Taylor’s formula with remainder.

Suppose

\[ f : [0, 1] \to \mathbb{R}, \]

(1) \quad \left| f' \right| \leq 1.

If \( \left| f'' \right| \leq C \) then on \([a, b] \subset [0, 1]\), then by Taylor’s formula,

(2) \quad \left| f(x) - (f(a) - f'(a)(x - a)) \right| \leq \frac{C}{2} \cdot (b - a)^2, \quad \text{(on } [a, b]) .

So \( f \) is well approximated by its tangent line at \( a \in [0, 1] \).

The shorter \([a, b]\) is, the better the linear approximation.
What can we say about *linear approximation of $f$ on short intervals* if $|f'| \leq 1$, but we *drop* the bound on $|f''|$?

The most basic quantitative differentiation theorem says that a 1-Lipschitz function $f : [0, 1] \to \mathbb{R}$ is as close as we like to being linear, at “most” locations and scales.

The principles underlying the proof have led to recent advances in the study of singularities of nonlinear elliptic and parabolic pde, which might possibly have been made much earlier if the elementary case had been widely known.
Let $|J|$ denote the length of the interval $J$.

**Definition.** $f$ is $\epsilon$-**linear** on $J$ if there exists a linear function $\ell$ such that

$$|J|^{-1} \cdot \sup_{x \in J} |f(x) - \ell(x)| \leq \epsilon.$$ 

This definition is *invariant under rescaling* of the domain and target, and the derivative of the rescaled function has absolute value $\leq 1$.

**Note.** We don’t require that $\ell(x)$ has any relation to the derivative of $f$ at any point of $J$. 

For every \( n \), we can divide the interval \( J \) into \( 2^n \) intervals, \( J_{n,j} \) of length \( 2^{-n} \cdot |J| \), whose interiors are disjoint.

These are called \textit{dyadic subintervals}.

We view each \( J_{n,j} \) as defining a \textit{location} and a \textit{scale}.

Given \( \epsilon > 0 \), let \( B_\epsilon(f) \) denote the collection of ”bad” locations and scales, those on which \( f \) fails to \( \epsilon \)-linear.
There is a *natural atomic measure* $\nu$ on the countable collection, $\{J_{n,j}\}$, namely,

$$\nu(J_{n,j}) := |J_{n,j}| = 2^{-n} \cdot |J|.$$  

For fixed $n$, we have $\sum_{n=1}^{2^n} |J_{n,j}| = |J|$, so

$$\text{Mass}(\nu) = \sum_{n=0}^{\infty} \sum_{j=1}^{2^n} |J_{n,j}| = \infty.$$ 

**Note.** So, $\text{Mass}(\nu)$ just means the sum of all the lengths.
**Theorem 1.** Let $f : [0, 1] \to \mathbb{R}$ satisfy $|f'| \leq 1$.

Then for all $\epsilon > 0$

$$(4) \quad \text{Mass}(\mathcal{B}_\epsilon(f)) \leq 4 \cdot (|\log_2 \epsilon| + 1) \cdot \epsilon^{-2}.$$  

Versions of this theorem in more sophisticated contexts were proved in the 1980’s and 90’s by Dorronsoro, Jones and David-Semmes, but were not at all widely known.

The $|\log_2 \epsilon|$ factor is not present in the above versions. It can easily be removed by modifying the proof below.
Basic example, not the worst possible (Semmes).

Define $f_{2^k}(x)$ with $|f'_{2^k}(x)| \leq 1$ by

$$f_{2^k}(x) = 2^{-k} \sin 2^k x.$$ 

The interesting case is when $k$ is large and so, $2^{-k}$ is small.

i) On a subinterval $J$ of length $|J|$, with $2^{-k} \ll |J|$, an excellent choice for a linear approximation of $f$ is:

$$\ell \equiv 0.$$ 

ii) If $|J| \sim 2^{-k}$ then $f \mid J$, does not look linear.

iii) If $|J| \ll 2^{-k}$, then $f$ is well approximated by its first order Taylor expansion.
Monotonicity of Riemann sums for $\int_J h^2$.

