Analysis II

January 11, 2017

Contents

1 Vector spaces

1.1 Vector spaces

If $a_n \in \mathbb{R}$, $(a_n) \to a$ if for every $\epsilon > 0$, $\exists N$ such that $|a_n - a| < \epsilon$ whenever $n > N$.
Now consider a general vector space:

Definition. Let V be a real vector space. A *norm* on V is a function $||\cdot|| : V \to \mathbb{R}$ satisfying:
• $||\mathbf{v}|| \geq 0 \ \forall \mathbf{v} \in V$, and $||\mathbf{v}|| = 0 \iff \mathbf{v} = \mathbf{0}$;
•$||\lambda \mathbf{v}|| = |\lambda| \cdot ||\mathbf{v}||$, $\forall \lambda \in \mathbb{R}$ and $\mathbf{v} \in V$;
•$||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$, $\forall \mathbf{v}, \mathbf{w} \in V$ (triangle inequality).

Example. $||\mathbf{v}||_2 = \left(\sum v_i^2\right)^{\frac{1}{2}}$, the Euclidean norm;
$||\mathbf{v}||_1 = \sum |v_i|$;
$||\mathbf{v}||_\infty = \max\{|v_1|, ..., |v_n|\}$.

Example. Let $V = C[0,1] = \{f : [0,1] \to \mathbb{R} | f \text{ is continuous}\}$. Then we can have the following norms:
• $||f||_1 = \int_0^1 |f(x)| dx$;
• $||f||_2 = \left(\int_0^1 f(x)^2 dx\right)^{\frac{1}{2}}$;
• $||f||_\infty = \max_{x \in [0,1]} |f(x)|$.

Notation. If $|| \cdot ||$ is a norm on V, we say the pair $(V, || \cdot ||)$ is a *normed space.*

Definition. Suppose $(V, || \cdot ||)$ is a normed vector space, and $(\mathbf{v}_n)$ is a sequence in V. We say $(\mathbf{v}_n)$ converges to $\mathbf{v} \in V$ if $\forall \varepsilon > 0$, $\exists N$ such that $\forall n > N$, $||\mathbf{v}_n - \mathbf{v}|| < \varepsilon$.
Equivalently, $(\mathbf{v}_n) \to \mathbf{v}$ if and only if $||\mathbf{v}_n - \mathbf{v}|| \to 0$ in $\mathbb{R}$.

Example. Let $V = \mathbb{R}^n$, $\mathbf{v}_k = (v_{k,1}, ..., v_{k,n})$.
(a) $(\mathbf{v}_k) \to \mathbf{v}$ with respect to $|| \cdot ||_\infty$
$\iff ||\mathbf{v}_k - \mathbf{v}||_\infty \to 0$
$\iff \max\{|v_{k,i} - v_i|\} \to 0$
$\iff |v_{k,i} - v_i| \to 0$ for all $1 \leq i \leq n$
$\iff v_{k,i} \to v_i$.

So sequence converges if and only if every component converges.

(b) $(\mathbf{v}_k) \to \mathbf{v}$ with respect to $|| \cdot ||_1$
$\iff \sum_{i=1}^n |v_{k,i} - v_i| \to 0$
$\iff |v_{k,i} - v_i| \to 0$ for all $1 \leq i \leq n$
$\iff v_{k,i} \to v_i$.

Note the two different norms in (a) and (b) give the same notion of convergence.

We set a convention that, when talking about convergence in $\mathbb{R}^n$ without mentioning a norm, then it's with respect to $|| \cdot ||_1$ (or $|| \cdot ||_\infty$ or $|| \cdot ||_2$) (these all give the same notion of convergence).

Example. Let $V = C[0,1]$,

$$f_n(x) = \begin{cases} 1 - nx & x \in \left[0, \frac{1}{n}\right) \\ 0 & x \in \left[\frac{1}{n}, 1\right] \end{cases}$$

So

$$||f_n||_1 = \int_0^1 |f_n(x)|dx = \frac{1}{2n} \to 0$$

as $n \to \infty$. So $(f_n) \to 0$ with respect to $||\cdot||_1$.

On the other hand, $||f_n||_\infty = 1 \not\to 0$, so $(f_n) \not\to 0$ with respect to $||\cdot||_\infty$. Here the two different norms give two different notions of convergence.

1.2 Continuity

Let $(V, ||\cdot||)$ be a normed vector space.

Recall: If $\mathbf{v}_n \in V$ and $\mathbf{v} \in V$, the sequence $(\mathbf{v}_n) \to \mathbf{v}$ if for every $\varepsilon > 0$, there exists n such that $||\mathbf{v}_n - \mathbf{v}|| < \varepsilon$ when $n > N$.

Definition. Suppose V and W are normed spaces, and $f : V \to W$. We say f is *continuous* if the sequence $(f(\mathbf{v}_n)) \to f(\mathbf{v})$ in W whenever $(\mathbf{v}_n) \to \mathbf{v}$ in V.

Example. (1)$f : V \to \mathbb{R}^n$, $f(\mathbf{v}) = (f_1(\mathbf{v}), ..., f_n(\mathbf{v}))$. Then f is continuous if and only if $f_1, ..., f_n$ are all continuous.
(2) $p_i : \mathbb{R}^n \to \mathbb{R}$ by $p_i(\mathbf{v}) = v_i$. Then p_i is continuous.
(3) $V = C[0,1]$, $x \in [0,1]$, $p_x : C[0,1] \to \mathbb{R}$ by $p_x(f) = f(x)$ (linear map). Then p_x is continuous with respect to the uniform norm on $C[0,1]$:

$$\begin{aligned} & (f_n) \to f \text{ wrt } ||\cdot||_\infty \\ \iff & \max_{y \in [0,1]} |f_n(x) - f(x)| \to 0 \\ \implies & |f_n(x) - f(x)| \to 0 \\ \implies & (f_n(x)) \to f(x) \end{aligned}$$

However, p_x is not continuous with respect to $||\cdot||_1$ on $C[0,1]$. See examples in M&T.

So linear maps may not be continuous.
(4) If $f : V_1 \to V_2$ and $g : V_2 \to V_3$ are continuous, so is $g \circ f : V_1 \to V_3$.
(5) $||\cdot|| : V \to \mathbb{R}$ is continuous.

Lemma. If $\mathbf{v}, \mathbf{w} \in V$, then $||\mathbf{w} - \mathbf{v}|| \geq |||\mathbf{w}|| - ||\mathbf{v}|||$.

Proof. Since $||\mathbf{v}|| + ||\mathbf{w} - \mathbf{v}|| \geq ||\mathbf{w}||$,
$||\mathbf{w} - \mathbf{v}|| \geq ||\mathbf{w}|| - ||\mathbf{v}||$.
Similarly, $||\mathbf{w} - \mathbf{v}|| = ||\mathbf{v} - \mathbf{w}|| \geq ||\mathbf{v}|| - ||\mathbf{w}||$. So $||\mathbf{w} - \mathbf{v}|| \geq |||\mathbf{w}|| - ||\mathbf{v}|||$. $\square$

Now we can prove the 5^{th} example above:

Proof. Let $f(\mathbf{v}) = ||\mathbf{v}||$. Then if $(\mathbf{v}_n) \to \mathbf{v}$, $(||\mathbf{v}_n - \mathbf{v}||) \to 0$. But $||\mathbf{v}_n - \mathbf{v}|| \geq |||\mathbf{v}_n|| - ||\mathbf{v}||| = |f(\mathbf{v}_n) - f(\mathbf{v})| \geq 0$.
So by squeeze rule, $(|f(\mathbf{v}_n) - f(\mathbf{v})|) \to 0$, i.e. $f(\mathbf{v}_n) \to f(\mathbf{v})$. □

Proposition. $f : V \to W$ is continuous if and only if for every $\mathbf{v} \in V$ and $\varepsilon > 0$, there exists $\delta > 0$ such that

$$||f(\mathbf{w}) - f(\mathbf{v})||_W < \varepsilon$$

whenever $||\mathbf{w} - \mathbf{v}||_V < \delta$.

Proof. Suppose the $\varepsilon - \delta$ condition hold. We'll show that f is continuous, i.e. if $(\mathbf{v}_n) \to \mathbf{v}$, then $(f(\mathbf{v}_n)) \to f(\mathbf{v})$.
Given $(\mathbf{v}_n) \to \mathbf{v}$ and $\varepsilon > 0$, pick $\delta > 0$ such that $||f(\mathbf{w}) - f(\mathbf{v})|| < \varepsilon$ whenever $||\mathbf{w} - \mathbf{v}|| < \delta$. Since $(\mathbf{v}_n) \to \mathbf{v}$, there exists N such that $||\mathbf{v}_n - \mathbf{v}|| < \delta$ whenever $n > N$, i.e. $||f(\mathbf{v}_n) - f(\mathbf{v})|| < \varepsilon$ when $n > N$. So $(f(\mathbf{v}_n)) \to f(\mathbf{v})$. So f is continuous.
If the $\varepsilon - \delta$ condition does not hold, then there exists $\mathbf{v} \in V$ and $\varepsilon > 0$ such that for every $n > 0$, there exists $\mathbf{v}_n$ with

$$||\mathbf{v} - \mathbf{v}_n|| < \frac{1}{n}$$

but

$$||f(\mathbf{v}) - f(\mathbf{v}_n)|| > \varepsilon$$

(Otherwise, take $\delta = \frac{1}{n}$ and we get a contradiction). Then $(\mathbf{v}_n) \to \mathbf{v}$, but $(f(\mathbf{v}_n)) \not\to f(\mathbf{v})$. So f is not continuous. □

1.2.1 Addendum

Suppose V, W are normed spaces and U_α is an open subset of V for all $\alpha \in A$. Let $U = \cup_{\alpha \in A} U_\alpha$.

Proposition. Suppose $f : U \to W$ and f is continuous on all U_α. Then f is continuous on U. It's important that U_α's are all open. For example, any $f : V \to W$ is continuous on $\{\mathbf{v}\}$, but may not be continuous on $\cup_{\mathbf{v} \in V} \{\mathbf{v}\} = V$.

Proof. Must show that given $\mathbf{v} \in U$ and $\varepsilon > 0$, $\exists \delta > 0$ s.t.

$$f(B_\delta(\mathbf{v}) \cap U) \subset B_\varepsilon(f(\mathbf{v}))$$

$\mathbf{v} \in \cup_{\alpha \in A} U_\alpha$, so $\mathbf{v} \in U_{\alpha_0}$ for some $\alpha_0 \in A$. f is continuous on U_{α_0}, so $\exists \delta_1 > 0$ s.t.

$$f(B_{\delta_1}(\mathbf{v}) \cap U_{\alpha_0}) \subset B_\varepsilon(f(\mathbf{v}))$$

U_{α_0} is open, so $\exists \delta_2 > 0$ s.t. $B_{\delta_2}(\mathbf{v}) \subset U_{\alpha_0}$.
Let $\delta = \min(\delta_1, \delta_2)$. Then $B_\delta(\mathbf{v}) \subset B_{\delta_1}(\mathbf{v})$ and $B_\delta(\mathbf{v}) \subset B_{\delta_2}(\mathbf{v}) \subset U_{\alpha_0}$.

So $B_\delta(\mathbf{v}) \subset B_{\delta_1}(\mathbf{v}) \cap U_{\alpha_0}$.
Thus

$$f(B_\delta(\mathbf{v}) \cap U) = f(B_\delta(\mathbf{v})) \subset f(B_{\delta_1}(\mathbf{v}) \cap U_{\alpha_0}) \subset B_\varepsilon(f(\mathbf{v}))$$

□

1.3 Open and Closed Subsets

Definition. If $\mathbf{v} \in V$ and $r > 0$,

$$B_r(\mathbf{v}) = \{\mathbf{w} \in V | ||\mathbf{v} - \mathbf{w}|| < r\}$$

is the *open ball* of radius r centered at $\mathbf{v}$,

$$B_r(\mathbf{v}) = \{\mathbf{w} \in V | ||\mathbf{v} - \mathbf{w}|| \leq r\}$$

is the *closed ball* of radius r centered at $\mathbf{v}$.

Now we can get an alternative definition of continuous:
• f is continuous if and only if for every $\mathbf{v} \in V$ and $\varepsilon > 0$, there exists $\delta > 0$ such that $f(B_\delta(\mathbf{v})) \subset B_\varepsilon(f(\mathbf{v}))$.

Definition. $U \subset V$ is an *open subset* of V if for every $\mathbf{u} \in U$, there exists $\varepsilon > 0$ such that $B_\varepsilon(\mathbf{u}) \subset U$.

Proposition. If $f : V \to W$ is continuous and $U \subset W$ is open, then $f^{-1}(U)$ is open in V.

Proof. Suppose $\mathbf{v} \in f^{-1}(U)$, i.e. $f(\mathbf{v}) \in U$.
U is open, so there exists $\varepsilon > 0$ such that $B_\varepsilon(f(\mathbf{v})) \subset U$.
f is continuous, so $\exists \delta > 0$ such that $f(B_\delta(\mathbf{v})) \subset B_\varepsilon(f(\mathbf{v})) \subset U$, i.e. $B_\delta(\mathbf{v}) \subset f^{-1}(U)$ so $f^{-1}(U)$ is open.
The converse is also true(see M&T). □

Definition. (Open subsets) Recall $U \subset V$ is *open* in V if for every $\mathbf{u} \in U$, $\exists \varepsilon > 0$ s.t. $B_\varepsilon(\mathbf{u}) \subset U$.

Proposition. If $f : V \to W$ is continuous and $U \subset W$ is open, then $f^{-1}(U)$ is open in V.

Example. Given $\mathbf{v} \in V$, define

$$f_{\mathbf{v}} : V \to \mathbb{R}$$
$$f_{\mathbf{v}}(\mathbf{w}) = ||\mathbf{v} - \mathbf{w}||$$

Then $f_{\mathbf{v}}$ is continuous, so

$$B_r(\mathbf{v}) = f_{\mathbf{v}}^{-1}((-r, r))$$

is open in V, i.e. open balls are open.

Definition. (Closed subsets) Recall if $C \subset V$, $V - C = \{\mathbf{v} \in V | \mathbf{v} \notin C\}$ is the *complement* of C. $C \subset V$ is *closed* if $V - C$ is an open subset of V.

Corollary. If $f : V \to W$ is continuous and C is closed in W, then $f^{-1}(C)$ is closed in V.

Example. Let
$$C = \{(x, f(x)) | x \in \mathbb{R}\}$$
where $f : \mathbb{R} \to \mathbb{R}$ is continuous. Then C is closed in $\mathbb{R}^2$.

Proof. Let $F : \mathbb{R}^2 \to \mathbb{R}$ by $F(x, y) = f(x) - y$ which is continuous.
Then $C = F^{-1}(\{0\})$ is closed, since $\{0\}$ is closed in $\mathbb{R}$. □

Example.
$$\overline{B}_r(\mathbf{v}) = f_{\mathbf{v}}^{-1}([0, r])$$
is closed in any normed space V.

Example. $\mathbb{Q} \subset \mathbb{R}$ is neither open nor closed.

Example. $V \subset V$, $\phi \subset V$ are both open and closed.

Proposition. C is closed in V if and only if for every sequence $(\mathbf{v}_n) \to \mathbf{v} \in V$ which satisfies $\mathbf{v}_n \in C$ for all n, we have $\mathbf{v} \in C$ as well.

Proof. Suppose C is closed in V, and $(\mathbf{v}_n) \to \mathbf{v}$ with $\mathbf{v} \notin C$.
Now $V - C$ is open, and $\mathbf{v} \in V - C$. So $\exists \varepsilon > 0$ s.t. $B_\varepsilon(\mathbf{v}) \subset V - C$.
Since $(\mathbf{v}_n) \to \mathbf{v}$, there exists N s.t. $\mathbf{v}_n \in B_\varepsilon(\mathbf{v}) \subset V - C$ for all $n > N$. So $\mathbf{v}_n \notin C$. Contradiction.

Conversely, suppose that C is not closed. Then $V - C$ is not open. So there exists $\mathbf{u} \in V - C$ such that for every $\varepsilon > 0$, $B_\varepsilon(\mathbf{v}) \not\subset V - C$, i.e. $B_\varepsilon(\mathbf{v}) \cap C \neq \phi$. Now pick $\mathbf{v}_n$ s.t. $\mathbf{v}_n \in B_{1/n}(\mathbf{v}) \cap C$. Then $||\mathbf{v}_n - \mathbf{v}|| < \frac{1}{n} \to 0$, so $(\mathbf{v}_n) \to \mathbf{v}$ for all $\mathbf{v}n \in C$, but $\mathbf{v} \notin C$. Contradiction. □

1.4 Lipschitz equivalence

We've seen in the first lecture that $|| \cdot ||_1$,$|| \cdot ||_2$,$|| \cdot ||_\infty$ all induce the same notion of convergence on $\mathbb{R}^n$. So $f : \mathbb{R}^n \to V$ is continuous with respect to $|| \cdot ||$ if and only if it's continuous with respect to $|| \cdot ||_\infty$.

