NOTES ON QUANTITATIVE RECTIFIABILITY AND DIFFERENTIABILITY

ROBERT YOUNG

Differentiability measures how well a function can be approximated by affine functions. Rectifiability likewise measures how well a set can be approximated by planes. But both of these notions operate at *infinitesimal* scales; they deal with the properties of limits. In these notes, we will study quantitative versions of these notions that operate at *local* scales; scales that are small but bounded away from zero. How well can a function or a set be approximated by affine functions or sets at local scales? How often can an function or set fail to be approximated by an affine function or set? How can we use questions like these to study the geometry and analysis of sets and functions?

Tentative outline:

- Coarse differentiation of curves
- From curves to spaces: Rademacher's theorem
- Pansu's theorem and embeddings
- Rectifiability and the Jones Traveling Salesman Problem
- Uniform rectifiability
- Surfaces in $\mathbb{R}^n$

1. Coarse differentiation of curves

Let's start with the simplest case: What can we say about the local or infinitesimal structure of maps $f: \mathbb{R} \to X$?

Clearly, we need some conditions on f and X. For example, we take f to be Lipschitz, absolutely continuous, or bounded variation, and we can take X to be a Hilbert space, a Banach space, or a metric space.

From analysis, we know that

Theorem 1.1. *If $f: \mathbb{R} \to \mathbb{R}$ is absolutely continuous map, then f is differentiable almost everywhere. That is, for almost every $x \in \mathbb{R}$, there is an $f'(x) \in X$ such that*

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x).$$

In particular, if f is Lipschitz, f is differentiable almost everywhere. But this is an infinitesimal result – what about local? What's the largest segment we can expect to find where f is ϵ–close to affine?

Let's start with a coarse differentiation theorem due to Eskin, Fisher, and Whyte.

Date: April 16, 2020.

R.Y. was supported by NSF grant 1612061.

Definition 1.2. Let X be a metric space and let $f: \mathbb{R} \to X$. Let $I = [a,b]$ be an interval of length L, let $\epsilon = \frac{1}{n}$ for some $n \in \mathbb{N}$. We say that f is *ϵ–efficient* on I if

$$\sum_{i=0}^{n-1} d_X(f(a+(i+1)\frac{L}{n}), f(a+i\frac{L}{n}))| \leq d_X(f(b), f(a)) + \epsilon L.$$

If $X = \mathbb{R}$, this is also known as ϵ–monotone – the inequality implies that graph can't backtrack by more than about ϵ

Then the following lemma is simple:

Lemma 1.3. *Let $f: [0,1] \to X$ be 1–Lipschitz (bounded variation also works) and let $\epsilon > 0$. There is an interval $I \subset [0,1]$ of length $L > \epsilon^{\frac{1}{\epsilon}}$ such that f is ϵ–efficient on I.*

Proof. Let

$$\ell_k = \sum_{i=0}^{\epsilon^{-k}} d(f(i\epsilon^k), f((i+1)\epsilon^k)|.$$

By the Lipschitz property, $\ell_k \leq 1$. If there is no ϵ–efficient interval, then $\ell_{k+1} > \ell_k + \epsilon$ and, by induction, $\ell_k \geq \epsilon k$. This is a contradiction when $k > \epsilon^{-1}$, so f is ϵ–efficient on some interval of length at most $\epsilon^{\epsilon^{-1}}$. □

This illustrates a basic principle: If we can break down some finite geometric quantity (in this case, length) into contributions from many different intervals, then there must be intervals that contribute less than ϵ. We can then prove theorems by showing that those intervals are geometrically or analytically nice in some way.

Ideally, the quantity should be *coercive*: intervals on which the quantity is zero should lie in some nice class, and intervals on which the quantity is ϵ should be close to that class. In a general metric space, we don't have that; it's easy to construct metric spaces that have ϵ–efficient curves that are far from 0–efficient (length-minimizing) curves. But if X is more specific, things are easier.

For example, a Lipschitz function to $\mathbb{R}^k$ is coarsely differentiable. For $\epsilon > 0$, we say that f is *ϵ–coarsely differentiable* on I if there is an affine function such that $|f(t) - \lambda(t)| \leq \epsilon \ell(I)$ for every $t \in I$.

Proposition 1.4. *Let $f: [0,1] \to \mathbb{R}$ be 1–Lipschitz, $\epsilon > 0$. There is an interval $I \subset [0,1]$ such that f is ϵ–*coarsely differentiable *on I.*

Proof. Consider the graph $g: [0,1] \to \mathbb{R}^2$, $g(t) = (t, f(t))$. This is Lipschitz, so for any $\delta > 0$, there is an interval I on which g is δ–efficient. By the Pythagorean Theorem, $g(I)$ is in the $\sim \sqrt{\delta}\ell(I)$–neighborhood of a line segment $L \in \mathbb{R}^2$, say $L = \{(t, \lambda(t)\}$, where $\lambda(t) = mt + b$. Since f was 1–Lipschitz, we can take $|m| \leq 1$. Then $d((x,y), L) \approx |y - \lambda(x)|$, so for $t \in I$, we have

$$|f(t) - \lambda(t)| \approx d((t, f(t)), L) \lesssim \sqrt{\delta}\ell(I).$$

□

It follows directly that the same is true for maps to $\mathbb{R}^k$.

In fact, intervals like this are abundant: the notions of ϵ–efficient, ϵ–coarse differentiable are scale-invariant, so both of these results apply to a Lipschitz function on any interval. So any subinterval of $[0,1]$ contains a smaller interval (not too small) on which f is ϵ–efficient or ϵ–coarse differentiable. And we can do better:

Definition 1.5. Let $D \subset \mathbb{R} \times \mathbb{R}^+$. Let $D_r = \{x \mid (x,r) \in D\}$. We say that D is a *(C-)Carleson set* if there is a $C > 0$ such that for every interval $I \subset \mathbb{R}$ of length L,

$$\int_0^L |D_r \cap I| \frac{\mathrm{d}r}{r} = \int_{-\infty}^{\log L} |D_{e^t} \cap I| \,\mathrm{d}t \leq CL, \tag{1}$$

where $|D_r \cap I|$ is the Lebesgue measure of $D_r \cap I$.

Thus, for example, for almost every point x, the set of $t > 0$ such that $(x, e^{-t}) \in S$ has finite measure, and on average, it has measure at most C. Roughly, "most" x's and "most" t's are not in S.

Some quick examples and properties that are easy to check:

(1) For $0 < a < b$, the set $\mathbb{R} \times [a,b]$ is Carleson
(2) For any $p \in \mathbb{R}$, the set $T = \{(x,r) \mid |x-p| < r\}$ is Carleson. That is, there can be points such that $(x, e^{-t}) \in T$ for many different t, but there can't be too many: $I \times \mathbb{R}^+$ is not Carleson.
(3) The union of finitely many Carleson sets is Carleson.

Basic principle: if you find a good coercive quantity, it should be small away from a Carleson set.

Theorem 1.6. *Let $f\colon \mathbb{R} \to X$ be 1–Lipschitz, $\epsilon > 0$ and let $S_\epsilon \subset \mathbb{R} \times \mathbb{R}^+$ be the set*

$$S_\epsilon = \{(x,r) \mid f \text{ is not } \epsilon\text{–efficient on } [x - \frac{r}{2}, x + \frac{r}{2}]\}.$$

Then S_ϵ is Carleson.

Proof. By symmetry, it suffices to check the Carleson condition for the interval $[0,1]$.

As before, let

$$\ell_k = \sum_{i=0}^{\epsilon^{-k}} d(f(i\epsilon^k), f((i+1)\epsilon^k)|.$$

Then

$$\ell_{k+1} \geq \ell_k + \epsilon^{k+1} \sum_i \mathbf{1}_{S_\epsilon}((i + \frac{1}{2})\epsilon^k, \epsilon^k),$$

so we have a sort of coarse Carleson condition:

$$\sum_{k=0}^{\infty} \sum_i \epsilon^{k+1} \mathbf{1}_{S_\epsilon}((i + \frac{1}{2})\epsilon^k, \epsilon^k) \leq 1.$$

The same thing holds under translation:

$$\sum_{k=0}^{\infty} \sum_i \epsilon^{k+1} \mathbf{1}_{S_\epsilon}(t + (i + \frac{1}{2})\epsilon^k, \epsilon^k) \leq 1,$$

and if we integrate with respect to t,

$$\int_{-1}^{1}\sum_{k=0}^{\infty}\sum_{i}\epsilon^{k+1}\mathbf{1}_{S_\epsilon}(t+(i+\frac{1}{2})\epsilon^k,\epsilon^k)$$
$$=\sum_{k=0}^{\infty}\sum_{i}\int_{-1}^{1}\epsilon^{k+1}\mathbf{1}_{S_\epsilon}(t+(i+\frac{1}{2})\epsilon^k,\epsilon^k)\le 2.$$

Let $S_{\epsilon,r}=\{x\mid (x,r)\in S_\epsilon\}$. Then for any $x_0\in[0,1]$,

$$\int_{-1}^{1}\mathbf{1}_{S_\epsilon}(t+x_0,\epsilon^k)\,\mathrm{d}t=|S_{\epsilon,\epsilon^k}\cap[x_0-1,x_0+1]|\ge|S_{\epsilon,\epsilon^k}\cap[0,1]|.$$

Then

$$\begin{aligned}2&\ge\sum_{k=0}^{\infty}\sum_{i}\epsilon^{k+1}|S_{\epsilon,\epsilon^k}\cap[0,1]|\\&\ge\sum_{k=0}^{\infty}\epsilon|S_{\epsilon,\epsilon^k}\cap[0,1]|.\end{aligned}$$

And the same thing holds under rescaling: for any $r>0$,

$$\sum_{k=0}^{\infty}\epsilon|S_{\epsilon,r\epsilon^k}\cap[0,r]|\le 2\epsilon^{-1}r$$

Therefore,

$$\begin{aligned}\int_0^1|S_{\epsilon,r}\cap[0,1]|\frac{\mathrm{d}r}{r}&=\sum_{k=1}^{\infty}\int_{\epsilon^k}^{\epsilon^{k-1}}|S_{\epsilon,r}\cap[0,1]|\frac{\mathrm{d}r}{r}\\&\le\int_1^{\epsilon^{-1}}\sum_{k=0}^{\infty}|S_{\epsilon,r\epsilon^k}\cap[0,r]|\frac{\mathrm{d}r}{r}\\&\le\int_1^{\epsilon^{-1}}2\epsilon^{-1}r\frac{\mathrm{d}r}{r}\\&\le 2\epsilon^{-2}r.\end{aligned}$$

□

Likewise, if $X=\mathbb{R}$, then we can replace ϵ–efficient with ϵ–coarsely differentiable:

Corollary 1.7. *Let $f:\mathbb{R}\to\mathbb{R}$ be 1–Lipschitz, $\epsilon>0$ and let $D_\epsilon\subset[0,1]\times(0,1]$ be the set*

$$D_\epsilon=\{(x,r)\mid f\textit{ is not }\epsilon\textit{–coarsely differentiable on }[x-\frac{r}{2},x+\frac{r}{2}]\}.$$

Then D_ϵ is Carleson.

We will take this a little further later on, but this is enough to start with. (Note, for instance, that this isn't quite enough to imply Rademacher's theorem: this implies that for almost every x, there is an affine λ and a sequence of radii r_k such that $\frac{\|f-\lambda\|_{L_\infty([x-r_k,x+r_k])}}{r_k}\to 0$, but λ need not be unique.)

