`richardwu.ca`

MATH 247 Course Notes

Calculus 3 (Advanced)

Spiro Karigiannis • Winter 2018 • University of Waterloo

Last Revision: April 11, 2018

Table of Contents

Abstract

These notes are intended as a resource for myself; past, present, or future students of this course, and anyone interested in the material. The goal is to provide an end-to-end resource that covers all material discussed in the course displayed in an organized manner. These notes are my interpretation and transcription of the content covered in lectures. The instructor has not verified or confirmed the accuracy of these notes, and any discrepancies, misunderstandings, typos, etc. as these notes relate to course's content is not the responsibility of the instructor. If you spot any errors or would like to contribute, please contact me directly.

1 January 3, 2018

1.1 Euclidean space $\mathbb{R}^n$

Most postulates and theorems apply to any n-dimensional real vector space with a positive-definite inner product.

$$\mathbb{R}^n = \{x = (x_1, x_2, \ldots, x_n); x_j \in \mathbb{R}, j = 1, \ldots, n\}$$

Some properties of vectors in $\mathbb{R}^n$ where $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, and $t \in \mathbb{R}$:

$$\begin{aligned}
&x + y = (x_1 + y_1, \ldots, x_n + y_n) \\
&tx = (tx_1, \ldots, tx_n) \\
&x + y = y + x \\
&(x + y) + z = x + (y + z) \\
&s(tx) = (st)x \\
&t\vec{0} = \vec{0} \\
&\vec{0}x = \vec{0} \\
&(t + s)x = tx + sx \\
&t(x + y) = tx + ty
\end{aligned}$$

1.2 Euclidean inner product

An important additional structure on $\mathbb{R}^n$ is the natural **Euclidean inner product** (aka the *dot product*).

$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

which can be written as $x \cdot y \in \mathbb{R}$.
Dot products are **billinear**, **symmetric**, and **positive-definite**. **Bilinearity** means

$$\begin{aligned}
(x + y) \cdot z &= x \cdot z + y \cdot z \\
x \cdot (y + z) &= x \cdot y + x \cdot z \\
(tx) \cdot y = x \cdot (ty) &= t(x \cdot y)
\end{aligned}$$

symmetric denotes

$$x \cdot y = y \cdot x$$

and **positive-definiteness** means $x \cdot x \geq 0$ with equality $\iff x = \vec{0}$.

Definition 1.1. The dot product is defined for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$

$$x \cdot y = \sum_{k=1}^{n} x_k y_k$$

Definition 1.2. The norm $\|x\|$ of $x \in \mathbb{R}^n$ (induced by some inner product $\langle x, x \rangle = x \cdot x$) is defined as

$$\|x\|^2 = x \cdot x$$
$$\|x\| = \sqrt{x \cdot x}$$

1.3 Triangle inequality

Proposition 1.1. Triangle inequality states

$$\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in \mathbb{R}^n$$

To prove the above, we need the **Cauchy-Schwarz Inequality**.

Theorem 1.1. The Cauchy-Schwarz inequality states that

$$|x \cdot y| \leq \|x\|\|y\|$$

with equality iff $x = ty$ or $y = tx$ for some $t \in \mathbb{R}$.

Proof. For the equality case, WLOG if $x = ty$

$$\begin{aligned} x \cdot y = ty \cdot y &= t\|y\|^2 \\ &= |t|\|y\|^2 \\ &= \|x\|\|y\| \end{aligned}$$

Let $t \in \mathbb{R}$. Note for all t

$$\begin{aligned} 0 \leq \|x - ty\|^2 &= (x - ty) \cdot (x - ty) \\ &= x \cdot x - ty \cdot x - tx \cdot y + t^2 y \cdot y \\ &= \|x\|^2 + t^2\|y\|^2 - 2t(x \cdot y) \end{aligned}$$

Thus we have

$$at^2 + bt + c \geq 0 \quad \forall t \in \mathbb{R}$$

where $a = \|y\|^2$, $b = -2x \cdot y$ and $c = \|x\|^2$. Note there can exist at most one root (positive parabola where all values are non-negative). For $at^2 + bt + c = 0$ to have at most one real root (such that t exists), we need $b^2 - 4ac \leq 0$ (from the quadratic formula).

$$\begin{aligned} 4(x \cdot y)^2 &\leq 4\|x\|^2\|y\|^2 \\ |x \cdot y| &\leq \|x\|\|y\| \end{aligned}$$

If we have equality $\exists t_0$ such that $at_0^2 + bt_0 + c = 0$ or $\|x - t_0 y\|^2 = 0$ so $x = t_0 y$. $\square$

Corollary 1.1. The triangle inequality

$$\begin{aligned}\|x+y\|^2 &= (x+y)\cdot(x+y)\\ &= \|x\|^2 + 2x\cdot y + \|y\|^2\\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\|+\|y\|)^2\end{aligned}$$

where the last line follows from the Cauchy-Schwarz inequality.

Definition 1.3. The **distance** between two points $x, y \in \mathbb{R}^n$ is defined to be

$$d(x,y) = \|x-y\|$$

which satisfies the properties

$$\begin{aligned}&d(x,y) = d(y,x)\\ &d(x,x) = 0\\ &d(x,y) \geq 0 \quad \text{with equality iff} \quad x = y\end{aligned}$$

so we can restate the triangle inequality as $d(x,y) \leq d(x,z) + d(z,x) \quad \forall x,y,z \in \mathbb{R}^n$.

1.4 Norms

There exists different "natural" norms on $\mathbb{R}^n$

Definition 1.4. A norm $\|\cdot\|$ on $\mathbb{R}^n$ is a map

$$\|\cdot\| : \mathbb{R}^n \to \mathbb{R}^{\geq 0}$$

such that

1. $\|x\| = 0 \iff x = \vec{0}$
2. $\|tx\| = |t|\|x\|$
3. $\|x+y\| \leq \|x\| + \|y\|$

All inner products determine a norm but not all norms are from inner products. We saw that the dot product determines a norm called the Euclidean norm.

l^1 norm $\|x\|_1 = \sum_{k=1}^n |x_k|$

l^p norm $\|x\|_p = (\sum_{k=1}^n |x_k|^p)^{\frac{1}{p}}$

sup norm (aka l^∞ norm) $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$

One can see that l^∞ norm is a "limit" of l^p norms as $p \to \infty$.
Note the l^2 norm is the Euclidean norm.

Why are norms important? **A norm determines a distance**. For example

$$d(x,y) = \|x-y\|$$

(all norms determine a distance but not all distances are from norms).
Distance is important to define a **limit** which is crucial for differentiability/integrability.

1.5 Angle between two vectors

A corollary to C-S for $x, y \neq \vec{0}$

$$-1 \leq \frac{x \cdot y}{\|x\|\|y\|} \leq 1$$

Define the angle $\theta \in [0, \pi]$ between x and y to be

$$\cos\theta = \frac{x \cdot y}{\|x\|\|y\|}$$

so we have another definition of the dot product

$$x \cdot y = \|x\|\|y\| \cos\theta$$

We say x, y are **orthogonal** if $\theta = \frac{\pi}{2} \iff x \cdot y = 0$.
Why is this the correct definition?

$$\begin{aligned}
\|y - x\|^2 &= (y - x) \cdot (y - x) \\
&= \|x\|^2 + \|y\|^2 - 2x \cdot y \\
&= \|x\|^2 + \|y\|^2 - 2\|x\|\|y\| \cos\theta
\end{aligned}$$

This aligns with the Law of Cosines $c^2 = a^2 + b^2 - 2ab\cos\theta$.

2 January 5, 2018

2.1 Linear maps

Definition 2.1. A map $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear if T takes linear combinations to linear combinations i.e.

$$T(\sum_{k=1}^{N} t_k v_k) = \sum_{k=1}^{N} t_k T(v_k) \quad t_i \in \mathbb{R} \quad v_j \in \mathbb{R}^n$$

We will see linear maps are closely related to **differentiability**.
Some facts about linear maps: let $e_1, \ldots, e_n$ be the standard basis.

$$x \in \mathbb{R}^n = (x_1, \ldots, x_n) = \sum_{k=1}^{n} x_k e_k$$

Let $f_1, \ldots, f_m$ be the standard basis of $\mathbb{R}^m$ where $f_j = (0, \ldots, 1, \ldots, 0) \in \mathbb{R}^m$.

$$y \in \mathbb{R}^m = (y_1, \ldots, y_n) = \sum_{k=1}^{m} y_k f_k$$

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear and let

$$\begin{aligned} y = \sum_{l=1}^{m} y_l f_l = T(x) &= T(\sum_{k=1}^{n} x_k e_k) \\ &= \sum_{k=1}^{n} x_k T(e_k) \\ &= \sum_{k=1}^{n} x_k (\sum_{l=1}^{m} A_{lk} f_l) \\ &= \sum_{k=1}^{n} (\sum_{l=1}^{m} A_{lk} x_k) f_l \end{aligned}$$

By uniqueness of the expansion of a vector in terms of a basis (f_js) we conclude that

$$y_l = \sum_{k=1}^{n} A_{lk} x_k \quad l = 1, \dots, m$$

or in matrix form

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \dots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

We've shown that any linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is necessarily **matrix multiplication**

$$y = T(x) = A \cdot x$$

for some unique $m \times n$ matrix A (with respect to some bases in $\mathbb{R}^n$ and $\mathbb{R}^m$).
The rule of matrix multiplication is automatic from the composition of linear maps. Let

$$\begin{aligned} &T : \mathbb{R}^n \to \mathbb{R}^m \\ &S : \mathbb{R}^m \to \mathbb{R}^p \\ &y = T(x) = A \cdot x \quad m \times n \\ &z = S(y) = B \cdot y \quad p \times m \end{aligned}$$

Therefore $S \circ T : \mathbb{R}^n \to \mathbb{R}^p$ is linear.

$$\begin{aligned} (S \circ T)(\sum t_k v_k) &= S(T(\sum_k t_k v_k)) \\ &= S(\sum_k x_k T(v_k)) \\ &= \sum_k x_k S(T(v_k)) \\ &= \sum_k t_k (S \circ T)(v_k) \end{aligned}$$

So we have

$$\begin{aligned} z_l = \sum_{j=1}^{m} B_{lj} y_j &= \sum_{j=1}^{m} B_{lj} (\sum_{i=1}^{n} A_{ji} x_i) \\ &= \sum_{i=1}^{n} (\sum_{j=1}^{m} B_{lj} A_{ji}) x_i \\ &= \sum_{i=1}^{n} C_{li} x_i \end{aligned}$$

where

$$z = (S \circ T)(x) = C \cdot x \quad p \times n$$

Recall the space $L(\mathbb{R}^n, \mathbb{R}^m)$ of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ is itself a finite dimensional real vector space of dimension nm (isomorphic to $\mathbb{R}^{nm}$).

$$T \in L(\mathbb{R}^n, \mathbb{R}^m) \iff A \in M_{m\times n}(\mathbb{R})$$

where $M_{m\times n}(\mathbb{R})$ is the space of real $m \times n$ matrices. There is a unique 1-1 correspondence between T and A (as shown before).

2.2 Operator norm

Note one can define norm on matrices. The natural Euclidean norm for matrix A can be defined as

$$\|A\|_2 = \sqrt{\sum_{i=1,\ldots,m; j=1,\ldots,n} (A_{ij})^2}$$

Definition 2.2. The **operator norm** is defined for a $T : \mathbb{R}^n \to \mathbb{R}^m$ linear map as

$$\|T\|_{op} = \inf\{C > 0, \|T(x)\| \le C\|x\| \quad \forall x \in \mathbb{R}^n\}$$

We need to show this norm is

1. Well-defined
2. $\|\cdot\|_{op}$ is a norm

1. Show well-defined

$$T(x) = A \cdot x \quad A \quad m \times n$$

$$\begin{bmatrix} A_{11} & \ldots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \ldots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} A_1 \cdot x \\ \vdots \\ A_m \cdot x \end{bmatrix} = T(x)$$

So the norm is

$$\begin{aligned} \|T(x)\|^2 &= (A_1 \cdot x)^2 + \ldots + (A_m \cdot x)^2 \\ &\le \|A_1\|^2\|x\|^2 + \ldots + \|A_m\|^2\|x\|^2 && \text{C-S} \\ &= (\|A_1\|^2 + \ldots + \|A_m\|^2)\|x\|^2 \end{aligned}$$

Case 1 Assume $\|A_1\|^2 + \ldots + \|A_m\|^2 = 0$.

$$\begin{aligned}\|A_1\|^2 + \ldots + \|A_m\|^2 = 0 &\iff A = 0_{m\times n} \\ &\iff T = 0 \in L(\mathbb{R}^n, \mathbb{R}^m)\end{aligned}$$

Then $T(x) = 0 \quad \forall x$ so $\|T(x)\| \leq C\|x\|$ holds $\forall C > 0$, thus the infimum of positive real numbers (0) implies $\|T\|_{op} = 0$.

Case 2 Assume $\|A_1\|^2 + \ldots + \|A_m\|^2 > 0$.
$\{C > 0, \|T(x)\| \leq C\|x\| \quad \forall x \in \mathbb{R}^n\}$ is non-empty because $\sqrt{\|A_1\|^2 + \ldots + \|A_m\|^2}$ is in there. By the completeness of $\mathbb{R}$, $\|T\|_{op}$ exists and is ≥ 0.

2. We've shown $\|T\|_{op}$ exists and is ≥ 0 for all $T \in L(\mathbb{R}^n, \mathbb{R}^m)$. It remains to shown $\|T\|_{op}$ is a norm:

   (a) $\|T\|_{op} = 0$ only for the zero map

   (b) $\|\lambda T\|_{op} = |\lambda|\|T\|_{op} \quad \forall \lambda \in \mathbb{R}$

   (c) $\|T + S\|_{op} \leq \|T\|_{op} + \|S\|_{op}$

   To see this, we note that since

$$\|T\|_{op} = \inf\{C > 0, \|T(x)\| \leq C\|x\| \quad \forall x \in \mathbb{R}^n\}$$

   $\exists$ a **decreasing sequence** $c_k \geq 0$ such that $\|T(x)\| \leq c_k\|x\| \quad \forall x \in \mathbb{R}^n$ and $\lim_{k\to\infty} c_k = \|T\|_{op}$.
   Take limit as $k \to \infty$ of the predicate in $\|T\|_{op}$.

$$\begin{aligned}\|T(x)\| &\leq (\lim_{k\to\infty} c_k)\|x\| \\ \|T(x)\| &\leq \|T\|_{op}\|x\|\end{aligned}$$

   So we have

$$\begin{aligned}\|T\|_{op} = 0 &\Rightarrow \|T(x)\| \leq 0 \quad \forall x \\ &\Rightarrow T(x) = 0 \quad \forall x \\ &\Rightarrow T = 0 \in L(\mathbb{R}^n, \mathbb{R}^m)\end{aligned}$$

   which proves (a).

$$\|\lambda T\|_{op} = |\lambda|\|T\|_{op}$$

   follows from

$$\begin{aligned}\|(\lambda T)(x)\| &= \|\lambda(T(x))\| \\ &= |\lambda|\|T(x)\| \quad \forall x\end{aligned}$$

   If $\lambda = 0$, $\lambda T = 0 \Rightarrow \|\lambda T\|_{op} = 0 = |\lambda|\|T\|_{op}$.

If $\lambda \neq 0$

$$\begin{aligned}
\|\lambda T\|_{op} &= \inf\{C > 0, \|(\lambda T)(x)\| \leq C\|x\|\} \\
&= \inf\{C > 0, |\lambda|\|T(x)\| \leq C\|x\|\} \\
&= \inf\{C > 0, \|T(x)\| \leq \frac{C}{|\lambda|}\|x\|\} \\
&= |\lambda| \inf\{\tilde{C} > 0, \|T(x)\| \leq \tilde{C}\|x\|\} && \tilde{C} = \frac{C}{\lambda} \\
&= |\lambda|\|T\|_{op}
\end{aligned}$$

which proves (b). (c) is similar.

3 January 8, 2018

3.1 Topology of $\mathbb{R}^n$

Topology is the study of **closeness** in a space.

3.2 Open and closed balls

Definition 3.1. Let $x \in \mathbb{R}^n$ and $r > 0$. The **open ball** at radius r centred at x is denoted

$$B_r(x) = \{y \in \mathbb{R}^n \mid \|x - y\| < r\}$$

It consists of all points in $\mathbb{R}^n$ whose distance from x is *strictly less than* r.

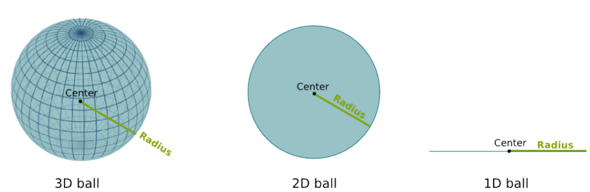

Figure 3.1: Open balls in $\mathbb{R}$, $\mathbb{R}^2$, and $\mathbb{R}^3$.

In $\mathbb{R}$, $B_r(x) = (x - r, x + r)$. In $\mathbb{R}^3$, $B_r(x)$ is the *interior* of a sphere of radius r centred at x.

Definition 3.2. Let $x \in \mathbb{R}^n$, $r > 0$. The **closed ball** of radius $r > 0$ centred at x is denoted

$$\overline{B_r(x)} = \{y \in \mathbb{R}^n \mid \|x - y\| \leq r\}$$

Remark 3.1. The notation will be explained in the following class/section. Note that

$$\overline{B_r(x)} = B_r(x) \cup \{\text{points exactly at distance } r\}$$

For $n = 1$, $\overline{B_r(x)} = [x - r, x + r]$.

3.3 Open sets

Definition 3.3. A subset $U \subseteq \mathbb{R}^n$ is called an **open set** (or open) iff $\forall x \in U$, $\exists r > 0$ (r depends on x) such that $B_r(x) \subseteq U$.

(Informally: a subset U is open if for every $x \in U$, all points sufficiently close to x are *also* in U).

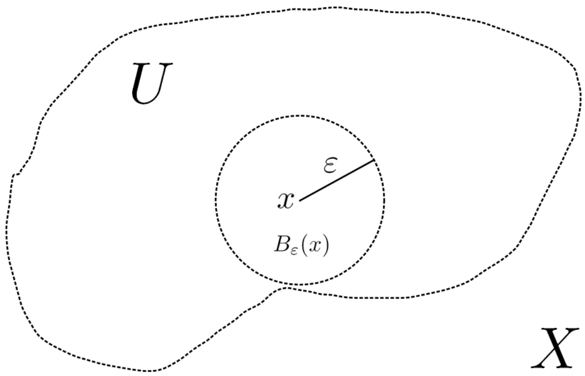

Figure 3.2: One can form an open ball for every point x in an open set U.

Example 3.1. Set that is not open

- $[0, 1] \subseteq \mathbb{R}$. Note: $\nexists r > 0$ for $x = 1$ such that $B_r(x) \subseteq [0, 1]$.

Sets that are open

- $\mathbb{R}^n$ since $x + \epsilon \in \mathbb{R}^n$ by definition.
- $\varnothing$ (vacuous: satisfied trivially $\varnothing$ has no points).

Proposition 3.1. An **open ball** is an **open set**.

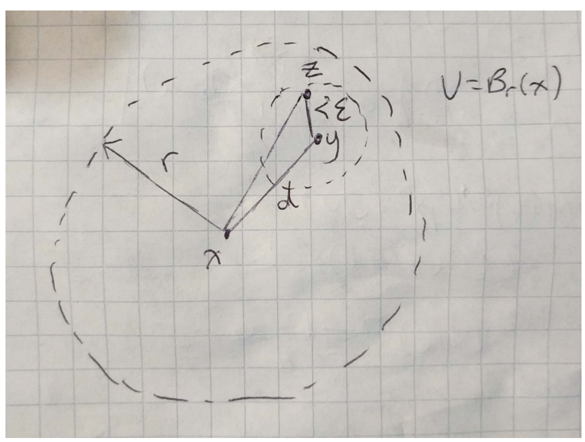

Figure 3.3: An open ball is an open set (see proof below).

Proof. Let $U = B_r(x)$ and $y \in U = B_r(x)$. We need to find some $\epsilon > 0$ such that $B_\epsilon(y) \subseteq U$.
Let $d = \|x - y\| < r$ since $y \in U = B_r(x)$.
Let $\epsilon = r - d > 0$.
Suppose $z \in B_\epsilon(y)$ thus $\|y - z\| < \epsilon$.
We thus have

$$\|z - x\| \overset{\triangle}{\leq} \|z - y\| + \|y - x\| < \epsilon + d = r$$

So $B_\epsilon(y) \subseteq U$ hence U is open. $\square$

We can construct more from open sets.

3.4 Properties of open sets

Lemma 3.1.

1. Let $U_\alpha \subseteq \mathbb{R}^n$ be open $\forall \alpha \in A$ (countably or uncountably many), then

$$\bigcup_{\alpha \in A} U_\alpha$$

is open.

2. Let $U_1, \ldots, U_k$ be open (**must be finite** number of sets). Then

$$\bigcap_{j=1}^{k} U_j$$

is open.

Informally, *arbitrary unions* of open sets are open. *Finite intersections* of open sets are open.

Proof.

1. We want to show $\bigcup_{\alpha \in A} U_\alpha$ is open.

   Let $x \in \bigcup_{\alpha \in A} U_\alpha$ so $\exists$ some $\alpha_0 \in A$ such that $x \in U_{\alpha_0}$ (holds since union of sets).

   But U_{α_0} is open so $\exists r > 0$ such that $B_r(x) \subseteq U_{\alpha_0} \subseteq \bigcup_{\alpha \in A} U_\alpha$.

2. Show $x \in \bigcap_{j=1}^{k} U_j$ so $x \in U_j$ for all $j = 1, \ldots, k$. Each U_j is open so $\forall j, \exists \epsilon_j > 0$ such that $B_{\epsilon_j}(x) \subseteq U_j$.

   Let $\epsilon = \min\{\epsilon_1, \ldots, \epsilon_k\} > 0$. $\forall j$ we have $B_\epsilon(x) \subseteq B_{\epsilon_j}(x) \subseteq U_j$ hence $B_\epsilon(x) \subseteq \bigcap_{j=1}^{k} U_j$.

   Remark 3.2. Arbitrary (e.g. nonfinite) intersections of open sets need not be open (the min. of infinite numbers is not well defined. An infimum of positive numbers need not be > 0 i.e. it could be 0).

   Even intersection of countably infinite sets may not be open. Suppose $U_k = (0, 1 + \frac{1}{k}) \subseteq \mathbb{R} \quad \forall k \in \mathbb{N}$. Note that $\bigcap_{k=1}^{\infty} U_k = (0, 1]$ is not open.

□

3.5 Closed sets

Definition 3.4. A subset $F \subseteq \mathbb{R}^n$ is called **closed** if $F^c = \mathbb{R} \setminus F$ is open (note: this definition is based on open's definition).

Proposition 3.2. A closed ball $\overline{B_r(x)} = \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$ is a closed set.

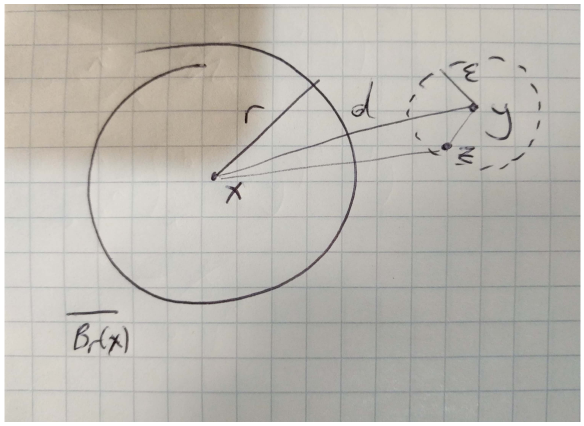

Figure 3.4: A closed ball is a closed set (see proof below).

Proof. Let $F = \overline{B_r(x)}$ and

$$F^c = (\overline{B_r(x)})^c = \{y \in \mathbb{R}^n \mid \|y - x\| > r\}$$

Let $y \in \overline{B_r(x)}^c$: need to find $\epsilon > 0$ such that $B_\epsilon(y) \subseteq F^c$.
Let $d = \|x - y\| > r$ and let $\epsilon = d - r > 0$.
If $z \in B_\epsilon(y)$, then

$$\begin{aligned} \|x - y\| &\overset{\triangle}{\leq} \|x - z\| + \|z - y\| \\ d &\leq \|x - z\| + \|z - y\| \\ \|x - z\| \geq d - \|z - y\| & \\ &> d - \epsilon = r \end{aligned}$$

Hence $z \in F^c$ so $B_\epsilon(y) \subseteq F^c$, thus F^c is open and by definition F is closed. □

3.6 Properties of closed sets

Lemma 3.2. Note: this lemma is the inverse of the equivalent for open sets.

1. If $F_1, \ldots, F_k$ is closed, then $\bigcup_{j=1}^k F_j$ is closed.
2. If F_α is closed $\forall \alpha \in A$, then $\bigcap_{\alpha \in A} F_\alpha$ is closed.

Finite unions of closed sets are closed. Arbitrary intersections of closed sets are closed.

Proof. By De Morgan's laws

$$(\bigcup_{j=1}^k F_j)^c = \bigcap_{j=1}^k (F_j)^c$$
$$(\bigcap_{\alpha \in A} F_\alpha)^c = \bigcup_{\alpha \in A} (F_\alpha)^c$$

□

3.7 Neither open nor closed

A subset V of $\mathbb{R}^n$ need not be either open or closed. It can be open, closed, neither or both!

Example 3.2. Examples of non-exclusive open or closed sets are

- $(0,1] \subseteq \mathbb{R}$ - neither
- $\mathbb{R}^n, \varnothing$ are *open and closed*

3.8 Interior

Sometimes a set is neither open nor closed, but there are always **natural open (interior) and closed (closure) sets** which can be associated to any subset of $\mathbb{R}^n$.

Definition 3.5. Let $A \subseteq \mathbb{R}^n$ (could be $\varnothing$).

$$\begin{aligned} A^o &= int(A) && \text{interior of } A \\ &= \bigcup_{\substack{V \subseteq A \\ V \text{open in } \mathbb{R}^n}} V && \text{union of \textbf{all} open subsets of } \mathbb{R}^n \text{ that are contained in } A \end{aligned}$$

Remark 3.3.

1. A^o is open (arbitrary union of open sets) and $A^0 \subseteq A$
2. if V is any open subset of R^n that is contained in A, then $V \subseteq A^o$ (A^o is the largest open subset of R^n that is contained in A)
3. A is open iff $A^o = A$

Proof. **Forwards:**

A is open and $A \subseteq A$ thus A must be a V in the union, but since all $V \subseteq A^o \subseteq A$ (where A is a V) then $A^o = A$.

Backwards:

$A^o = A$. Since A^o is open, A is open. □

3.9 Closure

Definition 3.6.

$$\begin{aligned} \overline{A} &= cl(A) && \text{closure of } A \\ &= \bigcap_{\substack{F \supseteq A \\ F \text{closed in } \mathbb{R}^n}} F && \text{intersection of \textbf{all} closed subsets of } \mathbb{R}^n \text{ that contains } A \end{aligned}$$

Remark 3.4.

1. $\overline{A}$ is closed (arbitrary intersection of closed sets) and $\overline{A} \supseteq A$
2. if F is any closed subset of R^n that contains A, then $F \supseteq \overline{A}$ ($\overline{A}$ is the smallest closed set of R^n containing A)
3. A is closed iff $\overline{A} = A$

4 January 10, 2018

4.1 Closure of open ball is closed ball

Proposition 4.1. The closure of the open ball $B_\epsilon(x)$ is the closed ball $\overline{B_\epsilon(x)}$ (hence the notation).

Proof. Remember

$$\overline{B_\epsilon(x)} = \{y \in \mathbb{R}^n \mid \|y - x\| \leq \epsilon\}$$

Let $A =$ closure of $B_\epsilon(x)$.
Let $F = \{y \in \mathbb{R}^n \mid \|x - y\| \leq \epsilon\}$.
We want to show $A = F$.
We know F is closed and $F \supset B_\epsilon(x)$, so F contains A: the closure of $B_\epsilon(x)$ (any closed set containing another set is in the intersection of the closure) or

$$F \supset A \supset B_\epsilon(x)$$

Suppose $F \neq A$, then $\exists y \in F$ with $y \notin A \Rightarrow y \notin B_\epsilon(x)$ so

$$\|x - y\| = \epsilon$$

(it's sandwiched between the closed ball ($\leq \epsilon$) and the open ball ($< \epsilon$), so it must hold with equality with ϵ by the Trichotomy property).

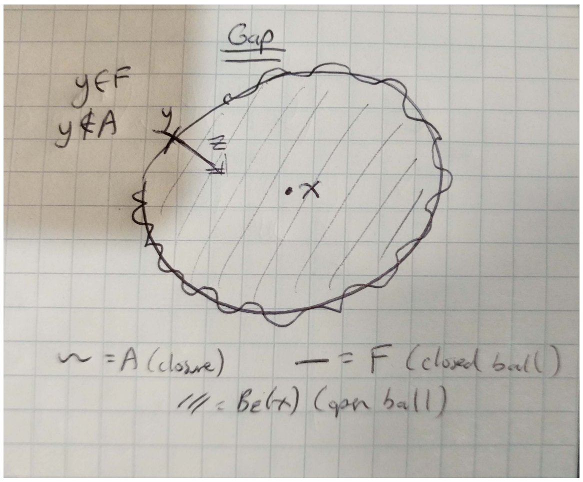

Figure 4.1: The closure of an open ball is the corresponding closed ball.

A is closed and $y \notin A$ so A^c is open and $y \in A^c$. So $\exists \delta > 0$ such that $B_\delta(y) \subseteq A^c$.
Let $t > 0$ with $t < \min\{\delta, \epsilon\}$.
Let

$$z = y + t\frac{(x - y)}{\|x - y\|}$$

(add t unit vectors from y to x). Note that

$$\|z - y\| = t < \delta$$

so $z \in B_\delta(y) \subseteq A^c$.

Also

$$\begin{aligned} x - z &= x - y - t\frac{(x-y)}{\|x-y\|} \\ &= (\|x-y\| - t)\frac{(x-y)}{\|x-y\|} \end{aligned}$$

where the left term is the norm of the vector and the right term is the unit vector.
Thus

$$\|x - z\| = |\|x-y\| - t| = |\epsilon - t| = \epsilon - t < \epsilon$$

So $z \in B_\epsilon(x) \subseteq A$, but we assumed $z \in A^c$ which is a contradiction.
So we must have $F = A$.

Remark 4.1. There is a much simpler proof of this using sequences and limit points.

□

4.2 Boundary

Definition 4.1. Let $A \subseteq \mathbb{R}^n$. We define the **boundary** of A denoted $\partial A = bd(A)$ to be

$$\partial A = bd(A) = \{x \in \mathbb{R}^n \mid B_\epsilon(x) \cap A \neq \varnothing, B_\epsilon(x) \cap A^c \neq \varnothing \quad \forall \epsilon > 0\}$$

That is, $x \in \partial A$ iff every open ball centred at x contains a point in A **and** a point in A^c.
Clearly

$$\begin{aligned} \partial B_\epsilon(x) &= \{y \in \mathbb{R}^n \mid \|y - x\| = \epsilon\} \\ &= \partial(\overline{B_\epsilon(x)}) \end{aligned}$$

4.3 Characterization of boundary

Proposition 4.2. Let $A \subseteq \mathbb{R}^n$, then

$$\begin{aligned} \partial A &= \overline{A} \setminus A^o \\ &= cl(A) \setminus int(A) \end{aligned}$$

Proof. The following two claims and proofs revolve around complements of sets and how if set A intersect a set B is the empty set, then A is a subset of B^c.

Claim 1

$$x \in \overline{A} \iff B_\epsilon(x) \cap A \neq \varnothing \quad \forall \epsilon > 0$$

Proof. **Forwards:**

Suppose $x \in \overline{A}$ but $\exists \epsilon_0 > 0$ $B_\epsilon(x) \cap A = \varnothing$.

So $B_\epsilon(x) \subseteq A^c \Rightarrow (B_\epsilon(x))^c \supset A$.

Since $(B_\epsilon(x))^c$ is closed, then $(B_\epsilon(x))^c \supset \overline{A}$ (by remark (2) after closure definition).

So $\overline{A} \cap B_\epsilon(x) = \varnothing$, but $x \in B_\epsilon(x) \Rightarrow x \notin \overline{A}$, which is a contradiction.

Backwards:

We prove the contrapositive

$$x \notin \overline{A} \Rightarrow B_\epsilon(x) \cap A = \varnothing \quad \exists \epsilon > 0$$

Assume $x \notin \overline{A} \Rightarrow x \in (\overline{A})^c$ which is open, so $\exists \epsilon_0 > 0$ such that $B_{\epsilon_0}(x) \subseteq (\overline{A})^c$. Therefore $B_{\epsilon_0}(x) \cap \overline{A} = \varnothing$ (where $\overline{A} \supset A$), which proves our claim). $\square$

Claim 2

$$x \notin A^o \iff B_\epsilon(x) \cap A^c \neq \varnothing \quad \forall \epsilon > 0$$

Proof. **Forwards:**

Suppose $x \notin A^o$. Assume (for contradiction) $\exists \epsilon_0 > 0$ such that

$$B_{\epsilon_0}(x) \cap A^c = \varnothing \Rightarrow B_{\epsilon_0}(x) \subseteq A$$

(nothing in A^c, thus all in A).

Ergo $x \in (A^o)^c$ and $B_{\epsilon_0}(x) \subseteq A^o$ (since $B_{\epsilon_0}(x)$ is an open set contained in A - remark (2) after interior definition).

So $B_{\epsilon_0}(x) \cap (A^o)^c = \varnothing$ but $x \in B_{\epsilon_0}(x) \cap (A^o)^c$ which is a contradiction.

Backwards:

(Contrapositive): suppose $x \in A^o$. A^o is open so $\exists \epsilon > 0$ such that

$$B_{\epsilon_0}(x) \subseteq A^o \subseteq A$$

so $B_{\epsilon_0}(x) \cap A^c = \varnothing$ for some $\epsilon_0 > 0$. $\square$

Putting the claims together:

$$\begin{aligned}
x \in \overline{A} &\iff B_\epsilon(x) \cap A \neq \varnothing \quad \forall \epsilon > 0 && (1)\\
x \in (A^o)^c &\iff B_\epsilon(x) \cap A^c \neq \varnothing \quad \forall \epsilon > 0 && (2)\\
x \in \partial A &\iff (1) + (2)\\
&\iff x \in \overline{A} \cap (A^o)^c = \overline{A} \setminus A^o
\end{aligned}$$

$\square$

4.4 Sequential characterization of limits

Definition 4.2. Let (x_k) be a sequence of points in $\mathbb{R}^n, k \in \mathbb{N}$. We say (x_k) **converges** to a point $x \in \mathbb{R}^n$ **iff** for any $\epsilon > 0$, $\exists N \in \mathbb{N}$ (N depends on ϵ in general)

$$k \geq N \Rightarrow \|x_k - x\| < \epsilon$$

(i.e. for any $\epsilon > 0$, **all** the elements of sequence x_k after some $k = N$ are within ϵ of x).

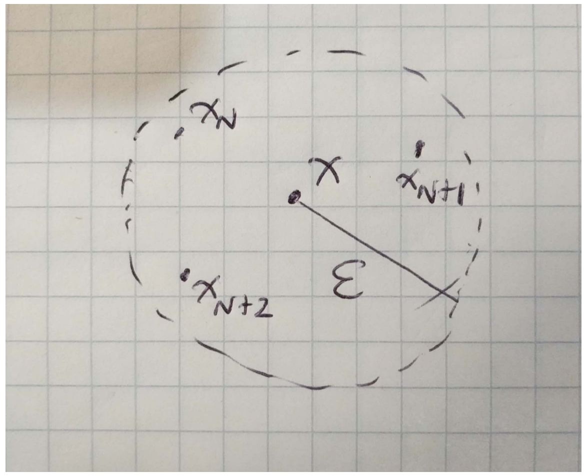

Figure 4.2: All points after $k = N$ for a converging sequence is within ϵ.

If (x_k) converges to x, we denote

$$\lim_{k\to\infty} x_k = x$$

where x is **the limit** of x_k.

4.5 Uniqueness of limits

Lemma 4.1. Suppose $\lim_{k\to\infty} x_k = x$ and $\lim_{k\to\infty} x_k = y$. Then $x = y$ (i.e. a sequence may not converge, but if it does the limit is unique).

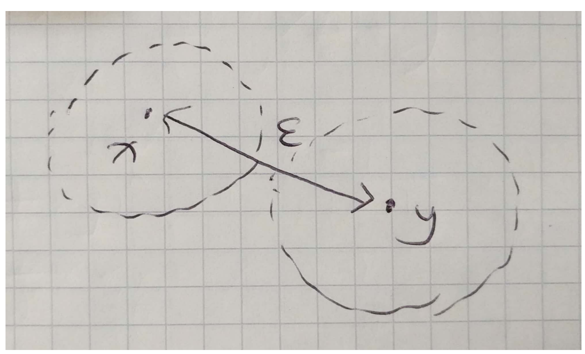

Figure 4.3: Sketch of proof with $x \neq y$ (see below).

Proof. Suppose $x \neq y$, so $\|x - y\| = \epsilon > 0$.
Since (x_k) converges to x, $\exists N_1 \in N$ such that $k \geq N_1$ and

$$\|x_k - x\| < \frac{\epsilon}{2}$$

Similarly for y $\exists k \geq N_2$.

Suppose $k \geq \max\{N_1, N_2\}$. Then

$$\begin{aligned}\|x-y\| &\overset{\triangle}{\leq} \|x-x_k\| + \|x_k - y\| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon\end{aligned}$$

So $x = y$ by contradiction. $\square$

4.6 Neighbourhood

Definition 4.3. Let $x \in \mathbb{R}^n$. A subset $U \in \mathbb{R}^n$ is called a **neighbourhood (n'h'd)** of x if $\exists \epsilon_0 > 0$ such that $B_{\epsilon_0}(x) \subseteq U$.

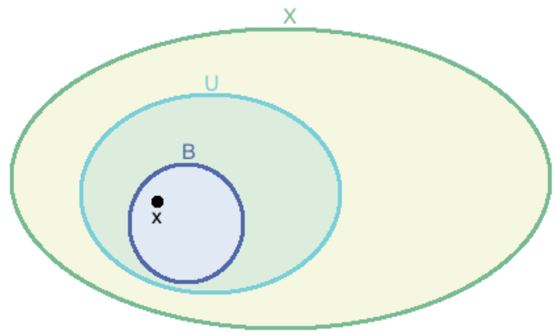

Figure 4.4: U is a neighbourhood of x since there exists an open set B of x contained in U.

(Equivalently, U is a n'h'd of $x \iff U$ contains an open set containing x.)

Definition 4.4. An open n'h'd of x is any open set containing x. (A set is an open n'h'd of x if it contains x **and** all points sufficiently close to x).

Lemma 4.2. Let (x_k) be a sequence in $\mathbb{R}^n$. Suppose $\lim_{k\to\infty} x_k$ exists and equal $x \in \mathbb{R}^n$. Then any n'h'd of x contains all x_k's for k sufficiently large, i.e. if U is a n'h'd of x, $\exists N \in \mathbb{N}$ (N depends on U) such that

$$k \geq N \Rightarrow x_k \in U$$

Proof. U is a n'h'd of x so $\exists \epsilon_0 > 0$ such that $B_\epsilon(x) \subseteq U$.
Since $\lim_{k\to\infty} x_k = x$, $\exists N \in N$ such that $k \geq N \Rightarrow \|x_k - x\| < \epsilon_0$ so $x_k \in B_\epsilon(x) \subseteq U \quad \forall k \geq N$. $\square$

5 January 12, 2018

5.1 Relationship between convergent sequences and open/closed sets

Recall: $x \in \overline{A} \iff B_\epsilon(x) \cap A \neq \varnothing \quad \forall \epsilon > 0$.

Proposition 5.1. Suppose $x \in \overline{A}$. Take $\frac{1}{k} > 0$. From above: $\exists x_k \in A$ such that $\|x_k - x\| < \frac{1}{k}$, then $\lim_{k\to\infty} x_k = x$.

Proof. Let $\epsilon > 0$ so $\exists N \in \mathbb{N}$ such that $\frac{1}{N} < \epsilon$ (Archimedean Principle). $\forall k \geq N$, $\frac{1}{k} \leq \frac{1}{N} < \epsilon$ so $\|x_k - x\| < \frac{1}{k} < \epsilon$. $\square$

To summarize, if $x \in \overline{A}$, then $\exists$ a sequence (x_k) such that $\lim_{k\to\infty} x_k = x$ **and** $x_k \in A \quad \forall \in \mathbb{N}$.

What about the converse?

Proposition 5.2. Suppose $x_k \in A$, $\lim_{k\to\infty} x_k = x$ **and** $x_k \in A\ \forall k \in \mathbb{N}$. Then $x \in \overline{A}$.