Proposition. Suppose $|| \cdot ||$,$|| \cdot ||'$ are two norms on V. The map $id : (V, || \cdot ||) \to (V, || \cdot ||')$ by $id(\mathbf{v}) = \mathbf{v}$ is continuous if and only if there exists some constants $C > 0$ such that
$$||\mathbf{v}||' \leq C||\mathbf{v}||$$
for all $\mathbf{v} \in V$.

Proof. Suppose $||\mathbf{v}||' \leq C||\mathbf{v}||$ for all $\mathbf{v} \in V$.
If $(\mathbf{v}_n) \to \mathbf{v}$ with respect to $||\cdot||$, then $(||\mathbf{v}-\mathbf{v}_n||) \to 0$. But then

$$0 \leq ||\mathbf{v}-\mathbf{v}_n||' \leq C||\mathbf{v}-\mathbf{v}_n||$$

By the squeeze law, $||\mathbf{v}-\mathbf{v}_n||' \to 0$ as well. So $(\mathbf{v}_n) \to \mathbf{v}$ with respect to $||\cdot||'$. This means $id : (V, ||\cdot||) \to (V, ||\cdot||')$ is continuous.

Conversely, suppose $id : (V, ||\cdot||) \to (V, ||\cdot||')$ is continuous. Then there exists $\delta > 0$ s.t. $B_\delta(\mathbf{0}, ||\cdot||) \subset B_1(\mathbf{0}, ||\cdot||')$.
For any $\mathbf{v} \in V, \mathbf{v} \neq 0$, there exists k s.t. $||k\mathbf{v}|| = \frac{\delta}{2}$. So $k\mathbf{v} \in B_\delta(\mathbf{0}, ||\cdot||)$, so $k\mathbf{v} \in B_1(\mathbf{0}, ||\cdot||')$, i.e. $||k\mathbf{v}||' < 1 = \frac{2}{\delta}||k\mathbf{v}||$. Divide by $|k|$ we get

$$||\mathbf{v}||' \leq \frac{2}{\delta}||\mathbf{v}||$$

for all $\mathbf{v} \neq \mathbf{0}$. So we can take $C = \frac{2}{\delta}$. The case $\mathbf{v} = \mathbf{0}$ is trivial. $\square$

Definition. If $||\cdot||$ and $||\cdot||'$ are two norms on V, we say they are *Lipschitz equivalent* if there exists $C > 0$ s.t.

$$\frac{1}{C}||\mathbf{v}|| \leq ||\mathbf{v}||' \leq C||\mathbf{v}||$$

for all $\mathbf{v} \in V$, or say there exists C_1, C_2 such that

$$||\mathbf{v}|| \leq C_1||\mathbf{v}||'$$

and

$$||\mathbf{v}||' \leq C_2||\mathbf{v}||$$

That is also equivalent to

$$id : (V, ||\cdot||) \to (V, ||\cdot||')$$

and

$$id : (V, ||\cdot||') \to (V, ||\cdot||)$$

being both continuous.

Corollary. If $||\cdot||$ and $||\cdot||'$ are Lipschitz equivalent, then:
(a) $(\mathbf{v}_n) \to \mathbf{v}$ with respect to $||\cdot||$ if and only if $(\mathbf{v}_n) \to \mathbf{v}$ with respect to $||\cdot||'$.
(b) $f : V \to W$ is continuous with respect to $||\cdot||$ if and only if $f : V \to W$ is continuous with respect to $||\cdot||'$.
(c) $g : W \to V$ is continuous with respect to $||\cdot||$ if and only if $g : W \to V$ is continuous with respect to $||\cdot||'$.

Example. $||\mathbf{v}||_\infty \leq ||\mathbf{v}||_2 \leq ||\mathbf{v}||_1 \leq n||\mathbf{v}||_\infty$ for all $\mathbf{v} \in \mathbb{R}^n$. So $||\cdot||_\infty$, $||\cdot||_2$, $||\cdot||_1$ are all Lipschitz equivalent.

Problem. Can we find a norm on $\mathbb{R}^n$ that is not Lipschitz equivalent to these?

2 Uniform Convergence

2.1 Notions of Convergence

Let $A \subset \mathbb{R}$, $f, f_n : A \to \mathbb{R}$.

We've known the definition of continuous and boundedness from Analysis I. Now define $C(A)$ to be the set of continuous functions $f : A \to \mathbb{R}$, and $B(A)$ to be the set of bounded functions $F : A \to \mathbb{R}$. Both of these are vector spaces.

We have $C[0,1] \subset B[0,1]$ by maximum value theorem, while $C(0,1) \not\subset B(0,1)$ (take $f(x) = \frac{1}{x}$).

Definition. If $f, f_n : A \to \mathbb{R}$, we say $(f_n) \to f$ *pointwise* if $(f_n(x)) \to f(x)$ for every $x \in A$.

Definition. The *uniform norm* $||\cdot||_\infty$ on $B(A)$ is given by

$$||f||_\infty = \sup_{x \in A} |f(x)|$$

If $f, f_n : A \to \mathbb{R}$, we say $(f_n) \to f$ *uniformly* if $||f - f_n||_\infty \to 0$.

Equivalently, if $(f_n) \to f$ pointwise, then for every $x \in A$ and $\epsilon > 0$, $\exists N$ s.t. $|f_n(x) - f(x)| < \varepsilon$ whenever $n > N$.
If $(f_n) \to f$ uniformly, given ε, we need to find some N that works for all $x \in A$.

Example. Let $A = \mathbb{R}$, $f_n(x) = x + \frac{1}{n}$, $f(x) = x$. Then $(f_n) \to f$ pointwise and uniformly.

Example. Let $A = \mathbb{R}$, $g_n(x) = \left(x + \frac{1}{n}\right)^2$, $g(x) = x^2$. Then $g(n) \to g$ pointwise, but $g_n - g = \frac{2x}{n} + \frac{1}{n^2}$ is not even bounded. So (g_n) does not converge to g uniformly. Nevertheless, $(g_n) \to g$ uniformly on $[a,b]$ for any $a, b \in \mathbb{R}$) (since convergence and uniform convergence is the same on compact sets).

Example. If $(f_n) \to f$ uniformly, then $(f_n) \to f$ pointwise (Immediate from definition).

Theorem. Suppose $f_n \in C(A)$ and $(f_n) \to f$ uniformly on A. Then $f \in C(A)$.

Proof. Given $x \in A$ and $\varepsilon > 0$, we need to find $\delta > 0$ s.t.

$$|f(x) - f(y)| < \varepsilon$$

whenever $|x - y| < \delta$ and $y \in A$.
Since $(f_n) \to f$ uniformly, $\exists N$ s.t.

$$|f_n(y) - f(y)| < \frac{\varepsilon}{4}$$

whenever $n \geq N$ and $y \in A$.
Since f_N is continuous, $\exists \delta > 0$ s.t.

$$|f_N(x) - f_N(y)| < \frac{\varepsilon}{2}$$

whenever $|x-y|<\delta$ and $y\in A$. Then for $|x-y|<\delta$ and $y\in A$,

$$\begin{aligned}|f(x)-f(y)| &\le |f(x)-f_N(x)|+|f_N(x)-f_N(y)|+|f_N(y)-f(y)|\\ &< \frac{\varepsilon}{4}+\frac{\varepsilon}{2}+\frac{\varepsilon}{4}=\varepsilon\end{aligned}$$

which is what we wanted to prove. □

Corollary. $C[a,b]$ is a closed subset of $B[a,b]$ with respect to $||\cdot||_\infty$.

Proof. Recall that C is closed if $c\in C$ whenever $(c_n)\to c$ and $c_n\in C$. □

Example. Let $A=[0,1]$, $f_n(x)=x^n$, $f(x)=\begin{cases}0 & x\in[0,1)\\ 1 & x=1\end{cases}$.
Then $(f_n)\to f$ pointwise but not uniformly, since $f_n\in C[0,1]$, but $f\notin C[0,1]$.

Example. Let $f_n(x)=(1-x)x^n$. Then $(f_n)\to 0$ pointwise. In fact $(f_n)\to 0$ uniformly.

Proof. Given $\varepsilon>0$, we must find N s.t. $|f_n(x)|<\varepsilon$ for all $x\in[0,1]$ whenever $n>N$.

We know $1-\varepsilon<1$, so $(1-\varepsilon)^n\to 0$. Pick N s.t. $(1-\varepsilon)^n<\varepsilon$ whenever $n>N$. Then for $n>N$,

$$|(1-x)x^n|<1\cdot(1-\varepsilon)^n<\varepsilon$$

for $x\in[0,1-\varepsilon]$, and

$$|(1-x)x^n|<\varepsilon\cdot 1^n=\varepsilon$$

for $x\in(1-\varepsilon,1]$. □

Everything so far in this chapter works for $f:A\to W$, where $A\subset V$ and V, W are both normed spaces. (exercise)

Recall that if $f,f_n\in C[a,b]$ with $a,b\in\mathbb{R}$, then $(f_n)\to f$ in L^1 (with respect to $||\cdot||_1$) if

$$||f_n-f||_1=\int_a^b|f_n(x)-f(x)|\to 0$$

Lemma. If $(f_n)\to f$ uniformly on $[a,b]$ and $f_n\in C[a,b]$, then $(f_n)\to f$ in L^1 on $[a,b]$.

Proof. $(f_n)\to f$ uniformly implies that $f\in C[a,b]$.
Given $\varepsilon>0$, pick N s.t.

$$|f_n(x)-f(x)|<\frac{\varepsilon}{(b-a)}$$

for $n>N$ and $x\in[a,b]$. Then

$$||f_n-f||_1=\int_a^b|f_n(x)-f(x)|\,dx<\int_a^b\frac{\varepsilon}{b-a}dx=\varepsilon$$

So $(f_n)\to f$ in L^1. □

Example. Let $A = [0,1]$,

$$f_n(x) = \begin{cases} nx & x \in \left[0, \frac{1}{n}\right] \\ 2 - nx & x \in \left[\frac{1}{n}, \frac{2}{n}\right] \\ 0 & x \in \left[\frac{2}{n}, 1\right] \end{cases}$$

Then $(f_n) \to 0$ pointwise, and in L^1, but not uniformly.

Example. Let $A = [0,1]$,

$$f_n(x) = \begin{cases} n^2 x & x \in \left[0, \frac{1}{n}\right] \\ 2n - n^2 x & x \in \left[\frac{1}{n}, \frac{2}{n}\right] \\ 0 & x \in \left[\frac{2}{n}, 1\right] \end{cases}$$

Then $(f_n) \to f$ pointwise, but not in L^1, nor uniformly.

We woud like to say that a sequence of bounded integrable functions on $[0,1]$ that converges pointwise converges in L^1. But for this to be true, we need a better definition of $\int$ (in measure and probability).

2.2 Power series

Recall some facts about series of complex numbers from Analysis I, for $\sum_{i=0}^{\infty} c_i$, $c_i \in \mathbb{C}$:
1) $\sum_{i=0}^{\infty} c_i = c$ means $\left(\sum_{i=0}^{n} c_i\right) \to c$;
2) $\sum_{i=0}^{\infty} c_i$ converges if and only if $\sum_{i=k}^{\infty} c_i$ converges;
3) $\sum_{i=k}^{\infty} \alpha^i = \frac{\alpha^k}{1-\alpha}$ if $|\alpha| < 1$;
4) If $\sum_{i=0}^{\infty} c_i$ converges, then $(c_n) \to 0$;
5) If $0 < a_i < b_i$ for all i (here $a_i, b_i \in \mathbb{R}$), and $\sum_{i=0}^{\infty} b_i$ converges, then $\sum_{i=0}^{\infty} a_i$ converges as well;
6) If $\sum_{i=0}^{\infty} |c_i|$ converges, then $\sum_{i=0}^{\infty} c_i$ converges.

Corollary. If $|c_i| < b_i$ for all i and $\sum_{i=0}^{\infty} b_i$ converges, then $\sum_{i=0}^{\infty} c_i$ converges.

Proof. Follows from (5) and (6). □

Definition. A *power series* is

$$\sum_{i=0}^{\infty} a_i (z_i)^i$$

where $a_i, c, z \in \mathbb{C}$. Call c the *center* of the series.

Proposition. Suppose $\sum_{i=0}^{\infty} a_i (z_0 - c)^i$ converges for some $z_0 \in \mathbb{C}$. Then the series $\sum_{i=0}^{\infty} a_i (z_0 - c)^i$ converges for all z with $|z - c| < |z_0 - c|$.

Proof. By (4), $\left(a_i (z_0 - c)^i\right) \to 0$. Pick N such that $|a_i (z_0 - c)^i| < 1$ for all $i \geq N$.

By (2), suffices to show that $\sum_{i=N}^{\infty} a_i (z-c)^i$ converges. Now

$$|a_i (z-c)^i| = |a_i (z_0 - c)^i| \cdot \left| \frac{z-c}{z_0-c} \right|^i \leq 1 \cdot \alpha^i$$

(call this 'Key Estimate', to be used later) for $i \geq N$ where $\alpha = \left| \frac{z-c}{z_0-c} \right|$.
For $|z-c| < |z_0 - c|$, $\alpha < 1$, so $\sum_{i=N}^{\infty} \alpha^i$ converges.
By corollary, it follows that $\sum_{i=0}^{\infty} a_i (z-c)^i$ converges. □

Definition.

$$R = \sup \left\{ |z-c| | \sum_{i=0}^{\infty} a_i (z-c)^i \text{ converges} \right\}$$

is the *radius of convergence* of this series.

The above proposition says that $\sum_{i=0}^{\infty} a_i (z-c)^i$ converges for all $z \in B_R(c) = \{z \in \mathbb{C} | |z-c| < R\}$.

We can define $f : B_R(c) \to \mathbb{C}$ by

$$f(z) = \sum_{i=0}^{\infty} a_i (z-c)^i$$

Let

$$p_n(z) = a_i (z-c)^i$$

Then $(p_n) \to f$ pointwise on $B_R(c)$.

Theorem. With notation as above, $(p_n) \to f$ uniformly on $\bar{B}_r(c) = \{z \in \mathbb{C} | |z-c| \leq r\}$ for any $r < R$.

Proof. Fix $z_0 \in \mathbb{C}$ with $r < |z_0 - c| < R$. Then $\sum_{i=0}^{\infty} a_i (z_0 - c)^i$ converges. Let

$$E_n(z) = f(z) - p_n(z) = \sum_{i=n+1}^{\infty} a_i (z-c)^i$$

We want to show that given $\varepsilon > 0$, $\exists N$ s.t. $|E_n(z)| < \varepsilon$ for all $n > N$ and $z \in \bar{B}_r(c)$.

Pick N_0 with $|a_i (z_0 - c)^i| < 1$ for all $i \geq N_0$ as in the proof of the previous proposition.

Now for $n > N_0$, Key Estimate says that

$$\begin{aligned} |E_n(z)| &= \left| \sum_{i=m}^{\infty} a_i (z-c)^i \right| \\ &\leq \sum_{i=n+1}^{\infty} |a_i (z-c)^i| \\ &\leq \sum_{i=n+1}^{\infty} \alpha(z)^i \end{aligned}$$

where $\alpha(z) = \frac{|z-c|}{|z_0-c|}$.
If $z \in \bar{B}_r(c)$, $\alpha(z) \leq \alpha_0 = \frac{r}{|z_0-c|} < 1$. So

$$|E_n(z)| \leq \sum_{i=1}^{\infty} \alpha^i = \frac{\alpha_0^{n+1}}{1-\alpha_0}$$

Now $\alpha_0 < 1$, so $\frac{\alpha_0^{n+1}}{1-\alpha_0} \to 0$ as $n \to \infty$. Pick $N > N_0$ s.t. $\frac{\alpha_0^{n+1}}{1-\alpha_0} < \varepsilon$ for $n > N$. Then $|E_n(z)| < \varepsilon$ for all $n > N$ and $z \in \bar{B}_r(c)$ which is what we wanted. $\square$

Remark. (p_n) may not converge uniformly on $B_R(c)$. For example, $\sum_{i=0}^{\infty} x^i$ has $R = 1$, and equals $f(x) = \frac{1}{1-x}$ on $B_1(0)$, but p_n is a polynomial, so bounded on $\bar{B}_1(0)$, so $f(x) - p_n(x)$ is not even a bounded function on $B_1(0)$.

Corollary.

$$f(z) = \sum_{i=0}^{\infty} a_i (z-c)^i$$

is a continuous map $f : B_R(c) \to \mathbb{C}$.