2. FROM CURVES TO SURFACES

Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be a Lipschitz function. Again, Rademacher's theorem implies that f is differentiable almost everywhere. What about coarse differentiability?

We can generalize the notion of a Carleson set:

Definition 2.1. Let $D \subset \mathbb{R}^k \times \mathbb{R}^+$. Let $D_r = \{x \mid (x,r) \in D\}$. We say that D is a *(C–)Carleson set* if there is a $C > 0$ such that for every ball $B \subset \mathbb{R}$ of radius r,

$$\int_0^L |D_r \cap B| \frac{\mathrm{d}r}{r} \le C|B|. \tag{2}$$

It follows from our work on curves that:

Corollary 2.2. *Let $v \in \mathbb{R}^k$ be a unit vector. Let*

$$D_{\epsilon,v} = \{(x,r) \mid f \text{ is not } \epsilon\text{–coarsely differentiable on the line segment } [x - \frac{r}{2}v, x + \frac{r}{2}v].$$

Then $D_{\epsilon,v}$ is Carleson.

So the density of the slices of $D_{\epsilon,v}$ is going down. Let B_1 be the unit ball. For any $\delta > 0$, there is an $r > 0$ and a ball $B_r(x)$ of radius r with $x \in B_1$ such that

$$\frac{\left|B_r(x) \cap (D_{\epsilon,v})_{4r}\right|}{|B_r(x)|} \le \delta.$$

That is, f is ϵ–coarsely differentiable on the line segment $[u-2rv, u+2rv]$ for all but a δ–fraction of $u \in B_r(x)$.

But, by the Lipschitz condition, if δ is small enough, then f is 2ϵ–coarsely differentiable on every line segment $[x-2rv, x+2rv]$, $x \in B_r(x)$. In fact,

Corollary 2.3. *Let $v_1, v_n \in \mathbb{R}^k$ be unit vectors and let $\epsilon > 0$. Then there is a ball B of radius r such that f is ϵ–coarsely differentiable on every line segment $[p - 2rv_i, p + 2rv_i]$, $p \in B$. In fact, the set of (x,r) such that $B(x,r)$ has this property is Carleson.*

We wanted coarse differentiability on the ball, not just these line segments, but that's easy to get.

Theorem 2.4 (Coarse Rademacher's Theorem)**.** *Let $\epsilon > 0$. The set of (x,r) such that f is not ϵ–coarsely differentiable on $B(x,r)$ is Carleson.*

Let's give two proofs. First, a proof by compactness.

Proof by compactness. [1]

First, an exercise:

[1]Proof by a compactness argument is fun: it avoids a lot of bookkeeping, but loses constants. We started by getting explicit bounds on the scale of the largest interval on which a curve is ϵ–efficient/coarsely-differentiable, but when we use a compactness argument, we throw those bounds out the window: the best we can say is that for any ϵ, there's some $r(\epsilon) > 0$ such that any 1–Lipschitz function is ϵ–coarsely differentiable on some ball of radius at least $r(\epsilon)$.

Exercise 1. Let $v_1,\dots,v_{k+1}\in\mathbb{R}^k$ be unit vectors in general position and let B_1 be the unit ball. There is an $r>0$ such that if $f|_{B_r}$ is linear on each line parallel to one of the v_i, then $f|_{B_1}$ is linear. (In fact, one can take $r=1$.)

To prove the theorem, it suffices to show that there is a $\delta>0$ such that if f is δ–coarsely differentiable on every line segment $[p-2rv_i, p+2rv_i]$, $p\in B_1$, then f is ϵ–coarsely differentiable on B_1.

Suppose not. Then there is a sequence of 1–Lipschitz functions f_i such that $f_i(0)=0$ and f_i is $\frac{1}{i}$–coarsely differentiable on each line segment but not ϵ–coarsely differentiable on B_1. By passing to a subsequence, there is a uniform limit $f=\lim f_i$ which is linear on each line segment but not $\frac{\epsilon}{2}$–coarsely differentiable. By the exercise, f is linear, which is a contradiction. □

Alternatively:

Alternative proof. [2] Let $e_1,\dots,e_k\in\mathbb{R}^k$ be the standard basis. Let $0<\delta<\frac{1}{100}$ and let $B=B(x,r)$ be a ball such that f is δ^2–coarsely differentiable on every line segment $[p-2re_i, p+2re_i]$, $p\in B$. Let $B'=B(x,\delta r)$. For any $p\in B'$, there are coarse partial derivatives $\partial_j(p)$, $j=1,\dots,k$ such that

$$|f(p+te_j)-f(p)-\partial_j(p)t|\le 8\delta^2 r$$

for all $|t|<2r$. Furthermore,

$$|\partial_j(p)-\frac{f(p+2re_j)-f(p-2re_j)}{4r}|\le 2\delta^2.$$

But $\|p-x\|\le\delta r$, so

$$|\partial_j(p)-\partial_j(x)|\le 4\delta^2+\frac{\delta}{2}\le\delta.$$

Let $v=(\partial_1(x),\dots,\partial_k(x)$ and let

$$\lambda(q)=f(x)+\vec{v}\cdot(q-x)$$

This is affine, and one can verify that for any $q\in B'$,

$$|f(q)-\lambda(q)|\lesssim\delta^2 kr.$$

So, there is a $C>0$ such that if f is δ^2–coarsely differentiable on every axis-parallel line segment in $B(x,r)$, then f is $C\delta k$–coarsely differentiable on $B(x,\delta r)$. Since the first condition is satisfied away from a Carleson set, so is the second. □

Exercise 2. There's a remarkable theorem of Kirchheim which generalizes Rademacher's Theorem to maps from $\mathbb{R}^n$ to a metric space:

[2]This proof is a sort of quantitative adaptation of the following argument. In the infinitesimal context, we know that the partial derivative of a Lipschitz function exists in any direction, and the partial derivative function is L_∞. By the Density Theorem, for almost every point, the partial derivatives are almost continuous on sufficiently small balls around that point, and a function with continuous partials is differentiable. The details of the translation are left as an exercise for the reader.

Theorem 2.5 (Kirchheim). *Let $f\colon \mathbb{R}^n \to X$ be a Lipschitz map. For almost every $x \in \mathbb{R}^n$, f is* metrically differentiable *at x. That is, there is a seminorm* $\mathrm{md}(f)_x$ *such that*

$$d(f(y), f(z)) = \mathrm{md}(f)_x(y,z) + o(\max\{|x-y|, |x-z|\}).$$

(1) State and prove a coarse version of this theorem for $n = 1$.
(2) Try using coarse methods to prove a coarse version of this theorem.

3. Application: Embeddings of the Heisenberg group

I originally got interested in differentiation in metric spaces because of geometric group theory. For an introduction to geometric group theory, look elsewhere (Löh, Druţu–Kapovich, Bowditch, etc.), but here's one result.

The *integer Heisenberg group* $\mathbb{H}$ is the group of matrices

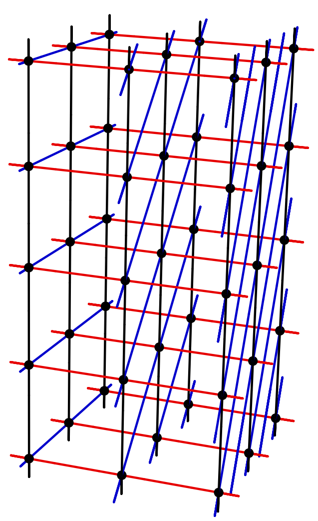
$$\mathbb{H} = \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{Z} \right\}$$

under multiplication. This can be given the presentation $\langle X, Y, Z \mid [X,Y] = Z, [X,Z] = [Y,Z] = 1\rangle$. The *Cayley graph* of a group G (with respect to a generating set S) is the graph Γ_G whose vertices correspond to group elements such that two vertices g_1 and g_2 are connected by an edge if there is an $s \in S$ such that $g_2 = g_1 s^{\pm 1}$. G acts on itself by left multiplication; this action extends to a left action on Γ_G. The restriction of the path metric on Γ_G to the vertex set, written d_w, is called the *word metric* on G (with respect to S), because

$$d_w(e,g) = \min\{k : g = s_1^{\pm 1} \dots s_k^{\pm 1} \text{ where } s_1, \dots, s_k \in S\}.$$

That is, $d_w(e,g)$ is the length of the shortest way to represent g as a product of generators.

The Cayley graph looks like this:



Here, the matrix $\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$ corresponds to the point (x, y, z), so the graph structure reflects the rules of matrix multiplication.

What does the metric look like? The key fact is that $d_w(I, Z^{n^2}) \approx n$, because $X^n Y^n X^{-n} Y^{-n} = Z^{n^2}$. (The phrase in geometric group theory is that the subgroup generated by Z is *quadratically distorted*.) It's not hard to show:

Lemma 3.1. $d(I, (x, y, z)) \approx \max\{|x|, |y|, \sqrt{|z|}\}$.

This is called the *ball-box inequality* because it says that the ball of radius r around the origin is shaped similarly to a box with dimensions $r \times r \times r^2$. In particular, for $r > 1$, $|B_r| \approx r^4$.

Geometric group theory studies the geometry of spaces like this. A natural question about a metric space is: can we draw this in a way that reflects the geometry better? Can we embed this in $\mathbb{R}^n$ by an isometry? By a bilipschitz embedding? By something weaker?

Theorem 3.2 (Pansu,Semmes)**.** *There is no bilipschitz embedding from $\mathbb{H}$ to $\mathbb{R}^n$ for any n.*

Key idea: Let $f : \mathbb{H} \to \mathbb{R}^n$. Consider the cosets of $\langle X \rangle$ (red lines in the figure). These partition $\mathbb{H}$ into a family of horizontal lines. The restriction of f to each of these lines satisfies the coarse differentiation theorem. Likewise, the restriction to cosets of $\langle Y \rangle$ (blue lines) satisfies the coarse differentiation theorem. So there is a large ball B of radius r such that f is coarsely differentiable on 99% of the red and blue intervals of length r with center in B. This implies strong restrictions on f.

Proof. Consider a 1–Lipschitz map $f : \mathbb{H} \to \mathbb{R}^n$. For any $g \in \mathbb{H}$, let $\gamma_g(t) = f(gX^t)$. Then

$$\|\gamma_g(s) - \gamma_g(t)\| = \|f(gX^s) - f(gX^t)\| \le d_w(gX^t, gX^s) = |s - t|,$$

so γ_g is a 1–Lipschitz map from $\mathbb{Z} \to \mathbb{R}^n$. We can extend by linear interpolation to get a Lipschitz map $\gamma_g : \mathbb{R} \to \mathbb{R}$.

