Mat 1501 lecture notesweek of September 30-Oct 4.

1. weak convergence

We say that a sequence (μ_k) of Radon measures on $\mathbb{R}^n$ *converges weakly* to a limiting measure μ if

$$\int f\ d\mu_k \to \int f d\mu \qquad \text{for all } f \in C_c(\mathbb{R}^n) \qquad \text{as } k \to \infty. \tag{1}$$

When this holds, we write $\mu_k \rightharpoonup \mu$.

There are a number of variants of this notion, listed below. However, the default definition for us will be that in (1), and we will always try to state clearly when we are considering one of the following alternatives.

- One variant arises if consider sequences such that (1) holds, but with $C_c(\mathbb{R}^n)$ replaced by the larger space $f \in C_0(\mathbb{R}^n)$. This is a slightly stronger notion of weak convergence, since we ask that (1) hold for a large class of functions, and is applicable to sequences of measures with uniformly bounded total variation.
- We can also define weak convergence of a sequence of signed measures, by *exactly* the same formula (1) as for Radon measures.
- We can also define weak convergence for vector-valued measures. For example, if $\mu_k = (\mu_k^1, \ldots, \mu_k^m)$ is a sequence of $\mathbb{R}^m$-valued measures, then $\mu_k \rightharpoonup \mu$ is

  $$\int F \cdot d\mu_k \to \int F \cdot d\mu \qquad \text{for all } F \in C_c(\mathbb{R}^n, \mathbb{R}^m)$$

  where

  $$\int F \cdot d\mu = \sum_{i=1}^m \int F^i \cdot d\mu^i, \qquad F = (F^1, \ldots, F^m).$$

- more generally still, recall that if Y is a separable Banach space, then we can think of continuous linear functionals on the space $C_c(\mathbb{R}^n, Y)$ as "Y^*-valued measures on $\mathbb{R}^n$". We say that a sequence of linear functionals $\lambda_k \in C_c(\mathbb{R}^n, Y)^*$ converges weakly to a limit λ if

  $$\lambda_k(F) \to \lambda(F) \text{ for all } F \in C_c(\mathbb{R}^n, Y).$$

  When this holds, we will again write $\lambda_k \rightharpoonup \lambda$.

  There is also a closely related notion in which we impose the stronger condition

  $$\lambda_k(F) \to \lambda(F) \text{ for all } F \in C_0(\mathbb{R}^n, Y).$$

Note that every example mentioned above is a special case of the last one(s). Thus we will sometimes state and prove results in the setting of $C_c(\mathbb{R}^n, Y)^*$, since it includes all possible special cases of interest for us.

1.1. weak compactness.

Theorem 1. *Assume Y is a separable Banach space, and that that λ_k is a sequence of linear functionals on $C_c(\mathbb{R}^n; Y)$ such that, for every compact $K \subset \mathbb{R}^n$, there exists some C_K such that*

$$\lambda_k(F) \le C_K \sup_{x\in K} \|F(x)\|_Y \qquad \textit{for all } F \in C_c(\mathbb{R}^n, Y) \textit{ with support in } K. \tag{2}$$

Then there exists a subsequence k_ℓ and some $\lambda \in C_c(\mathbb{R}^n, Y)^*$ such that

$$\lambda_{k_\ell}(F) \to \lambda(F) \qquad \text{as } \ell \to \infty, \text{ for every } F \in C_c(\mathbb{R}^n, Y). \tag{3}$$

PROOF. **Step 1**. First we claim that $C_c(\mathbb{R}^n; Y)$ is separable. To see this, fix dense subsets $D_1 \subset Y$ and $D_2 \subset C_c(\mathbb{R}^n)$, and define

$$\mathcal{D} := \{ \text{ finite sums of the form } \sum f_i(x) y_i, \text{ with } y_i \in D_1 \text{ and } f_i \in D_2 \}.$$

Then $\mathcal{D}$ is countable, and it is not hard to check that it is dense in $C_c(\mathbb{R}^n; Y)$.

Exercise 1. Prove that $\mathcal{D}$ is dense in $C_c(\mathbb{R}^n, Y)$.

(A very similar issue arose in our proof of the representation theorem for elements of $C_c(\mathbb{R}^n, Y)^*$. There I believe that I just asserted without proof, and without asking you to check it, the related fact that was needed.)

Step 2. We may assume that the countable dense subset $\mathcal{D} \subset C_c(\mathbb{R}^n, Y)$ found above is a vector space over Q, since if it is not, then we can replace it by a set, still countable, with this property.

Let us write the elements of $\mathcal{D}$ as $F_1, F_2, \ldots$.

