m$, $m \geq n$ be a parametrization of an n dimensional submanifold $\mathcal{M} \subset \mathbb{R}^m$. If $f : \Phi(\Omega) \to [0, \infty]$ is a nonnegative measurable function or if $f \in L^1(\Phi(\Omega))$ is integrable, then*

$$\int_{\Phi(\Omega)} f(y)\, d\sigma(y) = \int_\Omega (f \circ \Phi)(x) |J_\Phi(x)|\, dx.$$

For $f \geq 0$ on Ω and for $f|J_\Phi| \in L^1(\Omega)$ the change of variables formula takes the form

$$\int_{\Phi(\Omega)} (f \circ \Phi^{-1})(y)\, d\sigma(y) = \int_\Omega f(x) |J_\Phi(x)|\, dx. \tag{3.4}$$

According to Theorem 3.2 the Lebesgue measure in $\mathbb{R}^n$ coincides with the Hausdorff measure $\mathcal{H}^n$. One can prove that the surface measure on $\mathcal{M}$ also coincides with the Hausdorff measure $\mathcal{H}^n$ defined either with respect to the Euclidean metric of $\mathbb{R}^m$ restricted to $\mathcal{M}$ or with respect to the natural Riemannian metric on $\mathcal{M}$. We will not prove this fact, but this result should not be surprising; $\mathcal{M}$ is locally very well approximated by tangent spaces and this approximation allows one to deduce the result from Theorem 3.2. In particular in both theorems Theorem 3.16 and 3.17 we can replace dy and $d\sigma(y)$ by $d\mathcal{H}^n(y)$.

The purpose of this section is to prove a far reaching generalization the change of variables formula.

If $\Phi : \mathbb{R}^n \supset E \to \mathbb{R}^m$, $m \geq n$ is approximately differentiable a.e., then we can formally define the Jacobian of Φ at almost every point of E by

$$|J_\Phi(x)| = \sqrt{\det (D\Phi)^T D\Phi(x)}$$

Theorem 3.18 (Area formula)**.** *Let $\Phi:\mathbb{R}^n\supset E\to\mathbb{R}^m$, $m\geq n$ be approximately differentiable a.e. Then we can redefine it on a set of measure zero in such a way that the new mapping satisfies the Lusin property N. If Φ is approximately differentiable a.e., satisfies the Lusin property N and $f:E\to[0,\infty]$ is measurable or $f|J_\Phi|\in L^1(E)$, then*

$$\int_E f(x)|J_\Phi(x)|\,dx=\int_{\Phi(E)}\left(\sum_{x\in\Phi^{-1}(y)}f(x)\right)\,d\mathcal{H}^m(y). \tag{3.5}$$

Here we do not assume that the mapping Φ is one-to-one and this is why we have the sum on the right hand side, just to compensate the fact that the point y is the image of every point x in the set $\Phi^{-1}(y)$. Note that since $\mathcal{H}^0$ is the counting measure formula (3.5) can be rewritten as

$$\int_E f(x)|J_\Phi(x)|\,d\mathcal{H}^n(x)=\int_{\Phi(E)}\left(\int_{\Phi^{-1}(y)}f(x)\,d\mathcal{H}^0(x)\right)\,d\mathcal{H}^m(y)\,.$$

The reason why we want to write it this way will be apparent when we will discuss the co-area formula.

Proof. Lemma 2.8 shows that away from a set Z of measure zero Φ has the Lusin property since it consists of Lipschitz pieces. Now if we modify Φ on the set Z and send the set to a single point, a new mapping $\tilde{\Phi}$ will have the Lusin property and it will be equal to Φ almost everywhere. This proves the first part of the theorem. Assume now that Φ has the Lusin property. Note that if we remove from E a subset of measure zero both sides of (3.5) will not change its value. It is obvious for the left hand side, but regarding the right hand side it follows from the Lusin property of Φ. We can also remove the subset of E where $J_\Phi=0$. According to Theorem 3.13, Φ maps this set onto a set of measure zero and hence both sides of (3.5) will not change its value after such a removal. This combined with Theorem 2.7 allows us to assume that there are disjoint subsets $E_i\subset E$ such that $\bigcup_i E_i=E$ and C^1 mappings $\Phi_i:\mathbb{R}^n\to\mathbb{R}^m$ such that $\Phi_i=\Phi$ on E_i, $D\Phi_i=\mathrm{ap}D\Phi$ on E_i, $\operatorname{rank}D\Phi_i=n$ on E_i. Dividing the sets into small pieces, if necessary, we can also assume that Φ_i is one-to-one in an open set containing E_i, i.e. it is a parametrization of an n-dimensional submanifold of $\mathbb{R}^m$ on that open set. According to the classical change of variables formula (3.4) we have[24]

$$\int_{E_i}f(x)|J_{\Phi_i}(x)|\,dx=\int_{\Phi_i(E_i)}f(\Phi_i^{-1}(y))\,d\mathcal{H}^n(y)$$

which yields

$$\begin{aligned}\int_E f(x)|J_\Phi(x)|\,dx &= \sum_{i=1}^\infty\int_{E_i}f(x)|J_{\Phi_i}(x)|\,dx=\sum_{i=1}^\infty\int_{\Phi_i(E_i)}f(\Phi_i^{-1}(y))\,d\mathcal{H}^n(y)\\ &= \int_{\Phi(E)}\left(\sum_{x\in\Phi^{-1}(y)}f(x)\right)\,d\mathcal{H}^n(y).\end{aligned}$$

[24]We replace f in (3.4) by $f\chi_{E_i}$.

Indeed, if $f \geq 0$ we can change the order of integration ans summation by the monotone convergence theorem. In the case of $f \in L^1$ we consider separately the positive and negative parts of f. $\square$

Remark 3.19. It is necessary to require that Φ has the Lusin property. Indeed, if Φ maps a set of measure zero onto a set of positive measure, and $f = 1$, then the left hand side of the formula in Theorem 3.18 equals zero, but the right hand side is positive.

If f is a measurable function on $\mathbb{R}^m$, and $\Phi : E \to \mathbb{R}^m$ is approximately differentiable a.e. and has the Lusin property N, then Theorem 3.18 applies to the function $f \circ \Phi$ which is defined on E. Note that $f \circ \Phi$ is constant on the set $\Phi^{-1}(y)$ and hence the area formula takes the form

$$\int_E (f \circ \Phi)(x)|J_\Phi(x)|\, dx = \int_{\Phi(E)} f(y) N_\Phi(y, E)\, d\mathcal{H}^n(y),$$

where

$$N_\Phi(y, E) = \#(\Phi^{-1}(y) \cap E)$$

is the cardinality of the set $\Phi^{-1}(y) \cap E$. The function $N_\Phi(\cdot, E)$ is called the *Banach indicatrix* of Φ. This formula is true under the assumption that $f \geq 0$ or under the integrability assumption of $(f \circ \Phi)|J_\Phi|$. More precisely if $(f \circ \Phi)(x)|J_\Phi(x)|$ is integrable on E or if $f(y)N_\Phi(y, E)$ is integrable on $\Phi(E)$, then the other function is integrable too and the formula is true.

Remark 3.20. Suppose that $\Phi : Q \to Q$ is a homeomorphism of the unit cube $Q = [0, 1]^n$ that is identity on the boundary of the cube. Assume also that Φ is approximately differentiable a.e. and has the Lusin property. In this case the change of variables formula shows that $\int_Q |J_\Phi| = 1$. Since Φ is an orientation preserving homeomorphism, is it true that $J_\Phi \geq 0$ a.e.? Surprisingly, one can find such a homeomorphism with the property that $J_\Phi = -1$ a.e. or that it is positive on a subset of a cube and negative on another subset and it is reasonable to conjecture that the only real constrain for a construction of a mapping Φ with prescribed Jacobian is the condition that integral of $|J_\Phi|$ over the cube equals one.

The area formula generalizes to the case of mappings between Riemannian manifolds. Submanifolds of Euclidean spaces are examples of Riemannian manifolds.

Theorem 3.21. *The statement of Theorem 3.18 remains true if we replace $\mathbb{R}^n$ and $\mathbb{R}^m$ by n-dimensional and m-dimansional Riemannian manifolds respectively.*

3.4. **The co-area formula.** The area formula is a generalization of the change of variable formula to the case of mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$, where $m \geq n$. Surprisingly, it is also possible to generalize to change of variables formula to the case when $m \leq n$; this is so called the *co-area formula*. First we need to generalize the Jacobian to the case of mappings $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, $m \leq n$. Suppose that Φ is differentiable at $x \in \mathbb{R}^n$. If $m < n$, then

$$\sqrt{\det(D\Phi)^T(D\Phi)(x)} = 0 \tag{3.6}$$

because this is a formula for the n-dimensional volume of a parallelepiped which in our situation has the dimension $\leq m < n$. That means (3.6) is not a good notion of the Jacobian when$m < n$. Assume that rank of $D\Phi(x)$ is maximal, i.e. rank $D\Phi(x) = m \leq n$. If B is a ball in the tangent space $T_x\mathbb{R}^n$ centered at the origin, then $D\Phi(x)(B)$ is a non-degenerate m-dimensional ellipsoid in $T_{\Phi(x)}\mathbb{R}^m$. The kernel $\ker D\Phi(x)$ is an $n-m$ dimensional linear subspace of $T_x\mathbb{R}^n$ and $D\Phi(x)$ is a composition of two mappings; first we take the orthogonal projection of $T_x\mathbb{R}^n$ onto the m-dimensional space $(\ker D\Phi(x))^\perp$ and then we compose it with the linear isomorphism of m-dimensional spaces

$$D\Phi(x)|_{(\ker D\Phi(x))^\perp} : (\ker D\Phi(x))^\perp \to T_{\Phi(x)}\mathbb{R}^m. \tag{3.7}$$

Now we define $|J_\Phi(x)|$ as the absolute value of the Jacobain of the mapping (3.7), i.e. $|J_\Phi(x)|$ is factor by which the linear mapping (3.7) changes volume. Geometrically speaking the ellipsoid $D\Phi(x)(B)$ is the image of the m-dimensional ball $B \cap (\ker D\Phi(x))^\perp$ and hence

$$|J_\Phi(x)| = \frac{\mathcal{H}^m(D\Phi(x)(B))}{\mathcal{H}^m(B \cap (\ker D\Phi(x))^\perp)}\,.$$

If $\operatorname{rank} D\Phi(x) < n$ we set $|J_\Phi(x)| = 0$. Although we defined the Jacobian in geometric terms, there is a simple algebraic formula for $|J_\Phi(x)|$ which follows from the polar decomposition of the linear mapping $D\Phi(x)$.

$$|J_\Phi(x)| = \sqrt{\det(D\Phi)(D\Phi)^T(x)}.$$

Note that this is not the same formula as (3.6). However, the two formulas give the same value when $m = n$.

Exercise 3.22. *Prove this formula using the polar decomposition of $D\Phi(x)$.*

There is one more geometric interpretation of $|J_\Phi(x)|$ when $m \le n$ which easily follows from our geometric definition. Namely $|J_\Phi(x)|$ equals the supremum of m-dimensional measures of all ellipsoids $D\Phi(x)(B)$, where the supremum is over all m-dimensional balls B in $T_x\mathbb{R}^n$ of volume 1. This reminds us of the the geometric interpretation of the length of the gradient of a real valued function as the maximal rate of of change of a function. The function has maximal growth in the direction the gradient which is orthogonal to $\det D\Phi$. In our case the maximal growth of the m-dimensional measure of m-dimensional balls in $T_x\mathbb{R}^n$ is also in the direction orthogonal to $\ker D\Phi(x)$, see (3.7). This is the right intuition. If $\Phi : \mathbb{R}^n \to \mathbb{R}$, i.e. $m = 1$, one can easily see that

$$|J_\Phi(x)| = \sqrt{\det(D\Phi)(D\Phi)^T(x)} = |\nabla\Phi(x)|.$$

Now we can state the co-area formula. We will actually state both area and co-area formula in one theorem, because it will help to see similarities and differences between the two formulas.

Theorem 3.23 (The area and the co-area formulas)**.** *Let $\Phi : \mathbb{R}^n \supset E \to \mathbb{R}^m$ be a Lipschitz mapping defined on a measurable set $E \subset \mathbb{R}^n$. Let $f \ge 0$ be a measurable function on E or let $f|J_\Phi| \in L^1(E)$. Then*

- *(Area formula) If $n \le m$, then*

$$\int_E f(x)|J_\Phi(x)|\,d\mathcal{H}^n(x) = \int_{\mathbb{R}^m}\left(\int_{\Phi^{-1}(y)} f(x)\,d\mathcal{H}^0(x)\right)d\mathcal{H}^m(y).$$

- *(Co-area formula) If $n \geq m$, then*

$$\int_E f(x)|J_\Phi(x)|\, d\mathcal{H}^n(x) = \int_{\mathbb{R}^m} \left(\int_{\Phi^{-1}(y)} f(x)\, d\mathcal{H}^{n-m}(x) \right) d\mathcal{H}^m(y).$$

We will not prove the co-area formula here, but we will show that it contains results like integration in the spherical coordinates and the Fubini theorem as special cases!

Recall that if $\Phi : \mathbb{R}^n \to \mathbb{R}$, then $|J_\Phi(x)| = |\nabla \Phi(x)|$. Taking $\Phi(x) = |x|$ we have $|J_\Phi(x)| = 1$ everywhere except at the origin. Since the image of Φ is $[0, \infty)$, the co-area formula reads as

$$\int_{\mathbb{R}^n} f(x)\, d\mathcal{H}^n(x) = \int_0^\infty \left(\int_{\partial B(0,r)} f(x)\, d\mathcal{H}^{n-1}(x) \right) dr$$

which is the formula for the integration in the spherical coordinates.

Let now $\Phi : \mathbb{R}^n \to \mathbb{R}^m$, $m < n$ be the projection on the first m coordinates $\Phi(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$. Then $|J_\Phi(x)| = 1$ and we have

$$\int_{\mathbb{R}^n} f(x)\, d\mathcal{H}^n(x) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^{n-m}} f(x_1, \ldots, x_n)\, dH^{n-m}(x_{m+1}, \ldots, x_n) \right) d\mathcal{H}^m(x_1, \ldots, x_m)$$

which is the Fubini theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary Lipschitz function. Taking $\Phi = f$ we have $J_\Phi(x)| = |\nabla f(x)|$; taking the function f in the co-area formula to be equal[25] 1 we have

$$\int_{\mathbb{R}^n} |\nabla f(x)|\, dx = \int_{-\infty}^\infty \mathcal{H}^{n-1}(\{f = t\})\, dx.$$

As an application of the co-area formula we will prove

Theorem 3.24. *If $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous, then for a.e. $y \in \mathbb{R}^m$, $\Phi^{-1}(y)$ is countably $(n-m)$-rectifiable.*

Proof. If $m > n$, then $\mathcal{H}^m(f(\mathbb{R}^n)) = 0$, so $\Phi^{-1}(y) = \emptyset$ for a.e. $y \in \mathbb{R}^n$ and the empty set is countably rectifiable. Thus we can assume that $m \leq n$. Assume for a moment that $\Phi \in C^1$. Then according to the implicit function theorem

$$\Phi^{-1}(y) \cap \{\operatorname{rank} D\Phi = m\}$$

is a C^1, $(n-m)$-dimensional submanifold of $\mathbb{R}^n$ and it follows from the co-area formula that

$$\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap \{\operatorname{rank} D\Phi < m\}) = 0 \quad \text{for } \mathcal{H}^m\text{-a.e. } y \in \mathbb{R}^m.$$

Thus for almost all y, Φ^{-1} is a manifold plus a set of measure zero. Hence it is countably $(n-m)$-rectifiable. To prove the result in the case in which Φ is Lipschitz it suffices to use Theorem 2.2 which reduces the problem to the C^1-case. $\square$

[25]This might be slightly confusing since we have a double meaning of f.