**Proof.** (of Theorem 1.) First, consider any $h$, later $h = f'$. Let $h_J$ denote the average of $h$ on $J$ and put

$$V_n(h, J) := \sum_{j=1}^{2^n} (h_{J_{n,j}})^2 \cdot |J_{n,j}|.$$ \hfill (5)

$$D_N(h, J) := V_N(h, J) - V_0(h, J)$$ \hfill (6)

By inspection, for all $n, N$,

$$V_{n+N}(h, J) = \sum_{j=1}^{2^n} V_N(h, J_{n,j}).$$ \hfill (7)

$$V_{n+N}(h, J) - V_n(h, J) = \sum_{j=1}^{2^n} D_N(h, J_{n,j}).$$ \hfill (8)
Monotonicity and boundedness of $V_n(h, J)$.

\[
D_N(h, J) = -(h_J)^2 \cdot |J| + \sum_{j=1}^{2^N} (h_{J_{N,j}})^2 \cdot |J_{N,j}|
\]

(9)

\[
= \sum_{j=1}^{2^N} ((h_{J_{N,j}})^2 - (h_J)^2) \cdot |J_{N,j}|
\]

\[
= \sum_{j=1}^{2^N} ((h_{J_{N,j}})^2 - 2h_{J_{N,j}}h_J + (h_J)^2) \cdot |J_{N,j}|
\]

\[
= \sum_{j=1}^{2^N} (h_{J_{N,j}} - h_J)^2 \cdot |J_{N,j}| \downarrow 0.
\]

(10) \quad 0 \leq V_n(h, J) \nearrow \int_J h^2 \ dx.
Telescoping series and Markov’s inequality.

Using (8), (10), and taking into account cancellations,

\[
\sum_{n,j} D_N(h, J_{n,j}) = \sum_{n=0}^{\infty} V_{n+N}(h, J) - V_n(h, J)
\]

\[
= N \cdot \int_J h^2 \, dx - (V_0(h, J) + \cdots + V_{N-1}(h, J))
\]

\[
\leq N \cdot \int_J h^2 \, dx.
\]

By (9), *monotonicity*, the terms on the l.h.s. are *non-negative*, so for any \( \eta > 0 \), by Markov’s inequality:

**Theorem. 2** (Quantitative Lebesgue Differentiation)

(11) \[
\sum_{J_{n,j}} |J_{n,j}| < \eta^{-1} N \cdot \int_J h^2 \, dx.
\]

\( J_{n,j} : |D_N(h, J_{n,j})| : |J_{n,j}|^{-1} > \eta \)
Coercivity

If we take $h = f'$, with $|f'| \leq 1$ and $\eta = \varepsilon^2/4$, then together with (11), the following proposition completes the proof.

**Proposition.** (Coercivity) Let $|f'| \leq 1$. If

$$2^{-N} \leq \varepsilon,$$

then

$$D_N(f', J) \cdot |J|^{-1} \leq \frac{\varepsilon^2}{4}.$$  \hfill (12)

Then

$$\alpha(f', J) \leq \varepsilon.$$  \hfill (13)
Proof of coercivity.

It suffices to assume $J = [0, 1]$.

Define an affine linear function $\ell$ by

$$\ell(x) = f(0) + x(f(1) - f(0)).$$

Since $|f'| \leq 1$ and the set $x_j = j \cdot 2^{-N}$ is $\frac{\epsilon}{2}$-dense, it suffices to show

$$|f(x_j) - \ell(x_j)| \leq \frac{\epsilon}{2} \quad \text{(for all } i).$$
Conclusion of the proof.

We have $\ell(0) = f(0)$ and by integration,

$$f(x_j) = f(0) + \int_0^{x_j} f' \, dx = \sum_{j=1}^i f'_{J_{N,j}} \cdot |J_{N,j}|.$$

$$\ell(x_j) = \int_0^{x_j} f_J \, dx = f(0) + \sum_{j=1}^i f'_J \cdot |J_{N,j}|.$$

which imply

$$|f(x_i) - \ell(x_i)| \leq \sum_{j=1}^{2^{-N}} |f'_{J_{N,j}} - f'_J| \cdot |J_{N,j}|,$$

Relation (9) and the Schwarz inequality complete the proof.
Conditions implying quantitative differentiation.