The sequence $\lambda_k(F_1), k = 1, 2, \ldots$ is a bounded sequence of real numbers, as a result of (2), so there is a subsequence of the positive integers, which we can write $k^1_\ell, \ell = 1, 2, \ldots$, such that

$$\lambda_{k^1_\ell}(F_1) \to \text{ a limit, say } L_1 \qquad \text{as } \ell \to \infty.$$

Repeating this argument and repeatedly passing to subsequences (so that, say, $(k^{m+1}_\ell)_{\ell=1}^\infty$ is a a subsequence of $(k^m_\ell)_{\ell=1}^\infty$, then finally taking a diagonal subsequence, we can find a subsequence $(k_\ell)_{\ell=1}^\infty$ such that

$$\lambda_{k_\ell}(F_i) \to \text{ a limit } L_i \qquad \text{as } \ell \to \infty \qquad \text{for all } i.$$

We can thus define a function $L : \mathcal{D} \to \mathbb{R}$ by $L(F_i) = L_i$ for every i. It follows from the linearity of the λ_{k_ℓ} and the assumption (2) that L is linear and that

$$L(F_i) \le C_K \sup_{x \in K} \|F_i(x)\|_Y \qquad \text{if } \operatorname{supp}(F_i) \subset K. \tag{4}$$

Step 3. Thus, L defines a continuous map from a dense subset of $C_c(\mathbb{R}^n, Y)$ to $\mathbb{R}$, and so has a unique extension to a continuous map $\lambda : C_c(\mathbb{R}^n, Y) \to \mathbb{R}$. For every $F_i \in \mathcal{D}$,

$$\begin{aligned} \limsup |\lambda_{k_\ell}(F) - \lambda(F)| &\le \limsup \big(|\lambda_{k_\ell}(F - F_i)| + |\lambda_{k_\ell}(F_i) - \lambda(F_i)| + |\lambda(F_i - F)|\big) \\ &\le \limsup |\lambda_{k_\ell}(F - F_i)| + |\lambda(F_i - F)|. \end{aligned}$$

Since $\mathcal{D}$ is dense, it follows from this, (2), and (4) that λ satisfies (3). □

The theorem may also be proved by simply citing a version of the Banach-Alaoglu Theorem, which in fact is just a more general form of the same result (without any assumption of separability, and requiring some form of the axiom of choice in its proof.)

1.2. other descriptions of weak convergence of Radon measures. We now return to weak convergence of sequences of *Radon measures.*

Theorem 2. *Assume that* (μ_k) *is a sequence of Radon measures on* $\mathbb{R}^n$*. Then the following are equivalent.*

(1) $\mu_k \rightharpoonup \mu$.
(2) *For every open bounded* $O \subset \mathbb{R}^n$ *and compact* $K \subset \mathbb{R}^n$,
$$\liminf \mu_k(O) \ge \mu(O), \qquad \textit{and} \qquad \limsup \mu_k(K) \le \mu(K).$$
(3) *For every bounded Borel set* $A \subset \mathbb{R}^n$ *such that* $\mu(\partial A) = 0$,
$$\lim \mu_k(A) = \mu(A).$$

Proof.

(1) $\Rightarrow$ **(2)**: Suppose that $O \subset \mathbb{R}^n$ is open. Then
$$\mu(O) \;=\; \sup\{\int f d\mu \;:\; 0 \le f \le 1, f \in C_c(O)\}.$$
Since for any f with support in O and such that $0 \le f \le 1$, ,
$$\int f d\mu \;=\; \lim_k \int f d\mu_k \le \liminf_k \mu_k(O),$$
the first claim follows by taking the supremum over all such f. The second claim is proved similarly, starting from the observation that for K compact,
$$\mu(K) \;=\; \inf\{\int f d\mu \;:\; 0 \le f \le 1, f \in C_c(\mathbb{R}^n), f \ge 1 \text{ on } K\}.$$

(2) $\Rightarrow$ **(3)**: If $\mu(\partial A) = 0$, then (writing A° for the interior of A)
$$\mu(\bar{A}) \;=\; \mu(A^\circ) + \mu(\partial A) = \mu(A^\circ).$$
Also, since $\mu(A^\circ) \le \mu(A) \le \mu(\bar{A})$, (and similarly for μ_n) it follows that
$$\mu(A) = \mu(\bar{A}) \ge \limsup \mu_n(\bar{A}) \ge \liminf \mu_n(A^\circ) \ge \mu(A^\circ) = \mu(A).$$

(3) $\Rightarrow$ **(1)**: Assume (3), and fix $f \in C_c(\mathbb{R}^n)$. We want to prove that
$$\int f d\mu_k \to \int f d\mu. \tag{5}$$
We may assume that $f \ge 0$, since the general case follows easily.

Note that
$$\text{the set} \qquad \{t \ne 0 : \mu\big(f^{-1}\{t\}\big) > 0\} \qquad \text{is at most countable,} \tag{6}$$
since for any countable set C of nonzero real numbers,
$$\mu(\operatorname{supp}(f)) \ge \sum_{t_i \in C} \mu\big(f^{-1}\{t_i\}\big),$$
whereas, if (6) failed, the sum on the right-hand side above could be made arbitrarily large by a suitable choice of C.