3.5. **The Eilenberg inequality.**

Definition 3.25. A metric space is said to be *boundednly compact* if bounded and closed sets are compact.

An important step in the proof of the co-area formula is the following

Theorem 3.26 (Eilenberg inequality)**.** *Let $\Phi : X \to Y$ be a Lipschitz mapping between boundedly compact metric spaces. Let $0 \leq m \leq n$ be real numbers.*[26] *Assume that $E \subset X$ is $\mathcal{H}^n$-measurable with $\mathcal{H}^n(E) < \infty$. Then*

(1) $\Phi^{-1}(y) \cap E$ *is $\mathcal{H}^{n-m}$-measurable for $\mathcal{H}^m$-almost all $y \in Y$.*
(2) $y \mapsto \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)$ *is $\mathcal{H}^m$-measurable.*

Moreover

$$\int_Y \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)\, d\mathcal{H}^m(y) \leq (\mathrm{Lip}\,(\Phi))^m \frac{\omega_m \omega_{n-m}}{\omega_n}\, \mathcal{H}^n(E).$$

Observe that the left hand side corresponds to the right hand side in the co-area formula with $f = 1$. Observe also that $|J_\Phi|$ can be estimated by $\mathrm{Lip}\,(\Phi)^m$, and then the integral of the Jacobian over E can be estimated from above by $\mathrm{Lip}\,(\Phi)^m \mathcal{H}^n(E)$. This shows a deep connection between the co-area formula and the Eilenberg inequality. Since we used the estimate from the above we only have an inequality and one cannot expect equality in the Eilenberg inequality. What is remarkable is that the Eilenberg inequality is true in a great generality of boundedly compact metric spaces where differentiable structure is not available. We will prove Theorem 3.26 under the additional assumption that $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.

The measurability of the function $y \mapsto \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)$ is far from being obvious and we will want to integrate this function before proving its measurablility. To do this we will have to use the upper Lebesgue integral.

Definition 3.27. For a nonnegative function $f : X \to [0, \infty]$ defined μ-a.e. on a measure space (X, μ) the upper Lebesgue integral is defined as

$$\int^* f\, d\mu = \inf \left\{ \int \phi\, d\mu :\ 0 \leq f \leq \phi \text{ and } \phi \text{ is } \mu\text{-measurable} \right\}.$$

We do not assume measurability of f. Clearly if f is measurable the upper Lebesgue ineqgral equals the Lebesgue integral.

An important property of the upper integral is that if $\int^* f\, d\mu = 0$, then $f = 0$, μ-a.e. and hence it is measurable. Indeed, there is a sequence $0 \leq f \leq \phi_n$ such that $\int \phi_n\, d\mu \to 0$. That means $\phi_n \to 0$ in $L^1(\mu)$. Taking a subsequence we get $\phi_{n_k} \to 0$, μ-a.e. which proves that $f = 0$, μ-a.e.

[26]Not necessarily integers.

Proof of Theorem 3.26 when $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$. For ever positive integer $k > 0$ there is a covering

$$E \subset \bigcup_{i=1}^{\infty} A_{ik}, \quad A_{ik} \text{ is closed}, \quad \operatorname{diam} A_{ik} < \frac{1}{k}$$

such that

$$\frac{\omega_n}{2^n} \sum_{i=1}^{\infty} (\operatorname{diam} A_{ik})^n \le \mathcal{H}^n(E) + \frac{1}{k}. \tag{3.8}$$

It follows directly from the definition of the Hausdorff measure that

$$\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E) \le \frac{\omega_{n-m}}{2^{n-m}} \liminf_{k\to\infty} \sum_{i=1}^{\infty} \operatorname{diam} (\Phi^{-1}(y) \cap A_{ik})^{n-m}. \tag{3.9}$$

For any set $A \subset X$ we have

$$\operatorname{diam} (\Phi^{-1}(y) \cap A) = \operatorname{diam} (\Phi^{-1}(y) \cap A)\chi_{\overline{\Phi(A)}}(y) \le (\operatorname{diam} A)^{n-m}\chi_{\overline{\Phi(A)}}(y).$$

Hence (3.9) yields

$$\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E) \le \frac{\omega_{n-m}}{2^{n-m}} \liminf_{k\to\infty} \sum_{i=1}^{\infty} (\operatorname{diam} A_{ik})^{n-m}\chi_{\overline{\Phi(A_{ik})}}(y).$$

The function on the right hand side is measurable. Hence Fatou's lemma yields

$$\begin{aligned}
\int^*_{\mathbb{R}^m} \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)\, d\mathcal{H}^m(y) &\le \frac{\omega_{n-m}}{2^{n-m}} \liminf_{k\to\infty} \int_{\mathbb{R}^m} \sum_{i=1}^{\infty} (\operatorname{diam} A_{ik})^{n-m}\chi_{\overline{\Phi(A_{ik})}}(y)\, d\mathcal{H}^m(y) \\
&= \frac{\omega_{n-m}}{2^{n-m}} \liminf_{k\to\infty} \sum_{i=1}^{\infty} (\operatorname{diam} A_{ik})^{n-m}\mathcal{H}^m(\overline{\Phi(A_{ik})}).
\end{aligned}$$

If $p \in A_{ik}$, then

$$\overline{\Phi(A_{ik})} \subset \overline{B}(f(p), \operatorname{Lip}(\Phi) \operatorname{diam} A_{ik})$$

and hence

$$\mathcal{H}^m(\overline{\Phi(A_{ik})}) \le \frac{\omega_m}{2^m} (\operatorname{Lip}(\Phi))^m (\operatorname{diam} A_{ik})^m.$$

Thus

$$\begin{aligned}
\int^*_{\mathbb{R}^m} \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)\, d\mathcal{H}^m(y) &\le \frac{\omega_{n-m}}{2^{n-m}} \frac{\omega_m}{2^m} \frac{2^n}{\omega_n} \liminf_{k\to\infty} \frac{\omega_n}{2^n} \sum_{i=1}^{\infty} (\operatorname{diam} A_k)^n \\
&\le (\operatorname{Lip}(\Phi))^m \frac{\omega_{n-m}\omega_n}{\omega_n} \mathcal{H}^n(E),
\end{aligned}$$

by (3.8). It remains to prove the $\mathcal{H}^{n-m}$-measurability of the sets $\Phi^{-1}(y) \cap E$ for $\mathcal{H}^m$ almost all $y \in \mathbb{R}^m$ and the $\mathcal{H}^m$-measurability of the function $\varphi(y) = \mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E)$. Note that if $\mathcal{H}^n(E) = 0$, then $\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E) = 0$ for $\mathcal{H}^m$ almost every $y \in \mathbb{R}^m$ by the upper integral estimate. This observations shows that we can ignore subsets of E of $\mathcal{H}^n$ measure zero. Thus we can assume that E is the union of an increasing sequence of compact sets $E \bigcup_{k=1}^{\infty} E_k$, $E_k \subset E_{k+1}$. Note that the sets $\Phi^{-1}(y) \cap E_k$ are compact for every y and hence

$\Phi^{-1}(y)\cap E$ is Borel as the union of compact sets. Thus it suffices to prove that every function $y\mapsto\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E_k)$ is Borel measurable, because then the function

$$\varphi(y)=\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)=\lim_{k\to\infty}\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E_k)$$

will also be Borel. Hence we can assume that E is compact. It remains to prove that for every $t\in\mathbb{R}$ the set

$$\big\{y\in\mathbb{R}^m:\,\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)\leq t\big\} \tag{3.10}$$

is Borel. If $t<0$, then the set is empty, so we can assume that $t\geq 0$. Since the set in (3.10) can be written as

$$(\mathbb{R}^m\setminus\Phi(E))\cup\big(\Phi(E)\cap\big\{y\in\mathbb{R}^m:\,\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)\leq t\big\}\big)$$

and $\mathbb{R}^m\setminus\Phi(E)$ is open, it remains to prove that the set

$$\Phi(E)\cap\big\{y\in\mathbb{R}^m:\,\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)\leq t\big\}$$

is Borel. In the definition of the Hausdorff measure we may restrict to coverings by open sets. However this family of sets is uncountable and we would like to have a countable family of sets from which we would choose coverings. Let $\mathcal{F}$ be the family of all open sets in $\mathbb{R}^n$ that are finite unions of balls with rational centers and radii. The family $\mathcal{F}$ is countable and we claim that it can be used as the family of sets from which we choose coverings provided we define the Hausdorff measure of a compact set K. Indeed, first we cover the set K by open sets, $K\subset\bigcup_i U_i$ Each open set is the union of a family of balls with rational centers and radii. These balls form a covering of K and hence we can select a finite subcovering $K\subset\bigcup_{j=1}^N B_j$. Now we replace each set U_i by U_i' which is the union of all the balls B_j that are contained in U_i. We obtain a new covering $K\subset\bigcup_i U_i'$. Clearly $U_i'\subset U_i$ and $U_i'\in\mathcal{F}$.

Let $\mathcal{F}_i$ be the collection of *all* finite families $\{U_{i1},\ldots,U_{ik}\}\subset\mathcal{F}$ such that

$$\operatorname{diam}U_{ij}<\frac{1}{i},\ j=1,2,\ldots,\quad\text{and}\quad\frac{\omega_{n-m}}{2^{n-m}}\sum_{j=1}^k(\operatorname{diam}U_{ij})^{n-m}\leq t+\frac{1}{i}.$$

Note that each of the sets U_{ij} is the union of a finite family of balls. The family $\mathcal{F}_i$ is also countable. Clearly we define this family to deal with the coverings of the set $\Phi^{-1}(y)\cap E$ that satisfies $\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)\leq t$.

If $U\subset\mathbb{R}^n$ is open, then

$$\Phi(E)\cap\{y:\,\Phi^{-1}(y)\cap E\subset U\}=\Phi(E)\setminus\Phi(E\setminus U)$$

is Borel, because both of the sets $f(E)$ and $f(E\setminus U)$ are compact. In particular the set

$$V_i=\bigcup_{\{U_{i1},\ldots,U_{ik}\}\in\mathcal{F}_i}\left(\Phi(E)\setminus\Phi\left(E\setminus\bigcup_{j=1}^k U_{ij}\right)\right)$$

is Borel as a countable union over the entire family $\mathcal{F}_i$. We will prove that

$$\Phi(E)\cap\{y\in\mathbb{R}^m:\,\mathcal{H}^{n-m}(\Phi^{-1}(y)\cap E)\leq t\}=\bigcap V_i. \tag{3.11}$$

Clearly the set on the right hand side is Borel.

If $y \in E$ and $\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E) \le t$, then for any i we can find a covering

$$\Phi^{-1}(y) \cap E \subset U_{i1} \cup \ldots \cup U_{ik}, \quad \{U_{i1}, \ldots, U_{ik}\} \in \mathcal{F}_i.$$

Thus $y \notin \Phi(E \setminus \bigcup_{j=1}^k U_{ij})$ and hence $y \in V_i$. Since i can be chosen arbitrarily, $y \in \bigcap_{i=1}^\infty V_i$. On the other hand if $y \in \bigcap_{i=1}^\infty V_i$ then $y \in \Phi(E)$ and for all i, $y \in V_i$, i.e. there is $\{U_{i1}, \ldots, U_{ik}\} \in \mathcal{F}_i$ such that $y \notin \Phi(E \setminus \bigcup_{i=1}^k U_{ik})$, i.e. $\Phi^{-1}(y) \cap E \subset U_{i1} \cup \ldots \cup U_{ik}$, so

$$\mathcal{H}^{n-m}_{1/i}(\Phi^{-1}(y) \cap E) \le t + \frac{1}{i}.$$

Taking the limit as $i \to \infty$ we obtain $\mathcal{H}^{n-m}(\Phi^{-1}(y) \cap E) \le t$. The proof is complete. □

3.6. **Integral geometric measure.** We say that a metric space is *purely $\mathcal{H}^m$-unrectifiable* if for any Lipschitz mapping $f : \mathbb{R}^n \supset A \to X$ we have $\mathcal{H}^m(f(A)) = 0$. It easily follows from the definition that $E \subset \mathbb{R}^n$ is purely $\mathcal{H}^m$-unrectifiable if and only if for any countably $\mathcal{H}^m$ rectifiable set $F \subset \mathbb{R}^n$, $\mathcal{H}^m(E \cap F) = 0$.

Theorem 3.28. *If $\mathcal{H}^m(X) < \infty$, then there is a Borel countably rectifiable set $E \subset X$ such that $X \setminus E$ is purely $\mathcal{H}^m$-unrectifiable. Hence X has a decomposition into a rectifiable and a nonrectifiable parts $X = E \cup (X \setminus E)$. This decomposition is unique up to sets of $\mathcal{H}^m$-measure zero.*

Proof. Let M be the supremum of $\mathcal{H}^m(E)$ over all $\mathcal{H}^m$-countably rectifiable Borel sets $E \subset X$. Hence there are Borel countably $\mathcal{H}^m$-rectifiable sets $E_i \subset X$ such that $\mathcal{H}^m(E_i) > M - 1/i$. It is easily to see that $E = \bigcup_i E_i$ satisfies the claim of the theorem. Uniqueness is easy. □

Definition 3.29. Let $E \subset \mathbb{R}^n$ be a Borel set and let $1 \le m \le n$ be integers. If $m < n$, the *integral geometric measure* $\mathcal{I}^m$ of E is defined as

$$\mathcal{I}^m(E) = \frac{1}{\beta(n,m)} \int_{p \in O^*(n,m)} \int_{y \in \operatorname{Im} p} N_p(y, E)\, d\mathcal{H}^m(y)\, d\vartheta^*_{n,m}(p), \tag{3.12}$$

where $O^*(n,m)$ is the space of orthogonal projections p from $\mathbb{R}^n$ onto m-dimensional linear subspaces of $\mathbb{R}^n$, $\operatorname{Im} p$ is the image of the projection and $\vartheta^*_{n,m}$ is the Haar measure on $O^*(n,m)$ invariant under the action of $O(n)$, normalized to have total mass 1. Moreover $N_p(y, E)$ is the Banach indicatrix, i.e. $N_p(y, E) = \#(p^{-1}(y) \cap E)$. The coefficient $\beta(n,m)$ will be defined later. If $m = n$ we simply define $\mathcal{I}^m(E) = \mathcal{H}^m(E)$.

Thus roughly speaking $\mathcal{I}^m(E)$ is defined as follows. We fix an m-dimensional subspace of $\mathbb{R}^n$ and denote by p the orthogonal projection from $\mathbb{R}^n$ onto that subspace. Next we compute the measure of the projection of the set E onto that subspace taking into account the multiplicity function N_p and then we average resulting measures over all possible projections $p \in O^*(n,m)$. Note that since the measure $\vartheta^*_{n,m}$ is invariant under rotations $O(n)$, $\mathcal{I}^m(E_1) = \mathcal{I}^m(E_2)$ if $E_1, E_2 \subset \mathbb{R}^n$ are isometric. We still need to define the coefficient $\beta(n,m)$. Let $[0,1]^m \subset \mathbb{R}^m \times \mathbb{R}^{n-m} = \mathbb{R}^n$ be the m-dimensional unit cube in $\mathbb{R}^n$. We define

$$\beta(n,m) = \int_{p \in O^*(n,m)} \int_{y \in \operatorname{Im} p} N_p(y, [0,1]^m)\, d\mathcal{H}^m(y)\, d\vartheta^*_{n,m}(p).$$

Clearly $\beta(n,m)$ is a positive constant and with its definition $\mathcal{I}^m([0,1]^m)=1$. Note that $\mathcal{I}^m(Q)=\mathcal{H}^m(Q)$ for *any* m-dimensional cube in $\mathbb{R}^n$ regardless how the cube is positioned in the space. This follows from the $O(n)$ invariance of the measure $\vartheta^*_{n,m}$ and from the fact that both measures $\mathcal{I}^m$ and $\mathcal{H}^m$ scale in the same way under homothetic transformations. Hence $\mathcal{I}^m(E)=\mathcal{H}^m(E)$ if E is an m-dimensional polyhedron in $\mathbb{R}^n$. Indeed, up to a set of $\mathcal{H}^m$-measure zero such a polyhedron is the union of countably many m-dimensional cubes and if $\mathcal{H}^m(A)=0$, then $\mathcal{I}^m(A)=0$. This observation can be generalized to arbitrary countably $\mathcal{H}^m$-dectifiable sets.