In general to prove a quantitative differentiation theorem it suffices to have an energy (or energy density) which is:

1) Monotone with respect to scale. (as in (9)).
2) A priori bounded. (as in (10)).
3) Coercive. (as in (12) implies (13)).

New applications developed with Naber to partial regularity theory for geometric pde use several additional new ideas:

The quantitative stratification.

The quantitative cone-splitting principle.

Novel $\epsilon$-regularity theorems.
Part II; Generalized differentiation.

Let \((X_1, d_1)\) and \((x_2, d_2)\) be metric spaces.

**Definition** \(f : X_1 \rightarrow X_2\) is Lipschitz if there exists \(L\) such that for all \(x_1, y_1 \in X_1\),
\[
d_2(f(x_1), f(y_1)) \leq L \cdot d_1(x_1, y_1).
\]

**Theorem.** (Rademacher, \(\sim 1927\)) A Lipschitz function, \(f : \mathbb{R}^n \rightarrow \mathbb{R}\) is differentiable almost everywhere with respect to Lebesgue measure.

*Boundedness* of the difference quotients, \(\Delta y/\Delta x\), implies *almost everywhere existence of the limit* as \(\Delta x \rightarrow 0\).
What structure on the domain is needed?

The case \( n = 1 \) is due to Lebesgue.

Rademacher’s proof proceeded via reduction to 1-dimensional case, by considering families of parallel lines in \( \mathbb{R}^n \) with rational slopes and using Fubini’s theorem.

Below we will see an approach which does not require that the domain has such special structure.

This leads to an extension of Rademacher’s theorem to the case of domains in a large class of (typically fractal) metric measure spaces, now known as PI spaces.
Extensions to more general domains.

To say that a function $f$ is differentiable at $x$ says that for some unique linear function $Df_x$, 

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

Thus, $f$ converges to the linear function $Df_x$ under suitable blowup and rescaling of the domain and target.

“Lipschitz function” and “almost everywhere” make sense for a space $X$ equipped with a metric $d^X$ and a measure $\mu$.

But what might we mean by “unique” and “linear” if at the infinitesimal level, the geometry of $X$ is not Euclidean?
Linear functions on $\mathbb{R}^n$; intrinsic characterization.

**Theorem.** $\ell : \mathbb{R}^n \to \mathbb{R}$ is linear if and only if it is harmonic and the pointwise norm of the gradient is constant.

**Proof.**

\[
0 = \frac{1}{2} \Delta |\nabla \ell|^2 = \frac{1}{2} \sum_i \sum_j (\ell_j^2)_{ii} \\
= \sum_i \sum_j \ell_{ij}^2 + \ell_j \ell_{iij} \\
= \sum_j \sum_i \ell_{ij}^2 + \ell_j \ell_{iij} \\
= \sum_j \sum_i \ell_{ij}^2.
\]

Since the Hessian of $\ell$ vanishes identically, $\ell$ is linear.
Dirichlet’s principle.

For $U \subset \mathbb{R}^n$ open, $h : U \rightarrow \mathbb{R}$ satisfies,

(15) $\Delta h := \text{div}(\nabla h) = 0$,

if and only if for every metric ball $B_r(p) \subset U$ and every Lipschitz function $\phi$ such that

$\phi|\partial B_r(p) = h|\partial B_r(p)$,

we have

$$\int_{B_r(p)} |\nabla h(x)|^2 \leq \int_{B_r(p)} |\nabla \phi(x)|^2.$$

Since $\Delta h = \text{div}(\nabla h)$, if $|\nabla h(x)|$ is a constant function, then (15) is equivalent to $h$ being harmonic.
Generalized linear functions.

Define the pointwise Lipschitz constant $\text{Lip} f(x)$ by

$$\text{Lip} f(x) = \limsup_{r \to 0} \sup_{d^X(x, x) = r} \frac{|f(x) - f(x)|}{r}.$$ 

For functions $f$ on $\mathbb{R}^n$, we have

$$|\nabla f(x)| = \text{Lip} f(x).$$ 

For general metric measure spaces, $(X, d^X, \mu)$, in (15), we can replace $|\nabla h(x)|^2$ by $(\text{Lip} h(x))^2$ and define “harmonic” via Dirichlet’s principle.