Now fix $\varepsilon > 0$ and choose $0 = t_0 < t_1 < \ldots < t_N$ (for some N) such that
$$|t_i - t_{i-1}| < \varepsilon \quad \text{and} \quad \mu\big(f^{-1}\{t_i\}\big) = 0 \text{ for all } i, \qquad t_N > \sup_{\mathbb{R}^n} f. \tag{7}$$

Next, let
$$B_i := \{x \in \mathbb{R}^n : t_i < f(x) \le t_{i+1}\}$$

and define

$$f_\varepsilon := \sum_{i=1}^{N} t_i \mathbf{1}_{B_i}.$$

It follows from the choice of t_i and the continuity of f that $\mu(\partial B_i) = 0$ for every $i \geq 1$, and hence from (2) that

$$\int f_\varepsilon \; d\mu_k = \sum t_i \mu_k(B_i) \to \sum t_i \mu(B_i) = \int f \; d\mu$$

as $k \to \infty$. In addition, the definitions imply that $|f - f_\varepsilon| < \varepsilon \mathbf{1}_K$, where K is the support of f. Thus

$$\left| \int (f - f_\varepsilon) d\mu \right| \leq \varepsilon \mu(K)$$

and

$$\limsup \left| \int (f - f_\varepsilon) d\mu_k \right| \leq \varepsilon \limsup \mu_k(K) \leq \varepsilon \mu(K).$$

Since

$$\int f d\mu_k - \int f \; d\mu \leq \left| \int (f - f_\varepsilon) d\mu_k \right| + \left| \int f_\varepsilon d\mu_k - \int f_\varepsilon \; d\mu \right| + \left| \int (f_\varepsilon - f) d\mu \right|$$

it easily follows that

$$\limsup \left| \int f d\mu_k - \int f \; d\mu \right| \leq 2\varepsilon \mu(K)$$

and since ε is arbitrary, this implies (5). □

1.3. examples of weak convergence.

Example 1. Let (x_k) be a sequence of points in $\mathbb{R}^n$.

Let $\mu_k = \delta_{x_k}$. Thus,

$$\int f \; d\mu_k = f(x_k) \qquad \text{for } f \in C_c(\mathbb{R}^n).$$

Then it is more or less immediate that

$$\text{if } x_k \to \text{ some limit } x, \qquad \text{then } \mu_k \rightharpoonup \mu := \delta_x$$

as $k \to \infty$. Although this is sort of trivial, it illustrates the utility of weak convergence. In particular, if $x_k \to x$ and $x_k \neq x$, then it is *not* the case that $\delta_{x_k} \to \delta_x$ in certain natural stronger topologies. In particular, under these hypotheses, for every k we have

$$\begin{aligned} \|\delta_{x_k} - \delta_x\|_{C_0^*} &= \sup\{ \int f d(\delta_{x_k} - \delta_x) \; : \; f \in C_0(\mathbb{R}^n), \sup_x |f(x)| \leq 1 \} \\ &= \sup\{ f(x_k) - f(x) \; : \; f \in C_0(\mathbb{R}^n), \sup_x |f(x)| \leq 1 \} \\ &= 2 \end{aligned}$$

Example 2. One similarly checks that

$$\text{if } x_k \to \infty, \qquad \text{then } \mu_k \rightharpoonup 0.$$

Example 3. Let $\psi \in C_c^\infty(\mathbb{R}^n)$ be a function with support in the unit ball, and such that

$$\psi \geq 0, \qquad \int_{\mathbb{R}^n} \psi(x)\, dx = \int_{B_1} \psi(x)\, dx = 1$$

For $\varepsilon > 0$, define

$$\psi_\varepsilon(x) := \frac{1}{\varepsilon}\psi(\frac{x}{\varepsilon}), \qquad\qquad \mu_\varepsilon := \mathcal{L}^n \llcorner \psi_\varepsilon.$$

Then $\mu_\varepsilon \rightharpoonup \mu_0 := \delta_0$ as $\varepsilon \to 0$, or equivalently,

$$\int f \; d\mu_\varepsilon = \int f(x)\psi_\varepsilon(x)dx \to f(0) = \int f(x)d\mu_0(x) \qquad \text{as } \varepsilon \to 0.$$

Indeed, a change of variables shows that

$$\int_{B_\varepsilon(0)} \psi_\varepsilon(x) \; dx \;=\; \int_{B_1(0)} \psi(x) \; dx \;=\; 1.$$

Thus

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)\psi_\varepsilon(x)dx - f(0) &= \int_{B_\varepsilon(0)} \big(f(x) - f(0)\big)\psi_\varepsilon(x)dx \\ &\leq \sup_{|x|<\varepsilon} |f(x) - f(0)| \int_{B_\varepsilon(0)} |\psi_\varepsilon(x)| \; dx \\ &= \sup_{|x|<\varepsilon} |f(x) - f(0)| \to 0 \qquad \text{as } \varepsilon \to 0\end{aligned}$$

using the continuity of f.

Exercise 2. let us define μ_ε as above, but consider the limit $\varepsilon \to \infty$ rather than $\varepsilon \to 0$. Find all possible limits of any convergent subsequences of (μ_ε) as $\varepsilon \to \infty$. (If there is only one possible limit, then of course the whole sequence converges.)

Example 4. The last example illustrates that a sequence of measures, each of which is supported on an n-dimensional set, can converge weakly to a limting measure supported on a lower-dimensional set.

Now we give an example to illustrate the opposite phenomenon.