Proof. $p_n = \sum_{i=0}^{n} a_I (z-c)^i$ is a polynomial, so is continuous as a map $\mathbb{C} \to \mathbb{C}$. $(p_n) \to f$ uniformly on $\bar{B}_r(c)$ for any $r < R$, so $f : \bar{B}_r(c) \to \mathbb{C}$ is continuous for any $r < R$.
Given $z \in B_R(c)$, pick r with $z \in B_r(c)$. Then f is continuous at z. So f is continuous at all $z \in B_R(c)$, i.e. $f : B_R(c) \to \mathbb{C}$ is continuous. $\square$

We can now construct lots of continuous functions using power series.

Example.

$$\exp(z) = \sum_{i=0}^{\infty} \frac{z^i}{i!}$$

has $R = \infty$, so is a well defined, continuous function on $\mathbb{C}$.

Let $f(x) = \exp(x)$ for $x \in \mathbb{R}$. We want to show that $f'(x) = f(x)$:

$$\frac{d}{dx}\left(\sum_{i=0}^{\infty} \frac{x_i}{i!}\right) = \sum_{i=0}^{\infty} \frac{ix^{i-1}}{i!} = \sum_{i=1}^{\infty} \frac{x^{i-1}}{(i-1)!} = \exp(x)$$

this looks easy, but why does the first equality hold?

Example. Suppose

$$\sum_{i=0}^{\infty} a_i (z-c)$$

has radius of convergence R. Then if $p_n = \sum_{i=0}^{\infty} a_i (z-c)^i$, $(p_n) \to f(z) = \sum_{i=0}^{\infty} a_i (z-c)^i$ uniformly on $\bar{B}_r(c)$ for all $r < R \implies f$ is continuous on $\bar{B}_r(c)$ for $r \in R$.
Take $U_r = B_r(c)$, so f is continuous on U_r for $r < R$. U_r is open. So f is continuous on $\cup_{r<R} U_r = B_R(c)$.

2.3 Integration and Differentiation

Recall from Analysis I:

Theorem. (Fundamental Theorem of Calculus) If $f \in C[a,b]$, then

$$F(x) = \int_{x_0}^{x} f(y)\,dy$$

exists, and

$$F'(x) = f(x).$$

Some properties of integral:
Suppose $f, g \in C[a,b]$.
(1)

$$\int_{x_0}^{x} f(y) + \lambda g(y)\,dy = \int_{x_0}^{x} f(y)\,dy + \lambda \int_{x_0}^{x} g(y)\,dy$$

(2) If $f(y) \le g(y)$ for all $y \in [a,b]$, then

$$\int_{x_0}^{x} f(y)\,dy \le \int_{x_0}^{x} g(y)\,dy$$

(3)

$$\left|\int_{x}^{x_0} f(y)\,dy\right| \le \left|\int_{x}^{x_0} |f(y)|\,dy\right|$$

Suppose $f_n \in C[a,b]$ and $(f_n) \to f$ uniformly on $[a,b]$. So $f \in C[a,b]$. Thus

$$F(x) = \int_{x_0}^{x} f_n(y)\,dy$$

and

$$F(x) = \int_{x_0}^{x} f(y)\,dy$$

are defined.

Proposition. $(F_n) \to F$ uniformly on $[a,b]$.

Proof. $(f_n) \to f$ uniformly, so given $\varepsilon > 0$, $\exists N$ s.t.

$$|f_n(x) - f(x)| < \varepsilon$$

for all $n > N$ and $x \in [a,b]$. Choose N s.t.

$$|f_n(x) - f(x)| < \frac{\varepsilon}{b-a}$$

for all $n > N$ and $x \in [a, b]$. Then for $x \in [a, b]$,

$$\begin{aligned}
|F_n(x) - F(x)| &= \left| \int_{x_0}^{x} (f_n(y) - f(y))\, dy \right| \\
&\le \left| \int_{x_0}^{x} |f_n(y) - f(y)|\, dy \right| \\
&\le \left| \int_{x_0}^{x} \frac{\varepsilon}{b-a} dy \right| dy \\
&= \frac{\varepsilon |x - x_0|}{|b - a|} \\
&\le \varepsilon
\end{aligned}$$

So $(F_n) \to F$ uniformly on $[a, b]$. □

Note that $(f_n) \in C(\mathbb{R})$, $(f_n) \to f$ uniformly does not imply $(F_n) \to F$ uniformly on $\mathbb{R}$. (But does on $[a, b]$ for $a, b \in \mathbb{R}$).

Let

$$f(y) = \sum_{i=0}^{\infty} a_i (y - c)^i$$

be a real power series ($a_i, c, y \in \mathbb{R}$) with radius of convergence R. Then if the partial sum $p_n(y) = \sum_{i=0}^{n} a_i (y - c)^i$, then $(p_n) \to f$ uniformly on $[c - r, c + r]$ for any $r < R$.

Corollary.

$$\int_c^x f(y)\, dy = \sum_{i=0}^{\infty} \frac{a_i}{i+1} (x - c)^{i+1}$$

for all $x \in (c - R, c + R)$.

Proof. Given $x \in (c - R, c + R)$, pick r with $|x - c| < r < R$. Then $(p_n) \to f$ uniformly on $[c - r, c + r]$, so by proposition

$$(P_n) \to \int_c^x f(y)\, dy$$

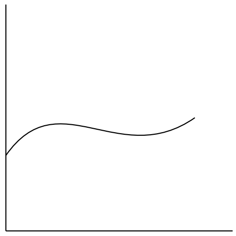
where

$$P_n = \int_c^x p_n(y)\, dy = \sum_{i=0}^{n} \frac{a_i}{i+1} (x - c)^{i+1}$$

□

Q: If $(f_n) \to f$ uniformly, what can I say about (f_n)?
A: Nothing, because:

Example. Take $f_n(x) = \frac{1}{n} \sin nx$, $x \in [0, \pi]$. Then $(f_n) \to 0$ uniformly on $[0, \pi]$, but $f_n'(x) = \cos nx$ doesn't converge for any $x \in (0, \pi)$.



Proposition. If

$$f(y) = \sum_{i=0}^{\infty} a_i (y-c)^i$$

converges on $(c-R, c+R)$, then

$$f(y) = \sum_{i=0}^{\infty} i a_i (y-c)^{i-1}$$

on $(c-R, c+R)$.

Proof.

Lemma.

$$\sum_{i=0}^{\infty} i a_i (y-c)^{i-1}$$

converges for all $y \in (c-R, c+R)$.

Pick y_0 with $|y-c| < |y_0-c| < R$.
$\sum_{i=0}^{\infty} a_i (y-c)^i$ converges, so by 'Key Estimate', $\exists N$ s.t.

$$|a_i (y-c)|^i < \alpha^i$$

for all $i \geq N$, where $\alpha = \left|\frac{y-c}{y_0-c}\right| < 1$.

If $y = c$, $\sum i a_i (y-c)^{i-1}$ obviously converges. If not, estimate

$$\left| i a_i (y-c)^{i-1} \right| < \frac{i}{|y-c|} \alpha^i$$

Now $\sum_{i=0}^{\infty} \frac{i}{|y-c|} \alpha^i$ converges by Ratio Test. So $\sum_{i=0}^{\infty} i a_i (y-c)^{i-1}$ converges as well. □

Now begin the proof of proposition:

$$g(y) = \sum_{i=0}^{\infty} i a_i (y-c)^{i-1}$$

is continuous on $(c-R, c+R)$. So by corollary,

$$\int_c^x g(y)\,dy = \sum_{i=1}^{\infty} a_i (x-c)^i = f(x) - f(c)$$

By Fundamental Theorem of Calculus, $f'(x) = g(x)$. $\square$

Application: Power series solutions of ODEs are legit (as long as we check the radius of convergence).

3 Compactness

3.1 Compact subsets of $\mathbb{R}^n$

Let V be a normed space. Then if $(\mathbf{v}_n) \to \mathbf{v} \in V$ and $\left(\mathbf{v}_{\mathbf{n_j}}\right)$ is a subsequence of $(\mathbf{v}_n)$, then $\left(\mathbf{v}_{\mathbf{n_j}}\right) \to \mathbf{v}$. We leave this as an exercise.

Definition. $A \subset V$ is bounded if $\exists M \in \mathbb{R}$ s.t. $||\mathbf{v}|| \leq M$ for all $\mathbf{v} \in A$.

If $||\cdot||$ and $||\cdot||'$ are Lipschitz equivalent, then boundedness with respect to the two norms are equivalent.

Corollary. (Bolzano-Weierstrass in $\mathbb{R}^n$) If $(\mathbf{v}_k)$ is a bounded sequence in $\mathbb{R}^n$, it has a converging subsequence.

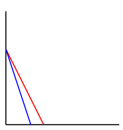
Proof. To prove this, simply pick a subsequence with the first coordinate convergent, then pick a subsequence of that subsequence with the second coordinate convergent, etc..

Let $\mathbf{v}_k = (v_{1,k}, ..., v_{n,k})$.
$(\mathbf{v}_k)$ is bounded, so $(v_{i,k})$ is bounded for all $1 \leq i \leq n$. By B-W theorem, there exists a convergent subsequence $\left(v_{1,k_j^1}\right)$ of $(v_{1,k})$. Now the sequence $\left(v_{2,k_j^1}\right)$ is bounded. So by B-W, there exists a subsequence $\left(v_{2,k_j^2}\right)$ which converges. Then by the previous exercise, $\left(v_{1,k_j^2}\right)$ converges.
Now consider the sequence $\left(v_{3,k_j^2}\right)$. By B-W, it has a convergent subsequence $\left(v_{3,k_j^3}\right)$. etc.

Apply B-W n times, we get $\left(\mathbf{v}_{k_j^n}\right)$ of original $(\mathbf{v}_n)$ s.t. $\left(v_{i,k_j^n}\right)$ converges for $1 \leq i \leq n$. So $\left(\mathbf{v}_{k_j^n}\right)$ converges. ☐

Example. Let $V = C[0,1]$ with $||\cdot||_\infty$, and

$$f_n(x) = \begin{cases} 1 - nx & x \in \left[0, \frac{1}{n}\right] \\ 0 & x \in \left[\frac{1}{n}, 1\right] \end{cases}$$



If

$$f(x) = \begin{cases} 1 & x = 0 \\ 0 & x > 0 \end{cases}$$

then $(f_n) \to f$ pointwise. Then (f_n) is bounded with respect to $||\cdot||_\infty$ but has no convergent subsequence.

Proof. Suppose $\left(f_{n_j}\right) \to g$ uniformly, then $\left(f_{n_j}\right) \to g$ pointwise, so $g = f$. But $f \notin C[0,1]$, so $\left(f_{n_j}\right) \not\to f$ uniformly. $\square$

Definition. We say $A \subset V$ is sequentially compact (s.compact) if any sequence $(\mathbf{v}_n)$ in A has a convergent subsequence $\left(\mathbf{v}_{n_j}\right) \to \mathbf{v} \in A$.

Example. R is not s.compact, since (n) has no convergent subsequence.

Example. $A = (0,2)$ is not s.compact, since $\left(\frac{1}{n}\right) \to 0 \notin A$.

Proposition. Suppose $A \subset V$ is s.compact. Then A is closed in V and bounded.

Proof. We prove the contrapositive:

If A is not closed, then there exists a sequence $(\mathbf{v}_n) \to \mathbf{v}$ with $\mathbf{v}_n \in A$ for all n but $\mathbf{v} \notin A$. By the exercise, any subsequence $\left(\mathbf{v}_{n_j}\right)$ converges to $\mathbf{v} \notin A$. So A is not s.compact.

If A is not bounded, then for all $n \in \mathbb{N}$ we can find $\mathbf{v}_n \in A$ with $||\mathbf{v}_n|| \geq n$. We claim that $\left(\mathbf{v}_{n_j}\right)$ has no convergent subsequence: if $\left(\mathbf{v}_{n_j}\right) \to \mathbf{v}$, then $\exists J$ s.t. $||\mathbf{v}_{n_j} - \mathbf{v}|| < 1$ for all $j > J$. So

$$||v_{n_j}|| \leq ||\mathbf{v}|| + ||\mathbf{v_{n_j}} - \mathbf{v}|| \leq ||\mathbf{v}|| + 1$$

for all $j > J$, but this is impossible since $n_j \geq j$, so $||v_{n_j}|| \geq j \to \infty$ as $j \to \infty$.

It follows that $\mathbf{v}_n$ has no convergent subsequence, so A is not s.compact. $\square$

Theorem. (Heine-Borel) $A \subset \mathbb{R}^n$ is s.compact if and only if A is closed and bounded.

Proof. By the proposition, A is s.compact $\implies$ A is closed and bounded.
Conversely, suppose A is closed and bounded, and $(\mathbf{v}_n)$ is a sequence in A. Then $(\mathbf{v}_n)$ is bounded (since A is). So by B-W, it has a convergent subsequence. Since A is closed, $\mathbf{v} \in A$. So A is s.compact. $\square$

Remark. By previous example, $\bar{B}_1(0)$ in $C[0,1]$ with $||\cdot||_\infty$ is closed and bounded but not s.compact since (f_n) has no convergent subsequence. So Heine-Borel theorem does not hold in general spaces.

Remark. If $A \subset V$ a normed space, then A is s.compact $\iff$ A is compact.

Proposition. Suppose $C \subset V$ is s.compact and $f : C \to W$ is continuous. Then $f(C)$ is s.compact.

Proof. Suppose $(\mathbf{w}_n)$ is a sequence in $f(C)$. Pick $\mathbf{v}_n \in C$ with $f(\mathbf{v}n) = \mathbf{w}_n$. We know C is s.compact, so $(\mathbf{v}_n)$ has a convergent subsequence $\left(\mathbf{v}_{n_j}\right) \to \mathbf{v} \in C$.

Now f is continuous, so $\left(\mathbf{w}_{n_j}\right) = \left(f\left(\mathbf{v}_{n_j}\right)\right) \to (f(\mathbf{v})) \in f(C)$. So $f(C)$ is s.compact. $\square$

We'll use the above to prove maximum value theorem.

Lemma. If $A \subset \mathbb{R}$ is closed and bounded, then $\sup A \in A$.

Proof. A is bounded, so $\sup A$ exists. Pick $x_n \in A$ with $\sup A - \frac{1}{n} \leq x_n \leq \sup A$. Then $(x_n) \to \sup A$. The result follows since A is closed. □

Theorem. (Maximum value theorem) Suppose C is s.compact, $f : C \to \mathbb{R}$ is continuous. Then there exists $\mathbf{v} \in V$ s.t.

$$f(\mathbf{v}) \geq f(\mathbf{v}')$$

for all $\mathbf{v}' \in C$.

Proof. We know $A = f(C)$ is a s.compact subset of $\mathbb{R}$, so it is closed and bounded. So by the lemma, $\sup A$ is in $A = f(C)$. So pick $\mathbf{v} \in C$ with $f(\mathbf{v}) = \sup A$. □

Application: Norms on $\mathbb{R}^n$:

Let $|| \cdot ||$ be a norm on $\mathbb{R}^n$.

Lemma. The map id:$(\mathbb{R}^n, || \cdot ||_1) \to (\mathbb{R}^n, || \cdot ||)$ is continuous.

Proof. Write $\mathbf{v} = (v_1, ..., v_n) = \sum_{i=1}^n v_i \mathbf{e}_i$. By the triangle inequality,

$$||\mathbf{v}|| \leq \sum_{i=1}^n ||v_i \mathbf{e}_i|| = \sum_{i=1}^n |v_i| ||\mathbf{e}_i|| \leq C \sum_{i=1}^n |v_i| = C||\mathbf{v}||_1$$

Where $C = \max_{1 \leq i} \leq n \{||\mathbf{e}_j||\}$. By criterion of section 1.4, the given map is continuous. □

Corollary. The map $f : (\mathbb{R}^n, || \cdot ||_1) \to \mathbb{R}$ given by $f(\mathbf{v}) = ||\mathbf{v}||$ is continuous.

Theorem. $|| \cdot ||$ is Lipschitz equivalent to $|| \cdot ||_1$.

Proof. Let $S = \{\mathbf{v} \in \mathbb{R}^n \mid ||\mathbf{v}_1 = 1\} = g^{-1}(\{1\})$, where $g(\mathbf{v}) = ||\mathbf{v}||_1$.

Now $g : (\mathbb{R}^n, || \cdot ||_1) \to \mathbb{R}$ is continuous, $\{1\}$ is closed in $\mathbb{R}$, so $g^{-1}(\{1\})$ is closed in $(\mathbb{R}^n, || \cdot ||_1)$. S is also obviously bounded in $(\mathbb{R}^n, || \cdot ||_1)$. So S is s.compact by Heine-Borel.