Theorem 3.30 (Federer)**.** *If $E\subset\mathbb{R}^n$ is countably $\mathcal{H}^m$-rectifiable, $m\leq n$, then $\mathcal{I}^m(E)=\mathcal{H}^m(E)$.*

Proof. We can assume that $m<n$, because $\mathcal{I}^n=\mathcal{H}^n$ by the definition. It suffices to assume that E is a subset of an m-dimensional C^1-submanifold $\mathcal{M}^m\subset\mathbb{R}^n$. Indeed, the general case will follow from Theorem 3.10 and the fact that $\mathcal{H}^m(A)=0$ implies that $\mathcal{I}^m(A)=0$. Let p' be the restriction of $p\in O^*(n,m)$ to $\mathcal{M}^m$. the area formula yields

$$\int_E |J_{p'}(x)|\,d\mathcal{H}^m(x)=\int_{\mathrm{Im}\,p} N_p(y,E)\,d\mathcal{H}^m(y). \tag{3.13}$$

Let L be an m-dimensional affine subspace of $\mathbb{R}^n$. Let p'' be the restriction of p to L and let $|J_{p''}|$ be the Jacobian of the orthogonal projection p'' of L onto $\mathrm{Im}\,p$. Clearly $|J_p(x)|=|J_{p''}|$,where $L=T_x\mathcal{M}^m$ is regarded as an affine subspace of $\mathbb{R}^n$. Observe that

$$\int_{p\in O^*(n,m)} |J_{p''}|\,d\vartheta^*_{n,m}(p)=C(n,m)$$

is a constant that depends on n and m only. Indeed, the measure $\vartheta^*_{n,m}$ is invariant under rotations $O(n)$ and hence we can rotate L without changing the value of the integral, so that L is parallel to $\mathbb{R}^m\times\{0\}\subset\mathbb{R}^n$. Hence (3.13) yields

$$\int_{p\in O^*(n,m)}\int_{y\in\mathrm{Im}\,p} N_p(y,E)\,d\mathcal{H}^m(y)\,d\vartheta^*_{n,m}(p)=C(n,m)\mathcal{H}^m(E).$$

Taking $E=[0,1]^m$ we see that $C(n,m)=\beta(n,m)$. $\square$

The next result is a celebrated structure theorem of Besicovitch-Federer which we state without proof.

Theorem 3.31 (Structure theorem)**.** *If $E\subset\mathbb{R}^n$, $\mathcal{H}^m(E)<\infty$, $m<n$ is purely $\mathcal{H}^m$-unrectifiable, then $\mathcal{I}^m(E)=0$.*

Thus any set $E\subset\mathbb{R}^n$ with $\mathcal{H}^m(E)<\infty$ can be decomposed into a rectifiable part on which $\mathcal{H}^m=\mathcal{I}^m$ and a non-rectifiable part on which $\mathcal{I}^m=0$. This says a lot about the structure of E, which explains the name of the theorem. This result also implies that $\mathcal{H}^m(E)\geq\mathcal{I}^m(E)$ for any Borel set $E\subset\mathbb{R}^n$.

One can construct Cantor type sets with $\mathcal{H}^m(E)>0$, but $\mathcal{I}^m(E)=0$. Clearly E must be purely $\mathcal{H}^m$-unrectifiable. However, the integral geometric measure can be used to detect the Hausdorff dimension of a set:

Theorem 3.32 (Mattila)**.** *If* $\dim_H E > m$*, then* $\mathcal{I}^m(E) = \infty$*. Hence* $\mathcal{I}^m(E) < \infty$ *implies that* $\dim_H E \leq m$*.*

The proof requires quite a lot of harmonic analysis and potential theory and we will not present it here.

4. The Sard theorem for mappings into ℓ^∞

In the theory of countably rectifiable metric spaces it is important to be able to verify whether the image of a Lipschitz mapping $f : \mathbb{R}^n \supset E \to X$ into a metric space satisfies $\mathcal{H}^n(f(E)) = 0$. The next result shows how to reduce this problem to the case of mappings into $\mathbb{R}^n$.

Definition 4.1. Let $f : Z \to X$ be a mapping between metric spaces and let $y_1, \ldots, y_n \in X$. The mapping $g : Z \to \mathbb{R}^n$ defined by

$$g(x) = (d(f(x), y_1), \ldots, d(f(x), y_n)),$$

where d denotes the metric in X, is called the *projection of* f *associated with points* $y_1, \ldots, y_n$.

Note that $\pi : X \to \mathbb{R}^n$, $\pi(y) = (d(y, y_1), \ldots, d(y, y_n))$ is Lipschitz. Hence if f is Lipschitz, then $g = \pi \circ f$ is Lipschitz too.

Theorem 4.2. *Let* X *be a metric space, let* $E \subset \mathbb{R}^n$ *be measurable, and let* $f : E \to X$ *be a Lipschitz mapping. Then the following statements are equivalent:*

1. $\mathcal{H}^n(f(E)) = 0$*;*
2. *For any Lipschitz mapping* $\varphi : X \to \mathbb{R}^n$*, we have* $\mathcal{H}^n(\varphi(f(E))) = 0$*;*
3. *For any collection of distinct points* $\{y_1, y_2, \ldots, y_n\} \subset X$*, the associated projection* $g : E \to \mathbb{R}^n$ *of* f *satisfies* $\mathcal{H}^n(g(E)) = 0$*;*
4. *For any collection of distinct points* $\{y_1, y_2, \ldots, y_n\} \subset X$*, the associated projection* $g : E \to \mathbb{R}^n$ *of* f *satisfies* $\operatorname{rank}(\operatorname{ap} Dg(x)) < n$ *for* $\mathcal{H}^n$*-a.e.* $x \in E$*.*

Remark 4.3. It follows from the proof that in conditions (3) and (4) we do not have to consider all families $\{y_1, y_2, \ldots, y_n\} \subset X$ of distinct points, but it suffices to consider such families with points y_i taken from a given countable and dense subset of $f(E)$.

The implications from (1) to (2) and from (2) to (3) are obvious. The equivalence between (3) and (4) easily follows from the change of variables formula (Theorem 3.18): if $g : \mathbb{R}^n \supset E \to \mathbb{R}^n$ is Lipschitz, then

$$\int_E |J_g(x)| \, d\mathcal{H}^n(x) = \int_{g(E)} N_g(y, E) \, d\mathcal{H}^n(y). \tag{4.1}$$

Therefore, it remains to prove the implication (4) to (1) which is the most difficult part of the theorem. We will deduce it from another result which deals with Lipschitz mappings into ℓ^∞, see Theorem 4.5.

Remark 4.4. [27] In general it may happen for a subset $A \subset X$ that $\mathcal{H}^n(A) > 0$, but for all Lipschitz mappings $\varphi : X \to \mathbb{R}^n$, $\mathcal{H}^n(\varphi(A)) = 0$. For example the Heisenberg group[28] $\mathbb{H}^k$ satisfies $\mathcal{H}^{2k+2}(\mathbb{H}^k) = \infty$, but $\mathcal{H}^{2k+2}(\varphi(\mathbb{H}^k)) = 0$ for all Lipschitz mappings $\varphi : \mathbb{H}^k \to \mathbb{R}^{2k+2}$. Hence the implication from (2) to (1) has to use in an essential way that the assumption that $A = f(E)$ is a Lipschitz image of a Euclidean set. Since the condition (2) is satisfied for $X = \mathbb{H}^k$ with $n = 2k+2$, we conclude that $\mathbb{H}^k$ is purely $(2k+2)$-unrectifiable.

Let $f = (f_1, f_2, \ldots) : \mathbb{R}^n \supset E \to \ell^\infty$ be an L-Lipschitz mappings. Then the components $f_i : E \to \mathbb{R}$ are also L-Lipschitz. Hence for $\mathcal{H}^n$-almost all points $x \in E$, all functions f_i, $i \in \mathbb{N}$ are approximately differentiable at $x \in E$. We define the approximate derivative of f componentwise

$$\operatorname{ap} Df(x) = (\operatorname{ap} Df_1(x), \operatorname{ap} Df_2(x), \ldots).$$

For each $i \in \mathbb{N}$, $\operatorname{ap} Df_i(x)$ is a vector in $\mathbb{R}^n$ with component bounded by L. Hence $\operatorname{ap} Df(x)$ can be regarded as an $n \times \infty$ matrix of real numbers bounded by L, i.e.

$$\operatorname{ap} Df(x) \in (\ell^\infty)^n, \qquad \|\operatorname{ap} Df\|_\infty \leq L,$$

where the norm in $(\ell^\infty)^n$ is defined as the supremum over all entries in the $n \times \infty$ matrix. The meaning of the rank of the $n \times \infty$ matrix $\operatorname{ap} Df(x)$ is clear. It is the dimension of the linear subspace of $\mathbb{R}^n$ spanned by the vectors $\operatorname{ap} Df_i(x)$, $i \in \mathbb{N}$. Hence $\operatorname{rank}(\operatorname{ap} Df(x)) \leq n$ a.e.

The next theorem is the main step in the proof of the remaining implication (4) to (1) of Theorem 4.2.

Theorem 4.5. *Let $E \subset \mathbb{R}^n$ be measurable and let $f : E \to \ell^\infty$ be a Lipschitz mapping. Then $\mathcal{H}^n(f(E)) = 0$ if and only if* $\operatorname{rank}(\operatorname{ap} Df(x)) < n$, *$\mathcal{H}^n$-a.e. in E.*

Before we prove this result we will show how to use it to complete the proof of Theorem 4.2.

Proof of Theorem 4.2. As we already pointed out, it remains to prove the implication from (4) to (1). Although we do not assume that X is separable, the image $f(E) \subset X$ is separable and hence it can be isometrically embedded into ℓ^∞ via the Kuratowski embedding (Theorem 1.18). More precisely let $\{y_i\}_{i=1}^\infty \subset f(E)$ be a dense subset and let $y_0 \in f(E)$. Then

$$f(E) \ni y \mapsto \kappa(y) = \{d(y, y_i) - d(y_i, y_0)\}_{i=1}^\infty \in \ell^\infty$$

is an isometric embedding of $f(E)$ into ℓ^∞. Clearly

$$\mathcal{H}^n_d(f(E)) = \mathcal{H}^n_{\ell^\infty}((\kappa \circ f)(E)),$$

[27]See Section 11.5 in David, G., Semmes, S.: *Fractured fractals and broken dreams. Self-similar geometry through metric and measure.* Oxford Lecture Series in Mathematics and its Applications, 7. The Clarendon Press, Oxford University Press, New York, 1997.

[28]Do not worry if you do not know what the Heisenberg groups are. This is just an example that will not be used in what is to follow.

where subscripts indicate metrics with respect to which we define the Hausdorff measures. It remains to prove that $\mathcal{H}^n_{\ell^\infty}((\kappa\circ f)(E))=0$. Since

$$(\kappa\circ f)(x)=\{d(f(x),y_i)-d(y_i,y_0)\}_{i=1}^\infty$$

it easily follows from the assumptions that

$$\operatorname{rank}(\operatorname{ap}D(\kappa\circ f))<n \quad \mathcal{H}^n\text{-a.e. in } E.$$

Hence (1) follows from Theorem 4.5. □

Thus it remains to prove Theorem 4.5. Before doing this let us make some comments explaining why it is not easy. Theorem 4.5 is related to the Sard theorem for Lipschitz mappings (Theorem 3.4) which states that if $f:\mathbb{R}^n\to\mathbb{R}^m$, $m\geq n$ is Lipschitz, then

$$\mathcal{H}^n(f(\{x\in\mathbb{R}^n:\operatorname{rank}Df(x)<n\}))=0.$$

The standard proof presented earlier is based on the observation that if $\operatorname{rank}Df(x)<n$, then for any $\varepsilon>0$ there is $r>0$ such that

$$|f(z)-f(x)-Df(x)(z-x)|<\varepsilon r \quad \text{for } z\in B(x,r)$$

and hence

$$\operatorname{dist}(f(z),W_x)\leq\varepsilon r \quad \text{for } z\in B(x,r),$$

where $W_x=f(x)+Df(x)(\mathbb{R}^n)$ is an affine subspace of $\mathbb{R}^m$ of dimension less than or equal to $n-1$. That means $f(B(x,r))$ is contained in a thin neighborhood of an ellipsoid of dimension no greater than $n-1$ and hence we can cover it by $C(L/\varepsilon)^{n-1}$ balls of radius $C\varepsilon r$, where L is the Lipschitz constant of f. Now we use the $5r$-covering lemma to estimate the Hausdorff content of the image of the critical set.

The proof described here employs the fact that f is Frechet differentiable and hence this argument *cannot* be applied to the case of mappings into ℓ^∞, because in general Lipschitz mappings into ℓ^∞ are not Frechet differentiable, i.e. in general the image of $f(B(x,r)\cap E)$ is not well approximated by the tangent mapping $\operatorname{ap}Df(x)$. To overcome this difficulty we need to investigate the structure of the set $\{\operatorname{ap}Df(x)<n\}$ using arguments employed in the proof of the general case of the Sard theorem for C^k mappings that will be presented in Section **??**. In particular we will need to use a version of the implicit function theorem.

In the proof of Theorem 4.5 we will also need the following result which is of independent interest.

Proposition 4.6. *Let $D\subset\mathbb{R}^n$ be a bounded and convex set with non-empty interior and let $f:D\to\ell^\infty$ be an L-Lipschitz mapping. Then*

$$\operatorname{diam}(f(D))\leq C(n)L\frac{(\operatorname{diam}D)^n}{\mathcal{H}^n(D)}\,\mathcal{H}^n(D\setminus A)^{1/n}$$

where

$$A=\{x\in D:\ Df(x)=0\}.$$

In particular if D is a cube or a ball, then

$$\operatorname{diam}(f(D))\leq C(n)L\mathcal{H}^n(D\setminus A)^{1/n} \tag{4.2}$$

Proof. We will need two well known facts.

Lemma 4.7. *If $E \subset \mathbb{R}^n$ is measurable, then*
$$\int_E \frac{dy}{|x-y|^{n-1}} \leq C(n)\mathcal{H}^n(E)^{1/n}.$$

Proof. Let $B = B(x,r) \subset \mathbb{R}^n$ be a ball such that $\mathcal{H}^n(B) = \mathcal{H}^n(E)$. Then
$$\int_E \frac{dy}{|x-y|^{n-1}} \leq \int_B \frac{dy}{|x-y|^{n-1}} = C(n)r = C'(n)\mathcal{H}^n(E)^{1/n}.$$
□

The next lemma will be proved in Section **??**.