Let μ_k be the measure on $\mathbb{R}$ defined by

$$\mu_k = \frac{1}{k}\sum_{i=1}^{k} \delta_{i/k}.$$

Thus,

$$\int f \; d\mu_k \;=\; \frac{1}{k}\sum_{i=1}^{k} f(\frac{i}{k}).$$

This is just a Riemann sum approximation to $\int_0^1 f(x)dx$, and if f is continuous, it is certainly the case that the Riemann sums converge to the integral. It follows that

$$\mu_k \rightharpoonup \mathcal{L}^1 \llcorner [0,1].$$

Example 5. Suppose that $g : \mathbb{R} \to \mathbb{R}$ is a continuous periodic function with period p, and for $\varepsilon \in (0,1]$, let $g_\varepsilon(x) := g(\frac{x}{\varepsilon})$. Thus g_ε has period εp, so oscillates rapidly as $\varepsilon \to 0$.

Let $\mu_\varepsilon := \mathcal{L}^1 \llcorner g_\varepsilon$, so that

$$\int f d\mu_\varepsilon = \int_{\mathbb{R}} f(x) g(\frac{x}{\varepsilon}) dx.$$

One can check that

$$\mu_\varepsilon \rightharpoonup \langle g \rangle \mathcal{L}^1, \qquad \text{where} \quad \langle g \rangle = \frac{1}{p} \int_0^p g(x) dx \ = \ \text{average of } g \text{ over one period.}$$

Equivalently,

$$\int_{\mathbb{R}} f(x) g(\frac{x}{\varepsilon}) dx \to \ \langle g \rangle \int_{\mathbb{R}} f(x) dx \qquad \text{for } f \in C_c(\mathbb{R}^n). \tag{8}$$

This is a real analysis exercise.

Exercise 3. Prove that (8) holds.

Perhaps the easiest way to *see* that (8) holds is to note that

$$\mu_\varepsilon(I) \rightharpoonup \langle g \rangle \mathcal{L}^1(I) \qquad \text{as } \varepsilon \to 0.$$

if I is any interval, and hence if I is any open set or any closed set. It is easy to persuade yourself that this is true, and this implies the conclusion, by Theorem 2. I do not know if this is the easiest way to go if you want to check all the details, and in fact Theorem 2 as stated does not apply unless $g(x) \geq 0$ for all x.

Suppose that I is an interval and that $\gamma : I \to \mathbb{R}^n$ is a smooth curve. There are two natural ways to associate a measure to γ.

First, we can define a *Radon measure* μ_γ by specifying that

$$\int f \ d\mu_\gamma \ = \ \int_I f(\gamma(s)) |\gamma'(s)| \ ds \qquad \text{for } f \in C_c(\mathbb{R}^n) \tag{9}$$

Second, we can define an $\mathbb{R}^n$-valued measure ν_γ by specifying that

$$\int F \cdot d\nu_\gamma := \int_I F(\gamma(s)) \cdot \gamma'(s) \ ds \qquad \text{for } F \in C_c(\mathbb{R}^n; \mathbb{R}^n) \tag{10}$$

Note that μ_γ depends only on the image of γ, and not on the parametrization, in the sense that

$$\mu_\gamma = \mu_{\gamma \circ \sigma}$$

if $\sigma : I \to I$ is a diffeomorphism (so that γ and $\gamma \circ \sigma$ are different parametrizations of the same curve).

Similarly, but not quite the same, $\nu_\gamma = \nu_{\gamma \circ \sigma}$ whenever $\sigma : I \to I$ is an *orientation-preserving* diffeomorphism (*i.e.*, $\sigma' > 0$ everywhere in I), in which case γ and $\gamma \circ \sigma$ are different parametrizations of the same *oriented* curve.

These two different ways of encoding geometric information in a measure behave differently with respect to weak convergence, as show in the following two examples.

Example 6. Suppose that $n = 2$.

For $\varepsilon \in [0, 1)$, let $\gamma_\varepsilon(s) := (\cos s, \varepsilon \sin s)$, for $s \in I := [0, 2\pi]$, and let us write μ_ε for μ_{γ_ε} and similarly ν_ε for ν_{γ_ε}, defined in (9) and (10) respectively.

Then it follows just by continuity that

$$\int f d\mu_\varepsilon \rightharpoonup \int f d\mu_0$$

and

$$\int f d\nu_\varepsilon \rightharpoonup \int f d\nu_0.$$

So this is not very dramatic. Note however that $\nu_0 = 0$ (that is, $\int F \cdot d\nu_0 = 0$ for every F) whereas

$$\int f d\mu_0 \;=\; 2\int_{-1}^{1} f(s,0)\; ds.$$

So, rather naturally, cancellation can occur in weak limits of vector-valued measures, but we do not see cancellation in weak limits of signed measures.

Example 7. A more interesting example in the same spirit arises from defining $\gamma_\varepsilon : \mathbb{R} \to \mathbb{R}^2$ by

$$\gamma_\varepsilon(s) = (s, \varepsilon \sin \frac{s}{\varepsilon})$$

for $\varepsilon \in (0,1]$. This is a curve that stays within an ε-neighborhood of the x axis but nonetheless oscillates a lot. As above we write μ_ε for μ_{γ_ε} and similarly ν_ε for ν_{γ_ε}.