$f : (\mathbb{R}^n, || \cdot ||_1) \to \mathbb{R}$, $f(\mathbf{v}) = ||\mathbf{v}||$ is continuous by corollary. So by maximum value theorem, there exists $\mathbf{v}_\pm \in S$ s.t.

$$C_- = f(\mathbf{v}_-) \leq f(\mathbf{v}) \leq f(\mathbf{v}_+) = C_+$$

for all $\mathbf{v} \in S$, i.e. $C_- \leq \mathbf{v} \leq \mathbb{C}_+$ for all $\mathbf{v} \in S$ where $C_- = ||\mathbf{v}_-|| > 0$ since $\mathbf{v}_- \in S \implies \mathbf{v}_- \neq \mathbf{0} \implies \mathbf{v}_- \neq 0$.

Then for $\mathbf{v} \neq 0$ in $\mathbb{R}^n$, $\mathbf{v}/||\mathbf{v}||_1 \in S$. So

$$0 < C_- \leq ||\frac{\mathbf{v}}{||\mathbf{v}||_1} \leq C_+$$

i.e.

$$C_-||\mathbf{v}||_1 \leq ||\mathbf{v}|| \leq C_+||\mathbf{v}||_1$$

where $C_-, C_+ > 0$. So the two norms are Lipschitz equivalent. □

Corollary. Any two norms on $\mathbb{R}^n$ are Lipschitz equivalent.

3.2 Completeness

Let V be a normed space, and let $(\mathbf{v}_n)$ be a sequence in V.

Definition. The sequence $(\mathbf{v})_n$ is *Cauchy* if given $\varepsilon > 0$, there exists N s.t. $||\mathbf{v}_n - \mathbf{v}_m|| < \varepsilon$ for all $n, m \geq N$.

Example. If $(\mathbf{v}_n) \to \mathbf{v}$, then $(\mathbf{v}_n)$ is Cauchy.

Proof. Given $\varepsilon > 0$, pick N s.t. $||\mathbf{v}_n - \mathbf{v}|| < \frac{\varepsilon}{2}$ for all $n \geq N$. Then for $n, m \geq N$, by triangle inequality,

$$||\mathbf{v}_n - \mathbf{v}_m|| \leq ||\mathbf{v}_n - \mathbf{v}|| + ||\mathbf{v} - \mathbf{v}_m|| < \varepsilon$$

i.e. $(\mathbf{v}_n)$ is Cauchy. □

Example. Let $s_n = \sum_{i=1}^{n} \frac{1}{i}$. Then s_n diverges. Also it is not Cauchy, even though $|s_n - s_{n+1}| \to 0$ as $n \to \infty$.

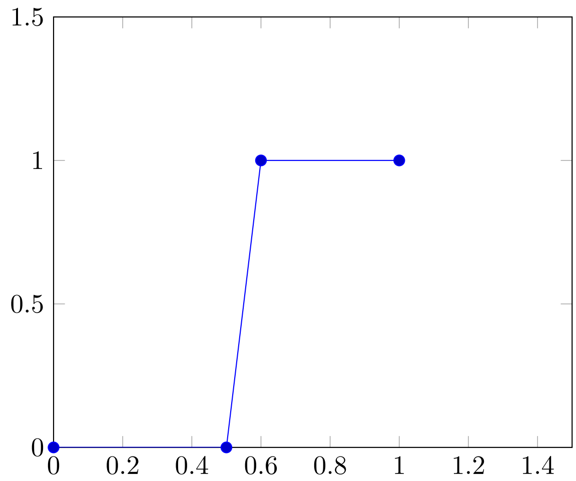
Cauchy sequences *want* to converge.

Example. Given $\varepsilon > 0$, pick N s.t. $||\mathbf{v}_n - \mathbf{v}_m|| < \varepsilon$ for all $n, m \geq N$. Then all but finitely many terms of $(\mathbf{v}_n)$ are contained in $B_\varepsilon(\mathbf{v}_N)$.

However they may not have an element of V to converge to.

Example. Let $V = C[0,1]$ with $|| \cdot ||_1$. Take

$$f_n = \begin{cases} 0 & x \in [0, 1/2] \\ n(x - 1/2) & x \in [1/2, 1/2 + 1/n] \\ 1 & x \in [1/2 + 1/n, 1] \end{cases}$$



$f(n)$ is Cauchy:
If $m, n \geq> N$, $|f_n(x) - f_m(x)| = 0$ if $x \notin A_n = [1/2, 1/2 + 1/N]$, and < 1 if $x \in A_N$. Then

$$||f_n - f_m||_1 = \int_0^1 |f_n(x) - f_m(x)|dx \leq \int_{1/2}^{1/2+1/N} 1dx = \frac{1}{N}$$

so (f_n) is Cauchy.

Now let

$$f(x) = \begin{cases} 0 & x \in [0, 1/2] \\ 1 & x \in (1/2, 1] \end{cases}$$

which is not in $C[0, 1]$.

If $(f_n) \to g \in C[0, 1]$ then $(f_n) \to g$ with respect to $|| \cdot ||_1$ on $[0, 1] - A_n$ for any $N > 0$. On the other hand, $(f_n) \to f$ uniformly on $[0, 1] - A_N$ for any $N > 0$.

On the other hand, $(f_n) \to f$ uniformly on $[0, 1] - A_N$ for any $N > 0$. So $(f_n) \to f$ with respect to $|| \cdot ||_1$ on $[0, 1] - A_N$ for all $N > 0$. Therefore $g(x) = f(x)$ for all $x \in [0, 1]$. Contradiction.

Definition. A normed space V is *complete* if every Cauchy sequence $(\mathbf{v}_n)$ in V converges to a limit $\mathbf{v} \in V$.

Example. $(C[0, 1], || \cdot ||_1)$ is not complete.

Application: Completeness of $\mathbb{R}^n$.

Let V be a normed vector space, and suppose $(\mathbf{v}_n)$ is a Cauchy sequence in V.

Lemma. $(\mathbf{v}_n)$ is bounded. (Exercise)

Lemma. If $(\mathbf{v}_n)$ has a convergent subsequence $(\mathbf{v}_{n_i}) \to \mathbf{v} \in V$, then $(\mathbf{v}_n) \to \mathbf{v}$.

Proof. Given $\varepsilon > 0$, pick M s.t. $||\mathbf{v}_n - \mathbf{v}_m|| < \frac{\varepsilon}{2}$ whenever $n, m > M$. Now $\mathbf{v}_{n_i}$ converges to $\mathbf{v}$, so pick I s.t. $||\mathbf{v}_{n_i} - \mathbf{v}|| < \frac{\varepsilon}{2}$ whenever $i > I$.
So choose $I' > I$ s.t. $n_{I'} \geq M$. Then for $n > n_{I'}$,

$$||\mathbf{v}_n - \mathbf{v}|| \leq ||\mathbf{v}_n - \mathbf{v}_{n_{I'}}|| + ||\mathbf{v}_{n_{I'}} - \mathbf{v}|| < \varepsilon$$

So $(\mathbf{v}_n) \to \mathbf{v}$. □

Theorem. $\mathbb{R}^n$ is complete.

Proof. Suppose $(\mathbf{v}_n)$ is a Cauchy sequence in $\mathbb{R}^n$. By lemma 1, $(\mathbf{v}_n)$ is bounded. By B-W, $(\mathbf{v}_n)$ has a convergent subsequence $(\mathbf{v}_{n_i}) \to \mathbf{v}$. By lemma 2, $(\mathbf{v}_n) \to \mathbf{v}$, i.e. every Cauchy sequence converges. So $\mathbb{R}^n$ is complete. □

Remark. If $|| \cdot ||$ and $|| \cdot ||'$ are Lipschitz equivalent, then $(\mathbf{v}_n)$ is Cauchy with respect to the two norms are equivalent. So Completeness with respect to the two norms are equivalent.

Since all norms on $\mathbb{R}^n$ are Lipschitz equivalent, the the theorem holds for any norm.

We saw $(C[0,1], || \cdot ||_1)$ is not complete. What about $(C[0,1], || \cdot ||_\infty)$?

Bounded sequences need not have convergent subsequences.

Theorem. $C[0,1]$ is complete with respect to $|| \cdot ||_\infty$.

Proof. Given a Cauchy sequence (f_n), we must find $f \in C[0,1]$ s.t. $(f_n) \to f$ uniformly.

Given $\varepsilon > 0$, choose N s.t. $||f_n - f_m|| < \varepsilon/2$ for all $n, m \geq N$. Then if $x \in [0,1]$,

$$\begin{aligned} |f_n(x) - f_m(x) &\leq \max_{x \in [0,1]} |f_n(x) - f_m(x)| \\ &= ||f_n - f_m||_\infty \\ &< \varepsilon/2 < \varepsilon \end{aligned}$$

For $n, m \geq N$.

So $(f_n(x))$ is a Cauchy sequence in $\mathbb{R}$. But $\mathbb{R}$ is complete. So $\lim_{n \to \infty} f_n(x)$ exists.

Define $f(x) = \lim_{n \to \infty} f_n(x)$. Then $(f_n) \to f$ pointwise.

Now we want to prove $(f_n) \to f$ uniformly. Given $\varepsilon > 0$, and $x \in [0,1]$, pick M (depending on x) s.t. $|f_n(x) - f(x) < \varepsilon/2$ whenever $n \geq M$.

Let $R = \max(N, M)$, then for $n \geq N$,

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x) - f_R(x)| + |f_R(x) - f(x)| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon \end{aligned}$$

for $n, R \geq N$. i.e. $|f_n(x) - f(x)| < \varepsilon$ for all $x \in [0,1]$ i.e. $||f_n - f||_\infty < \varepsilon$. So $(f_n) \to f$ uniformly.

$f_n \in C[0,1] \implies f \in C[0,1]$. So $(f_n) \to f \in C[0,1]$ uniformly. □

3.3 Uniform continuity

Suppose V, W are normed spaces, $A \subset V$.

Definition. $f : A \to W$ is *uniformly continuous* if for every $\varepsilon > 0$, $\exists \delta > 0$ s.t. $||f(\mathbf{v}) - f(\mathbf{v}')|| < \varepsilon$ whenever $||\mathbf{v} - \mathbf{v}'|| < \delta$.

Example. Let $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Then $f(x+\delta) - f(x) = 2x\delta + \delta^2$. For fixed δ, $2x\delta + \delta^2 \to \infty$ as $x \to \infty$. So $f(x) = x^2$ is not uniformly continuous.

Example. Let $f : (0,1] \to \mathbb{R}$ with $f(x) = \frac{1}{x}$. This is not uniformly continuous as well (consider $x \to 0$).

Theorem. If C is s.compact, and $f : C \to W$ is continuous, then f is uniformly continuous.

Proof. Suppose f is not uniformly continuous. Then there exists $\varepsilon > 0$ s.t. for all $n > 0$ we can find $\mathbf{v}_n, \mathbf{w}_n \in C$ with $||\mathbf{v}_n - \mathbf{w}_n|| < \frac{1}{n}$, and $||f(\mathbf{v}_n) - f(\mathbf{w}_n)|| \geq \varepsilon$ (else f is uniformly continuous).

Since C is s.compact, $(\mathbf{v}_n)$ has a convergent subsequence $(\mathbf{v}_{n_i}) \to \mathbf{v}^* \in C$.

f is continuous and $\mathbf{v}^* \in C$, so $\exists \delta > 0$ s.t. $||f(\mathbf{v}) - f(\mathbf{v}^*)|| < \varepsilon/2$ whenever $\mathbf{v} \in B_\delta(\mathbf{v}^*)$.

If $\mathbf{v}, \mathbf{v}' \in B_\delta(\mathbf{v}^*)$, then

$$\begin{aligned} ||f(\mathbf{v}) - f(\mathbf{v}')|| &\leq ||f(\mathbf{v}) - f(\mathbf{v}^*)|| + ||f(\mathbf{v}^*) - f(\mathbf{v}')|| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon \end{aligned}$$

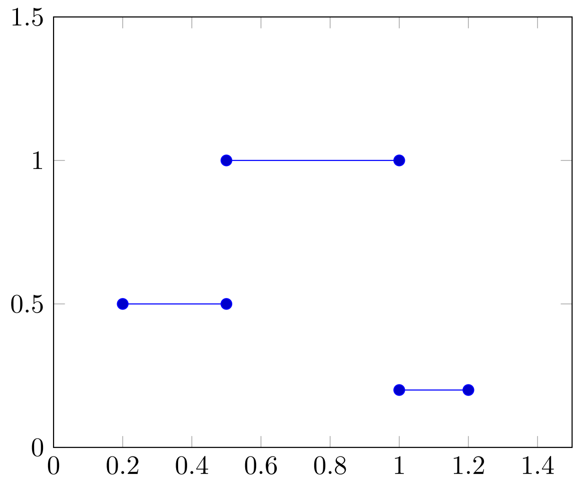
$(\mathbf{v}_{n_i}) \to \mathbf{v}^*$, so pick I_1 s.t. $||\mathbf{v}_{n_i} - \mathbf{v}^*|| < \delta/2$ when $i \geq I_1$.

Pick I_2 s.t. $1/I_2 < \delta/2$. Then for $i \geq \max(I_1, I_2)$, we have $||\mathbf{v}_{n_i} - \mathbf{v}^*|| < \delta/2$ and $||\mathbf{v}_{n_i} - \mathbf{w}_{n_i}|| < \frac{1}{n_i} < \frac{1}{i} < \frac{1}{I_2} < \frac{\delta}{2}$.

So $||\mathbf{w}_{n_i} - \mathbf{v}^*|| < ||\mathbf{w}_{n_i} - \mathbf{v}_{n_i}|| + ||\mathbf{v}_{n_i} - \mathbf{v}^*|| < \delta/2 + \delta/2 = \delta$, i.e. $\mathbf{w}_{n_i}, \mathbf{v}_{n_i} \in B_\delta(\mathbf{v}^*)$, $||f(\mathbf{v}_{n_i}) - f(\mathbf{w}_{n_i})|| \geq \varepsilon$. Contradiction. So f must be uniformly continuous. □

3.4 Application: Integration

Recall from Analysis I: We say $f : [a,b] \to \mathbb{R}$ is piecewise constant if $\exists a = a_0 < a_1 < ... < a_n = b$ and $c_1, ..., c_n \in \mathbb{R}$ s.t. $f(x) = c_i$ if $x \in (a_{i-1}, a_i)$.



Let $P[a,b] = \{f : [a,b] \to \mathbb{R} \mid f \text{ is piecewise constant}\}$. If $f \in P[a,b]$ is as above, then

$$I(f) = \sum_{i=1}^{n} c_i (a_i - a_{i-1}) = \text{"} \int \text{"} f$$

Lemma. If $f, g \in P[a,b]$, $\lambda \in \mathbb{R}$, then

$$f - \lambda g \in P[a,b]$$

and $I(f - \lambda g) = I(f) - \lambda I(g)$.

Write $f \geq g$ if $f(x) \geq g(x)$ for all $x \in [a,b]$.

Lemma. If $f \geq 0$, $I(f) \geq 0$.

So if $f, g \in P[a,b], f \geq g$, then $I(f) \geq I(g)$.

Definition. (Riemann Integral) Suppose $f : [a,b] \to \mathbb{R}$ is bounded. Let

$$\mathcal{U}(f) = \{g \in P[a,b] \mid g \geq f\},$$
$$\mathcal{L}(f) = \{g \in P[a,b] \mid g \leq f\}$$

since f is bounded, these are not empty.
Let

$$U(f) = \{I(g) \mid g \in \mathcal{U}(f)\},$$
$$L(f) = \{I(g) \mid g \in \mathcal{L}(f)\}$$

If $g^+ \in \mathcal{U}(f)$ and $g^- \in \mathcal{L}(f)$, then $g^+ \geq f \geq g^-$. So $I(g^+) \geq I(g^-)$. If $u \in U(f)$ and $l \in L(f)$, then $u \geq l$. So $U(f)$ is bounded below, $L(f)$ is bounded above.

Now let

$$u(f) = \inf U(f),$$
$$l(f) = \inf L(f)$$

Note that $u(f) \geq l(f)$.

We say f is Riemann integrable if $u(f) = l(f)$, in which case we define

$$\int_a^b f(x)\,dx = u(f) = l(f)$$

If $f \in P[a,b]$, then $u(f) = I(f) = l(f)$, so f is RI.

Theorem. If $f \in C[a,b]$, then f is RI.

Lemma. Given $\varepsilon > 0$, $\exists g^+ \in \mathcal{U}(f)$ and $g^- \in \mathcal{L}(f)$ s.t. $I(g^+) - I(g^-) < \varepsilon$.

Proof. $[a,b]$ is closed and bounded in $\mathbb{R}$, so it is s.compact. By last lecture's theorem, $f : [a,b] \to \mathbb{R}$ is uniformly continuous.
So pick δ s.t.

$$|f(x) - f(y)| < \frac{\varepsilon}{b-a}$$

whenever $|x-y| < \delta$. Choose $a = a_0 < a_1 < ... < a_n = b$ such that $a_{i+1} - a_i < \delta$ for all i.