Proof. If not, $x \in (\overline{A})^c$ so $\exists \epsilon > 0$ such that $B_\epsilon(x) \subseteq (\overline{A})^c$. But $\exists N \in \mathbb{N}$ such that

$$k \geq N \Rightarrow x_k \in B_\epsilon(x)$$

and so $x_k \in (\overline{A})^c$. But from our hypothesis we have $x_k \in A \subseteq \overline{A}$ which is a contradiction. Thus $x \in \overline{A}$. □

(i.e. whenever (x_k) is a convergent sequence of points all of whose elements are in A, then the limit is in $\overline{A}$).
Special case: If A is closed ($\overline{A} = A$) then if $(x_k) \to x$ and $x_k \in A\ \forall k$ then $x \in A$; this is **not** true for open sets A.

5.2 Bounded and Cauchy sequences

Definition 5.1. A sequence (x_k) in $\mathbb{R}^n$ is called **bounded** if $\exists M > 0$ such that

$$\|x_k\| \leq M \quad \forall k \in \mathbb{N}$$

(that is: all the x_k's lie in some closed ball $\overline{B_M(x)}$ centred at 0).

Definition 5.2. A sequence (x_k) is called **Cauchy** if for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that

$$k, l \geq N \Rightarrow \|x_k - x_l\| < \epsilon$$

(eventually all points in the sequence are close to each other).

5.3 Convergent $\iff$ Cauchy

Proposition 5.3. Let (x_k) be a convergent sequence. Then (x_k) is Cauchy.

Proof. Let $x = \lim_{k\to\infty} x_k$. Let $\epsilon > 0$, then $\exists N$ such that

$$\|x_k - x\| < \frac{\epsilon}{2}$$

If $k, l \geq N$ then

$$\|x_k - x_l\| \overset{\triangle}{\leq} \|x_k - x\| + \|x - x_l\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

□

Recall from MATH 147: In $\mathbb{R}$ every Cauchy sequence converges (equivalent to **completeness** of $\mathbb{R}$ or the real line). We show Cauchy converges in $\mathbb{R}^n$ in assignment 2 by showing that each j-th component $x^{(j)}$ converges then by the completeness of $\mathbb{R}$ this follows for $\mathbb{R}^n$.

5.4 Convergence implies bounded

Lemma 5.1. Every convergent sequence is bounded.

Proof. Let $x = \lim_{k\to\infty} x_k$. Let $M_0 = \|x\| + \epsilon$ for $\epsilon > 0$. Then $\exists N$ such that

$$k \geq N \Rightarrow \|x_k - x\| < \epsilon$$

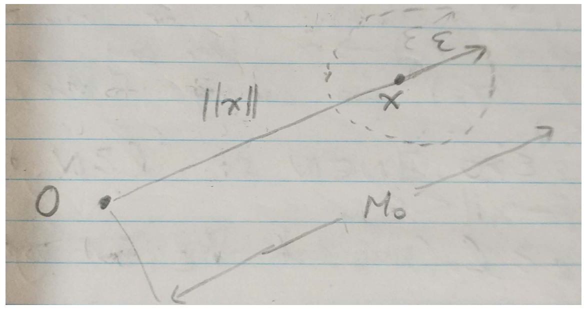

Figure 5.1: Convergent sequences can be bounded by the limit and ϵ and finite points in the sequence.

Note that

$$k \geq N \Rightarrow \|x_k\| \overset{\triangle}{\leq} \|x_k - x\| + \|x\| < \epsilon + \|x\| = M_0$$

Thus we let $M = \max\{\|x_1\|, \ldots, \|x_{N-1}\|, M_0\}$ then $\|x_k\| \leq M \quad \forall k \in \mathbb{N}$. $\square$

Note: not every bounded sequence is Cauchy. Consider $x_k = (-1)^{k+1}$ is $\mathbb{R}$, which is bounded but not convergent. Can we find a weaker statement that's true i.e. given a bounded sequence, can we somehow obtain from it a convergent sequence?

5.5 Subsequences

Definition 5.3. Let (x_k) be a sequence in $\mathbb{R}^n$. Let $1 \leq k_1 < k_2 < \ldots < k_e < k_{e+1} < \ldots$ be a sequence of $1, 2, 3, 4, \ldots$. Then the corresponding sequence (y_l) (or (x_{k_l})) in $\mathbb{R}^n$ given by $y_l = x_{k_l}$ is called a **subsequence** of (x_k).

Example 5.1. The subsequence given by $k_l = 2l - 1$ (odd numbers) is

$$(x_{2l-1}) = x_1, x_3, x_5, \ldots$$

Proposition 5.4. Suppose $(x_k) \to x$. Then any subsequence (x_{k_l}) of (x_k) also converges to the same limit x.

Proof. Let $\epsilon > 0$. $\exists N \in \mathbb{N}$ such that $l \geq N$ then $\|x_l - x\| < \epsilon$, but $k_l \geq l$ (since each k_e must be strictly larger $> k_{e-1}$), so $\|x_{k_l} - x\| < \epsilon \quad \forall l \geq N$ hence $\lim_{k_l \to \infty} x_{k_l} = x$. $\square$

Note: A sequence (x_k) that does not converge can have

1. Subsequences that don't converge (e.g. $k_l = l$ so $x_{k_l} = x_l$).
2. Distinct subsequences with different limits.

   For example, $x_k = (-1)^{k+1}$ which is $1, -1, 1, -1, \ldots$, we can have two subsequences

$$x_{2l-1} = (-1)^{2l} = 1, 1, 1, \ldots \qquad (x_{2l-1}) \to 1$$
$$x_{2l} = (-1)^{2l-1} = -1, -1, -1, \ldots \qquad (x_{2l}) \to -1$$

5.6 Bolzano-Weierstrass (B-W) Theorem

Theorem 5.1. In $\mathbb{R}^n$, every bounded sequence has a convergent subsequence.

Remark 5.1. This convergent subsequence is **not** unique. We'll see in the proof that we make many arbitrary choices.

Proof. By induction on n.
Case $n = 1$: Let (x_k) be a sequence in $\mathbb{R}$ that is **bounded**. So $\exists M > 0$ such that $|x_k| \leq M \quad \forall k \in \mathbb{N} \iff x_k \in [-M, M]$.

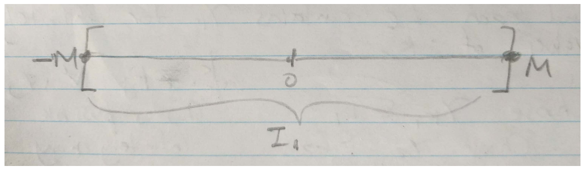

Figure 5.2: I_1 is the interval of our bounded sequence in $\mathbb{R}$.

Define
$$I_1 = [-M, M] = [-M, 0] \cup [0, M]$$
At least one (maybe both) of $[-M, 0]$ and $[0, M]$ contains x_k for infinite many values of k (the x_k's could initially be all in one side then infinitely many in the other, or the x_k's could jump back and forth so both would have infinitely many).
Let I_2 denote the one with infinitely many. That is $x_k \in I_2$ for infinitely many x_k's. Note that
$$I_2 \subseteq I_1$$
$$I_2 = [a, b] = [a, \frac{a+b}{2}] \cup [\frac{a+b}{2}, b]$$
Again, at least one of these halves contains infinitely many x_k's. Let I_3 be that one.
Keep subdividing in this way and choosing a half which contains x_k for infinitely many k's. We have
$$\text{length } I_1 = 2M$$
$$\text{length } I_2 = M$$
$$\text{length } I_3 = \frac{M}{2}$$
$$\vdots$$
$$\text{length } I_l = \frac{2M}{2^{l-1}}$$
moreover,
$$I_1 \supseteq I_2 \supseteq \ldots \supseteq I_e \supseteq I_{e+1} \supseteq \ldots$$
and each I_l contains x_k for infinitely many values of k.
We can thus choose some $x_{k_1} \in I_1, x_{k_2} \in I_2, \ldots, x_{k_l} \in I_l \quad \forall l \in \mathbb{N}$ where $1 \leq k_1 < k_2 < \ldots < k_e < k_{e+1} < \ldots$. This is possible since there are infinitely many x_k's in each interval.
We claim:

1. $$\bigcap_{l=1}^{\infty} I_l \neq \varnothing$$

   and in fact contains **exactly one point** x.

   Note that
   $$I_l = [a_l, b_l] \quad \text{for some } a_l < b_l$$

and also

$$I_l \supset I_{l+1} \Rightarrow a_1 \leq a_l \leq a_{l+1} < b_{l+1} \leq b_l \leq b_1 \quad \forall l$$

(i.e. either endpoints move inwards for each successive interval).
So (a_l) is an increasing sequence bounded by b_1, therefore $\exists a$ such that $\lim_{l\to\infty} a_l = a$ and $a_l \leq a \leq b_1 \quad \forall l$.
Similarly (b_l) is a decreasing sequence bounded by a_1, so $\exists b$ such that $\lim_{l\to\infty} b_l = b$ and $a_1 \leq b \leq b_l \quad \forall l$.
We have $a_l < b_l \quad \forall l$. Taking the limit we have $a \leq b$ (limit can only be guaranteed with potential for equality).

$$a_1 \leq a_l \leq a_{l+1} \leq a \leq b \leq b_{l+1} \leq b_l \leq b_1$$

Note that

$$\begin{aligned} 0 \leq b - a \leq b_l - a_l &= length(I_l) \\ &= \frac{2M}{2^{l-1}} \to 0 \text{ as } l \to \infty \end{aligned}$$

hence $a = b$ (call this x).
By construction $x = a = b \in [a_l, b_l] = I_l \quad \forall l$ so

$$x \in \bigcap_{l=1}^{\infty} I_l$$

so there exists an element. Suppose $y \in \bigcap_{l=1}^{\infty} I_i$ then $x, y \in I_l \quad \forall l$ and

$$|x - y| \leq \frac{2M}{2^{l-1}} \quad \forall l \Rightarrow x = y \quad (\text{as } l \to \infty)$$

2.
$$\lim_{l\to\infty} x_{k_l} = x$$

Assume $x_{k_l} \in I_l$ and $x \in I_l \quad \forall l$ (from claim 1). So

$$|x_{k_l} - x| \leq \frac{2M}{2^{l-1}} \to 0 \text{ as } l \to \infty$$

thus $\lim_{l\to\infty} x_{k_l} = x$.

The above two claims prove the theorem for $n = 1$.
Now suppose the thoerem is true for n, we show it is true for $n + 1$.
Let (x_k) be a bounded sequence in $\mathbb{R}^{n+1}$, so $\exists M$ such that $\|x_k\| \leq M \quad \forall k$.
We write $x_k = (x_k^1, x_k^2, \ldots, x_k^{n+1})$ where x_k^j is the j-th component of vector $x_k \in \mathbb{R}^{n+1}$.
So

$$\|x_k\|^2 = |x_k^1|^2 + |x_k^2|^2 + \ldots + |x_k^n|^2 + |x_k^{n+1}|^2 \leq M^2 \tag{5.1}$$

Define a sequence (y_k) in $\mathbb{R}^n$ as the first n components of x_k

$$y_k = (x_k^1, \ldots, x_k^n)$$

therefore $\|y_k\| \leq M \quad \forall k$ by equation 5.1.
By the inductive hypothesis, $\exists$ a subsequence (y_{k_l}) of (y_k) that converges to some point $x = (x^1, \ldots, x^n) \in \mathbb{R}^n$.

Consider the sequence $(x_{k_l}^{n+1})$ in $\mathbb{R}^1$ **(TODO(richardwu): why can't we just use (x_k^{n+1}) here instead?)**. By equation 5.1, $|x_{k_l}^{n+1}| \leq M \quad \forall l$, so $(x_{k_l}^{n+1})$ is a bounded sequence in $\mathbb{R}$. By B-W for $n = 1$, $\exists$ subsequence $(x_{k_{l_j}}^{n+1})$ that converges to some $x^{n+1} \in \mathbb{R}$.
Consider the subsequence $(y_{k_{l_j}})$ of (y_{k_l}), which also converges to $(x^1, \dots, x^n) \in \mathbb{R}^n$.
So $x_{k_{l_j}}^a \to x^a$ as $j \to \infty$ for $a = 1, \dots, n$ and $a = n+1$.
Thus the sequence $x_{k_{l_j}} \to x$ as $j \to \infty$.

Remark 5.2. We used the IH/B-W on the first n components and then the $n+1$ component to find corresponding convergent subsequences. In order to "meld" them together, we needed to take the subsequence of either subsequence (to have a 2-level subsequence) to ensure it converges for the same k_{l_j}'s as the other 1-level subsequence.
TODO(richardwu): see the above TODO for why we don't just take k_l's instead of k_{l_j}'s.

□

6 Janaury 15, 2017

6.1 Connectedness

Definition 6.1. Let E be a non-empty subset of $\mathbb{R}^n$.
We say E is **disconnected** if there exists a **separation** for E. A separation of E is a pair U, V open sets in $\mathbb{R}^n$ such that

1. $E \cap U \neq \varnothing$
2. $E \cap V \neq \varnothing$
3. $E \cap U \cap V = \varnothing$
4. $E \subseteq U \cup V$

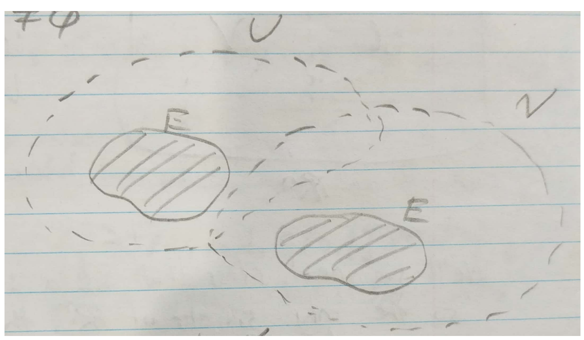

Figure 6.1: E is disconnected since there are open sets U, V that form a separation.

Note that $U \cap V$ need not be empty, but it must be disjoint from E.
(intuitively a set is disconnected if it is more than one piece).

Definition 6.2. E is **connected** if $\nexists$ any separation of E.

Remark 6.1. Connectedness and disconnectedness do not apply to $\varnothing$.

6.2 Is R^n connected?

(Yes it is).
Suppose $\exists$ a separation of $\mathbb{R}^n$ of open sets U, V such that

1. $$\varnothing \neq U \cap \mathbb{R}^n = U$$
$$\varnothing \neq V \cap \mathbb{R}^n = V$$

   which implies U, V both non-empty. Furthermore

2. $$U \cap V \cap \mathbb{R}^n = U \cap V = \varnothing$$

   which implies U, V are disjoint.

3. $$\mathbb{R}^n \subseteq U \cup V \subseteq \mathbb{R}^n$$

   so $\mathbb{R}^n = U \cup V$. Since $U \cap V = \varnothing$, then $U^c = V$ and $V^c = U$.

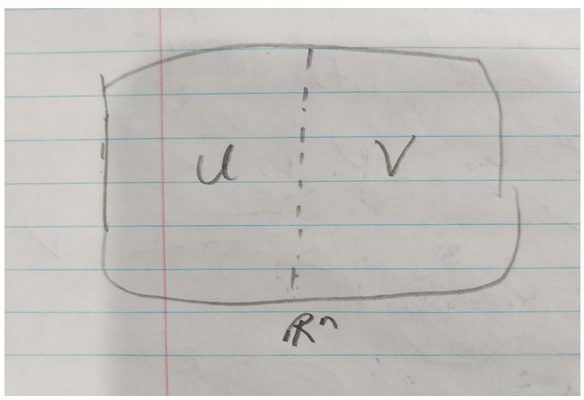

Figure 6.2: Sketch of what disconnected $\mathbb{R}^n$ would look like.

This would mean U, V are both **non-empty** subsets that are **both open and closed** and $U, V \neq \mathbb{R}^n$ (since they are non-empty disjoint).
In other words, if $\exists U$ such that $U \neq 0, U \neq \mathbb{R}^n$ and U is both open and closed, then $U, V = U^c$ gives a separation of $\mathbb{R}^n$.
We'll see (next class) that $\nexists$ a separation of $\mathbb{R}^n$ for any n, so the only subsets of $\mathbb{R}^n$ that are both open and closed are $\varnothing, \mathbb{R}^n$.

6.3 $[0, 1]$ is connected

This is an example of a connected subset in $\mathbb{R}$ and will be used next time to prove $\mathbb{R}^n$ is connected and *more*.

Theorem 6.1. Let $E = [0, 1] \subseteq \mathbb{R}$. Then E is connected.
(Aside: in fact: *any* interval $[a, b], [a, b), (a, b], (a, b)$ in $\mathbb{R}$ is connected and these are the **only** connected subsets in $\mathbb{R}$ i.e. connectedness $\Rightarrow$ interval).

Proof. By contradiction.
Suppose $[0,1]$ is not connected. $\exists$ a separation U, V open subsets of $[0,1]$ where

1. $U \cap [0,1] \neq \varnothing$
2. $V \cap [0,1] \neq \varnothing$
3. $U \cap V \cap [0,1] = \varnothing$
4. $[0,1] \subseteq U \cup V$

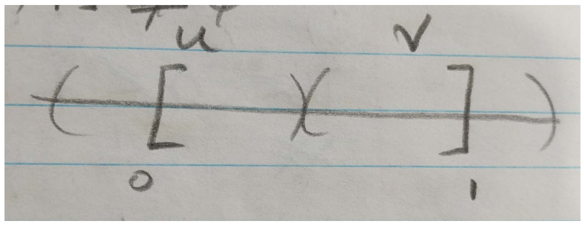

Figure 6.3: Sketch of U, V open sets as (potential) separation for $[0,1]$.

WLOG $0 \in U$. Since U is open and $0 \in U$, $\exists \epsilon_0 > 0$ such that $B_{\epsilon_0}(0) = (-\epsilon_0, \epsilon_0) \subseteq U$.
WLOG, $\epsilon_0 < 1$ so $[0, \epsilon_0) \subseteq U \cap [0,1]$.
Define t_0 as

$$\sup\{\epsilon \in (0,1) \mid [0,\epsilon) \subseteq U \cap [0,1]\}$$

note: the above is a non-empty subset of $\mathbb{R}$ since ϵ_0 is in the set. It's bounded above by 1, so the supremum or t_0 must exist.
We have $0 < \epsilon_0 \leq t_0 \leq 1$ so $t_0 \in (0,1]$, thus $t_0 \in U$ or $t_0 \in V$.

Case 1: $t_0 \in U$ Since U is open (all open sets have some open ball around every point) $\exists \delta > 0$ such that

$$(t_0 - \delta, t_0 + \delta) \subseteq U \tag{6.1}$$

WLOG $\delta < t_0$ but $0 < t_0 - \delta < t_0$ so by definition of t_0 (as supremum), $\exists \hat{\epsilon} > 0$ with $t_0 - \delta < \hat{\epsilon} < t_0$ such that

$$[0, \hat{\epsilon}) \subseteq U \cap [0,1] \tag{6.2}$$

Combining equation 6.1 and 6.2 (joining the two intervals together since we do not know if either separately are in U), we have

$$[0, t_0 + \delta) \subseteq U \cap [0,1] \tag{6.3}$$

We have two subcases:

$t_0 < 1$ Then we can shrink $\delta > 0$ further to ensure $t_0 + \delta < 1$ ($\delta < 1 - t_0$).
Then $0 < t_0 + \delta < 1$ and $[0, t_0 + \delta) \subseteq U \cap [0,1]$ which contradicts t_0 as the supremum.

$t_0 = 1$ This implies $U \cap [0,1] = [0,1]$ by equation 6.3 but then $V \cap [0,1] = \varnothing$ (since $U \cap V \cap [0,1] = \varnothing$), which is a contradiction since V must be non-empty.

Case 2: $t_0 \in V$ Since V is open $\exists \zeta > 0$ such that

$$(t_0 - \zeta, t_0 + \zeta) \subseteq V \tag{6.4}$$

WLOG $\zeta < t_0$ but $0 < t_0 - \zeta < t_0$ so by definition of t_0 (as supremum) $\exists \tilde{\epsilon} > 0$ with $t_0 - \zeta < \tilde{\epsilon} \leq t_0$ such that

$$[0, \tilde{\epsilon}) \subseteq U \cap [0, 1] \tag{6.5}$$

(it's U since that was the set t_0 was defined with).

Pick $s \in (t_0 - \zeta, \tilde{\epsilon})$. Then $s \in U \cap [0, 1]$ by equation 6.5 but also $s \in V \cap [0, 1]$ by equation 6.4, which is a contradiction.

By the contradiction of the two cases above, $[0, 1]$ is connected. □

7 January 17, 2017

7.1 Convex sets

Definition 7.1. A **non-empty** subset E of $\mathbb{R}^n$ is called **convex** if whenever $x, y \in E$ then

$$tx + (1 - t)y \in E \quad \forall t \in [0, 1]$$

i.e. the line segment between any 2 points in E is contained in E.

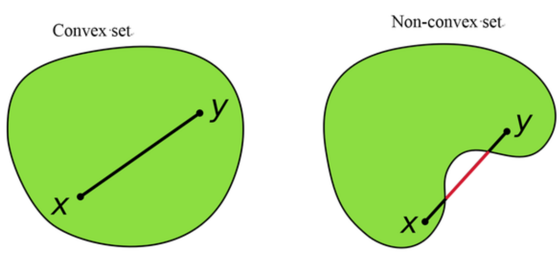

Figure 7.1: Convex and non-convex sets in $\mathbb{R}^2$.

7.2 Convex ⇒ connected

Corollary 7.1. Any convex subset E of $\mathbb{R}^n$ is connected. This implies two corollaries:

Corollary 7.2. $\mathbb{R}^n$ is connected $\forall n$ since $\mathbb{R}^n$ is trivially convex.

Corollary 7.3. The only subsets of $\mathbb{R}^n$ that are both open and closed are $\varnothing, \mathbb{R}^n$ (see the remark about $\mathbb{R}^n$ connectedness from above).

Proof. Let E be convex and suppose E is *not* connected. $\exists$ open subsets U, V such that

1. $U \cap E \neq \varnothing$
2. $V \cap E \neq \varnothing$
3. $U \cap V \cap E = \varnothing$
4. $E \subseteq U \cup V$

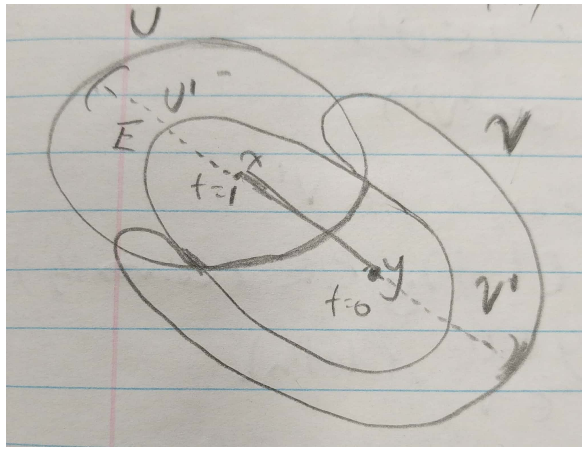

Figure 7.2: Suppose convex E is not connected and there exists a separation U, V.

Let $x \in U \cap E$ and $y \in V \cap E$ (therefore $x \neq y$ since $U \cap V \cap E = \varnothing$). Since E is convex,

$$tx + (1-t)y \in E \quad \forall t \in [0,1]$$

Define U', V' subsets of $\mathbb{R}^n$ by

$$U' = \{t \in \mathbb{R} : tx + (1-t)y \in U\} \qquad V' = \{t \in \mathbb{R} : tx + (1-t)y \in V\}$$

(note: U', V' is not restricted to elements $[0,1]$: t could extend arbitrarily into E^c).
Claim: U', V' are open subsets of $\mathbb{R}$. Let $t_0 \in U'$ so $x_0 = t_0 + (1-t_0)y \in U$. Since U is open in $\mathbb{R}^n$ $\exists \epsilon_0 > 0$ such that $B_{\epsilon_0}(x_0) \in U$. We pick $t \in \mathbb{R}$ such that

$$|t - t_0| < \frac{\epsilon_0}{\|x\| + \|y\|}$$

then

$$\begin{aligned} B_{\epsilon_0}(x_0) \Rightarrow \|(tx + (1-t)y) - x_0\| &= \|tx + (1-t)y - t_0 x - (1-t_0)y\| \\ &= \|(t-t_0)x + (t_0 - t)y\| \\ &\overset{\triangle}{\leq} |t - t_0|(\|x\| + \|y\|) \\ &< \epsilon_0 \end{aligned}$$

But $B_{\epsilon_0}(x_0) \subseteq U$ so if $|t - t_0| < \frac{\epsilon_0}{\|x\|+\|y\|}$ then $t \in U'$ (we want our choice of t to imply $t \in U'$).
So $\frac{B_{\epsilon_0}(t_0)}{\|x\|+\|y\|} \subseteq U'$ so U' is open.
Similarly, V' is open.
Thus here are the properties of U', V'. They are both open in $\mathbb{R}$ and

1. $U' \cap [0,1] \neq \varnothing$ (since $1 \in U'$)
2. $V' \cap [0,1] \neq \varnothing$ (since $0 \in V'$)
3. $U' \cap V' \cap [0,1] = \varnothing$

   Given some $t \in [0,1]$ (since $tx \in (1-t)y \in E$ from convexity), note that either $t \in U'$ from $tx + (1-t)y \in U$ or $t \in V'$ from $tx + (1-t)y \in V$ (we know from before that $U \cap V \cap E = \varnothing$ thus this must hold for the subsets

U', V').

4. $[0,1] \subseteq U' \cup V'$

   If $t \in [0,1]$, then $z = tx + (1-t)y \in E$ so $z \in U \cup V$ from before, so $z \in U$ or $z \in V$, thus by their definitions $t \in U'$ or $t \in V'$.

Then U', V' is a separation for $[0,1]$, which is a contradiction. Thus E is connected. □

Remark 7.1. In general, to prove a set E is connected it is generally easier to assume it is *not* connected and there exists a separation, then derive a contradiction.

7.3 Open cover and compactedness

Definition 7.2. Let E be a subset of $\mathbb{R}^n$. An **open cover** of E is a collection of open subsets U_α $\alpha \in A$ of $\mathbb{R}^n$ such that

$$E \subseteq \bigcup_{\alpha \in A} U_\alpha$$

(finite or infinite union of open subsets).

Definition 7.3. The subset E is called **compact** *iff* every open cover of E admits a **finite subcover**.
That is: if $\bigcup U_\alpha$ $\alpha \in A$ is an open cover of E, then $\exists$ a finite subset A_0 of A such that

$$E \subseteq \bigcup_{\alpha \in A_0} U_\alpha$$

Informally, whenever a compact E is covered by a collection of open sets, it is actually covered by just finitely many of those open sets.

Remark 7.2. This definition is not very useful for checking if a subset is compact (because you would have to check every open cover of E).

7.4 Bounded sets

Definition 7.4. A subset $E \subseteq \mathbb{R}^n$ is called **bounded** if $\exists M > 0$ such that $E \subseteq \overline{B_M(0)}$. That is $\|x\| \leq M$ $\forall x \in E$.

8 January 19, 2018

8.1 Heine-Borel theorem

Theorem 8.1. Let E be a subset of $\mathbb{R}^n$. E is **compact** *iff* E is both **closed and bounded**.
The following proof uses the *density of rationals*.

Proof. **Step 1:** Suppose E is *compact*. We want to show that E is *bounded*.
Let $U_k = B_k(0) = \{x \in \mathbb{R}^n \mid \|x\| < k\}$. U_k is open $\forall k$, thus $U_k \subseteq U_{k+1}$ $\forall k \in \mathbb{N}$. Therefore

$$E \subseteq \bigcup_{k=1}^{\infty} U_k = \mathbb{R}^n$$

$\{U_k, k \in \mathbb{N}\}$ is an open cover of E. Since E is compact, $\exists$ a finite subcover so $\exists\ k_1 < k_2 < \ldots < k_N \in \mathbb{N}$ such that

$$E \subseteq \bigcup_{j=1}^{N} U_{k_j} = U_{k_N} = B_{k_N}(0)$$

since they're nested so E is bounded.

Corollary 8.1. $\mathbb{R}^n$ is not compact because it is not bounded.

Step 2: Let E be *compact*, we show it is *closed*.
To do this: we show E^c is open (aside if $E^c = \varnothing$ then we are done. This never happens since E is not $\mathbb{R}^n$).
Let $x \in E^c$. We need to find an open ball centred at x lying completely in E^c. Note $E \subseteq \mathbb{R}^n \setminus \{x\}$ since $x \notin E$.
Let (different U_k from before)

$$U_k = (\overline{B_{\frac{1}{k}}(x)})^c = \{x \in \mathbb{R}^n \mid \|x - y\| > \frac{1}{k}\}$$

which is open (complement of closed ball). We can use this as covers.

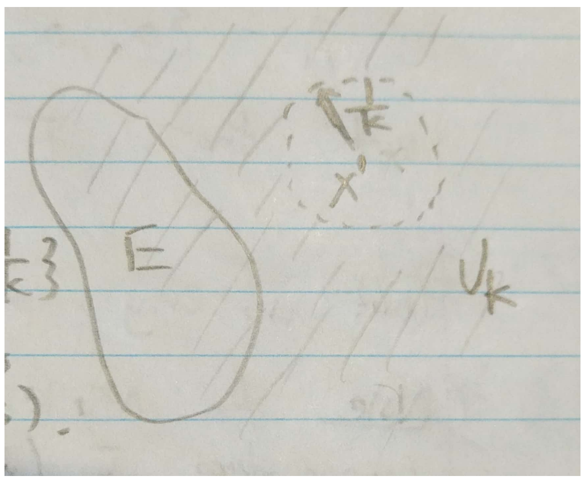

Figure 8.1: U_k is the complement of the closed ball centred at $x \in E^c$ with radius $\frac{1}{k}$ for some $k \in \mathbb{N}$.

If $l > k$, then $\frac{1}{l} < \frac{1}{k}$. Thus if $y \in U_k$, then $\|y - x\| > \frac{1}{k} > \frac{1}{l}$ so $y \in U_l$. That is

$$U_k \subseteq U_l \quad k < l \tag{8.1}$$

Note that we have

$$E \subseteq \mathbb{R}^n \setminus \{x\} = \bigcup_{k=1}^{\infty} U_k$$

where the infinite union of U_k is an open cover of E. Since E is compact, we have a finite subcover $U_{k_1}, \ldots, U_{k_N}$ such that

$$\begin{aligned} E &\subseteq \bigcup_{j=1}^{N} U_{k_j} \\ &= U_{k_N} && \text{equation 8.1} \\ &= (\overline{B_{\frac{1}{k_N}}(0)})^c \end{aligned}$$

Take complements (from $A \subseteq B \Rightarrow B^c \subseteq A^c$)

$$x \in B_{\frac{1}{k_N}}(x) \subseteq \overline{B_{\frac{1}{k_N}}(x)} \subseteq E^c$$

So $\exists$ an open ball for x thus E^c is open and E is closed.
Before we prove the converse:

Lemma 8.1. Let E be any subset of $\mathbb{R}^n$. Let $\{U_\alpha \mid \alpha \in A\}$ be an open cover of E (so U_α open $\forall \alpha \in A$). That is

$$E \subseteq \bigcup_{\alpha \in A} U_\alpha$$

Thus $\exists$ a countable subset $\tilde{A}$ of A

$$\tilde{A} = \{\alpha_1, \alpha_2, \ldots\} = \{\alpha_k \mid k \in \mathbb{N}\}$$

such that $E \subseteq \bigcup_{k=1}^{\infty} U_{\alpha_k}$.
That is: **every open cover admits a countable subcover**. (Note: an infinite set is countable **iff** $\exists$ bijective correspondence with $\mathbb{N}$. Rational numbers are countable whereas $\mathbb{R}$ is not).

Proof. Assume $E \subseteq \bigcup_{\alpha \in A} U_\alpha$.
Let $x \in E$. Then $\exists \alpha(x) \in A$ such that $x \in U_{\alpha(x)}$.
Since $U_{\alpha(x)}$ is open $\exists \epsilon(x) > 0$ such that

$$B_{\epsilon(x)}(x) \subseteq U_{\alpha(x)} \tag{8.2}$$

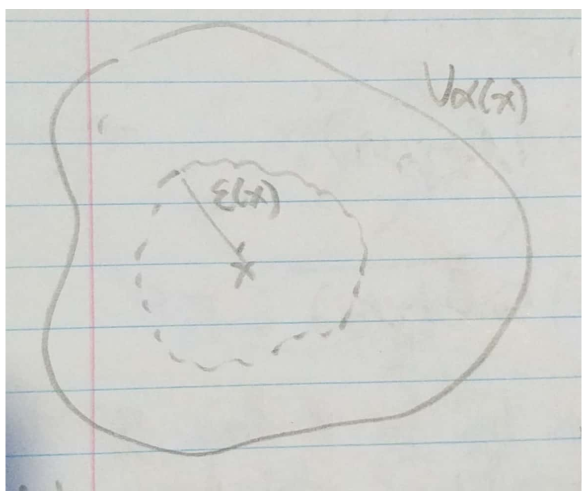

Figure 8.2: We can construct an open ball $B_{\epsilon(x)}(x)$ within some $U_{\alpha(x)}$ for every $x \in E$.

Then $E \subseteq \bigcup_{x \in E} B_{\epsilon(x)}(x)$ (all x). Since the rational numbers $\mathbb{Q}$ are dense in $\mathbb{R}$, $\exists q(x) \in \mathbb{Q}^n$ and $\zeta(x) \in \mathbb{Q}$ such that

$$\|x - q(x)\| < \frac{\epsilon(x)}{4}$$
$$\frac{\epsilon(x)}{4} < \zeta(x) < \frac{\epsilon(x)}{2}$$

Therefore

$$\|x - q(x)\| < \frac{\epsilon(x)}{4} < \zeta(x)$$

So $x \in B_{\zeta(x)}(q(x))$.

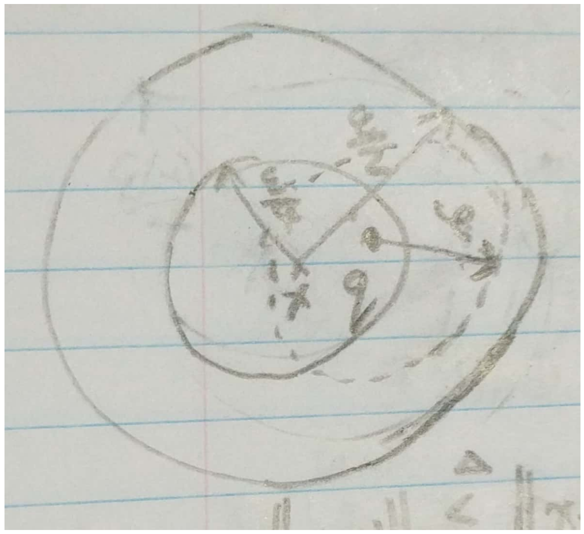

Figure 8.3: We find some open ball centred at $q(x) \in \mathbb{Q}^n$ with radius $\zeta(x) \in \mathbb{Q}$ that contains x.

Suppose $y \in B_{\zeta(x)}(q(x))$, then we have

$$\begin{aligned}
\|x-y\| &\overset{\triangle}{\leq} \|x-q(x)\| + \|q(x)-y\| \\
&< \frac{\epsilon(x)}{4} + \zeta(x) \\
&< \frac{\epsilon(x)}{4} + \frac{\epsilon(x)}{2} \\
&< \epsilon(x)
\end{aligned}$$

So $y \in B_{\epsilon(x)}(x)$ therefore $B_{\zeta(x)}(q(x)) \subseteq B_{\epsilon(x)}(x)$.
We have shown for every $x \in E$, $\exists q(x) \in \mathbb{Q}^n$ and $\zeta(x) \in \mathbb{Q}$ such that

$$x \in B_{\zeta(x)}(q(x)) \subseteq B_{\epsilon(x)}(x)$$

So $E \subseteq_{x \in E} B_{\zeta(x)}(q(x))$ but $\mathbb{Q}$ and $\mathbb{Q}^n$ are countable so $\exists q_j \in \mathbb{Q}^n$ and $\zeta_j \in \mathbb{Q}$ where $j \in \mathbb{N}$ such that

$$\begin{aligned}
E &\subseteq \bigcup_{j=1}^{\infty} B_{\zeta_j}(q_j) \\
&\subseteq \bigcup_{j=1}^{\infty} B_{\epsilon(x_j)}(x_j) \\
&\subseteq \bigcup_{j=1}^{\infty} U_{\alpha}(x_j) && \text{by equation 8.2}
\end{aligned}$$

so we have a countable subcover. □

Back to proving the converse: **Step 3:** Let E be *closed and bounded.* We want to show E is *compact.*

Let $\{U_\alpha \mid \alpha \in A\}$ be an open cover of E. We showed in the above lemma that $\exists q_j \in \mathbb{Q}^n$, $\zeta_j \in \mathbb{Q}$ where $j \in \mathbb{N}$ such that

$$E \subseteq \bigcup_{j=1}^{\infty} B_{\zeta_j}(q_j)$$

Claim: $\exists N \in \mathbb{N}$ such that $E =\subseteq \bigcup_{j=1}^{N} B_{\zeta_j}(q_j)$. (i.e. we need only need finitely many of these balls).
If the claim is true, then

$$\begin{aligned} E &\subseteq \bigcup_{j=1}^{N} B_{\zeta_j}(q_j) \\ &\subseteq \bigcup_{j=1}^{N} B_{\epsilon(x_j)}(x_j) \\ &\subseteq \bigcup_{j=1}^{N} U_\alpha(x_j) \end{aligned}$$

so we would have a finite subcover.
It remains to prove the claim. Suppose the claim is false (proof by contradiction). Then for every $k \in \mathbb{N}$, we have $E \setminus \bigcup_{j=1}^{k} B_{\zeta_j}(q_j) \neq \varnothing$.
We choose $x_k \in E \setminus \bigcup_{j=1}^{k} B_{\zeta_j}(q_j)$. Note that (x_k) is a sequence in E and E is bounded, so (x_k) is a bounded sequence.
By Bolzano-Weierstrass, there exists a subsequence (x_{k_l}) that converges.
$(x_{k_l}) \in E\ \forall l \in \mathbb{N}$ and E is closed so $x = \lim_{l\to\infty} x_{k_l} \in E$ as well.
$x \in E$, so $\exists$ some $J \in \mathbb{N}$ such that

$$x \in B_{\zeta_J}(q_J) \tag{8.3}$$

from our lemma.
But since $\lim_{l\to\infty} x_{k_l} = x$ then $\exists N \in \mathbb{N}$ such that $\forall l \geq N$

$$x_{k_l} \in B_{\zeta_j}(q_j)$$

(definition of convergent sequence).
But by construction of our sequence

$$x_{k_l} \notin \bigcup_{j=1}^{N} B_{\zeta_j}(q_j)$$

for any $k_l \geq N$.
So if $k_l > J$, then

$$x_{k_l} \notin B_{\zeta_J}(q_J) \tag{8.4}$$

for $l \geq \max(N, J) \Rightarrow k_l \geq \max(N, J)$.
From equation 8.3 and equation 8.4, we have a contradiction.
(The idea of this proof revolves around showing that all $x \in E$ must be in some open ball with rational parameters. By assuming the contrary of the claim that there is a finite subcover, we choose some sequence outside of all finite subcovers (made of the rational parameters) and show that it is not in an open ball with rational parameters. Thus we have a contradiction so there must be some finite subcover with the open balls of rational parameters). $\square$

9 January 22, 2018

9.1 Limits of functions

Definition 9.1. Let $V \subseteq \mathbb{R}^n$ be an *open set* with $x_0 \in V$. Let $f : V \setminus \{x_0\} \to \mathbb{R}^m$ for some m (i.e. f is defined at all points of V except *possibly* at x_0).
We say $\lim_{x \to x_0} f(x)$ exists and equals $L \in \mathbb{R}^m$ **iff** $\forall \epsilon > 0$, $\exists \delta > 0$ such that

$$0 < \|x - x_0\| < \delta \Rightarrow \|f(x) - L\| < \epsilon$$

(note that $B_\delta(x_0) \subseteq V$ must hold). In general, δ depends on both ϵ and x_0 (and on f as well if suppose it has a different domain).

Remark 9.1. When $n > 1$, things get more complicated since in $n = 1$, there exists only 2 ways to approach x_0: from left or right (i.e. $\lim_{x\to\infty} f(x)$ exists *iff* both left and right limits exist and are equal in $n = 1$).
In $n > 1$, $\exists$ *infinitely many ways* to approach x_0. This is what makes establishment of the existence of limits harder for $n > 1$.

Example 9.1. Example where different linear paths result in a different limit:
Let $n = 2, m = 1$ ($f : \mathbb{R}^2 \to \mathbb{R}$) where we denote $(x, y) \in \mathbb{R}^2$.
Suppose we wish to find

$$\lim_{(x,y)\to(2,3)} \frac{(x-2)^2}{(x-2)^2 + (y-3)^2}$$

where $f(x, y)$ defined everywhere except $(2, 3)$.

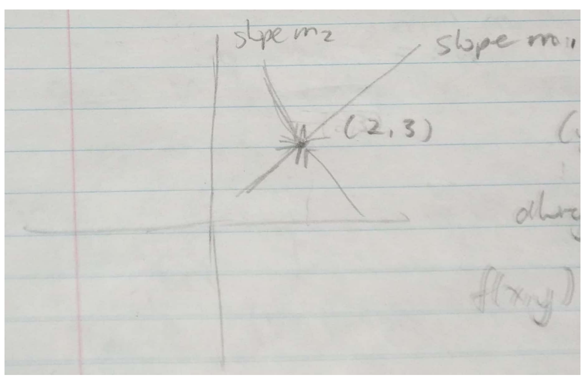

Figure 9.1: There exists many paths to approach $x_0 = (2, 3)$ in f, $m_1 \neq m_2$.