By our previous results, for any $\epsilon > 0$, the set of (t, r) such that γ_g is ϵ–coarse differentiable on $[t - r, t + r]$ is Carleson. In fact:

Lemma 3.3. *Let $\epsilon > 0$ and let*

$$S_\epsilon = \{(g, r) \in \mathbb{H} \times \mathbb{Z} : \gamma_g \textit{ is not } \epsilon\textit{–coarse differentiable on } [-r, r]\}$$

Then S_ϵ is Carleson in the sense that there is a $C > 0$ such that for any $h \in \mathbb{H}$ and any $R > 1$,

$$\sum_{r=1}^{R} r^{-1} \left|\{(g, r) \in S_\epsilon : g \in B_R(h)\}\right| \le C|B_R(h)| \approx CR^4.$$

Proof. Let $L_1, \dots, L_k$ be the cosets of $\langle X \rangle$ that intersect $B_R(h)$. By the ball-box inequality, $k \approx R^3$. Let $g_1, \dots, g_k \in B_R(h)$ be representatives of the L_i. By the coarse differentiation theorem, there is a C_0 such that

$$|\{(g,r)\in S_\epsilon : g\in L_i\cap B_R(h)\}| \le |\{(t,r) : \gamma_{g_i} \text{ not } \epsilon\text{–CD on } [t-r,t+r]\}|$$

$$\sum_{r=1}^{R} r^{-1}|\{(g,r)\in S_\epsilon : g\in L_i\cap B_R(h)\}| \le C_0 R$$

$$\sum_{r=1}^{R} r^{-1}|\{(g,r)\in S_\epsilon : g\in B_R(h)\}| \le C_0 k R \lesssim R^4.$$

□

Let $\delta>0$ and let $R=2^m$ where $m>400C\delta^{-1}$. Then there is an $\frac{m}{2}\le s\le m$ such that

$$\sum_{r=2^s}^{2^{s+1}} r^{-1}\left|\{(g,r)\in S_\epsilon : g\in B_R(h)\}\right| \le \frac{C|B_R(h)|}{200C\delta^{-1}} = \frac{\delta|B_R(h)|}{100}.$$

If $2^s\le r\le 2^{s+1}$, then $r\approx 2^s$, so

$$2^{-s}\sum_{r=2^s}^{2^{s+1}} \left|\{(g,r)\in S_\epsilon : g\in B_R(h)\}\right| \le \frac{\delta|B_R(h)|}{100},$$

so there is an $2^s\le r_0\le 2^{s+1}$ such that

$$\left|\{(g,r_0)\in S_\epsilon : g\in B_R(h)\}\right| \le \frac{\delta|B_R(h)|}{100}.$$

By a covering argument, there is an h_0 such that

$$\left|\{(g,r_0)\in S_\epsilon : g\in B_{r_0}(h_0)\}\right| \lesssim \delta|B_{r_0}(h_0)|.$$

If δ is sufficiently small, then f is 2ϵ–coarsely differentiable on every line through $B=B_{r_0}(h_0)$. For $g\in B$, define

$$\partial_X(g) = \text{coarse derivative of } \gamma_g|_{[-r,r]}.$$

Likewise, define $\partial_Y(g)$ in a similar fashion using the cosets of $\langle Y\rangle$.

Translate so that $h_0=\mathbf{0}$. Let $t=\sqrt{\epsilon}r_0$. (We just need $\epsilon r_0 \ll t \ll r_0$.) Then $Z^{t^2}=X^tY^tX^{-t}Y^{-t}$. So

$$\begin{aligned}\frac{f(Z^{t^2})-f(0)}{t} &= \frac{t\partial_X(0)+t\partial_Y(X^t)-t\partial_X(X^tY^t)-t\partial_Y(X^tY^tX^{-t})+O(\epsilon r_0)}{t}\\ &= \partial_X(0)-\partial_X(X^tY^t)+\partial_Y(X^t)-\partial_Y(X^tY^tX^{-t})+O(\sqrt{\epsilon}).\end{aligned}$$

The endpoints of the line segments of length $2r_0$ through 0 and X^tY^t are separated by $\approx t+\sqrt{tr_0}$, so

$$|\partial_X(0)-\partial_X(X^tY^t)| \lesssim \frac{t+\sqrt{tr_0}+\epsilon r_0}{r_0} \lesssim \sqrt[4]{\epsilon}.$$

Likewise,

$$|\partial_Y(X^t)-\partial_X(X^tY^tX^{-t})| \lesssim \sqrt[4]{\epsilon},$$

so

$$\frac{f(Z^{t^2})-f(0)}{t} = O(\sqrt[4]{\epsilon}+\sqrt{\epsilon}) = O(\sqrt[4]{\epsilon}).$$

Therefore, f is not bilipschitz. □

Exercise 3. Why can't we prove the theorem using the coarse differentiation theorem for the restriction of f to vertical line segments?

Note that quantitative differentiation is crucial here: since the γ_g's are constructed to be piecewise linear, they are automatically differentiable almost everywhere. The fact that we use is not that they are linear on small intervals, it's that they are linear on large intervals.

Exercise 4. Prove a similar result for maps $\mathbb{Z}^2 \to \mathbb{H}$.

Pansu originally proved this using a differentiability result for Lipschitz maps between sub-Riemannian manifolds, which we'll sketch briefly. It's natural to consider a scaling limit of the Heisenberg group. For $t \in \mathbb{Z}$, $t > 1$, the map $\delta_t(x, y, z) = (tx, ty, t^2 z)$ is a group automorphism that roughly scales the metric by t, i.e., $d_w(\delta_t(g), \delta_t(h)) \approx t$. In fact, we can define a metric

$$d_{CC}(g,h) = \lim_{t\to\infty} \frac{d_w(\delta_t(g),\delta_t(h))}{t}.$$

This extends naturally to the rational Heisenberg group and by continuity to the real Heisenberg group $\mathbb{H}_\mathbb{R}$, with the property that

$$d_{CC}(\delta_t(g),\delta_t(h)) = t d_{CC}(g,h).$$

In particular, $d_{CC}(0, Z^t) = d_{CC}(0,(0,0,t)) = 4\sqrt{|t|}$ for all t.

Let $f: (\mathbb{H}, d_W) \to \mathbb{R}^n$ be 1–Lipschitz. By the Arzela-Ascoli theorem, there is a sequence $t_1, t_2, \dots$ such that

$$f_\infty(x,y,z) = \lim_{i\to\infty} \frac{f(\delta_{t_i}(x,y,z))}{t_i}$$

converges to a 1–Lipschitz map from $(\mathbb{H}_\mathbb{R}, d_{CC}) \to \mathbb{R}^n$. Pansu showed that any such map is *Pansu differentiable* almost everywhere. That is, for almost every $p \in \mathbb{H}_\mathbb{R}$, there is a differential $\alpha: \mathbb{H}_\mathbb{R} \to \mathbb{R}^n$ such that α is a homomorphism and

$$\lim_{q\to p} \frac{\|f_\infty(q) - f_\infty(p) - \alpha(p^{-1}q)\|}{d_{CC}(p,q)} = 0.$$

Any homomorphism from $\mathbb{H}_\mathbb{R}$ to $\mathbb{R}^n$ must send Z to 0, so f_∞ is not bilipschitz; it follows that f is not bilipschitz either.

This is closely related to the following open question: Suppose G and H are connected, simply-connected nilpotent Lie groups equipped with a left-invariant Riemannian metric and suppose that G and H are bilipschitz equivalent. Are G and H necessarily isomorphic? Pansu's theorem implies that G and H must have the same scaling limit, but many different groups can have the same scaling limit.

4. Rectifiability and the Analyst's Traveling Salesman Problem

Let's switch gears from functions to sets.

A classic problem in computer science is the Traveling Salesman Problem: given a finite set of points (in the plane, in a metric space), what's the shortest path that goes through every point. Jones generalized this to infinite sets; the Analyst's Traveling Salesman Problem asks:

> Let $K \subset \mathbb{R}^n$ be a bounded set of points. Is there a Lipschitz curve $\gamma\colon [0,1] \to \mathbb{R}^n$ such that $S \subset \gamma([0,1])$?

Jones proved a beautiful criterion for answering this question based on how closely K can be approximated by lines. For any axis-aligned cube Q, let $l(Q)$ be its side length. Let $\omega(Q)$ be the smallest radius of a cylinder (or line) that contains $Q \cap K$. Let S_Q be the center line of that cylinder. Define

$$\beta_K(Q) = \frac{\omega(Q)}{l(Q)}.$$

This is scale-invariant and bounded, i.e., $0 \le \beta_K(Q) \le \sqrt{n}$.

A *dyadic interval* in $\mathbb{R}$ is an interval of the form $[a2^{-k}, (a+1)2^{-k}]$ for some $a, k \in \mathbb{Z}$. A *dyadic cube* in $\mathbb{R}^n$ is a cube formed by taking the product of n dyadic intervals of the same length. Let $\mathscr{D}$ be the set of dyadic cubes in $\mathbb{R}^n$ and let $\mathscr{D}_k$ be the set of dyadic cubes of length 2^{-k}; the cubes in $\mathscr{D}_k$ tile $\mathbb{R}^n$.

For any cube Q and any $r > 0$, let rQ be the concentric cube of side length $r\,l(Q)$.

Theorem 4.1 (Jones, Okikiolu)**.** *If $\Gamma \subset \mathbb{R}^n$ is a connected set, then*

$$\sum_{\mathscr{D}} \beta_\Gamma(3Q)^2 l(Q) \lesssim \mathscr{H}^1(\Gamma). \tag{3}$$

Conversely, if $K \subset \mathbb{R}^n$ and $\sum_{\mathscr{D}} \beta_K(3Q)^2 l(Q) < \infty$, then there is a connected set Γ such that $K \subset \Gamma$ and

$$\mathscr{H}^1(\Gamma) \lesssim \operatorname{diam}(K) + \sum_{\mathscr{D}} \beta_K(3Q)^2 l(Q) < \infty. \tag{4}$$

Jones [Jon90] originally proved this theorem in $\mathbb{R}^2$ using complex analysis; his proof of (3) is based on the following lemma:

Lemma 4.2. *There is a $C > 0$ with the following property. Let $S \subset \mathbb{C}$ be a bounded simply-connected open set. Then S can be decomposed (up to measure-zero sets) as a disjoint union $S = \bigcup S_i$ such that each S_i is a Lipschitz domain (a rescaling and translation of a set of the form $\{re^{i\theta} \mid r < f(\theta)\}$ where f is a C–Lipschitz function and $f(\theta) \in [C^{-1}, C]$ for all θ) and*

$$\sum \mathscr{H}^1(\partial S_i) \lesssim \mathscr{H}^1(\partial S).$$

The proof is based on careful estimates of the derivatives of a conformal map $F\colon D^2 \to S$; the pieces of the decomposition are the images of dyadic squares and unions of dyadic squares. Given a bounded connected set Γ of diameter D, Jones adds a circle of radius $2D$ and a line segment to obtain a connected set Γ' such that $\mathbb{C} \setminus \Gamma'$ consists of some simply-connected bounded connected components and a single unbounded component. By applying the lemma to each bounded connected component, Jones obtains a union of Lipschitz domains whose boundaries contain Γ; the bound on Γ follows from a bound on the Lipschitz domains.

Jones's proof of (4) was also originally stated in $\mathbb{R}^2$, but it holds for subsets of $\mathbb{R}^n$ as well (and a variation holds in Hilbert space [Sch07]). Okikiolu [Oki92] generalized (3) to subsets of $\mathbb{R}^n$.

Before we sketch these proofs, it's instructive to try some simple examples:

- K is a line segment: Then $\beta_K = 0$ for any cube; the left side of (3) is 0 and the right side of (4) is $\operatorname{diam}(K)$.
- K is an L–shaped curve of length 1. There are a few cases for $\beta_K(3Q)$. If $3Q$ doesn't contain the corner, then $\beta_K(3Q) = 0$. If $l(Q) \geq 1$, then $\omega(3Q) \leq 1$, so $\beta_K(3Q) \leq l(Q)^{-1}$. If Q contains the corner, then the corner is in the middle of $3Q$, so $\beta_K(3Q) > \epsilon$.

  There are at most C_n cubes in $\mathscr{D}_k$ such that $3Q$ contains the corner; the rest of the cubes can be ignored. If $k \leq 0$, then these each contribute at most $\beta_K(3Q)^2 l(Q) \approx (l(Q)^{-1})^2 l(Q) = 2^k$ to the sum.