Define

$$c_i^+ = \max_{x \in [a_{i-1}, a_i]} f(x),$$
$$c_i^- = \min_{x \in [a_{i-1}, a_i]} f(x)$$

(These exist by Maximum value theorem) So

$$c_i^+ = f\left(x^+\right) \geq f\left(x^-\right) \forall x \in [a_{i-1}, a_i],$$
$$c_i^- = f\left(x^-\right) \leq f(x) \forall x \in [a_{i-1}, a_i]$$

$x^+, x^- \in [a_{i-1}, a_i] \implies |x^+ - x^-| < \delta$.

Define

$$g^+(x) = c_i^+ \text{ if } x \in [a_{i-1}, a_i),$$
$$g^-(x) = c_i^- \text{ if } x \in [a_{i-1}, a_i)$$

Then $|x^+ - x^-| < \delta \implies c_i^+ - c_i^- < \frac{\varepsilon}{b-a}$ for all i. So to sum up, $g^+ \geq f \geq g^-$ and $g^+ - g^- \leq \frac{\varepsilon}{b-a}$.
Thus $g^+ \in \mathcal{U}(f)$, $g^- \in \mathcal{L}(f)$, and

$$I\left(g^+\right) - I\left(g^-\right) = I\left(g^+ - g^-\right) \leq I\left(\frac{\varepsilon}{b-a}\right) = \varepsilon$$

□

Now prove the theorem:

Proof. $I(g^+) \geq u(f) \geq l(f) \geq I(g^-)$. So $u(f) - l(f) \leq I(g^+) - I(g^-) < \varepsilon$ for all $\varepsilon > 0$, which implies $u(f) = l(f)$. □

Corollary. If $f \in C[a,b]$, $\exists f_k \in P(a,b)$ s.t. $(f_k) \to f$ uniformly on $[a,b]$.

Proof. For each k, choose g_k^+ as in the proof of lemma with $\varepsilon = \frac{1}{k}$. Then $\left(g_k^+\right) \to f$ uniformly. □

Example. (Speed and Distance) Suppose $f[a,b] \to \mathbb{R}^n$ is continuous. $f(t) = (f_1(t), ..., f_n(t))$ where all f_i are continuous.
Define $\int_a^b f(t)\,dt = \left(f_1(t)\,dt, ..., \int_a^b f_n(t)\,dt\right)$ (Integrating pointwise).

If $f(t) = \mathbf{v}(t)$ =velocity of a particle in $\mathbb{R}^n$ at time t, then $\mathbf{p}(b) - \mathbf{p}(a) = \int_a^b f(t)\,dt$ is the displacement of particle from its position at $t = a$. $||\mathbf{v}(t)||$ is the speed of particle.

Proposition. If $f : [a,b] \to \mathbb{R}^n$ is continuous, then

$$||\int_a^b f(t)\,dt|| \le \int_a^b ||f(t)||dt$$

Lemma. If $x_i, y_i \in \mathbb{R}$ satisfy:
(1) $x_i \le y_i$ for all i;
(2) $(x_i) \to x$ and $(y_i) \to y$
Then $x \le y$.

Proof. $y_i - x_i \ge 0, (y_i - x_i) \to y - x \implies y - x \ge 0.$ □

Lemma. The proposition holds if f is piecewise constant (maybe not continuous).

Proof. Suppose $f(t) = \mathbf{v}_i$ for $t \in (a_{i-1}, a_i)$. Then

$$\begin{aligned}||\int_a^b f(t)\,dt|| &= ||I(f)|| \\ &= ||\sum_{i=1}^n (a_{i+1} - a_i)\,\mathbf{v}_i|| \\ &\le \sum_{i=1}^n (a_i - a_{i-1})\,||\mathbf{v}_i|| \\ &= I(||f||) \\ &= \int_a^b ||f||dt.\end{aligned}$$

□

Proof of proposition:

Proof. Choose a sequence of piecewise constant functions $f_k : [a,b] \to \mathbb{R}^n$ s.t. $(f_k) \to f$ uniformly.
Then

$$\int_a^b f_k \to \int_a^b f$$

(uniformly convergence $\implies$ L^1 convergence) and

$$\left(\left|\left|\int_a^b f_k\right|\right|\right) \to \left(\left|\left|\int_a^b f\right|\right|\right)$$

since $||\cdot||$ is continuous.

Also $(||\mathbf{f}_k||) \to ||f||$ uniformly ($||\cdot||$ is continuous). So

$$\left(\int_a^b ||f_k||\right) \to \int_a^b ||f||$$

So now take $x_k = ||\int_a^b f_k||, x = ||\int_a^b f||, y_k = \int_a^b ||f_k||, y = \int_a^b ||f||$.

Then $x_k \leq y_k$, so $x \leq y$. $\square$

4 Differentiation

Slogan: *The derivative is a linear map.*

4.1 Derivative

Definition. Let $U \subset \mathbb{R}^n$ be open, $f : U - \{x_0\} \to \mathbb{R}^m$. We say

$$\lim_{x \to x_0} f(x) = y$$

if the function $\bar{f} : U \to \mathbb{R}^m$ given by

$$\bar{f}(x) = \begin{cases} f(x) & x \neq x_0 \\ y & x = x_0 \end{cases}$$

is continuous at x_0.

Note that we don't care which norms on $\mathbb{R}^n$ or $\mathbb{R}^m$ we use: all the norms on $\mathbb{R}^n$ are Lipschitz equivalent, so they determine the same continuous functions.

Definition. Suppose $U \subset \mathbb{R}^n$ is open, $x_0 \in U$ and $f : U \to \mathbb{R}^m$. We say f is differentiable at x_0 if there is a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ s.t.

$$\lim_{v \to 0} \frac{f(x_0 + v) - (f(x_0) + L(v))}{||v||} = 0$$

If such an L exists, it is unique.

Proof. Suppose L_1, L_2 exist. Subtracting the two limit equations gives

$$\lim_{v \to 0} \frac{L_2(v) - L_1(v)}{||v||} = 0$$

If $v \in \mathbb{R}^n$, $v \neq 0$, then $tv \to 0$ as $t \to 0^+$. So

$$\lim_{t \to 0^+} \frac{L_2(tv) - L_1(tv)}{||tv||} = 0$$

Since L_1, L_2 are linear maps, simplify that and we get $L_2(v) = L_1(v)$. But v is arbitrary. So $L_1 = L_2$. □

When the equation in the definition of differentiability holds, we say

$$Df|_{x_0} = L$$

is the *derivative* of f at x_0. Note that $Df|_{x_0}$ is a *linear map* from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Equivalently, f is differentiable at x_0 with $Df|_{x_0} = L$ if

$$f(x_0 + v) = f(x_0) + L(v) + ||v||\alpha(v)$$

where $\lim_{v \to 0} \alpha(v) = 0$.

Proposition. Suppose $f : U \to \mathbb{R}^m$ is differentiable at $x_0 \in U$. Then f is continuous at x_0.

Lemma. Suppose $L : \mathbb{R}^n \to (W, ||\cdot||)$ is a linear map where W is a normed space. Then $\lim_{v\to 0} L(v) = 0$.

Note that the lemma is false if $\mathbb{R}^n$ is replaced by an arbitrary normed space.

Proof. Let $v = (v_1, ..., v_n) = \sum_{i=1}^n v_i e_i$. Then

$$\begin{aligned} ||L(v)|| &= ||\sum_{i=1}^n v_i L(e_i)|| \\ &\leq \sum_{i=1}^n |v_i| \cdot ||L(e_i)|| \\ &\leq C \sum_{i=1}^n |v_i| \\ &= C||v||_1 \end{aligned}$$

Where $C = \max\{||L(e_1)||, ..., ||L(e_n)||\}$.

Given $\varepsilon > 0$, pick $\delta > \varepsilon/C$. If $||v||_1 < \delta$ then $||L(v)|| < \varepsilon$, so $\lim_{v\to 0} L(v) = 0$. □

Prove of proposition:

Proof. Since f is differentiable at x_0, we have

$$f(x_0 + v) = f(x_0) + L(v) + ||v||\alpha(v)$$

where $\lim_{v\to 0} \alpha(v) = 0$. Now take the limit $v \to 0$ of both sides we have

$$\lim_{v\to 0} f(x_0 + v) = f(x_0)$$

So f is continuous at x_0. □

4.2 The derivative as a matrix

Suppose $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^m$.

We say f is differentiable if f is differentiable at all $x \in U$.

If so, we have $Df : U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$.

From Linear Algebra we know that there is a bijection between $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ and the set of $m \times n$ real matrix:

$$[a_{ij}] \longleftrightarrow L(e_j) = \sum a_{ij} e_i$$

Now let's consider given $f : U \to \mathbb{R}^m$, how we compute $Df = [a_{ij}(x)]$. We first reduce to the case $m = 1$ by writing

$$f(x) = (f_1(x), ..., f_n(x))$$

Then think about $F : U \to \mathbb{R}$.

Proposition. f is differentiable at x_0 if and only if f_i is differentiable for all $1 \leq i \leq m$. If so,

$$Df|_{x_0} = (Df_1|_{x_0}, ..., Df_m|_{x_0}).$$

Proof. Suppose $g : U \to \mathbb{R}^m$. Using the uniform norm on $\mathbb{R}^m$, we see that $\lim_{v\to 0} g(v) = 0$ iff $\lim_{v\to 0} g_i(v) = 0$ for all $1 \leq i \leq m$.

Now let $L : \mathbb{R}^n \to \mathbb{R}^m$. Then

$$\lim_{v\to 0} \frac{f(x_0 + v) - (f(x_0) + L(v))}{||v||} = 0$$

if and only if

$$\lim_{v\to 0} \frac{f_i(x_0 + v) - (f_i(x_0) + L_i(v))}{||v||} = 0$$

for all $1 \leq i \leq m$, i.e. f_i is differentiable at x_0 and $Df_i|_{x_0} = L_i$. □

Summary:

$$Df|_{x_0} = \begin{bmatrix} Df_1|_{x_0} \\ ... \\ Df_m|_{x_0} \end{bmatrix}$$

where $Df_i|_{x_0} : \mathbb{R}^n \to \mathbb{R}$ is a $1 \times n$ matrix $[a_1, ..., a_n]$.

Definition. (Directional Derivative)
Suppose $F : U \to \mathbb{R}$. If $v \in \mathbb{R}^n$, the *directional derivative* of F in direction v at x is

$$\begin{aligned} D_v F|_x &= \lim_{t\to 0} \frac{F(x + tv) - F(x)}{t} \\ &= \frac{d}{dt}(F(x + tv))|_{t=0} \end{aligned}$$

$D_v F$ measures the rate of change of F if I walk away from x at velocity v.

It's also helpful to consider

$$D_v^+ F = \lim_{t\to 0^+} \frac{F(x + tv) - F(x)}{t}$$

and similarly for $D_v^- F$. We can prove that

$$D_v^- F|_x = -D_{-v}^+ F|_x$$

Note: $D_v F$ exists iff $D_v^+ F, D_v^- F$ both exist and are equal.

Example. Consider a special case $v = e_i$. Then

$$\begin{aligned} D_i F|_x &= \frac{\partial F}{\partial x_i}|_x \\ &= D_{e_i} F|_x \\ &= \frac{d}{dt}(F(x_1, ..., x_i + t, ..., x_n))|_{t=0} \\ &= \frac{d}{dt}(F(x_1, ..., x_{i-1}, t, x_{i+1}, ..., x_n))|_{t=x} \end{aligned}$$

is the ith partial derivative of F.

Proposition. If $F : U \to \mathbb{R}$ is differentiable at x, then $D_v F|_x = DF|_x(v)$.

Proof. If $v = 0$ then both sides are 0.

If $v \neq 0$, then $tv \to 0$ as $t \to 0^+$, so differentiability of F implies

$$\lim_{t \to 0^+} \frac{F(x + tv) - ((F(x) + L(tv))}{||tv||} = 0$$

where $L = DF|_x$. So

$$\lim_{t \to 0^+} \frac{F(x + tv) - F(x)}{t} - L(v) = 0$$

i.e. $D_v^+ F|_x = DF|_x(v)$. Then $D_v^- F|_x = -D_{-v}^+ F|_x = -L(-v) = L(v)$. □

If $DF|_x = [a_1, ..., a_n]$ then $a_i = DF|_x(e_i) = D_{e_i} F|_x = D_i F|_x$. So we have

$$DF|_x = [D_1 F|_x, ..., D_n F|_x]$$

Summary: if $f : \mathbb{R}^n \to \mathbb{R}^m$, then

$$Df = \begin{bmatrix} Df_1 \\ ... \\ Df_m \end{bmatrix} = [D_j f_i]$$

Example. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ with $f(x, y, z) = (x^2 + y^2 + z^2, xyz)$. Then

$$Df = \begin{bmatrix} 1 & 2y & 3z^2 \\ yz & xz & xy \end{bmatrix}$$

Note: Just because $D_i F_x$ all exists doesn't mean that F is differentiable at x.

Example. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$F(x, y) = \begin{cases} 0 & xy = 0 \\ H(x, y) & \text{otherwise} \end{cases}$$

where $H(x, y)$ is any arbitrary horrible function. Then

$$D_1 F|_0 = D_2 F|_0 = 0$$

but F may not even be continuous.

We can even have D_vF well defined for every v, but F is not differentiable.

Example. Let $S^1 = \{v \in \mathbb{R}^2 | ||v|| = 1\}$. Choose $h : S \to \mathbb{R}$ to be any function. Define $F : \mathbb{R}^2 \to \mathbb{R}$ by

$$F(v) = \begin{cases} ||v||h(\frac{v}{||v||}) & v \neq 0 \\ 0 & v = 0 \end{cases}$$

Then $D_v^+F|_0 = ||v||h(\frac{v}{||v||})$. If we let $h(-v) = -h(v)$ then $D_v^+F = D_v^-F$, so D_vF is well defined. Now if F is differentiable, $D_vF|_0 = DF|_0(v)$, so $h(v)$ would have to be a linear function on S^1; but h is arbitrary except the one condition above.

A criterion for differentiability: Let $U \subset \mathbb{R}^n$ be open.

Definition. $C^1(U) = \{f : U \to \mathbb{R}|$ for $1 \leq i \leq n$, the partial derivative $D_if|_x$ exists for all $x \in U$ and is a continuous function of $x\}$.

Example.

$$F(x, y, z) = e^{\cos x^2 y + z} - y^2 z \in C^1(\mathbb{R}^3)$$

Theorem. If $F \in C^1(U)$, then F is differentiable on U. Tools used in proof:
- Alternative characterisation of differentiability in 4.1;
- If $\lim_{v\to 0} g(v) = w_0$ and $\lim_{v\to w_0} f(v) = z$, then $\lim_{v\to 0} f(g(v)) = z$;
- Suppose $b : U \to \mathbb{R}$ is bounded on $B_r(v_0)$ for some $r > 0$. Then $\lim_{v\to v_0} b(v)\alpha(v) = 0$ if $\lim_{v\to v_0} \alpha(v) = 0$.

Proof. (of the bullet point):
Since b is bounded, there exists $M \in \mathbb{R}$ s.t. $|b(x)| \leq M$ for all $v \in B_r(v_0)$. Since $\lim_{v\to 0} \alpha(v) = 0$, given $\varepsilon > 0$, there exists $\delta > 0$ s.t. $||\alpha(v)|| < \frac{\varepsilon}{M}$ whenever $v \in B_\delta(v_0)$. Then let $\delta' = \min(\delta, r)$. We have $||b(v)\alpha(v)|| = ||b(v)||||\alpha(v)|| < \varepsilon$ for $v \in B_\delta(v_0)$. So $\lim_{v\to v_0} b(v)\alpha(v) = 0$. □

Proof. (for $n = 2$)
We want to estimate $F(x + v) - F(x)$ for small v. Since U is open, $\exists r > 0$ s.t. $B_r(x) \subset U$.

From now on we assume $||v|| < r$ (since v is small that's reasonable). So $x' \in U$. Since D_1F exists, we write

$$F(x') - F(x) = F(x_1 + v_1, x_2) - F(x_1, x_2) = v_1 DF|_x + |v_1|\alpha_1(v_1)$$

where $\lim_{v_1\to 0} \alpha_1(v_1) = 0$. Similarly

$$F(x + v) - F(x') = v_2 \cdot D_2F|_x + |v_2|\alpha_2(v_2)$$

where $\lim_{v_2\to 0} \alpha_2(v_2) = 0$.
Mistake! Here $\alpha_2(v_2)$ depends on v_1.