Note that

$$\gamma'(s) = (1, \cos\frac{s}{\varepsilon}), \qquad |\gamma'(s)| = (1+\cos^2(\frac{s}{\varepsilon}))^{1/2}.$$

Thus one sees that for $f \in C_c(\mathbb{R}^2)$,

$$\begin{aligned}\int f d\mu_\varepsilon \;&=\; \int_{\mathbb{R}} f(\gamma_\varepsilon(s))(1+\cos^2(\frac{s}{\varepsilon}))^{1/2}\; ds \\ &=\; \int_{\mathbb{R}} f(s,0)(1+\cos^2(\frac{s}{\varepsilon}))^{1/2}\; ds \\ &\qquad + \; \int_{\mathbb{R}} \left[f(s,\varepsilon\sin(\frac{s}{\varepsilon})) - f(s,0)\right](1+\cos^2(\frac{s}{\varepsilon}))^{1/2}\; ds.\end{aligned}$$

A function $f \in C_c(\mathbb{R}^n)$ is uniformly continuous, which makes it easy to check that

$$\int_{\mathbb{R}} \left[f(s,\varepsilon\sin(\frac{s}{\varepsilon})) - f(s,0)\right](1+\cos^2(\frac{s}{\varepsilon}))^{1/2}\; ds \to 0 \tag{11}$$

as $\varepsilon \to 0$. And by Example 5 above,

$$\int_{\mathbb{R}} f(s,0)(1+\cos^2(\frac{s}{\varepsilon}))^{1/2}\; ds \to \langle (1+\sin^2)^{1/2}\rangle \int_{\mathbb{R}} f(s,0)\; ds.$$

It follows that

$$d\mu_\varepsilon \rightharpoonup \mu_0, \text{ where } \int f d\mu_0 := \langle (1+\sin^2)^{1/2}\rangle \int_{\mathbb{R}} f(s,0)\; ds \quad \text{for } f \in C_c(\mathbb{R}^n).$$

The factor of $\langle (1+\sin^2)^{1/2}\rangle$ reflects the presence of oscillations in the curves γ_ε. Thus, a memory of these oscillations is recorded by the weak limit.

On the other hand, similar arguments show that for $F = (F^1, F^2) \in C_c(\mathbb{R}^2, \mathbb{R}^2)$,

$$\begin{aligned}\int F \cdot d\nu_\varepsilon \;&=\; \int_{\mathbb{R}} [F^1(\gamma_\varepsilon(s)) + F^2(\gamma_\varepsilon(s))\cos(\frac{s}{\varepsilon})]\; ds \\ &\to \int_{\mathbb{R}} F^1(s,0)\; ds \\ &=: \int F d\nu_0\end{aligned}$$

where $\nu_0 = \nu_{\gamma_0}$ for $\gamma_0(s) = (s, 0) = \lim_{\varepsilon\to 0} \gamma_\varepsilon(s)$. Thus the limit of the sequence of vector-valued measures, in this example at least, does not remember the oscillations in the sequence of curves γ_ε.

2. differerntiability properties of Lipschitz functions

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *Lipschitz continuous*, or just *Lipschitz*, if there exists some constant L such that

$$|f(x) - f(y)| \leq L|x - y|$$

for all $x, y \in \mathbb{R}^n$. Clearly, the same definition can be formulated if $\mathbb{R}^n$ and $\mathbb{R}^m$ are replaced by any two metric spaces.

The *Lipschitz constant* of f is

$$\mathrm{Lip}(f) := \sup_{x,\neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

We will show that every Lipschitz function is differentiable in two distinct senses. We now introduce the first of these notions of differentiability.

2.1. weak derivatives. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *weakly differentiable* if there exists a function $v \in L^1_{loc}(\mathbb{R}^n; \mathbb{R}^n)$ such that

$$\int f(x)\nabla \cdot \varphi(x)\ dx \ = \ -\int v(x) \cdot \varphi(x) dx \qquad \text{for all } \varphi \in C^1_c(\mathbb{R}^n)$$

When this holds, we call v a "weak gradient" of f. Note that if v_1 and v_2 are two weak gradients of f, then

$$\int (v_1 - v_2) \cdot \varphi = 0 \qquad \text{for all } \varphi \in C^1_c(\mathbb{R}^n)$$

from which it follows that $v_1 = v_2$ a.e.. So in fact we can speak without ambiguity (ignoring sets of measure zero) of *the* weak gradient.

We typically write Df to denote the weak gradient of f.

It follows from the divergence theorem that if f is C^1, then f is weakly differentiable, and $Df = \nabla f$ (where the right-hand side denotes the classical gradient of f.)

We rewrite the definition of weak derivative using this notation:

$$\int f(x)\nabla \cdot \varphi(x)\ dx \ = \ -\int Df(x) \cdot \varphi(x) dx \qquad \text{for all } \varphi \in C^1_c(\mathbb{R}^n)$$

Proposition 1. *A Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is weakly differentiable, and the weak derivative Df satisfies*

$$Df \in L^\infty(\mathbb{R}^n; \mathbb{R}^n).$$

Proof. **Step 1**. Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz, and fix some $\varphi \in C^1_c(\mathbb{R}^n; \mathbb{R}^n)$. We will write φ as $(\varphi^1, \ldots, \varphi^n)$.