Suppose we have paths/lines with slope m where $(y - 3) = m(x - 2)$. Along this line we have

$$\begin{aligned} f(x,y) &= \frac{(x-2)^2}{(x-2)^2 + (y-3)^2} \\ &= \frac{1}{1+m^2} \end{aligned}$$

So f is a constant function which depends on the slope of the line/path (it depends on m). Since we found at least 2 paths towards $(2, 3)$ along which f approaches different limiting values, then the limit **DNE**.

Example 9.2. Example where linear paths converge but quadratic paths do not:

We wish to find

$$\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^4}$$

where the domain is $\mathbb{R}^2 \setminus \{0,0\}$.

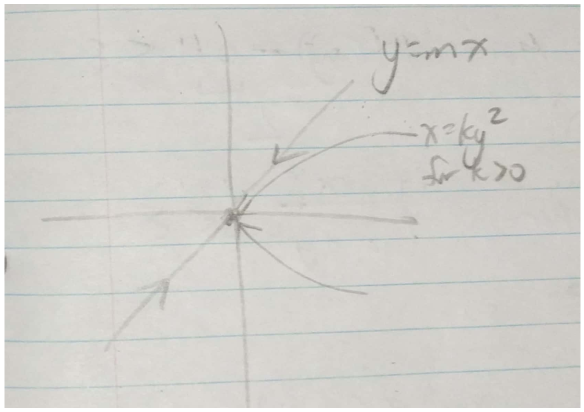

Figure 9.2: There are linear and non-linear paths that approaches $x_0 = (0,0)$.

Note that unlike previously, linear paths $y = mx$ do converge

$$\begin{aligned} f(x,y) &= \frac{x(mx)^2}{x^2+(mx)^4} \\ &= \frac{m^2x^3}{x^2+m^4x^4} \\ &= \frac{m^2x}{1+m^4x^2} \end{aligned}$$

So as $x \to 0$, then $\frac{m^2x}{1+m^4x^2} \to 0$ for any m.
Along $x = 0$ we still have

$$f(x,y) = \frac{0 \cdot y^2}{0^2+y^4} = 0 \quad \forall y \neq 0$$

(this is important since a vertical line is not explicit). So f approaches 0 as $(x,y) \to (0,0)$ for linear paths.
We must consider other non-linear paths as well e.g. along $x = ky^2$ we have

$$f(x,y) = \frac{(ky^2)y^2}{(ky^2)^2+y^4} = \frac{k}{k^2+1}$$

which is a constant that depends on k, thus the limit **DNE**.

Example 9.3. Example where the limit does exist in $n > 1$ space:
We wish to find

$$\lim_{(x,y)\to(0,0)} \frac{x^4}{x^2+y^2}$$

We *expect the limit to converge* since the degree of the numerator is > degree of denominator, thus numerator $\to 0$ "much faster" than the denominator so the quotient should go to zero.
Let $\epsilon > 0$. We want to find $\delta > 0$ (depends on ϵ) such that if

$$\|(x,y) - (0,0)\| < \delta \Rightarrow \|f(x,y) - 0\| < \epsilon$$

where $L = 0$.
Rewriting the above: we have if $x^2 + y^2 < \delta^2$, then

$$|\frac{x^4}{x^2+y^2}| < \epsilon$$

Observe that $x^2 \le x^2 + y^2$ so

$$\frac{x^2}{x^2+y^2} \le 1 \quad (x,y) \ne (0,0)$$

Furthermore note

$$\begin{aligned} |\frac{x^4}{x^2+y^2}| = \frac{x^4}{x^2+y^2} &= x^2 \left(\frac{x^2}{x^2+y^2}\right) \\ &\le x^2 && \frac{x^2}{x^2+y^2} \le 1 \\ &\le x^2 + y^2 \\ &< \delta^2 = \epsilon \end{aligned}$$

Thus we can take $\delta = \sqrt{\epsilon}$ such that

$$x^2 + y^2 < \delta^2 = \epsilon \Rightarrow |f(x,y) - 0| < \epsilon$$

9.2 Uniqueness of limits

Remark 9.2. A given limit may not exist, but if it does it's **unique** (same proof as uniqueness of limits of sequences).

9.3 Sequential characterization of limits of functions

Proposition 9.1. For $f : V \setminus \{x_0\} \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $\lim_{x \to x_0} f(x) = L$ **iff** the sequence $f(x_k)$ converges to L for every sequence (x_k) in $V \setminus \{x_0\}$ converging to x_0.
i.e. this states the path heuristic from before works formally with sequences too.

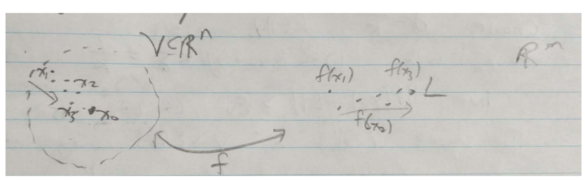

Figure 9.3: Any given sequence (x_k) in V that converges to x_0 must have $f(x_k)$ converge to l in $\mathbb{R}^m$.

Proof. **Forwards:** Suppose $\lim_{x \to x_0} f(x) = L$.
Let (x_k) be a sequence in $\mathbb{R}^n$ with $x_k \in V \setminus \{x_0\}$ $\forall k$ and $\lim_{k \to \infty} x_k = x_0$.
We need to show $\lim_{k \to \infty} f(x_k) = L$.
Let $\epsilon > 0$. From our premise $\exists \delta > 0$ such that

$$\|x - x_0\| < \delta \Rightarrow \|f(x) - L\| < \epsilon$$

Since $x_k \to x_0$ as $k \to \infty$, $\exists N \in \mathbb{N}$ such that

$$k \geq N \Rightarrow 0 < \|x_k - x_0\| < \delta$$

(definition of convergence).
So $k \geq N \Rightarrow \|f(x_k) - L\| < \epsilon$ so the forwards direction holds.
Backwards: Conversely, suppose the sequence $f(x_k)$ converges to L for every (x_k) in $V \setminus \{x_0\}$ converging to x_0. We want to show $\lim_{x\to x_0} f(x_0) = L$.
Suppose the limit does not converge to L (contradiction). Negation of the statement is: $\exists\epsilon_0 > 0$ such that $\forall\delta > 0$, $\exists x_\delta$ such that

$$0 < \|x_\delta - x_0\| < \delta \text{ but } \|f(x_\delta) - L\| \geq \epsilon_0$$

(this is the negation of 1) $\forall\epsilon > 0$, 2) $\exists\delta > 0$, and 3) $\forall x$ $\|x - x_0\| < \delta \Rightarrow \|f(x) - L\| < \epsilon$).
Choose $\delta = \frac{1}{k}$, $k \in \mathbb{N}$. $\exists x_k$ with

$$0 < \|x_k - x_0\| < \delta = \frac{1}{k} \tag{9.1}$$

but $\|f(x_k) - L\| \geq \epsilon_0$ (4).
The sequence (x_k) in $V \setminus \{x_0\}$ converges to x_0 by the premise but $f(x_k) \not\to L$ by equation 9.1. This contradicts the premise. □

9.4 Properties of limits of functions

Let $f, g : V \setminus \{x_0\} \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and suppose

$$\lim_{x\to x_0} f(x) = L \qquad \lim_{x\to x_0} g(x) = M$$

Then (the above limits **must exist**)

$$\begin{aligned}
&\lim_{x\to x_0} (f(x) + g(x)) = L + M && \text{(additive)}\\
&\lim_{x\to x_0} cf(x) = cL && \text{(scalar multiplicative)}\\
&\lim_{x\to x_0} \frac{f(x)}{g(x)} = \frac{L}{M} && \text{if } m = 1, M \neq 0\\
&\lim_{x\to x_0} (f(x)g(x)) = LM && \text{if } m = 1 \text{ (same proof as for n = 1)}
\end{aligned}$$

Proofs are left as exercises.

10 January 24, 2018

10.1 Component functions

Definition 10.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U is open. Then for $x \in U$

$$f(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m$$

$f_i : U \to \mathbb{R}$, $1 \leq i \leq m$ are the **component functions** of f (real-valued).

Lemma 10.1. $x_0 \in V$ open in $\mathbb{R}^n$. Let $f : V \setminus \{x_0\} \to \mathbb{R}^m$. Then $\lim_{x\to x_0} f(x) = L = (L_1, \dots, L_m)$ **iff** $\lim_{x\to x_0} f_i(x) = L_i$ $\forall i = 1, 2, \dots, m$.

Proof. By property of convergence of limits of sequences in assignment 3 and sequence characterization of limits of functions. That is

$$\lim_{x\to x_0} f(x) = L \overset{seq.char.}{\iff} \lim_{k\to\infty}(x_k) = L \overset{a3\#2}{\iff} \lim_{k\to\infty} f_i(x_k) = L_i \overset{seq.char.}{\iff} \lim_{x\to x_0} f_i(x) = L_i$$

We can also prove this using $\epsilon - \delta$. □

10.2 Squeeze theorem

Theorem 10.1. Suppose $f, g, h : V \setminus \{x_0\} \to \mathbb{R}$ ($m = 1!$). If $f(x) \leq g(x) \leq h(x)$ $\forall x \in V \setminus \{x_0\}$ (this only really needs to hold in a n'h'd of x_0) and $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} h(x) = L \in \mathbb{R}$, then

$$\lim_{x\to x_0} g(x) = L$$

Proof. Same as proof in $n = 1$ case. □

10.3 Norm properties of limits

Proposition 10.1. Suppose $f : V \setminus \{x_0\} \to \mathbb{R}^m$ and $\lim_{x\to x_0} f(x) = L$ then

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

Proof. Let $\epsilon > 0$, $\exists \delta > 0$ such that

$$0 < \|x - x_0\| < \delta \Rightarrow \|f(x) - L\| < \epsilon$$

Note that

$$\|f(x)\| \overset{\triangle}{\leq} \|f(x) - L\| + \|L\|$$
$$\|L\| \overset{\triangle}{\leq} \|L - f(x)\| + \|f(x)\|$$

So rearranging each of the inequalities above and using the premise we see that

$$|\|f(x)\| - \|L\|| < \epsilon$$

if $0 < \|x - x_0\| < \delta$ so $\lim_{x\to x_0} \|f(x)\| = \|L\|$. □

10.4 Continuity

Definition 10.2. Let $U \subseteq \mathbb{R}^n$ be *open* and $f : U \to \mathbb{R}^m$.
Let $x_0 \in U$. We say f is **continuous at** x_0 if

1. $\lim_{x\to x_0} f(x)$ exists
2. The limit equals $f(x_0)$

Explicitly, f is **continuous** at x_0 **iff** $\forall \epsilon > 0$ $\exists \delta > 0$ such that

$$\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| < \epsilon$$

Equivalently, by the sequential characterization of limits, f is continuous at x_0 iff whenever (x_k) is a sequence in U converging to x_0, then $f(x_k)$ is a sequence in $\mathbb{R}^m$ converging to $f(x_0)$.

10.5 Continuity on a set

Definition 10.3. f is **continuous on** U (an open set) if it is continuous at every $x \in U$.

Example 10.1. $n = m$ and $U = \mathbb{R}^n$ and $f(x) = x$ (identity map). Then $\forall \epsilon > 0$, let $\delta = \epsilon > 0$ such that

$$\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| = \|x - x_0\| < \delta = \epsilon$$

Example 10.2. Let $c \in \mathbb{R}^m$ be fixed, $U = \mathbb{R}^n$. Then $f(x) = c$ is a constant function and is continuous on $\mathbb{R}^n$ since $\forall \epsilon > 0$, take *any* $\delta > 0$ we have

$$\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0)\| = \|c - c\| = 0 < \epsilon$$

Example 10.3. Let $m = 1$, $U = \mathbb{R}^n$, and $f(x, y) = xy$ for $x = (x_1, x_2) = (x, y) \in \mathbb{R}^2$.
We claim $f(x)$ is continuous on $\mathbb{R}^n$.
Before we prove this example, for the *component functions*:

Remark 10.1. If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, f is continuous at $x_0 \in U$ **iff** $f_i : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is continuous at x_0 for all $i = 1, \ldots, n$.

Proof. Let $h(x, y) = (x, y)$ (identity map, continuous. on $\mathbb{R}^2$ by example 10.1).
So $h_1(x, y) = x$ and $h_2(x, y) = y$ are continuous on $\mathbb{R}^2$.
$f(x, y) = xy = h_1(x, y)h_2(x, y)$ is continuous on $\mathbb{R}^2$ because limits of product equals products of limits. □

10.6 Composition of continuous functions is continuous

Proposition 10.2. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be continuous on U. Let $g : V \subseteq \mathbb{R}^m \to \mathbb{R}^p$ be continuous on V. Suppose $f(U) = \{f(x) \mid x \in U\} \subseteq V$ so the composition

$$h = g \circ f : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$$

is defined $g(f(x))$. Then $h = g \circ f$ is continuous on U.

Proof. Assignment 4. □

More generally, if f is continuous at $x_0 \in U$, $f(x_0) \in V$ and g is continuous at $f(x_0)$ then $h = g \circ f$ is continuous at x_0.

10.7 Dot product of continuous functions is continuous

Proposition 10.3. Suppose $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Define $f \cdot g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ by

$$(f \cdot g)(x) = f(x) \cdot g(x) = f_1(x)g_1(x) + f_2(x)g_2(x) + \ldots + f_m(x)g_m(x)$$

If f, g continuous at x_0, then $f \cdot g$ is continuous at x_0 (by addition and product of continuous functions on $\mathbb{R}$).

10.8 Inverse image

Definition 10.4. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U is open. Let $A \subseteq \mathbb{R}^m$.
The **inverse image** of A under f is denoted $f^{-1}(A)$ and is defined to be

$$f^{-1}(A) = \{x \in U \mid f(x) \in A\}$$

(i.e. the domain of f which maps to the image A).

Remark 10.2. The notation is a bit ambiguous. Suppose we *restrict* f to be a proper subset $V \subset U$ that is still open.
Call this $f_{|V} : V \subseteq \mathbb{R}^n \to \mathbb{R}^m$. So $f_{|V}(x) = f(x)\ \forall x \in V$ then if $A \subseteq \mathbb{R}^m$

$$f_{|V}^{-1}(A) = \{x \in V \mid f(x) \in A\} = f^{-1}(A) \cap V$$

Remark 10.3. **Note:** $f^{-1}(A)$ could be **empty**.

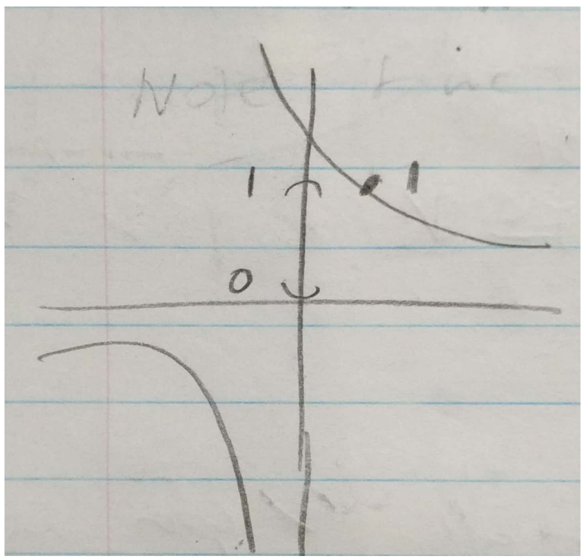

Figure 10.1: Inverse images of $f(x) = \frac{1}{x}$ for $A = (0, 1)$ correspond to $f^{-1}(A) = (1, \infty)$. It may be empty however (e.g. for $A = \{0\}$).

Let $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $f(x) = \frac{1}{x}$. Note that

$$f^{-1}((0, 1)) = (1, \infty)$$
$$f^{-1}(\{0\}) = \varnothing$$

11 January 26, 2018

11.1 Continuity and open/closed sets

Proposition 11.1. $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U is open. Then f is continuous on U **iff** $f^{-1}(V)$ is open in $\mathbb{R}^n$ whenever V is open in $\mathbb{R}^m$.
(that is: continuous iff the inverse image of any open set is open).

Proof. **Forwards:** Suppose f is continuous on U. Let $V \subseteq \mathbb{R}^m$ be open. WLOG $f^{-1}(V) \neq \varnothing$. Let $x_0 \in f^{-1}(V) \subseteq U \Rightarrow f(x_0) \in V$.
Since V is open, $\exists \epsilon > 0$ such that $B_\epsilon(f(x_0)) \subseteq V$. But f is continuous at x_0 so $\exists \delta > 0$ such that

$$\|x - x_0\| < \delta \Rightarrow \|f(x) - f(x_0\| < \epsilon$$

(we take δ small enough such that $B_\delta(x_0) \subseteq U$).
Thus $f(B_\delta(x_0)) \subseteq B_\epsilon(f(x_0)) \subseteq V$. So $B_\delta(x_0) \subseteq f^{-1}(V)$ hence $\forall x_0 \in f^{-1}(V)$ $\exists \delta > 0$ such that $B_\delta(x_0) \subseteq f^{-1}(V)$ so $f^{-1}(V)$ is open.

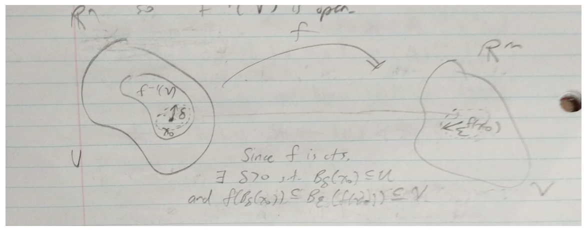

Figure 11.1: $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is continuous **iff** the inverse image of any open set $V \in \mathbb{R}^m$ is open.

Backwards: Suppose $f^{-1}(V)$ is open in $\mathbb{R}^n$ for all V open in $\mathbb{R}^m$. We need to show that f is continuous on U. Let $x_0 \in U$. Let $\epsilon > 0$, $B_\epsilon(f(x_0))$ is open in $\mathbb{R}^m$ so by our assumption

$$f^{-1}(V) = f^{-1}(B_\epsilon(f(x_0)))$$

is open in $\mathbb{R}^n$.
Also $x_0 \in f^{-1}(V)$ since $f(x_0) \in V = B_\epsilon(f(x_0))$ so $\exists \delta > 0$ such that

$$B_\delta(x_0) \subseteq f^{-1}(V)$$

(since $f^{-1}(V)$ is open).
Hence $f(B_\delta(x_0)) \subseteq V = B_\epsilon(f(x_0))$ so f is continuous at x_0. $\square$

Remark 11.1. One can also show that $f^{-1}(\textit{closed}) = \textit{closed}$ is also equivalent to continuity (on assignment 4).

Remark 11.2. Question: Is the reverse the open set property true? That is, suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ U open, f continuous on U. Let $V \subseteq U$ defined $f(V) = \{f(x) \mid x \in V\}$ the image of V under f. If V is open in $\mathbb{R}^n$, is $f(V)$ necessarily open in $\mathbb{R}^m$?
No: here's a counter-example.

Example 11.1. $n = m = 1$, $U = \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ (continuous on $\mathbb{R}$).
Take $V = (-1, 1)$ open in $\mathbb{R}$. Then $f(V) = [0, 1)$ which is not open in $\mathbb{R}$.

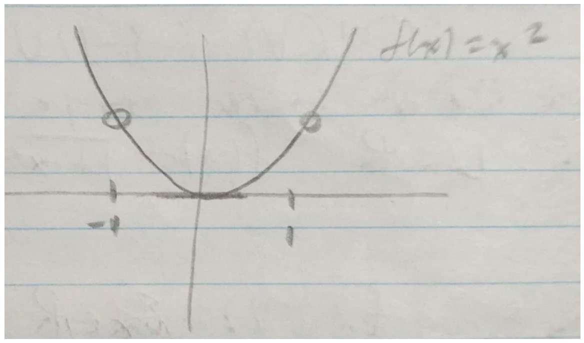

Figure 11.2: An open domain on a continuous $f(x) = x^2$ may not admit an open image.

Similarly, if V is closed in $\mathbb{R}^n$, $f(V)$ *need not* be closed in $\mathbb{R}^m$.

Example 11.2. $n = m = 1$, $U = \mathbb{R} \setminus \{0\}$, and $f(x) = \frac{1}{x}$.
Let $V = [1, \infty)$ which is closed on $\mathbb{R}$ (although this is unbounded there is a closed boundary so this is still closed). Then $f(V) = (0, 1]$ is not closed.

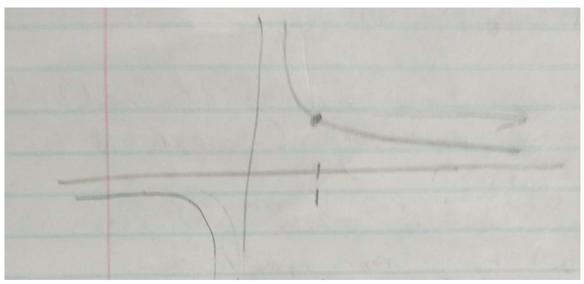

Figure 11.3: A closed domain on a continuous $f(x) = \frac{1}{x}$ may not admit a closed image.

Two other types of subsets were **compact and connected**.

11.2 Continuity and compact sets

Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ continuous on U, U open.
Does the same property hold for compact/connected sets as it did for open/closed sets?
That is, if $V \subseteq \mathbb{R}^m$ is **compact**, is $f^{-1}(V)$ **compact** on $\mathbb{R}^n$? If $V \subseteq \mathbb{R}^m$ is **connected**, is $f^{-1}(V)$ **connected** on $\mathbb{R}^n$?
No to both!

Example 11.3. Counter-example for compact set:
$n = m = 1$, $U = \mathbb{R}$, and $f(x) = \frac{1}{1+x^2}$. Let $V = [0, 1]$ which is compact. Then $f^{-1}(V) = \{x \in \mathbb{R} \mid \frac{1}{1+x^2} \in [0, 1]\} = \mathbb{R}$ is not compact.

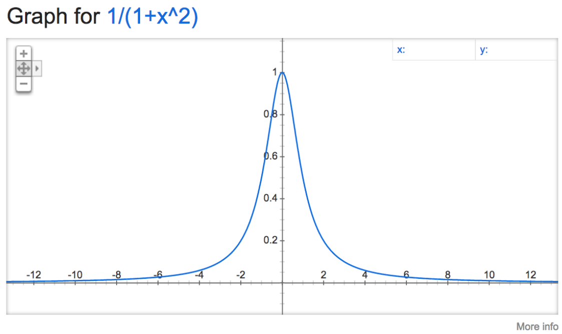

Figure 11.4: A compact image on a continuous $f(x) = \frac{1}{1+x^2}$ may not admit a compact inverse image.

Example 11.4. Counter-example for connected set:
$n = m = 1$, $U = \mathbb{R}$, and $f(x) = x^2$. Let $V = (1, 9)$ which is connected. Then $f^{-1}(V) = (-3, -1) \cup (1, 3)$ is not connected.

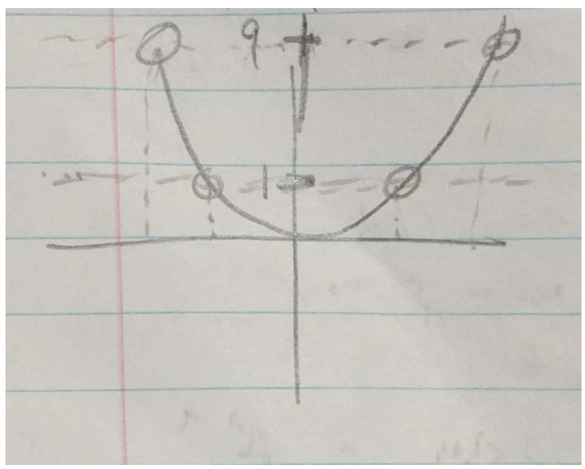

Figure 11.5: A connected image on a continuous $f(x) = x^2$ may not admit a connected inverse image.

Proposition 11.2. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ continuous on U which is open. Let $K \subseteq U$ be **compact**. Then $f(K) = \{f(x) \mid x \in K\}$ is **compact** in $\mathbb{R}^m$.

Proof. Let $\{U_\alpha \mid \alpha \in A\}$ be an open cover of $f(K)$ i.e. $U_\alpha \subseteq \mathbb{R}^m$ is open $\forall \alpha \in A$ and

$$f(K) \subseteq \bigcup_{\alpha \in A} U_\alpha$$

Claim: $K \subseteq \bigcup_{\alpha \in A} f^{-1}(U_\alpha)$.
If $x \in K$, then $f(x) \in f(K)$ so $f(x) \in U_\alpha$ for some $\alpha \Rightarrow x \in f^{-1}(U_\alpha)$.
Since f is continuous, U_α open $\Rightarrow f^{-1}(U_\alpha)$ open in $\mathbb{R}^n$ $\forall \alpha$ (by previous propositions).
So $\{f^{-1}(U_\alpha) \mid \alpha \in A\}$ is an open cover of K which is compact.
So $\exists \alpha_1, \ldots, \alpha_n \in A$ such that

$$K \subseteq \bigcup_{j=1}^{N} f^{-1}(U_{\alpha_j})$$

Then $f(K) \subseteq \bigcup_{j=1}^{N} U_{\alpha_j}$ because if $y \in f(K)$ where $y = f(x)$ for some $x \in K$ where $x \in f^{-1}(U_{\alpha_j})$ for some $j \in \{1, \ldots, N\}$ so $f(x) \in U_{\alpha_j}$.
So $f(K)$ is compact. □

Remark 11.3. By Heine-Borel, compact $\iff$ closed and bounded. But we've seen that if f is continuous $f(\textit{closed}) \neq \textit{closed}$ in general. This implies the additional bounded property makes it valid.
What about $f(\textit{bounded}) = \textit{bounded}$ for a continuous f? **No:** see this counter-example:

Example 11.5. $U = (0, \infty) \subseteq \mathbb{R}$ for $n = m = 1$ and $f(x) = \log x$.
Let $V = (0, 1)$ which is bounded. Then $f(V) = (-\infty, 0)$ which is not bounded.

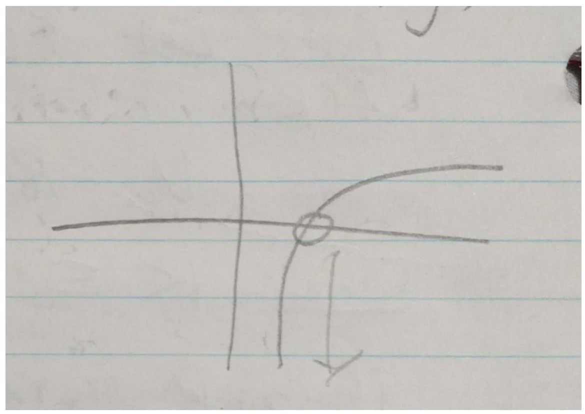

Figure 11.6: A bounded domain on a continuous $f(x) = \log x$ may not admit a bounded image.

But as we proved before, $f(\textit{closed} + \textit{bound}) = \textit{closed} + \textit{bounded}$ so both conditions are sufficient.

11.3 Extreme value theorem (EVT)

Corollary 11.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U is open ($m = 1!$) and f is continuous on U.
Let $K \subseteq U$ be **compact**. Then $\exists x_1, x_2$ in K

$$f(x_1) \leq f(x) \leq f(x_2) \quad \forall x \in K$$

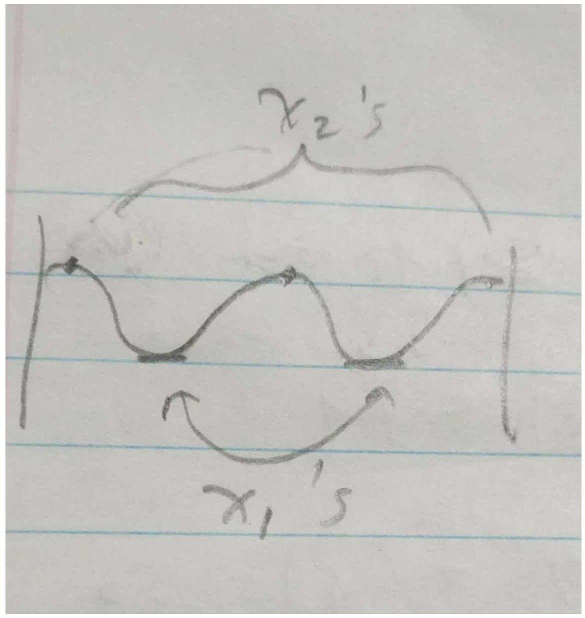

Figure 11.7: x_1, x_2 may not be unique in the Extreme Value Theorem.

This means a continuous. real-valued function on a compact set attains a global maximum value and global minimum value.
Clearly this wouldn't work in $\mathbb{R}^m, m > 1$ since there is no notion of min/max of vectors).

Proof. Assume K compact, f is continuous on U so $f(K)$ is a compact subset of $\mathbb{R}$ (by previous proposition).
By Heine-Borel $f(K)$ is **closed and bounded** in $\mathbb{R}$, so

$$M = \sup_{x \in K} f(x)$$
$$m = \inf_{x \in K} f(x)$$

both exists and is finite (bounded intervals has an infima and suprema).
Let $k \in \mathbb{N}$ such that $M - \frac{1}{k} < M$ so $\exists x_k \in K$ such that

$$M - \frac{1}{k} < f(x_k) \leq M$$

K is bounded so (x_k) is a bounded sequence in $\mathbb{R}$, so by Bolzano-Weierstrass $\exists$ convergent subsequence $(x_{k_l}) \to x$ as $l \to \infty$.
But K is closed so $\lim_{l \to \infty} x_{k_l} = x \in K$.
Since f is continuous, so

$$\lim_{l \to \infty} f(x_{k_l}) = f(\lim_{l \to \infty} (x_{k_l})) = f(x) \in f(K)$$

since $x \in K$. Thus

$$M - \frac{1}{k_l} < f(x_{k_l}) \leq M$$

then as $l \to \infty$, we have $M \leq f(x) \leq M$.
So $x \in K$ and $f(x) = \sup_{y \in K} f(y)$ (some y) so global max is attained.
For global min, similarly $m \leq f(x_k) < m + \frac{1}{k}$.

Remark 11.4. The non-uniqueness of extreme values come from choosing an arbitrary convergent subsequence from (x_k).

□

This generalizes EVT from 147: if f is continuous on $[a,b]$ then f attains a global max/min.

Remark 11.5. Note that f must be continuous for this to hold. Otherwise we could have

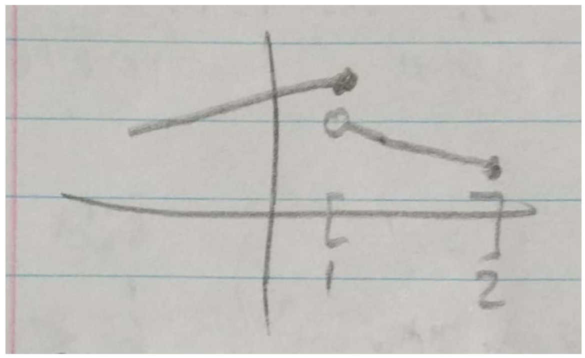

Figure 11.8: There is a global min on $[1,2]$ but no global max.

12 January 29, 2018

12.1 Continuity and connected sets

(See above for connected image does not imply connected inverse image example.)

Proposition 12.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ continuous on U which is open.
Let $E \subseteq U$ be **connected** on $\mathbb{R}^n$. Then $f(E)$ is **connected** in $\mathbb{R}^m$ (i.e. continuous image of connected set is connected).

Proof. Suppose $f(E)$ is not connected. Let $V_1, V_2 \in \mathbb{R}^m$ open sets be a separation of $f(E)$

1. $f(E) \subseteq V_1 \cup V_2$
2. $f(E) \cap V_1 \neq \varnothing$
3. $f(E) \cap V_2 \neq \varnothing$
4. $f(E) \cap V_1 \cap V_2 = \varnothing$

Since f is continuous, $f^{-1}(V_1), f^{-1}(V_2)$ are open in $\mathbb{R}^n$. If $x \in E$, $f(x) \in f(E) \subseteq V_1 \cup V_2$. So $f(x) \in V_1$ or $f(x) \in V_2$ which implies $x \in f^{-1}(V_1) \cup f^{-1}(V_2)$, that is

$$E \subseteq f^{-1}(V_1) \cup f^{-1}(V_2)$$

Let $y \in f(E) \cap V_1 \neq \varnothing$. So $\exists x \in E$ such that $y = f(x) \in V$, that is

$$x \in f^{-1}(V) \cap E \neq \varnothing$$

Similarly $f^{-1}(V_2) \cap E \neq \varnothing$.

If $x \in E \cap f^{-1}(V_1) \cap f^{-1}(V_2)$ then $f(x) \in f(E) \cap V_1 \cap V_2 \neq \varnothing$ which is a contradiction of our initial assumptions that V_1, V_2 is a separation.
So $f^{-1}(V_1) \cap f^{-1}(V_2) \cap E = \varnothing$, which means $\{f^{-1}(V_1), f^{-1}(V_2)\}$ is a separation of R which is a contradiction since E is connected.
Thus $f(E)$ must be connected. □

12.2 Intermediate value theorem (IVT)

Corollary 12.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, where U open ($m = 1!$).
Suppose f is continuous on U and let $E \subseteq U$ be connected. Let $x, y \in E$ such that $f(x) < f(y)$. Then for **each** $w \in (f(x), f(y))$, $\exists z \in E$ such that $f(z) = w$.
(i.e. a continuous real-valued fn on a connected set admits all values between any two of its values).

Proof. Assume the contrary: that is $\exists w_0 \in (f(x), f(y))$ such that $w_0 \notin f(E)$.
Let

$$V_1 = \{w \in \mathbb{R} \mid w < w_0\} = (-\infty, w_0)$$
$$V_2 = \{w \in \mathbb{R} \mid w > w_0\} = (w_0, \infty)$$

then V_1, V_2 is a spearation of $f(E)$ but $f(E)$ is connected by previous proposition, which is a contradiction. □

Aside: If $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ for U open is continuous on U then

$$\|f\| : U \to \mathbb{R}$$

where $\|f\|(x) = \|f(x)\|$ is continuous. real-valued so we can apply EVT or IVT to $\|f\|$.

12.3 Uniform continuity

Definition 12.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U is open, and let $D \subseteq U$.
We say that f is **uniformly continuous on** D **iff** $\forall \epsilon > 0\ \exists \delta > 0$ such that

$$x, y \in D \quad \|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \epsilon$$

Remark 12.1.

1. Uniformly continuous only makes sense with respect to a particular subset D of U. f may be unif. continuous on $D_1 \subseteq U$ but not unif. continuous on $D_2 \subseteq U$.
2. If f is unif continuous on D, this means given any $\epsilon > 0$, we can find **a single** $\delta > 0$ depending only on ϵ that works to establish continuity of $f_{|D}$ at $x \in D$ for all $x \in D$.
3. If f is unif continuous on D, then $f_{|D}$ is continuous at $x \in D$ $\forall x \in D$ (but the converse is not necessarily true).

Example 12.1. Let $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ and $f(x) = \frac{1}{x}$ (f is continuous).
Let $D = (0, 1]$.

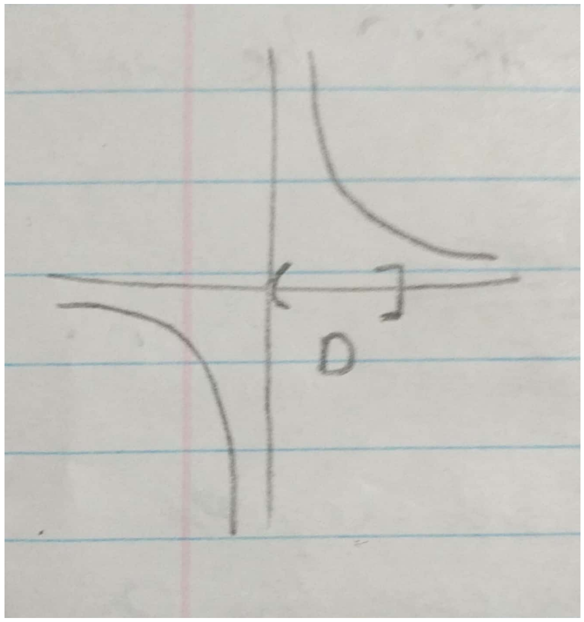

Figure 12.1: For the subdomain $D = (0, 1]$, $f(x) = \frac{1}{x}$ is not uniformly continuous on D. We can find x, y arbitrary close to 0 such that for a given $\epsilon > 0$, there is not single $\delta > 0$.

Claim: f is NOT unif continuous on D (so we find an ϵ where no single δ works).
Let $\epsilon = \frac{1}{2}$. Let $\delta > 0$ be arbitrary. Let $n \in \mathbb{N}$ be large enough so that

$$\frac{1}{n(n+1)} < \delta$$

Let $x = \frac{1}{n}$ and $y = \frac{1}{n+1} \in D$. So we have

$$|x - y| = |\frac{1}{n} - \frac{1}{n+1}| = \frac{1}{n(n+1)} < \delta$$

but we have

$$|f(x) - f(y)| = |n - (n+1)| = 1 > \epsilon$$

Example 12.2. Let $f : \mathbb{R} \to \mathbb{R}$ and $f(x) = x^2$ (f is continuous).
Let $D = [0, \infty)$.
Claim: f is not unif continuous on D.
Let $\epsilon = 1$, let $\delta > 0$ be arbitrary. Let $x = \frac{2}{\delta}$ and $y = \frac{2}{\delta} + \frac{\delta}{2} \in D$ (close to each other).
Note

$$|x - y| = |\frac{2}{\delta} - (\frac{2}{\delta} + \frac{\delta}{2})| = \frac{\delta}{2} < \delta$$

and

$$|f(x) - f(y)| = |(x - y)(x + y)| = \frac{\delta}{2}(\frac{4}{\delta} + \frac{\delta}{2}) = 2 + \frac{\delta^2}{4} > 2 > \epsilon$$

12.4 Uniform continuity and compact sets

Theorem 12.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be continuous on U open. Let $K \subseteq U$ be **compact**.
Then f is unif continuous on K.

Remark 12.2. Aside: In example 12.1, we chose $D = (0, 1]$ and a proof with a counter n when x, y gets arbitrarily close to 0.
If $D = [\zeta, 1]$ such that is compact, our counter argument with n would fail because we can't get arbitrarily close to 0.

Proof. Let $\epsilon > 0$ and let $x \in K$. Since f is continuous at x, $\exists \delta(x) > 0$ such that

$$\|y - x\| < \delta(x) \Rightarrow \|f(y) - f(x)\| < \frac{\epsilon}{2}$$

i.e. $f(B_{\delta(x)}(x)) \subseteq B_{\frac{\epsilon}{2}}(f(x))$.
Also $K \subseteq \bigcup_{x \in K} B_{\frac{\delta(x)}{2}}(x)$ (this is an arbitrary union for every point in K).
By compactness of K, $\exists$ finite set $x_1, \ldots, x_N \in K$ such that

$$K \subseteq \bigcup_{j=1}^{N} B_{\frac{\delta(x_j)}{2}}(x_j) \tag{12.1}$$

Let $\delta = \min\{\frac{\delta(x_1)}{2}, \ldots, \frac{\delta(x_N)}{2}\} > 0$.
Suppose $x, y \in K$ and $\|x - y\| < \delta$, and $x \in B_{\frac{\delta(x_j)}{2}}(x_j)$ for some $j \in [1, \ldots, N]$ by equation 12.1.
Then we have

$$\begin{aligned}
\|y - x_j\| &\overset{\triangle}{\leq} \|y - x\| + \|x - x_j\| \\
&< \delta + \frac{\delta(x_j)}{2} \\
&< \frac{\delta(x_j)}{2} + \frac{\delta(x_j)}{2} \\
&= \delta(x_j)
\end{aligned}$$

That is $\|y - x_j\| < \delta_(x_j)$ so $y \in B_{\delta(x_j)}(x_j)$, thus we have

$$\begin{aligned}
\|f(x) - f(y)\| &\overset{\triangle}{\leq} \|f(x) - f(x_j)\| + \|f(x_j) - f(y)\| \\
&< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

so we found a $\delta > 0$ (single δ) such that $x, y \in K$ where $\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \epsilon$. $\square$

12.5 Differentiability

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U open and let $a \in U$. We want to define what it means for f to be differentiable at a and what is the derivative of f at a (denoted $(Df)_a$ OR $(Df)(a)$).
We'll see soon that if f is differentiable at a, then $(Df)_a$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ ($m \times n$ matrix).