**Definition.** $\ell$ is generalized linear if it is harmonic, $\text{Lip} \ell(x)$ is a constant function and $\text{range}(\ell) = \mathbb{R}$. 
Thus, at least part of the statement of Rademacher’s theorem makes sense for general metric measure spaces.

**Question:** For which metric measure spaces \((X, d^X, \mu)\), is it the case that at \(\mu\)-a.e. \(x \in X\), if we rescale the metric by \(d^X \to r^{-1}d^X\), then \(r^{-1}(f(x) - f(x))\) looks more and more like a generalized linear function?

Next, we describe an interesting class of spaces, PI spaces, for which the above can be shown to hold.

**Remark.** As we will see, the full generalization of Rademacher for PI spaces includes a version of uniqueness.
Definition. \((X, d^X, \mu)\) is a PI space if \(\mu\) is doubling and satisfies a Poincaré inequality.

The simplest PI space is \(\mathbb{R}^n\) with its standard metric and Lebesgue measure.

There are also fractal examples such as:

- Lakkso and inverse limit type spaces.
- The Heisenberg group \((\mathbb{H}, d^\mathbb{H}, \mu)\) group, Carnot groups.
- Certain inverse limits of sequences of finite graphs.
- Spaces quasisymmetrically equivalent to certain Sierpinski carpets.
Doubling measures.

**Definition.** The measure $\mu$ on $(X, d^X)$ is *doubling* if for all $R > 0$ there exists $\beta = \beta(R)$, such that for all $0 < r \leq R$,

$$\mu(B_{2r}(q)) \leq \beta \cdot \mu(B_r(q)).$$

The standard proof of the *Vitali covering theorem* works for doubling measures.

Therefore, consequences such as the *Lebesgue differentiation theorem* hold for doubling measures.
The Poincaré inequality.

Let $f_A$ denote the average value of $f$ on the set $A$.

Let $p \geq 1$.

**Definition.** A $p$-Poincaré inequality holds on $(X, d^X, \mu)$ if there exists $\Lambda > 0$, $\tau = \tau(R)$, such that for all $B_r(q)$ with $r \leq R$,

$$\int_{B_r(q)} |f - f_{B_r(q)}| \, d\mu \leq \tau \cdot r \left( \int_{B_{\Lambda r}(q)} (\text{Lip } f)^p \, d\mu \right)^{1/p}.$$
The almost everywhere asymptotic harmonicity of $f$ follows from a contradiction/replacement argument based on the fact that the Dirichlet energy is lower semicontinuous under $L_p$ convergence $f_j \to f$ i.e.

\begin{equation}
\liminf_j \int_U (\text{Lip } f_j)^p \, d\mu \geq \int_U (\text{Lip } f)^p \, d\mu .
\end{equation}

For $\mathbb{R}^n$, relation (16) is known as Rellich’s theorem.

For general PI spaces, it requires an entirely new proof.
Note. The Lebesgue differentiation theorem implies that asymptotically under blow up (in the $L_1$ sense):

*Any $L^1$ function, is becoming constant $\mu$-a.e.*

In particular, this holds for Lip $f(x)$. 
Suppose that for some Lipschitz function \( f : X \to \mathbb{R} \), asymptotic harmonicity fails on a set \( A \) with \( \mu(A) > 0 \).

Then there exists \( 0 < \eta < 1 \), such that for all \( \epsilon > 0 \), we can cover \( \mu \)-a.e. of \( A \) by disjoint balls, \( \{ B_{r_j}(x_i) \} \), with \( r_i \leq \epsilon \), such that on each \( B_{r_i}(x_i) \), we can replace \( f \) by a function \( \psi_i \), with the same boundary values, whose Dirichlet energy is smaller than that of \( f \), by a factor \( \eta \).

Here, disjointness comes from the Vitali covering theorem.
By truncating each $\psi_i$ if necessary (which doesn’t increase Dirichlet energy) without loss of generality, we can assume

$$\sup_{B_{r_i}(x_i)} \psi_i \leq \sup_{\partial B_{r_i}(x_i)} f,$$

$$\inf_{B_{r_i}(x_i)} \psi_i \geq \inf_{\partial B_{r_i}(x_i)} f.$$  

In this way, we construct a sequence, $f_j \xrightarrow{L_p} f$, which contradicts lower semicontinuity of the Dirichlet energy.