Lemma 4.8. *If $D \subset \mathbb{R}^n$ is a bounded and convex set with non-empty interior and if $u : D \to \mathbb{R}$ is Lipschitz continuous, then*
$$|u(x) - u_D| \leq \frac{(\operatorname{diam} D)^n}{n\mathcal{H}^n(D)} \int_D \frac{|\nabla u(y)|}{|x-y|^{n-1}}\, dy \quad \textit{for all } x \in D,$$
where
$$u_D = \frac{1}{\mathcal{H}^n(D)} \int_D u(x)\, dx.$$

Now we can complete the proof of Proposition 4.6. If $Df(x) = 0$, then $\nabla f_i(x) = 0$ for all $i \in \mathbb{N}$. For each $i \in \mathbb{N}$ we have
$$\begin{aligned} |f_i(x) - f_{iD}| &\leq \frac{(\operatorname{diam} D)^n}{n\mathcal{H}^n(D)} \int_D \frac{|\nabla f_i(y)|}{|x-y|^{n-1}}\, dy \leq \frac{L(\operatorname{diam} D)^n}{n\mathcal{H}^n(D)} \int_{D\setminus A} \frac{dy}{|x-y|^{n-1}} \\ &\leq C(n)L\frac{(\operatorname{diam} D)^n}{\mathcal{H}^n(D)}\, \mathcal{H}^n(D\setminus A)^{1/n}. \end{aligned}$$
Hence for all $x, y \in D$
$$|f_i(x) - f_i(y)| \leq |f_i(x) - f_{iD}| + |f_i(y) - f_{iD}| \leq 2C(n)L\frac{(\operatorname{diam} D)^n}{\mathcal{H}^n(D)}\, \mathcal{H}^n(D\setminus A)^{1/n}.$$
Taking supremum over $i \in \mathbb{N}$ yields
$$\|f(x) - f(y)\|_\infty \leq 2C(n)L\frac{(\operatorname{diam} D)^n}{\mathcal{H}^n(D)}\, \mathcal{H}^n(D\setminus A)^{1/n}$$
and the result follows upon taking supremum over all $x, y \in D$. □

Proof of Theorem 4.5. The implication from left to right is easy. Suppose that $\mathcal{H}^n(f(E)) = 0$. For any positive integers $i_1 < i_2 < \ldots < i_n$ the projection
$$\ell^\infty \ni (y_1, y_2, \ldots) \to (y_{i_1}, y_{i_2}, \ldots, y_{i_n}) \in \mathbb{R}^n$$
is Lipschitz continuous and hence the set
$$(f_{i_1}, \ldots, f_{i_n})(E) \subset \mathbb{R}^n$$
has $\mathcal{H}^n$-measure zero. It follows from the change of variables formula (4.1) that the matrix $[\partial f_{i_j}/\partial x_\ell]_{j,\ell=1}^n$ of approximate partial derivatives has rank less than n almost everywhere in E. Since this is true for any choice of $i_1 < i_2 < \ldots < i_n$ we conclude that rank $(\operatorname{ap} Df(x)) < n$ a.e. in E.

Suppose now that rank $(\operatorname{ap} Df(x)) < n$ a.e. in E. We need to prove that $\mathcal{H}^n(f(E)) = 0$. This implication is more difficult. Since $f_i : E \to \mathbb{R}$ is Lipschitz continuous, for any $\varepsilon > 0$ there is $g_i \in C^1(\mathbb{R}^n)$ such that

$$\mathcal{H}^n(\{x \in E :\ f_i(x) \neq g_i(x)\}) < \varepsilon/2^i.$$

Moreover $\operatorname{ap} Df_i(x) = Dg_i(x)$ for almost all points of the set where $f_i = g_i$ (Theorem 2.7(d)). Hence there is a measurable set $F \subset E$ such that $\mathcal{H}^n(E \setminus F) < \varepsilon$ and

$$f = g \quad \text{and} \quad \operatorname{ap} Df(x) = Dg(x) \quad \text{in } F$$

where

$$g = (g_1, g_2, \ldots), \quad Dg = (Dg_1, Dg_2, \ldots).$$

It suffices to prove that $\mathcal{H}^n(f(F)) = 0$, because we can exhaust E with sets F up to a subset of measure zero and f maps sets of measure zero to sets of measure zero. Let

$$\tilde{F} = \{x \in F :\ \operatorname{rank}(\operatorname{ap} Df(x)) = \operatorname{rank} Dg(x) < n\}.$$

Since $\mathcal{H}^n(F \setminus \tilde{F}) = 0$, it suffices to prove that $\mathcal{H}^n(f(\tilde{F})) = 0$. For $0 \leq j \leq n-1$ let

$$K_j = \{x \in \tilde{F} : \operatorname{rank} Dg(x) = j\}.$$

Since $\tilde{F} = \bigcup_{j=0}^{n-1} K_j$, it suffices to prove that $\mathcal{H}^n(f(K_j)) = 0$ for any $0 \leq j \leq n-1$. Again, by removing a subset of measure zero we can assume that all points of K_j are density points of K_j. To prove that $\mathcal{H}^n(f(K_j)) = 0$ we need to make a change of variables in $\mathbb{R}^n$, but only when $j \geq 1$.

If $x \in \mathbb{R}^n \setminus F$, the sequence $(g_1(x), g_2(x), \ldots)$ is not necessarily bounded. Let V be the linear space of all real sequences $(y_1, y_2, \ldots)$. Clearly $g : \mathbb{R}^n \to V$. We do not equip V with any metric structure. Note that $g|_F : F \to \ell^\infty \subset V$, because g coincides with f on F.

Lemma 4.9. *Let $1 \leq j \leq n-1$ and $x_0 \in K_j$. Then there exists a neighborhood $x_0 \in U \subset \mathbb{R}^n$, a diffeomorphism $\Phi : U \subset \mathbb{R}^n \to \Phi(U) \subset \mathbb{R}^n$, and a composition of a translation (by a vector from ℓ^∞) with a permutation of variables $\Psi \colon V \to V$ such that*

- *$\Phi^{-1}(0) = x_0$ and $\Psi(g(x_0)) = 0$;*
- *There is $\varepsilon > 0$ such that for $x = (x_1, x_2, \ldots, x_n) \in B(0, \varepsilon) \subset \mathbb{R}^n$ and $i = 1, 2, \ldots, j$,*

$$\left(\Psi \circ g \circ \Phi^{-1}\right)_i(x) = x_i,$$

i.e., $\Psi \circ g \circ \Phi^{-1}$ fixes the first j variables in a neighborhood of 0.

Proof. By precomposing g with a translation of $\mathbb{R}^n$ by the vector x_0 and postcomposing it with a translation of V by the vector $-g(x_0) = -f(x_0) \in \ell^\infty$ we may assume that $x_0 = 0$ and $g(x_0) = 0$. A certain $j \times j$ minor of $Dg(x_0)$ has rank j. By precomposing g with a permutation of j variables in $\mathbb{R}^n$ and postcomposing it with a permutation of j variables in V we may assume that

$$\operatorname{rank} \left[\frac{\partial g_m}{\partial x_\ell}(x_0)\right]_{1 \leq m, \ell \leq j} = j. \tag{4.3}$$

Let $H : \mathbb{R}^n \to \mathbb{R}^n$ be defined by

$$H(x) = (g_1(x), \ldots, g_j(x), x_{j+1}, \ldots, x_n).$$

It follows from (4.3) that $J_H(x_0)\neq 0$ and hence H is a diffeomorphism in a neighborhood of $x_0=0\in\mathbb{R}^n$. It suffices to observe that for all $i=1,2,\ldots,j$,

$$\left(g\circ H^{-1}\right)_i(x)=x_i.$$

□

In what follows, by cubes we will mean cubes with edges parallel to the coordinate axes in $\mathbb{R}^n$. It suffices to prove that any point $x_0\in K_j$ has a cubic neighborhood whose intersection with K_j is mapped onto a set of $\mathcal{H}^n$-measure zero. Since we can take cubic neighborhoods to be arbitrarily small, the change of variables from Lemma 4.9 allows us to assume that

$$K_j\subset(0,1)^n,\quad g_i(x)=x_i\quad\text{for } i=1,2,\ldots,j \text{ and } x\in[0,1]^n. \tag{4.4}$$

Indeed, according to Lemma 4.9 we can assume that $x_0=0$ and that g fixes the first j variables in a neighborhood of 0. The neighborhood can be very small, but a rescaling argument allows us to assume that it contains a unit cube Q around 0. Translating the cube we can assume that $Q=[0,1]^n$. If $x\in K_j$, since $\operatorname{rank} Dg(x)=j$ and g fixes the first j coordinates, the derivative of g in directions orthogonal to the first j coordinates equals zero at x, $\partial g_k(x)/\partial x_i=0$ for $i=j+1,\ldots,n$ and any k.

Lemma 4.10. *Under the assumptions* (4.4) *there exists a constant $C=C(n)>0$ such that for any integer $m\geq 1$, and every $x\in K_j$, there is a closed cube $Q_x\subset[0,1]^n$ with edge length d_x centered at x with the property that $f(K_j\cap Q_x)=g(K_j\cap Q_x)$ can be covered by m^j balls in ℓ^∞ each of radius CLd_xm^{-1}, where L is the Lipschitz constant of f.*

The theorem is an easy consequence of this lemma through a standard application of the $5r$-covering lemma (Theorem 3.5); we used a similar argument in the proof of the Sard theorem. First of all observe that cubes with sides parallel to coordinate axes in $\mathbb{R}^n$ are balls with respect to the ℓ^∞_n metric

$$\|x-y\|_\infty=\max_{1\leq i\leq n}|x_i-y_i|.$$

Hence the $5r$-covering lemma applies to families of cubes in $\mathbb{R}^n$. By $5^{-1}Q$ we will denote a cube concentric with Q and with 5^{-1} times the diameter. The cubes $\{5^{-1}Q_x\}_{x\in K_j}$ form a covering of K_j. Hence we can select disjoint cubes $\{5^{-1}Q_{x_i}\}_{i=1}^\infty$ such that

$$K_j\subset\bigcup_{i=1}^\infty Q_{x_i}.$$

If d_i is the edge length of Q_{x_i}, then $\sum_{i=1}^\infty(5^{-1}d_i)^n\leq 1$, because the cubes $5^{-1}Q_{x_i}$ are disjoint and contained in $[0,1]^n$. Hence

$$\mathcal{H}^n_\infty(f(K_j))\leq\sum_{i=1}^\infty\mathcal{H}^n_\infty(f(K_j\cap Q_{x_i}))\leq\sum_{i=1}^\infty m^j(CLd_im^{-1})^n\leq 5^nC^nL^nm^{j-n}.$$

Since the exponent $j-k$ is negative, and m can be arbitrarily large we conclude that $\mathcal{H}^n_\infty(f(K_j))=0$ and hence $\mathcal{H}^n(f(K_j))=0$.

Thus it remains to prove Lemma 4.10.

Proof of Lemma 4.10. Various constants C in the proof below will depend on n only. Fix an integer $m \geq 1$. Let $x \in K_j$. Since every point in K_j is a density point of K_j, there is a closed cube $Q \subset [0,1]^n$ centered at x of edge length d such that

$$\mathcal{H}^n(Q \setminus K_j) < m^{-n}\mathcal{H}^n(Q) = m^{-n}d^n. \tag{4.5}$$

By translating the coordinate system in $\mathbb{R}^n$ we may assume that

$$Q = [0,d]^j \times [0,d]^{n-j}.$$

Each component of $f : Q \cap K_j \to \ell^\infty$ is an L-Lipschitz function. Extending each component to an L-Lipschitz function on Q results in an L-Lipschitz extension $\tilde{f} : Q \to \ell^\infty$. This is well known and easy to check.

Divide $[0,d]^j$ into m^j cubes with pairwise disjoint interiors, each of edge length $m^{-1}d$. Denote the resulting cubes by Q_ν, $\nu \in \{1, 2, \ldots, m^j\}$. It remains to prove that

$$f((Q_\nu \times [0,d]^{n-j}) \cap K_j) \subset \tilde{f}(Q_\nu \times [0,d]^{n-j})$$

is contained in a ball (in ℓ^∞) of radius $CLdm^{-1}$. It follows from (4.5) that

$$\mathcal{H}^n((Q_\nu \times [0,d]^{n-j}) \setminus K_j) \leq \mathcal{H}^n(Q \setminus K_j) < m^{-n}d^n.$$

Hence

$$\mathcal{H}^n((Q_\nu \times [0,d]^{n-j}) \cap K_j) > (m^{-j} - m^{-n})d^n.$$

This estimate and the Fubini theorem imply that there is $\rho \in Q_\nu$ such that

$$\mathcal{H}^{n-j}((\{\rho\} \times [0,d]^{n-j}) \cap K_j) > (1 - m^{j-n})d^{n-j}.$$

Hence

$$\mathcal{H}^{n-j}((\{\rho\} \times [0,d]^{n-j}) \setminus K_j) < m^{j-n}d^{n-j}.$$

It follows from (4.2) with n replaced by $n-j$ that

$$\operatorname{diam}_{\ell^\infty}(\tilde{f}(\{\rho\} \times [0,d]^{n-j})) \leq CL\mathcal{H}^{n-j}(\{\rho\} \times [0,d]^{n-j}) \setminus K_j)^{1/(n-j)} \leq CLm^{-1}d. \tag{4.6}$$

Indeed, the rank of the derivative of g restricted to the slice $\{\rho\} \times [0,d]^{n-j}$ equals zero at the points of $(\{\rho\} \times [0,d]^{n-j}) \cap K_j$ and this derivative coincides a.e. with the approximate derivative of $\tilde{f}$ restricted to $(\{\rho\} \times [0,d]^{n-j}) \cap K_j$ which by the property of g must be zero as well.

Since the distance of any point in $Q_\nu \times [0,d]^{n-j}$ to $\{\rho\} \times [0,d]^{n-j}$ is bounded by $Cm^{-1}d$ and $\tilde{f}$ is L-Lipschitz, (4.6) implies that $\tilde{f}(Q_\nu \times [0,d]^{n-j})$ is contained in a ball of radius $CLdm^{-1}$, perhaps with a constant C bigger than that in (4.6). The proof is the lemma is complete. □

This also completes the proof of Theorem 4.5. □

5. The Whitney extension theorem

Let us recall the statement of the Whitney extension theorem that we used in Section 2.

Theorem 5.1 (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*

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

if and only if

$$\lim_{\substack{x,y\in K,\, x\neq y\\ |x-y|\to 0}} \frac{|f(y)-f(x)-L(x)(y-x)|}{|y-x|} = 0. \tag{5.1}$$

The necessity of the condition (5.1) follows from Taylor's theorem with the integral form of the remainder which we recall below. Condition (5.1) simply means that the formal Taylor remainder of the first order is $o(|x-y|)$ uniformly on K.

Theorem 5.2 (Taylor's formula). *If $f \in C^m(\Omega)$, where $\Omega \subset \mathbb{R}^n$ is open and convex, then for $x, y \in \Omega$ we have*

$$f(y) - \sum_{|\alpha|\leq m-1} D^\alpha f(x) \frac{(y-x)^\alpha}{\alpha!} = m \sum_{|\alpha|=m} \frac{(y-x)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} D^\alpha f(x+t(y-x))\, dt$$

and

$$f(y) - \sum_{|\alpha|\leq m} D^\alpha f(x) \frac{(y-x)^\alpha}{\alpha!} = m \sum_{|\alpha|=m} \frac{(y-x)^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} (D^\alpha f(x+t(y-x)) - D^\alpha f(x))\, dt.$$

Here $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i \geq 0$ are integers, $|\alpha| = \alpha_1 + \ldots + \alpha_n$, $\alpha! = \alpha_1! \cdots \alpha_n!$ and $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.

Remark 5.3. There are examples showing that the condition

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

is not sufficient for the existence of a C^1 extension F; we need a uniform convergence.

Actually Whitney proved a more general result that Theorem 5.1 that characterizes restrictions of $C^m(\mathbb{R}^n)$ functions to K.