Remark 12.3. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Choose basis β of $\mathbb{R}^n$, basis γ of $\mathbb{R}^m$. Then T is represented wrt these two bases by an $m \times n$ basis $[T]_{\gamma,\beta}$.
If $\tilde{\beta}$ is another basis of $\mathbb{R}^n$ and $\tilde{\gamma}$ is another basis of $\mathbb{R}^m$, then let P (invertible $n \times n$) and Q (invertible $m \times m$) be the change of bases matrices from basis β to $\tilde{\beta}$ and from basis γ to $\tilde{\gamma}$, respectively. Thus

$$[T]_{\tilde{\gamma},\tilde{\beta}} = Q^{-1}[T]_{\gamma,\beta}P$$

If $n = m$ and we choose $\beta = \gamma$ and $\tilde{\beta} = \tilde{\gamma}$ then

$$[T]_{\tilde{\beta}} = P^{-1}[T]_\beta P$$

If $n = m = 1$ such that $\beta = \gamma, \tilde{\beta} = \tilde{\gamma}$ we have

$$[T]_{\tilde{\beta}} = [T]_\beta \quad \text{since the } 1 \times 1 \text{ matrices commute}$$

i.e. the matrices representing a linear map $\mathbb{R} \to \mathbb{R}$ is **unique** (doesn't depend on β).

13 January 31, 2018

13.1 Single variable differentiability

Definition 13.1. Let $f : U \subseteq \mathbb{R} \to \mathbb{R}$, U open, and $a \in U$. We say f is **differentiable** at a **iff**

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

exists. If so, we call the limit the **derivative** of f at a and we denote it

$$f'(a) = \frac{df(a)}{dx} = (Df)_a$$

Remark 13.1. Claim: If f is differentiable at a then f is continuous at a.
We have

$$f(a+h) - f(a) = \frac{f(a+h) - f(a)}{h} \cdot h$$

Taking the limit of both sides we get

$$\lim_{h\to 0} f(a+h) - f(a) = 0 = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} \cdot h$$

So we get (from the right equality)

$$\lim_{h\to 0} f(a+h) = f(a)$$
$$\Rightarrow \lim_{x\to a} f(x) = f(a)$$

so f is continuous at a (limit exists and $= f(a)$).
The converse is false: e.g. $f(x) = |x|$.

When we choose our definition of differentiable in general, we'll want the property that "differentiable at a" $\Rightarrow$ "continuous at a".

13.2 Partial derivatives

Definition 13.2. Let $i \in \{1, \ldots, n\}$. The **partial derivative** of f *in the x_i-direction* at the point a is defined to be

$$\lim_{h\to 0} \frac{f(a_1, \ldots, a_{i-1}, a_i + h, a_{i+1}, \ldots, a_n) - f(a_1, \ldots, a_n)}{h}$$

if it exists.

This is the ordinary derivative at $x_i = a_i$ of f thought of as only a function of x_i with all other $x_j = a_j$ constant.

Notation: when the partial derivative exists, it is denoted

$$\frac{\partial f}{\partial x_i}(a) = f_{x_i}(a)$$

The shorthand definition: let $e_1, \dots, e_n$ be the standard basis of $\mathbb{R}^n$. Then

$$\frac{\partial f}{\partial x_i}(a) = \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h}$$

e.g. in $\mathbb{R}^1$, $e_1 = (1) \Rightarrow a + he_1 = a + h \in \mathbb{R}^1$.

Example 13.1.

1. $f(x, y) = \sin(xy)$
2. $g(x, y, z) = e^{x^2 z} \log(y + z)$
3. $h(x, y, z) = y^3 \sin(xz) + e^{13z + x^3 \log(z^5+1)}$
4. $\lambda(x, y, z) = x^2 \sin(y) - \frac{yz}{x}$

Then we have

$$\frac{\partial f}{\partial x} = y\cos(xy) \qquad \frac{\partial f}{\partial y} = x\cos(xy)$$

$$\frac{\partial g}{\partial x} = 2xze^{x^2 z} \log(y + z) \qquad \frac{\partial g}{\partial y} = \frac{e^{x^2 z}}{y + z}$$

$$\frac{\partial g}{\partial z} = \frac{e^{x^2 z}}{y + z} + x^2 e^{x^2 z} \log(y + z)$$

$$\frac{\partial h}{\partial y} = 3y^2 \sin(xz)$$

$$\frac{\partial \lambda}{\partial x} = 2x \sin(y) + \frac{yz}{x^2}$$

13.3 Wrong definitions of differentiability

Remark 13.2. A reasonable guess for the definition of differentiability of f at a is to say all the partial derivatives $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ all exist at a (but not continuous).
This is **WRONG!** because there exists examples where $\frac{\partial f}{\partial x_i}$ all exist at a but f is not continuous at a (see assignment 5)!
From intuition: f may not be continuous *in between* x_i-directions.

Definition 13.3. You can also consider the rate of change of f at a in the direction of *any* unit vector u (i.e. in between the standard vectors e_i).
This is called the **directional derivative** of f at a in the u-direction and is denoted

$$(D_u f)_a$$

(On assignment 5: show $(D_{e_i} f)(a) = \frac{\partial f}{\partial x_i}(a)$.)

Remark 13.3. Another reasonable guess; f is differentiable at a if **all** the directional derivatives $(D_u f)(a)$ exists at a for all unit vectors u.

This is **also WRONG!**, since there exists examples where all $(D_u f)(a)$ exists at a but f is not continuous at a! From intuition: one can take more complicated paths to a where contunity may not hold, e.g. as before with x^2 or x^3 paths.

13.4 Second partial derivatives

Suppose $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U open, and suppose $\frac{\partial f}{\partial x_i}$ exists everywhere on U.
So $\frac{\partial f}{\partial x_i} : U \subseteq \mathbb{R}^n \to \mathbb{R}$ (it's a function on U).
So we can ask about the existence of $\frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i})$ or

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j}$$

(remark the order of the notation).
For example if $n = 2$, there are 4 (n^2) second partial derivatives

$$\begin{aligned} f_{xx} &= \frac{\partial^2 f}{\partial x \partial x} = \frac{\partial^2 f}{\partial x^2} \\ f_{yy} &= \frac{\partial^2 f}{\partial y^2} \\ f_{xy} &= \frac{\partial^2 f}{\partial y \partial x} \\ f_{yx} &= \frac{\partial^2 f}{\partial x \partial y} \end{aligned}$$

Example 13.2. For $f(x, y) = \sin(xy)$, we have

$$f_x = y\cos(xy) \quad f_y = x\cos(xy)$$

thus we have

$$\begin{aligned} f_{xx} &= -y^2 \sin(xy) & f_{xy} &= \cos(xy) - xy\sin(xy) \\ f_{yy} &= -x^2 \sin(xy) & f_{yx} &= \cos(xy) - xy\sin(xy) \end{aligned}$$

Notice that $f_{xy} = f_{yx}$ everywhere!
For $\lambda(x, y, z) = x^2 \sin(y) - \frac{yz}{x}$.

$$\lambda_x = 2x\sin(y) + \frac{yz}{x^2} \quad \lambda_y = x^2\cos(y) - \frac{z}{x} \quad \lambda_z = \frac{-y}{x}$$

Thus we have

$$\begin{aligned} \lambda_{xy} &= 2x\cos(y) + \frac{z}{x^2} & \lambda_{yx} &= 2x\cos(y) + \frac{z}{x^2} \\ \lambda_{xz} &= \frac{y}{x^2} & \lambda_{zx} &= \frac{y}{x^2} \\ \lambda_{yz} &= \frac{-1}{x} & \lambda_{zy} &= \frac{-1}{x} \end{aligned}$$

So again

$$\frac{\partial^2 \lambda}{\partial x_i \partial x_j} = \frac{\partial^2 \lambda}{\partial x_j \partial x_i} \quad \forall i, j$$

Question: is this always true? **NO!** There exists examples where $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ $(n > 1)$ such that

$$\frac{\partial^2 \lambda}{\partial x_i \partial x_j}(a) \neq \frac{\partial^2 \lambda}{\partial x_j \partial x_i}(a)$$

for certain i, j and a (for a certain point a usually).

13.5 $C^k(U)$ (class of continuous functions)

Definition 13.4. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U open. We say f is in $C^0(U)$ if f is continuous on U.
We say f is in $C^1(U)$ if f is in $C^0(U)$ and all $\frac{\partial f}{\partial x_i}$'s exist and are continuous on U.
We say f is in $C^2(U)$ if f is in $C^1(U)$ and all $\frac{\partial^2 f}{\partial x_i \partial x_j}$'s exist and are continuous on U.
In general, for $k \in \mathbb{N}$, f is in $C^k(U)$ if f is in $C^{k-1}(U)$ and all $\frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_1}}$ exist and are continuous on U.

Definition 13.5. f is in $C^\infty(U)$ if $f \in \bigcap_{k=0}^{\infty} C^k(U)$ i.e. if $f \in C^k(U)$ $\forall k \in \mathbb{N}$.

Remark 13.4.

$$C^0(U) \supset C^1(U) \supset \dots \supset C^k(U) \supset C^{k+1}(U) \supset \dots \supset C^\infty(U)$$

14 February 2, 2018

14.1 Mean Value Theorem (MVT)

Theorem 14.1. Let $f : U \subseteq \mathbb{R} \to \mathbb{R}$, U open, be continuous on $[a, b] \in U$ and differentiable on (a, b). There $\exists c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a)$$

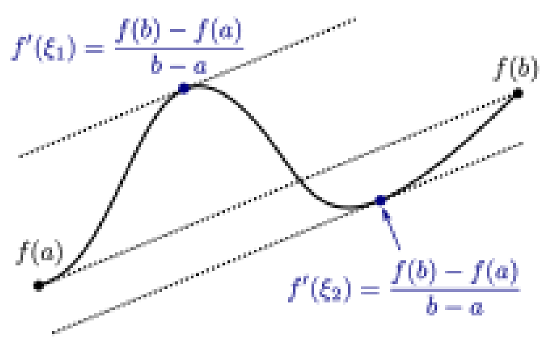

Figure 14.1: There may be multiple $c \in (a, b)$ that satisfy the MVT property.

14.2 "Commutativity" of mixed partial derivatives

Theorem 14.2. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U open. Let $a \in U$. Suppose $\frac{\partial f}{\partial x_j}$, $\frac{\partial f}{\partial x_k}$ **exist and are continuous** $(j \neq k, j, k \in \{1, \dots, n\})$ on a *neighbourhood of* a.
Furthermore, suppose that $\frac{\partial^2 f}{\partial x_j \partial x_k}$ exists in a *neighbourhood of* a and is continuous on a.

Then $\frac{\partial^2 f}{\partial x_k \partial x_j}$ exists at a and

$$\frac{\partial^2 f}{\partial x_k \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_k}(a)$$

Remark 14.1. In the examples above: all the first and second partial derivatives existed and were continuous everywhere on the domain of f. So we have much more than we need (we only require they exist and continuous at a) to apply the above theorem and conclude that the mixed partials are equal.

Remark 14.2. The partial derivatives need only be continuous on *a neighbourhood of* a: nothing needs said about the space away from a.

Proof. We will require 3 applications of the single variable Mean Value Theorem (MVT).
First we'll show we can reduce the problem to $n = 2$ ($x_j = x, x_k = y$) and $a = (0, 0)$.
Let $s, t \in \mathbb{R}$ be small enough such that

$$h(s,t) = f(a + se_j + te_k)$$

is defined (this is possible since $a \in U$ and U is open so open ball).
Let's compute $\frac{\partial h}{\partial s}(s_0, t_0)$

$$\begin{aligned}
\frac{\partial h}{\partial s}(s_0, t_0) &= \lim_{\epsilon \to 0} \frac{h(s_0 + \epsilon, t_0) - h(s_0, t_0)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{f(a + (s_0 + \epsilon)e_j + t_0 e_k) - f(a + s_0 e_j + t_0 e_k)}{\epsilon} \\
&= \frac{\partial f}{\partial x_j}(a + s_0 e_j + t_0 e_k) && \text{definition of } \frac{\partial f}{\partial x_j}
\end{aligned}$$

Similarly,

$$\frac{\partial h}{\partial t}(s_0, t_0) = \frac{\partial f}{\partial x_k}(a + s_0 e_j + t_0 e_k)$$

Note that these hold for **any** $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, U open therefore (recall the partial derivatives are themselves functions on $U \to \mathbb{R}$, so we can apply the above recursively on itself)

$$\begin{aligned}
\frac{\partial^2 f}{\partial s \partial t}(s_0, t_0) &= \frac{\partial^2 f}{\partial x_j x_k}(a + s_0 e_j + t_0 e_k) \\
\frac{\partial^2 f}{\partial t \partial s}(s_0, t_0) &= \frac{\partial^2 f}{\partial x_k x_j}(a + s_0 e_j + t_0 e_k)
\end{aligned}$$

So

$$\begin{aligned}
\frac{\partial^2 f}{\partial s \partial t}(0, 0) &= \frac{\partial^2 f}{\partial x_j x_k}(a) \\
\frac{\partial^2 f}{\partial t \partial s}(0, 0) &= \frac{\partial^2 f}{\partial x_k x_j}(a)
\end{aligned}$$

(assuming these all exist). Therefore it is enough to consider the case when $n = 2$, $a = (0, 0)$ (we reduced our problem into a $\mathbb{R}^2 \to \mathbb{R}$ problem using arbitrary s, t).
Let's prove our thoerem with $h : U \subseteq \mathbb{R}^2 \to \mathbb{R}$, U open, and $0 \in U$.
Note that $\frac{\partial h}{\partial s}, \frac{\partial h}{\partial t}$ exists and are continuous on a neighbourhood of 0 (these are from our initial premise after converting to our reduced problem). Also, $\frac{\partial^2 h}{\partial t \partial s}$ exists on a n'h'd of 0 and is continuous. at 0.
We want to show $\frac{\partial^2 h}{\partial s \partial t}$ exists at 0 and equals $\frac{\partial^2 h}{\partial t \partial s}(0)$ at 0.

Let us define

$$H(s,t) = (h(s,t) - h(s,0)) - (h(0,t) - h(0,0))$$

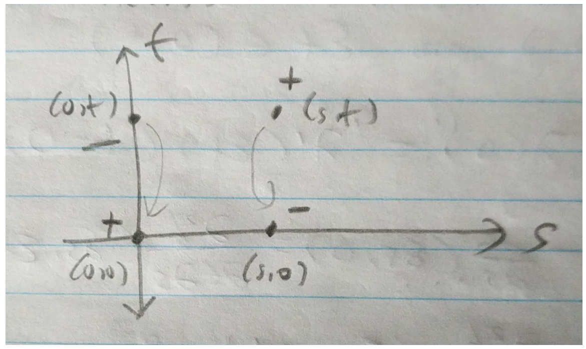

Figure 14.2: Sketch of how we set up $H(s,t)$.

Fix t sufficiently close to 0. Define $k(s) = h(s,t) - h(s,0)$. So we have

$$H(s,t) = k(s) - k(0)$$

Also $k'(s) = \frac{\partial h}{\partial s}(s,t) - \frac{\partial h}{\partial s}(s,0)$ exists and continuous on n'h'd of 0 (by hypothesis). So we can apply the MVT to k on $[0,s]$, then

$$\exists \delta \in (0,1) \Rightarrow \delta s \in (0,s)$$

such that

$$k(s) - k(0) = k'(\delta s)(s-0)$$

So we have

$$H(s,t) = s[\frac{\partial h}{\partial s}(\delta s,t) - \frac{\partial h}{\partial s}(\delta s,0)]$$

Fix s (and hence δ). We define

$$\lambda(t) = \frac{\partial h}{\partial s}(\delta s,t)$$
$$\lambda'(t) = \frac{\partial^2 h}{\partial t \partial s}(\delta s,t)$$

exists near $t=0$ (by hypothesis) so $\lambda(t)$ is continuous. and differentiable on $[0,t]$ for t small enough.
Applying the MVT to λ on $[0,t]$

$$\exists \epsilon \in (0,1) \Rightarrow \epsilon t \in (0,t)$$

such that

$$\lambda(t) - \lambda(0) = \lambda'(\epsilon t)(t-0)$$

So we have (from definition of λ)

$$\frac{\partial h}{\partial s}(\delta s,t) - \frac{\partial h}{\partial s}(\delta s,0) = \frac{\partial^2 h}{\partial t \partial s}(\delta s,\epsilon t)(t)$$

Substituting this into $H(s,t)$, we get

$$H(s,t) = st\frac{\partial^2}{\partial t \partial s}(\delta s,\epsilon t)$$

for some $\delta, \epsilon \in (0,1)$. Therefore when we take the limit

$$\lim_{(s,t)\to(0,0)} \frac{1}{st} H(s,t) = \frac{\partial^2 h}{\partial t \partial s}(0,0)$$

since $\frac{\partial^2 h}{\partial t \partial s}$ is assumed to be continuous. at $(0,0)$. If we can show the LHS is equivalent to $\frac{\partial^2 h}{\partial s \partial t}(0,0)$, then we are done.
Recall that

$$H(s,t) = h(s,t) - h(s,0) - h(0,t) + h(0,0)$$

We can also write

$$H(s,t) = (h(s,t) - h(0,t)) - (h(s,0) - h(0,0))$$

(notice the regrouping: in the graph, we are now subtracting in the other direction).
We define $\mu(t) = h(s,t) - h(0,t)$ so $H(s,t) = \mu(t) - \mu(0)$. Therefore

$$\mu'(t) = \frac{\partial h}{\partial t}(s,t) - \frac{\partial h}{\partial t}(0,t)$$

exists $\forall t$ sufficiently close to 0 (from hypothesis where $a = 0$). So $\mu(t)$ is continuous on $[0,t]$ and is differentiable on $(0,t)$ for t small.
Applying the MVT to μ on $[0,t]$, then

$$\exists \theta \in (0,1) \Rightarrow \theta t \in (0,t)$$

such that

$$\mu(t) - \mu(0) = \mu'(\theta t)(t - 0)$$

Thus we have

$$H(s,t) = t[\frac{\partial h}{\partial t}(s,\theta t) - \frac{\partial h}{\partial t}(0,\theta t)]$$

We can rewrite this as

$$\frac{H(s,t)}{st} = \frac{1}{s}[\frac{\partial h}{\partial t}(s,\theta t) - \frac{\partial h}{\partial t}(0,\theta t)]$$

Since $\frac{\partial h}{\partial t}$ is continuous. on a n'h'd of $(0,0)$ (hypothesis), then we let $t \to 0$ **first** so that we get

$$\frac{\partial h}{\partial t}(s,\theta t) \to \frac{\partial h}{\partial t}(s,0)$$
$$\frac{\partial h}{\partial t}(0,\theta t) \to \frac{\partial h}{\partial t}(0,0)$$

Thus we have

$$\lim_{(s,t)\to(0,0)} = \lim_{s\to 0}[\frac{1}{s}(\frac{\partial h}{\partial t}(s,0) - \frac{\partial h}{\partial t}(0,0))]$$

This is precisely the definition of the partial derivative with respect to s, thus we have

$$\lim_{(s,t)\to(0,0)} \frac{H(s,t)}{st} = \frac{\partial^2 h}{\partial s \partial t}(0,0) = \frac{\partial^2 h}{\partial t \partial s}(0,0)$$

□

14.3 Defining multivariable differentiability

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ (general m), U open. Let $a \in U$. We want to define what it means for f to be differentiable at a.
We *expect* that if f is differentiable at a, then

1. f should be **continuous on** a
2. all the partial derivatives of f (if $m = 1$) should exist at a

Example 14.1. Example of where partial derivatives exist but is not continuous at given $a = (0, 0)$:
Let $n = 2$ where $U = \mathbb{R}^2$ and

$$f(x, y) = \begin{cases} x + y & \text{if } x = 0 \text{ or } y = 0 \\ 1 & \text{otherwise} \end{cases}$$

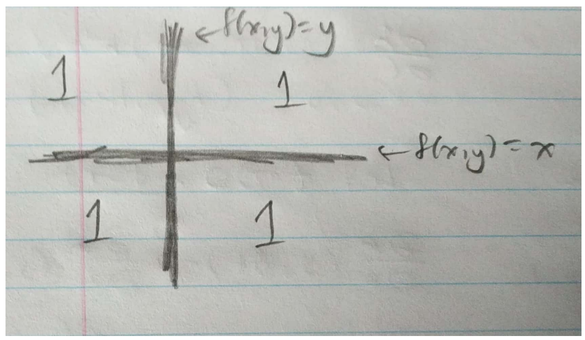

Figure 14.3: A function where the partial derivatives exist but are not continuous at a given point $(0, 0)$.

Note that

$$\begin{aligned} \frac{\partial f}{\partial x}(0, 0) &= \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} \\ &= \lim_{h \to 0} \frac{h - 0}{h} = 1 \end{aligned}$$

and similarly, $\frac{\partial f}{\partial y}(0, 0) = 1$.
Both partials exist at $(0, 0)$ but clearly f is not continuous at $(0, 0)$ ($\lim_{x \to 0} f(x) \neq f(x)$).

Remark 14.3. This shows that our previous bad definition that if partials exist, then differentiable is *wrong*.

14.4 Differentiability with linear maps

Going back to the $n = m = 1$ simple case, where $f : U \subseteq \mathbb{R} \to \mathbb{R}$ for U open and $x_0 \in U$. Then f is **differentiable at** $x_0 \iff \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h}$ exists (previously defined).
Note that $f'(x_0) \in \mathbb{R}$ are 1×1 matrices i.e. a linear map from $\mathbb{R}$ to $\mathbb{R}$.
Assuming f is differentiable at x_0, we define the linear map $T : \mathbb{R} \to \mathbb{R}$ by

$$T(h) = f'(x_0)h$$

such that

$$\begin{aligned}\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} &= f'(x_0)\\ &= \lim_{h\to 0}\frac{f'(x_0)h}{h}\\ &= \lim_{h\to 0}\frac{T(h)}{h}\end{aligned}$$

So (rewriting the above)

$$\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)-T(h)}{h}=0$$

Note that for any limits on an arbitrary $g(x)$ that approach to 0

$$\lim_{h\to 0} g(x)=0 \iff \lim_{h\to 0}|g(x)|=0$$

so we have

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

We've shown that if f is differentiable at x_0, $\exists$ a linear map $T:\mathbb{R}\to\mathbb{R}$ such that the above holds.
We've thus motivated the general definition of differentiability:

Definition 14.1. For $f:U\subseteq\mathbb{R}^n\to\mathbb{R}^m$, U open, let $x_0\in U$.
We say f is **differentiable** at x_0 if $\exists$ a linear map $T:\mathbb{R}^n\to\mathbb{R}^m$ such that

$$\lim_{h\to 0}\frac{\|f(x_0+h)-f(x_0)-T(h)\|}{\|h\|}=0 \tag{14.1}$$

where we take the norm of an $\mathbb{R}^m$ vector in the numerator and the norm of an $\mathbb{R}^n$ vector in the denominator (this is also the reason why we needed to take the norm to be able to divide the two).

15 February 5, 2018

15.1 Differential (Jacobian matrix) $(Df)_a$

Remark 15.1. We'll show soon that if such a T linear map exists, it is necessarily **unique** and it's called the *derivative of f at a* and is denoted $(Df)_a$ OR $(Df)(a)$ (i.e. $(Df)_a$ is an $m\times n$ matrix of real numbers).
$(Df)_a$ is also called the **differential** of f at a, or the **linearization** of f at a, or the **Jacobian matrix**.

Notice: If $(Df)_a=T$ exists satisfying equation 14.1 then

1. $T:\mathbb{R}^n\to\mathbb{R}^m$ is linear
2. T is a very good approximation to the map $h\mapsto f(a+h)-f(a)$ near $h=0$ in the following sense (recall $T(h)=f'(a)h=\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}\cdot h$).

$$\begin{aligned}&h\mapsto f(a+h)-f(a)\\ &h\mapsto T(h)\end{aligned}$$

Both agree at $h=\vec{0}$ (i.e. both send $\vec{0}\to\vec{0}$) and moreover the difference

$$\|f(a+h)-f(a)-T(h)\|\to 0 \quad \text{as } h\to 0$$

so fast that the quotient in equation 14.1 still goes to 0 as $h \to \vec{0}$ (sends numerator to 0 faster than $h \to \vec{0}$).

15.2 Differentiability implies continuity with $T(h)$

Proposition 15.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, U open, and $a \in U$. Suppose f is differentiable at a. Then f is **continuous** at a (we show this is true for the linear map definition now).

Proof. We need to show that $\lim_{x \to a} f(x) = f(a)$ or $\lim_{h \to 0} f(a+h) = f(a)$ (where $x = a + h$).
Note that

$$\begin{aligned}\lim_{h \to 0} \|f(a+h) - f(a) - T(h)\| &= \lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} \cdot \|h\| \\ &= 0 \cdot 0 \\ &= 0\end{aligned}$$

where the second equality holds from products of existing limits. So we have

$$\|f(a+h) - f(a)\| \overset{\triangle}{\leq} \|f(a+h) - f(a) - T(h)\| + \|T(h)\|$$

Since T is linear, we have $\|T(h)\| \leq \|T\|_{op}\|h\|$ which $\to 0$ as $h \to \vec{0}$. Thus combining the above we have

$$\|f(a+h) - f(a)\| \leq \|f(a+h) - f(a) - T(h)\| + \|T\|_{op}\|h\|$$

which by the squeeze theorem we have

$$\lim_{h \to 0} f(a+h) = f(a)$$

and thus f is continuous at a. $\square$

15.3 Differential is matrix of partial derivatives

Theorem 15.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ and $a \in U$. Suppose f is differentiable at a.
We have

$$f(x) \in \mathbb{R}^m = (f_1(x), \ldots, f_m(x))$$

where $f_j : U \subseteq \mathbb{R}^n \to \mathbb{R}$ are the component functions of f, $1 \leq j \leq m$.
Then all the partial derivatives $\frac{\partial f_i}{\partial x_j}$ exists at a for $1 \leq i \leq m$, $1 \leq j \leq n$. Moreover,

$$T = (Df)_a$$

is the $m \times n$ matrix whose (i,j)-entry is $\frac{\partial f_i}{\partial x_j}(a)$. This shows $(Df)_a$ is unique if it exists.

Proof. By assumption, $\exists$ $m \times n$ matrix T such that

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

Recall from linear algebra we have

$$\begin{bmatrix} T(e_1) & T(e_2) & \vdots & T(e_n) \end{bmatrix}$$

which is an $m \times n$ matrix (each column is the image of e_i, the standard basis vector).
Since the above limit exists and is zero, we get 0 if $h \to \vec{0}$ along *any path*.

Choose the path

$$h = te_j \quad j \in \{1, \ldots, n\}$$

as $t \in \mathbb{R}$ goes to 0, then $h \to \vec{0}$.
We have

$$\lim_{t\to 0} \frac{\|f(a+te_j) - f(a) - T(te_j)\|}{\|te_j\|} = 0$$
$$\iff \lim_{t\to 0} \| \frac{f(a+te_j) - f(a) - T(te_j)}{t} \| = 0 \qquad \therefore \|e_j\| = 1$$

T is linear so $T(te_j) = tT(e_j)$ thus

$$\lim_{t\to 0} \| \frac{f(a+te_j) - f(a)}{t} - T(e_j)\| = 0$$

Recall that $\lim_{x\to x_0} \|g(x) - L\| = 0 \iff \lim_{x\to x_0} g(x) = L$ (trivial by epsilon-delta, true for all $\epsilon > 0$). Thus

$$\lim_{t\to 0} \frac{f(a+te_j) - f(a)}{t} = T(e_j)$$

so the i-th component of the quotient above is $\frac{\partial f_i}{\partial x_j}(a)$. Therefore we've shown

$$T_{ij} = \frac{\partial f_i}{\partial x_j}(a)$$

exists and holds. $\square$

Remark 15.2. If f is differentiable at a, then all $\frac{\partial f_i}{\partial x_j}$ exist at a (as above). So if *even one* $\frac{\partial f_i}{\partial x_j}$ does not exist at a, then f is *not differentiable* at a.
Warning: Just because all $\frac{\partial f_i}{\partial x_j}(a)$ exist **DOES NOT** necessarily imply that f is differentiable at a, because with $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T_{ij} = \frac{\partial f_i}{\partial x_j}(a)$$

it may not be true that

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

15.4 Gradient notation

For $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ (note $m = 1$!), $a \in U$, and f differentiable at a, then $(Df)_a$ is a $1 \times n$ matrix

$$(Df)_a = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) & \ldots & \frac{\partial f}{\partial x_n}(a) \end{bmatrix}$$

This is called the **gradient** of f at a and is also denoted

$$(\nabla f)(a) = (\nabla f)_a = (Df)_a = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) & \ldots & \frac{\partial f}{\partial x_n}(a) \end{bmatrix}$$

Now for general m, if $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$ then

$$(Df)_a = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \dots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix} = \begin{bmatrix} (\nabla f_1)(a) \\ (\nabla f_2)(a) \\ \vdots \\ (\nabla f_m)(a) \end{bmatrix}$$

15.5 Differentiable $\iff$ all components are differentiable

Lemma 15.1. Let $f : U : \mathbb{R}^n \to \mathbb{R}^m$, $a \in U$. Then f is differentiable at a **iff** each component function $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is differentiable at a $\forall i = 1, \dots, m$.

Proof. f is differentiable at a iff

$$\begin{aligned} &\lim_{h \to 0} \frac{\|f(a+h) - f(a) - T(h)\|}{\|h\|} = 0 \\ \iff &\lim_{h \to 0} \|\frac{f(a+h) - f(a) - T(h)}{\|h\|}\| = 0 \\ \iff &\lim_{h \to 0} \|\frac{f_i(a+h) - f_i(a) - T_i(h)}{\|h\|}\| = 0 \quad \forall i = 1, \dots, m \end{aligned}$$

where the last $\iff$ follows from the fact that the vector inside the outer $\|\cdot\|$ is an $\mathbb{R}^m$ vector, and any vector converges $\iff$ its components converges (shown before). □

15.6 Linear combination is differentiable

Proposition 15.2. Let $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$. Suppose f, g both differentiable at $a \in U$. Let $\lambda, \mu \in \mathbb{R}$. Then $\lambda f + \mu g : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ or

$$(\lambda f + \mu g)(x) = \lambda f(x) + \mu g(x)$$

is differentiable at a and

$$(D(\lambda f + \mu g))_a = \lambda (Df)_a + \mu (Dg)_a$$

Proof. Assignment 6 (use triangle inequality and the fact that $(Dh)_a$ is linear, then squeeze theorem). □

16 February 7, 2018

16.1 Partial derivatives exist and continuous implies differentiability

Theorem 16.1 (Partials exist and continuous implies differentiability)**.** (Sufficient but *NOT NECESSARY* condition for differentiability).
Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$, $a \in U$. Suppose all $\frac{\partial f_i}{\partial x_j}$ **exists on a n'h'd of** a and are **continuous at** a.
Then f is differentiable at a.

Proof. From last time f is differentiable $\iff$ every f_i is also differentiable at a.

Hence it's enough to prove the theorem for $m = 1$. Let

$$a = (a_1, \dots, a_n)$$
$$h = (h_1, \dots, h_n)$$

Define for $j = 1, \dots, n$

$$v_j = \sum_{k=1}^{j} h_k e_k = (h_1, h_2, \dots, h_j, 0, \dots, 0)$$

Therefore $v_n = h$. We set $v_0 = \vec{0} = (0, \dots, 0)$.
Note that

$$f(a+h) - f(a) = f(a+v_n) - f(a+v_0) \tag{16.1}$$
$$= \sum_{k=1}^{n} [f(a+v_k) - f(a+v_{k-1})] \tag{16.2}$$

Note $v_k = v_{k-1} + h_k e_k$. By hypothesis, $\frac{\partial f}{\partial x_k}$ exists in a n'h'd of a so for h sufficient close to 0 the function

$$\begin{aligned}\mu_k(t) &= f(a + v_{k-1} + te_k)\\ &= f(a + h_1, \dots, a_{k-1} + h_{k-1}, a_k + t, \dots, a_{k+1}, \dots, a_n)\end{aligned}$$

is a differentiable function of t on $[0, h_k)$ for h_k sufficient small. Thus

$$\mu_k(t) = \frac{\partial f}{\partial x_k}(a + v_{k-1} + te_k)$$

We apply MVT to μ_k, $\exists \epsilon_k \in (0,1)$ so $e_k h_k = (0, h_k)$ such that

$$\begin{aligned}&\mu_k'(\epsilon_k h_k)(h_k - 0) = \mu_k(h_k) - \mu_k(0)\\ \Rightarrow &h_k[\frac{\partial f}{\partial x_k}(a + v_{k-1} + \epsilon_k h_k e_k)] = f(a+v_k) - f(a+v_{k-1})\end{aligned}$$

Hence equation 16.1 becomes

$$f(a+h) - f(a) = \sum_{k=1}^{n} h_k \cdot \frac{\partial f}{\partial x_k}(a + v_{k-1} + \epsilon_k h_k e_k) \tag{16.3}$$

For

$$(Df)_a = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) & \dots & \frac{\partial f}{\partial x_n}(a) \end{bmatrix}$$

this exists (by hypothesis). We need to show

$$\frac{\|f(a+h) - f(a) - (Df)_a(h)\|}{\|h\|} \to 0$$

as $h \to 0$.
Recall that

$$(Df)_a(h) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a) h_k$$

(where LHS is a $1 \times n$ matrix multiplied by an $n \times 1$ vector). So from equation 16.3

$$\begin{aligned} f(a+h) - f(a) - (Df)_a(h) &= \sum_{k=1}^{n} [\frac{\partial f}{\partial x_k}(a + v_{k-1} + \epsilon_k h_k e_k) - \frac{\partial f}{\partial x_k}(a)] h_k \\ &= L \cdot h \end{aligned}$$

where L_k is the stuff inside the summation and $L = (L_1, \ldots, L_n)$.
Therefore we have

$$\begin{aligned} \frac{|f(a+h) - f(a) - (Df)_a(h)|}{\|h\|} &= \frac{\|L \cdot h\|}{\|h\|} \\ &\leq \frac{\|L\|\|h\|}{\|h\|} && \text{Cauchy-Schwarz} \\ &= \|L\| \end{aligned}$$

so enough to show that

$$\lim_{h \to 0} L = 0$$

(where we then apply squeeze theorem).
So we've reduced the problem to show

$$\lim_{h \to \vec{0}} L_k = 0 \quad \forall k = 1, \ldots, n$$

or

$$\lim_{h \to \vec{0}} \frac{\partial f}{\partial x_k}(a + v_{k-1} + \epsilon_k h_k e_k) - \frac{\partial f}{\partial x_k}(a) = 0$$

Note that $v_{k-1} = \sum_{j=1}^{k-1} h_j e_j \to 0$ as $h \to \vec{0}$. Furthermore for $0 < \epsilon_k < 1$ we have $\epsilon_k h_k e_k \to 0$ as $h \to \vec{0}$ thus

$$a + v_{k-1} + \epsilon_k h_k e_k \to a$$

as $h \to \vec{0}$. Note that $\frac{\partial f}{\partial x_k}$ is assumed to be *continuous at* a, so $\lim_{h \to L_k} \to 0$ as desired. □

16.2 Summary about differentiability

To check if $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$

1. If f is **not continuous at** a, then f is **not differentiable** at a

2. If any of $\frac{\partial f_i}{\partial x_j}$ do not exist at a, f is **not differentiable** at a

3. Let $(Df)_a$ be the $m \times n$ matrix whose i, j entry is $\frac{\partial f_i}{\partial x_j}(a)$. Then f is differentiable at $a \iff$

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

4. We can avoid step 3 if we know all $\frac{\partial f_i}{\partial x_j}$ exist on a n'h'd of a and are continuous at a (this implies f is differentiable at a by theorem 16.1).

16.3 Differentiability and C^1

Let $U \subseteq \mathbb{R}^n$ open. We say f is in $C^1(U)$ if all $\frac{\partial f_i}{\partial x_j}$ exist and are continuous everywhere on U. by the previous theorem, if $f \in C^1(U)$ then f is differentiable at any point in U.
Also $C^0(U)$ implies continuous function on U. Note from before

$$C^1(U) \subseteq C^0(U)$$

So we have the desired property that $C^1 \Rightarrow$ differentiable $\Rightarrow$ continuous . Functions in C^1 are sometimes called **continuously differentiable**.

Example 16.1. To show conditions of theorem 16.1 are sufficient but not necessary, let $n = 2$, $U \subseteq \mathbb{R}^2$

$$f(x, y) = (x^2 + y^2)\sin(\frac{1}{\sqrt{x^2 + y^2}})$$

for $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$.

Step 1 f is continuous on at $(0, 0)$ (by squeeze).

Step 2 Compute $f_x(0, 0)$ and $f_y(0, 0)$

$$\begin{aligned} f_x(0,0) &= \lim_{h \to 0} \frac{f(h,0) - f(0,0)}{h} \\ &= \lim_{h \to 0} \frac{h^2}{h} \sin(\frac{1}{\sqrt{h^2}}) \\ &= 0 \end{aligned}$$

by squeeze. Similarly $f_y(0, 0) = 0$. Thus we have $(Df)_{(0,0)} = [0, 0]$.

Step 3 Need to check

$$\lim_{(h_1,h_2)\to(0,0)} \frac{|f((0,0) + (h_1, h_2)) - f(0,0) - (Df)_{(0,0)}((h_1, h_2))|}{\sqrt{h_1^2 + h_2^2}} = 0$$

Thus we have

$$\begin{aligned} &\lim_{(h_1,h_2)\to(0,0)} \frac{(h_1^2 + h_2^2)\sin(\frac{1}{\sqrt{h_1^2+h_2^2}})}{\sqrt{h_1^2 + h_2^2}} \\ =&\lim_{(h_1,h_2)\to(0,0)} \sqrt{h_1^2 + h_2^2}\sin(\frac{1}{\sqrt{h_1^2 + h_2^2}}) \\ =&0 \end{aligned}$$

by squeeze.

So f is differentiable at $(0, 0)$.

Follow-up: we show that $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ (which exists everywhere) are not necessarily continuous at $(0, 0)$ (to show that our previous conditions are sufficient but not necessary)

$$f(x, y) = (x^2 + y^2)\sin(\frac{1}{\sqrt{x^2 + y^2}})$$

Recall that $f_x(0,0) = f_y(0,0) = 0$. So at a point $(x,y) \neq (0,0)$

$$\begin{aligned} f_x &= 2x\sin(\frac{1}{\sqrt{x^2+y^2}}) + (x^2+y^2)\cos(\frac{1}{\sqrt{x^2+y^2}}) \cdot (\frac{-1}{2})(x^2+y^2)^{\frac{-3}{2}} \cdot 2x \\ &= 2x\sin(\frac{1}{\sqrt{x^2+y^2}}) - \frac{x}{\sqrt{x^2+y^2}}\cos(\frac{1}{\sqrt{x^2+y^2}}) \end{aligned}$$

We want to check if

$$\lim_{(x,y)\to(0,0)} f_x(x,y) = f_x(0,0) = 0$$

and similarly for f_y. Note the first term $\to 0$ by squeeze. We thus want to show (to show it's not continuous)

$$\lim_{(x,y)\to(0,0)} \frac{x}{\sqrt{x^2+y^2}}\cos(\frac{1}{\sqrt{x^2+y^2}}) \quad \text{DNE}$$

Remark 16.1. One can imagine approaching 0 from the y-axis (fix $x = 0$) which obviously goes to 0, but one can also approach from the x-axis (where we have $\frac{x}{|x|}\cos(\frac{1}{|x|})$). Although $\cos(\frac{1}{|x|})$ is bounded we do not know what happens when the two terms are put together so we can't say it obviously exists.

By sequential characterization of limits

$$\lim_{(x,y)\to(0,0)} h(x,y) = 0 \iff \lim_{k\to\infty} h(x_k,y_k) = 0$$

for all sequences $(x_k,y_k) \in \mathbb{R}^2$ converging to $(0,0)$.
Thus consider $(x_k,y_k) = (\frac{(-1)^k}{k\pi}, 0)$, so we have

$$\begin{aligned} h(x_k,y_k) &= \frac{(-1)^k\frac{1}{k\pi}}{\sqrt{\frac{1}{k^2\pi^2}}}\cos(\frac{1}{\sqrt{\frac{1}{k^2\pi^2}}}) \\ &= (-1)^k\cos(k\pi) \\ &= 1 \quad \forall k \end{aligned}$$

Similarly when $(x_k,y_k) = (\frac{(-1)^{k+1}}{k\pi}, 0)$, we have the limit approaching to -1. Since they have different limits, then the limit DNE so f_x is not continuous at $(0,0)$.

Upshot: We have

$$\text{continuous } \supset \text{ differentiable } \supset C^1$$

where the rightmost inequality highlights the condition that $\frac{\partial f_i}{\partial x_j}$ exists and continuous is not necessary for differentiability.

17 February 9, 2018

17.1 Product rule for differentiability

Proposition 17.1. Let $U \subseteq \mathbb{R}^n$, $f,g : U \to \mathbb{R}^m$, $a \in U$.
Suppose f,g are both differentiable at a. Then we claim $f \cdot g : U \to \mathbb{R}$, where $(f\cdot g)(x) = f(x)\cdot g(x)$ is differentiable at a and

$$D(f\cdot g)_a = f(a)^T(Dg)_a + g(a)^T(Df)_a \tag{17.1}$$

where we have $1 \times n$ matrix on the LHS and $1 \times m$, $m \times n$, $1 \times m$, and $m \times n$ matrices on the right.