  If $k > 0$, then the cubes in $\mathscr{D}_k$ each contribute at most $\beta_K(3Q)^2 l(Q) \leq (3\sqrt{n})^2 l(Q) \lesssim 2^{-k}$. At least one of the cubes contains the corner, so the total contribution satisfies
  $$\epsilon^2 2^{-k} \leq \sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q) \lesssim 2^{-k},$$
  i.e., $\sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q) \approx 2^{-k}$.

  Therefore,
  $$\sum_{Q \in \mathscr{D}} \beta_K(3Q)^2 l(Q) = \sum_{k \leq 0} \sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q) + \sum_{k>0} \sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q) \lesssim \sum_{k \leq 0} 2^k + \sum_{k>0} 2^{-k} \approx 1.$$
  And the contributions are largest when k is close to 0 and fall off exponentially away from zero.
- K is a zig-zag curve of length 1 made up of segments of length $\frac{1}{1024}$.

  Exercise 5. Show that $\sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q) \approx 2^{-|k-10|}$
- K is two zig-zag curves joined into an L shape.

  Exercise 6. Estimate $\sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q)$.
- K is the four-corners Cantor set.

  Exercise 7. Estimate $\sum_{Q \in \mathscr{D}_k} \beta_K(3Q)^2 l(Q)$.

The proof of Theorem 4.1 has two parts. First, we construct Γ by perturbing a connected set at smaller and smaller scales. Second, we prove (3) by showing that a connected set can be approximated by lines and unions of lines at "most" scales and locations.

4.1. **General considerations.** The use of dyadic cubes here is partly notational convenience, and $3Q$ can be replaced by rQ for any $r > 1$. For $S \subset \mathbb{R}^n$, let
$$\beta_K(S) = \omega(S) \operatorname{diam}(S)^{-1},$$
where $\omega(S)$ is the radius of the smallest cylinder containing $S \cap K$. (This disagrees with $\beta_K(Q)$ by a constant, but ignore that.)

We say that S is a *C–quasiball* (or just a *quasiball*) if there is an $x \in S$ such that $B_{C^{-1}\operatorname{diam}(S)} \subset S$. For example, rectangles with bounded aspect ratio, triangles with angles bounded below, cubes in $\mathbb{R}^n$, etc.

Let $Q(x,r)$ be the cube of side length r centered at x. Then

Exercise 8. Let $d \in \mathbb{R}$. Show that if $Q(x,r) \subset Q(y,s)$ and $r \approx s$, then $r^d \beta_K(Q(x,r)) \approx s^d \beta_K(Q(y,s))$. Show that

$$\int_0^\infty \int_{\mathbb{R}^n} r^d \beta_K(Q(x,r))^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r} \approx \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}^n} 2^{-kd} \beta_K(Q(x,2^{-k}))^2 \,\mathrm{d}x.$$

Exercise 9. For $k \in \mathbb{Z}$, let $\mathscr{C}_k$ be a collection of c–quasiballs such that $\operatorname{diam} S \leq c2^{-k}$ for every $S \in \mathscr{C}_k$. Let $\mathscr{C} = \bigcup_k \mathscr{C}_k$.

(1) Show that if for every k, $\mathscr{C}_k$ has multiplicity at most c, then

$$\sum_{S\in\mathscr{C}} \beta_K(S)^2 \operatorname{diam} S \lesssim_c \int_0^\infty \int_{\mathbb{R}^n} r^{-n+1} \beta_K(Q(x,r))^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r}$$

(2) Show that if for every $k \in \mathbb{Z}$ and every $x \in \mathbb{R}^n$, there is an $S \in \mathscr{C}_k$ such that $B_{c^{-1}2^{-k}}(x) \subset S$, then

$$\sum_{S\in\mathscr{C}} \beta_K(S)^2 \operatorname{diam} S \gtrsim_c \int_0^\infty \int_{\mathbb{R}^n} r^{-n+1} \beta_K(Q(x,r))^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r}.$$

(3) Conclude that for any $r, s > 1$,

$$\sum_{Q\in\mathscr{D}} \beta_K(rQ)^2 l(Q) \approx_\epsilon \sum_{Q\in\mathscr{D}} \beta_K(sQ)^2 l(Q).$$

(4) Find a set K such that

$$\sum_{Q\in\mathscr{D}} \beta_K(3Q)^2 l(Q) \not\lesssim \sum_{Q\in\mathscr{D}} \beta_K(Q)^2 l(Q).$$

4.2. **Constructing Γ (after Jones).** Let $K \subset \mathbb{R}^d$ be a set such that

$$\sum_{\mathscr{D}} \beta_K(50Q)^2 l(Q) < \infty.$$

We will construct Γ as a limit of connected graphs that approximate K at smaller and smaller scales.

Let $r_0 > 0$ be a small number to be chosen later. By rescaling, suppose that $\operatorname{diam} K = 1$ and that $\mathbf{0} \in K$. (If K is a single-point set, the theorem is trivial.) For $n \geq 0$, let N_n be a maximal $r_0 2^{-n}$–net and suppose that $N_0 \subset N_1 \subset \dots$. That is, if $z, z' \in N_n$ and $z \neq z'$, then $|z - z'| \geq r_0 2^{-n}$, and if $w \in K$, there is a $z \in N_n$ such that $|w - z| < r_0 2^{-n}$. We will define a sequence of connected graphs (in the sense of graph theory) Γ_n with vertex set $V(\Gamma_n) = N_n$ and

$$\ell(\Gamma_n) \lesssim 1 + \sum_{\mathscr{D}} \beta_K(50Q)^2 l(Q).$$

Any finite graph admits a path that goes over each edge exactly twice, so if we can find such graphs Γ_n, then there is a Lipschitz $f_n \colon [0,1] \to \mathbb{R}^d$ such that $\Gamma_n \subset f([0,1])$ and $\operatorname{Lip}(f_n) \leq 2\ell(\Gamma_n)$. By Arzela–Ascoli, a subsequence of these maps converge to a Lipschitz map $f \colon [0,1] \to \mathbb{R}^d$; we let $\Gamma = f([0,1])$.

Exercise 10. Let $\Gamma \subset \mathbb{R}^n$ be a connected set with $\mathscr{H}^1(\Gamma) < \infty$. Then there is a Lipschitz $f\colon [0,1] \to \mathbb{R}^n$ such that $\Gamma \subset f([0,1])$ and $\mathrm{Lip}(f) \lesssim \mathscr{H}^1(\Gamma)$.

Let $0 < \epsilon_0 < \frac{r_0}{100}$ be a small number to be chosen later. We call a dyadic cube Q *good* if $\beta_K(50Q) < \epsilon_0$ (i.e., $50Q \cap K$ is contained in a cylinder of radius $50\epsilon_0 \ell(Q)$) and *bad* otherwise.

We construct the Γ_i's inductively. Let Γ_0 be the complete graph on N_0. Let $\mathscr{D}_i$ be the set of dyadic cubes of side length 2^{-i}. Suppose that we have constructed Γ_i with the following properties.

(1) If $Q \in \mathscr{D}_i$ is bad, then Γ_i contains the complete graph on $50Q \cap N_i$.
(2) If $Q \in \mathscr{D}_i$ is good, then $50Q \cap K$ is contained in a cylinder C of radius $50\epsilon_0 \ell(Q) = 50\epsilon_0 2^{-n}$ and $30Q \cap N_i$ consists of points in C spaced at least $r_0 2^{-n}$ apart. Since $\epsilon_0 < \frac{r_0}{100}$, we can label these points $z_1, \dots, z_k$ in order. (That is, if we rotate C so that it lies along the x–axis, then the z_i's are in order of increasing x–coordinate.) Then Γ_i contains the path $[z_1, \dots, z_k]$.

We construct Γ_{i+1} from Γ_i as follows.

(1) For each bad cube $Q \in \mathscr{D}_{i+1}$, add the complete graph on $50Q \cap N_{i+1}$ to Γ_{i+1}.
(2) Suppose $Q \in \mathscr{D}_{i+1}$ is a good cube. Let $P(Q) \in \mathscr{D}_i$ be the *parent* of Q, i.e., the unique cube such that $Q \subset P(Q)$. Note that $50Q \subset 30P(Q)$.

    If $P(Q)$ is bad, add the complete graph on $50Q \cap N_{i+1}$ to Γ_{i+1}. Otherwise, label the points in $50Q \cap N_i$ by $z_1, \dots, z_k$ in order. Since $50Q \cap N_i \subset 30P(Q) \cap N_i$, Γ_i contains the path $[z_1, \dots, z_k]$. We orient the cylinder so that z_1 is on the left and z_k is on the right.

    If there are new points $v \in 50Q \cap N_{i+1} \setminus N_i$ that lie between z_1 and z_k, we label them $y_1, \dots, y_m$, in order and replace the path $[z_1, \dots, z_k] \subset \Gamma_i$ by the path $[y_1, \dots, y_m] \subset \Gamma_{i+1}$.

    If one of the endpoints of $[z_1, \dots, z_k]$ lies in $30Q$, then we might need to extend the path to one side or the other. If $z_1 \in 30Q$ and there are new points $v \in 40Q \cap N_{i+1} \setminus N_i$ that are left of z_1, then $d(v, z_1) < r_0 2^{-i}$, so in fact, $v \in 31Q$; otherwise, there would be points in $50Q \cap N_i$ that are left of z_1, which contradicts the minimality of z_1. Denote these points by $y_1, \dots, y_m$, in order (in this case, m must be small, say $m \le 5$) and extend the path $[z_1, \dots, z_k]$ by adding $[y_1, \dots, y_m]$. Do likewise if $z_k \in 30Q$. The result satisfies the desired conditions.

Exercise 11. What happens if we change the choice of Q here?

What happens to the length of Γ_i when we execute this construction? It depends on Q. If Q is bad, then we add a complete graph to Γ_i. Each complete graph has ≈ 1 vertices (a $r_0 2^{-i}$–net can only have boundedly many points in any 2^{-i}–cube) connected by edges of length at most $\approx 2^{-i}$, so each bad cube contributes length $\approx 2^{-i}$. Correspondingly, Q contributes $\beta_K(50Q)^2 l(Q) \ge \epsilon_0^2 2^{-i}$ to the sum, so the contribution is roughly the same as the length increase (up to a multiplicative factor depending on ϵ_0).

If Q is good, things are more complicated. In this case, we perturb a path (replacing $[z_1,\dots,z_k]$ by $[y_1,\dots,y_m]$ and/or extend a path at one of its endpoints (if $v < z_1$ or $v > z_k$). The effect of perturbation is straightforward: when we perturb, the Pythagorean theorem implies that the length increases by a multiplicative factor of $\approx \beta_K(50Q)^2$, so the net increase is $\approx \beta_K(50Q)^2 l(Q)$.

Extension is more complicated and requires a more holistic approach. We consider the special case when $Q_0 \in \mathscr{D}_0$ is a good cube with only good descendants, $\beta_K(50Q') < \epsilon_0$ for all $Q \subset Q_0$; the general case is similar. Then there is a thin cylinder C such that $K \cap 50Q_0 \subset C$.