We will use the notation

$$D^h_i g(x) := \frac{g(x + he_i) - g(x)}{h}, \qquad D^h \cdot \varphi := \sum_{i=1}^n D^h_i \varphi^i.$$

where e_i is the standard unit vector in the ith coordinate direction. By the Dominated Convergence Theorem,

$$\int f \ \nabla \cdot \varphi \ dx \ = \ \lim_{h \to 0} \int f \ D^h \cdot \varphi \ dx.$$

And for every h, by a change of variables,

$$\begin{aligned} \int f \ D^h \cdot \varphi \ dx &= \frac{1}{h} \left(\int f(x)\varphi(x + he_i)dx - \int f(x)\varphi(x)dx \right) \\ &= \frac{1}{h} \left(\int f(x - he_i)\varphi(x)dx - \int f(x)\varphi(x)dx \right) \\ &= - \int \frac{f(x - he_i) - f(x)}{-h} \varphi(x)dx \\ &= - \int \ D^{-h} f \cdot \varphi \ dx \end{aligned}$$

Thus

$$\int f \ \nabla \cdot \varphi \ dx \ = \ \lim_{h \to 0} \int \ D^{-h} f \cdot \varphi \ dx \qquad \text{for all } \varphi \in C^1_c(\mathbb{R}^n, \mathbb{R}^n). \tag{12}$$

Step 2. For every $x \in \mathbb{R}^n$ and $h \in \mathbb{R}$,

$$|D^{-h} f(x)|^2 = \sum_{i=1}^n |D_i^{-h} f(x)|^2 = \frac{1}{h^2} \sum_{i=1}^n |f(x + he_i) - f(x)|^2 \le n \operatorname{Lip}(f)^2.$$

(The left-hand side above denotes the square of the euclidean norm of the vector $(D_1^{-h} f, \dots, D_n^{-h} f)$.) Thus, the family of (vector-valued) functions $(D^{-h} f)_{h \in (0,1]}$ is uniformly bounded with respect to the L^∞ norm. Thus it follows from standard facts about weak compactness, stated and proved in Lemma 1 below, that there is a subsequence $h_k \to 0$ as $k \to \infty$, and a weak limit, which we can denote $Df \in L^\infty(\mathbb{R}^n; \mathbb{R}^n)$, such that

$$\int D^{-h} f \cdot \varphi \to \int Df \cdot \varphi \qquad \text{for all } \varphi \in C^0_c(\mathbb{R}^n, \mathbb{R}^n).$$

In view of (12), it follows that f is weakly differentiable, and its weak derivative belongs to $L^\infty(\mathbb{R}^n, \mathbb{R}^n)$.

□

Lemma 1. *Assume that $(v_h)_{h \in (0,1]}$ is a family of functions such that*

$$\sup_{h \in (0,1]} \|v_h\|_{L^\infty} := M < \infty.$$

Then there exists a sequence $h_k \to 0$ and a function $v \in L^\infty(\mathbb{R}^n; \mathbb{R}^m)$ such that

$$\int v_h \cdot \varphi \to \int v \cdot \varphi \qquad \textit{for all } \varphi \in C^0_c(\mathbb{R}^n, \mathbb{R}^m) \tag{13}$$

In fact, (13) still holds if we replace $C^0_c(\mathbb{R}^n, \mathbb{R}^m)$ by the larger space of functions $L^1(\mathbb{R}^n; \mathbb{R}^m)$.

Proof. For every h, we define a linear functional $\lambda_h : C_c(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}$ by

$$\lambda_h(\varphi) := \int v_h \cdot \varphi \, dx.$$

Then $|v_h(x) \cdot \varphi(x)| \le |v_h(x)| \, |\varphi(x)| \le M|\varphi(x)|$ for a.e. x, so

$$|\lambda_h(\varphi)| \le M \int |\varphi| \, dx \ \le \ M \sup_{x\in K} |\varphi(x)| \mathcal{L}^n(K) \qquad \text{if } \operatorname{supp}(\varphi) \subset K. \tag{14}$$

It therefore follows from Theorem 1 that there exists some $\lambda : C_c(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}$ such that

$$\lambda_h(\varphi) \to \lambda(\varphi) \qquad \text{for all } \varphi \in C^0_c(\mathbb{R}^n, \mathbb{R}^m)$$

Then (14) implies that

$$|\lambda(\varphi)| \le M \int |\varphi| \, dx \qquad \text{for all } \varphi \in C^0_c(\mathbb{R}^n, \mathbb{R}^m)$$

As a result, Corollary 3 in the Week 3 notes implies that there exists some $v \in L^\infty(\mathbb{R}^n;\mathbb{R}^m)$ such that

$$\lambda(\varphi) = \int v \cdot \varphi \, dx \qquad \text{for all } \varphi \in C^0_c(\mathbb{R}^n, \mathbb{R}^m).$$

The conclusion of the lemma follows by combining the above points. □

We will need another fact about weak differentiability. This may be interpreted as showing that weak differentiability is not *too* weak.

Lemma 2. *Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous, weakly differentiable function, and that $Df = 0$ a.e..*

Then f is constant.