Remark 17.1. Let $h = f \cdot g$ so $h = \sum_{k=1}^{m} f_k g_k$. If h is differentiable at a, its derivative $(Dh)_g$ would be

$$(Dh)_a = \begin{bmatrix} \frac{\partial h}{\partial x_1}(a) & \dots & \frac{\partial h}{\partial x_n}(a) \end{bmatrix}$$

But

$$\frac{\partial h}{\partial x_i} = \frac{\partial}{\partial x_i}(\sum_k f_k g_k) = \sum_k \frac{\partial f_k}{\partial x_i} g_k + f_k \frac{\partial g_k}{\partial x_i}$$

So the above two equations are just equation 17.1 in components.

Proof. We need to prove that

$$\lim_{t \to 0} \frac{\|f(x_0 + t) - f(x_0) - T(t)\|}{\|t\|} = 0 \quad t \in \mathbb{R}$$

Note that

$$h(a+t) - h(a) - (Dh)_a(t) = (f \cdot g)(a+t) - (f \cdot g)(a) - f(a)^T (Dy)_a(t) - g(a)^T (Df)_a(t)$$

(so we assume the product rule and show it implies differentiability since our theorem is an $\iff$). Note the above can be rewritten as

$$\begin{aligned}
&= (f(a+t) - f(a) - (Df)_a(t)) \cdot g(a+t) && \text{name this } T_1 \\
&+ f(a) \cdot (g(a+t) - g(a) - (Dg)_a(t)) && \text{name this } T_2 \\
&+ (Df)_a(t) \cdot (g(a+t) - g(a)) && \text{name this } T_3
\end{aligned}$$

By triangle inequality we have

$$|h(a+t) - h(a) - (Dh)_a(t)| \leq |T_1| + |T_2| + |T_3|$$

We thus show that as $t \to 0$, then $\frac{|T_i|}{\|t\|} \to 0$ for $i = 1, 2, 3$.
f, g differentiable at a so f, g are continuous at a. Therefore we have

$$\begin{aligned}
\frac{|T_1(t)|}{\|t\|} &\leq \frac{\|f(a+t) - f(a) - (Df)_a(t)\|}{\|t\|} \cdot \|g(a+t)\| \\
\frac{|T_2(t)|}{\|t\|} &\leq \|f(a)\| \cdot \frac{\|g(a+t) - g(a) - (Dg)_a(t)\|}{\|t\|} \\
\frac{|T_3(t)|}{\|t\|} &\leq \frac{\|(Df)_a(t)\| \|g(a+t) - g(a)\|}{\|t\|} \\
&\leq \frac{\|(Df)_a\|_{op} \|t\|}{\|t\|} \cdot \|g(a+t) - g(a)\|
\end{aligned}$$

where the inequalities are from Cauchy-Schwarz. These all $\to 0$ as $\|t\| \to 0$ (since they are all products of existing limits). ☐

Special case when $m = 1$: We have $f, g : U \subseteq \mathbb{R}^n \to \mathbb{R}$ and $f \cdot g = fg$. Then

$$D(fg)_a^T = \nabla(fg)(a) = f(a) \cdot (\nabla g)(a) + g(a) \cdot (\nabla f)(a)$$

Informally, $\nabla(fg) = f \nabla g + g \nabla f$.

17.2 Chain rule

Theorem 17.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be differentiable at $a \in U$. Let $g : V \subseteq \mathbb{R}^n \to \mathbb{R}^p$ be differentiable at $b = f(a) \in V$. Assume $f(U) \subseteq V$.
Then $g \circ f : U \subseteq \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a and

$$D(g \circ f)_a = (Dg)_{f(a)}(Df)_a$$

where we have matrices of size $p \times n$ on the left and $p \times m$ and $m \times n$ on the right (note that the linear map is a composition of linear maps: that is the derivative of a composition is the composition of the derivatives).

Proof. Let $Q_1(h) = f(a+h) - f(a) - (Df)_a(h)$ (defined for h small). Similarly, let $Q_2(k) = g(b+k) - g(b) - (Dg)_b(k)$ (k small). By hypothesis we have

$$\lim_{h\to 0} \frac{\|Q_1(h)\|}{\|h\|} = 0$$
$$\lim_{k\to 0} \frac{\|Q_2(k)\|}{\|k\|} = 0$$

For k small, set $k = f(a+h) - f(a) = f(a+h) - b$ (small by continuity). So we have

$$\begin{aligned} g(f(a+h)) - g(f(a)) &= g(b+k) - g(b) \\ &= (Dg)_b(k) + Q_2(k) \\ &= (Dg)_b(f(a+h) - f(a)) + Q_2(k) \\ &= (Dg)_b((Df)_a(h) + Q_1(h)) + Q_2(k) \\ &= (Dg)_b((Df)_a(h)) + (Dg)_b(Q_1(h)) + Q_2(k) && \text{linearity} \end{aligned}$$

Thus we have

$$\begin{aligned} &\frac{\|g(f(a+h)) - g(f(a)) - (Dg)_{f(a)}(Df)_a(h)\|}{\|h\|} \\ =&\frac{(Dg)_{f(a)}(Q_1(h)) + Q_2(k)}{\|h\|} \\ \leq&\|(Dg)_b\|_{op}\frac{\|Q_1(h)\|}{\|h\|} + \frac{\|Q_2(k)\|}{\|h\|} && \text{triangle inequality and op norm} \end{aligned}$$

where the left term $\to 0$ as $h \to 0$ by hypothesis. We want o prove that

$$\lim_{h\to 0} \frac{\|Q_2(k)\|}{\|h\|} = 0$$

to finish the proof.
Let $\epsilon_1 > 0$ be arbitrary, since $\lim_{h\to 0} \frac{\|Q_1(h)\|}{\|h\|} = 0$. Then $\exists \delta_1 > 0$ such that

$$0 < \|h\| < \delta_1 \Rightarrow \frac{\|Q_1(h)\|}{\|h\|} < \epsilon_1$$

by definitions of limits. Thus $\|Q_1(h)\| \leq \epsilon_1 \|h\|$ $\forall h$ where $0 < \|h\| < \delta_1$.

Since $\lim_{k\to 0}\frac{\|Q_2(k)\|}{\|k\|}=0$ for any arbitrary $\epsilon_2>0$, then $\exists\delta_2>0$ such that

$$0<\|k\|<\delta_2\Rightarrow\frac{\|Q_1(k)\|}{\|k\|}<\epsilon_2$$

We claim that

$$\|h\|<\delta_1\Rightarrow\|Q_2(k)\|\leq\epsilon_2\|k\|\leq\epsilon_2\|h\|$$

Note that

$$\begin{aligned}\|k\|&=\|f(a+h)-f(a)\|\\&=\|(Df)_a(h)+Q_1(h)\|\\&\leq\|(Df)_a\|_{op}\|h\|+\|Q_1(h)\| &&\text{triangle and op norm}\\&(\|(Df)_a\|_{op}+\epsilon_1)\|h\|\end{aligned}$$

for $\|h\|<\delta_1$. So $\exists C>0$ such that $\|k\|\leq C\|h\|$ for $\|h\|<\delta_1$ thus our claim holds.
From our claim, we have

$$\frac{\|Q_2(k)\|}{\|h\|}\leq\epsilon_2 C$$

and since $\epsilon_2>0$ is arbitrary we have

$$\lim_{h\to 0}\frac{\|Q_2(k)\|}{\|h\|}=0$$

□

18 February 12, 2018

18.1 Explicit form of chain rule

Writing out the chain rule explicitly with components:

$$\begin{aligned}y&=(y_1,\ldots,y_m)=(f_1(x_1,\ldots,x_n),\ldots,f_m(x_1,\ldots,x_n))\\z&=(z_1,\ldots,z_p)=(g_1(y_1,\ldots,y_m),\ldots,g_p(y_1,\ldots,y_m))\end{aligned}$$

where $z=g(y)=g(f(x))=(g\circ f)(x)=h(x)$.
Furthermore, let $a=(a_1,\ldots,a_n)$ and $b=(b_1,\ldots,b_m)$.
Recall that the chain rule is given as

$$(Dh)_a=(Dg)_{f(a)}(Df)_a$$

Let $1\leq i\leq p$, $1\leq j\leq n$. Then

$$\text{(i,j)-th entry of }(Dh)_a=\frac{\partial h_i}{\partial x_j}(a)$$

which corresponds to

$$\text{(i,j)-th entry of }(Dg)_b(df)_a=\sum_{k=1}^m[(Dg)_b]_{ik}[(Df)_a]_{kj}$$

Example 18.1. For the simple, single variable case where $m = n = p = 1$, we have

$$h'(a) = g'(f(a)) \cdot f'(a)$$
$$\Rightarrow \frac{dh}{dx}(a) = \frac{dg}{dy}(f(a)) \cdot \frac{df}{dx}(a)$$

Remark 18.1. We commonly abuse notation when discussing derivatives. In the $m = n = p = 1$ example, we write

$$y = f(x)$$
$$z = g(y)$$

where y is really $y(x)$ (it's a function of x, similarly z). So really we are referring to

$$\frac{df}{dx} \text{ is "equivalent" to } \frac{dy}{dx}$$
$$\frac{dg}{dx} \text{ is "equivalent" to } \frac{dz}{dx}$$

So we may see

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}$$

(these are not fractions!). In components

$$\frac{\partial z_i}{\partial x_i} = \sum_{k=1}^{m} \frac{\partial z_i}{\partial y_k}\frac{\partial y_k}{\partial x_j}$$

Example 18.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$, differentiable on $\mathbb{R}^2$ and let $(x, y) = h(r, \theta) = (r\cos\theta, r\sin\theta)$.
Let $h : \mathbb{R}^2 \to \mathbb{R}^2$ also differentiable on $\mathbb{R}^2$ (in C^∞ actually, but we only need C^1).
$f \circ h : \mathbb{R}^2 \to \mathbb{R}$ should be differentiable on $\mathbb{R}^2$.
Again, we use the *abuse of notation* where we write

$$f(r, \theta) = f(x(r, \theta), y(r, \theta)) = f(h(r, \theta))$$

So from the chain rule we have

$$f_r = \frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial r}$$
$$= f_x \cos\theta + f_y \sin\theta$$

similarly we have

$$f_\theta = \frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \theta}$$
$$- r\sin\theta \cdot f_x + r\cos\theta \cdot f_y$$

18.2 The derivative is a linearization

For the $n = 1$ case, suppose $f : U \subseteq \mathbb{R} \to \mathbb{R}$ where $x_0 \in U$, $y_0 = f(x_0)$ and f differentiable at x_0.

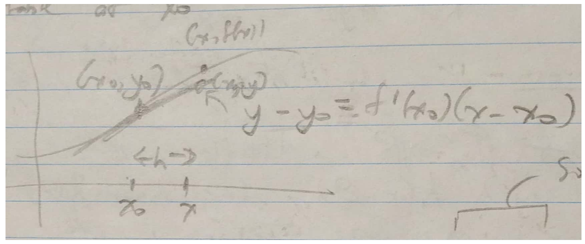

Figure 18.1: We can express the function at x_0 as a linearization expressed by the tangent line.

We can express f as a linear function

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + R_{x_0}(h)$$

where $h = x - x_0$. Also $f(x) - f(x_0)$ is the *change in the function*, $f'(x_0)(x - x_0)$ is the *change in the tangent line*, and $R_{x_0}(h)$ is *some remainder term*.
We can say that f **is differentiable at** x_0 **iff** $\lim_{h\to 0} \frac{R_{x_0}(h)}{h} = 0$ (this follows from the definition of the derivative: also, the remainder term approaches 0 faster than the horizontal distance).
Let $\delta y = f(x) - f(x_0)$ (change in function between x_0 and x), $\mathrm{d}y = f'(x_0)(x - x_0)$ (change in the linearization). Then $\delta y \approx \mathrm{d}y$ for x close to x_0 then

$$\frac{\delta y - \mathrm{d}y}{h} \to 0 \text{ as } h \to 0$$

For $n > 1$, let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ differentiable at x_0. Then

$$f(x) - f(x_0) = (Df)_{x_0}(x - x_0) + R_{x_0}(h)$$

where $h = x - x_0 \in \mathbb{R}^n$ ($(Df)_{x_0}$ and $(x - x_0)$ are $1 \times n$ (row) and $n \times 1$ (column) matrices, respectively).
where

$$\lim_{h\to\vec{0}} \frac{R_{x_0}(h)}{\|h\|} = 0$$

For $h \approx \vec{0}$, $\mathrm{d}y = (Df)_{x_0}(x - x_0)$ is a good approximation of $\delta y = f(x) - f(x_0)$ because

$$\frac{\delta y - \mathrm{d}y}{\|h\|} \to 0 \text{ as } h \to \vec{0}$$

When $n = 1$, the graph of the linear approximation to $f(x)$ is $L(x) = f(x_0) + f'(x_0)(x - x_0)$ (this follows by dropping the remainder term as it goes to 0). It corresponds to the tangent line

$$\{(x, y) \mid y = f(x_0) + f'(x_0)(x - x_0)\}$$

When $n = 2$, the graph of the linear approximation to $f(x)$ is the graph

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

If we let $\vec{x}_0 = (x_0, y_0)$ then we have

$$L(\vec{x}) = f(\vec{x}_0) + (Df)_{\vec{x}_0}(\vec{x} - \vec{x}_0)$$

which is the **tangent plane** at (x_0, y_0), i.e. the set of points

$$\{(x, y, z) \in \mathbb{R}^3 \mid z = L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)\}$$

or

$$\{(x, y, z) \in \mathbb{R}^3 \mid z = Ax + By + C\}$$

a plane in $\mathbb{R}^3$ passing through $(x_0, y_0, f(x_0, y_0))$.

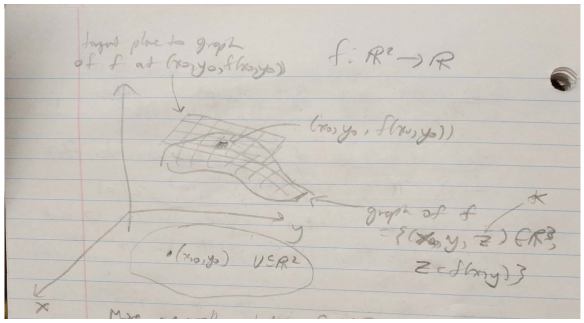

Figure 18.2: Graph of the linear approximation of $f : \mathbb{R}^2 \to \mathbb{R}$ ($n = 2$). The curve is a wavy plane in $\mathbb{R}^3$ and we can create a tangent plane at (x_0, y_0). The set at the bottom of the graph is our domain $U \subseteq \mathbb{R}^2$.

More generally, let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ differentiable at $x_0 \in U$. Then **the graph of** f is

$$\begin{aligned}\Gamma_f &= \{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} \mid y = f(x_1, \ldots, x_n)\} \\ &= \{(x_1, \ldots, x_n, f(x_1, \ldots, x_n) \mid (x_1, \ldots, x_n) \in U\}\end{aligned}$$

The **linear approximation of** f **at** x_0 is the function $L : \mathbb{R}^n \to \mathbb{R}$ where

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

or explicitly

$$\begin{aligned}L(x_1, \ldots, x_n) \in \mathbb{R} &= f(x_0) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(x_0)(x_k - (x_0)_k) \\ &= A_1x_1 + A_2x_2 + \ldots + A_nx_n + B\end{aligned}$$

where the summation term is B.
The graph of L is

$$\begin{aligned}\Gamma_L &= \{(x_1, \ldots, x_n, L(x_1, \ldots, x_n)), (x_1, \ldots, x_n) \in \mathbb{R}^n\} \\ &= \{(x_1, \ldots, x_n, y), y = A_1x_1 + \ldots + A_nx_n + B\}\end{aligned}$$

is a **hyperplane** in $\mathbb{R}^{n+1}$ i.e. it is almost exactly the same thing as an n-dimensional subspace of $\mathbb{R}^{n+1}$ except it need not pass through the origin.
The graph Γ_L is the **tangent space** to the graph of f at $(x_0, f(x_0)) \in \mathbb{R}^n \times \mathbb{R} = \mathbb{R}^{n+1}$.

Remark 18.2. f is just an $\mathbb{R}^n$ plane "moves around" in $\mathbb{R}^{n+1}$. L approximates the $\mathbb{R}^n$ plane at a specific point at x_0 for $f(x_0)$.

19 February 14, 2018

19.1 Taylor's Theorem for one variable

Theorem 19.1 (Taylor's Theorem for one variable)**.** Let $I \subseteq \mathbb{R}$ be an *interval*, let p be a non-negative integer. Let $h : I \to \mathbb{R}$ be $(p+1)$-times differentiable on I (in particular this means $h^{(k)}(t) = \frac{d^k h}{dt^k}$ is continuous $\forall k = 0, \ldots, p$ while the $(p+1)$th derivative *may not be continuous*).
Let $t_0 \neq t \in I$. Then $\exists \theta$ between t_0 and t such that

$$h(t) = \sum_{k=0}^{p} \frac{h^{(k)}(t_0)}{k!}(t-t_0)^k + \frac{h^{(p+1)}(\theta)}{(p+1)!}(t-t_0)^{p+1}$$

where the summation is the pth Taylor polynomial of h at t_0 and the last term is $R_p(t)$ the remainder term (**Note: θ is not unique**).

Proof. Define $y \in \mathbb{R}$ by

$$h(t) = \sum_{k=0}^{p} \frac{h^{(k)}(t_0)}{k!}(t-t_0)^k + \frac{y}{(p+1)!}(t-t_0)^{p+1}$$

This can be solved uniquely for y (since we know all of h, p, t_0, t; that is y depends on h, t, t_0).
Define $H : I \to \mathbb{R}$ as

$$H(s) = h(t) - [\sum_{k=0}^{p} \frac{h^{(k)}(s)}{k!}(t-s)^k + \frac{y}{(p+1)!}(t-s)^{p+1}]$$

H is **continuous** (all parts of it are continuous.) on I and **differentiable**.
By construction, $H(t_0) = 0$. Also $H(t) = h(t) - h(t) = 0$ (where all terms disappear except when $k = 0$).
By Rolle's Theorem, $\exists \theta$ between t_0 and t such that $H'(\theta) = 0$.

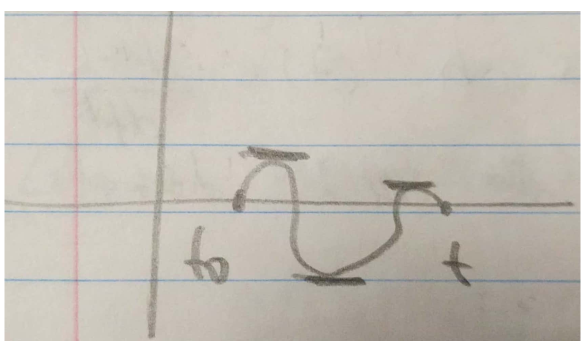

Figure 19.1: Rolle's theorem states there is some θ between t_0 and t where the gradient is 0.

Taking the derivative of the function

$$\begin{aligned}
H'(s) &= 0 - h'(s) + \sum_{k=1}^{p} \frac{h^{(k+1)}(s)}{k!}(t-s)^k + \sum_{k=1}^{p} \frac{h^{(k)}(s)}{k!} \cdot k \cdot (t-s)^{k-1} + \frac{y}{(p+1)!}(p+1)(t-s)^p \\
&= -h'(s) - \sum_{k=1}^{p} \frac{h^{(k+1)}(s)}{k!}(t-s)^k + \sum_{j=0}^{p-1} \frac{h^{(j+1)}(s)}{j!}(t-s)^j + \frac{y}{p!}(t-s)^p \\
&= -h'(s) - \frac{h^{(p+1)}(s)}{p!}(t-s)^p + h'(s) + \frac{y}{p!}(t-s)^p \\
&= -\frac{h^{(p+1)}(s)}{p!}(t-s)^p + \frac{y}{p!}(t-s)^p
\end{aligned}$$

From Rolle's Theorem, $\exists \theta$ between t_0, t such that

$$\begin{aligned}
& H'(\theta) = 0 \\
\iff & \frac{-h^{(p+1)}(\theta)}{p!}(t-\theta)^p + \frac{y}{p!}(t-\theta)^p = 0 \\
\iff & y = h^{(p+1)}(\theta)
\end{aligned}$$

□

Remark 19.1. When $p = 0$, the theorem is just the **Mean Value Theorem (MVT)**. That is: if h is differentiable on I, $\exists \theta$ between t_0, t such that

$$h(t) = h(t_0) + h'(\theta)(t - t_0)$$

Remark 19.2. Taylor's Theorem says that if h is $(p+1)$-times differentiable on I, then for any $t_0 \in I$, we can approximate h by a pth order polynomial in $t - t_0$, namely

$$h_p(t) = \sum_{k=0}^{p} \frac{h^{(k)}(t_0)}{k!}(t-t_0)^k$$

with an *error term (remainder)* "of order $(t-t_0)^{p+1}$"

$$R_p(t) = \frac{h^{(p+1)}(\theta)}{(p+1)!}(t-t_0)^{p+1}$$

in particular

$$\lim_{t \to t_0} \frac{h(t) - h_p(t)}{(t-t_0)^p} = 0 \Rightarrow \lim_{t \to t_0} \frac{R_p(t)}{(t-t_0)^p} = 0$$

if $h^{(p+1)}$ is **continuous** at t_0.

19.2 Taylor's Theorem for C^∞ (not on exam)

Remark 19.3. If $h \in C^\infty(U)$ i.e. $h^{(k)}$ exists on U $\forall k \in \mathbb{N}$ then h has a pth Taylor polynomial at $t_0 \in I$ $\forall p \in \mathbb{N}$

$$h_{p,t_0}(t) = \sum_{k=0}^{p} \frac{h^{(k)}(t_0)}{k!}(t-t_0)^k$$

Question: Is $\lim_{p\to\infty} h_{p,t_0}(t) = h(t)$ always true?
Answer: **Not always** (it always holds for $t = t_0$ but may not hold for any other t).
Note that if $\exists t \neq t_0$ in I such that the above holds, then it holds $\forall s \in I$ with $|s - t_0| < |t - t_0|$.
A function for which this is true for some $t \neq t_0$ is called **real analytic at** t_0.
Such functions have a convergent power series expansion at t_0 with a positive radius of convergence.

20 February 16, 2018

20.1 Taylor's Theorem for n variables

We will use Taylor's Theorem for one variable to prove it for n variables.

Lemma 20.1. $U \subseteq \mathbb{R}^n$ open and non-empty. Let $f \in C^p(U)$ (all partial derivatives of order at most p exist and are continuous on U) for $p \geq 0$.
Let $a \in U$, $\xi \in \mathbb{R}^n$ such that

$$\{a + t\xi \mid t \in [0, 1]\} \subseteq U$$

(line segment from a to $a + \xi$).

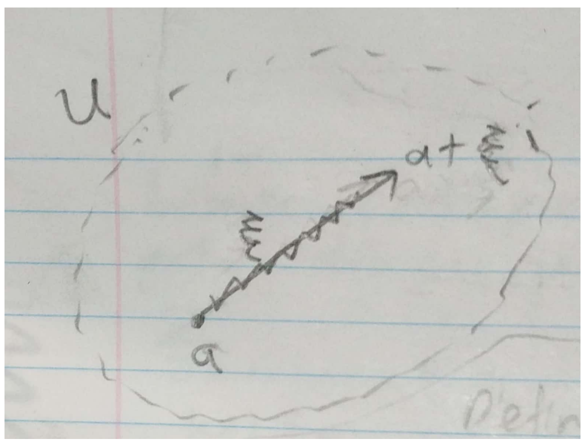

Figure 20.1: We use some vector $a \in U$, $\xi \in \mathbb{R}^n$, and $a + \xi \in U$.

Note that $\{a + t\xi\} \subseteq U$ for $t \in (-\epsilon, 1 + \epsilon)$ for some $\epsilon > 0$ since U is open.
Define $g : (-\epsilon, 1 + \epsilon) \to \mathbb{R}$ by

$$g(t) = f(a + t\xi)$$

(restriction of f to the line).
Then $g \in C^p(I)$ i.e. $\frac{d^k g}{dt^k}$ exists and is continuous on I $\forall k = 0, \dots, p$ and

$$\frac{d^k g}{dt^k} = \sum_{j_1,\dots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_1} \dots \partial x_{j_k}}(a + t\xi)\xi_{j_1} \cdot \dots \cdot \xi_{j_k}$$

for $1 \leq k \leq p$ ($\xi = (\xi_1, \dots, \xi_n)$) (note we use abuse of notation here: the kth derivative is really a function of the t in $a + t\xi$).

Example 20.1. For $n = 2, p = 3$, we have

$$\begin{aligned}\frac{dg}{dt} &= \sum_{j=1}^{2} \frac{\partial f}{\partial x_j}(a + t\xi)\xi_j \\ &= \xi_1 \frac{\partial f}{\partial x}(a + t\xi) + \xi_2 \frac{\partial f}{\partial y}(a + t\xi)\end{aligned}$$

and

$$\begin{aligned}\frac{d^2 g}{dt^2} &= \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(a + t\xi)\xi_i \cdot \xi_j \\ &= f_{xx}(a + t\xi)\xi_1^2 + 2f_{xy}(a + t\xi)\xi_1\xi_2 + f_{yy}(a + t\xi)\xi_2^2\end{aligned}$$

Proof. By induction on p (with k) where $0 \leq k \leq p$.
Let $h(t) = a + t\xi$ where $h : I \to \mathbb{R}^n$ is differentiable and we have $h_i(t) = a_i + t\xi_i$.
Let $g(t) = (f \circ h)(t) = f(a + t\xi)$ which is continuous on I, so g is continuous on U.
Suppose $p \geq 1$, $f \in C^1(U)$ so f is differentiable on U. By chain rule, $g = f \circ h$ is differentiable on I and

$$\begin{aligned}\frac{dg}{dt}(t) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(h(t)) \frac{dh_i}{dt}(t) \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a + t\xi)\xi_i\end{aligned}$$

This proves the $k = 1$ case.
Assume true for $0 \leq k < p$. We'll show it's true for $k + 1 \leq p$.
By hypothesis,

$$\frac{d^k g}{dt^k}(t) = \sum_{j_1,\ldots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_1} \ldots \partial x_{j_k}}(a + t\xi)\xi_{j_1} \cdot \ldots \cdot \xi_{j_k}$$

We'd like to the derive our $k + 1$ derivative

$$\frac{d^{k+1} g}{dt^{k+1}}(t) = \frac{d}{dt}\left(\frac{d^k g}{dt^k}(t)\right) = \sum_{j_1,\ldots,j_k=1}^{n} \frac{d}{dt}\left(\frac{\partial^k f}{\partial x_{j_1} \ldots \partial x_{j_k}}(a + t\xi)\xi_{j_1} \cdot \ldots \cdot \xi_{j_k}\right)$$

We claim that $\frac{\partial^k f}{\partial x_{j_1} \ldots \partial x_{j_k}}$ (function F) is in C^1: the first partial derivatives of F are the $(k+1)$th partial derivatives of f so $\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_n}$ are continuous on U since $f \in C^p(U)$ and $k + 1 \leq p$.
So $F \in C^1(U) \Rightarrow F$ is differentiable by the chain rule

$$\frac{d}{dt}F(x_1(t), \ldots, x_n(t)) = \sum_{j_{k+1}=1}^{n} \frac{\partial F}{\partial x_{j_{k+1}}}(x_1(t), \ldots, x_n(t)) \frac{dx_{j_{k+1}}}{dt}$$

where $x(t) = a + t\xi \Rightarrow \frac{dx_i}{dt} = \xi_i$.
So we have

$$\frac{d^{k+1}}{dt^{k+1}}g(t) = \sum_{j_1, j_{k+1}=1}^{n} \frac{\partial^{k+1} f}{\partial x_{j_1} \ldots \partial x_{j_{k+1}}}(a + t\xi)\xi_{j_1} \cdot \ldots \cdot \xi_{j_{k+1}}$$

and $g^{(k+1)}(t)$ is continuous on I because it's a composition of continuous functions ($k+1 \leq p$). $\square$

We'll use the lemma to prove the general Taylor's Theorem (for multivariables). First some notation

Definition 20.1. We denote

$$(D^{(k)}f)_a(\xi) = \sum_{j_1,\dots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_1}\dots\partial x_{j_k}}(a)\xi_1\dots\xi_k$$

for $k \geq 1$ and $(D^{(0)}f)_a = f(a)$.

Theorem 20.1 (Taylor's Theorem for n variables)**.** Let $U \subseteq \mathbb{R}^n$ open, $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be in $C^{p+1}(U)$. Let $a \in U$, $\xi \in \mathbb{R}^n$ such that $\{a + t\xi \mid t \in [0,1]\} \subseteq U$.
Then $\exists\theta \in (0,1)$ such that

$$f(a+\xi) = \sum_{k=0}^{p} \frac{(D^{(k)}f)_a(\xi)}{k!} + \frac{1}{(p+1)!}(D^{(p+1)}f)_{a+\theta\xi}(\xi)$$

Proof. Define $g(t) = f(a+t\xi)$ as before on $I = (-\epsilon, 1+\epsilon)$.
By Lemma, $g \in C^{p+1}(I)$, $g^{(k)}(t) = (D^{(k)}f)_{a+t\xi}(\xi)$ $\forall k = 0,\dots,p+1$.
By 1-D Taylor's Theorem with $t_0 = 0, t = 1$ (and $\theta \in (t_0, t) = (0,1)$) we have

$$\begin{aligned} f(a+(1)\xi) = g(1) &= \sum_{k=0}^{p} \frac{g^{(k)}(0)}{k!}(1-0)^k + \frac{g^{(p+1)}(\theta)}{(p+1)!}(1-0)^{p+1} \\ &= \sum_{k=0}^{p} \frac{(D^k f)_a(\xi)}{k!} + \frac{(D^{p+1}f)_{a+\theta\xi}(\xi)}{(p+1)!} \end{aligned}$$

as desired. $\square$

Example 20.2. Explicitly for $p = 0$ where $p+1 = 1$

$$\begin{aligned} f(a+\xi) &= f(a) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a+\theta\xi)\xi_k \\ &= f(a) + (\nabla f)(a+\theta\xi)\cdot\xi \end{aligned}$$

Example 20.3. Explicitly for $p = 1$ where $p+1 = 2$

$$f(a+\xi) = f(a) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)\xi_k + \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(a+\theta\xi)\xi_j\xi_k$$

where $(\nabla f)(a)\cdot\xi = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)\xi_k$.

Example 20.4. Explicitly for $p = 2$ where $p+1 = 3$

$$f(a+\xi) = f(a) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)\xi_k + \frac{1}{2}\sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j\partial x_k}(a)\xi_j\xi_k + \frac{1}{6}\sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i\partial x_j\partial x_k}(a+\theta\xi)\xi_i\xi_j\xi_k$$

20.2 Hessian matrix

Definition 20.2. The **Hessian** of f at $a \in U$ is the $n \times n$ symmetric matrix $(\text{Hess } f)_a$ whose i, j entry is

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a)$$

So $\sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(a)\xi_j\xi_k = \sum_{j,k=1}^{n}[(\text{Hess } f)_a]_{jk}\xi_j\xi_k = (\text{Hess } f)_a(\xi, \xi) = \xi^T(\text{Hess } f)_a\xi$.

If A is a symmetric real $n \times n$ matrix the **bilinear form** associated to A is the map $A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$A(x, y) = \sum_{j,k=1}^{n} A_{jk}x_jy_k = x \cdot (Ay) = y \cdot (Ax) = x^TAy = y^TAx$$

where x, y are $n \times 1$ column matrices and A is $n \times n$. A symmetric implies that $A(x, y) = A(y, x)$

20.3 Example of Taylor's Theorem

Example 20.5. For $n = 2$, $U = \mathbb{R}^2$ and $f : \mathbb{R}^2 \to \mathbb{R}$ where $f(x, y) = \sin(xy)$ let $a = (a_1, a_2) = (\sqrt{\frac{\pi}{2}}, \sqrt{\frac{\pi}{2}})$.
With Taylor's formula for $p = 1$ $(p + 1 = 2)$

$$\begin{aligned} f_x &= y\cos(xy) & f_{xx} &= -y^2\sin(xy) \\ f_y &= x\cos(xy) & f_{yy} &= -x^2\sin(xy) \\ f_{xy} &= \cos(xy) - xy\sin(xy) \end{aligned}$$

Taylor says $\exists$ some $\theta \in (0, 1)$ such that

$$a + \theta\xi = [a_1 + \theta\xi_1, a_2 + \theta\xi_2] = [c_1, c_2]$$

Let $(x, y) = a + \xi = (a_1 + \xi_1, a_2 + \xi_2)$ where $\xi_1 = x - a_1, \xi_2 = y - a_2$.
Then we have

$$\begin{aligned} f(x, y) = {} & f(a_1, a_2) + f_x(a_1, a_2)(x - a_1) + f_y(a_1, a_2)(y - a_2) \\ & + \frac{1}{2}[f_{xx}(c_1, c_2)(x - a_1)^2 + 2f_{xy}(c_1, c_2)(x - a_1)(y - a_2) + f_{yy}(c_1, c_2)(y - a_2)^2] \end{aligned}$$

For this example, $a_1 = a_2 = \sqrt{\frac{\pi}{2}}$. So we have

$$\begin{aligned} f_x(a_1, a_2) &= f_y(a_1, a_2) = 0 \\ f(a_1, a_2) &= \sin(\frac{\pi}{2}) = 1 \end{aligned}$$

Thus we end up with

$$f(x, y) = 1+0+0+\frac{1}{2}[-c_2^2\sin(c_1c_2)(x-\sqrt{\frac{\pi}{2}})^2+2(\cos(c_1c_2)-c_1c_2\sin(c_1c_2))(x-\sqrt{\frac{\pi}{2}})(y-\sqrt{\frac{\pi}{2}})-c_1^2\sin(c_1c_2)(y-\sqrt{\frac{\pi}{2}})^2]$$

20.4 Application of Taylor's Theorem

Proposition 20.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$. Suppose $f \in C^1(U)$.

Let K be a **compact** subset of $\mathbb{R}^n$ with $K \subseteq U$. If $E \subseteq K$ is **convex**, $\exists$ a constant $M > 0$ (depending on f and on K but not on E) such that

$$\|f(x) - f(y)\| \leq M\|x - y\| \quad \forall x, y \in E$$

Proof. $f \in C^1$ so each $\frac{\partial f}{\partial x_k}$ is continuous on U. $K \subseteq U$ is compact so $\frac{\partial f}{\partial x_k}$ are bounded on K by EVT.
So $\exists M > 0$ such that

$$\|(\nabla f)(a)\|^2 = \sum_{k=1}^{n} (\frac{\partial f}{\partial x_k}(a))^2 \leq M^2 \quad \forall a \in K$$

So

$$|(\nabla f)(a) \cdot v| \overset{C-S}{\leq} \|(\nabla f)(a)\|\|v\| \leq M\|v\|$$

for all $a \in K$ and all $v \in \mathbb{R}^n$.
By Taylor's for $p + 1 = 1$, let $x, y \in E$ and let $x = y + \xi \Rightarrow x - y = \xi$. Then

$$f(y + \xi) = f(y) + (\nabla f)(a) \cdot \xi$$

For some a between x, y. Since E is convex then $a \in E$ so then

$$\begin{aligned} &f(x) - f(y) = (\nabla f)(a) \cdot (x - y) \\ \Rightarrow &\|f(x) - f(y)\| \leq M\|x - y\| \end{aligned}$$

as desired. $\square$

21 February 26, 2018

21.1 Lipschitz functions

From before, we showed that

$$\|f(x) - f(y)\| \leq M\|x - y\| \quad \forall x, y \in E$$

on some $E \subseteq K \subseteq U$. This says the *restriction* of f on E is **Lipschitz**: in particular any Lipshitz function on a set E is **uniformly continuous** on E (for any ϵ, choose $\delta = \frac{\epsilon}{M}$). Note however that uniform continuity *does not imply* Lipschitz.

21.2 Slightly more general version of Taylor's theorem

Theorem 21.1 (More general 1-D Taylor's theorem)**.** Let $I \subseteq \mathbb{R}$ be an interval, $p \in \mathbb{N}$. Let $h : I \subseteq \mathbb{R} \to \mathbb{R}$ be p times differentiable (previously we had $p + 1$). Then for $t \neq t_0 \in I$

$$h(t) = \sum_{k=0}^{p} \frac{h^{(k)}(t_0)}{k!}(t - t_0)^k + R_{t_0,p}(t)$$

where the summation is the p**th Taylor polynomial** of h at t_0 (as before) and the $R_{t_0,p}(t)$ is some general remainder term (previously we knew this was a term in terms of $h^{(p+1)}(\theta)$ for some $\theta \in (t_0, t)$).
Note that

$$\lim_{t \to t_0} \frac{R_{t_0,p}(t)}{(t - t_0)^p} = 0$$

(prove using L'Hopital's rule).

Theorem 21.2 (More general multivariable Taylor's theorem)**.** $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ (U open, as always). Suppose $f \in C^p(U)$ (previously had $C^{p+1}(U)$). Let $a \in U$, $\xi \in \mathbb{R}^n$ such that

$$\{a + t\xi, t \in [0, 1]\} \subseteq U$$

Let

$$(D^{(k)} f)_a(\xi) = \sum_{j_1,\ldots,j_k=1}^{n} \frac{\partial^k f}{\partial x_{j_1} \ldots \partial x_{j_k}}(a)\xi_1 \ldots \xi_k$$

(as before). Define $h(t) = f(a + t\xi)$. Apply (more general) 1-D case to $t_0 = 0, t = s$. One can show that

$$f(x) = \sum_{k=0}^{p} \frac{D^{(k)} f)_a(\xi)}{k!} + R_{a,p}(x)$$

where $x = a + \xi$ and where

$$\lim_{x \to a} \frac{R_{a,p}(x)}{\|x - a\|^p} = 0$$

21.3 Optimization (min/max) for real-valued functions of several variables

Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ ($m = 1$) be differentiable on U (e.g. this is automatic if $f \in C^1(U)$).

Definition 21.1. Let $a \in U$. We say f has a **local minimum** at a if $\exists \epsilon > 0$ such that

$$f(x) \geq f(a) \quad \forall x \in B_\epsilon(a)$$

(i.e. all points around it in some open ball of radius $\epsilon > 0$ are greater than it).
We say f has a **local maximum** at a if $\exists \epsilon > 0$ such that

$$f(x) \leq f(a) \quad \forall x \in B_\epsilon(a)$$

Claim. If f has a *local max or a local min* at $a \in U$, then

$$(\nabla f)(a) = \vec{0}$$

Proof. Fix $j \in \{1, \ldots, n\}$, $a = (a_1, \ldots, a_n)$. Consider the one variable function

$$g(t) = f(a + te_j) = f(a_1, \ldots, a_{j-1}, a_j + t, a_{j+1}, \ldots, a_n)$$

$g(t)$ has a local max or local min at $t = 0$.
So by single variable calculus, $g'(0) = 0$. So we have

$$\frac{\partial f}{\partial x_j}(a) = \lim_{t \to 0} \frac{f(a + te_j) - f(a)}{t} = \lim_{t \to 0} \frac{g(t) - g(0)}{t} = 0$$

So $\frac{\partial f}{\partial x_j}(a) = 0$ for all $j = 1, \ldots, n$. □

21.4 Critical points and saddle points

Definition 21.2. A point $a \in U$ such that $(\nabla f)(\vec{a}) = 0$ is called a **critical point** of f.

We've shown if f has a local max or local min at a, then a is a critical point. But the *converse is not true*: not all critical points correspond to local extrema.

Example 21.1.

$$f(x,y) = x^2 - y^2$$

We have

$$(\nabla f) = (2x, -2y) = (0,0)$$

at $(x,y) = (0,0)$ (this is a critical point). It is however not a minimum or a maximum.

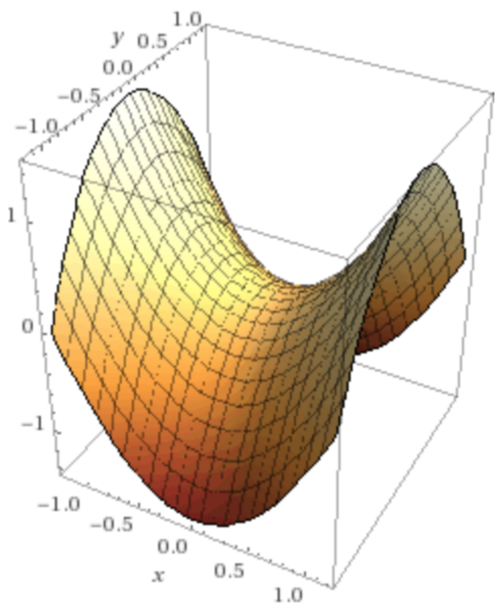

Figure 21.1: A graph of $f(x,y) = x^2 - y^2$ where $(0,0)$ is a saddle point.

Definition 21.3. A critical point $a \in U$ of f is called a **saddle point** if $\exists \epsilon > 0$ such that $\forall \epsilon' \in (0, \epsilon)$, $\exists x, y B_{\epsilon'}(a)$

$$f(x) < f(a) < f(y)$$

(i.e. one side goes up and one side goes down).

21.5 Second derivative test

This was covered in class but was re-visited (with corresponding proof) later on.

22 February 28, 2018

22.1 Bilinear and quadratic forms

Definition 22.1. H is **bilinear** on $\mathbb{R}^n$ i.e. $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that

$$H(av + bw, u) = aH(v,u) + bH(w,u)$$
$$H(v, aw + bu) = aH(v,w) + bH(v,u)$$

where $a, b \in \mathbb{R}$ and $u, v, w \in \mathbb{R}^n$.