Theorem 5.4 (Whitney extension theorem). *Let $K \subset \mathbb{R}^n$ be a compact set and let $(f^\alpha)_{|\alpha|\leq m}$ be a family of continuous functions on K indexed by multiindices α, $|\alpha| \leq m$. Then there is a function $g \in C^m(\mathbb{R}^n)$ such that*

$$D^\alpha g\big|_K = f^\alpha \quad \textit{for all } \alpha,\ |\alpha| \leq m$$

if and only if for any α, $|\alpha| \leq m$

$$\lim_{\substack{x,y\in K,\, x\neq y\\ |x-y|\to 0}} \frac{\left|f^\alpha(y) - \sum_{|\beta|\leq m-|\alpha|} f^{\alpha+\beta}(x) \frac{(y-x)^\beta}{\beta!}\right|}{|x-y|^{m-|\alpha|}} = 0. \tag{5.2}$$

Again necessity of the condition follows from Taylor's theorem. If $f = f^0$, then g is a C^m extension of f. If $m = 1$, Theorem 5.4 reduces to Theorem 5.1. Indeed, if $|\alpha| = 0$, then the conditions (5.1) and (5.2) are equivalent; if $|\alpha| = 1$, then the condition (5.2) simply means that f^α is uniformly continuous which is trivially true since f^α is a continuous function on a compact set.

In order to prove Theorem 5.4 we need to develop terminology that will be helpful when dealing with various constructions arising from the family $(f^\alpha)_{|\alpha|\leq m}$.

Definition 5.5. A family $F = (f^\alpha)_{|\alpha|\leq m}$ of continuous functions on a compact set $K \subset \mathbb{R}^n$ is called a *jet of order* m or simply an m*-jet*. The space of all m-jets on K is denoted by $J^m(K)$. It is a Banach space with respect to the norm

$$|F|_m^K = \sup_{\substack{x\in K\\ |\alpha|\leq m}} |f^\alpha(x)|.$$

Note that the space $J^m(K)$ is nothing else, but the Banach space of all continuous functions $C(K, \mathbb{R}^N)$, where N equals the number of multiindices α with $|\alpha| \leq m$.

Sometimes for simplicity we will write $|F|_m^K = |F|_m$. There is a natural restriction operator

$$J^m : C^m(\mathbb{R}^n) \to J^m(K), \qquad J^m(g) = (D^\alpha g|_K)_{|\alpha|\leq m}.$$

The Whitney extension theorem provides a characterization of the subspace $J^m(C^m(\mathbb{R}^n)) \subset J^m(K)$.

If $F = (f^\alpha)_{|\alpha|\leq m}$, we will write $F(x) = f^0(x)$. Thus the functions f^α play a role of derivatives of $F(x)$. We define a formal derivative on the space $J^m(K)$ by

$$D^\alpha : J^m(K) \to J^{m-|\alpha|}(K), \qquad D^\alpha (f^\beta)_{|\beta|\leq m} = (f^{\alpha+\beta})_{|\beta|\leq m-|\alpha|}$$

and a formal Taylor polynomial

$$T_a^m F(x) = \sum_{|\alpha|\leq m} f^\alpha(a)\frac{(x-a)^\alpha}{\alpha!}, \quad x \in \mathbb{R}^n,\ a \in K.$$

Clearly $T_a^m F(x)$ is a polynomial of degree m as a function of x. Since $T_a^m F(x)$ is a smooth function of x, we can associate an m-jet with it using the operation J^m

$$\tilde{T}_a^m F = J^m(T_a^m F) \in J^m(K),$$

i.e.

$$\tilde{T}_a^m F = \left(\sum_{|\beta|\leq m-|\alpha|} f^{\alpha+\beta}(a)\frac{(x-a)^\beta}{\beta!} \right)_{|\alpha|\leq m}.$$

In other words $\tilde{T}_a^m F$ is the collection of formal Taylor polynomials of all the functions f^α, $|\alpha| \leq m$. We will write

$$(\tilde{T}_a^m F)^\alpha(x) = \sum_{|\beta|\leq m-|\alpha|} f^{\alpha+\beta}(a)\frac{(x-a)^\beta}{\beta!}$$

for the formal Taylor polynomial of f^α. Finally we define a formal Taylor remainder

$$R_a^m F = F - \tilde{T}_a^m \in J^m(K),$$

i.e.

$$R_a^m F = \left(f^\alpha(x) - \sum_{|\beta| \le m-|\alpha|} f^{\alpha+\beta}(a) \frac{(x-a)^\beta}{\beta!} \right)_{|\alpha|\le m}$$

is the collection of all formal Taylor remainders of all functions f^α, $|\alpha| \le m$. We will write

$$(R_a^m F)^\alpha(x) = f^\alpha(x) - \sum_{|\beta| \le m-|\alpha|} f^{\alpha+\beta}(a) \frac{(x-a)^\beta}{\beta!}\,.$$

The condition (5.2) from Whitney's theorem can be rewritten as: For all α, $|\alpha| \le m$

$$\frac{(R_x^m F)^\alpha(y)}{|x-y|^{m-|\alpha|}} \rightrightarrows 0 \quad \text{as } |x-y| \to 0,\ x \ne y, \text{ uniformly on } K,$$

or

$$(R_x^m F)^\alpha(y) = o(|x-y|^{m-|\alpha|}) \quad \text{for } x, y \in K \text{ as } |x-y| \to 0.$$

Now we will show how to reformulate this conditions in two other equivalent ways.

Definition 5.6. A *modulus of continuity* is a continuous, nondecreading and concave function $\eta : [0,\infty) \to [0,\infty)$ such that $\eta(0) = 0$.

If a continuous function f satisfies $|f(x) - f(y)| \le \eta(|x-y|)$, then f is uniformly continuous and the modulus of continuity tells us how quickly $f(y)$ converges to $f(x)$ as $y \to x$.

Example 5.7. If $\eta(t) = t$, then f is Lipschitz. If $\eta(t) = t^s$, $0 < s \le 1$, then f is s-Hölder continuous.

Proposition 5.8. *If f is a continuous function defined on a compact metric space (X, d), then there is a modulus of continuity η such that $|f(x) - f(y)| \le \eta(d(x,y))$ for all $x, y \in X$.*

Proof. Let

$$\tilde\eta(t) = \sup\{|f(y) - f(x)| :\ d(x,y) \le t\}.$$

Clearly $\tilde\eta$ is nondecreasing and $\tilde\eta(t) \to 0$ as $t \to 0^+$ by uniform continuity of f. Now one can show that η defined as the infimum of all concave functions that are greater than or equal to $\tilde\eta$ satisfies the claim.[29] □

Remark 5.9. It is easy to see that the modulus of continuity constructed in the proof of Proposition 5.8 is the least modulus of continuity of f.

Theorem 5.10. *Let $F \in J^m(K)$. Then the following conditions are equivalent.*

(1) *For all α, $|\alpha| \le m$*

$$\frac{(R_x^m F)^\alpha(y)}{|x-y|^{m-|\alpha|}} \rightrightarrows 0 \quad \textit{as } |x-y| \to 0,\ x \ne y, \textit{ uniformly on } K.$$

(2) *There is a modulus of continuity η such that for all α, $|\alpha| \le m$*

$$|(R_x^m F)^\alpha(y)| \le \eta(|x-y|)|x-y|^{m-|\alpha|}, \quad x, y \in K.$$

[29] To construct η we take the intersection of all convex sets that contain the graph of $\tilde\eta$ and the positive t-axis. The resulting set is bounded from above by the graph of η.

(3) *There is a modulus of continuity η_1 such that*
$$|T^m_x F(z) - T^m_y F(z)| \le \eta_1(|x-y|)(|x-z|^m + |y-z|^m)$$
for all $x, y \in K$, $z \in \mathbb{R}^n$.

Moreover if η satisfies (2), then we can take $\eta_1 = C_1(m,n)\eta$ and if η_1 satisfies (3), then we can take $\eta = C_2(m,n)\eta_1$.

Proof. The implication from (2) to (1) is obvious. Suppose now that (1) is satisfied. Let
$$\tilde{\eta}(t) = \sup\left\{ \frac{|(R^m_x F)^\alpha(y)|}{|x-y|^{m-|\alpha|}} : |\alpha| \le m,\ x,y \in K,\ 0 < |x-y| \le t \right\}.$$
Clearly $\tilde{\eta}$ is nondecreasing and $\tilde{\eta}(t) \to 0$ as $t \to 0^+$. Now (2) is satisfied with η defined as the infimum over all concave functions that are greater than or equal to $\tilde{\eta}$. Thus we proved that the conditions (1) and (2) are equivalent and it remains to prove equivalence of the confitions (2) and (3). First we will prove that (2) implies (3).

Lemma 5.11.
$$T^m_x F(z) - T^m_y F(z) = \sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} (R^m_y F)^\alpha(x).$$

Proof. We have
$$(R^m_y F)^\alpha(x) = f^\alpha(x) - \sum_{|\beta| \le m-|\alpha|} f^{\alpha+\beta}(y) \frac{(x-y)^\beta}{\beta!} = f^\alpha(x) - D^\alpha_x \sum_{|\beta| \le m} f^\beta(y) \frac{(x-y)^\beta}{\beta!}.$$
Hence
$$\sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} (R^m_y F)^\alpha(x) = \underbrace{\sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} f^\alpha(x)}_{T^m_x F(z)} - \sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} D^\alpha_x \underbrace{\left(\sum_{|\beta| \le m} f^\beta(y) \frac{(x-y)^\beta}{\beta!} \right)}_{P(x)}.$$
Since $P(x)$ is a polynomial of degree m it equals to its Taylor polynomial of degree m
$$P(z) = \sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} D^\alpha P(x).$$
Hence
$$\sum_{|\alpha| \le m} \frac{(z-x)^\alpha}{\alpha!} D^\alpha_x \left(\sum_{|\beta| \le m} f^\beta(y) \frac{(x-y)^\beta}{\beta!} \right) = \sum_{|\beta| \le m} f^\beta(y) \frac{(z-y)^\beta}{\beta!} = T^m_y F(z).$$
□

It follows from the lemma that
$$|T^m_x F(z) - T^m_y F(z)| \le \sum_{|\alpha| \le m} \frac{|z-x|^{|\alpha|}}{\alpha!} |x-y|^{m-|\alpha|} \eta(|x-y|) \le \heartsuit.$$

Recall

Lemma 5.12 (Young's inequality)**.** *If $a, b > 0$ and $p^{-1} + q^{-1} = 1$, then*

$$ab \le \frac{a^p}{p} + \frac{b^q}{q}.$$

Hence $ab \le a^p + b^q$. Taking $p = m/|\alpha|$ and $q = m/(m - |\alpha|)$ we obtain

$$|z - x|^{|\alpha|}|x - y|^{m-|\alpha|} \le |z - x|^m + |x - y|^m.$$

Hence

$$\heartsuit \le C\eta(|x - y|)(|z - x|^m + |x - y|^m), \quad \text{where} \quad C = \sum_{|\alpha|\le m} \frac{1}{\alpha!}$$

depends on m and n only. It suffices to observe now that $|x - y| \le |z - x| + |z - y|$. This proves the implication from (2) to (3) with $\eta_1 = C\eta$. The proof of the implication from (3) to (2) is more tricky. According to Lemma 5.11

$$\left|\sum_{|\alpha|\le m} \frac{(z - x)^\alpha}{\alpha!}(R^m_y F)^\alpha(x)\right| \le \eta_1(|x - y|)(|x - z|^m + |y - z|^m).$$

Let z' be defined by $z - x = |x - y|(z' - x)$ and let $\lambda = |x - y|$. Then $|x - z|^m = \lambda^m|z' - x|^m$. Since $|y - z| \le |x - z| + |x - y|$ we have

$$|y - z|^m \le C(|x - z|^m + |x - y|^m) = C\lambda^m(|z' - x|^m + 1).$$

Hence

$$|x - z|^m + |y - z|^m \le C\lambda^m(1 + |z' - x|^m).$$

Thus

$$\left|\sum_{|\alpha|\le m} \frac{\lambda^{|\alpha|}}{\alpha!}(z' - x)^\alpha(R^m_y F)^\alpha(x)\right| \le C\eta_1(|x - y|)\lambda^m(1 + |z' - x|^m). \tag{5.3}$$

Fix $x, y \in K$, $x \ne y$. On the left hand side we have a polynomial of variable $(z' - x)$. If P is a polynomial of one real variable and $a_0, \dots, a_m$ are distinct real numbers, then the coefficients of P can be represented as a linear combinations of numbers $P(a_0), \dots, P(a_m)$. Indeed, we obtain a system of $m + 1$ linear equations for coefficients whose determinant (Vandermonde) is non-zero. Similar result is true for polynomials P of degree m of a multidimensional variable like $(z' - x) \in \mathbb{R}^n$. There is a set of points $a_0, \dots, a_N \in \mathbb{R}^n$ independent of a polynomial such that coefficients of P are linear combinations of numbers $P(a_0), \dots, P(a_N)$. In particular coefficients of the polynomial (5.3), i.e. numbers

$$\frac{\lambda^{|\alpha|}}{\alpha!}(R^m_y F)^\alpha(x)$$

are linear combinations of values of the polynomial at fixed points $a_0, \dots, a_N$. Hence (5.3) yields

$$\left|\frac{\lambda^{|\alpha|}}{\alpha!}(R^m_y F)^\alpha(x)\right| \le C\eta_1(|x - y|)\lambda^m.$$

Since $\lambda = |x - y|$, (2) follows. $\square$

Definition 5.13. $C^m(K)$ is a subspace of $J^m(K)$ consisting of all m-jets that satisfy one of the equivalent conditions from Theorem 5.10. The elements of $C^m(K)$ are called *Whitney m-jets.* The modulus of continuity η from Theorem 5.10(2) is called a modulus of continuity of $F \in C^m(K)$.

The space $C^m(K)$ is equipped with two equivalent norms

$$\|F\|_m^K = |F|_m^K + \sup_{\substack{x,y\in K,\, x\neq y\\ |\alpha|\leq m}} \frac{|(R_x^m F)^\alpha(y)|}{|x-y|^{m-|\alpha|}}\, ,$$

$$\|F\|_m^{\prime K} = |F|_m^K + \sup_{\substack{x,y\in K,\, x\neq y\\ z\in\mathbb{R}^n}} \frac{|T_x^m F(z) - T_y^m F(z)|}{|x-z|^m + |y-z|^m}\, .$$

The equivalence of the norms follows from the estimates that we obtained in the proof of Theorem 5.10 when we proved that η and η_1 can be made comparable.

Note that we can take η and η_1 in Theorem 5.10 such that

$$\|F\|_m^K = |F|_m^K + \eta(\operatorname{diam} K), \quad \|F\|_m^{\prime K} = |F|_m^K + \eta_1(\operatorname{diam} K). \tag{5.4}$$

Indeed, for $t \geq \operatorname{diam} K$,

$$\tilde{\eta}(t) = \tilde{\eta}(\operatorname{diam} K) = \sup_{\substack{x,y\in K,\, x\neq y\\ |\alpha|\leq m}} \frac{|(R_x^m F)^\alpha(y)|}{|x-y|^{m-|\alpha|}}\, .$$

Hence η constructed in the proof of Theorem 5.10 satisfies

$$\eta(\operatorname{diam} K) = \tilde{\eta}(\operatorname{diam} K) = \sup_{\substack{x,y\in K,\, x\neq y\\ |\alpha|\leq m}} \frac{|(R_x^m F)^\alpha(y)|}{|x-y|^{m-|\alpha|}}\, .$$

A similar argument applies to the second equality in (5.4).

Theorem 5.14 (Whitney extension theorem)**.** *There is a linear operator $W : C^m(K) \to C^m(\mathbb{R}^n)$ such that if $F = (f^\alpha)_{|\alpha|\leq m} \in C^m(K)$, then*

$$D^\alpha(W(F))(x) = f^\alpha(x) \quad \textit{for all } x \in K \textit{ and } |\alpha| \leq m$$

and $W(F)|_{\mathbb{R}^n\setminus K} \in C^\infty(\mathbb{R}^n \setminus K)$. Moreover if K is contained in the interior of a cube Q, and

$$\lambda = \sup_{x\in Q} \operatorname{dist}(x, K),$$

then there is a constant $C = C(m, n, \lambda)$ such that

1. $\|W(F)\|_m^Q \leq C\|F\|_m^K$.
2. *If η is a modulus of continuity of F, then $C\eta$ is a modulus of continuity of $W(F)|_{\overline{Q}} \in C^m(\overline{Q})$, i,e, $C\eta$ is a modulus of continuity of all derivatives $D^\alpha W(F)|_{\overline{Q}}$, $|\alpha| \leq m$.*

The proof is based on an explicit construction of the extension which requires the so called Whitney partition of unity.