Proof. Fix a C^∞ function $\psi : \mathbb{R}^n \to \mathbb{R}$ with the following properties:

$$\operatorname{supp}(\psi) \subset B_1(0), \qquad \psi(x) \ge 0 \text{ for all } x,. \qquad \int_{\mathbb{R}^n} \psi \, dx \ = \int_{B_1(0)} \psi \, dx \ = 1.$$

For $\varepsilon > 0$, we define $\psi_\varepsilon(x) := \frac{1}{\varepsilon^n}\psi(\frac{x}{\varepsilon})$, so that

$$\operatorname{supp}(\psi) \subset B_\varepsilon(0), \qquad \psi(x) \ge 0 \text{ for all } x,. \qquad \int_{\mathbb{R}^n} \psi_\varepsilon \, dx \ = \int_{B_e(0)} \psi_\varepsilon \, dx \ = 1.$$

Finally, for $\varepsilon > 0$ define

$$f_\varepsilon := \psi_\varepsilon * f$$

(here using the standard notation for the convolution integral

$$\psi_\varepsilon * f(x) := \int_{\mathbb{R}^n} \psi_\varepsilon(x-y) f(y) \, dy \ = \ \int_{\mathbb{R}^n} \psi_\varepsilon(y) f(x-y) \, dy. \qquad)$$

We have essentially proved in Example 3 above that $f_\varepsilon \to f$ uniformly as $\varepsilon \to 0$, so it suffices to show that f_ε is constant for every $\varepsilon > 0$.

It follows from the smoothness of ψ_ε and standard properties of convolutions that

$$f_\varepsilon \text{ is smooth, and } \nabla f_\varepsilon = \nabla(\psi_\varepsilon * f) = (\nabla \psi_\varepsilon) * f$$

Hence, to show that f_ε is constant, it suffices to show that $\nabla f_\varepsilon = 0$.

To do this, we use further standard properties of convolutions to compute, for some arbitrary $\varphi \in C^1_c(\mathbb{R}^n, \mathbb{R}^n)$,

$$\begin{aligned} -\int \nabla f_\varepsilon \cdot \varphi \ = \ \int f_\varepsilon \nabla \cdot \varphi \ = \ \int (\psi_\varepsilon * f) \nabla \cdot \varphi \ &= \ \int f(\psi_\varepsilon * \nabla \cdot \varphi) \\ &= \ \int f \, \nabla \cdot (\psi_\varepsilon * \varphi) \\ &= \ \int Df \ \cdot (\psi_\varepsilon * \varphi). \end{aligned}$$

But $Df = 0$ a.e., and thus

$$\int \nabla f_\varepsilon \cdot \varphi \ = \ 0 \qquad \text{for all } \varphi \in C^1_c(\mathbb{R}^n, \mathbb{R}^n).$$

It easily follows that $\nabla f_\varepsilon = 0$, which completes the proof of the lemma. □

Exercise 4. Persuade yourself that the following variants of Lemma 2 are true.

- if $f : \mathbb{R}^n \to \mathbb{R}$ is a locally (Lebesgue) integrable function such that $Df = 0$ a.e.., then f is constant a.e..
- if $O \subset \mathbb{R}^n$ is a connected open set, and $Df = 0$ a.e. in O, then f is constant in O.

Exercise 5. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is a weakly differentiable function and $Df \in L^\infty(\mathbb{R}^n, \mathbb{R}^n)$, then in fact f is Lipschitz continuous.

Hint: It suffices to show that f is the uniform limit of a sequence of functions with uniformly bounded Lipschitz constants.

In combination with the facts we have proved above, this exercise shows that a function $\mathbb{R}^n \to \mathbb{R}$ is Lipschitz if and only if if is weakly differentiable. with weak derivative in L^∞.

2.2. almost everywhere differentiability. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be differentiable at a point x if there exists a vector, denoted $\nabla f(x)$, such that

$$\frac{f(x + hy) - f(x)}{h} - \nabla f(x) \cdot y \to 0$$

as $h \to 0$, *uniformly* for y in the unit ball $B_1 \subset \mathbb{R}^n$.

For $f : \mathbb{R}^n \to \mathbb{R}^m$, the definition of differentiable is exactly the same, except that $\nabla f(x)$ is then a linear map from $\mathbb{R}^n \to \mathbb{R}^m$, which we can write as a $m \times n$ matrix.

Theorem 3 (Rademacher's Theorem)**.** *A Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathcal{L}^n$ almost every $x \in \mathbb{R}^n$.*

Proof. We will prove the theorem for $m = 1$; the general case then follows easily.

If $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, we will prove that f is differentiable, with $\nabla f(x) = Df(x)$, at every point x at which

$$\lim_{r \to 0} ⨍_{B_r(x)} |Df(z) - Df(x)| \, dz \ = \ 0. \tag{15}$$

Since we already know that (15) holds $\mathcal{L}^n$ almost everywhere, this will prove the theorem.

We henceforth fix some x satisfying (15). It is convenient to define

$$g_h(y) := \frac{f(x+hy) - f(x)}{h} - Df(x)\cdot y,$$

so that our goal is now to show that $g_h(y) \to 0$ uniformly for $y \in B_1$ as $h \to 0$.