Let $e_1, \ldots, e_n$ be standard basis of $\mathbb{R}^n$. So

$$x = \sum_{i=1}^{n} x_i e_i \quad y = \sum_{j=1}^{n} y_j e_j$$

where $x, y \in \mathbb{R}^n$.

$$H(x,y) = \sum_{i,j=1}^{n} H(e_i, e_j) x_i y_j$$

by bilinearity (we can factor out each e_k).
Define $H_{ij} = H(e_i, e_j)$ (entries of an $n \times n$ real matrix that represents the bilinear form H with respect to the basis $\{e_1, \ldots, e_n\}$).

$$\begin{aligned} H(x,y) &= \sum_{i,j=1}^{n} H_{ij} x_i y_j = x^T H y \\ H(y,x) &= \sum_{i,j=1}^{n} H_{ij} y_i x_j = \sum_{i,j=1}^{n} H_{ji} y_j x_i \end{aligned}$$

We say H is **symmetric** if $H(x,y) = H(y,x)\ \forall x, y \in \mathbb{R}^n$. Clear: H is symmetric iff $H_{ij} = H_{ji}$.
From now on suppose H is a **symmetric bilinear form**. Q is a homogeneous 2nd degree polynomial (no terms have degree < 2) on $\mathbb{R}^n$.
In particular it's **continuous** on $\mathbb{R}^n$. **Notice:** if $\lambda \in \mathbb{R}$

$$Q(\lambda x) = \lambda^2 Q(x) \tag{22.1}$$

Aside: if we know Q is the quadratic form of some *symmetric bilinear form* H, then H can be recovered from Q via the "polarization" identity:

$$\begin{aligned} Q(x+y) = H(x+y, x+y) &= H(x,x) + 2H(x,y) + H(y,y) \\ &= Q(x) + Q(y) + 2H(x,y) \end{aligned}$$

Definition 22.2. Notice $Q(0) = 0$ **always**.

1. We say Q is **positive definite** if $Q(x) > 0\ \forall x \neq \vec{0}$.
2. We say Q is **positive semi-definite** if $Q(x) \geq 0\ \forall x \in \mathbb{R}^n$.
3. We say Q is **negative definite** if $Q(x) < 0\ \forall x \neq \vec{0}$.
4. We say Q is **negative semi-definite** if $Q(x) \leq 0\ \forall x \in \mathbb{R}^n$.
5. We say Q is **indefinite** if $\exists x, y \in \mathbb{R}^n$ such that $Q(x) > 0$, $Q(y) < 0$.

   For indefinite, *non-degenerate* means no $z \neq \vec{0} \Rightarrow Q(z) = 0$. *Degenerate* if there is such a z.

Aside: This property of Q is connected to the signs of the eigenvalues of any matrix representing the bilinear form H. The number of positive, negative, and 0 eigenvalues are independent of choice of basis (the quantity is the same, but they may be different valued). So

1. positive definite $\iff$ only positive eigenvalues
2. positive semi-definite $\iff$ only non-negative eigenvalues
3. Similarly for negative cases.
4. indefinite $\iff$ $\exists$ at least one positive and one negative eigenvalue (there may exist 0 eigenvalues too).

Lemma 22.1. Let Q be a quadratic form associated to symmetric bilinear form of H.

1. If Q is positive definite, $\exists M > 0$ such that $Q(x) \geq M\|x\|^2$ $\forall x \in \mathbb{R}^n$.
2. If Q is negative definite, $\exists M > 0$ such that $Q(x) \leq -M\|x\|^2$ $\forall x \in \mathbb{R}^n$.

Proof. Let

$$S^{n-1} = \partial(B_1(\vec{0})) = \{x \in \mathbb{R}^n \mid \|\vec{x}\| = 1\}$$

(called unit sphere in $\mathbb{R}^n$).
S^{n-1} is **compact** (closed and bounded). Why is it closed? We have $F : \mathbb{R}^n \to \mathbb{R}$ and $F(x) = \|x\|^2$ where F is continuous. Note that $S^{n-1} = F^{-1}(\{1\})$ where $\{1\}$ is closed in $\mathbb{R}$. Thus the inverse image of closed is closed.
Q is continuous on compact set S^{n-1}. By EVT $\exists$ some $M \in \mathbb{R}$ such that

$$Q(x) \geq M \quad \forall x \in S^{n-1}$$

and some $x_0 \in S^{n-1}$ where $Q(x_0) = M$.
But since Q is positive definite and $x_0 \neq \vec{0}$ (it's on unit sphere), then $M = Q(x_0) > 0$. Let $x \neq \vec{0} \in S^{n-1}$, so we have $\frac{x}{\|x\|} \in S^{n-1}$ ($\|x\| = 1$). By eq. (22.1) we have

$$M \leq Q\left(\frac{x}{\|x\|}\right) = \frac{1}{\|x\|^2}Q(x) \Rightarrow Q(x) \geq M\|x\|^2$$

for $x \neq \vec{0}$ and trivial for $x = \vec{0}$.
Negative definite case similar. $\square$

22.2 Second derivative test

From the above lemma, we can prove the second derivative test.

Theorem 22.1. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(U)$ (we usually only consider C^2 functions for 2nd derivative tests). Let a be a critical point for f ($(\nabla f)(a) = \vec{0}$).
Let $H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$ and H be the Hessian of f at a (symmetric $n \times n$ matrix represents symmetric bilinear form with associated quadratic form Q).

1. If Q is **positive definite**, then f has a **local min** at a.
2. If Q is **negative definite**, then f has a **local max** at a.
3. If Q is **indefinite**, then a is a **saddle point** of f.

(otherwise test fails and *any of the 3* can happen).

Proof. By Taylor's theorem, since $f \in C^2(U)$ we have

$$f(x) = f(a) + (\nabla f)(a) \cdot (x - a) + \frac{1}{2}\sum_{i,j}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)(x_i - a_i)(x_j - a_j) + R(x)$$

note that $(\nabla f)(a) = 0$ since a is a critical point. We also have

$$\lim_{x \to a} \frac{R(x)}{\|x - a\|^2} = 0$$

for all x in some ball $B_r(a) \subseteq U$.
Let $y = x - a$. Then

$$f(x) = f(a) + \frac{1}{2}\sum_{i,j}^{n} H_{ij} y_i y_j + R(y)$$

for $x \in B_r(a0$ and

$$\lim_{y \to 0} \frac{R(y)}{\|y\|^2} = 0 \tag{22.2}$$

Rewriting with $Q(y) = \sum_{i,j}^{n} H_{ij} y_i y_j$ we have

$$f(x) = f(a) + \frac{1}{2}Q(y) + R(y)$$

By eq. (22.2) $\exists \delta > 0$ (WLOG $\delta < r$) such that

$$\|y\| < \delta \Rightarrow \frac{|R(y)|}{\|y\|^2} < \frac{M}{2}$$

So we have

$$\frac{-M}{2} < \frac{R(y)}{\|y\|^2} < \frac{M}{2} \Rightarrow \frac{-M}{2}\|y\|^2 < R(y) < \frac{M}{2}\|y\|^2$$

Case 1: Q is positive definite By lemma, $\exists M > 0$ such that $Q(y) \geq M\|y\|^2$ $\forall y$. Thus we have

$$\begin{aligned} f(x) &= f(a) + \frac{1}{2}Q(y) + R(y) \\ &\geq f(a) + \frac{M}{2}\|y\|^2 - \frac{M}{2}\|y\|^2 \\ &= f(a) \quad \forall x \in B_\delta(a) \end{aligned}$$

so f has a local min at a.

Case 2: Q is negative definite $\exists M > 0$ such that $Q(x) =\leq -M\|y\|^2$ by lemma. So

$$\begin{aligned} f(x) &= f(a) + \frac{1}{2}Q(y) + R(y) \\ &\leq f(a) - \frac{M}{2}\|y\|^2 + \frac{M}{2}\|y\|^2 \\ &= f(a) \end{aligned}$$

so f has a local max at a.

Case 3: Q is indefinite $\exists y, \tilde{y}$ non-zero in $\mathbb{R}^n$ with

$$Q(y) = A > 0$$
$$Q(\tilde{y}) = -\tilde{A} < 0$$

where $A, \tilde{A} > 0$.

Define for $t \in \mathbb{R}$

$$y_t = ty$$
$$\tilde{y}_t = t\tilde{y}$$

By eq. (22.2) $\exists \delta > 0$ such that for any $z \in \mathbb{R}^n$ and $\|z\| < \delta$ and

$$\frac{|R(z)|}{\|z\|^2} < \frac{1}{2}\min\{\frac{A}{\|y\|^2}, \frac{\tilde{A}}{\|\tilde{y}\|^2}\} = \epsilon$$

(i.e. $-\|z\|^2\epsilon < R(z) < \epsilon\|z\|^2$).

Let $|t|$ be sufficient small so $y_t, \tilde{y}_t \in B_\delta(0)$ i.e.

$$|t| < \min\{\frac{\delta}{\|y\|}, \frac{\delta}{\|\tilde{y}\|}\}$$

Let $x = a + y$ such that

$$\begin{aligned} f(a + y_t) &= f(a) + \frac{1}{2}Q(y_t) + R(y_t) \\ &= f(a) + \frac{t^2}{2}A + R(y_t) \end{aligned}$$

similarly $f(a + \tilde{y}_t) = f(a) - \frac{t^2}{2}A + R(\tilde{y}_t)$.

Then we have

$$\begin{aligned} f(a + y_t) &= f(a) + \frac{t^2}{2}A + R(y_t) \\ &> f(a) + \frac{t^2}{2}A - \epsilon\|y_t\|^2 \qquad > f(a) + \frac{t^2}{2}A - \frac{1}{2}\frac{A}{\|y\|^2}\|y_t\|^2 \\ &= f(a) + \frac{t^2}{2}A - \frac{t^2}{2}A \\ &= f(a) \end{aligned}$$

Similarly

$$
\begin{aligned}
f(a+\tilde{y}_t) &= f(a) - \frac{t^2}{2}A + R(\tilde{y}_t) \\
&< f(a) - \frac{t^2}{2}A + \frac{1}{2}\frac{\tilde{A}}{\|\tilde{y}\|^2}\|\tilde{y}_t\|^2 \\
&= f(a)
\end{aligned}
$$

So a is a saddle point of f.

$\square$

23 March 2, 2018

23.1 Examples where 2nd derivative test fails

Example 23.1.

$$
\begin{aligned}
f(x,y) &= x^4 + y^2 \\
g(x,y) &= -x^4 - y^2 \\
h(x,y) &= x^3 + y^2
\end{aligned}
$$

all have **one** critical point at $(0,0)$.
Their Hessians at $(0,0)$ are

$$
\begin{aligned}
(\text{Hess } f)_{(0,0)} &= \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \\
(\text{Hess } g)_{(0,0)} &= \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \\
(\text{Hess } h)_{(0,0)} &= \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}
\end{aligned}
$$

they are all neither positive definite, negative definite, nor indefinite ($\exists x \neq \vec{0}$ such that $x^T H x = 0$ so no definite cases; all either positive or negative elements so not indefinite).
Test fails so we cannot apply it. The point $(0,0)$ for any of the three functions is either a local min, local max, or a saddle point (unknown).

23.2 Matrix norms

Let $\mathbb{R}^{n\times n}$ space of $n \times n$ real matrices, which $\approx \mathbb{R}^{n^2}$ as a vector space.

Definition 23.1. Define the **norm** on $\mathbb{R}^{n\times n}$ by taking the usual *Euclidean norm* on $\mathbb{R}^{n^2}$

$$
\|A\|^2 = \sum_{i,j=1}^{n} A_{ij}^2
$$

Claim. Let $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n \approx \mathbb{R}^{n\times 1}$, so $Ax \in \mathbb{R}^{n\times 1} \approx \mathbb{R}^n$.

We claim

$$\|Ax\| \leq \|A\|\|x\| \quad \forall x \in \mathbb{R}^n$$

(just like for op norms).

Proof. We have

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Thus we have

$$Ax = \begin{bmatrix} A_1 x \\ A_2 x \\ \vdots \\ A_n x \end{bmatrix} \in \mathbb{R}^{n \times 1}$$

So we can express

$$\begin{aligned} \|Ax\|^2 &= \sum_{i=1}^{n} (A_i \cdot x)^2 \\ &\overset{C-S}{\leq} \sum_{i=1}^{n} \|A_i\|^2 \|x\|^2 \\ &= \|x\|^2 \sum_{i=1}^{n} \|A_i\|^2 \\ &= \|x\|^2 \|A\|^2 \end{aligned}$$

as desired. □

23.3 Inverse function theorem

Theorem 23.1 (Inverse function theorem)**.** Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$ for some $k \geq 1$.
Let $V = f(U)$, let $a \in U$ such that $(Df)_a$ is *invertible* (note that $n = m$ since we require square matrices for invertibility).
Then $\exists$ open set $\tilde{U} \subseteq U$ containing a, an open set $\tilde{V} \subseteq V$ contain $f(a)$, and a map $g : \tilde{V} \to \tilde{U}$ (with $g(\tilde{V}) = \tilde{U}$) such that $g(f(x)) = x \; \forall x \in \tilde{U}$ and $f(g(y)) = y \; \forall y \in \tilde{V}$.
Moreover, $g \in C^k(\tilde{V})$ for the *same* k **and if** $b \in \tilde{V}$ then

$$(Dg)_b = [(Df)_{f^{-1}(b)}]^{-1}$$

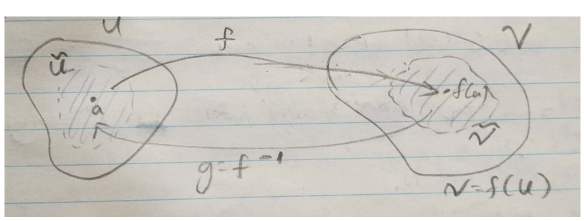

Figure 23.1: Diagram of a function that satisfies the inverse function theorem for some $\tilde{U}, \tilde{V}$.

Lemma 23.1. We claim that

$$f\Big|_{\tilde{U}} : \tilde{U} \to \tilde{V}$$

is a bijection.

23.4 Lemma 1 for inverse function theorem

Lemma 23.2. Lemma 1 (inverse function theorem) Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ (U open) be in $C^k(U)$ for $k \geq 1$. Let $a \in U$. If $(Df)_a$ is **invertible**, then $\exists$ open n'h'd $\tilde{U}$ of a with $\tilde{U} \subseteq U$ and $M > 0$ such that

$$\|f(x_1) - f(x_2)\| \geq M\|x_1 - x_2\| \quad \forall x_1, x_2 \in \tilde{U}$$

Consequence: The restriction of f to $\tilde{U} \subseteq U$ is one-to-one (set $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$).

Proof. Let $L = (Df)_a$. By hypothesis L is invertible. Thus

$$\begin{aligned}
x_1 - x_2 &= L^{-1}L(x_1 - x_2) \\
&= L^{-1}(Lx_1 - Lx_2) \\
\|x_1 - x_2\| &= \|L^{-1}(Lx_1 - Lx_2)\| \\
&\leq \|L^{-1}\|\|Lx_1 - Lx_2\| \\
&= \frac{1}{2M}\|Lx_1 - Lx_2\| \qquad \text{let } 2M = \frac{1}{\|L^{-1}\|}
\end{aligned}$$

So we have

$$\|Lx_1 - Lx_2\| \geq 2M\|x_1 - x_2\| \tag{23.1}$$

Since $f \in C^k(U)$, $k \geq 1$, each $\frac{\partial f_i}{\partial x_j}$ is **continuous** on U. $(Df)_x$ is the $n \times n$ matrix whose (i,j)th entry is $\frac{\partial f_i}{\partial x_j}(x)$. So the entries of the matrix $(Df)_x$ are **continuous functions** on U.
Recall that $L = (Df)_a$ where $a \in U$, so $\exists \epsilon > 0$ such that if $\|x - a\| < \epsilon$ then

$$\sum_{i,j=1}^{n} (\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(a))^2 \|(Df)_x - L\|^2 < \frac{M^2}{n} \tag{23.2}$$

(take $\|x - a\| < \delta_{ij}$ which implies $\|\frac{\partial f_i}{\partial x_j}(x) - \frac{\partial f_i}{\partial x_j}(a)\|^2 < \frac{M^2}{n^2}$, then $\epsilon = \min\{\delta_{ij}\}$ and the result follows).
Define $h(x) = f(x) - Lx$, where $h : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$.
Note that $(DL)_x = L$ (for any linear maps as shown on midterm Q3).
So

$$\begin{aligned}
(Dh)_x &= (Df)_x - (DL)_x \\
(Dh)_x &= (Df)_x - L
\end{aligned}$$

Thus we have

$$\|(Dh)_x\| = \|(Df)_x - L\| < \frac{M}{\sqrt{n}} \tag{23.3}$$

if $x \in B_\epsilon(a)$ by eq. (23.2).
Apply MVT to $h \in C^1(U)$ on $B_\epsilon(a)$ (i.e. Taylor's theorem for $p + 1 = 1$ applied to each component of $h \in C^1$):

Given $x_1, x_2 \in B_\epsilon(a)$, $\exists$ c_i's on line segment between x_1 and x_2 such that

$$h_i(x_2) - h_i(x_1) = (\nabla h_i)(c_i)(x_2 - x_1)$$
$$\|h_i(x_2) - h_i(x_1)\|^2 \overset{C-S}{\leq} \|(\nabla h_i)(c_i)\|^2 \|x_2 - x_1\|^2$$

Summing over all components h_i

$$\begin{aligned}
\sum_{i=1}^{n} \|h_i(x_2) - h_i(x_1)\|^2 &\leq \sum_{i=1}^{n} \|(\nabla h_i)(c_i)\|^2 \|x_2 - x_1\|^2 \\
&\overset{eq.\ (23.3)}{\leq} \sum_{i=1}^{n} \frac{M^2}{N} \|x_2 - x_1\|^2 \qquad\qquad \|(\nabla h_i)(c_i)\|^2 \leq \|(Dh)_{c_i}\|^2 \\
&= M^2 \|x_2 - x_1\|^2
\end{aligned}$$

Therefore $\|h(x_2) - h(x_1)\| \leq M\|x_2 - x_1\|$ for all $x_1, x_2 \in B_\epsilon(a)$. This can be written as

$$\|f(x_2) - Lx_2 - (f(x_1) - Lx_1)\| \leq M\|x_2 - x_1\|$$
$$\|(f(x_2) - f(x_1)) - (Lx_2 - Lx_1)\| \leq M\|x_2 - x_1\|$$

Note that $\|A - B\| = \|B - A\| \geq \|B\| - \|A\|$ (by triangle inequality). Thus for $A = f(x_2) - f(x_1)$ and $B = Lx_2 - Lx_1$

$$\begin{aligned}
M\|x_2 - x_1\| \geq \|A - B\| &\geq \|B\| - \|A\| \\
&= \|L(x_2 - x_1)\| - \|f(x_2) - f(x_1)\| \\
&\overset{eq.\ (23.1)}{\geq} 2M\|x_2 - x_1\| - \|f(x_2) - f(x_1)\|
\end{aligned}$$

Thus the lemma immediately follows. □

24 March 5, 2018

24.1 Lemma 2 for inverse function theorem

Lemma 24.1 (Lemma 2 (inverse function theorem)). Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$, $k \geq 1$. Let $V = f(U)$. If f is **injective (1-1)** on U *and* $(Df)_a$ is invertible $\forall x \in U$, then V is open.
(Note: on A8 Q5 we use Lemma 1 and Lemma 2 to show that you can remove the injective hypothesis from Lemma 2 and the result still holds).

Proof. We want to show V is open.
Let $b \in V$. We want to find an open ball centred at b completely contained in V.
Since $f : U \to V$ is an injection then $\exists$ one $a \in U$ with $f(a) = b$ ($a = f^{-1}(b)$).
Because U is open, $\exists \epsilon > 0$ such that

$$K = \overline{B_\epsilon(a)} \subseteq U$$

(take an open ball by openness, then take a closed ball slightly and strictly smaller).

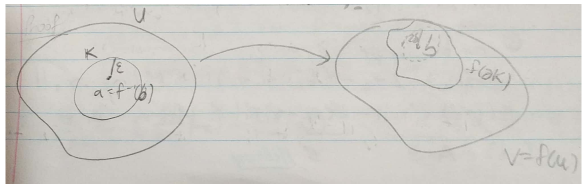

Figure 24.1: Diagram of proof for Lemma 2 for the inverse function theorem.

Note ∂K is compact because it's closed and bounded. Since f is continuous, $f(\partial K)$ is a compact subseteq of $\mathbb{R}^n$ contained in V.
Since $a \notin \partial K$ so $b \notin f(\partial K)$ because f is 1-1 on U.
Since $f(\partial K)$ is compact and hence closed, and $b \notin f(\partial K)$ so $\exists \delta > 0$ such that

$$B_{2\delta}(b) \cap f(\partial K) = \varnothing \tag{24.1}$$

($b \in (f(\partial K))^c$ which is open so open ball).

Claim. For any $y \in B_\delta(b)$, $y = f(x)$ for some $x \in K$.
If the claim holds, then $B_\delta(b) \subseteq f(K) \subseteq f(U) = V$ then V is open. It remains to prove the claim.
Let $y \in B_\delta(b)$.
Define $\phi : U \to \mathbb{R}$ by $\phi(x) = \|f(x) - y\|^2$. Since $f \in C^k(U)$ for $k \geq 1$ and $x \mapsto \|x\|^2$ is in $C^\infty(\mathbb{R}^n)$ by the chain rule $\phi \in C^k(U)$ for $k \geq 1$ so ϕ is differentiable.
Since $K \subseteq U$ is compact by EVT $\exists$ point $x_0 \in K$ such that $\phi(x_0) \leq \phi(x)$ $\forall x \in K$ (global min).
TODO(richardwu): Why is anything from below until the part where we know x_0 is already a local minimum necessary? Thus

$$\phi(a) = \|f(a) - y\|^2 = \|b - y\|^2 < \delta^2$$

since $y \in B_\delta(b)$ and $a \in K$, then $\phi(x_0) < \delta^2$ since x_0 is where our global minimum is.
Suppose $x_0 \in \partial K$. Then $f(x_0) \in f(\partial K)$ so $f(x_0) \notin B_{2\delta}(b)$ by eq. (24.1) so $\|f(x_0) - b\| > 2\delta$.
Thus we have

$$\begin{aligned} \|f(x_0) - y\| &\overset{\triangle}{\geq} \|f(x_0) - b\| - \|b - y\| \\ &> 2\delta - \delta > \delta \end{aligned}$$

So $\phi(x_0) > \delta^2$ which is a contradiction.
Hence $x_0 \notin \partial K$ so $x_0 \in int(K)$.
ϕ has a local minimum at x_0 (since it is a global min).
Hence x_0 is an optimal point of ϕ so $(\nabla\phi)(x_0) = 0$.

$$\begin{aligned} \phi(x) &= \|f(x) - y\|^2 \\ &= \sum_{k=1}^{n} (f_k(x) - y_k)^2 \\ \frac{\partial \phi}{\partial x_j} &= \sum_{k=1}^{n} 2(f_k(x) - y_k) \cdot \frac{\partial f_k}{\partial x_j} \end{aligned}$$

Thus

$$(\nabla\phi)(x_0) = 2(f(x_0) - y)^T (Df)_{x_0} = 0$$

By hypothesis, $(Df)_x$ is invertible $\forall x \in U$, $x_0 \in K \subseteq U$ so $(Df)_{x_0}$ is invertible which implies that $f(x_0) - y = 0_{1\times n}$.

So $f(x_0) = y$ which proves the claim.

$\square$

24.2 Lemma 3 for inverse function theorem

Lemma 24.2 (Lemma 3 (inverse function theorem))**.** (Same hypothesis as Lemma 1 and 2).
Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$, $k \geq 1$, $V = f(U)$, f is **injective (1-1)** on U and $(Df)_a$ is invertible $\forall x \in U$. Then

$$g = f^{-1} : V = f(U) \to U \subseteq \mathbb{R}^n$$

is **continuous**.

Proof. Let $W \subseteq \mathbb{R}^n$ be open. We needto show $g^{-1}(W) \subseteq V$ is an open subset.

$$\begin{aligned} g^{-1}(W) &= \{y \in V, g(y) \in W\} \\ &= \{y \in V, g(y) \in W \cap U\} \\ &= g^{-1}(W \cap U) \\ &= f(W \cap U) \end{aligned}$$

So we need to show $f(W \cap U)$ is open for all W open on $\mathbb{R}^n$, but this follows by applying Lemma 2 with U replaced by $U \cap W$. $\square$

24.3 Lemma 4 for inverse function theorem

Lemma 24.3 (Lemma 4 (inverse function theorem))**.** (Same hypothesis as Lemma 1, 2, and 3).
Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$, $k \geq 1$, $V = f(U)$, f is **injective (1-1)** on U and $(Df)_a$ is invertible $\forall x \in U$. Then $g : f^{-1} : V \to U$ is **differentiable** at b for all $b \in V$ and

$$(Dg)_b = [(Df)_{f^{-1}(b)}]^{-1}$$

Proof. Let $b \in V$. We will show g is differentiable at b and $(Dg)_b = (Df)_a^{-1}$ where $f(a) = b$.
We need to show $\exists$ linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\lim_{h \to \vec{0}} \frac{g(b+h) - g(b) - T(h)}{\|h\|} = \vec{0}$$

Let $L = (Df)_a$. We'll show $T = L^{-1}$ works (which proves the theorem), that is

$$G(h) = \frac{g(b+h) - g(b) - L^{-1}h}{\|h\|} \to \vec{0}$$

as $h \to \vec{0}$.
Define $\delta(h) = g(b+h) - g(b)$ for $\|h\|$ small. We claim $\exists \epsilon > 0$ such that $\frac{\|\delta(h)\|}{\|h\|}$ is **bounded** for $0 < \|h\| < \epsilon$.
Notice

$$\delta(h) = \|h\| G(h) + L^{-1}(h) \Rightarrow \frac{\|\delta(h)\|}{\|h\|} \leq \|L^{-1}\| + \|G(h)\|$$

(if we knew $\lim_{h \to \vec{0}} G(h) = \vec{0}$ then it follows it is bounded for $0 < \|h\| < \epsilon$ for ϵ sufficient small, but we don't know this: this is exactly what we're trying to prove!)

By Lemma 1, $\exists \tilde{U} \subseteq U$, open n'h'd of a and $M > 0$ such that

$$\|f(x_1) - f(x_2)\| \geq M\|x_1 - x_2\| \quad \forall x_1, x_2 \in \tilde{U} \tag{24.2}$$

By Lemma 2 since f is 1-1 on $\tilde{U}$, $\tilde{V} = f(\tilde{U})$ is open.
Choose $\epsilon > 0$ such that $b + h \in \tilde{V} = f(\tilde{U})$ when $\|h\| < \epsilon$.
So $b + h = f(x)$ for some $x \in \tilde{U}$ thus $g(b + h) = x$.
Applying eq. (24.2) with $x_1 = g(b + h) \Rightarrow f(x_1) = b + h$ and $x_2 = f(b) \Rightarrow f(x_2) = b$ we get

$$\|(b + h) - b\| \geq M\|g(b + h) - g(b)\|$$

So

$$\frac{\|\delta(h)\|}{\|h\|} = \frac{\|g(b + h) - g(b)\|}{\|h\|} \leq \frac{1}{M}$$

if $0 < \|h\| < \epsilon$, thus we've proven the claim that it is bounded.
Thus

$$\begin{aligned} G(h) &= \frac{\delta(h) - L^{-1}h}{\|h\|} \\ &= \frac{-L^{-1}(h - L\delta(h))}{\|\delta(h)\|} \cdot \frac{\|\delta(h)\|}{\|h\|} \end{aligned}$$

We know $\delta(h) \neq 0$ if $h \neq 0$ because g is 1-1 on V, so

$$\|G(h)\| \leq \|L^{-1}\| \cdot \|\frac{h - L\delta(h)}{\|\delta(h)\|}\| \cdot \frac{\|\delta(h)\|}{\|h\|}$$

So it is enough to show that the second term $\to 0$ since the other terms are bounded.
Note that

$$b + h = f(g(b + h)) = f(g(b) + \delta(h)) = f(a + \delta(h))$$

So $h = f(a + \delta(h)) - b = f(a + \delta(h)) - f(a)$.
So we have

$$\frac{h - L\delta(h)}{\|\delta(h)\|} = \frac{f(a + \delta(h)) - f(a) - L\delta(h)}{\|\delta(h)\|}$$

Note as $h \to \vec{0}$ then $\|\delta(h)\| \to 0$ because by Lemma 3 g is continuous on V so $\lim_{h\to 0} g(b + h) - g(b) = \delta(h) = 0$, so the above $\to 0$ since $L = (Df)_a$ and f is differentiable at a.
Thus g is differentiable at b and $(Dg)_b = (Df)_a^{-1} = [(Df)_{g(b)}]^{-1}$. □

25 March 7, 2018

25.1 Lemma 5 for inverse function theorem

(Same hypothesis as Lemma 1, 2, 3, and 4).
Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$, $k \geq 1$, $V = f(U)$, f is **injective (1-1)** on U and $(Df)_a$ is invertible $\forall x \in U$.
For $g : f^{-1} : V \to U$, $g \in C^k(V)$ for the **same** k.

Proof. Lemma 2 says V is open.
Lemma 3 says $g : f^{-1} : V \to U$ is continuous.
Lemma 4 says g is differentiable on V.

$g \circ f : U \to U$ is the identity map. By the chain rule

$$(Dg)_b \circ (Df)_a = I$$

where $b = f(a)$, so $(Dg)_b = [(Df)_{g(b)}]^{-1}$ (we gave another proof last time) where $f(a) = b$ and $a = g(b)$.
Define a map $Dg : V \to \mathbb{R}^{n\times n}$ where

$$Dg : V \xrightarrow{g} U \xrightarrow{Df} W \xrightarrow{i(inverse)} W \tag{25.1}$$
$$b \mapsto g(b) \mapsto (Df)_{g(b)} \mapsto [(Df)_{g(b)}]^{-1} \tag{25.2}$$

where $i = A \to A^{-1}$ is the inverse and $i : W \to W$ is in $C^\infty(W)$ (this follows since $i(A) = \frac{adj(A)}{det(A)}$ where $det(A)$ are entries of degree $n-1$ homoegenous polynomials which are in C^∞). ($W \subseteq \mathbb{R}^{n\times n}$ in fact open subset of invertible matrices).
We want to show $f \in C^k \to g \in C^k$ for $k \geq 1$.
Base case ($k = 1$): Since $f \in C^1 \Rightarrow Df \in C^0$. $g \in C^0$ by Lemma 3, so $Dg \in C^0$ by eq. (25.1) so $g \in C^1$.
We've shown this for $k = 1$.
Inductive case: suppose $f \in C^{r-1}(U)$, $r - 1 \geq 1$ implies $g \in C^{r-1}(V)$.
We want to show this is true for r.
Let $f \in C^r \Rightarrow f \in C^{r-1} \Rightarrow g \in C^{r-1}$ by inductive hypothesis.
Also $Df \in C^{r-1}$ so $Dg = i \circ (Df) \circ g \in C^{r-1}$ (they're all in C^{r-1}).
Thus $g \in C^r$. □

25.2 Proof of inverse function theorem

Recall the theorem:

Theorem 25.1 (Inverse function theorem)**.** Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be in $C^k(U)$ for some $k \geq 1$.
Let $V = f(U)$, let $a \in U$ such that $(Df)_a$ is *invertible* (note that $n = m$ since we require square matrices for invertibility).
Then $\exists$ open set $\tilde{U} \subseteq U$ containing a, an open set $\tilde{V} \subseteq V$ contain $f(a)$, and a map $g : \tilde{V} \to \tilde{U}$ (with $g(\tilde{V}) = \tilde{U}$) such that $g(f(x)) = x \ \forall x \in \tilde{U}$ and $f(g(y)) = y \ \forall y \in \tilde{V}$.
Moreover, $g \in C^k(\tilde{V})$ for the *same* k **and if** $b \in \tilde{V}$ then

$$(Dg)_b = [(Df)_{f^{-1}(b)}]^{-1}$$

Proof. Since $f \in C^1$ and $det((Df)_a) \neq 0$, $\exists$ open n'h'd of W of a, $W \subseteq U$ such that $(Df)_x$ is invertible $\forall x \in W$ (this follows since $det((Df)_a)$ is some function of a and is not 0. Since $det((Df)_a)$ is continuous there is an open n'h'd where it is also not 0 and thus making Df invertible on this W).
Lemma 1 says $\exists \tilde{U} \subseteq W$, $\tilde{U}$ open n'h'd of a such that $f\Big|_{\tilde{U}}$ is one-to-one.
Lemma 2 says $\tilde{V} = f(\tilde{U})$ is open.
Lemma 3 says $f^{-1} : \tilde{V} \to \tilde{U}$ is continuous.
Lemma 4 says f^{-1} is differentiable.
Lemma 5 says $f^{-1} \in C^k(\tilde{V})$ for the same k. □

25.3 Example of inverse function theorem

Example 25.1. Let $(x, y) = f(u, v) = (uv, u^2 + v^2)$ where $f : \mathbb{R}^2 \to \mathbb{R}^2$. Note that $f \in C^\infty(\mathbb{R}^2)$ since f_i are polynomials.

We want to prove f^{-1} exists and is C^∞ in some nonempty open set containing $(2,5)$.
For $f(a,b)=(2,5)$, find all points $(u,v)\in\mathbb{R}^2$ such that $f(u,v)=(2,5)$.

$$\begin{aligned} uv = 2 &\Rightarrow v = \frac{2}{u} \\ u^2+v^2 = 5 &\Rightarrow u^2 + \frac{4}{u^2} = 5 \\ &\Rightarrow u^4 - 5u^2 + 4 = 0 \\ &\Rightarrow (u^2-1)(u^2-4) = 0 \end{aligned}$$

So $(u,v) = \{(1,2),(-1,-2),(2,1),(-2,-1)\}$ (4 points).
Note that

$$(Df)_{(u,v)} = \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix} = \begin{bmatrix} v & u \\ 2u & 2v \end{bmatrix}$$

Thus $det((Df)_{(u,v)} = 2v^2 - 2u^2 = 2(v^2-u^2) \neq 0$ for any of our points.
So by the Inverse Function Theorem (IFT), for any of these 4 points (a,b) there is an open n'h'd $\tilde{U}$ of (a,b) and an open n'h'd of $\tilde{V}$ of $(2,5)$ such that $f:\tilde{U}\to\tilde{V}$ is invertible and $f^{-1}\in C^\infty(\tilde{V})$.
Aside: Let's see this explicitly. For $x = uv$ and $y = u^2+v^2$, we solve for u, v as functions of x, y.

$$\begin{aligned} v &= \frac{x}{u} \\ y &= u^2 + \frac{x^2}{u^2} \Rightarrow u^4 - u^2 y + x^2 = 0 \end{aligned}$$

So

$$\begin{aligned} u^2 &= \frac{y\pm\sqrt{y^2-4x^2}}{2} \Rightarrow u = \pm\sqrt{\frac{y\pm\sqrt{y^2-4x^2}}{2}} \\ v &= \pm x\sqrt{\frac{2}{y\pm\sqrt{y^2-4x^2}}} \end{aligned}$$

Note we have 4 solutions for u and v (4 combinations of $\pm$).
Each are defined on a n'h'd of exactly one of 4 points (a,b) we found before: note when $x=2$ and $y=5$ so $u^2 = \frac{5\pm3}{2} = 4$ or 1 which means $u = 2,-2,1,-1$.

25.4 Informal motivation for implicit function theorem

Informal motivation: Let $f: W\subseteq\mathbb{R}^{n+m}\to\mathbb{R}^q$ (where $\mathbb{R}^n\times\mathbb{R}^m = \mathbb{R}^{n+m}$).
We write $z\in\mathbb{R}^q = f(y,x)$ where $y\in\mathbb{R}^n$ and $x\in\mathbb{R}^m$.
Question: Can we solve the equation $f(y,x)=0$ for y as a function of x? (we want to find $y=h(x)$ such that $f(h(x),x)=0$ for all x).
This would say that $f(y,x)=0$ **implicitly** defines y as a function of x.
Simplest case: suppose f is linear where $\mathbb{R}^{n+m}\to\mathbb{R}^q$. f is multiplication by a $q\times(n+m)$ matrix.

$$f(y,x) = A\begin{bmatrix} y \\ x\end{bmatrix} = \begin{bmatrix} A_n & A_m\end{bmatrix}\begin{bmatrix} y \\ x\end{bmatrix} = 0$$

where $\begin{bmatrix} y \\ x \end{bmatrix}$ is an $(n+m) \times 1$ matrix (y and x is $n \times 1$ and $m \times 1$, respectively, and A_n and A_m are $q \times n$ and $q \times m$ matrices, respectively).
From linear algebra: this system of linear equations will have each $y_1, \ldots, y_n$ uniquely determined by each (arbitrary) choice of $x_1, \ldots, x_m$ if the $q \times n$ matrix A_n is square (i.e. $q = n$) and invertible (full rank) i.e. $det(A_n) \neq 0$.
Idea for general case: if $f \in C^k(W)$, $k \geq 1$, we approximate near $(y_0, x_0) \in W$ by its *linearization* at (y_0, x_0) where $f(y_0, x_0) = 0$:

$$\begin{aligned} L(y,x) &= f(y_0, x_0) + (Df)_{(y_0,x_0)} \begin{bmatrix} y - y_0 \\ x - x_0 \end{bmatrix} \\ &= A \begin{bmatrix} y - y_0 \\ x - x_0 \end{bmatrix} \end{aligned}$$

We'll see that *locally* (near (y_0, x_0)) the solvability of $f(y,x) = 0$ for y as a function of $y = h(x)$ of x is **equivalent** to the solvability of the linear system $L(y,x) = 0$.
That is if $\frac{\partial f_i}{\partial y_j}(y_0, x_0)$ is invertible ($q = n$) then we expect $\exists \tilde{W} \subseteq W$ open n'h'd of (y_0, x_0) such that $f(y,x) = 0$ can be solved for $y = h(x)$ for all $(y,x) \in \tilde{W}$.

26 March 9, 2018

26.1 Implicit function theorem

Theorem 26.1 (Implicit function theorem)**.** Let $f : W \subseteq \mathbb{R}^{n+m} \to \mathbb{R}^n$ (note $q = n$!) be in $C^k(W)$ for $k \geq 1$. Suppose $f(y_0, x_0) = 0$ for some $(y_0, x_0) \in W$.
Let A be the $n \times n$ matrix where $A_{ij} = \frac{\partial f_i}{\partial y_j}(y_0, x_0)$ (A_n from before).
If $det(A) \neq 0$ (i.e. A invertible) then $\exists W' \subseteq W$ open n'h'd of (y_0, x_0) and an open n'h'd U of x_0 in $\mathbb{R}^m$ and a function $h : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$, $h \in C^k(U)$ for the same k such that

$$\{(y,x) \in W' \mid f(y,x) = 0\} = \{(h(x), x), x \in U\}$$

i.e. on W', the points where $f = 0$ can be expressed as y as a function of x.

Proof. Define $F : W \subseteq \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ by

$$F(y,x) = (f(y,x), x)$$

where $F \in C^k(W)$ (components are C^k since $f(y,x) \in C^k$ and $x \in C^\infty$).
Note that

$$(DF)_{(y_0,x_0)} = \begin{bmatrix} \frac{\partial f_i}{\partial y_j}(y_0, x_0) & \frac{\partial f_i}{\partial x_j}(y_0, x_0) \\ \frac{\partial x_i}{\partial y_j}(y_0, x_0) & \frac{\partial x_i}{\partial x_j}(y_0, x_0) \end{bmatrix} = \begin{bmatrix} A & * \\ 0 & I_{m \times m} \end{bmatrix}$$

wher DF is an $n + m \times n + m$ matrix, the top quadrant has n rows and the left quadrant has n columns (bottom and right quadrants have m rows and columns respectively).
Note that $det((DF)_{(y_0,x_0)}) = det(A)det(I) \neq 0$ (determinant of triangular matrix is the product of the diagonals).
We can apply the inverse function theorem on F. By InvFT, $\exists W' \subseteq W$ open n'h'd of (y_0, x_0) and $V' \subseteq \mathbb{R}^{n+m}$ open n'h'd of $F(y_0, x_0) = (0, x_0)$ such that $V' = F(W')$ and $F : W' \to V'$ is invertible with $F^{-1} \in C^k(V')$.
Define $U \subseteq \mathbb{R}^m$ by $U = \{x \in \mathbb{R}^m \mid (0,x) \in V'\}$.
Note $x_0 \in U$.

Claim. U is open.
Let $\bar{x} \in U \iff (0, \bar{x}) \in V'$, but V' is open so $\exists \epsilon > 0$ such that $B_\epsilon(0, \bar{x}) \subseteq V$. Note that

$$(y, x) \in B_\epsilon(0, \bar{x}) \iff \|y - 0\|^2 + \|x - \bar{x}\|^2 < \epsilon^2$$

So

$$(0, x) \in B_\epsilon(\bar{x}) \iff \|x - \bar{x}\|^2 < \epsilon^2 \iff x \in B_\epsilon(\bar{x})$$

where the second $\iff$ follows from before since $\|y - 0\|^2 \geq 0$.
Since $B_\epsilon(\bar{x}) \subseteq U$ so U is open.