Theorem 5.15 (Whitney covering). *Let $E \subset \mathbb{R}^n$ be closed. Then its complement $\mathbb{R}^n \setminus E$ admits a cover $\mathcal{V} = \{Q_i\}_{i\in I}$ by closed dyadic cubes*[30] *with pairwise disjoint interiors such that*

(1) $\bigcup_{i\in I} Q_i = \mathbb{R}^n \setminus E$;
(2) $\operatorname{diam} Q_i \leq \operatorname{dist}(Q_i, E) \leq 4 \operatorname{diam} Q_i$ *for all* $i \in I$;
(3) *If* Q_i *and* Q_j *touch, then* $\operatorname{diam} Q_i \leq 4 \operatorname{diam} Q_j$;
(4) *Any cube* Q_i *touches at most* $N = 12^n$ *other cubes in* $\mathcal{V}$;
(5) *Fix* $1 < \varepsilon < 1/4$ *and let* $\mathcal{V}^*$ *be the collection of cubes obtained by expanding each cube* Q_i *by the factor* $(1+\varepsilon)$ *around the center. Then* $\mathcal{V}^*$ *is a cover of* $\mathbb{R}^n \setminus E$ *such that each point in* $\mathbb{R}^n \setminus E$ *has a neighborhood which intersects at most* N *cubes* $Q_i^* \in \mathcal{V}^*$.

Proof. Let us first recall the definition of dyadic cubes. Let $\mathcal{M}_0$ be the collection of cubes in $\mathbb{R}^n$ with edges of length 1 and integer vertices. For $k \in \mathbb{Z}$ let $\mathcal{M}_k = 2^{-k}\mathcal{M}_0$. The cubes in the family

$$\mathcal{M} = \bigcup_{k\in\mathbb{Z}} \mathcal{M}_k$$

are called *dyadic cubes*. The dyadic cubes have an important property: If $Q_1, Q_2 \in \mathcal{M}$, then either they have disjoint interiors or one is contained in another. Note also that the diameter of each cube in the family $\mathcal{M}_k$ equals $2^{-k}\sqrt{n}$.

Let

$$\Omega_k = \{x \in \mathbb{R}^n : 2\sqrt{n}2^{-k} < \operatorname{dist}(x, E) \leq 2\sqrt{n}2^{-k+1}\}.$$

Clearly $\mathbb{R}^n \setminus E = \bigcup_{k\in\mathbb{Z}} \Omega_k$ is a decomposition of the complement of E into 'layers' according to the distance to the set E. Let

$$\mathcal{V}_0 = \bigcup_{k\in\mathbb{Z}} \{Q \in \mathcal{M}_k :\ Q \cap \Omega_k \neq \emptyset\}.$$

The cubes in $\mathcal{V}_0$ are not necessarily disjoint, but they have a property that we want:

$$\operatorname{diam} Q \leq \operatorname{dist}(Q, E) \leq 4 \operatorname{diam} Q \text{ for } Q \in \mathcal{V}_0.$$

Indeed, let $Q \in \mathcal{M}_k$ be such that $Q \cap \Omega_k \neq \emptyset$. Fix $x \in Q \cap \Omega_k$. We have

$$\operatorname{diam} Q = 2^{-k}\sqrt{n} = 2 \cdot 2^{-k}\sqrt{n} - \operatorname{diam} Q < \operatorname{dist}(x, E) - \operatorname{diam} Q \leq \operatorname{dist}(Q, E)$$

and

$$\operatorname{dist}(Q, E) \leq \operatorname{dist}(x, E) \leq 2\sqrt{n}2^{-k+1} = 4 \operatorname{diam} Q.$$

The cubes in $\mathcal{V}_0$ are disjoint from E and hence they cover $\mathbb{R}^n \setminus E$. Since the cubes in $\mathcal{V}_0$ are not necessarily disjoint we need to select a subfamily of pairwise disjoint cubes. For each $Q \in \mathcal{V}_0$ there is a unique largest in volume cube $Q' \in \mathcal{V}_0$ that contains Q. Indeed, since diameters of cubes in $\mathcal{V}_0$ that contain Q are bounded by $\operatorname{dist}(Q, E)$, a largest cube containing Q exist. It is unique, because if two dyadic cubes contain Q, then one is contained in another. For the same reason cubes in the family $\mathcal{V} \subset \mathcal{V}_0$ of all such largest cubes have pairwise disjoint interiors. Clearly

$$\bigcup_{Q\in\mathcal{V}} Q = \mathbb{R}^n \setminus E.$$

[30] Dyadic cubes will be defined at the beginning of the proof.

Thus we proved properties (1) and (2). Property (3) follows from (2). Indeed, if Q_i and Q_j touch, then

$$\operatorname{diam} Q_i \le \operatorname{dist}(Q_i, E) \le \operatorname{dist}(Q_j, E) + \operatorname{diam} Q_j \le 5 \operatorname{diam} Q_j,$$

but diameters of the cubes are powers of 2 times $\sqrt{n}$, so $\operatorname{diam} Q_i \le 4 \operatorname{diam} Q_j$. To prove (4) assume that $Q_i \in \mathcal{M}_k$. By (3) no smaller cubes than those in $\mathcal{M}_{k+2}$ can touch Q_i. Q_i touches 3^n cubes in $\mathcal{M}_k$ (including Q_i). Since each cube in $\mathcal{M}_k$ splits into 4^n cubes in $\mathcal{M}_{k+2}$, the number of cubes in $\mathcal{M}_{k+2}$ that can touch Q_i is bounded by $3^n \cdot 4^n = 12^n$.[31] If some of the cubes in $\mathcal{V}$ that touch Q_i are larger than $\mathcal{M}_{k+2}$, the number of touching cubes is smaller.

It remain to prove (5). By (2) each cube Q_i^* is contained in $\mathbb{R}^n \setminus E$. Since the cubes Q_i cover $\mathbb{R}^n \setminus E$, their expansions Q_i^* also cover $\mathbb{R}^n \setminus E$. Let $x \in \mathbb{R}^n \setminus E$. Then $x \in Q_i$ for some $i \in I$. There is $\delta > 0$ such that if $B(x,\delta) \cap Q_j^* \neq \emptyset$, then $B(x,\delta) \subset Q_j^{**}$, where Q_j^{**} is the expansion of Q_i by the factor $1 + 1/4 > 1 + \varepsilon$, and hence $Q_j^{**} \cap Q_i \neq \emptyset$. Now it suffices to show that the number of indices $j \in I$ such that $Q_j^{**} \cap Q_i \neq \emptyset$ is bounded by 12^n. To this end it suffices to show that if $Q_j^{**} \cap Q_i \neq \emptyset$, then Q_j touches Q_i. This however, follows from the observation: diameters of cubes that touch Q_j are at least $\operatorname{diam} Q_j/4$ and hence Q_j^{**} is contained in the union of cubes that touch Q_j (including Q_j.) $\square$

From now on we fix the expansion factor $1 + \varepsilon$, where $0 < \varepsilon < 1/4$ and the exact value of ε will not be important.

Theorem 5.16 (Whitney partition of unity)**.** *Let $E \subset \mathbb{R}^n$ be closed and let $\mathcal{V} = \{Q_i\}_{i\in I}$ be the Whitney cover of $\mathbb{R}^n \setminus E$ by cubes. Then there is a C^∞ partition of unity $\{\varphi_i\}_{i\in I}$ subordinated to $\mathcal{V}^* = \{Q_i^*\}_{i\in I}$, i.e.*

$$\operatorname{supp} \varphi_i \subset Q_i^*, \quad \sum_{i\in I} \varphi_i(x) \equiv 1 \text{ on } \mathbb{R}^n \setminus E$$

such that

$$|D^\alpha \varphi_i(x)| \le \frac{C_\alpha}{\operatorname{dist}(x,E)^{|\alpha|}} \quad \text{for all } \alpha \text{ and all } x \in \mathbb{R}^n \setminus E. \tag{5.5}$$

Proof. Let $h \in C_0^\infty(-(1+\varepsilon), 1+\varepsilon)$, $0 \le h \le 1$, $h \equiv 1$ on $[-1,1]$. Then

$$\varphi(x) = \prod_{k=1}^n h(x_k) \in C_0^\infty\big((-(1+\varepsilon),(1+\varepsilon)^n\big)$$

and

$$\psi_i(x) = \varphi\left(\frac{2(x-z_i)}{\ell_i}\right) \in C_0^\infty(Q_i^*),$$

where z_i is the center of Q_i and ℓ_i is the edge length of Q_i. Note that $\psi_i \equiv 1$ on Q_i. Since $\ell_i \sim \operatorname{diam} Q_i \sim \operatorname{dist}(x,E)$ for $x \in Q_i^*$ we easily obtain that

$$|D^\alpha \psi_i(x)| \le \frac{C_\alpha}{\operatorname{dist}(x,E)^{|\alpha|}} \quad \text{for all } \alpha \text{ and all } x \in \mathbb{R}^n \setminus E. \tag{5.6}$$

[31] This estimate is clearly too large, but we are not interested in getting the best possible estimate. It is only important that the number $N = 12^n$ depends on n only and not on the location or size of a cube.

Finally we define

$$\varphi_i(x) = \frac{\psi_i(x)}{\sum_{j\in I}\psi_j(x)}.$$

The dominator is greater than or equal to 1 on $\mathbb{R}^n \setminus E$. Since in a neighborhood of any point in $\mathbb{R}^n \setminus E$ we have a sum of no more than 12^n terms in the denominator, $\varphi_i \in C_0^\infty(Q_i^*)$ and clearly $\sum_{i\in I}\varphi_i \equiv 1$ on $\mathbb{R}^n \setminus E$. It remains to prove the estimate (5.5). We have $\psi_i = \varphi_i \sum_j \psi_j$ and hence

$$D^\alpha \psi_i = \sum_{\beta\leq\alpha} \binom{\alpha}{\beta} D^\beta \varphi_i \left(\sum_j D^{\alpha-\beta}\psi_j \right),$$

$$D^\alpha \varphi_i = \frac{D^\alpha\psi_i - \sum_{\beta<\alpha}\binom{\alpha}{\beta} D^\beta\varphi_i \left(\sum_j D^{\alpha-\beta}\psi_j\right)}{\sum_j \psi_j}.$$

Again, the dominator is no less than 1. Hence (5.5) follows from (5.6) by induction over $|\alpha|$. □

Proof of Theorem 5.14. Let $\mathcal{V} = \{Q_i\}_{i\in I}$ be the Whitney covering of $\mathbb{R}^n \setminus K$ and let $\{\varphi_i\}_{i\in I}$ be the Whitney partition of unity subordinated to $\mathcal{V}^* = \{Q_i^*\}_{i\in I}$. For each $i \in I$ let $a_i \in K$ be such that[32]

$$\operatorname{dist}(a_i, Q_i) = \operatorname{dist}(K, Q_i).$$

Let $F \in C^m(K)$. We define the Whitney extension $W(F) := \tilde{f}$ by the formula

$$\tilde{f}(x) = \begin{cases} f^0(x) & \text{if } x \in K, \\ \sum_{i\in I}\varphi_i(x) T^m_{a_i}F(x) & \text{if } x \in \mathbb{R}^n \setminus K. \end{cases}$$

Clearly $\tilde{f} \in C^\infty(\mathbb{R}^n \setminus K)$, $\tilde{f} = f^0$ on K, and the operator W is linear. We still need to prove that $\tilde{f} \in C^m(\mathbb{R}^n)$, $D^\alpha \tilde{f} = f^\alpha(x)$ for $x \in K$ and that the operator W is continuous in the sense explained in the theorem. For $|\alpha| \leq m$ let

$$\tilde{f}^\alpha(x) = \begin{cases} f^\alpha(x) & \text{if } x \in K, \\ D^\alpha \tilde{f}(x) & \text{if } x \in \mathbb{R}^n \setminus K. \end{cases}$$

Let η be a modulus of continuity of F such that

$$\|F\|^K_m = |F|^K_m + \eta(\operatorname{diam} K), \qquad \eta(t) = \eta(\operatorname{diam} K) \quad \text{for } t \geq \operatorname{diam} K. \tag{5.7}$$

Assume that K is contained in the interior of a cube Q and let

$$\lambda = \sup_{x\in Q} \operatorname{dist}(x, K).$$

We will prove that there is a constant $C = C(m, n, \lambda)$ such that

$$|\tilde{f}^\alpha(x) - D^\alpha T^m_a F(x)| \leq C\eta(|x-a|)|x-a|^{m-|\alpha|} \quad \text{for } |\alpha| \leq m,\ x \in Q,\ a \in K. \tag{5.8}$$

Note that the estimate is obvious if $x \in K$ with $C = 1$.

[32] a_i is not necessarily unique.

Before proving (5.8) we will show how part (1) of the theorem follows from it. Let δ_j be the multiindex corresponding to the partial derivative $\partial/\partial x_j = D^{\delta_j}$, i.e. $\delta_j = (0,\ldots,0,1,0,\ldots,1)$ with 1 on j-th position. For $|\alpha|<m$, $x\in Q$ and $a\in K$ we have

$$\Big|\tilde{f}^\alpha(x)-\tilde{f}^\alpha(a)-\sum_{j=1}^n(x_j-a_j)f^{\alpha+\delta_j}(a)\Big| \leq |\tilde{f}^\alpha(x)-D^\alpha T^m_a F(x)|+\Big|\sum_{2\leq|\beta|\leq m-|\alpha|} f^{\alpha+\beta}(a)\frac{(x-a)^\beta}{\beta!}\Big|.$$

The first term on the right hand side if $o(|x-y|)$ by (5.8); clearly the second term is $O(|x-a|^2)$. Hence

$$\frac{\partial\tilde{f}^\alpha}{\partial x_j}(a) = f^{\alpha+\delta_j}(a) = \tilde{f}^{\alpha+\delta_j}(a). \tag{5.9}$$

Thus

$$\frac{\partial\tilde{f}^\alpha}{\partial x_j}(x) = \tilde{f}^{\alpha+\delta_j}(x) \quad \text{for all } x\in\mathbb{R}^n,$$

Indeed, if $x\in K$, it follows from (5.9), but if $x\in\mathbb{R}^n\setminus K$ it follows from the very definition of $\tilde{f}^\alpha$. Since $\tilde{f}^0=\tilde{f}$ we obtain by induction that $\tilde{f}$ is m-times differentiable on $\mathbb{R}^n$ and

$$D^\alpha\tilde{f} = \tilde{f}^\alpha \quad \text{on } \mathbb{R}^n.$$

To see that $\tilde{f}\in C^m(\mathbb{R}^n)$ it suffices to observe that for $|\alpha|=m$, $\tilde{f}^\alpha$ is continuous on $\mathbb{R}^n$. Indeed, (5.8) reads as

$$|\tilde{f}^\alpha(x)-\tilde{f}^\alpha(a)| = |\tilde{f}^\alpha(x)-f^\alpha(a)| \leq C\eta(|x-a|).$$

We will now show how to prove continuity of W from (5.8). Let $x\in Q$ and let $a\in K$ be such that $|x-a| = \operatorname{dist}(x,K)$. It follows from (5.8) that

$$|\tilde{f}^\alpha(x)| \leq |D^\alpha T^m_a F(x)| + C\eta(\lambda)\lambda^{m-|\alpha|} \leq \sum_{|\beta|\leq m-|\alpha|}\frac{\lambda^{|\beta|}}{\beta!}|F|^K_m + C\lambda^{m-|\alpha|}\|F\|^K_m,$$

because

$$\eta(\lambda)\leq\eta(\operatorname{diam}K)\leq|F|^K_m+\eta(\operatorname{diam}K) = \|F\|^K_m.$$