Instead, apply 1-variable mean value theorem to $f(t) = F(x_1 + v_1, x_2 + t)$ to write

$$F(x + v) - F(x') = v_2 D_2F|_{x''(v)}$$

where $x''(v) = (x_1 + v_1, x_2 + h(v))$ where $0 < h(v) < v_2$. Then as before, we can add to get

$$F(x+v) - F(x) = L(v) + ||v||E(v)$$

where

$$\begin{aligned} E(v) &= \frac{|v_1|}{||v||}\alpha_1(v_1) \quad &+\frac{|v_2|}{||v||}(D_2F|_{x'(v)} - D_2F|_x) \\ &= E_2(v) &+E_1(v) \end{aligned}$$

Note that $||x''(v) - x||_2 = (v_1^2 + h(v)^2)^{0.5} \leq ||v||_2$. So $\lim_{v\to 0} x''(v) = x$.

Now D_2F is continuous, so $\lim_{v\to 0} D_2F|_{x''} - D_2F|_x = 0$.

We'll show that as $v \to 0$, $E_1(v), E_2(v) \to 0$, then we are done.

• E_1: As $v \to 0$, $x' \to x$. Now D_2F is continuous, so

$$\lim_{x'\to x}(D_2F|_{x'} - D_2F|_x) = 0$$

So

$$\lim_{v\to 0}(D_2F|_{x'} - D_2F|_x) = 0$$

Now $\frac{|v_2|}{||v||} < 1$ for all $v \in \mathbb{R}^2\backslash\{0\}$, so by lemma $E_1 \to 0$.

• E_2: $\lim_{v\to 0} v_1 = 0$ and $\lim_{v_1\to 0}\alpha(v_1) = 0$, so $\lim_{v\to 0}\alpha_1(v_1) = 0$. Same as above we get $E_2 \to 0$.

(Refer to DC notes last page of Section 6.1 (p66).)

□

Example. Let $V = M_{n\times n}(\mathbb{F}) \cong \mathbb{R}^{n^2}$, $f : V \to V$ by $f(x) = x^2$. Then

$$f(x+v) = (x+v)^2 = x^2 + xv + vx + v^2 = f(x) + L_x(v) + v^2$$

where

$$L_x(v) = xv + vx$$

is linear in V. Compare with the definition we get

$$DF|_x = Lx.$$

4.3 The Chain Rule

Theorem. (Chain Rule)
Suppose $g : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x, and $f : \mathbb{R}^m \to \mathbb{R}^l$ is differentiable at $g(x)$. Then $f \circ g : \mathbb{R}^n \to \mathbb{R}^l$ is differentiable at x, and

$$D(f \circ g)|_x = Df|_{g(x)} \circ Dg|_x$$

Example. Suppose $r : \mathbb{R} \to \mathbb{R}^n$ by $r(t) = (r_1(t), ..., r_n(t))$, $F : \mathbb{R}^n \to \mathbb{R}$, $F \circ r : \mathbb{R} \to \mathbb{R}$.

Then $D(F \circ r)|_t$ is a linear map $\mathbb{R} \to \mathbb{R}$ given by 1×1 matrix $[\frac{d}{dt}(F \circ r)]$, $Dr|_t : \mathbb{R} \to \mathbb{R}^n$ is given by

$$\begin{bmatrix} r_1'|_t \\ ... \\ r_n'|_t \end{bmatrix}$$

and $DF|_t : \mathbb{R}^n \to \mathbb{R}$ given by

$$[D_1 F|_{r(t)}, ..., D_n F|_{r(t)}]$$

So $D(F \circ r)$ is given by matrix multiplication:

$$\begin{aligned} D(F \circ r) &= \sum_{i=1}^{n} D_i F|_{r(t)} \cdot r_i'(t)|_t \\ &= \sum \frac{\partial F}{\partial x_i} r_i' \end{aligned}$$

Now back to the theorem. Since g is differentiable, $g(x+v) = g(x) + (L_1(v) + ||v||\alpha(v))(= e(v))$ at x where $L_1 = Dg|_x : \mathbb{R}^n \to \mathbb{R}^m$ and $\lim_{v \to 0} \alpha(v) = 0$.

Lemma. $\lim_{v \to 0} e(v) = 0$.

Proof. g is differentiable at $x \implies g$ is continuous at x. Done. □

Lemma. $\exists r > 0$, s.t. $\frac{||e(v)||}{||v||}$ is bounded on $B_r(0)$.

Proof.

$$\begin{aligned} \frac{||e(v)||}{||v||} &= ||L_1(\frac{v}{||v||}) + \alpha(v)|| \\ &\le ||L_1(\frac{v}{||v||})|| + ||\alpha(v)|| \end{aligned}$$

write $v' = \frac{v}{||v||}$, so $||v'|| = 1$.

L_1 is linear, so continuous. $\{v \in \mathbb{R}^n | ||v|| = 1\}$ is closed and bounded in $\mathbb{R}^n$, so by MVT, $\exists M$ s.t. $||L_1(v')|| \le M$ for all v' with $||v'|| = 1$.

$\lim_{v \to 0} \alpha(v) = 0$, so $\exists r$ s.t. $||\alpha(v)|| < 1$ for $v \in B_r(0)$.

Then for $v \in B_r(0), \frac{||e(v)||}{||v||} \le< M + 1$. □

Proof. (of Chain Rule)
f is differentiable at $g(x)$, so

$$f(g(x) + w) = f(g(x)) + LL_2(w) + ||w||B(w)$$

where $L_2 = Df|_{g(x)}$ and $\lim_{w \to 0} B(w) = 0$.

$$\begin{aligned}
f(g(x+v)) &= f(g(x)+e(v)) \\
&= fg(x) + L_2(e(v)) + ||e(v)||B(e(v)) \\
&= fg(x) + L_2(L_1(v)) + L_2(||v||\alpha(v)) + ||e(v)||B(e(v)) \\
&= fg(x) + (Df|_{g(x)} \cdot Dg|_x)(v) + ||v||E(v)
\end{aligned}$$

where

$$E(v) = L_2(\alpha(v)) + \frac{||e(v)||}{||v||}B(e(v))$$

we must show that $\lim_{v\to 0} E(v) = 0$ and then we are done.

We know $\lim_{v\to 0}\alpha(v) = 0$. L_2 is linear, hence continuous, so $\lim_{w\to 0} L_2(w) = L_2(0) = 0$. Thus $\lim_{v\to 0} L_2(\alpha(v)) = 0$.

By the above second lemma, $\exists r > 0$ s.t. $\frac{||e(v)||}{||v||}$ is bounded on $B_r(0)$. By the above first lemma $\lim_{v\to 0} e(v) = 0$.

We know $\lim_{w\to 0} B(w) = 0 \implies \lim_{v\to 0} B(e(v)) = 0$.

Then by last lecture's lemma,

$$\lim_{v\to 0}\frac{||e(v)||}{||v||}B(e(v)) = 0$$

So $\lim_{v\to 0} E(v) = 0$. □

Application of Chain Rule:
• The gradient.

Suppose $F : U \to \mathbb{R}$, where $U \subset \mathbb{R}^n$ is open. $DF|_x \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$.

Recall from LA that $\mathbb{R}^n \cong \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ by $v \to \phi_v : \phi_v(w) = v \cdot w$. That sends $\nabla F|_x$ to $DF|_x = [D_1F|_x, ..., D_nF|_x]$ where $\nabla F|_x = (D_1F|_x, ..., D_nF|_x)$ is the *gradient* of F at x.

So

$$D_vF|_x = DF|_x(v) = \nabla F|_x \cdot v$$

• Mean value inequality.

Definition. (Convex)
lüe

Proposition. Suppose $U \subset \mathbb{R}^n$ is open and convex, and $F : U \to \mathbb{R}$ is differentiable. If $||\nabla F|_x||_2 \le M \ \forall x \in U_1$. Then

$$|F(x_1) - F(x_0)| \le M||x_1 - x_0||_2$$

for all $x_0, x_1 \in U$.

Proof. Let $\gamma : [0,1] \to \mathbb{R}^n$ be given by

$$\gamma(t) = (1-t)x_0 + tx_1$$

then γ is differentiable and $\gamma'(t) = x_1 - x_0$.

Let $f(t) = F(\gamma(t))$. By Chain rule, f is differentiable and $f'(t) = \nabla F|_{\gamma(t)} \cdot \gamma'(t)$.

By Cauchy-Schwartz,

$$\begin{aligned} |f'(t)| &\leq ||\nabla F|_{\gamma(t)}|| \cdot ||x_1 - x_0|| \\ &\leq M||x_1 - x_0|| \end{aligned}$$

Apply 1-variable MVT to $f(t)$, we see that

$$|F(x_1) - F(x_0)| = |f(1) - f(0)| = |f'(c)|$$

for some $c \in [0,1]$

$$\leq M||x_1 - x_0||$$

□

Corollary. If $U \subset \mathbb{R}^n$ is open and convex, $F : U \to \mathbb{R}$ has $D_i F \equiv 0$ for $1 \leq i \leq n$. Then $F(x) \equiv c$ for some $c \in \mathbb{R}$.

Proof. $D_i F \equiv 0 \implies F$ differentiable $\implies$

$$|F(x_1) - F(x_0)| \leq 0 \cdot ||x_1 - x_0|| = 0$$

for all $x_1, x_0 \in U$. □

Remark. The hypothesis that U is convex is needed for the proposition, but can be weakened for the corollary.

Example. Suppose any 2 points x_1, x_0 in U can be joined by a differentiable path $\gamma : [0,1] \to U$ with $\gamma(0) = x_0$, $\gamma(1) = x_1$. Then the corollary still holds.

Proof. Consider $f(t) = F(\gamma(t))$. Then $f'(t) = DF|_{\gamma(t)}(\gamma'(t))$ by the chain rule. $D_i F \equiv 0 \implies DF \equiv 0 \implies f'(t) \equiv 0 \implies f(t)$ is constant. So $F(x_0) = f(0) = f(1) = F(x_1)$ for any x_0, x_1 in U. □

However, the corollary does not hold if U is disconnected. In fact it holds whenever $U \subset \mathbb{R}^n$ is open and connected.

4.4 Higher Derivatives

Q: If the derivative is a linear map, what is the 2nd derivative?
A: 2nd derivative is a symmetric bilinear form.

Suppose $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^m$ is differentiable.

Fix $v \in \mathbb{R}^n$ and define $g_v : U \to \mathbb{R}^m$ by

$$g_v(x) = Df|_x(v).$$

Definition. f is twice differentiable if all g_v are differentiable. If so, define $D^2f|_x(v, w) = D_{g_v}(w)$, i.e.

$$D^2f|_x : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$$

Example. $V = M_{n\times n}(\mathbb{R}) \cong \mathbb{R}^{n^2}$. $f : V \to V$ is given by $f(x) = x^2$. Then from previous section we know

$$g_A(X) = DF|_X(A) = XA + AX$$

Differentiate $g_A(X)$, get

$$\begin{aligned} g_A(X + B) &= A(X + B) + (X + B)A \\ &= (AX + XA) + AB + BA \\ &= g_A(X) + L_A(B) \end{aligned}$$

where $L_A(B) = AB + BA$ is linear in B.

So $D_{g_A}|_X(B) = AB + BA = D^2f|_X(A, B)$.

Note: $D^2f_X(A, B) = D^2f|_X(B, A)$.

Lemma. Suppose $f : U \to \mathbb{R}^m$ is twice differentiable, let $B(v, w) = D^2f|_x(v, w)$. Then B is a bilinear form.

Proof.

$$\begin{aligned} g_{v_1+\lambda v_2}(x) &= Df|_x(v_1 + \lambda v_2) \\ &= Df|_x(v_1) + \lambda Df|_x(v_2) = g_{v_1}(x) + \lambda g_{v_2}(x) \end{aligned}$$

So differentiating we get linearity in the first argument. Similarly we can prove linearity in the second argument. □

Suppose $F : U \to \mathbb{R}$ is differentiable. Then the partial derivatives $D_iF : U \to \mathbb{R}$ are all defined.

Notation. Write $D_{ij}F = D_i(D_jF)$ if it exists.

Definition. $C^2(U) = \{F : U \to \mathbb{R}|$ all 1st and 2nd order partial derivatives of F are defined and continuous $\}$.

Proposition. If $F \in C^2(U)$, then F is twice differentiable and

$$D^2F|_x(v, w) = \sum_{1\le i,j\le n} v_i w_j D_{ji}F(x)$$

Proof. Let $G{\cdot}i = D_iF$. Then all 1st order partial derivatives of G_i are defined and continuous so G_i is differentiable.

Then for $v \in \mathbb{R}^n$, $G_{v(x)} = DF|_x(v) = \sum_{1\le i\le n} v_iD_iF|_x = \sum v_iG(x)$.

So for a fixed value of v, $G_v(x)$ is a linear combination of the G_is. Since all of them are differentiable, G_v is differentiable. So F is twice differentiable, and $D^2F|_x(v, w) = DG_v|_x(w) = \sum_{1\le j\le n} w_jD_jG_v|_x = \sum_{1\le j,i\le n} v_iw_jD_{ji}F|_x$. □

Now $D_j(G_v) = D_j(\sum_{i=1}^n v_i G_i) = \sum_{i=1}^n v_i D_j G_i = \sum_{i=1}^n v_i D_{ji} F$.

Equivalently, $D^2F|_x(v,w) = W^t B v$ where $B = [D_{ij}F|_x]$ is the *Hessian* matrix of 2nd order partial derivatives.

Example. $F(x,y) = x^2y^3$. Then

$$B = \begin{pmatrix} 2y^3 & 6xy^2 \\ 6xy^2 & 6x^2y \end{pmatrix}$$

Recall that if $U \subset \mathbb{R}^n$ is open and $F \in C^2(U)$, then $D^2F|_x : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is bilinear and given by

$$D^2F|_x(v,w) = \sum_{1 \le i,j \le n} v_i w_j D_{ji} F|_x = w^T H(x) v$$

where $H(x) = [D_{ji}F|_x]$ is the *Hessian matrix.*

Theorem. (symmetry of mixed partials)
Suppose $U \subset \mathbb{R}^2$ is open and $F \in C^2(U)$. Then $D_{12}F = D_{21}F$.

Note that it's not enough for the partial derivatives to be defined. They must be continuous or the theorem may fail (see example sheet).

Lemma.
$$D_{12}F|_{(x_0,y_0)} = \lim_{v \to 0} \frac{S(v)}{v^2}$$
where

$$S(v) = F(x_0+v, y_0+v) - F(x_0+v, y_0) - F(x_0, y_0+v) + F(x_0, y_0)$$

Proof. Since U is open, there exists $\varepsilon > 0$ s.t. $B_\varepsilon((x_0,y_0), ||\cdot||_\infty) \subset U$.

From now on, assume $|v| < \varepsilon/2$.

Consider $A(y) = F(x_0+v, y) - F(x_0, y)$. Fix v with $|v| < \varepsilon/2$. Then A is differentiable on $(y_0 - \varepsilon/2, y_0 + \varepsilon/2)$, and

$$A'(y) = D_2F(x_0+v, y) - D_2F(x_0, y)$$

Note that $S(v) = A(y_0+v) - A(y_0)$. So by MVT,

$$S(v) = vA'(y^*)$$

for some $y^* \in [y_0, y_0+v]$

$$\begin{aligned} &= v[D_2F(x_0+v, y^*) - D_2F(x_0, y^*)] \\ &= v[B(x_0+v) - B(x_0)] \end{aligned}$$

where $B(x) = D_2F(x, y^*)$.

B is differentiable on $(x_0 - \varepsilon/2, x_0 + \varepsilon/2)$, aget$B'(x) = D_{12}F(x, y^*)$. Applying MVT to B we get

$$S(v) = v^2 B'(x^*) = v^2 D_{12}F(x^*, y^*)$$

for some $x^* \in [x_0, x_0 + v]$. Note that we have

$$||(x^*(v), y^*(v)) - (x_0, y_0)||_\infty \leq ||v||_\infty$$

So

$$\lim_{v \to 0} (x^*(v), y^*(v)) = (x_0, y_0)$$

Then

$$\begin{aligned} \lim_{v \to 0} \frac{S(v)}{v^2} &= \lim_{v \to 0} D_{12}F(x^*(v), y^*(v)) \\ &= D_{12}F(x_0, y_0) \end{aligned}$$

Since D_{12} is continuous. □

Proof. (of theorem)
The expression $S(v)$ is symmetric under interchanging roles of x and y. Similar arguments as in the above proof shows

$$D_{21}F(x_0, y_0) = \lim_{v \to 0} \frac{S(v)}{v^2} = D_{12}F(x_0, y_0)$$

So they are equal.

□

Corollary. If $U \subset \mathbb{R}^n$ is open, $G \in C^2(U)$, then $D_{ij}G = D_{ji}G$ for all $1 \leq i, j \leq n$.