We define a family of sets G_i of paths in Γ_i, which we call *tracks* as follows. Let $z_1,\dots,z_k \in N_0 \cap 50Q_0$, arranged in order, and let $G_0 = \{\gamma_0 = [z_1,\dots,z_k]\}$. For each track $\gamma = [x_1,\dots,x_k] \in G_i$, we construct a track in G_{i+1} by replacing segments of $[x_1,\dots,x_k]$ by paths $[y_1,\dots,y_m]$ as in the construction of Γ_{i+1}. We extend γ when the extension forms a small angle with the existing curve, i.e., if $Q \in \mathscr{D}_{i+1}$ and $x_1, x_2 \in 50Q$ and $v \in 30Q \cap N_{i+1} \setminus N_i$ is left of x_1, then we extend γ as in the construction (likewise if $x_{k-1}, x_k \in 50Q$ and v is right of x_k).

Otherwise, we create a new track. This happens when $50Q \cap N_i$ contains only a single point x and there is a $v \in 30Q \cap N_{i+1} \setminus N_i$. In this case, we label the points in $\{x\} \cup 30Q \cap N_{i+1}$ by $y_1,\dots,y_m$ in order and add the track $[y_1,\dots,y_m]$ to G_{i+1}. Note that in this case, $d(x,v) < r_0 2^{-i}$, so $x \in 32Q$; otherwise, there would be another point in $50Q \cap N_i$. In this case, we call x the parent vertex of $[y_1,\dots,y_m]$ and we call the track containing x the *parent track* of $[y_1,\dots,y_m]$.

By induction, every point in $Q \cap N_i$ is contained in one of the tracks of G_i. If a track is created in G_i, it is created with diameter $\approx 2^{-i}$ and then extended by at most $\approx 2^{-j}$ at the jth step, so a track created at the ith step never has a diameter more than $\approx 2^{-i}$. We must thus estimate: how many tracks can be created at each stage?

Let $[y_1,\dots,y_m]$ be a track created at the ith step with parent track γ_0 and parent vertex $x_a \in N_i$. Let x_{a-1} and x_{a+1} be the nearest points in N_i to x_a. As noted above, there is a $Q \in \mathscr{D}_{i+1}$ such that $x_a \in 32Q$ but $x_{a-1}, x_{a+1} \not\in 50Q$, so $|x_a - x_{a\pm1}| \geq 9 \cdot 2^{-i}$. That is, there is a gap of width at least $7 \cdot 2^{-i}$ in the projection of $Q \cap \Gamma$ to the axis of C. Since each track with parent track γ_0, corresponds to a gap in the projection, the total diameter of these child tracks is at most, say $\frac{1}{2} \operatorname{diam} \gamma_0$. The same is true for the next generation of grandchild tracks and so on, so the total diameter of all the tracks is $\approx \operatorname{diam} \gamma_0 \lesssim 1$. Combining this with the Pythagorean argument, we get the desired bound.

4.3. **Bounding $\sum \beta^2 l$ (after Okikiolu).** First, some reductions.

Suppose $K \subset \mathbb{R}^n$ is a connected set and $\mathscr{H}^1(K) < \infty$. By Exercise 10, there is a Lipschitz curve $\gamma\colon [0,1] \to \mathbb{R}^n$ such that $K \subset \gamma([0,1])$, so it suffices to consider images of Lipschitz curves (also known as *rectifiable curves*).

Let $W = \{0, \frac{1}{3}, \frac{2}{3}\}^n$. The sets of cubes

$$\mathscr{D}_k + W = \{Q + w \mid Q \in \mathscr{D}_k, w \in W\}$$

are covers of $\mathbb{R}^n$ that satisfy Exercise 9,

$$\sum_{Q\in\mathcal{D}} \beta_K(3Q)^2 l(Q) \approx \sum_{Q\in\mathcal{D}+W} \beta_K(Q)^2 l(Q) = \sum_{w\in W}\sum_{Q\in\mathcal{D}} \beta_{K-w}(Q)^2 l(Q).$$

Combining this with the previous result, it suffices to show that for any 1–Lipschitz $\gamma\colon [0,1]\to\mathbb{R}^n$,

$$\sum_{Q\in\mathcal{D}} \beta^2_{\gamma([0,1])}(Q)\, l(Q) = \sum_{Q\in\mathcal{D}} \frac{\omega(Q)^2}{l(Q)} \lesssim l(\gamma).$$

Recall that $\omega(Q)$ is the radius of the smallest cylinder containing $\Gamma\cap Q$.

Now we describe the bound. Let $Q_0=[0,1]^n$ be the unit cube. By rescaling, suppose that $\Gamma=\gamma([0,1])\subset Q_0$ and $l(\gamma)\approx 1$. We may append line segments so that $\gamma(0)=\gamma(1)=0$. For each cube Q, let $\langle Q\rangle$ be the set of descendants (subcubes) of Q and for each i, let $\langle Q\rangle_i$ be the set of ith–generation descendants of Q (subcubes of side length 2^{-k-i}).

Consider the following sequence of approximations of γ. Let $T_k=\bigcup_{Q\in\langle Q_0\rangle_k}\partial Q$. Let γ_i agree with γ on $\gamma^{-1}(T_i)$. The complement of $\gamma^{-1}(T_i)$ is an open set, i.e., a union of disjoint intervals; extend γ over each of these intervals by linear interpolation.

Then $L_k=l(\gamma_k)\le 1$ for every k and L_k is an increasing sequence. Let $\Gamma=\gamma([0,1])$ and $\Gamma_k=\gamma_k([0,1])$.

For each cube $Q\in\langle Q_0\rangle_k$, let

$$U_Q=\gamma_k(\gamma^{-1}(\operatorname{int}(Q))),$$

where $\operatorname{int}(Q)$ is the interior of Q. This is a union of line segments connecting points in ∂Q. Let

$$s(Q)=\sup_{x\in\operatorname{int}(Q)\cap\Gamma} d(x,U_Q)=\sup_{y\in\gamma^{-1}(\operatorname{int}(Q))} d(\gamma(y),U_Q)$$

or $s(Q)=0$ if $\operatorname{int}(Q)\cap\Gamma=\emptyset$. Let

$$t(Q)=\sup_{y\in\gamma^{-1}(\operatorname{int}(Q))} d(\gamma_{k+1}(y),U_Q).$$

By the Pythagorean Theorem,

$$L_{k+1}-L_k\gtrsim \sum_{Q\in\langle Q_0\rangle_k}\frac{t(Q)^2}{l(Q)};$$

it follows that

$$\sum_{Q\in\langle Q_0\rangle}\frac{t(Q)^2}{l(Q)}\lesssim l(\gamma) \tag{5}$$

We have $t(Q)\le s(Q)\le\omega(Q)$. In order to prove the theorem, we need to replace $t(Q)$ in (5) by $s(Q)$, then by $\omega(Q)$.

Replacing $t(Q)$ by $s(Q)$ is straightforward. For each i, let $\alpha_{Q,i} \in \langle Q\rangle_i$ be the subcube that maximizes $t(\alpha_{Q,i})$. Then $s(Q) \le \sum_{i=0}^{\infty} t(\alpha_{Q,i})$. It follows that

$$\left(\sum_{Q\in\langle Q_0\rangle} \frac{s(Q)^2}{l(Q)}\right)^{\frac{1}{2}} \lesssim \sum_{i=0}^{\infty}\left(\sum_{Q\in\langle Q_0\rangle} \frac{t(\alpha_{Q,i})^2}{l(Q)}\right)^{\frac{1}{2}}$$

by the triangle inequality for Hilbert space. If $Q \neq Q'$, then $\alpha_{Q,i} \neq \alpha_{Q',i}$, so

$$\sum_{Q\in\langle Q_0\rangle} \frac{t(\alpha_{Q,i})^2}{l(Q)} = 2^{-i} \sum_{Q\in\langle Q_0\rangle} \frac{t(\alpha_{Q,i})^2}{\alpha_{Q,i}} \le 2^{-i} \sum_{Q\in\langle Q_0\rangle} \frac{t(Q)^2}{l(Q)} \lesssim 2^{-i} l(\Gamma).$$

Therefore,

$$\left(\sum_{Q\in\langle Q_0\rangle} \frac{s(Q)^2}{l(Q)}\right)^{\frac{1}{2}} \lesssim \sum_{i=0}^{\infty} \sqrt{2^{-i} l(\Gamma)} \lesssim \sqrt{l(\Gamma)},$$

as desired.

Replacing $s(Q)$ by $\omega(Q)$ is harder and we will just sketch the argument. Fix a small $\delta > 0$ and a large $\lambda > 0$. We can ignore cubes where $s(\lambda Q) > \delta\omega(Q)$; those cubes contribute at most $\approx \delta^{-1} l(\Gamma)$ to the sum. We thus consider the cubes where $s(\lambda Q) \le \delta\omega(Q)$: cubes that are much closer to a union of lines than to a single line. Call the set of such cube $\mathscr{A}$.

We will show that $\sum_{Q\in\mathscr{A}} \omega(Q) \lesssim l(\gamma)$. We use a projection argument. Let L be a line in $\mathbb{R}^n$, let π_L be orthogonal projection to L. By the coarea formula,

$$\int_L |\pi_L^{-1}(x)|\,\mathrm{d}x \le l(\gamma). \tag{6}$$

We claim that there are lines $L_1, \dots, L_k$ such that

$$\sum_{Q\in\mathscr{A}} \omega(Q) \lesssim \sum_i \int_{L_i} |\pi_{L_i}^{-1}(x)|\,\mathrm{d}x \le l(\gamma) \le k l(\gamma).$$

The key is the following lemma.

Lemma 4.3. *For each $Q \in \mathscr{A}$, there are many lines L such that $\pi_L|_{2Q\cap\Gamma}$ has multiplicity > 1 on an interval of length $\approx l(Q)$.*

If we choose enough lines L_i, then for every cube in $\mathscr{A}$, the lemma holds for some L_i. So, if there are a lot of cubes in $\mathscr{A}$, the projection π_L should have high multiplicity, but the multiplicity is bounded by (6).

The difficulty is bookkeeping. Each cube in $\mathscr{A}$ corresponds to an interval on which $\pi_L|_{2Q}$ has multiplicity > 1, but we have to avoid overcounting these intervals — this is nontrivial. For example, suppose Γ contains a subset P consisting of two parallel line segments of length 1 and distance ρ. For any k such that $2^k \ge \rho$, there are cubes Q containing segments of both lines, so the lines could be counted roughly $-\log_2 \rho$ times. Nevertheless, there are exponentially fewer large cubes that intersect P, so

$$\sum_Q \frac{\omega_P(Q)^2}{l(Q)} \lesssim \sum_Q \omega_P(Q) \lesssim l(P).$$

Okikiolu solves this by a careful weighting procedure; see [Oki92] for the full argument or [Sch07] for a generalization.

5. Uniform rectifiability and singular integrals

Jones's Traveling Salesman Theorem links the parametrizability of a 1–dimensional subset of $\mathbb{R}^2$ and its approximability by lines. It's natural to ask how to generalize these notions to higher-dimensional spaces and subsets.

David and Semmes [DS91, DS93] developed a remarkable generalization of these ideas called *uniform rectifiability*. Uniform rectifiability can be defined in several equivalent ways, several of which are based on:

(1) Approximability by planes and generalizations of β–numbers
(2) Parametrizability by Lipschitz and bilipschitz maps
(3) The boundedness of singular integral operators

Part of the power of uniform rectifiability is the way that it links approximability, parametrizability, and singular integrals. In the next sections, we will briefly sketch some aspects of the study of singular integrals and the Cauchy transform in $\mathbb{R}^2$ and the links to the Jones Traveling Salesman Theorem, then give some of the equivalent definitions of uniform rectifiability. Proofs of the equivalence can be found in David and Semmes monograph [DS91].