Let $G = F^{-1} \in C^k(V')$ such that

$$G(v, u) = F^{-1}(v, u) = (y, x) = (G_1(v, u), G_2(v, u)) \tag{26.1}$$

where $F(y, x) = (f(y, x), x) = (v, u)$.
Since the second component of F just takes $x \mapsto x$ (identity), $G_2(v, u) = u \Rightarrow G_2(v, u) = x$.
So $F^{-1}(0, u) = (G(0, u), u)$. Let $h : U \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be $h(u) = G(0, u)$, $h \in C^k(U)$.
So $F^{-1}(0, x) = (h(x), x)$ for all $x \in U$. Note that

$$\begin{aligned}
&\{(y, x) \in W' \mid f(y, x) = 0\} \\
=&(F^{-1} \circ F)\{(y, x) \in W' \mid f(y, x) = 0\} \\
=&F^{-1}\{F(y, x) \mid (y, x) \in W' \text{ and } f(y, x) = 0\} \\
=&F^{-1}\{(0, x) \mid (0, x) \in V'\} \\
=&\{F^{-1}(0, x) \mid x \in U\} \\
=&\{(h(x), x) \mid x \in U\}
\end{aligned}$$

□

26.2 Example of implicit function theorem

Example 26.1. Given $x_0, y_0, u_0, v_0, s_0, t_0$ **nonzero** real numbers that satisfy the simultaneous equations

$$\begin{aligned}
&u^2 + sx + ty = 0 \\
&v^2 + tx + sy = 0 \\
&2s^2x + 2t^2y - 1 = 0 \\
&s^2x - t^2y = 0
\end{aligned}$$

(this is almost impossible to solve explicitly: we may only want to know it exists).
Show that $\exists$ smooth (C^∞) functions $u(x, y), v(x, y), s(x, y), t(x, y)$ defined on an open n'h'd of (x_0, y_0) such that u, v, s, t satisfy the equations and

$$\begin{aligned}
u(x_0, y_0) &= u_0 \\
v(x_0, y_0) &= v_0 \\
s(x_0, y_0) &= s_0 \\
t(x_0, y_0) &= t_0
\end{aligned}$$

We'll apply the implicit function theorem. Define $f : \mathbb{R}^6 = \mathbb{R}^{4+2} \to \mathbb{R}^4$ where

$$f(u, v, s, t, x, y) = \begin{bmatrix} u^2 + sx + ty \\ v^2 + tx^s y \\ 2s^2x + 2t^2y - 1 \\ s^2x - t^2y \end{bmatrix} \in \mathbb{R}^4$$

By hypothesis, $f(u_0, v_0, s_0, t_0, x_0, y_0) = 0$. Also

$$Df = \begin{bmatrix} 2u & 0 & x & y & \dots \\ 0 & 2v & y & x & \dots \\ 0 & 0 & 4sx & 4ty & \dots \\ 0 & 0 & 2sx & -2ty & \dots \end{bmatrix}$$

So we have

$$A = \begin{bmatrix} 2u_0 & 0 & x_0 & y_0 \\ 0 & 2v_0 & y_0 & x_0 \\ 0 & 0 & 4s_0x_0 & 4t_0y_0 \\ 0 & 0 & 2s_0x_0 & -2t_0y_0 \end{bmatrix}$$

where $det(A) = (2u_0)(2v_0)(-8s_0x_0t_0y_0 - 8s_0x_0t_0y_0) = 64u_0v_0s_0t_0x_0y_0 \neq 0$ since they're all non-zero.
So u, v, s, t exist by be the implicit function theorem in a n'h'd of (x_0, y_0) and are in C^∞ (since f is in C^∞, polynomials).

26.3 Constraint optimization (methods of Lagrange multiplier)

Suppose we want to optimize (maximize or minimize) a real-valued function $f(x_1, \dots, x_n)$ subject to k constraints

$$g_1(x_1, \dots, x_n) = 0$$
$$\vdots$$
$$g_k(x_1, \dots, x_n) = 0$$

where $1 \leq k \leq n$, $f, g_1, \dots, g_k : U \subseteq \mathbb{R}^n \to \mathbb{R}$.

Example 26.2. Optimize $f(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2$ subject to the constraint $x_1 \cdot \ldots \cdot x_n = 1$ and $x_1 + \ldots + x_n = 0$.
Idea: Suppose $a \in U$ was a *local max or local min* of f subject to the constraints $g_i(x) = 0$, $i = 1, \dots, k$.
It's **no longer true that** $(\nabla f)(a) = \vec{0}$. This time, we do not permit moving in arbitrary directions at a because of the constraints.

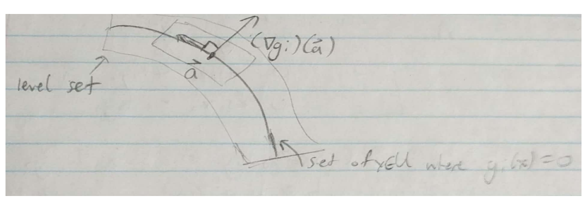

Figure 26.1: Example of a given level set (a given g_i) as a "wavy plane" and the intersection of all constraints i.e. $x \in U$ restricted to all constraints as the line.

(on assignment 7, we showed that $(\nabla g_i)(a)$ is $\perp$ to any tangent vector to the level set $\{x \mid g_i(x) = 0\}$ at a).

So we only consider direction u such that u is **tangent** to all level sets $g_i(x) = 0$, $i = 1, \dots, k$, that is

$$(\nabla g_i)(a) \cdot u = 0 \quad \forall i = 1, \dots, k$$

In these directions we've assumed f has a local extrema at a i.e.

$$(\nabla f)(a) \cdot u = 0$$

for all such u's.
So we must have

$$(\nabla f)(a) = \lambda_1 (\nabla g_1)(a) + \ldots + \lambda_k (\nabla g_k)(a)$$

for some $\lambda_1, \dots, \lambda_k$. These are called **Lagrange multipliers** (there are actually n equations here, one for each x_i/a_i).
Along with our k constraints $g_i = 0$, we have $k+n$ equations for $k+n$ unknowns $(a_1, \dots, a_n, \lambda_1, \dots, \lambda_k)$.

Example 26.3. Find extrema of $f(x,y) = 4x^2 - 3xy$ subject to constraint $g(x,y) = x^2 + y^2 - 1 = 0$ ($g(x,y)$ is the constraint that (x,y) has to be on the unit circle).
(in EVT, $\exists$ a global maximm of f on this set).
We have

$$\begin{aligned}
&\nabla f + \lambda \nabla g = \vec{0} \\
\Rightarrow &(8x - 3y, -3x) + \lambda(2x, 2y) = 0
\end{aligned}$$

So

$$\begin{aligned}
&8x - 3y + 2x\lambda = 0 \\
&-3x + 2y\lambda = 0 \Rightarrow x = \frac{2}{3}\lambda y \\
&x^2 + y^2 = 1
\end{aligned}$$

So we get (pluggin the 2nd result into the 1st equation)

$$\begin{aligned}
&\frac{16}{3}\lambda y - 3y + \frac{4\lambda^2 y}{3} = 0 \\
\Rightarrow &(4\lambda^2 + 16\lambda - 9)y = 0
\end{aligned}$$

So either $y = 0 \to x = 0$ which is a contradiction of our constraint, or

$$4\lambda^2 + 16\lambda - 9 = 0 \Rightarrow (2\lambda - 1)(2\lambda + 9) = 0$$

so $\lambda = \frac{1}{2}$ or $\lambda = \frac{-9}{2}$.
When $\lambda = \frac{1}{2}$, then $(x,y) \in \{(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}), (\frac{-1}{\sqrt{10}}, \frac{-3}{\sqrt{10}})\}$, which both have values $f(x,y) = \frac{-1}{2}$.
When $\lambda = \frac{-9}{2}$, then $(x,y) \in \{(\frac{-3}{\sqrt{10}}, \frac{1}{\sqrt{10}}), (\frac{3}{\sqrt{10}}, \frac{-1}{\sqrt{10}})\}$, which both have values $f(x,y) = \frac{9}{2}$.
Thus we have our two global mins and two global maxes, respectively (by EVT).

27 March 12, 2018

27.1 Lagrange multipliers

Theorem 27.1 (Lagrange multipliers)**.** Let $1 \leq k \leq n$. Let $W \subseteq \mathbb{R}^n$ (open). Let $f : W \subseteq \mathbb{R}^n \to \mathbb{R}$ and $g : W \subseteq \mathbb{R}^n \to \mathbb{R}^k$ (component functions $g_1, \ldots, g_k$ are the constraint functions).
Let $S = \{w \in W \mid g(x) = 0\}$ (the "constraint" set). Let $a \in S$.
Suppose

1. f has a local extrema at a subject to the constraints $g(x) = 0$ (i.e. f restricted to S has a local extrema at a).
2. $rank((Dg)_a) = k$ (where $(Dg)_a$ is $k \times n$ thus maximal rank).

Then $\exists \lambda \in \mathbb{R}^k$ such that
$$(Df)_a + \lambda(Dg)_a = \vec{0}$$
where $(Df)_a$, λ and $(Dg)_a$ are $1 \times n$, $1 \times k$ and $k \times n$ matrices.
Equivalently,
$$\frac{\partial f_i}{\partial x_j}(a) + \sum_{i=1}^{k} \lambda_i \frac{\partial g_i}{\partial x_j}(a) = 0 \quad \forall j = 1, \ldots, n$$
or
$$(\nabla f)(a) + \sum_{i=1}^{k} \lambda_i (\nabla g_i)(a) = 0$$

Proof. Note that $k \times n$ matrix $(Dg)_a$ has rank k by hypothesis
$$\begin{aligned} (Dg)_a &= \left[\tfrac{\partial g_i}{\partial x_j}(a)\right] \\ &= \begin{bmatrix} (\nabla g_1)(a) \\ \vdots \\ (\nabla g_k)(a) \end{bmatrix} \end{aligned}$$
where $(\nabla g_i)(a)$ row vectors are linearly independent. From linear algebra, $\exists$ $k \times k$ minor (matrix) of $(Dg)_a$ with non-zero determinant.
WLOG by re-ordering the coordinates we can assume
$$(Dg)_a = \begin{bmatrix} A & B \end{bmatrix}$$
where A and B is $k \times k$ and $k \times (n-k)$, respectively, and $det(A) \neq 0$.
By the implicit function theorem, $\exists U \subseteq \mathbb{R}^{n-k}$ open containing $(a_{k+1}, \ldots, a_n) = \bar{a}$ and $h : U \subseteq \mathbb{R}^{n-k} \to \mathbb{R}^k$ with $h \in C^1(U)$ such that
$$h(x_{k+1}, \ldots, x_n) = (x_1, \ldots, x_k)$$
and
$$\begin{aligned} & g(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n) = 0 && \forall x \in S \\ \iff & g(h(x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n) = 0 && \bar{x} = (x_{k+1}, \ldots, x_n) \in S \end{aligned}$$
and $(a_1, \ldots, a_k) = h(a_{k+1}, \ldots, a_n) = h(\bar{a})$. (i.e. on some n'h'd of a, the points $x \in S$ can be written with respect to $x_1, \ldots, x_k$ as C^1 functions of $x_{k+1}, \ldots, x_n$).

Define $H : U \subseteq \mathbb{R}^{n-k} \to \mathbb{R}^n$ (where $H \in C^1(U)$)

$$\begin{aligned}
&H(x_{k+1}, \ldots, x_n) = (h(x_{k+1}, \ldots, x_n), x_{k+1}, \ldots, x_n) \\
&H(\bar{x}) = (h(\bar{x}), \bar{x}) \quad \forall \bar{x} \in U \\
&H(\bar{a}) = a
\end{aligned}$$

Define $\tilde{f}, \tilde{g} \in C^1(U)$ as

$$\begin{aligned}
\tilde{f} &= f \circ H : U \subseteq \mathbb{R}^{n-k} \to \mathbb{R} \\
\tilde{g} &= g \circ H : U \subseteq \mathbb{R}^{n-k} \to \mathbb{R}^k
\end{aligned}$$

(they are restrictions of f, g to the constraint set S).
By construction, $\tilde{g} = 0$ on U since $g(x) = 0 \; \forall x \in S$. Also, $\tilde{f}$ has a local extrema at $\bar{a} \in U$ (this is now unconstrained).
So we have $(D\tilde{g})_{\bar{a}} = 0$ (because $\tilde{g} = 0$ on U and $\bar{a} \in U$), and $(D\tilde{f})_{\bar{a}} = 0$ (because $\bar{a}$ local extrema of $\tilde{f}$ at $\bar{a}$).
Note that $H(\bar{x}) = (h(\bar{x}), \bar{x})$, so

$$(DH)_a = \begin{bmatrix} (Dh)_{\bar{a}} \\ I_{n-k} \end{bmatrix}$$

where $(Dh)_{\bar{a}}$ is $k \times (n-k)$ and I is $(n-k) \times (n-k)$.
By chain rule, we have

$$\begin{aligned}
(D\tilde{g})_a &= (Dg)_{H(\bar{a})}(Dh)_{\bar{a}} && \text{where } H(\bar{a}) = a \\
&= (Dg)_a \begin{bmatrix} (Dh)_{\bar{a}} \\ I \end{bmatrix} \\
&= \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} (Dh)_{\bar{a}} \\ I \end{bmatrix} \\
&= A(Dh)_{\bar{a}} + B = 0
\end{aligned}$$

where $A \in \mathbb{R}^{k \times k}$, $B \in \mathbb{R}^{k \times (n-k)}$, $(Dh)_{\bar{a}} \in \mathbb{R}^{k \times (n-k)}$ and $I \in \mathbb{R}^{(n-k) \times (n-k)}$.
So we have $(Dh)_{\bar{a}} = -A^{-1}B$.
Let $(Df)_{\bar{a}} = \begin{bmatrix} C & D \end{bmatrix}$ (where $C \in \mathbb{R}^{1 \times k}$ and $D \in \mathbb{R}^{1 \times (n-k)}$). Thus we have

$$\begin{aligned}
(D\tilde{f})_a &= (Df)_{H(\bar{a})}(Dh)_{\bar{a}} && \text{where } H(\bar{a}) = a \\
&= \begin{bmatrix} C & D \end{bmatrix} \begin{bmatrix} (Dh)_{\bar{a}} \\ I \end{bmatrix} \\
&= C(Dh)_{\bar{a}} + D = 0
\end{aligned}$$

Thus $D = -C(Dh)_{\bar{a}} = -CA^{-1}B$ from before.
We want to show that

$$(Df)_a + \lambda(Dg)_a = 0$$

Note that

$$\begin{aligned}
\begin{bmatrix} 0 & 0 \end{bmatrix} &= \begin{bmatrix} C & D \end{bmatrix} - \begin{bmatrix} CA^{-1}A & CA^{-1}B \end{bmatrix} \\
&= \begin{bmatrix} C & D \end{bmatrix} - CA^{-1} \begin{bmatrix} A & B \end{bmatrix} \\
\Rightarrow &(Df)_a + \lambda(Dg)_a = 0
\end{aligned}$$

where $\lambda = -CA^{-1}$. $\square$

27.2 Examples of Lagrange multipliers

Example 27.1. Find all extrema of $f(x, y, z) = x^2 + y^2 + z^2$ (square of distance from origin) subject to the 2 constraints: $x - y = 1$ and $y^2 - z^2 = 1$.

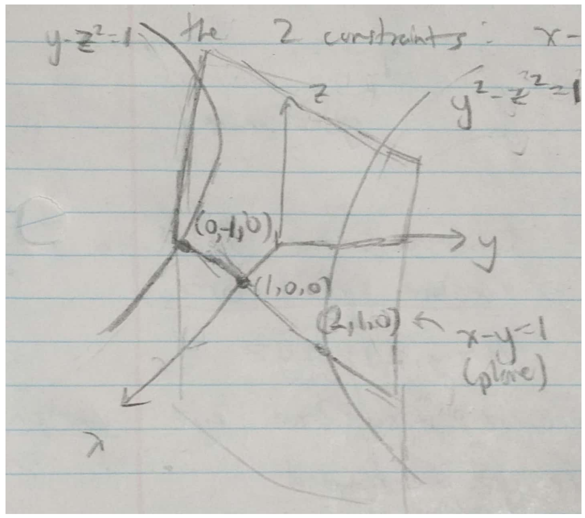

Figure 27.1: Diagram of $\mathbb{R}^3$ with constraints $x - y = 1$ (plane) and $y^2 - z^2 = 1$ (hyperbola). Actual values in the constraint set must be in the intersection of the two graphs.

There exists points on constraint set with arbitrary large distance from origin (no global max).
We know there **will exist** a global min (which will also be a local min). We expect 2 local minima since $y^2 - z^2 = 1$ cuts twice into the other constraint plane.
We have

$$g_1(x, y, z) = x - y - 1 = 0$$
$$g_2(x, y, z) = y^2 - z^2 - 1 = 0$$

and from Lagrange multipliers we know $\nabla f + \lambda \nabla g_1 + \mu \nabla g_2 = 0$, thus

$$(2x, 2y, 2z) + \lambda(1, -1, 0) + \mu(0, 2y, -2z) = 0$$

or

$$2x + \lambda = 0$$
$$2y - \lambda + 2\mu y = 0$$
$$2z - 2\mu z = 0 \Rightarrow z(1 - \mu) = 0$$

From the last constraint, either $\mu = 1$ or $z = 0$:

$\mu = 1$ Then the second equation becomes $4y = \lambda$ and the first equation becomes $2x + 4y = 0$ so $x = -2y$.
From our original constraint equations, we have from g_1 $-3y = 1 \Rightarrow y = \frac{-1}{3}$ and from g_2 $\frac{1}{9} - z^2 = 1 \Rightarrow z^2 = \frac{-8}{9}$ which is a **contradiction** since squares are always positive.

$z = 0$ From g_2 we have $y = \pm 1$ and from g_1 we have $x = y + 1$.

Thus we have two solutions $(2, 1, 0)$ and $(0, -1, 0)$ (which satisfy all the other equations too).

Thus we have $f(2, 1, 0) = 5$ (some local min) and $f(0, -1, 0 = 1$ (global min).

Example 27.2. Extremize volume subject to *constant surface area.* Let $V(x, y, z) = xyz$ and $S(x, y, z) = 2xy + 2xz + 2yz = c$ for some constant $c > 0$.
Note that xy, xz, yz are *bounded.* Also it is easy to see that as $x \to \infty$, then $y \to 0$ and $z \to 0$ (for xy and xz, as $x \to \infty$, y and z must go to 0 since they are constant). Thus $V \to 0$.
So there's no minimum (V can go towards 0 as we move to infinity in any direction). But there will be a global max (think of constricting the space to a ball where everything outside has $f(x, y, z) < k$ for some constant k. Then by EVT local max inside ball is global max).
From Lagrange multipliers, we have

$$\nabla V + \lambda \nabla S = 0$$

Thus we have

$$(yz, xz, xy) + \lambda(2(y + z), 2(x + z), 2(x + y)) = 0$$

or

$$\begin{aligned} yz + 2\lambda(y + z) &= 0 \\ xz + 2\lambda(x + z) &= 0 \\ xy + 2\lambda(x + y) &= 0 \end{aligned}$$

where $\lambda \neq 0$ since $x, y, z > 0$.
Taking x times the first equation subtracting y times the second equation we get $2\lambda xz = 2\lambda yz$ thus $x = y = z$. Therefore the maximum volume occurs for a **cube**.

Example 27.3. Opposite problem as above: extremize surface area subject to constant volume.

Claim. No global max, but there will be a global min.
Note that $V(x, y, z) = xyz = C$ and $S(x, y, z) = 2xy + 2xz + 2yz$. We let $y = z$ (arbitrary) (also $y^2 = \frac{C}{x}$). As $x \to \infty$, then $y = z \to 0$. So we get $S(x, y, z) = 4xy + \frac{2C}{x} = 4\sqrt{C}x^{1/2} + \frac{2C}{x} \to \infty$ as $x \to \infty$.
Then by Lagrange multipliers we have $\nabla S + \lambda \nabla V = 0$. After solving, we see that $x = y = z$ which means a cube has minimal surface area given constant volume.

28 March 14, 2018

28.1 Boxes and size of boxes in $\mathbb{R}^n$

Definition 28.1. Let $I = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n] \subseteq \mathbb{R}^n$. I is the Cartesian product of closed bounded intervals, i.e. $x \in I \iff a_i \leq x \leq b_i$ for all $i = 1, \ldots, n$.
We'll call I a box in $\mathbb{R}^n$ (interval in $\mathbb{R}$, rectangle in $\mathbb{R}^2$, box in $\mathbb{R}^3$).
It is clear I is **compact** because it's closed and bounded.

Definition 28.2. Define the **size of a box** I $\mu(I) \in \mathbb{R}$ (scalar) to be

$$\mu(I) = (b_1 - a_1)(b_2 - a_2) \ldots (b_n - a_n) = \prod_{k=1}^{n} (b_k - a_k)$$

(length in $\mathbb{R}$, area in $\mathbb{R}^2$, volume in $\mathbb{R}^3$).

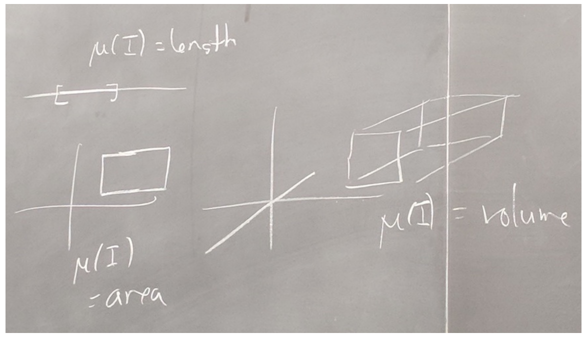

Figure 28.1: Boxes and their size interpretation in $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$.

Remark 28.1. This should be thought of as an n-dimensional "volume". This is also called the **Jordan content**, sometimes "Jordan-measurable" (bad terminology since even if it has a Jordan content it may not be "measurable").

28.2 Zero size

Definition 28.3. Let $E \subseteq \mathbb{R}^n$. We say E has **zero size** (and write $\mu(E) = 0$) iff $\forall \epsilon > 0$, $\exists$ boxes $I_1, \ldots, I_N$ with $E \subseteq \bigcup_{k=1}^{N} I_k$ and $\sum_{k=1}^{N} \mu(I_k) < \epsilon$ (i.e. we can cover E by finitely many boxes whose sizes sum to as small as we want). Note $I_1, \ldots, I_k$ need not be disjoint.
Clearly: If E has size zero then any $U \subseteq E$ has size zero.
The empty set clearly has size zero (any box will vacuously cover it). We will see other sets also have size zero.

Example 28.1. Let $E = \{x\}$ (singleton). Let $I_\delta = [x_1 - \delta, x_1 + \delta] \times \ldots \times [x_n - \delta, x_N + \delta]$. Note that $\{x\} \subseteq I_\delta$ and $\mu(I_\delta) = (2\delta)^n$.
For any $\epsilon > 0$, choose $\delta > 0$ such that $(2\delta)^n < \epsilon$ therefore $E = \{x\}$ has size zero.

Example 28.2. Let $E_1, \ldots, E_m$ be subsets of $\mathbb{R}^n$ with size zero. Then $\bigcup_{k=1}^{m} E_k$ has size zero.

Proof. For each $k = \{1, \ldots, m\}$ $\exists N_k$ boxes $I_{k,1}, I_{k,2}, \ldots, I_{k,N_k}$ such that $E_k \subseteq \bigcup_{j=1}^{N_k} I_{k,j}$ and $\sum_{j=1}^{N_k} \mu(I_{k,j}) < \frac{\epsilon}{m}$. Then $\{I_{k,j} \mid 1 \leq k \leq m, \quad 1 \leq j \leq N_k\}$ covers $E = \bigcup_{j=1}^{m} E_j$ and $\sum_{1 \leq k \leq m} \sum_{j=1}^{N_k} \mu(I_{k,j}) < \epsilon$. ☐

28.3 Continuous graphs of compact sets have zero size

Proposition 28.1. Let $K \subseteq \mathbb{R}^n$ be **compact**. Let $f : U \subseteq \mathbb{R}^n \to \mathbb{R}$ be continuous on U with $K \subseteq U$. Define

$$\Gamma_{f,K} = \{(x, f(x)) \in \mathbb{R}^{n+1} \mid x \in \mathbb{R}^n\}$$

or the "graph of f over the set K". Then $\Gamma_{f,k}$ has size zero.
Intuition: note that there is a bijective correspondence between $\Gamma_{f,K}$ and K (we can project each point in $\Gamma_{f,K}$ back to its point in K). Since K is finite size (compact) then K has finite size. Note $\Gamma_{f,K}$ is in one higher dimension so it's actually a very "thin" plane that corresponds to K in $\mathbb{R}^{n+1}$, so it should have size zero.

Proof. Let $\epsilon > 0$ be arbitrary. Since K is compact, it's bounded so $\exists M > 0$ such that $K \subseteq [-M, M] \times [-M, M] \times \ldots \times [-M, M]$ where there are n intervals. That is $x \in K \Rightarrow |x_k| \leq M$ for all $k = 1, \ldots, n$.
Since f is continuous on K and K is compact we know f is **uniformly continuous** on K so $\exists \delta > 0$ (WLOG $\delta < 1$)) such that if $x, y \in K$ and $\|x - y\| < \delta$, then $|f(x) - f(y)| < \frac{\epsilon}{2(2M+1)^n}$
Let $\tilde{\delta} = \frac{\delta}{\sqrt{n}}$ so $\tilde{\delta} < 1$. Also choose $N \in \mathbb{N}$ such that $N\tilde{\delta} < 2M$ and $(N+1)\tilde{\delta} \geq 2M$ (by Archimedean Principle).
For $k_1, \ldots, k_n \in \{0, 1, \ldots, N\}$ define $I_{k_1,\ldots,k_n} = [-M + k_1\tilde{\delta}, -M + (k_1 + 1)\tilde{\delta}] \times \ldots \times [-M + k_n\tilde{\delta}, -M + (k_n + 1)\tilde{\delta}]$.

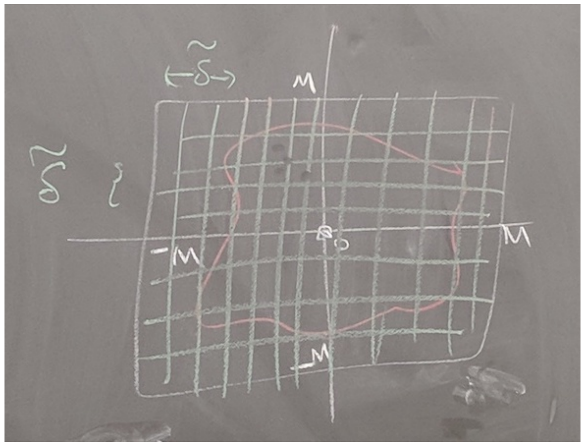

Figure 28.2: Each set of $k_1, \dots, k_n$ corresponds to a $\tilde{\delta}$ grid square in the diagram.

$$x \in I_{k_1,\dots,k_n} \iff -\mu + k_j\tilde{\delta} \le x_j \le -\mu + (k_j+1)\tilde{\delta}$$

Each $I_{k_1,\dots,k_n}$ is a box in $\mathbb{R}^n$, so $\mu(I_{k_1,\dots,k_n}) = (\tilde{\delta})^n$.
Given $k_1, \dots, k_n$, choose $y_{k_1,\dots,k_n} \in I_{k_1,\dots,k_n}$ such that $y_{k_1,\dots,k_n} \in K$. If no such point exists, we won't need that index $k_1, \dots, k_n$ (i.e. in the diagram, we choose a $y_{k_1,\dots,k_n}$ in each box of the grid inside the red K compact set; red dots in diagram).
For those $k_1, \dots, k_n$ such that a $y_{k_1,\dots,k_n}$ exists, define

$$J_{k_1,\dots,k_n} = I_{k_1,\dots,k_n} \times [f(y_{k_1,\dots,k_n}) - \frac{\epsilon}{2(2M+1)^n}, f(y_{k_1,\dots,k_n}) + \frac{\epsilon}{2(2M+1)^n}]$$

in $\mathbb{R}^{n+1}$, so $(x,t) \in J_{k_1,\dots,k_n} \iff x \in I_{k_1,\dots,x_n}$ and $|f(y_{k_1,\dots,k_n)-t}| < \frac{\epsilon}{2(2M+1)^n}$.
$J_{k_1,\dots,k_n}$ is a box in $\mathbb{R}^{n+1}$ and

$$\begin{aligned}\mu(J_{k_1,\dots,k_n}) &= \mu(I_{k_1,\dots,k_n}) \cdot \frac{\epsilon}{(2M+1)^n} \\ &= \frac{(\tilde{\delta})^n \epsilon}{(2M+1)^n} \\ &= \left(\frac{\tilde{\delta}}{2M+1}\right)^n \epsilon\end{aligned}$$

Let $x \in K$ then $x \in I_{k_1,\dots,k_n}$ for at least one $k_1, \dots, k_n$ so $\|x - y_{k_1,\dots,k_n}\|^2 \le n(\tilde{\delta})^2 < \delta^2$.
So $\|x - y_{k_1,\dots,k_n}\| < \delta \Rightarrow |f(x) - f(y_{k_1,\dots,k_n})| < \frac{\epsilon}{2(2M+1)^n}$ so $(x, f(x)) \in J_{k_1,\dots,k_n}$.

Hence $\Gamma_{f,K} \subseteq \bigcup_{k_1,\dots,k_n} J_{k_1,\dots,k_n}$. Note that the size of $\Gamma_{f,K}$ is

$$\begin{aligned}
\sum_{k_1,\dots,k_n} \mu(J_{k_1,\dots,k_n}) &= \sum_{k_1,\dots,k_n} \left(\frac{\tilde{\delta}}{2M+1}\right)\epsilon \\
&\leq (N+1)^n \left(\frac{\tilde{\delta}}{2M+1}\right)\epsilon \qquad k_j = \{0,\dots,N\} \Rightarrow (N+1)^n \text{ combinations of them} \\
&= \left((N+1)\frac{\tilde{\delta}}{2M+1}\right)\epsilon \\
&< \epsilon
\end{aligned}$$

where the last line follows since $N\tilde{\delta} < 2M$ and $\tilde{\delta} < 1$, so $(N+1)\tilde{\delta} < 2M+1$ thus $\frac{(N+1)\tilde{\delta}}{2M+1} < 1$. □

28.4 Boundary of boxes have zero size

Corollary 28.1. (corollary to proposition): Let I be a box in $\mathbb{R}^n$ then ∂I has **zero size**.

Proof. ∂I is a finite union of graphs of continuous functions over compact sets. E.g. for a interval in $\mathbb{R}$ we have two points. For a rectangle in $\mathbb{R}^2$ we have line segments (similarly for a box in $\mathbb{R}^3$).
It is thus the union of $2n$ sets, each of which is a graph of a continuous function over a box (compact) in $\mathbb{R}^{n-1}$ which is zero size, thus ∂I is of zero size. □

29 March 16, 2018

29.1 Non-zero size

We'll need a technical lemma that gives us an equivalent characterization for what it means for a set to have zero size.
Recall: $E \subseteq \mathbb{R}^n$ has zero size if $\forall \epsilon > 0$ if $\exists$ $I_1,\dots,I_N$ boxes (N depends on E,ϵ) such that $E \subseteq \bigcup_{j=1}^N I_j$ and $\sum_{j=1}^N \mu(I_j) < \epsilon$.

Definition 29.1. A $E \subseteq \mathbb{R}^n$ **does not have zero size** iff $\exists \epsilon_0 > 0$ such that $\forall$ finite collections of boxes $I_1,\dots,I_n$ with $E \subseteq \bigcup_{j=1}^N I_j$, we have $\sum_{j=1}^n \mu(I_j) \geq \epsilon_0$ (this is the definition of non-zero size).

We first show a weaker technical lemma that is similar to the definition.

Lemma 29.1 (Technical lemma for zero size). A subset $E \subseteq \mathbb{R}^n$ **does not have size zero** iff $\exists \tilde{\epsilon_0} > 0$ such that $\forall$ finite collections of boxes $I_1,\dots,I_n$ with $E \subseteq \bigcup_{j=1}^N I_j$ where $int(I_j) \cap E \neq \varnothing$, we have $\sum_{j=1}^n \mu(I_j) \geq \tilde{\epsilon_0}$ (we only include the I_j's that matter unlike the definition).

Proof. **Backwards:** Clear: take $\tilde{\epsilon_0} = \epsilon_0$ so $\sum \mu(I_i) \geq \sum \mu(I_j) \geq \epsilon_0$ (since $int(I_j) \cap E \neq \varnothing$ so I_j must be contained within all I_i's, all possible collection of boxes that cover E).
Forwards:
Suppose E does not have size zero. $\exists \epsilon_0 > 0$ such that the statement from the definition holds.
Let $\tilde{\epsilon_0} = \frac{\epsilon_0}{2} > 0$. Suppose $I_1,\dots,I_N$ boxes in $\mathbb{R}^n$ where $E \subseteq \bigcup_{j=1}^N I_j$.
By reordering, WLOG

$$\begin{aligned}
&int(I_j) \cap E \neq 0 && j = 1,\dots,m \\
&int(I_j) \cap E = 0 && j = m+1,\dots,N
\end{aligned}$$

This implies that $I_j \cap E \subseteq \partial I_j$ for $j = m+1, \ldots, N$.
From last time, ∂I_j has size zero so

$$\bigcup_{j=m+1}^{N} I_j \cap E \subseteq \bigcup_{j=m+1}^{N} \partial I_j$$

so this has size zero. So $\exists$ boxes $J_1, \ldots, J_m$ in $\mathbb{R}^n$ such that

$$\bigcup_{j=m+1}^{N} I_j \cap E \subseteq \bigcup_{l=1}^{M} J_l$$

and

$$\sum_{l=1}^{M} \mu(J_l) < \frac{\epsilon_0}{2}$$

Also note that

$$E \subseteq \left(\bigcup_{j=1}^{m} I_j\right) \cup \left(\bigcup_{l=1}^{M} J_l\right)$$

This is a cover of E by finitely many boxes. By our stronger definition of non-zero size we have

$$\sum_{j=1}^{m} \mu(I_j) + \sum_{l=1}^{M} \mu(I_l) \geq \epsilon_0$$

thus

$$\sum_{j=1}^{m} \mu(I_j) \geq \frac{\epsilon_0}{2} = \tilde{\epsilon_0}$$

since $\sum_{l=1}^{M} \mu(I_l) < \frac{\epsilon_0}{2}$. □

Remark 29.1. Not having size zero **does not mean** it has positive size. It may not have any well-defined size!

29.2 Partitions of boxes

Definition 29.2. Let $I = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n]$ be a box in $\mathbb{R}^n$.
For $j \in \{1, \ldots, n\}$, choose $t_{j,0}, t_{j,1}, \ldots, t_{j,N_j}$ $N_j \geq 1$ such that $a_j = t_{j,0} < t_{j,1} < t_{j,2} < \ldots < t_{j,N_j-1} < t_{j,N_j} = b_j$.

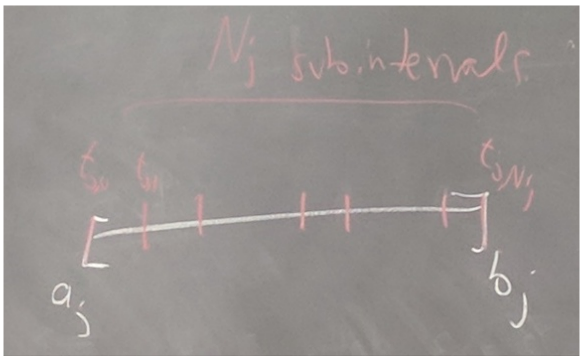

Figure 29.1: Each interval can be partitioned by $t_{j,0}, \ldots, t_{j,N_j}$ into N_j subintervals.

Let $P_j = \{t_{j,l} \mid l = 0, 1, \ldots, N_j\}$ and $P = P_1 \times P_2 \times \ldots \times P_n$.
Then $x \in P \iff x_j \in P_j$ for all $j = 1, \ldots, n$.
Such a P is a called a **partition** of the box.
P has $N_1 \times N_2 \times \ldots \times N_n$ elements.
A partition P of I determines a **subdivision** of I into $N_1 \times \ldots \times N_n$ boxes of the form

$$I_{k_1,\ldots,k_n} = [t_{1,k_1}, t_{1,k_1+1}] \times \ldots \times [t_{n,k_n}, t_{n,k_n+1}]$$

where $k_j \in \{0, 1, \ldots, N_j - 1\}$.

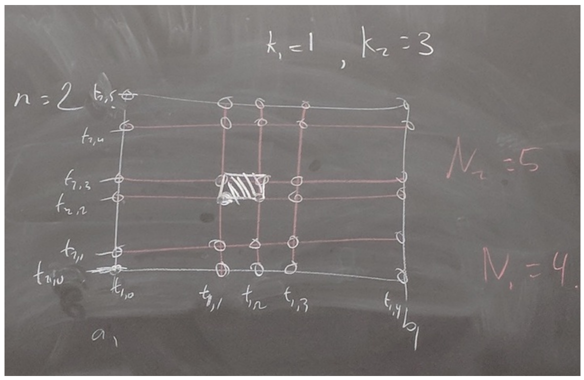

Figure 29.2: A 2D box $I \in \mathbb{R}^2$ is partitioned where $N_1 = 4$ and $N_2 = 5$ (for 20 boxes). The shaded in box is $I_{1,2} = [t_{1,1}, t_{1,2}] \times [t_{2,2}, t_{2,3}]$ where $k_1 = 1$ and $k_2 = 2$ (not 3).

The boxes in the subdivision of I corresponding to the partition P only **only intersect at most along their boundaries**.

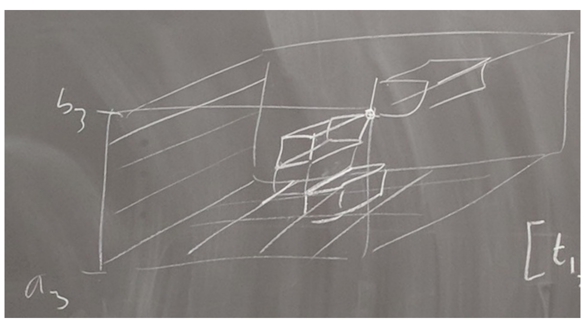

Figure 29.3: A 3D box $I \in \mathbb{R}^3$ where we have subdivisions that only intersect at corners, edges and faces.

29.3 Riemann sum (in terms of partitions and boxes)

Definition 29.3. Let $f : I \to \mathbb{R}^m$, $I \subseteq \mathbb{R}^n$ be a box and P be a partition of I. Let I_α where $\alpha \in P$ be the corresponding subdivision of I. Then

$$I = \bigcup_{\alpha \in P} I_\alpha$$

For each I_α choose $x_\alpha \in I_\alpha$. Then

$$S(f, P) = \sum_{\alpha \in P} f(x_\alpha)\mu(I_\alpha)$$

is called a **Riemann sum** for f with respect to the partition P (note that $\mu(I_\alpha) \in \mathbb{R}$ are scalars hence $S(f, P) \in \mathbb{R}^m$).

Remark 29.2. Given f and P, $\exists$ infinitely many Riemann sums $S(f, P)$ corresponding to different choices of $x_\alpha \in I_\alpha$.

29.4 Refinement of partitions

Definition 29.4. Let P and Q be two partitions of the same box I. We say that Q is a **refinement** of P if $P_j \subseteq Q_j$ for all $j = 1, \ldots, n$ (P_j and Q_j are the partition points $t_{j,l}$: so Q_j contains all partition points of P_j and more!).
That is, all the subboxes J_B of I corresponding to Q are themselves subboxes of one of the subboxes I_α of I corresponding to P.
Note that

$$I = \bigcup_\alpha I_\alpha = \bigcup_\beta J_\beta$$

each I_α is a union of some J_β's.

Remark 29.3. Let P and Q be any two partitions of I. There always exists a partition R of I that is a common refinement (i.e. it is a refinement of both P and Q). Just take $R_j = P_j \cup Q_j$ for all j.

29.5 Riemann integral of box

(not in the textbook, see notes by *V. Runde* linked in syllabus). The other integration theory is the Lebesgue integration theory.

Definition 29.5. Let $I \subseteq \mathbb{R}^n$ be a box. Let $f : I \to \mathbb{R}^m$. f is **integrable** on I **iff** $\exists y \in \mathbb{R}^m$ with the following property: $\forall \epsilon > 0$, there exists a partition P_ϵ of I such that, for **any** refinement P of P_ϵ and **any** Riemann sum $S(f, P)$ corresponding to P, we have

$$\|S(f, P) - y\| < \epsilon$$

(i.e. Riemann sums converge).
If this holds, we say f is **Riemann integrable** on I and $y \in \mathbb{R}^m$ is called the **Riemann integral** of f over I and we write

$$y = \int_I f$$

Remark 29.4. On assignment 10, we prove that such a y is unique if it exists.