**Remark.** This argument is quite general.
Example: A simple inverse limit space \((X, d^X, \mu)\).

View \(\mathbb{R}\) as a metric graph with edges \(e_i, i \in \mathbb{Z}\), each of which has length 1.

Double each odd numbered edge and then subdivide all edges of length 1 of the resulting graph into thirds.

Double the middle third of each old edge of length 1 and subdivide all edges of length 1/3 into thirds.

Iterate \(n\) times, to get \((X_n, d^{X_n})\) and let \((X, d^X)\) denote the (Gromov-Hausdorff) limit as \(n \to \infty\).
Measures on inverse limits.

Start with Lebesgue measure on \( \mathbb{R} \).

Every time we double an edge, assign half its measure to each of its ”children”.

More generally, fix \( 0 < \alpha < 1 \).

Every time we double and edge \( e_i \), we assign any fractions \( \alpha_i \) and \( (1 - \alpha_i) \) to its children, subject only to \( \alpha \leq \alpha_i \leq 1 - \alpha \).

This gives *uncountably many mutually singular measures* \( \mu \), *for which* \( (X, d^X, \mu) \) *is a PI space*, all of which arise as limits of suitable sequences of measures \( \mu_n \) on \( X_n \).
We say \((X, d^X, \mu)\) admits a differentiable structure if there is a countable collection of measurable sets \(\{U_\alpha\}\), with

\[
\mu(X \setminus \bigcup_{\alpha} U_\alpha) = 0,
\]

and Lipschitz maps, \(\phi_\alpha : U_\alpha \to \mathbb{R}^{k_\alpha}\), such that for every Lipschitz function \(f : X \to \mathbb{R}\), there exist \(\mu\)-a.e. unique bounded measurable functions \(h_\alpha : U_\alpha \to \mathbb{R}^{k_\alpha}\), such that for \(\mu\)-a.e \(x\), one has the “first order Taylor expansion”

\[
f(x) = f(x) + h_\alpha(x) \cdot (\phi_\alpha(x) - \phi_\alpha(x)) + o(d^X(x, x)).
\]

When such a structure exists, in natural sense, it is unique.
Generalized Rademacher theorem.

**Theorem.** ([Ch, 1999]) A PI space admits a differentiable structure.

**Remark.** There exist examples of spaces admitting a differentiable structure which can not be written as unions of positive measure subsets of PI spaces.

For PI spaces, Rademacher’s theorem can be extended to all Lipschitz $f : X \to L^p$, for all $1 < p < \infty$.

For the examples in the above remark, this holds only for some $p$ with $2 < p < 3$. 
The inverse limit example. revisited.

For inverse limit type spaces \((X, d^X, \mu)\) as above, there is a natural 1-Lipschitz projection map \(\pi : X \to \mathbb{R}\).

Let \(I : \mathbb{R} \to \mathbb{R}\) denote the identity map viewed as a 1-Lipschitz function and put \(\phi = I \circ \pi\).

Then the differentiable structure is given by a single “chart”

\[(U, \phi) = (X, \phi) .\]

This is not completely trivial to verify.
Using the full Rademacher theorem for PI spaces, we showed that “most” such spaces e.g. inverse limit type spaces do not bi-Lipschitz embed in $\mathbb{R}^N$, for any $N < \infty$.

With Kleiner, we extended the generalized Rademacher theorem and bi-Lipschitz nonembedding theorem to the infinite dimensional targets i.e. to Lipschitz maps

$$f : X^n \to L^p, \quad (1 < p < \infty).$$
Idea of the proof.

If $f$ is an embedding, then so is its blow up at a typical point and in appropriate cases, it can easily be seen that this simpler map can’t be a bi-Lipschitz embedding.

The inverse limit examples do bi-Lipschitz embed in $L^1$; [CK, 2011].
Metric differentiation and the Heisenberg group.

Certain PI spaces including the Heisenberg group with its Carnot-Caratheodory metric group do not even bi-Lipschitz embed in $L^1$.

Lipschitz maps to $L^1$ need not be differentiable.

However, one can apply a novel differentiation theory we developed with Kleiner and made quantitative with Kleiner and Naor.

It implies that in a certain sense, the metric induced by a Lipschitz map to $L^1$ is differentiable.