Hence

$$|W(F)|^Q_m \leq C(m,n,\lambda)\|F\|^K_m.$$

It suffices to observe now that $\|W(F)\|^Q_m\leq C(m,n)|W(F)|^Q_m$. This is an easy consequence of the Taylor formula with the integral remainder which implies that the Taylor remainder of $\tilde{f}^\alpha = D^\alpha\tilde{f} = D^\alpha W(F)$ on Q can be estimated by

$$\frac{\Big|\tilde{f}^\alpha(y)-\sum_{|\beta|\leq m-|\alpha|}\tilde{f}^{\alpha+\beta}(x)\frac{(y-x)^\beta}{\beta!}\Big|}{|x-y|^{m-|\alpha|}} \leq C\sup_{z\in Q,|\gamma|=m}|\tilde{f}^\gamma(z)| \leq C|W(F)|^Q_m.$$

Thus it remains to prove (5.8). Observe that η is also a modulus of continuity of $D^\alpha F$. Hence the implication from (2) to (3) in Theorem 5.10 applied to F replaced by $D^\alpha F$ yields the existence of $C=C(m,n)$ such that

$$|D^\alpha T^m_a F(x)-D^\alpha T^m_b F(x)| \leq C\eta(|a-b|)(|x-a|^{m-|\alpha|}+|x-b|^{m-|\beta|}) \tag{5.10}$$

for all $a,b\in K$, $x\in\mathbb{R}^n$ and $|\alpha|\leq m$. Since (5.8) is true for $x\in K$ (with $C=1$) we can assume that $x\notin K$. We have

$$\tilde{f}(x)-T^m_aF(x)=\sum_{i\in I}\varphi_i(x)\big(T^m_{a_i}F(x)-T^m_aF(x)\big)$$

and the Leibniz rule gives

$$\tilde{f}^\alpha(x)-D^\alpha T^m_aF(x)=\sum_{\beta\leq\alpha}\binom{\alpha}{\beta}\underbrace{\Big(\sum_{i\in I}D^\beta\varphi_i(x)D^{\alpha-\beta}\big(T^m_{a_i}F(x)-T^m_aF(x)\big)\Big)}_{S_\beta(x)}.$$

We will estimate $|S_0(x)|$ first. For $x\in\operatorname{supp}\varphi\subset Q_i^*$ we have

$$|x-a_i|\leq\operatorname{dist}(x,Q_i)+\operatorname{dist}(Q_i,a_i)\leq\frac{\operatorname{diam}Q_i}{4}+\operatorname{dist}(Q_i,K)\leq\frac{17}{4}\operatorname{diam}Q_i\leq\frac{17}{3}|x-a|,$$

because

$$\operatorname{diam}Q_i\leq\operatorname{dist}(Q_i,K)\leq\operatorname{dist}(Q_i,a)\leq|x-a|+\frac{\operatorname{diam}Q_i}{4}.$$

Hence

$$|a-a_i|\leq|a-x|+|x-a_i|\leq\Big(1+\frac{17}{3}\Big)|x-a|<7|x-a|.$$

Concavity of η yields $\eta(|a-a_i|)\leq 7\eta(|x-a|)$ and hence

$$|S_0(x)|\leq C(m,n)\eta(|x-a|)|x-a|^{m-|\alpha|}.$$

Now let $\beta\neq 0$. Since $\sum_i\varphi_i=1$, $\sum_i D^\beta\varphi_i=0$ and hence

$$\sum_{i\in I}D^\beta\varphi_i(x)T^m_aF(x)=\sum_{i\in I}D^\beta\varphi_i(x)T^m_bF(x)=0$$

for all $b\in K$. Thus

$$S_\beta(x)=\sum_{i\in I}D^\beta\varphi_i(x)D^{\alpha-\beta}\big(T^m_{a_i}F(x)-T_bF(x)\big).$$

Take $b\in K$ such that $|x-b|=\operatorname{dist}(x,K)$. The estimates

$$|x-a_i|\leq\frac{17}{3}|x-a|,\qquad\eta(|a-a_i|)\leq 7\eta(|x-a|)$$

were proved for any $a\in K$. In particular they are true if we replace a by b, so

$$|x-a_i|\leq\frac{17}{3}\operatorname{dist}(x,K),\qquad\eta(|b-a_i|)\leq 7\eta(\operatorname{dist}(x,K)).$$

Hence (5.5) and (5.10) yield

$$\begin{aligned}
&|D^\beta\varphi_i(x)D^{\alpha-\beta}(T^m_{a_i}F(x)-T^m_bF(x))|\\
&\leq\ \frac{C}{\operatorname{dist}(x,K)^{|\beta|}}\eta(|a_i-b|)(|x-a_i|^{m-|\alpha-\beta|}+|x-b|^{m-|\alpha-\beta|})\\
&\leq\ C'\eta(\operatorname{dist}(x,K))\operatorname{dist}(x,K)^{m-|\alpha|} \qquad (5.11)\\
&\leq\ C'\eta(|x-a|)|x-a|^{m-|\alpha|}.
\end{aligned}$$

Note that this estimate is true for any $\beta>0$. Indeed, the above calculation requires that $|\alpha-\beta|\leq m$, but if $|\alpha-\beta|>m$, then $D^{\alpha-\beta}(T^m_{a_i}F(x)-T^m_bF(x))=0$ and the above inequality is trivially true. Thus for any $\beta > 0$, $|S_\beta(x)| \leq C\eta(|x-a|)|x-a|^{m-|\alpha|}$, because for any $x \in \mathbb{R}^n \setminus K$, no more than 12^n terms in the sum that defines S_β are non-zero.

We still need to prove part (2).

Lemma 5.17. *For $|\alpha| > m$ there is a constant $C(m,n,|\alpha|,\lambda)$ such that*

$$|D^\alpha W(F)(x)| \leq C\frac{\eta(\operatorname{dist}(x,K))}{\operatorname{dist}(x,K)^{|\alpha|-m}} \quad \textit{for all } x \in Q \setminus K.$$

Proof. For $x \in Q \setminus K$, $a \in K$, $||x-a| = \operatorname{dist}(x,K)$ we have

$$W(F)(x) = T^m_a F(x) + \sum_{i\in I} \varphi_i(x)(T^m_{a_i}F(x) - T_a F(x)).$$

If $|\alpha| > m$, then $D^\alpha T^m_a F(x) = 0$ and hence

$$|D^\alpha W(F)(x)| \leq \sum_{\beta\leq\alpha}\binom{\alpha}{\beta}\sum_{i\in I}|D^\beta\varphi_i(x)D^{\alpha-\beta}(T^m_{a_i}F(x) - T^m_a F(x))|.$$

Now the result follows from the estimate (5.11) which we proved for all β. □

In order to prove (2) it suffices to show that

$$|D^\alpha\tilde{f}(x) - D^\alpha\tilde{f}(y)| \leq C(m,n,\lambda)\eta(|x-y|) \quad \text{for } x,y \in \overline{Q},\ |\alpha| \leq m. \tag{5.12}$$

Indeed, the estimated for lower order derivatives will follow from the Taylor formula with the remainder in the integral form. If $x \in K$ or $y \in K$, then (5.12) follows from (5.8). Thus we may assume that $x, y \in \overline{Q} \setminus K$. We need to consider two cases.

CASE 1. $\operatorname{dist}(x,K) \geq 2|x-y|$. This implies that the interval $\overline{xy}$ does not touch K. Hence the mean value theorem yields

$$D^\alpha\tilde{f}(x) - D^\alpha\tilde{f}(y) = \nabla(D^\alpha\tilde{f})(z)\cdot(x-y), \quad \text{for some } z \in \overline{xy}$$

which together with Lemma 5.17 lead to

$$|D^\alpha\tilde{f}(x) - D^\alpha\tilde{f}(y)| \leq C(m,n,\lambda)\frac{\eta(\operatorname{dist}(z,K))}{\operatorname{dist}(z,K)}\,|x-y|.$$

Since $\operatorname{dist}(z,K) \geq |x-y|$, concavity of η yields

$$\frac{\eta(\operatorname{dist}(z,K))}{\operatorname{dist}(z,K)}\,|x-y| = \frac{\eta(\frac{\operatorname{dist}(z,K)}{|x-y|}|x-y|)}{\operatorname{dist}(z,K)}\,|x-y| \leq \eta(|x-y|).$$

CASE 2. $\operatorname{dist}(x,K) < 2|x-y|$. Let $a, b \in K$ be such that $|x-a| = \operatorname{dist}(x,K)$, $|y-b| = \operatorname{dist}(y,K)$. It easily follows from the triangle inequality that

$$|x-a| \leq 2|x-y|, \quad |y-b| \leq 3|x-y|, \quad |a-b| \leq 6|x-y|.$$

We have

$$\begin{aligned}|D^\alpha\tilde{f}(x) - D^\alpha\tilde{f}(y)| &\leq |D^\alpha\tilde{f}(x) - D^\alpha\tilde{f}(a)| + |D^\alpha\tilde{f}(a) - D^\alpha\tilde{f}(b)| + |D^\alpha\tilde{f}(y) - D^\alpha\tilde{f}(b)| \\ &\leq C(\eta(|x-a|) + \eta(|a-b|) + \eta(|y-b|)) \leq C\eta(|x-y|),\end{aligned}$$

because (5.12) is satisfied if at least one of the points is in K. The last inequality follows from concavity of η. □

Example 5.18. In Subsection 2.2 we constructed a C^1 mapping $F:\mathbb{R}^n\to\mathbb{R}^n$ which maps the van Koch snowflake K homeomorphically onto a circle, but $DF=0$ on K. It directly follows from the estimates in Subsection 2.2 that the modulus of continuity of the Whitney 1-jet $(f,L)=(f,0)$ is $\eta(t)=Ct^\alpha$, where $\alpha=(\log 4/\log 3)-1$. Hence the Whitney extension is not only C^1, but it is $C^{1,\alpha}$. More general results of this type will be provided in Section **??**.

5.1. **Extension from domains.** Suppose $f\in C^m(\Omega)$, where Ω is abounded and open set. It is an interesting problem to find conditions that will guarantee existence of $F\in C^m(\mathbb{R}^n)$ such that $F|_\Omega=f$. Clearly the problem must be related to the Whitney extension theorem.

6. Isoperimetric inequality

6.1. **The Brunn-Minkowski inequality.** The isoperimetric theorem states that *among all sets with given volume, a ball has the least surface area.* Since this result deals with all compact subsets of $\mathbb{R}^n$ (possibly very irregular), the notion of the surface area needs to be clarified.

For a compact set $A\subset\mathbb{R}^n$ let

$$A_h=\{x:\ \mathrm{dist}\,(x,A)\le h\}.$$

Then $|A_h|-|A|$ is the Lebesgue measure of the h-strip around A.

Definition 6.1. The *Minkowski content* of A is

$$\mu_+(A)=\liminf_{h\to 0^+}\frac{|A_h|-|A|}{h}\,.$$

If A is a closure of a bounded open set with smooth boundary, then

$$\mu_+(A)=\mathcal{H}^{n-1}(\partial A)$$

Thus it is reasonable to regard $\mu_+(A)$ as a generalization of the measure of the boundary of A.

The isoperimetric theorem can be deduced from the isoperimetric inequality.

Theorem 6.2 (Isoperimetric inequality)**.** *For any compact set $A\subset\mathbb{R}^n$ we have*

$$|A|^{\frac{n-1}{n}}\le n^{-1}\omega_n^{-1/n}\mu_+(A).$$

Indeed, for a ball B we have equality in the isoperimetric inequality. Now if A is a set with $|A|=|B|$, then

$$\mu_+(A)\ge n\omega_n^{1/n}|A|^{\frac{n-1}{n}}=n\omega_n^{1/n}|B|^{\frac{n-1}{n}}=\mu_+(B).$$

Hence the measure of the boundary of A (i.e. $\mu_+(A)$) is greater than or equal to the measure of the boundary of a ball of equal ball which is the isoperimetric theorem.

We will prove the isoperimetric inequality as a consequence of the Brunn-Minkowski inequality. For sets $A,B\subset\mathbb{R}^n$ we define

$$A+B=\{x+y:\ x\in A,\ y\in B\}.$$

Theorem 6.3 (Brunn-Minkowski inequality)**.** *For compact sets $A, B \subset \mathbb{R}^n$ we have*
$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

Before proving this result we will show how to conclude the isoperimetric inequality. Since $A_h = A + \overline{B}(0,h)$, the Brunn-Minkowski inequality yields
$$|A_h|^{1/n} \geq |A|^{1/n} + \omega_n^{1/n} h$$
and hence
$$\omega_n^{1/n} \leq \liminf_{h\to 0} \frac{|A_h|^{1/n} - |A|^{1/n}}{h} \overset{\heartsuit}{=} n^{-1}|A|^{\frac{n-1}{n}} \liminf_{h\to 0} \frac{|A_h| - |A|}{h} = n^{-1}|A|^{\frac{n-1}{n}} \mu_+(A)$$
which readily implies the isoperimetric inequality.

Exercise 6.4. *Prove $\heartsuit$.*

Proof of the Brunn-Minkowski inequality. We split the proof in three steps.

Step 1. Suppose that A and B are rectangular boxes with edge lengths $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ respectively. Then $A+B$ is also a rectangular box with sides $\{a_i+b_i\}_{i=1}^n$. In this case the Brunn-Minkowski inequality reduces to
$$\Big(\prod_{i=1}^n (a_i+b_i)\Big)^{1/n} \geq \Big(\prod_{i=1}^n a_i\Big)^{1/n} + \Big(\prod_{i=1}^n b_i\Big)^{1/n}.$$
This inequality easily follows from the arithmetic-geometric mean inequality after we rewrite it as
$$\Big(\prod_{i=1}^n \frac{a_i}{a_i+b_i}\Big)^{1/n} + \Big(\prod_{i=1}^n \frac{b_i}{a_i+b_i}\Big)^{1/n} \leq 1.$$

Step 2. Assume now that A and B are finite unions of rectangular boxes with disjoint interiors.[33] We will prove this case using induction with respect to the total number k of boxes in $A \cup B$. The case $k=2$ was proved in Step 1. Assume now that the claim is true for some $k \geq 2$; it remains to prove it for $k+1$. Thus suppose that the total number of boxes in $A\cup B$ is $k+1 \geq 3$. One of the sets, say A has at least two boxes. Let π by a hyperplane such that two fixed boxes in A are on opposite sides of half-spaces π^pm generated by π. By translating the sets[34] we may assume that $0 \in \pi$. Let $A^\pm = \pi^\pm \cap A$ and $B^\pm = \pi^\pm \cap B$. Let $\lambda = |A^+|/|A|$. Note that a parallel translation of B does not change $|A|$, $|B|$ and $|A+B|$, so we can move B to such a position that $|B^+|B| = \lambda$. The total number of boxes in $A^+ \cup B^+$ is no more than k.[35]. The same is true for $A^- \cup B^-$. Thus we can apply the induction hypothesis to $A^\pm \cup B^\pm$. Note that $A^+ + B^+ \subset \pi^+$ and $A^- + B^- \subset \pi^-$,[36] so the two sets have disjoint interiors and hence
$$\begin{aligned} |A+B| &\geq |A^+ + B^+| + |A^- + B^-| \geq (|A^+|^{1/n} + |B^+|^{1/n})^n + (|A^-|^{1/n} + |B^-|^{1/n})^n \\ &= \lambda(|A|^{1/n} + |B|^{1/n})^n + (1-\lambda)(|A|^{1/n} + |B|^{1/n})^n = (|A|^{1/n} + |B|^{1/n})^n \end{aligned}$$

[33] We do not assume that boxes in A have disjoint interiors from boxes in B.

[34] Translation does not change volumes of A, B and $A+B$.

[35] Because at least one of the boxes from A is in π^-.