5.1. Rectifiability and the Cauchy transform.

This is mostly outside my area of expertise and primarily motivational. If you're interested in learning more, most of the material in this section is drawn from [Tol14], and full details can be found therein.

A classical question in complex analysis is the Painlevé problem, which asks what compact subsets of $\mathbb{C}$ are *removable* for bounded analytic functions. A compact subset $E \subset \mathbb{C}$ is removable if for every open set Ω containing E, every bounded analytic function $f : \Omega \setminus E \to \mathbb{C}$ extends to an analytic function on E. By the Riemann extension theorem, any point is removable, but the disc $B_r(0)$ is not (the function $\frac{1}{z}$ is analytic and bounded by $\frac{1}{r}$), and any connected set with at least two points is not (it contains a curve segment γ; consider a Riemann map from $\mathbb{C} \setminus \gamma \to D^2$). Thus any removable set must be totally disconnected.

When E is a set with Hausdorff dimension $\dim_H(E) > 1$, Frostman's Lemma can be used to show that E is nonremovable. Conversely, if $\mathscr{H}^1(E) = 0$, then E is removable; in fact, the *analytic capacity* $\gamma(E)$, which measures the maximum of

$$\frac{|f'(\infty)|}{\|f\|_\infty} = \frac{\left|\lim_{z\to\infty} z(f(z) - f(\infty))\right|}{\|f\|_\infty}$$

for bounded analytic functions on $\mathbb{C} \setminus E$ (i.e., how far from constant an analytic function with L_∞–norm 1 can be), satisfies $\gamma(E) \le \mathscr{H}^1(E)$.

Remarkably, the converse inequality does not hold: Vitushkin showed that if E is the four-corners Cantor set obtained by replacing the unit square by four squares of side length $\frac{1}{4}$ at its corners, then each of those by four squares of side length $\frac{1}{16}$, etc., then $\mathscr{H}^1(E) > 0$, but $\gamma(E) = 0$ and thus E is removable.

Denjoy's conjecture, however, states that the converse does hold for rectifiable sets:

Theorem 5.1. *Let $\Gamma \subset \mathbb{C}$ be a rectifiable curve and $E \subset \Gamma$ a compact subset. Then E is removable if and only if $\mathcal{H}^1(E) = 0$.*

The theorem follows from a result of Calderón on the boundedness of the Cauchy transform, which we now sketch.

By the above, it suffices to show that a positive-measure subset of a rectifiable curve is nonremovable; indeed, it suffices to consider curves that are the graphs of Lipschitz functions. We consider the Cauchy transform on such curves.

Given a finite complex measure or a compactly supported complex distribution μ on $\mathbb{C}$, we define the *Cauchy transform* as

$$\mathscr{C}\mu(z) = \int \frac{1}{w-z}\,\mathrm{d}\mu(w). \tag{7}$$

This integral is defined almost everywhere on $\mathbb{C}$ with respect to Lebesgue measure. For a fixed Radon measure μ and an $f \in L_1(\mu)$, let $\mathscr{C}_\mu(f) = \mathscr{C}(f\mu)$. This is closely related to the *Cauchy integral formula*, which can be stated in terms of $\mathscr{C}$. Namely, if f is a holomorphic function on a neighborhood $\Omega \subset \mathbb{C}$ of a region E bounded by a closed rectifiable Jordan curve γ, then

$$f(z) = \int_\gamma \frac{1}{2\pi i}\frac{f(w)}{w-z}\,\mathrm{d}z = \int \frac{1}{w-z}\,\mathrm{d}\mu_\gamma(w)$$

for an appropriate complex measure μ_f supported on γ.

The function $\mathscr{C}\mu$ is analytic outside the support of μ; conversely, if $f \in L_{1,\mathrm{loc}}$ is analytic outside some compact set K and $f(\infty) = 0$, then $f = \mathscr{C}\mu$ for some distribution μ supported in K [Tol14, Thm 1.14]. We can thus think of the Cauchy transform as a way to express functions that are analytic outside the support of μ. Thus, in order to show that a set E is nonremovable, it suffices to find a nontrivial distribution μ supported on E such that $\mathscr{C}\mu$ is bounded.

By results of Nazarov, Treil, and Volberg [Tol14, Thm. 2.16] and Davie and Øksendal [Tol14, Thm. 4.6], this is closely connected to the L_2–boundedness of the Cauchy transform $\mathscr{C}_\mu\colon L_2(E) \to L_2(E)$ for measures μ supported on E. This takes some definition, since (7) may not converge absolutely when $z \in \operatorname{supp}\mu$. Let

$$\mathscr{C}_\epsilon\mu(z) = \int_{|w-z|>\epsilon} \frac{1}{w-z}\,\mathrm{d}\mu(w)$$

and likewise $\mathscr{C}_{\mu,\epsilon}(f) = \mathscr{C}_\epsilon(f\mu)$ be the *truncated Cauchy transform*, so that $\lim_{\epsilon\to 0}\mathscr{C}_\epsilon\mu(z)$ is the *principal value* of the integral (7) if it exists. We say that $\mathscr{C}_\mu$ is $L_2(\mu)$–*bounded* if the $\mathscr{C}_\epsilon$ are uniformly bounded as operators from $L_2(\mu)$ to $L_2(\mu)$. (If so, then the limits $\lim_{\epsilon\to 0}\mathscr{C}_\epsilon f\mu(z)$ exist a.e. w.r.t. μ, but this takes some effort to prove and does not hold for other similar operators [Tol14, Ch. 8].)

Then:

Theorem 5.2 ([Tol14, Rem. 4.8])**.** *If E supports a non-zero Radon measure μ with linear growth (i.e., $\mu(B_r(z_0)) \lesssim r$ for every $z_0 \in \mathbb{C}$, $r > 0$) such that $\mathscr{C}_\mu$ is bounded* *as an operator from* $L_2(\mu)$ *to* $L_2(\mu)$*, then there is a non-zero function* h *with* $0 \le h \le \mathbf{1}_E$ *such that* $\|\mathscr{C}(h\mu)\|_{L_\infty(\mathbb{C}\setminus E)} < \infty$ *and thus* E *is nonremovable.*

So, when is the Cauchy transform bounded? When $\mu = \mathscr{H}^1_{|\mathbb{R}}$ is Hausdorff 1–measure on the real line, we have

$$\mathscr{C}_\mu(f)(t) = \frac{1}{2\pi i}\,\text{p. v.} \int_{-\infty}^{\infty} \frac{f(\tau)}{\tau - t}\,\mathrm{d}\tau,$$

which is, up to a constant, the *Hilbert transform* $H(f)$ of f. This transform is $L_2(\mathbb{R})$–bounded (in fact, it acts on the Fourier transform as $\widehat{H(f)}(\omega) = -i\,\text{sign}(\omega)\hat{f}(\omega)$, where $\text{sign}(\omega) = \frac{\omega}{|\omega|}$). It follows that any subset of $\mathbb{R}$ with positive measure is nonremovable (but there are easier ways to prove this).

Since the Hilbert transform is a convolution with an odd kernel, it has a lot of cancellation; for instance, $H(c) = 0$ for any constant function c. The boundedness of $\mathscr{C}_\mu$ depends, in part on the amount of cancellation involved in (7), which depends in turn on the amount of local symmetry of μ. Smooth curves are locally linear, leading to bounds on $\mathscr{C}_\mu$ by standard techniques. Calderón showed L_2–boundedness for arclength measure on the graph of a Lipschitz function with sufficiently small Lipschitz constant [Cal77], and Coifman, McIntosh and Meyer for an arbitrary Lipschitz graph [CMM82]. Since a rectifiable curve can be covered by rotations of countably many Lipschitz graphs, any positive-measure subset of a rectifiable curve is nonremovable, which implies Denjoy's conjecture.

The original motivation for Jones's formulation of the Traveling Salesman Problem was to study the Cauchy transform; Jones [Jon89] gave a new proof that the Cauchy transform on a Lipschitz graph is bounded by using Theorem 4.1 to show that a Lipschitz graph is close to a line at most points and scales.

Conversely, if $\mu = \mathscr{H}^1_{|E}$ for some set $E \subset \mathbb{C}$, then $L_2(\mu)$–boundedness of the Cauchy transform implies that E is close to a line at most points and most scales. Theorems of Jones and Melnikov and Verdera [MV95] imply:

Theorem 5.3 ([Tol14, Prop. 3.3])**.** *If* $E \subset \mathbb{C}$ *is Ahlfors* 1*–regular (i.e.,* $\mathscr{H}_1(E \cap B_r(x)) \approx r$ *for every* $x \in E$ *and every* $0 < r < \operatorname{diam}(E)$*), then* $\mathscr{C}_{\mathscr{H}^1_{|E}}$ *is* $L_2(E)$*–bounded if and only if*

$$\sum_{Q \in \mathscr{D}} \beta_E(Q)^2\, l(Q) < \infty.$$

That is, $\mathscr{C}_{\mathscr{H}^1_{|E}}$ *is* $L_2(E)$*–bounded if and only if* E *is contained in the image of a Lipschitz curve.*

This is an ingredient in David's solution of the Painlevé problem for sets of finite $\mathscr{H}^1$–measure:

Theorem 5.4 (David)**.** *Let* $E \subset \mathbb{C}$ *be compact and* $\mathscr{H}^1(E) < \infty$*. Then* E *is removable if and only if* E *is purely unrectifiable. (That is, if* γ *is any Lipschitz curve, then* $\mathscr{H}^1(\gamma \cap E) = 0$*.)*

5.2. **Uniform rectifiability.** The necessary and sufficient conditions in the TST don't translate directly to higher dimensions. Among other things, while every connected set with $\mathcal{H}^1(S) < \infty$ can be parametrized by a curve and thus a limit of connected sets with uniformly bounded length has bounded length, the same is not true in higher dimensions (for example, consider the unit disc with n^2 cones of height 1 and base of radius $\frac{1}{n^2}$ attached, like a patch of grass). Instead, they lead to uniform rectifiability. We give three main classes of definitions of uniform rectifiability; all of these are equivalent.

In the following, let $E \subset \mathbb{R}^n$ be an Ahlfors d–regular set. That is, for every $x \in E$ and every $0 < r < \operatorname{diam}(E)$ (including $\operatorname{diam} E = \infty$), $\mathcal{H}^d(E \cap B(x,r)) \approx r^d$. Throughout this section, keep in mind the example of Lipschitz graphs. A *Lipschitz graph* is a translation and rotation of a set of the form $\Gamma_f = \{(x, f(x)) \in \mathbb{R}^n \mid x \in \mathbb{R}^{n-1}\}$, where $f\colon \mathbb{R}^{n-1} \to \mathbb{R}$ is a Lipschitz function, and these will be the prototypical examples of uniformly rectifiable sets.

5.2.1. *β–numbers.* The Jones TST gives a characterization of rectifiable sets using β–numbers. In higher dimensions, a similar bound characterizes uniform rectifiability.

For $1 \le p < \infty$, let

$$\beta_{p,E}(x,r) = \inf_P \left(\frac{1}{r^d} \int_{E \cap B(x,r)} \left(\frac{d(y,P)}{r} \right)^p \mathrm{d}\mathcal{H}^d(y) \right)^{\frac{1}{p}},$$

where the infimum is taken over all n–planes $P \subset \mathbb{R}^{n+1}$. When $p = \infty$, the L_p norm on the distance becomes an L_∞ norm, i.e.,

$$\beta_{\infty,E}(x,r) = \inf_P \sup_{y \in E \cap B(x,r)} \frac{d(y,P)}{r}.$$

The β_K's used in the TST are most similar to $\beta_{\infty,K}$. For all p, this is bounded (because of the Ahlfors regularity) and scale-invariant.