Proof. Apply the theorem to $F(z_1, z_2 = G(x_1, x_2, ...z_1(i), ..., z_2(j), ..., x_n)$. □

In other words, if $G \in C^2(U)$, the Hessian matrix $H = [D_{ji}G|_x]$ is symmetric: $H^T = H$.

Corollary. $D^2G|_x$ is symmetric. i.e. $D^2G|_x(v, w) = D^2G|_x(w, v)$.

Proof. $D^2G|_x(v, w) = w^T H v$ is a 1×1 matrix, so symmetric. Take its transpose and we get the other side of the equation. □

Higher derivatives are defined inductively: If $F : U \to \mathbb{R}$ is $(k - 1)$ times differentiable, then

$$D^k F|_x(v_1, ..., v_k) = DG|_x(v_k)$$

(if exists) where

$$G(x) = D^{k-1}F|_x(v_1, ..., v_{k-1})$$

The same proof as for $k = 2$ shows that if $F \in C^k(U)$ then F is k times differentiable, and

$$D^k F|_x(v_1, ..., v_k) = \sum_{\alpha \in \{1,...,n\}^k} v^\alpha D_\alpha F|_x$$

where

$$v^\alpha = \prod_{i=1}^{k} v_{i,\alpha_i}$$

where

$$v_i = (v_{i,1}, ..., v_{i,n})$$
$$\alpha = (\alpha_1, ..., \alpha_k)$$

If we let $A(v_1, ..., v_k) = D^k F|_\alpha(v_1, ..., v_k)$, then A is
1) Symmetric: $A(v_1, ..., v_k) = A(v_i, ..., v_{i_k})$ where $(i_1, ..., i_k)$ is any permutation of $(1, ..., k)$;
2) Multilinear: $(v_1 + \lambda v_1', v_2, ..., v_k) = A(v_1, v_2, ..., v_k) + \lambda A(v_1', v_2, ..., v_k)$.

Proposition. Suppose $F \in C^k(U)$ and define

$$f(t) = F(x_0 + tv)$$

for $x_0 \in U$. Since U is open, f is defined on $(-\varepsilon, \varepsilon)$.

Then f is $k-$times differentiable and

$$f^k(t) = D^k F|_{x_0+tv}(v, v, ..., v)(\text{ k times})$$

Proof. Recall that if $G \in C^1(U)$ and $g = G(x_0 + tv)$ then $g'(t) = D_v G|_{x_0+tv} = DG|_{x_0+tv}(v)$.

The proof is by induction on k:

$k = 1$ is exactly the above equation applied to $G = F$.

For the general case, suppose the proposition holds for $k - 1$. Then let

$$\begin{aligned} h(t) &= f^{(k-1)}(t) \\ &= D^{k-1} F|_{x_0+tv}(v, ..., v) \\ &= H(x_0 + tv) \end{aligned}$$

where $H(x) = D^{k-1} F|_x(v, ..., v)$.

Apply the above equation to $G = H$, get

$$f^k(t) = h'(t) = DH|_{x_0+tv}(v) = D^k F|_{x_0+tv}(v, ..., v) \text{ (k times)}$$

□

Theorem. (Taylor's Theorem)
If $F \in C^k(\mathbb{R}^n)$, then

$$F(x_0 + v) = \sum_{i=0}^{k-1} \frac{1}{i!} D^i F|_{x_0}(v, ..., v) + \frac{1}{k!} D^k F|_{x_0+tv}(v, ..., v)$$

for some $t \in [0, 1]$.

Proof. Consider $f(t) = F(x_0 + tv)$ as above. Then by Taylor's theorem in 1 variable, we have

$$f(1) = \sum_{i=0}^{k-1} \frac{1}{i!} f^{(i)}(0) \cdot 1^i + \frac{1}{k!} f^{(k)}(t) 1^k$$

for some $t \in [0, 1]$, i.e.

$$F(x + v) = \sum_{i=0}^{k-1} \frac{1}{i!} D^i F|_{x_0}(v, ..., v) + \frac{1}{k!} D^k|_{x_0+tv}(v, ..., v)$$

□

Remark. $D^k F|_{x_0}(v, ..., v)$ is a degree k polynomial in the coefficients of v s.t. all the k^{th} order partial derivatives agree with k^{th} order partial derivatives of F at x_0.

5 Metric spaces

5.1 Basics

Definition. lüe

Example. lüe

Definition. (open and closed sets)
lüe

5.2 Lipschitz Maps

Suppose (X, d_X) and (Y, d_Y) are metric spaces.

Definition. $f : X \to Y$ is $k-$Lipschitz ($k \in \mathbb{R}^+$) if $d_Y(f(x_1), f(x_2)) \leq k d_X(x_1, x_2)$ for all $x_1, x_2 \in X$.
f is Lipschitz if it's $k-$Lipschitz for some $k \in \mathbb{R}^+$.

Proposition. f is Lipschitz implies that f is uniformly continuous.

Proof. Suppose f is $k-$Lipschitz. If $d(x_1, x_2) < \varepsilon/k$, then $d(f(x_1), f(x_2)) < \varepsilon$. □

Proposition. Suppose $U \subset \mathbb{R}^n$ is open, $F \in C^1(U)$, and $k = \bar{B}_r(\mathbf{x}_0) \subset U$. Then $F|_k$ is Lipschitz.

Proof. F is C^1, so the map

$$\begin{aligned} U \to \mathbb{R}^n & \quad \to \mathbb{R} \\ \mathbf{x} \to \nabla F|_{\mathbf{x}} & \quad \to ||\nabla F|_{\mathbf{x}}|| \end{aligned}$$

is continuous.

$k = \bar{B}_r(\mathbf{x}_0)$ is a closed and bounded subset of $\mathbb{R}^n$. By the Maximum Value Theorem, $\exists M \in \mathbb{R}$ s.t. $||\nabla F|_{\mathbf{x}}|| \leq M$ for all $\mathbf{x} \in k$. $k = \bar{B}_r(\mathbf{x}_0)$ is convex, so by Mean Value Inequality,

$$|F(\mathbf{x}_1) - F(\mathbf{x}_2)| \leq M ||\mathbf{x}_1 - \mathbf{x}_2||_2$$

i.e.

$$d(F(\mathbf{x}_1), F(\mathbf{x}_2)) \leq M d(\mathbf{x}_1, \mathbf{x}_2)$$

for $\mathbf{x}_1, \mathbf{x}_2 \in k$. □

Proposition. If $f : Y \to Z$ is k_1-Lipschitz, $g : X \to Y$ is k_2-Lipschitz, then $f \circ g : X \to Z$ is $k_1 k_2-$Lipschitz.

Proof.

$$\begin{aligned} d_2\left(f\left(g\left(x_1\right)\right), f\left(g\left(x_2\right)\right)\right) &\leq k_1 d_y\left(g\left(x_1\right), g\left(x_2\right)\right) \\ &\leq k_1 k_2 d_x\left(x_1, x_2\right) \end{aligned}$$

So The composition of Lipschitz maps is Lipschitz. □

Proposition. If $||\cdot||$ and $||\cdot||'$ are two norms on a vector space V, then $||\cdot||$ is Lipschitz equivalent to $||\cdot||'$ if and only if both the identity maps from V equipped with one norm to the other norm are Lipschitz.

Definition. Suppose V, W are finite dimensional normed vector spaces.
If $L \in \mathcal{L}(V, W)$, the operator norm

$$||L||_{op} = \sup_{\mathbf{v}\in V, \mathbf{v}\neq 0} \frac{||L(\mathbf{v})||_W}{||\mathbf{v}||_V} = \max_{||\mathbf{v}||=1} ||L(\mathbf{v})||_W.$$

The maximum exists since S^1 is closed and bounded in $V = \mathbb{R}^n$.

Lemma. $||\cdot||_{op}$ is a norm on $\mathcal{L}(V, W)$.

Proof. Omitted. □

Proposition. If $||L_1||_{op} = k$, then L_1 is $k-$Lipschitz.

Proof.

$$\begin{aligned} ||L(\mathbf{v}_1) - L(\mathbf{v}_2)|| &= ||L(\mathbf{v}_1\mathbf{v}_2)|| \\ &\leq k||\mathbf{v}_1 - \mathbf{v}_2|| \end{aligned}$$

Since $||L||_{op} = k$. □

5.3 Contraction maps

Suppose X is a metric space and $f : X \to X$.

Definition. $x \in X$ is a *fixed point* of f if $f(x) = x$.

Definition. $f^n = f \circ f \circ ... \circ f$ (n times) :$X \to X$it the composition of f with itself n times.

If f is $k-$Lipschitz, then f^n is k^n-Lipschitz.

Definition. $f : X \to X$ is a *contraction map* if f is $k-$Lipschitz for some $k < 1$.

Theorem. Suppose X is a complete metric space, $f : X \to X$ is a contraction map. Then f has a unique fixed point.

Proof. Suppose f is $k-$Lipschitz for some $k < 1$.

Lemma. If $x \in X$, then $d(x, f^n(x)) \leq \frac{1}{1-k} d(x, f(x))$ regardless of n.

Proof. f^n is k^n Lipschitz, so

$$\begin{aligned} d\left(f^n(x), f^{(n+1)}(x)\right) &= d\left(f^n(x), f^n(f(x))\right) \\ &\leq k^n d(x, f(x)) \end{aligned}$$

So

$$\begin{aligned} d(x, f^n(x)) &\leq d(x, f(x)) + ... + d\left(f^{n-1}(x), f^n(x)\right) \\ &\leq d(x, f(x)) + kd(x, f(x)) + ... + k^{n-1} d(x, f(x)) \\ &= \frac{1-k^n}{1-k} d(x, f(x)) \\ &\leq \frac{1}{1-k} d(x, f(x)) \end{aligned}$$

□

Proof of Theorem:
Pick $x \in X$ and consider $(f^n(x))$.

This sequence is Cauchy: if $m \geq n$, then

$$\begin{aligned} d(f^m(x), f^n(x)) &= d\left(f^n(x), f^n\left(f^{m-n}(x)\right)\right) \\ &\leq k^n d\left(x, f^{m-n}(x)\right) \\ &\leq \frac{k^n}{1-k} d(x, f(x)) \end{aligned}$$

We know $k < 1$, so

$$\lim_{n \to \infty} k^n \left(\frac{d(x, f(x))}{1-k}\right) = 0$$

So pick N s.t. the above is less than ε for all $n \geq N$. Then if $m \geq n \geq N$,

$$d(f^n(x), f^m(x)) \leq \frac{k^n}{1-k} d(x, f(x)) < \varepsilon$$

So $(f^n(x))$ is Cauchy. So it converges to some x^*.

We claim that $f(x^*) = x^*$: since f is Lipschitz, f is continuous, and $f^n(x) \to x^*$, so $f(f^n(x)) \to f(x^*)$. But $f^{n+1}(x) \to x^*$. So $f(x^*) = x^*$.

We also claim that x^* is the only fixed point: Suppose $f(y) = y$. Then $d(f(x^*), f(y)) = d(x^*, y)$. But $d(f(x^*), f(y)) \leq kd(x^*, y)$, since f is a contraction where $k < 1$, this can only happen if $d(x^*, y) = 0$, i.e $x^* = y$. □

6 Solving Equations

Problem: Suppose $U \subset \mathbb{R}^n$ is open. $f : U \to \mathbb{R}^m$ is C^1 and $f(x_0) = y_0$. Can we solve $f(x) = y$ for y close to y_0?

If so, what does the set of x close to x_0 solution look like?

There are three cases:
a) $n < m$. For 'most' $y \in \mathbb{R}^m$, there is no solution. Idea: $\dim(U) \leq n < m$.

b) $m = n$. If y is sufficiently close to y_0 and $Df|_{x_0}$ is an isomorphism, then there is a unique solution near x_0 (inverse function theorem).

c) $m < n$. If $Df|_{x_0}$ is surjective and y is close to y_0, set of solutions near x_0 looks like $B_\varepsilon(0) \subset \mathbb{R}^{n-m}$ (implicit function theorem).

We'll prove (b) and use it to prove (c).

6.1 Newton's method

$n = 1$: solve $f(x) = y^* = y$.

Approximate f by graph of it's tangent line at $(x_0, f(x_0))$.

$$g(x) = f(x_0) + f'(x_0)(x - x_0)$$

solve $g(x_1) = y^*$:

$$x_1 = x_0 + \frac{y^* - f(x_0)}{f'(x_0)}$$

Now repeat:

$$x_2 = x_1 + \frac{y^* - f(x_1)}{f'(x_1)}$$

and etc. Hope that $(x_n) \to x^*$ with $f(x^*) = y^*$.

General case: $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$: approximate $f(x)$ near $x = x_0$ by

$$g(x) = f(x_0) + Df|_{x_0}(x - x_0)$$

solve equation $g(x_1) = y$: $x_1 = x_0 + (Df|_{x_0})^{-1}(y - f(x_0))$.

Repeat: $x_2 = x_1 + (Df|_{x_1})^{-1}(y - f(x_1))$ etc.

Equivalently: define $n_y(x) = x + (Df|_x)^{-1}(y - f(x))$. Then

$$(x_k) = (n_y^k(x_0))$$

(the k^{th} iterate of n_y).

If x is a fixed point of n_y, then

$$\begin{aligned} x &= x + (Df|_x)^{-1}(y - f(x)) \\ \implies 0 &= (Df|_x)^{-1}(y - f(x)) \\ \implies 0 &= y - f(x) \\ \implies f(x) &= y \end{aligned}$$

so we have a solution.

So if we knew n_y was a contraction map, we would get a solution.

Problem: This only makes sense if $Df|_x$ is invertible. Analyzing $(Df|_x)^{-1}$ term is painful.

Modified Newton's method:

Suppose $f : U \to \mathbb{R}^n$ is C^1, $f(x_0) = y_0$ and $Df|_{x_0} = A$ is invertible where $A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map.

Approximate $Df|_x$ by $Df|_{x_0} = A$, i.e. consider

$$N_y(x) = x + A^{-1}(y - f(x))$$

If $N_y(x) = x$, then $f(x) = y$ so we found a solution.

Is N_y a contraction map when x is close to x_0?

Compute

$$\begin{aligned} N_y(x) - N_y(x') &= x + A^{-1}(y + f(x)) - (x' + A^{-1}(y - f(x')) \\ &= x - x' + A^{-1}(f(x') - f(x)) \\ &= A^{-1}(A(x) - f(x) - (A(x') - f(x'))) \\ &= A^{-1}(h(x) - h(x')) \end{aligned}$$

where $h(x) = A(x) - f(x)$.

Notice:

$$\begin{aligned} Dh|_x &= DA|_x - Df|_x \\ &= A|_x - Df|_x \\ &= Df|_{x_0} - Df|_x \end{aligned}$$

C^1 maps: $Df = U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) = M_{n\times n}(\mathbb{R}) = \mathbb{R}^{n^2}$.

f is C^1 if Df is continuous.

Note: Since all norms on $\mathbb{R}^{n^2}$ are equivalent, we can use whatever norm on $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ we like.

For applications: use operator norm $||\cdot||_{op}$.

Lemma. If $||Dh|_x||_{op} < M$ for all $x \in B_r(x_0)$, then $||h(x)-h(x')||_2 \leq M\sqrt{n}||x-x'||$ for $x, x' \in B_r(x_0)$ (n is the dimension of space).

Proof. let h_i be the i^{th} component of h. Then

$$||Dh_i|_x|| = ||Dh|_x(e_i)|| \leq M \cdot ||e_i||_2 = M$$

So by Mean value inequality, $|h_i(x) - h_i(x')| \leq M \cdot ||x - x'||_2$. So

$$||h(x) - h(x') \leq \sqrt{n}M \cdot ||x - x'||$$

□

Proposition. Given $\varepsilon > 0$, there exists $\delta > 0$ s.t. N_y is $\varepsilon-$lipschitz on $B_\delta(x_0)$.

Proof. f is C^1, so choose $\delta > 0$ s.t.

$$||Dh|_x|| = ||Df|_x - Df|_{x_0}||_{op} \leq \frac{\varepsilon}{||A^{-1}||_{op} \cdot \sqrt{n}}$$

for $x \in B_\delta(x_0)$, so

$$\begin{aligned} ||N_y(x) - N_y(x')|| &= ||A^{-1}(h(x) - h(x'))|| \\ &\leq ||A^{-1}||_{op}||h(x) - h(x')|| \\ &\leq ||A^{-1}||_{op} \cdot \sqrt{n}(\varepsilon/\sqrt{n} \cdot ||A^{-1}||_{op}) \cdot (||x - x'||) \end{aligned}$$

□

6.2 The Inverse Function Theorem (See alternative notes)

Let $U \subset \mathbb{R}^n$ be open, $f(\mathbf{x}_0) = \mathbf{y}_0$, and $f : U \to \mathbb{R}^n$ is C^1, $A = Df|_{\mathbf{x}_0}$ is invertible.