[36] Because $0 \in \pi$.

which proves the inequality we wanted to prove.

STEP 3. Now we are ready to prove the inequality in the general case. If K is compact, K_ε is its ε-neighborhood. Since the intersection of all such neighborhoods is K it follows that $|K_\varepsilon| \to |K|$ as $\varepsilon \to 0$. We can find sets $\hat{A}_\varepsilon$ and $\hat{B}_\varepsilon$ that are unions of finite numbers of boxes with disjoint interiors such that $A \subset \hat{A}_\varepsilon \subset A_\varepsilon$, $B \subset \hat{B}_\varepsilon \subset B_\varepsilon$. Hence $\hat{A}_\varepsilon + \hat{B}_\varepsilon \subset (A+B)_{2\varepsilon}$. According to Step 2, the Brunn-Minkowski inequality is true for $\hat{A}_\varepsilon + \hat{B}_\varepsilon$, so

$$|(A+B)_{2\varepsilon}|^{1/n} \geq |\hat{A}_\varepsilon + \hat{B}_\varepsilon|^{1/n} \geq |\hat{A}_\varepsilon|^{1/n} + |\hat{B}_\varepsilon|^{1/n} \geq |A|^{1/n} + |B|^{1/n}$$

and the result follows upon taking the limit as $\varepsilon \to 0$. □

As another application of the Brunn-Minkowski inequality we will prove the isodiamteric inequality. This inequality will play a fundamental role in Section **??** when we will prove that the Hausforff measure $\mathcal{H}^n$ in $\mathbb{R}^n$ coincides with the Lebesgue measure.

Theorem 6.5 (Isodiamteric inequality)**.** *If $E \subset \mathbb{R}^n$ is compact, then*

$$|E| \leq |\mathrm{conv}\,(E)| \leq \omega_n \Big(\frac{\mathrm{diam}\, E}{2}\Big)^n,$$

where conv (E) *is the convex hall of E.*

Proof. Since taking the convex hall of a set does not increase its diameter, we can assume that E is convex, $E = \mathrm{conv}\,(E)$. It suffices to show that there is a compact central symmetric set F centered at 0 such that

$$|F| \geq |E|, \qquad \mathrm{diam}\, F \leq \mathrm{diam}\, E \tag{6.1}$$

Indeed, since F is central symmetric, it is contained in the ball $\overline{B}(0, \mathrm{diam}\, F/2)$[37] and hence

$$|E| \leq |F| \leq \omega_n \Big(\frac{\mathrm{diam}\, F}{2}\Big)^n \leq \Big(\frac{\mathrm{diam}\, E}{2}\Big)^n$$

Let E' be the symmetric image of E with respect to 0 and let $F = (E+E')/2$. Clearly F is compact and symmetric with respect to 0. It suffices to show that it satisfies (6.1). It follows from the Brunn-Minkowski inequality that

$$|F|^{1/n} = \Big|\frac{E}{2} + \frac{E'}{2}\Big|^{1/n} \geq \Big|\frac{E}{2}\Big|^{1/n} + \Big|\frac{E'}{2}\Big|^{1/n} = \frac{|E|^{1/n}}{2} + \frac{|E'|^{1/n}}{2} = |E|^{1/n}.$$

Let now $x, y \in F$ be such that $\mathrm{diam}\, F = |x-y|$. Then there are points $x', y' \in E$ and $x'', y'' \in E'$ such that $x = (x'+x'')/2$, $y = (y'+y'')/2$. We have

$$\begin{aligned} \mathrm{diam}\, F &= |x-y| = \frac{1}{2}\,|x'+x''-y'-y''| \leq \frac{1}{2}(|x'-y'| + |x''-y''|) \\ &\leq \frac{1}{2}(\,\mathrm{diam}\, E + \,\mathrm{diam}\, E') = \,\mathrm{diam}\, E. \end{aligned}$$

□

[37] Why?

6.2. **Boxing inequality.** The next result can also be regarded as a version of the isoperimetric inequality in a sense that it gives an estimate of the size of a set (Hausdorff content) in terms of the Hausdorff measure of the boundary.[38]

Theorem 6.6 (Boxing inequality)**.** *Let Ω be a bounded domain in $\mathbb{R}^n$. Then there is a covering of Ω by balls of radii r_i, $i = 1, 2, \dots$ such that*

$$\sum_{i=1}^{\infty} r_i^{n-1} \leq C(n)\mathcal{H}^{n-1}(\partial\Omega).$$

In other words

$$\mathcal{H}^{n-1}_{\infty}(\Omega) \leq C(n)\mathcal{H}^{n-1}(\partial\Omega).$$

Proof.

Lemma 6.7. *Let $\Omega \subset \mathbb{R}^n$ be open. If B_r is a ball of rdius r such that $|B_r \cap \Omega| = \frac{1}{2}|B_r|$, then*

$$\mathcal{H}^{n-1}(B_r \cap \partial\Omega) \geq C(n) r^{n-1}.$$

Proof. Let $\chi = \chi_{B_r\cap\Omega}$ and $\psi = \chi_{B_r\setminus\Omega}$. For $0 \neq z \in \mathbb{R}^n$, let P_z be the orthogonal projection onto the $(n-1)$-dimensional hyperplane $z^{\perp}$ orthogonal to z. For any set $E \subset \mathbb{R}^n$ we have

$$\mathcal{H}^{n-1}(P_z(E)) \leq \mathcal{H}^{n-1}(E), \tag{6.2}$$

because the projection P_z is 1-Lipschitz. According to the Fubini theorem we have

$$\begin{aligned}
\frac{\omega_n^2 r^{2n}}{4} &= |\Omega \cap B_r|\,|B_r \setminus \Omega| = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \chi(x)\psi(y)\,dy\,dx \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \chi(x)\psi(x+y)\,dy\,dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \chi(x)\psi(x+y)\,dx\,dy \\
&= \int_{|y|\leq 2r} |\{x:\, x \in B_r \cap \Omega \text{ and } x+y \in B_r \setminus \Omega\}|\,dy = \heartsuit
\end{aligned}$$

Indeed, if $|y| > 2r$, then $\chi(x)\psi(x+y) = 0$. Hence we can restrict the domain of the y-integration to the ball $|y| \leq 2r$. The integrand equals 1 on the set of all x such that $x \in B_r \cap \Omega$ and $x + y \in B_r \setminus \Omega$, zero otherwise.

Each interval $\overline{x(x+y)}$ intersects with $B_r \cap \partial\Omega$. Hence

$$P_y(\{x:\, x \in B_r \cap \Omega \text{ and } x+y \in B_r \setminus \Omega\}) \subset P_y(B_r \cap \partial\Omega) \subset y^{\perp}. \tag{6.3}$$

For each $z \in y^{\perp}$ let ℓ_z be the line through z, orthogonal to $y^{\perp}$. Clearly

$$\mathcal{H}^1(\ell_z \cap \{x:\, x \in B_r \cap \Omega \text{ and } x+y \in B_r \setminus \Omega\}) \leq 2r, \tag{6.4}$$

because $x \in B_r$ and $2r$ is the diameter of the ball. Hence (6.3), (6.4) and the Fubini theorem yield

$$\begin{aligned}
&|\{x:\, x \in B_r \cap \Omega \text{ and } x+y \in B_r \setminus \Omega\}| \\
&\quad\leq\ 2r \cdot \mathcal{H}^{n-1}(P_y(\{x:\, x \in B_r \cap \Omega \text{ and } x+y \in B_r \setminus \Omega\})) \\
&\quad\leq\ 2r \cdot \mathcal{H}^{n-1}(P_y(B_r \cap \partial\Omega)) \\
&\quad\leq\ 2r \cdot \mathcal{H}^{n-1}(B_r \cap \partial\Omega).
\end{aligned}$$

[38] I have no idea why it is called the boxing inequality.

The last inequality follows from (6.2). Thus

$$\heartsuit \leq \omega_n (2r)^n \cdot 2r \cdot \mathcal{H}^{n-1}(B_r \cap \partial\Omega)$$

which easily imples

$$\mathcal{H}^{n-1}(B_r \cap \partial\Omega) \geq \frac{\omega_n r^{n-1}}{2^{n+3}}.$$

Clearly the constant in this inequality is not optimal. $\square$

Now we can complete the proof of th boxing inequality. For each $x \in \Omega$ there is $\rho_x > 0$ such that[39]

$$|B(x,\rho_x) \cap \Omega| = \frac{1}{2}|B(x,\rho_x)|.$$

The balls $\{B(x,\rho_x)\}_{x\in\Omega}$ form a covering of Ω. By Theorem 3.5 we can select a countable number of pairwise disjoint balls $\{B(x_i,\rho_{x_i})\}_{i=1}^\infty$ such that

$$\Omega \subset \bigcup_{i=1}^\infty B(x_i, 5\rho_{x_i}).$$

Let $r_i = 5\rho_{x_i}$. We have

$$\mathcal{H}^{n-1}(\partial\Omega) \geq \sum_{i=1}^\infty \mathcal{H}^{n-1}(B(x_i,\rho_{x_i}) \cap \partial\Omega) \geq C(n) \sum_{i=1}^\infty \rho_{x_i}^{n-1} = C(n) 5^{-(n-1)} \sum_{i=1}^\infty r_i^{n-1}.$$

The first inequality follows from the fact that the balls $B(x_i,\rho_{x_i})$ are pairwise disjoint, while second inequality is a direct consequence of Lemma 6.7. The proof is complete. $\square$

6.3. **Sobolev inequality.** In this section we will prove the Sobolev inequality; in the next section we will see that the Sobolev inequality with the best constant for $p=1$ is equivalent to the isoperimetric inequality.

Theorem 6.8 (Sobolev inequality)**.** *Let $1 \leq p < n$ and $p^* = np/(n-p)$. If u is a compactly supported Lipschitz function, then*

$$\left(\int_{\mathbb{R}^n} |u(x)|^{p^*}\, dx\right)^{1/p^*} \leq C(n,p) \left(\int_{\mathbb{R}^n} |\nabla u(x)|^p\, dx\right)^{1/p}. \tag{6.5}$$

Proof. First of all we can assume that $u \in C_0^\infty$. Indeed, if we know the result for C_0^∞ functions and u is a compactly supported and Lipschitz, then applying the result to the approximation by convolution $u_\varepsilon = u * \varphi_\varepsilon \in C_0^\infty$ and letting $\varepsilon \to 0$ yields the result for u, because u_ε converges to u uniformly and $\nabla u_\varepsilon = \nabla u * \varphi_\varepsilon$ converges to u in L^q for any $1 \leq q < \infty$.

Step 1. $p=1$. This is the crucial step in the proof. As we will see later, the general case $1 \leq p < n$ easily follows from the case $p=1$. We have

$$|u(x)| \leq \int_{-\infty}^{x_1} |D_1 u(t_1,x_2,\ldots,x_n)|\, dt_1 \leq \int_{-\infty}^{\infty} |D_1 u(t_1,x_2,\ldots,x_n)|\, dt_1.$$

[39] Why?

Here by D_i we denote the partial derivative with respect to i-th coordinate. Analogous inequalities hold with x_1 replaced by $x_2, \ldots, x_n$. Hence

$$|u(x)|^{\frac{n}{n-1}} \leq \prod_{i=1}^{n} \left(\int_{-\infty}^{\infty} |D_i u| \, dt_i \right)^{\frac{1}{n-1}}.$$

Now we integrate both sides with respect to $x_1 \in \mathbb{R}$. Note that exactly one integral in the product on the right hand side does not depend on x_1, so

$$\int_{-\infty}^{\infty} |u(x)|^{\frac{n}{n-1}} \, dx_1 \leq \left(\int_{-\infty}^{\infty} |D_1 u| \, dt_1 \right)^{\frac{1}{n-1}} \int_{-\infty}^{\infty} \prod_{i=2}^{n} \left(\int_{-\infty}^{\infty} |D_i u| \, dt_i \right)^{\frac{1}{n-1}} dx_1.$$

To estimate the product on the right hand side we need to recall the Hölder inequality:

Lemma 6.9. *If $p_1, \ldots, p_k > 0$, $p_1 + \ldots + p_k = 1$, then*

$$\int_X \Big| \prod_{i=1}^{k} f_i \Big| \, d\mu \leq \prod_{i=1}^{k} \Big(\int_X |f_i|^{p_i} \, d\mu \Big)^{1/p_i}.$$

This result is an easy consequence of a standard version of Hölder's inequality via an induction argument. In particular if $p_1 = \ldots = p_{n-1} = 1/(n-1)$, then

$$\int_X \prod_{i=1}^{n-1} |f_i|^{1/(n-1)} \, d\mu \leq \prod_{i=1}^{n-1} \Big(\int_X |f_i| \, d\mu \Big)^{1/(n-1)}.$$

Hence

$$\int_{-\infty}^{\infty} |u(x)|^{\frac{n}{n-1}} \, dx_1 \leq \left(\int_{-\infty}^{\infty} |D_1 u| \, dt_1 \right)^{\frac{1}{n-1}} \prod_{i=2}^{n} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |D_i u| \, dt_i \, dx_1 \right)^{\frac{1}{n-1}}.$$

Again, on the right hand side we have a product of n integrals with one integral independent of x_2. We integrate both sides with respect to $x_2 \in \mathbb{R}$ and apply Hölder's inequality in a similar way as above. This leads to an inequality which we then integrate with respect to $x_3 \in \mathbb{R}$ etc. In the end, we obtain the inequality

$$\int_{\mathbb{R}^n} |u(x)|^{\frac{n}{n-1}} \, dx \leq \prod_{i=1}^{n} \left(\int_{\mathbb{R}^n} |D_i u| \, dx \right)^{\frac{1}{n-1}},$$

which readily implies (6.5) with $p = 1$.

Step 2. General case. Let $f = |u|^{p(n-1)/(n-p)}$. The function f is of class C^1, because $\alpha = p(n-1)/(n-p) > 1$.

Exercise 6.10. *Prove that if $u \in C_0^\infty(\mathbb{R}^n)$ and $\alpha > 1$, then $|u|^\alpha \in C^1$.*

It follows from the definition of f that $f^{n/(n-1)} = |u|^{np/(n-p)}$. Applying (6.5) with $p = 1$ to f yields

$$\left(\int_{\mathbb{R}^n} |u|^{\frac{np}{n-p}} \right)^{\frac{n-1}{n}} = \left(\int_{\mathbb{R}^n} f^{\frac{n}{n-1}} \right)^{\frac{n-1}{n}} \leq C \int_{\mathbb{R}^n} |\nabla f|.$$

We have[40]

$$|\nabla f| = \frac{p(n-1)}{n-p}|u|^{\frac{n(p-1)}{n-p}}|\nabla u|,$$

so the Hölder inequality implies

$$\begin{aligned}\Big(\int_{\mathbb{R}^n}|u|^{\frac{np}{n-p}}\Big)^{\frac{n-1}{n}} &\leq C\int_{\mathbb{R}^n}|u|^{\frac{n(p-1)}{n-p}}|\nabla u| \leq C\Big(\int_{\mathbb{R}^n}|u|^{\frac{n(p-1)}{n-p}\frac{p}{p-1}}\Big)^{\frac{p-1}{p}}\Big(\int_{\mathbb{R}^n}|\nabla u|^p\Big)^{1/p}\\ &= C\Big(\int_{\mathbb{R}^n}|u|^{\frac{np}{n-p}}\Big)^{\frac{p-1}{p}}\Big(\int_{\mathbb{R}^n}|\nabla u|^p\Big)^{1/p}\end{aligned}$$

and the theorem follows, because

$$\frac{n-1}{n} - \frac{p-1}{p} = \frac{n-p}{np}\,.$$

$\square$

The constant C obtained in the above proof is not the best possible. As we will see now, the Sobolev inequality for $p=1$ with the best constant is related to the isoperimetric inequality and it actually can be regarded as an integral version of the isoperimetric inequality.

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

[40]Chain rule.