Definition (C3). [3] *An Ahlfors d–regular subset $E \subset \mathbb{R}^n$ is* uniformly rectifiable *if and only if*

$$\beta_1(x,r)^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r}$$

is a Carleson measure. *That is, for any $x \in E$ and any $R > 0$,*

$$\int_0^R \int_{E \cap B(x,r)} \beta_1(x,r)^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r} \lesssim R^d. \tag{8}$$

When $d = 1$ and E is a bounded set, this is similar to the Jones TST, but the bound uses β_1 instead of β_∞, and the Jones TST only bounds the global integral

$$\int_0^\infty \int_{\mathbb{R}^n} \beta_1(x,r)^2 \,\mathrm{d}x \frac{\mathrm{d}r}{r}$$

rather than providing bounds on each of the $B(x,R)$'s.

[3] The numbering of definitions follows [DS91].

The fact that (8) holds for Lipschitz graphs is a theorem of Dorronsoro [Dor85]. For $x \in \mathbb{R}^n$, $r > 0$, define

$$\alpha_{p,f}(x,r) = \inf_{\lambda \in \mathsf{Aff}} \left(\frac{1}{r^n} \int_{B(x,r)} \left(\frac{f(y) - \lambda(y)}{r} \right)^p \mathrm{d}y \right)^{\frac{1}{p}}$$
$$= \inf_{\lambda \in \mathsf{Aff}} \frac{\|f - \lambda\|_{L_p(B(x,r))}}{r^{\frac{n}{p}+1}}$$

and

$$\alpha_{\infty,f}(x,r) = \inf_{\lambda \in \mathsf{Aff}} \frac{\|f - \lambda\|_{L_\infty(Q)}}{r},$$

where Aff is the set of affine functions from $\mathbb{R}^n$ to $\mathbb{R}$. These are analogues of β_K for functions rather than graphs; they measure the distance from $f|_{B(x,r)}$ to an affine function in a scale-invariant way.

Theorem 5.5 ([Dor85])**.** *Let $f \in W^{1,2}(\mathbb{R}^n)$ (i.e., $\|f\|_2 < \infty$ and $\|\nabla f\|_2 < \infty$). Then for any $p \in [1,2]$,*

$$\|\nabla f\|_2^2 \approx \int_{\mathbb{R}^n} \int_0^\infty \alpha_{f,p}(B(x,r))^2 \frac{\mathrm{d}r}{r} \, \mathrm{d}x.$$

A nice proof of this theorem by Fourier analysis can be found in [Azz16].

5.2.2. *Parametrizability.* While connected 1–dimensional sets can be parametrized by Lipschitz curves, uniformly rectifiable sets satisfy a weaker notion of parametrizability.

Definition (C6). *A d–regular set $E \subset \mathbb{R}^n$ is said to have* big pieces of lipschitz images (BPLI) *if there are $\lambda, M > 0$ such that for every $x \in E$, $0 < r < \operatorname{diam} E$, there is an M–Lipschitz map $f \colon B_{\mathbb{R}^d}(0,r) \to \mathbb{R}^n$ such that if $F = f(B_{\mathbb{R}^d}(0,r))$, then*

$$\mathcal{H}^d(F \cap E \cap B(x,r)) \geq \lambda \mathcal{H}^d(E \cap B(x,r)) \approx r^d. \tag{9}$$

We say E is uniformly rectifiable *if it has BPLI.*

In particular, if E has BPLI, then it is rectifiable in the usual sense. That is, there are Lipschitz maps $f_1, f_2, \ldots : \mathbb{R}^d \to \mathbb{R}^n$ such that

$$\mathcal{H}^d(E \setminus \bigcup_i f_i(\mathbb{R}^d)) = 0.$$

(One can construct such maps by taking a cover of E by disjoint balls, then finding Lipschitz images that cover a positive fraction of each ball. By repeating on the uncovered regions of E, one produces Lipschitz images that cover more and more of E.)

As before, it is clear that Lipschitz graphs are uniformly rectifiable; a Lipschitz graph can be covered entirely by a single Lipschitz image.

5.2.3. *Singular integrals.* Recall that when $E \subset \mathbb{C}$ is a 1–dimensional set and $f \in L_2(E)$, the Cauchy transform of f is the principal value of the convolution of f with the *Cauchy kernel* $K_C(w) = \frac{1}{w}$; in real coordinates,

$$K_C(x_1, x_2) = \Big(\frac{x_1}{\|(x_1,x_2)\|^2}, \frac{-x_2}{\|(x_1,x_2)\|^2}\Big).$$

The simplest substitutes for the Cauchy kernel for a d–dimensional set $E \subset \mathbb{R}^n$ are the Riesz kernels, $K_i(x) = \frac{x_i}{\|x\|^{d+1}}$. Like K_C, these are odd kernels (so if E is a plane and f is constant, then $K_i * f = 0$). Further, K_i is degree–d homogeneous, so the corresponding transforms are scale-invariant. It is an open question whether the boundedness of the Riesz transform implies uniform rectifiability in general, though recent work of Nazarov, Tolsa, and Volberg [NTV14] proves that it does when $d = n-1$.

David and Semmes defined the following class of *Cauchy-type kernels.* Let $\mathcal{K}_d(\mathbb{R}^n)$ be the set of smooth real-valued functions K on $\mathbb{R}^n \setminus \{0\}$ K such that K is odd and

$$|x|^{d+j} |\nabla^j K(x)| \in L_\infty(\mathbb{R}^n \setminus \{0\})$$

for $j = 0, 1, 2 \dots$. This includes, in particular, smooth odd functions that are degree–d homogeneous. Then:

Definition (C1). *Suppose E is $\mathcal{H}^d$–measurable. We say that E is* good for the kernel K *if the family of operators*

$$T_\epsilon f(x) = \int_{\substack{y \in E \\ |y-x|>\epsilon}} K(x-y) f(y)\, \mathrm{d}y$$

determines a family of uniformly bounded linear operators from $L_2(E)$ to $L_2(E)$.

We say that E is uniformly rectifiable *if it is good for any kernel $K \in \mathcal{K}_d(\mathbb{R}^n)$.*

Again, Lipschitz graphs are uniformly rectifiable; this is a consequence of Coifman, McIntosh and Meyer's work on Lipschitz graphs in $\mathbb{R}^2$ [CMM82].

5.2.4. *Approximability by Lipschitz graphs.* Finally, we present one more definition, which is more complicated than the others, but is surprisingly useful in many situations. A regular set E is uniformly rectifiable if any only if it admits a *corona decomposition*, which consists of a collection of Lipschitz graphs that approximates E at most points and most scales.

Formalizing this relies on the following proposition; this is due to David [Dav88] for subsets of $\mathbb{R}^n$ and to Christ [Chr90] for metric-measure spaces.

Proposition 5.6. *Let E be a d–regular set in $\mathbb{R}^n$ of diameter $R \le \infty$. Then E admits decompositions into dyadic "cubes" in the following sense. For each $k \in \mathbb{Z}$ such that $2^{-k} \le R$, there is a partition Δ_k of E (i.e., a collection of sets such that $\bigcup_{Q \in \Delta_k} Q = E$ and such that if $Q, Q' \in \Delta_k$, $Q \ne Q'$, then Q is disjoint from Q') and there is a $C > 1$ such that:*

1. *If $R < \infty$ and $k = \lfloor \log_2 R \rfloor$, then $\Delta_k = \{E\}$.*
2. *For all $Q \in \Delta_k$, $C^{-1} 2^k \le \operatorname{diam} Q \le C 2^k$ and $C^{-1} 2^{kd} \le \mathcal{H}^d(Q) \le C 2^{kd}$.*

(3) *For each k, the partition Δ_{k+1} is a refinement of Δ_k. That is, if $Q \in \Delta_j$, $Q' \in \Delta_k$, $k \geq j$, then either $Q' \subset Q$ or $Q \cap Q' = \emptyset$.*

(4) *For any $k \in \mathbb{Z}$, $Q \in \Delta_k$, and any $r > 0$, the boundary of Q is small in the following sense. Let*

$$\partial Q(r) = \{x \in Q \mid d(x, E \setminus Q) \leq r\} \cup \{x \in E \setminus Q \mid d(x, Q) \leq r\}.$$

For any $0 < t < 1$,

$$\mathcal{H}^d(\partial Q(t2^k)) \leq C t^{1/C} 2^{kd} \tag{10}$$

for each $Q \in \Delta_k$.

The last condition is a little subtle; it implies, for instance, that if ϵ is sufficiently small, depending on C, then

$$\mathcal{H}^d(\partial Q(\epsilon 2^k)) \leq \frac{\mathcal{H}^d(Q)}{2}$$

for any Q.

When $\operatorname{diam} E < \infty$, one can arrange the Δ_k's into a rooted tree. Let Δ be the tree with one vertex at height k for each element of Δ_k. This has a single vertex at height $\lfloor \log_2 R \rfloor$, which we mark as the root. For each vertex v $in \mathcal{V}(\Delta)$ at height k, there is a corresponding cube $Q_v \in \Delta_k$, and we let $h(v) = k$ and $\sigma(v) = 2^k$. We connect vertices v and w if $h(v) = h(w) + 1$ and $Q_w \subset Q_v$, in which case we say that v is the *parent* of w. One can check that each vertex (except the root) has a single parent and that the tree has bounded degree. For each $v \in \mathcal{V}(\Delta)$, let $\mathcal{C}(v)$ be the *children* of v (the adjacent vertices of height $h(v)+1$) and we write $v \leq w$ if v is a descendant of w. We call Δ and the collection of associated cubes a *cubical patchwork* for E.

Given a cubical patchwork Δ, a *coronization* is a partition of $\mathcal{V}(\Delta)$ into singletons and *coherent subsets* (a subset $S \subset \mathcal{V}(\Delta)$ is coherent if S is the vertex set of a connected subtree of Δ and such that for every $v \in S$, either $\mathcal{C}(v) \subset S$ or $\mathcal{C}(v) \cap S = \emptyset$). Let $\mathcal{B}$ be the set of singletons, which we call *bad cubes* and let $\mathcal{F}$ be the set of coherent subsets, which we call *stopping-time regions.* We require the following properties:

(1) $\mathcal{B}$ satisfies a Carleson packing condition. That is, for any $v \in \Delta$,
$$\sum_{\substack{w \in \mathcal{B} \\ w \leq v}} \mathcal{H}^d(Q_w) \lesssim \mathcal{H}^d(Q_v).$$

(2) For each $S \in \mathcal{F}$, let $M(S)$ be the maximal vertex (root) of S. Then $M(\mathcal{F})$ satisfies a Carleson packing condition.

Roughly, this says that for almost every $x \in E$, the number of bad cubes and stopping-time regions containing x is finite and bounds the average number of bad cubes and stopping-time regions containing x.

Definition (C4). *For any cube Q and any $r > 0$, let*

$$rQ = \{x \in E \mid d(x, Q) \leq (r-1) \operatorname{diam}(Q)\}$$

(analogously to our notation $3Q$ for concentric cubes above).

A d–regular set E admits a corona decomposition *if for every $\eta > 0$ and $\theta > 0$, there is a coronization $(\mathscr{B}, \mathscr{F})$ such that for every $S \in F$, there is a d–dimensional Lipschitz graph Γ_S with Lipschitz constant η such that $d(x, \Gamma_S) \leq \theta \operatorname{diam}(Q_v)$ whenever $x \in 2Q_v$ and $v \in S$.*

E is uniformly rectifiable if and only if it admits a corona decomposition.