Step 1. First note that for any $y_1, y_2 \in \mathbb{R}^n$ and any $h > 0$,

$$\begin{aligned} |g_h(y_1) - g_h(y_2)| &\le \frac{|f(x-hy_1) - f(x-hy_2)|}{h} + |Df(x)|\,|y_1 - y_2| \\ &\le \big(\mathrm{Lip}(f) + |Df(x)|\big)\ |y_1 - y_2|. \end{aligned}$$

Also, it is clear that $g_h(0) = 0$ for all h, so

$$\begin{aligned} |g_h(y)| \ = \ |g_h(y) - g_h(0)| &\le \big(\mathrm{Lip}(f) + |Df(x)|\big)\ |y - 0| \\ &\le R\big(\mathrm{Lip}(f) + |Df(x)|\big) \qquad \text{if } |y| \le R. \end{aligned}$$

Thus the family of functions $(g_h)_{h\in(0,1]}$ is uniformly bounded and equicontinuous in any bounded subset of $\mathbb{R}^n$, and in particular in the unit ball B_1.

We can thus appeal to the Arzela-Ascoli Theorem to find that for any sequence h_k tending to zero, there is a further subsequence (which we will still denote h_{k_ℓ} and a limit g such that

$$g_{h_{k_\ell}} \to g \text{ uniformly in } B_1 \subset \mathbb{R}^n. \tag{16}$$

If f were not differentiable at x, one could find a subsequence g_{h_k} such that $\liminf_{k\to\infty} \sup_{x\in B_1} |g_{h_k}(x)| > 0$. We could then pass to a further subsequence for which (16) holds, and it would necessarily be the case that $\sup_{x\in B_1}|g(x)| > 0$. To rule out this possibility, it therefore suffices to show that any limit of a uniformly convergent subsequence must equal zero.

Step 2. To finish the proof, we will now show that any limit g of a uniformly convergent subsequence, which we will write[1] as (g_{h_k}) must equal zero.

Let g be such a limit. Note that $g(0) = \lim_{k\to\infty} g_{h_k}(0) = 0$, so it suffices to show that g is constant in $B_1(0)$. and for this it suffices (in view of Lemma 2 above, or more precisely, a variant of Lemma 2 stated in Exercise 4) to show that g is weakly differentiable, with $Dg = 0$ a.e. in B_1. This will certainly follow if we can show that

$$\int_{B_1} g\nabla\cdot\varphi = 0 \qquad \text{for all } \varphi \in C^1_c(B_1). \tag{17}$$

We will see that this follows from the fact that x is a Lebesgue point of Df (in the strong sense of (15) above). Indeed, for $\varphi \in C^1_c(B_1)$,

(18)

$$\begin{aligned} \int_{B_1} g(y)\nabla\cdot\varphi(y)\ dy &= \lim_{k\to\infty}\int_{B_1} g_{h_k}(y)\nabla\cdot\varphi(y)\ dy \\ &= \lim_{k\to\infty}\int_{B_1} \frac{f(x+h_k y) - f(x) - Df(x)\cdot(h_k y)}{h_k}\nabla\cdot\varphi(y)\ dy. \end{aligned}$$

For every k, we make a change of variables, defining $z = x + h_k y$, so that $y = \frac{z-x}{h_k}$. We define φ_k so that $\varphi(y) = \varphi_k(z)$. Then by the chain rule,

$$\nabla_y\cdot\varphi(y) = h_k\nabla_z\varphi_k(z).$$

[1] rather than $g_{h_{k_\ell}}$

Also, note that if $y \in B_1$ then $z \in B_{h_k}(x)$. So for every k,

$$\int_{B_1} \frac{f(x+h_k y) - f(x) - Df(x)\cdot(h_k y)}{h_k} \nabla \cdot \varphi(y)\ dy$$
$$= \frac{1}{h_k^n} \int_{B_{h_k}(x)} \Big[f(z) - f(x) - Df(x)\cdot(z-x)\Big] \nabla \cdot \varphi_k(z)\ dz$$
$$= \frac{1}{h_k^n} \int_{B_{h_k}(x)} \big[Df(z) - Df(x)\big] \cdot \varphi_k(z)\ dz$$

using the fact that Df is a weak derivative of f.

In general, $\mathcal{L}^n(B_h(x)) = h^n \mathcal{L}^n(B_1)$, so

$$\frac{1}{h^n} \int_{B_h(x)} \cdots\ dx \ = \ \mathcal{L}^n(B_1) ⨍_{B_h(x)} \cdots\ dx.$$

Thus, noting that $\|\varphi_k\|_\infty = \|\varphi\|_\infty$ for every k,

$$\left| \int_{B_1} g_{h_k}(y) \nabla \cdot \varphi(y)\, dy \right| \le \mathcal{L}^n(B_1)\|\varphi\|_\infty ⨍_{B_{h_k}(x)} |Df(z) - Df(x)|\, dx.$$

Then (17) follows from this together with (15) and (18).

□

We also stated the following result.

Theorem 4. *Assume that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a Lipschitz continuous function. Then for any $\varepsilon > 0$, there exists a C^1 function $g : \mathbb{R}^n \to \mathbb{R}^m$ such that the set*

$$G := \{s \in \mathbb{R}^n : \ f(x) = g(x), \ \text{and } \nabla f(x) \text{ exists and equals } \nabla g(x)\}$$

satisfies

$$\mathcal{L}^n(\mathbb{R}^n \setminus G) < \varepsilon$$

In the lecture, we briefly discussed the proof, which relies on Rademacher's Theorem and the Whitney Extension Theorem.