Let $a = ||A^{-1}||_{op}$, so that

$$||A^{-1}(v)|| \leq a||\mathbf{v}||$$

for all $v \in \mathbb{R}^n$.

Lemma. $\exists n > 0$ s.t. $Df|_{\mathbf{x}}$ is invertible for all $\mathbf{x} \in B_n(\mathbf{x}_0)$.

Proof. $f : U \to \mathbb{R}^n$ is C^1, so the map

$$\begin{aligned} \alpha : U &\to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) && \to \mathbb{R} \\ \mathbf{x} &\to Df|_{\mathbf{x}} && \to \det(Df|_{\mathbf{x}}) \end{aligned}$$

is continuous.

$\mathbb{R} - \{0\}$ is open in $\mathbb{R}$, so $\alpha^{-1}(\mathbb{R} - 0)$ is open in U.

$\alpha^{-1}(\mathbb{R} - \{0\}) = \{\mathbf{x} \in U | Df_{\mathbf{x}} \text{ is invertible}\}$. $\mathbf{x}_0 \in \alpha^{-1}(\mathbb{R} - \{0\})$, so $\exists n > 0$ s.t. $B_n(\mathbf{x}_0) \subset \alpha^{-1}(\mathbb{R} = 0)$. □

Consider $N_{\mathbf{y}}(\mathbf{x}) = \mathbf{x} + A^{-1}(\mathbf{y} - f(\mathbf{x}))$ (Modified Newton's Method).

Fix r_0 s.t. $0 < r_0 < n$ and $N_{\mathbf{y}}$ is $\frac{1}{2}$-Lipschitz on $\bar{B}_{r_0}(\mathbf{x}_0)$. By the lemma, $Df|_{\mathbf{x}}$ is invertible for $\mathbf{x} \in \bar{B}_{r_0}(x_0)$.

We want N_y to be a contraction map.

Problem: $N_y(B_{r_0}(\mathbf{x}_0))$ need not be in the region where N_y is contracting.

Solution: require $\mathbf{y}$ to be close to y_0.

Proposition. (2) Let $r(\mathbf{y}) = 2a||\mathbf{y} - \mathbf{y}_0||$. If $r(\mathbf{y}) \leq r_0$, then $N_{\mathbf{y}} : \bar{B}_{r(\mathbf{y})}(\mathbf{x}_0) \to \bar{B}_{r(\mathbf{y})}(\mathbf{x}_0)$.

Proof. Suppose $\mathbf{x} \in B_{r(\mathbf{y})}(\mathbf{x}_0)$. Then

$$\begin{aligned}
||N_{\mathbf{y}}(\mathbf{x}_0)|| &\leq N_{\mathbf{y}}(\mathbf{x}) - N_{\mathbf{y}}(\mathbf{x}_0)|| + ||N_y(\mathbf{x}_0) - \mathbf{x}_0|| \\
&\leq \frac{1}{2}||\mathbf{x} - \mathbf{x}_0|| + ||A^{-1}(\mathbf{y} - \mathbf{y}_0)|| \\
&\leq \frac{1}{2}r(\mathbf{y}) + a||\mathbf{y} - \mathbf{y}_0|| \\
&= \frac{1}{2}r(\mathbf{y}) + \frac{1}{2}r(\mathbf{y}) \\
&= r(\mathbf{y})
\end{aligned}$$

So $N_{\mathbf{y}}(\mathbf{x}_0) \in B_{r(\mathbf{y})}(\mathbf{x}_0)$. □

Proposition. (3) Suppose $r \leq r_0$. If $\mathbf{y} \in B_{\frac{r}{2a}}(\mathbf{y}_0)$, then there is a unique $\mathbf{x} \in B_r(\mathbf{x}_0)$ s.t. $f(\mathbf{x}) = \mathbf{y}$.

Proof. $f(\mathbf{x}) = \mathbf{y} \iff N_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$.

$N_{\mathbf{y}}$: If $\mathbf{y} \in \bar{B}_{r/2a}(\mathbf{y}_0)$, then $r(\mathbf{y}) = r$, so by the previous proposition,

$$N_{\mathbf{y}} : \bar{B}_r(\mathbf{x}_0) \to \bar{B}_r(\mathbf{x}_0)$$

$r \leq r_0$, so $N_{\mathbf{y}}$ is $\frac{1}{2}-$ Lipschitz on $\bar{B}_r(\mathbf{x}_0)$, i.e. $N_{\mathbf{y}} : \bar{B}_r(\mathbf{x}_0) \to \bar{B}_r(\mathbf{x}_0)$ is a contraction.

$\bar{B}_r(\mathbf{x}_0)$ is a closed subset of $\mathbb{R}^n$, which is complete, so $\bar{B}_r(x_0)$ is complete. Thus there is a unique $\mathbf{x} \in \mathbf{B}_r(\mathbf{x}_0)$ such that $N_{\mathbf{y}}(\mathbf{x}) = \mathbf{x}$, i.e. there exists a unique $\mathbf{x} \in \bar{B}_r(\mathbf{x}_0)$ s.t. $f(\mathbf{x}) = \mathbf{y}$. □

Note: $r(\mathbf{y}) = 2a||\mathbf{y} - \mathbf{y}_0||$, so $r(\mathbf{y}) \leq r \implies y \in \bar{B}_{r/2a}(\mathbf{y}_0)$.

Remark. If $y \in \bar{B}_{r/2a}(\mathbf{y}_0)$ for $r < r_0$, then it's in $\bar{B}_{r_0/2a}(\mathbf{y}_0)$, so the proposition implies that there is a unique $\mathbf{x} \in \bar{B}_{r_0/2a}(\mathbf{y}_0)$ with $f(\mathbf{x}) = y$ and $\mathbf{x} \in B_r(\mathbf{x}_0)$.

Proposition. (4) There are open sets $V \subset U$, $\mathbf{x}_0 \in V$, $W \subset \mathbb{R}^n$, $\mathbf{y}_0 \in W$ s.t. $f|_V : V \to W$ bijectively.

Proof. Take $W = B_{r_0/4a}(\mathbf{y}_0)$. f is continuous, so $f^{-1}(W)$ is open. Take

$$]V = f^{-1}(W) \cap B_{r_0}(\mathbf{x}_0)$$

which is open.

Given $\mathbf{y} \in W$, there exists $x \in \bar{B}_{r_0/2}(\mathbf{x}_0)$ with $f(\mathbf{x}) = y$ by the previous proposition. Moreover, this $\mathbf{x}$ is the unique $\mathbf{x}$ in that open ball. So $\mathbf{x} \in V = f^{-1}(W) \cap B_r(\mathbf{x}_0)$ and is the unique such element. □

Proposition. (5) Let $g : W \to V$ be the inverse of f. Then g is continuous at $\mathbf{y}_0$.

Proof. If $\mathbf{y} \in \bar{B}_{r/2a}(\mathbf{y}_0)$, then $g(\mathbf{y}) \in \bar{B}_r(\mathbf{x}_0)$ by proposition 3, i.e. if $||\mathbf{y} - \mathbf{y}_0|| \leq \delta$, then $||g(\mathbf{y}) - g(\mathbf{y}_0)|| \leq 2a\delta$. So g is continuous at $\mathbf{y}_0$. □

Proposition. (6) g is differentiable at $\mathbf{y}_0$ and $Dg|_{y_0} = A^{-1}$.

Proof. Note that $g(\mathbf{y})$ satisfies $N_{\mathbf{y}}(g(\mathbf{y})) = g(\mathbf{y})$, so if $\mathbf{y} \in \bar{B}_{r/2a}(\mathbf{y}_0)$, $g(\mathbf{y}) \in N_y\left(\bar{B}_{r(\mathbf{y})}(\mathbf{x}_0)\right)$.

Now $N_y : \bar{B}_{r(\mathbf{y})}(\mathbf{x}_0) \to \bar{B}_{r(\mathbf{y})}(\mathbf{x}_0)$ is $\varepsilon(r(\mathbf{y}))-$ Lipschitz, by proposition 1 where $\varepsilon(r(\mathbf{y})) \to 0$ as $r(\mathbf{y}) \to 0$.

So $N_y\left(B_{r(\mathbf{y})}(\mathbf{x}_0)\right) \subset B_{\varepsilon(r(\mathbf{y}))\cdot r(\mathbf{y})}(N_y(\mathbf{x}_0))$, i.e. $g(y) = N_{\mathbf{y}}(\mathbf{x}_0) + E(\mathbf{y})$, where

$$\begin{aligned}
||E(\mathbf{y})|| &\leq \varepsilon(r(\mathbf{y})) \cdot r(\mathbf{y}) \\
&= \varepsilon(r(\mathbf{y}))2a||\mathbf{y} - \mathbf{y}_0|| \\
&= \mathbf{x}_0 + A^{-1}(\mathbf{y} - \mathbf{y}_0) + E(\mathbf{y})
\end{aligned}$$

where

$$\frac{||E(\mathbf{y})||}{||\mathbf{y} - \mathbf{y}_0||} \leq 2a\varepsilon(r(\mathbf{y}))$$

and $\varepsilon(r(\mathbf{y})) \to 0$ as $||\mathbf{y} - \mathbf{y}_0|| \to 0$. So the above equation says that g is differentiable at $\mathbf{y}_0$, and $Dg|_{\mathbf{y}_0} = A^{-1}$. □

Definition. Suppose $V, W \subset \mathbb{R}^n$ are open. $f : V \to W$ is a *diffeomorphism* if
- f is bijective;
- f and f^{-1} are both C^1.

Theorem. (Inverse function theorem) Suppose $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^n$ is C^1 with $f(\mathbf{x}_0) = \mathbf{y}_0$ and $Df|_{\mathbf{x}_0}$ is invertible. Then there are open subsets $V \subset U$ and $\mathbf{x}_0 \in V$, $W \subset R^n$ and $\mathbf{y}_0 \in W$ s.t. $f|_V : V \to W$ is a diffeomorphism.

Proof. Let V and W be as in Proposition 4. Then $f : V \to W$ bijectively. Let $g = f^{-1} : W \to V$. Must show g is C^1.
We know $V \subset B_{r_0}(\mathbf{x}_0)$ where $Df|_{\mathbf{x}}$ is invertible for all $\mathbf{x} \in B_r(\mathbf{x}_0)$ (hypothesis of this subsection).
Apply proposition 6 with $\mathbf{x}$ in place of $\mathbf{x}_0$, we see that g is differentiable at $\mathbf{x}$, and $Dg|_{\mathbf{x}} = (Df|_{\mathbf{x}})^{-1}$.

g is differentiable implies that g is continuous. To see g is C^1, note that $Dg : W \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is a composition

$$\begin{array}{cccc} W \to & V \to & \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \to & \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n) \\ \mathbf{y} \to & g(\mathbf{y}) & & \\ & x \to & Df|_{\mathbf{x}} & \\ & & A \to & A^{-1} \end{array}$$

$\square$

Reformulation: Local change of coordinates:
Suppose $U_i = f_i(x_1, ..., x_n)$ for $1 \le i \le n$.
Consider $J = \left(\frac{\partial u_i}{\partial x_j}\right) = (D_j f_i) =$ matrix representing Df.
If $\det(J|_{\mathbf{x}_0}) \ne 0$ (i.e. $Df|_{\mathbf{x}_0}$) then we can use $(U_1, ..., u_n)$ as a local system of coordinates near $\mathbf{x}_0$.
i.e. we can solve for x_j's in terms of u_i's:

$$x_j = g_j(u_1, ..., u_n).$$

Example. Polar coordinates: $x = r\cos\theta$, $y = r\sin\theta$,

$$J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

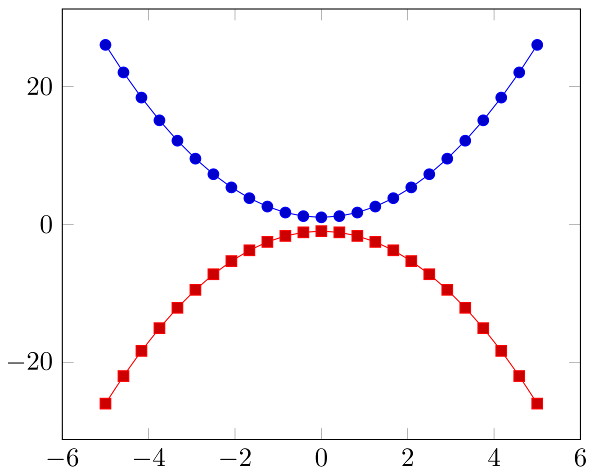
$\det J = r$ i.e. there's a good change of coordinates between polar and rectangular coordinates except when $r = 0$.

6.3 The implicit function theorem

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be C^1 $(n \ge m)$, $F(\mathbf{x}_0) = \mathbf{y}_0$.

Problem: Describe $F^{-1}(\mathbf{y}_0)$ near $\mathbf{x}_0$.

Example. $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x, y) = x^2 - y^2$.



Notation. $B_\varepsilon^k = B_\varepsilon(\mathbf{0}) \subset \mathbb{R}^k = k-$dimensional open ball.

Theorem. Suppose $F : \mathbb{R}^n \to \mathbb{R}^m$ is C^1, $f(\mathbf{x}_0) = \mathbf{y}_0$, and $DF|_{\mathbf{x}_0}$ is surjective. Then there's an open set $V \subset \mathbb{R}^n$, $\mathbf{x}_0 \in V$, and a C^1 map $G : B_\varepsilon^{n-m} \to \mathbb{R}^n$ such that:
1) $F^{-1}(\mathbf{y}_0) \cap V = \operatorname{im} G$;
2) G is injective;
3) $DG|_{\mathbf{z}}$ is injective for all $\mathbf{z} \in B_\varepsilon^{n-m}$.
i.e. if $n - m = 1$, $B'_\varepsilon = (-\varepsilon, \varepsilon)$, $F^{-1}(\mathbf{y}_0) \cap V$ is a paramterized curve;
if $n - m = 2$ then this is a parametrized surface.
For general $n - m$, we call this is a parametrized $(n - m)$-manifold.

Example. $F(x, y) = x^2 - y^2$, $DF|_{(x,y)} = [2x, -2y]$ is surjective $\iff$ $(x, y) \neq (0, 0)$.

Definition. $F^{-1}(\mathbf{y}_0)$ is *smooth* at $\mathbf{x}_0$ if $DF|_{\mathbf{x}_0}$ is surjective, *singular* at $\mathbf{x}_0$ otherwise.
$F^{-1}(\mathbf{y}_0)$ is smooth if it is smooth at all $\mathbf{x} \in F^{-1}(\mathbf{y}_0)$.

Proof of theorem:

Proof. $DF|_{\mathbf{x}_0} : \mathbb{R}^n \to \mathbb{R}^m$ is surjective. So $K := \ker DF|_{\mathbf{x}_0}$ has dimension $(n - m)$. Choose any $\pi \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ with $\pi(K) = \mathbb{R}^{n-m}$. Define $f : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m} = \mathbb{R}^n$ by $f(\mathbf{x}) = (F(\mathbf{x}), \pi(\mathbf{x}))$. So $Df : \mathbb{R}^n \to \mathbb{R}^m \oplus \mathbb{R}^{n-m}$.
$Df_{\mathbf{x}_0}(\mathbf{v}) = (DF|_{\mathbf{x}_0}(\mathbf{v}), \pi(\mathbf{v}))$ since $D\pi = \pi$.
Claim: $Df|_{\mathbf{x}_0}$ is an isomorphism: If $Df|_{\mathbf{x}_0}(\mathbf{v}) = \mathbf{0}$, then $DF|_{\mathbf{x}_0}(\mathbf{v}) = 0 \implies \mathbf{v} \in K$. But $\pi : K \to \mathbb{R}^{n-m}$ is an isomorphism, so $\pi(\mathbf{v}) = 0 \implies \mathbf{v} = 0$. So $\ker Df|_{\mathbf{x}_0} = \{\mathbf{0}\} \implies Df|_{\mathbf{x}_0}$ is an isomorphism.

By the inverse function theorem, there exists $V \subset \mathbb{R}^n$, $\mathbf{x}_0 \in V$, $W \subset \mathbb{R}^m \times \mathbb{R}^{n-m}$, $(\mathbf{y}_0, \pi(\mathbf{x}_0)) \in W$, s.t. $f : V \to W$ is an diffeomorphism. Let $g = f^{-1} : W \to V$

Then $F^{-1}(\mathbf{y}_0) \cap V = f^{-1}(\mathbf{y}_0 \times \mathbb{R}^{n-m}) \cap V$, so $g(\mathbf{y}_0 \times \mathbb{R}^{n-m}) \cap W = F^{-1}(\mathbf{y}_0) \cap V$.

Define $G(\mathbf{z}) = g(\mathbf{y}_0, \mathbf{z}_0)$, g is injective implies that G is injective, and D_g injective $\implies DG$ injective. $\square$