Note that it is nontrivial to show that a Lipschitz graph admits a corona decomposition, since one must approximate a λ–Lipschitz graph by a family of η–Lipschitz graphs and η may be much less than λ.

If a set has a corona decomposition, it can be constructed by a stopping-time argument. First, we construct the set of bad cubes; for some large $r > 0$ and small $\epsilon > 0$, we say that Q is bad if $\beta_{1,E}(rQ) > \epsilon$, where

$$\beta_{1,E}(rQ) = \inf_P \frac{1}{\operatorname{diam}(rQ)^d} \int_{E \cap rQ} \frac{d(y,P)}{\operatorname{diam}(rQ)} \, \mathrm{d}\mathscr{H}^d(y).$$

The rest of the cubes are good. (Note that because of Ahlfors regularity, if $\beta_{1,E}$ is small, then so is $\beta_{p,E}$ for any $1 \leq p \leq \infty$.)

Consider what happens when every cube is good. Then every cube can be approximated by a plane – in fact, a large neighborhood in E around every cube can be approximated by a plane. If ϵ is sufficiently, small, one would hope to find a surface interpolating between these planes, but this surface need not be a Lipschitz graph, because the slope of the planes may increase at smaller and smaller scales. We thus partition the set of good cubes into coherent subsets such that in each coherent subset, the approximating planes form angles of roughly η with each other. These are the stopping-time regions; it remains to show that they can be approximated by Lipschitz graphs with small Lipschitz constant and to prove the Carleson packing condition. (See Chapters 7–14 of [DS91], or see Chapter 4 of [DS93] for a proof that Lipschitz graphs admit a corona decomposition)

Exercise 12. Construct a corona decomposition for:

- a square
- the graph of $f(x) = \sin(x)$ (where $\eta < 1$)
- the graph of $f(x) = x\sin(x^2)$ (where $\eta < 1$)

5.3. **Equivalence of definitions.** Remarkably, all of the definitions of uniform rectifiability given above are equivalent. Indeed, one of the reasons that uniform rectifiability is so powerful is that there are so many different definitions; in addition to the four formulated above, [DS91] gives three more,[4] and many more have been described since David and Semmes's original monographs.

Some of the implications:

(C6) ⇒ (C1): By Coifman–McIntosh–Meyer, a Lipschitz graph is good for any Cauchy-type kernel. David showed that if a class $\mathscr{S}$ of sets is "good for K", then so is the class of sets with big pieces of $\mathscr{S}$ – that is, sets E that

[4](C2) is a Carleson condition on a measure related to kernels of Cauchy type, (C5) is the very big pieces of bilipschitz images condition, which strengthens the BPLI condition, and (C7) describes parametrizations by maps satisfy a condition weaker than Lipschitz.

satisfy (9) for some $F \in \mathscr{S}$. In particular, a set with BPLI is good for any Cauchy-type kernel.

(C1) ⇒ (C3): I have not verified this, but I suspect that given (C1), one can prove (C3) by considering the Riesz kernels. I may be entirely wrong about this, because the proof in [DS91] goes through some other definitions.

(C3) ⇒ (C4): See above.

(C4) ⇒ (C6): When $S \subset \mathcal{V}(\Delta)$ is a stopping-time region containing infinitely many vertices, the approximating Lipschitz graph covers part of $Q(S)$. By patching together such maps, one can cover S by Lipschitz images.

6. Application: Surfaces in $\mathbb{R}^n$

Two of my recent works [You18, NY18] have used uniform rectifiability to study arbitrary surfaces in $\mathbb{R}^n$ and the Heisenberg group. I'd like to briefly sketch some of the ideas to demonstrate how UR can be applied to geometry.

The first obstacle in applying UR to the geometry of surfaces is that many surfaces are not uniformly rectifiable. We have seen this already with the four-corners Cantor set K, but this set is not the best example because K is not really a surface – it isn't rectifiable.

To see surfaces that are rectifiable but not uniformly rectifiable, it's better to look at UR from the quantitative viewpoint. Namely, for any choice of constants in any of the definitions of uniform rectifiability, there is a d–regular rectifiable subset of $\mathbb{R}^n$ that fails to be uniformly rectifiable with those constants; for instance, while the ith step K_i in the construction of the four-corners set has BPLI for some λ_i and M_i, as $i \to \infty$, either $\lambda_i \to 0$ or $M_i \to \infty$.

An advantage of this perspective is that we can use uniform rectifiability to study surfaces with bounded curvature and cellular surfaces (surfaces that can be written as a union of cells in the unit grid). While every such surface which is bounded is uniformly rectifiable for some constants, those constants depend on the size of the surface – large surfaces may have multiscale structure (like the four-corners set) or may fail to be Ahlfors regular. Regardless, one can prove decompositions along the lines of Lemma 4.2:

Theorem 6.1 ([You18])**.** *If $A \in C_d(\tau;\mathbb{Z}_2)$ is a d-cycle in the unit grid in $\mathbb{R}^N$, then there are cycles $M_1,\dots,M_k \in C_d(\tau;\mathbb{Z}_2)$ and uniformly rectifiable sets $E_1,\dots,E_k \subset \mathbb{R}^N$ with bounded uniform rectifiability constants such that $A = \sum_i M_i$, $\operatorname{supp} M_i \subset E_i$, and $\sum_i \mathscr{H}^d(E_i) \lesssim \operatorname{mass} A$.*

The main ingredient in this result is a theorem of David and Semmes on quasiminimizing sets. A quasiminimizing set, or quasiminimizer, is a set whose volume cannot be reduced too much by a small deformation. Specifically,

Definition 6.2. Let $0 < d < N$ be an integer. If $\phi\colon \mathbb{R}^N \to \mathbb{R}^N$ is a Lipschitz map such that $\phi(x) = x$ for all x outside some compact set, let $W = \{x \in \mathbb{R}^n \mid \phi(x) \neq x\}$. We say that ϕ is a deformation of $\mathbb{R}^N$ supported on the set $\operatorname{supp}\phi = W \cup \phi(W)$.

If $k \geq 1$ and $0 < r \leq \infty$ and $S \subset \mathbb{R}^N$ is a nonempty closed set with Hausdorff dimension d, we say that S is a (k,r)-*quasiminimizer* if:

- $\mathcal{H}^d(S \cap B) < \infty$ for every ball $B \subset \mathbb{R}^N$, and
- if ϕ is a deformation supported on a set of diameter $\leq r$ and W is as above, we have

$$|\mathcal{H}^d(S \cap W) \leq k\mathcal{H}^s(\phi(S \cap W)).$$

Theorem 6.3 ([DS00, Thm. 2.11])**.** *Let $S \subset \mathbb{R}^N$ be a (k,r)-quasiminimizer. For each $x \in S^*$ and each $0 < R < r$, there is a uniformly rectifiable, Ahlfors regular set E of dimension d such that*

$$S^* \cap B(x,R) \subset E \subset S^* \cap B(x,2R).$$

The uniform rectifiability constants of E can be taken to depend only on N and k.

Theorem 6.1 is used to prove the main results of [You18], including the following lifting property for cycles mod n:

Theorem 6.4. *For every $n,d,N > 0$, there is a $c > 0$ such that if τ is the unit grid in $\mathbb{R}^N$ and $A \in C_d(\tau;\mathbb{Z}_n)$ is a mod-n cellular cycle in τ, then there is a cellular cycle $R \in C_d(\tau;\mathbb{Z})$ such that $A \equiv R \pmod{n}$ and* $\operatorname{mass} R \leq c \operatorname{mass} A$.

A theorem similar to Theorem 6.1 for codimension–1 cycles in the Heisenberg group is used in [NY18] to prove quantitative bounds on the embeddability of the Heisenberg group into L_1 (a descendant of Theorem 3.2).

References

[Azz16] Jonas Azzam, *Bi-Lipschitz parts of quasisymmetric mappings*, Rev. Mat. Iberoam. **32** (2016), no. 2, 589–648. MR 3512428

[Cal77] A.-P. Calderón, *Cauchy integrals on Lipschitz curves and related operators*, Proc. Nat. Acad. Sci. U.S.A. **74** (1977), no. 4, 1324–1327. MR 466568

[Chr90] M. Christ, *A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral*, Colloq. Math. **60/61** (1990), no. 2, 601–628. MR 1096400 (92k:42020)

[CMM82] R. R. Coifman, A. McIntosh, and Y. Meyer, *L'intégrale de Cauchy définit un opérateur borné sur L^2 pour les courbes lipschitziennes*, Ann. of Math. (2) **116** (1982), no. 2, 361–387. MR 672839

[Dav88] G. David, *Morceaux de graphes lipschitziens et intégrales singulières sur une surface*, Rev. Mat. Iberoamericana **4** (1988), no. 1, 73–114. MR 1009120 (90h:42026)

[Dor85] José R. Dorronsoro, *A characterization of potential spaces*, Proc. Amer. Math. Soc. **95** (1985), no. 1, 21–31. MR 796440

[DS91] G. David and S. Semmes, *Singular integrals and rectifiable sets in $\mathbf{R}^n$: Beyond Lipschitz graphs*, Astérisque (1991), no. 193, 152. MR 1113517

[DS93] Guy David and Stephen Semmes, *Analysis of and on uniformly rectifiable sets*, Mathematical Surveys and Monographs, vol. 38, American Mathematical Society, Providence, RI, 1993. MR 1251061

[DS00] G. David and S. Semmes, *Uniform rectifiability and quasiminimizing sets of arbitrary codimension*, Mem. Amer. Math. Soc. **144** (2000), no. 687, viii+132. MR 1683164 (2000i:49062)

[Jon89] Peter W. Jones, *Square functions, Cauchy integrals, analytic capacity, and harmonic measure*, Harmonic analysis and partial differential equations (El Escorial, 1987), Lecture Notes in Math., vol. 1384, Springer, Berlin, 1989, pp. 24–68. MR 1013815

[Jon90] ———, *Rectifiable sets and the traveling salesman problem*, Invent. Math. **102** (1990), no. 1, 1–15. MR 1069238

[MV95] Mark S. Melnikov and Joan Verdera, *A geometric proof of the L^2 boundedness of the Cauchy integral on Lipschitz graphs*, Internat. Math. Res. Notices (1995), no. 7, 325–331. MR 1350687

[NTV14] Fedor Nazarov, Xavier Tolsa, and Alexander Volberg, *On the uniform rectifiability of AD-regular measures with bounded Riesz transform operator: the case of codimension 1*, Acta Math. **213** (2014), no. 2, 237–321. MR 3286036

[NY18] Assaf Naor and Robert Young, *Vertical perimeter versus horizontal perimeter*, Ann. of Math. (2) **188** (2018), no. 1, 171–279. MR 3815462

[Oki92] Kate Okikiolu, *Characterization of subsets of rectifiable curves in* $\mathbf{R}^n$, J. London Math. Soc. (2) **46** (1992), no. 2, 336–348. MR 1182488

[Sch07] Raanan Schul, *Subsets of rectifiable curves in Hilbert space—the analyst's TSP*, J. Anal. Math. **103** (2007), 331–375. MR 2373273

[Tol14] Xavier Tolsa, *Analytic capacity, the Cauchy transform, and non-homogeneous Calderón-Zygmund theory*, Progress in Mathematics, vol. 307, Birkhäuser/Springer, Cham, 2014. MR 3154530

[You18] Robert Young, *Quantitative nonorientability of embedded cycles*, Duke Math. J. **167** (2018), no. 1.