29.6 Cauchy criterion for Riemann integrable

Theorem 29.1 (Cauchy criterion)**.** Let $I \subseteq \mathbb{R}^n$ be a box, $f : I \to \mathbb{R}^m$. The following are equivalent (TFAE):

1. f is Riemann integrable on I
2. $\forall \epsilon > 0$, $\exists$ partition P_ϵ of I such that for any refinements P, Q of P_ϵ and any Riemann sums $S(f, P), S(f, Q)$ we have
$$\|S(f, P) - S(f, Q)\| < \epsilon$$

Proof. **(1) $\Rightarrow$ (2):**
Let $y = \int_I f$. Given $\epsilon > 0$ there exists P_ϵ such that $\|S(f, P) - y\| < \frac{\epsilon}{2}$ for all refinements P of P_ϵ.

So

$$\begin{aligned}\|S(f,P)-S(f,Q)\| &\overset{\triangle}{\leq} \|S(f,P)-y\| + \|y-S(f,Q)\| \\ &< \frac{\epsilon}{2}+\frac{\epsilon}{2} \\ &= \epsilon\end{aligned}$$

Conversely, suppose (2) holds.
For all $k \in \mathbb{N}$, there exists P_k of I such that

$$\|S(f,P)-S(f,Q)\| < \frac{1}{2^k}$$

for all refinements P, Q of P_k.
By taking their common refinement WLOG P_{k+1} is a refinement of P_k for all k.
For all $k \in \mathbb{N}$, fix a particular Riemann sum $S(f, P_k) = s_k \in \mathbb{R}^m$.
If $l > k$ then

$$\begin{aligned}\|s_l - s_k\| &\overset{\triangle}{\leq} \sum_{j=k}^{l-1} \|s_{j+1}-s_j\| \\ &< \sum_{j=k}^{l-1} \frac{1}{2^j} && \text{because } P_{j+1} \text{ is a refinement of } P_j\end{aligned}$$

which $\to 0$ as $k \to \infty$.
So s_j is a Cauchy sequence in $\mathbb{R}^m$ so it converges to some $y \in \mathbb{R}^m$. Need to show $y = \int_I f$. Let $\epsilon > 0$, choose K large enough so $\frac{1}{2^K} < \frac{\epsilon}{2}$ and $\|s_K - y\| < \frac{\epsilon}{2}$. □

30 March 19, 2018

30.1 Refined/simplified Cauchy criterion lemma

Lemma 30.1 (Refined/simplified Cauchy criterion)**.** $f : I \to \mathbb{R}^m$ is integrable on I **iff** $\forall \epsilon > 0$, $\exists$ partition P_ϵ of I such that if $S_1(f, P_\epsilon)$, $S_2(f, P_\epsilon)$ are **any two Riemann sums** for f with respect to P_ϵ, then

$$\|S_1(f,P_\epsilon) - S_2(f,P_\epsilon)\| < \epsilon$$

(this is equivalent to the previous two statements mentioned in section 29.6). i.e. we just need to know that the Riemann sums corresponding to a **single partition** P_ϵ are within ϵ.

Remark 30.1. We'll use this lemma many times to prove results.

Proof. **Forwards:** Since f is integrable, we use (2) of the Cauchy criterion and take $P = Q = P_\epsilon$.
Backwards: Suppose second part of lemma holds. We need to show (2) of the Cauchy criterion holds.
For P and Q refinements of P_ϵ, P has subcover $I = \bigcup_\alpha I_\alpha$ and Q has subcover $I = \bigcup_\beta J_\beta$.
Observe $f(x) = (f_1(x), \ldots, f_m(x))$. So we have

$$S(f,P) = \sum_\alpha f(x_\alpha)\mu(I_\alpha) = v = (v_1, \ldots, v_m)$$

where $v_j = \sum_\alpha f_j(x_\alpha)\mu(I_\alpha) = S(f_j, P)$. Similarly $S(f,Q) = \sum_\beta f(y_\beta)\mu(I_\beta) = w = (W_1, \ldots, w_m)$ where $w_j = S(f_j, Q)$.
Thus we have

$$\begin{aligned}\|S(f,P) - S(f,Q)\|^2 &= \|v-w\|^2 \\ &= \sum_{j=1}^m |v_j - w_j|^2 \\ &= \sum_{j=1}^m |S(f_j,P) - S(f_j,Q)|^2\end{aligned}$$

If we can show each of $|S(f_j,P) - S(f_j,Q)|^2$ are $< \frac{\epsilon}{m}$, then we are done.
This means WLOG we can assume $m = 1$ (if our lemma implies (2) of Cauchy criterion holds for $m = 1$, let $\epsilon > 0$. Apply our lemma for $\tilde{\epsilon} = \frac{\epsilon}{m} > 0$. Our lemma implies (2) for $m = 1$ with $\tilde{\epsilon}$ yields our lemma implies (2) for a general m from the summation).
Assume $m = 1$. Let P_ϵ be given by our lemma where $I = \bigcup_\alpha I_\alpha$.
Let P, Q be refinements of P_ϵ. P has decomposition $I = \bigcup_\beta J_\beta$ and Q has decomposition $I = \bigcup_\gamma K_\gamma$. Each I_α is a union of finitely many J_β's and also a union of finitely many K_γ's. Thus

$$I_\alpha = \bigcup_{\beta, J_\beta \subseteq I_\alpha} J_\beta = \bigcup_{\gamma, K_\gamma \subseteq I_\alpha} K_\gamma$$

So we have

$$\begin{aligned}S(f,P) - S(f,Q) &= \sum_\beta f(x_\beta)\mu(J_\beta) - \sum_\gamma f(y_\gamma)\mu(K_\gamma) \\ &= \sum_\alpha \Big(\sum_{\beta, J_\beta \subseteq I_\alpha} f(x_\beta)\mu(J_\beta)) - \sum_\alpha \Big(\sum_{\gamma, K_\gamma \subseteq I_\alpha} f(y_\gamma)\mu(K_\gamma))\end{aligned}$$

for some $x_\beta \in J_\beta$ and $y_\gamma \in K_\gamma$ (just one of each since we are in $m = 1$).
For each α, choose $z_\alpha, w_\alpha \in I_\alpha$ such that

$$\begin{aligned}f(z_\alpha) &= \max\{f(x_\beta), f(y_\gamma), \forall \beta, \gamma \text{ such that } J_\beta \subseteq I_\alpha \text{ and } K_\gamma \subseteq I_\alpha\} \\ f(w_\alpha) &= \min\{f(x_\beta), f(y_\gamma), \forall \beta, \gamma \text{ such that } J_\beta \subseteq I_\alpha \text{ and } K_\gamma \subseteq I_\alpha\}\end{aligned}$$

(we have finitely many β, γ as above). By construction $f(w_\alpha) \le f(x_\beta) \le f(z_\alpha)$ for all β such that $J_\beta \subseteq I_\alpha$. Also $-f(z_\alpha) \le -f(y_\gamma) \le -f(w_\alpha)$ for all γ such that $K_\alpha \subseteq I_\alpha$. Then

$$\begin{aligned}[f(w_\alpha) - f(z_\alpha)]\mu(I_\alpha) &= f(w_\alpha)\left(\sum_{\beta, J_\beta \subseteq I_\alpha} \mu(J_\beta)\right) - f(z_\alpha)\left(\sum_{\gamma, K_\gamma \subseteq I_\alpha} \mu(K_\gamma)\right) \\ &\le \sum_{\beta, J_\beta \subseteq I_\alpha} f(x_\beta)\mu(J_\beta)) - \sum_{\gamma, K_\gamma \subseteq I_\alpha} f(y_\gamma)\mu(K_\gamma) \\ &\le f(z_\alpha)\left(\sum_{\beta, J_\beta \subseteq I_\alpha} \mu(J_\beta)\right) - f(w_\alpha)\left(\sum_{\gamma, K_\gamma \subseteq I_\alpha} \mu(K_\gamma)\right) \\ &= [f(z_\alpha) - f(w_\alpha)]\mu(I_\alpha)\end{aligned}$$

Thus we have established an inequality that sandwiches our $S(f,P) - S(f,Q)$ for a given α (second line). Summing over all α we have

$$-\left(\sum_\alpha (f(z_\alpha)\mu(I_\alpha)) - \sum_\alpha (f(w_\alpha)\mu(I_\alpha))\right) \leq S(f,P) - S(f,Q) \leq \sum_\alpha f(z_\alpha)\mu(I_\alpha) - \sum_\alpha f(w_\alpha)\mu(I_\alpha)$$

Thus going back to our initial equation

$$\begin{aligned}
|S(f,P) - S(f,Q)| &\leq \sum_\alpha f(z_\alpha)\mu(I_\alpha) - \sum_\alpha f(w_\alpha)\mu(I_\alpha) \\
&= |\sum_\alpha f(z_\alpha)\mu(I_\alpha) - \sum_\alpha f(w_\alpha)\mu(I_\alpha)| \\
&= |S_1(f,P_\epsilon) - S_2(f,P_\epsilon)| \\
&< \epsilon
\end{aligned}$$

where the last line follows from our initial assumption of the second part of the lemma. □

30.2 Bounded and continuous "almost everywhere" ⇒ integrable

Theorem 30.1 (Integrable with size zero discontinuity)**.** Let $I \subseteq \mathbb{R}^n$ be a box. Let $f : I \to \mathbb{R}^m$ be **bounded**. Let $S \subseteq I$ be the points in I where f **fails to be continuous**. If S has **size zero**, then f is Riemann integrable on I. (informally, a function which is bounded and continuous "almost everywhere" (except S size zero) is integrable. Note "almost everywhere" means something similar but different in measure theory).

Remark 30.2. If $S = \varnothing$ then boundedness is automatic for f by EVT (since I is compact and there are no gaps).

Remark 30.3. If f is integrable on I, then f must be bounded (assignment 11 question 2; hint: integrable f means all components are integrable).

Remark 30.4. We usually define integrability on interior points, but note that ∂I has size zero so we can always include it into our set S, thus we can say all of I is integrable.

Proof. Since f is bounded, $\exists C > 0$ such that

$$\|f(x)\| < C \quad \forall x \in I$$

Claim. Because S has size zero, $\exists$ a partition P of I such that

$$\sum_{\alpha, I_\alpha \cap S \neq \varnothing} \mu(I_\alpha) < \frac{\epsilon}{4C}$$

Proof of claim: Start with a cover of S by boxes with sum of sizes $< \frac{\epsilon}{4C}$. Note that S is a subset of I. Take any partition P of I. Refine P using the boxes from our cover for S (we take the **endpoints** of every box in the cover of S and use that in our refinement). The subdivisions in the partition that actually overlap S has equal or smaller size.

Let $K = \bigcup_{\alpha, I_\alpha \cap S \neq \varnothing} I_\alpha$. This is **compact**.
Since K is compact and $f\mid_K$ is continuous, then $f\mid_K$ is **uniformly continuous** on K.
$\exists \delta > 0$ such that $x, y \in K$ such that

$$\|x - y\| < \delta \Rightarrow \|f(x) - f(y)\| < \frac{\epsilon}{2\mu(I)}$$

Choose a refinement Q of P such that for each subbox J_β of Q where $J_\beta = [a_1^{(\beta)}, b_1^{(\beta)}] \times \ldots \times [a_n^{(\beta)}, b_n^{(\beta)}]$ with $\max_{j=1,\ldots,n}|b_j^{(\beta)} - a_j^{(\beta)}| < \frac{\delta}{\sqrt{n}}$.
Let $S_1(f,Q)$, $S_2(f,Q)$ be any two Riemann sums for f wrt Q. Then

$$\begin{aligned}
S_1(f,Q) - S_2(f,Q) &= \sum_\beta f(x_\beta)\mu(J_\beta) - \sum_\beta f(y_\beta)\mu(J_\beta) \\
\|S_1(f,Q) - S_2(f,Q)\| &\overset{\triangle}{\leq} \sum_\beta \|f(x_\beta) - f(y_\beta)\|\mu(J_\beta) \\
&= \sum_{\beta, J_\beta \not\subseteq K} \|f(x_\beta) - f(y_\beta)\|\mu(J_\beta) + \sum_{\beta, J_\beta \subseteq K} \|f(x_\beta) - f(y_\beta)\|\mu(J_\beta) \\
&\|f(x_\beta) - f(y_\beta)\| < 2C \text{ since } f \text{ is bounded so we have} \\
&< \sum_{\beta, J_\beta \not\subseteq K} (2C)\mu(J_\beta) + \sum_{\beta, J_\beta \subseteq K} \frac{\epsilon}{2\mu(I)}\mu(J_\beta) \\
&< 2C(\frac{\epsilon}{4C}) + \frac{\epsilon}{2\mu(I)}\mu(I) \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}$$

Thus f is integrable by Refined/simplified Cauchy criterion. □

31 March 21, 2018

31.1 Riemann integral on general set $D \subseteq I$

Definition 31.1. Let $D \subseteq \mathbb{R}^n$ be **bounded**. Let $f : D \to \mathbb{R}^m$. Choose any box I in $\mathbb{R}^n$ such that $D \subseteq I$ (possible since D is bounded).
Define $\tilde{f} : I \to \mathbb{R}^m$ by $\tilde{f}(x) = f(x)$ if $x \in D$ and $\tilde{f}(x) = 0$ if $x \notin D$ (i.e. $\tilde{f}$ is the extension by zero of f from D to I).
We say f is **Riemann integrable** on D iff $\tilde{f}$ is Riemann integrable on I and we write

$$\int_D f = \int_I \tilde{f}$$

Claim. This is well-defined (independent of choice of I).

Proof. Exercise (key point: $\tilde{f}$ is identically zero outside of D). □

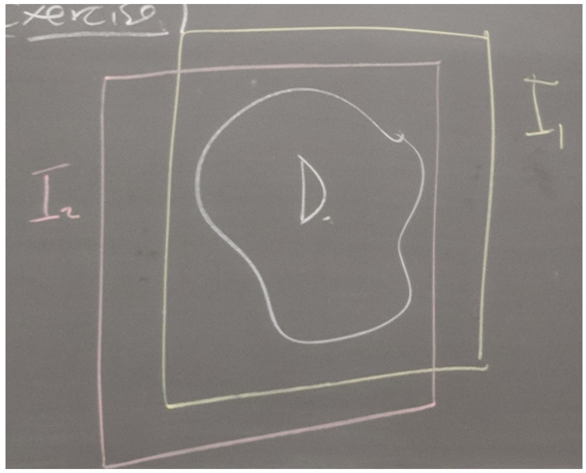

Figure 31.1: We can show that any box I_1 or I_2 works for the above definition.

□

31.2 Bounded arbitrary set with size zero boundary implies integrability

Theorem 31.1. Let $D \neq \varnothing$ and bounded subset of $\mathbb{R}^n$ with $\mu(\partial D) = 0$ (i.e. the boundary ∂D has size zero). Suppose $f : D \to \mathbb{R}^m$ is **bounded** and **continuous**, then f is (Riemann) integrable on D.

Remark 31.1. Boundedness of f is automatic if D is closed since f is continuous.

Proof. Let I be a box such that $I \supseteq D$. Let $\tilde{f}$ be extension by zero of f from D to I.
$\tilde{f}$ is continuous at each point in $int(D)$ by assumption. $\tilde{f}$ is continuous at each point in $int(I \setminus D)$ because it's identically zero.
So $S = \{x \in I \text{ such that } \tilde{f} \text{ is not continuous}\} \subseteq \partial D \cup \partial I$ has size zero.
So by previous theorem $\tilde{f}$ is integrable on I, so f is integrable on D. □

31.3 Indicator function

Definition 31.2. Let $D \subseteq \mathbb{R}^n$ be **any subset**. We define the **indicator function** $X_D : \mathbb{R}^n \to \mathbb{R}$ of D to be

$$X_D(x) \begin{cases} 1 & \text{, if } x \in D \\ 0 & \text{, if } x \notin D \end{cases}$$

31.4 Size of general sets

Now we can define the size of more general sets.

Definition 31.3. Let $D \subseteq \mathbb{R}^n$ be **bounded**. We say that D is **sizeable** *iff* X_D is integrable on D (equivalently if $f : D \to \mathbb{R}$, $f(x) = 1\ \forall x \in D$ is integrable on D).
If D is sizeable, then the **size** of D is

$$\mu(D) = \int_D 1 = \int_D X_D = \int_I X_D \mid_I$$

for any $I \supseteq D$.

Remark 31.2. In books where μ (size) is called "Jordan content", then to say a set D is sizeable means D has a well-define Jordan content aka "contentable".

Remark 31.3. A set D may or may not be sizeable. If it is *sizeable*, the size *may be zero or positive* (i.e. integral is non-negative for non-negative functions; we'll see this later).

31.5 Characterization of sizeability

Theorem 31.2. Let $D \neq \varnothing$ and bounded. Then D is **sizeable** iff ∂D **has size zero**.

Proof. **Backwards:**
Suppose $\mu(\partial D) = 0$. Then X_D is continuous except on a set of size zero, so X_D is integrable on D, so D is sizeable.
Forwards:
Suppose D is sizeable. We must show ∂D has size zero. By contradiction, suppose ∂D does not have size zero.
So by Technical lemma for zero size from last week, $\exists \epsilon_0 > 0$ such that if $I_1, \ldots, I_N$ are boxes in $\mathbb{R}^n$ with $\bigcup_{j=1}^{I_N} \supseteq \partial D$ then

$$\sum_{j=1, int(I_j)\cap \partial D \neq \varnothing}^{N} \mu(I_j) \geq \epsilon_0$$

By hypothesis, D is sizeable so X_D is integrable on D. Let $I \supseteq D$ (box). $\exists$ a partition P of I such that

$$|S(X_D, P) - \int_I X_D| < \frac{\epsilon_0}{2}$$

(where $\int_I X_D = \mu(D)$).
Let's consider two different Riemann sums for X_D with respect to the partition P (i.e. choose $x_\alpha \in I_\alpha$ for all α for Riemann sum 1; $y_\alpha \in I_\alpha$ for all α for Riemann sum 2).
For Riemann sum 1, if $int(I_\alpha) \cap \partial D \neq \varnothing$, choose any $x_\alpha \in D$ (by property of ∂D, where $\partial D \cap D \neq \varnothing$). If $int(I_\alpha) \cap \partial D = \varnothing$, choose $x_\alpha \in I_\alpha$ arbitrary.
For Riemann sum 2, If $int(I_\alpha) \cap \partial D \neq \varnothing$, choose $y_\alpha \in D^c$ (by property of ∂D). If $int(I_\alpha) \cap \partial D = \varnothing$, let $y_\alpha = x_\alpha$.
So we have

$$\begin{aligned} S_1(X_D, P) &= \sum_\alpha X_D(x_\alpha)\mu(I_\alpha) \\ S_2(X_D, P) &= \sum_\alpha X_D(y_\alpha)\mu(I_\alpha) \end{aligned}$$

Thus

$$S_1(X_D, P) - S_2(X_D, P) = \sum_{\alpha, int(I_\alpha)\cap D \neq \varnothing} \mu(I_\alpha) \geq \epsilon_0$$

from before (note that for $int(I_\alpha \cap D = \varnothing)$, $y_\alpha = x_\alpha$ so they cancel out; for the above case, $X_D(y_\alpha) = 0$).
But

$$\begin{aligned} |S_1(X_D, P) - S_2(X_D, P)| &\leq |S_1(X_D, P) - \int_I X_D| + |\int_I X_D - S_2(X_D, P)| \\ &< \frac{\epsilon_0}{2} + \frac{\epsilon_0}{2} \\ &= \epsilon_0 \end{aligned}$$

Contradiction: so ∂D has size zero. □

31.6 Properties of Riemann integrals

Let $D \subseteq \mathbb{R}^n$ be bounded.

1. Let $f, g : D \to \mathbb{R}^m$ be integrable on D. Let $\lambda, \mu \in \mathbb{R}$. Then $\lambda f + \mu g$ is integrable on D and

$$\int_D (\lambda f + \mu g) = \lambda \int_D f + \mu \int_D g$$

Proof. Assignment 10 question 3. □

Let V be the space of **all** functions from D to $\mathbb{R}^m$ (inf. dimensional real vector space).

Let W be the set of all **integrable functions** from D to $\mathbb{R}^m$.

The property says W is a *subspace* of V and $\int : W \to \mathbb{R}^m$ i.e. $f \mapsto \int f$ is **linear**.

2. Let $f : D \to \mathbb{R}$ ($m = 1!$) be integrable on D and **non-negative**. Then $\int_D f \geq 0$.

Proof. Suppose $\int_D f < 0$. Let $\epsilon = -\int_D f > 0$. Choose a partition P of $I \supseteq D$ such that

$$\begin{aligned} &|S(f,P) - \int_D f| < \frac{\epsilon}{2} \\ \Rightarrow &|S(f,P) + \epsilon| < \frac{\epsilon}{2} \\ \Rightarrow &S(f,P) + \epsilon < \frac{\epsilon}{2} \\ \Rightarrow &S(f,P) < \frac{-\epsilon}{2} < 0 \end{aligned}$$

But f is non-negative so $S(f,P) \geq 0$ for any Riemann sum, a contradiction. □

Corollary 31.1. If D is sizeable, $\mu(D) \geq 0$.

3. Let D be sizeable and $f : D \to \mathbb{R}$ ($m = 1!$) be integrable on D and $\exists M_1, M_2 \in \mathbb{R}$ such that $M_1 \leq f(x) \leq M_2$ $\forall x \in D$. Then $M_1 \mu(D) \leq \int_D f \leq M_2 \mu(D)$.

Proof. Let $h : D \to \mathbb{R}$ where $h(x) = M_2 - f(x) \geq 0$ on D.

Also $M_2 = M_2 \cdot X_D$ is integrable on D by property (1).

So
$$0 \leq \int_D h = \int_D M_2 X_D - \int_D f = M_2 \mu(D) - \int_D f$$
so h is integrable by property (1) (the case for M_1 is similar). □

4. Let $D_1, D_2 \subseteq \mathbb{R}^n$ be bounded with $\mu(D_1 \cap D_2) = 0$.

Let $f : D_1 \cup D_2 \to \mathbb{R}^m$. Suppose f is integrable on D_1 and D_2. Then f is integrable on $D_1 \cup D_2$ and

$$\int_{D_1 \cup D_2} f = \int_{D_1} f + \int_{D_2} f$$

Proof.

Claim. If f is integrable on some subset E, then f is **bounded** on E (exercise of assignment 11).

Claim. If $\mu(E) = 0$, then f is integrable on E and $\int_E f = 0$.

Proof of claim: When $m = 1$ (real-valued), use property (3) i.e. $M_1 \leq f(x) \leq M_2$ on D, then $M_1\mu(D) \leq \int_D f \leq M_2\mu(D)$.

Choose a box $I \supset D_1 \cup D_2$. Thus

$$\int_{D_1} f = \int_I fX_{D_1} = \int_{D_1 \cup D_2} fX_{D_1}$$

Where the first equality holds since f is integrable on D_1, so fX_{D_1} is integrable on $D_1 \cup D_2$.

Similarly fX_{D_2} is integrable on $D_1 \cup D_2$.

Also, $\mu(D_1 \cap D_2) = 0$, so by the first part of claim #2, $fX_{D_1 \cap D_2}$ is integrable on $D_1 \cup D_2$ (continuous except possible on a set of size zero).

Note that

$$fX_{D_1 \cup D_2} = fX_{D_1} + fX_{D_2} - f_{D_1 \cap D_2}$$

Thus

$$\begin{aligned} \int_I fX_{D_1 \cup D_2} &\overset{\text{property (1)}}{=} \int_I fX_{D_1} + \int_I fX_{D_2} - \int_I fX_{D_1 \cap D_2} \\ \Rightarrow \int_{D_1 \cup D_2} f &= \int_{D_1} f + \int_{D_2} f - 0 \end{aligned}$$

by second part of claim 2. ☐

5. Let $D \subseteq \mathbb{R}^n$ be bounded. If $f : D \to \mathbb{R}^m$ is integrable on D, then $\|f\| : D \to \mathbb{R}$ where $\|f\|(x) = \|f(x)\|$ is integrable on D and

$$\|\int_D f\| \leq \int_D \|f\|$$

Proof. Let I be a box containing D. Let $\epsilon > 0$. We'll use the refined Cauchy criterion for integrability.

Since f is integrable on D, each component function $f_j : D \to \mathbb{R}$ is integrable on D. So $\exists$ a partition P_ϵ of I such that

$$|S_1(f_j, P_\epsilon) - S_2(f_j, P_\epsilon)| < \frac{\epsilon}{m}$$

for all $j = 1, \ldots, m$ (by taking a common refinment of the m partitions corresponding to the m component functions).

Let $\{I_\alpha\}$ be the subboxes of I determined by P_ϵ. Fix $j \in \{1, \ldots, m\}$.

Let $x_\alpha, y_\alpha \in I_\alpha$ arbitrary.

Choose $z_\alpha, w_\alpha \in \{x_\alpha, y_\alpha\}$ where

$$\begin{aligned} f_j(z_\alpha) &= \max\{f_j(x_\alpha), f_j(y_\alpha)\} \\ f_j(w_\alpha) &= \min\{f_j(x_\alpha), f_j(y_\alpha)\} \end{aligned}$$

(i.e. z_α, y_α correspond to whichever that maps to a larger and smaller value on f_j, respectively).

Thus we can rewrite

$$\begin{aligned}
\sum_\alpha |f_j(x_\alpha) - f_j(y_\alpha)|\mu(I_\alpha) &= \sum_\alpha (f_j(z_\alpha) - f_j(w_\alpha))\mu(I_\alpha) \\
&= \sum_\alpha f_j(z_\alpha)\mu(I_\alpha) - \sum_\alpha f_j(w_\alpha)\mu(I_\alpha) \\
&< \frac{\epsilon}{m}
\end{aligned}$$

for any j (from our premise).

So

$$\begin{aligned}
|S_1(\|f\|, P_\epsilon) - S_2(\|f\|, P_\epsilon)| &= |\sum_\alpha \|f_j(x_\alpha)\|\mu(I_\alpha) - \sum_\alpha \|f_j(y_\alpha)\|\mu(I_\alpha)| \\
&\overset{\triangle}{\leq} \sum_\alpha \|f(x_\alpha) - f(y_\alpha)\|\mu(I_\alpha) && |\|a\| - \|b\|| \leq \|a - b\| \\
&\leq \frac{\epsilon}{m} m && \|v\| \leq \sum_{j=1}^{n} |v_j| \\
&= \epsilon
\end{aligned}$$

So $|S_1(\|f\|, P_\epsilon) - S_2(\|f\|, P_\epsilon)| < \epsilon$ For any two Riemann sums of $\|f\|$ wrt P_ϵ, so by the refined Cauchy criterion $\|f\|$ is integrable on D.

Since f, $\|f\|$ both integrable on D, for all $\epsilon > 0$, $\exists$ partition Q_ϵ of $I \supseteq D$ such that $x_\alpha \in I_\alpha$ for all α satisfies both

$$\begin{aligned}
\|\sum_\alpha f(x_\alpha)\mu(I_\alpha) - \int_I f\| &< \frac{\epsilon}{2} \\
\|\sum_\alpha \|f(x_\alpha)\|\mu(I_\alpha) - \int_I \|f\|\| &< \frac{\epsilon}{2}
\end{aligned}$$

So we have

$$\begin{aligned}
\|\int_I f\| &\overset{\triangle}{\leq} \|\int_I f - \sum_\alpha f(x_\alpha)\mu(I_\alpha)\| + \|\sum_\alpha f(x_\alpha)\mu(I_\alpha)\| \\
&< \frac{\epsilon}{2} + \|\sum_\alpha f(x_\alpha)\mu(I_\alpha)\| \\
&\overset{\triangle}{\leq} \frac{\epsilon}{2} + \sum_\alpha \|f(x_\alpha)\|\mu(I_\alpha) \\
&< \frac{\epsilon}{2} + (\int_I \|f\| + \frac{\epsilon}{2}) \\
&= \int_I \|f\| + \epsilon
\end{aligned}$$

Therefore $\|\int_D f\| \leq \int_D \|f\| + \epsilon$ for any $\epsilon > 0$, so we have $\|\int_D f\| \leq \int_D \|f\|$. $\square$

32 March 23, 2018

32.1 Mean value theorem for integration

Theorem 32.1 (Mean value theorem for integration)**.** Let $D \subseteq \mathbb{R}^n$ be **compact, connected, and sizeable**. Let $f : D \to \mathbb{R}$ be **continuous** on D. Then $\exists$ at least one point $x_0 \in D$ such that

$$\int_D f = f(x_0)\mu(D) \tag{32.1}$$

Proof. If $\mu(D) = 0$ then both sides of eq. (32.1) are zero for any $x_0 \in D$ ($\int_D f = 0$ by claim #2 in property (4)). So assume $\mu(D) > 0$. Let

$$M_1 = \min\{f(x); x \in D\}$$
$$M_2 = \max\{f(x); x \in D\}$$

(exists because D compact and f continuous so $f(D)$ compact and hence bounded).
In fact, by the extreme value theorem, these extrema are *attained*, that is $\exists x_1, x_2 \in D$ such that $M_1 = f(x_1)$ and $M_2 = f(x_2)$ (not unique).
By property (3), $M_1\mu(D) \leq \int_D f \leq M_2\mu(D)$, thus

$$M_1 = f(x_1) \leq \frac{1}{\mu(D)} \int_D f \leq f(x_2) = M_2$$

Since D is connected $\exists$ (not unique) x_0 by the intermediate value theorem such that

$$f(x_0) = \frac{1}{\mu(D)} \int_D f$$

Remark 32.1. Why is it called the mean valued theorem?
Suppose $f(x_1), \ldots, f(x_M) \in \mathbb{R}$, $x_1, \ldots, x_M \in D$ then

$$Mean(y_i) = \frac{\sum_{k=1}^M f(x_k))}{\sum_{k=1}^M 1} \to \frac{\int_D f}{\int_D 1}$$

(this is of course not rigorous but rather a heuristic for intuition).

□

33 March 26, 2018

33.1 Fubini's theorem for evaluating integrals

How do we actually **evaluate** a Riemann integral? We use **Fubini's theorem**, which allows us to decompose an integral over a region in $\mathbb{R}^n$ into n iterated single-variable integrals over regions in $\mathbb{R}^1$, then applying the fundamental theorem of calculus.

Theorem 33.1 (Fubini's theorem)**.** Let I be a box in $\mathbb{R}^n$. Let J be a box in $\mathbb{R}^m$. Then $I \times J$ is a box in $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$.

Let $f : I \times J \to \mathbb{R}^p$. Suppose that f is integrable on $I \times J$ **and** for each $x \in I$ the function where $y \mapsto f(x, y)$ i.e. $f(x, \cdot) : J \to \mathbb{R}^p$ is integrable on J, that is

$$\int_J f(x, \cdot) = F(x) \in \mathbb{R}^m$$

exists for all $x \in I$ where $F : I \to \mathbb{R}^m$.
Then F is integrable on I and

$$\int_I F = \int_{I \times J} f$$

or $\int_I \left(\int_J f(x, \cdot)\right) = \int_{I \times J} f$.

Remark 33.1. This means if the hypotheses are satisfied, to integrate over $I \times J$, we can first integrate over J, then over I.
Of course if f is integrable on $I \times J$ and $f(\cdot, y)$ is integrable on I for all $y \in J$, let $G(y) = \int_I f(\cdot, y)$. Then G is integrable on J, and $\int_J G = \int_{I \times J} f$.

Proof. Let $\epsilon > 0$. Since $f : I \times J \to \mathbb{R}^p$ is integrable on $I \times J$, there exists a partition P_ϵ of I such that

$$\left\| S(f, P) - \int_{I \times J} f \right\| < \frac{\epsilon}{2} \tag{33.1}$$

for any Riemann sum of f wrt any refinement P of P_ϵ.
Let Q_ϵ and R_ϵ be the partitions of I and J respectively such that $P_\epsilon = Q_\epsilon \times R_\epsilon$ (i.e. we have $(P_\epsilon)_1 \times \ldots \times (P_\epsilon)_{n+m}$ where $(Q_\epsilon)_k = (P_\epsilon)_k$ for $k = 1, \ldots, n$ and $(R_\epsilon)_k = (P_\epsilon)_{n+k}$ for $k = 1, \ldots, m$).
Let Q be any refinement of Q_ϵ (partition of I) with corresponding subdivision I_α. Choose $x_\alpha \in I_\alpha$. For each α, the function $y \mapsto f(x_\alpha, y)$ is integrable on J. So $\exists$ a partition T_α of J such that for each refinement T of T_α with corresponding subdivision J_β we have

$$\left\| \sum_\beta f(x_\alpha, y_\beta)\mu(J_\beta) - F(x_\alpha) \right\| < \frac{\epsilon}{2\mu(I)} \tag{33.2}$$

where $y_\beta \in J_\beta$ the left term is $S(f(x_\alpha, \cdot), T)$ and $F(x_\alpha) = \int_J f(x_\alpha, \cdot)$.
Let T_ϵ (partition of J) be a common refinement of R_ϵ and T_α for all α with corresponding subdivision J_β of J.
Then $Q_\epsilon \times T_\epsilon$ is a refinment of P_ϵ with corresponding subdivision $I_\alpha \times J_\beta$ of $I \times J$.
Choose $y_\beta \in J_\beta$ arbitrary, thus from (33.1) we have

$$\left\| \sum_{\alpha,\beta} f(x_\alpha, y_\beta)\mu(I_\alpha)\mu(J_\beta) - \int_{I \times J} f \right\| < \frac{\epsilon}{2} \tag{33.3}$$

where $\mu(I_\alpha)\mu(J_\beta) = \mu(I_\alpha \times J_\beta)$.

Thus we have

$$
\begin{aligned}
\left\| \sum_{\alpha} F(x_\alpha)\mu(I_\alpha) - \int_{I\times J} f \right\| &\leq \left\| \sum_{\alpha} F(x_\alpha)\mu(I_\alpha) - \sum_{\alpha,\beta} f(x_\alpha, y_\beta)\mu(I_\alpha)\mu(J_\beta) \right\| + \left\| \sum_{\alpha,\beta} f(x_\alpha, y_\beta)\mu(I_\alpha)\mu(J_\beta) - \int_{I\times J} f \right\| \\
&\overset{(33.3) \text{ and } \triangle}{\leq} \sum_{\alpha} \left\| F(x_\alpha) - \sum_{\beta} f(x_\alpha, y_\beta)\mu(J_\beta) \right\| \mu(I_\alpha) + \frac{\epsilon}{2} \\
&\overset{(33.2)}{\leq} \sum_{\alpha} \left(\frac{\epsilon}{2\mu(I)} \right) \mu(I_\alpha) + \frac{\epsilon}{2} \\
&= \frac{\epsilon}{2\mu(I)} \left(\sum_{\alpha} \mu(I_\alpha) \right) + \frac{\epsilon}{2} \\
&= \frac{\epsilon}{2} + \frac{\epsilon}{2} \\
&= \epsilon
\end{aligned}
$$

We've shown $F : I \to \mathbb{R}^p$ is integrable on I and $\int_I F = \int_{I\times J} f$. We found a partition Q_ϵ of I such that for any Riemann sum of F converges for any refinement Q of Q_ϵ i.e. $\|S(F,Q) - \int_{I\times J} f\| < \epsilon$. $\square$

33.2 1-D Fubini's corollary

Corollary 33.1. ($p = 1$, $m = n = 1$). If $\int_a^b f(x,y)\,\mathrm{d}x$ exists $\forall y \in [c,d]$ and $\int_{[a,b]\times[c,d]} f(x,y)$ exists, then $\int_{[a,b]\times[c,d]} f(x,y) = \int_c^d \left(\int_a^b f(x,y)\,\mathrm{d}x \right) \mathrm{d}y$.

Example 33.1. Let $D = [0,1]\times[0,1] \subseteq \mathbb{R}^2$. Let $f(x,y) = y^3 e^{xy^2}$.
f is continuous on D. Also $y \mapsto f(x,y)$ continuous on $[0,1]$ for all $x \in [0,1]$ and $x \mapsto f(x,y)$ continuous on $[0,1]$ for all $y \in [0,1]$.
By our corollary, we have

$$
\begin{aligned}
\int_D f &= \int_0^1 \left(\int_0^1 y^3 e^{xy^2}\,\mathrm{d}y \right) \mathrm{d}x \\
&= \int_0^1 \left(\int_0^1 y^3 e^{xy^2}\,\mathrm{d}x \right) \mathrm{d}y \\
&= \int_0^1 \left(\frac{y^3 e^{xy^2}}{y^2} \bigg|_{x=0}^{x=1} \right) \mathrm{d}y \\
&= \int_0^1 y e^{y^2} - y\,\mathrm{d}y \\
&= \left(\frac{e^{y^2}}{2} - \frac{y^2}{2}\right)\bigg|_0^1 \\
&= \frac{e}{2} - 1
\end{aligned}
$$

33.3 1-D Fubini's corollary with bounding functions

Corollary 33.2. Let $\phi, \psi : [a,b] \to \mathbb{R}$ **continuous** with $\phi(x) \leq \psi(x)$ for all $x \in [a,b]$.
Let $D = \{(x,y) \in \mathbb{R}^2, x \in [a,b], \phi(x) \leq y \leq \psi(x)\}$.
Let $c \leq d$ such that $c \leq \phi(x) \leq \psi(x) \leq d$ for all $x \in [a,b]$.

Let $f : D \to \mathbb{R}$ be bounded such that the set $D_0 \in \mathbb{R}^2$ of its points of discontinuity has size zero, **and**, $\{y \in [c,d], (x,y) \in D_0\}$ also has size zero as a subset of $\mathbb{R}^1$, for each $x \in [a,b]$.
Then f is integrable on D and

$$\int_D f = \int_a^b \left(\int_{\phi(x)}^{\psi(x)} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

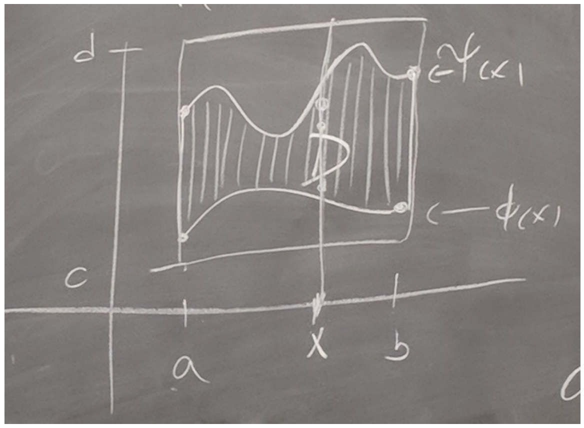

Figure 33.1: D is the area between $x \in [a,b]$ and $\phi(x) \le \psi(x)$. There may exist points of discontinuity (e.g. at x in the diagram) but the region is still integrable.

Proof. $D \subseteq [a,b] \times [c,d] = I$. Let $\tilde{f}$ be extension by zero of f to I. Since D_0 has size zero, by theorem 30.1 $\tilde{f}$ is integrable on I so f is integrable on D and

$$\int_D f = \int_I \tilde{f}$$

By our premise (where $\{y \in [c,d], (x,y) \in D_0\}$ has size zero), for each $x \in [a,b]$ the function $y \mapsto f(x,y)$ has discontinuities on a set of size zero.
Thus $y \mapsto f(x,y)$ is integrable on $[c,d]$ for all $x \in [a,b]$ and

$$\int_c^d f(x,y)\,\mathrm{d}y = \int_{\phi(x)}^{\psi(x)} f(x,y)\,\mathrm{d}y$$

So by Fubini's, we have

$$\int_D f = \int_a^b \left(\int_{\phi(x)}^{\psi(x)} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

□

Example 33.2. Let $D = \{(x,y) \in \mathbb{R}^2, 1 \le x \le 3, x^2 \le y \le x^2 + 1\}$.

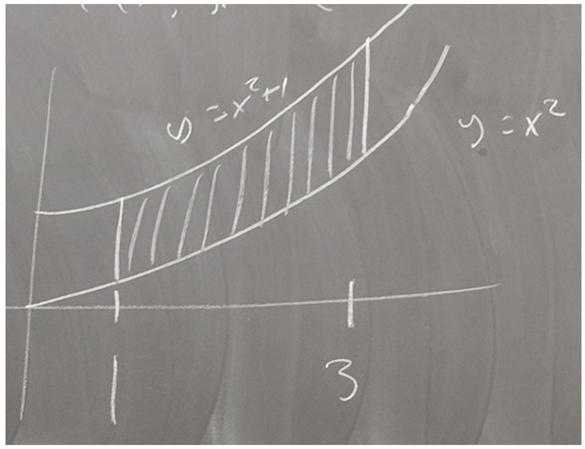

Figure 33.2: D is the area between x^2 and x^2+1 and between $x = 1$ and $x = 3$.

We have

$$\begin{aligned}\mu(D) &= \int_D 1 \\ &= \int_1^3 \left(\int_{x^2}^{x^2+1} 1 \, \mathrm{d}y \right) \mathrm{d}x \\ &= \int_1^3 \left(y \Big|_{y=x^2}^{y=x^2+1} \right) \mathrm{d}x \\ &= \int_1^3 1 \, \mathrm{d}x \\ &= 2\end{aligned}$$

34 March 28, 2018

34.1 Examples of evaluating integrals using Fubini's theorem

Example 34.1. Let D be the subset of $\mathbb{R}^3$ bounded by the planes $x = 0$, $y = 0$, $z = 0$, and $x + y + z = 1$. Compute $\int_D f$ where $f(x, y, z) = y$ (i.e. the "density" is y).

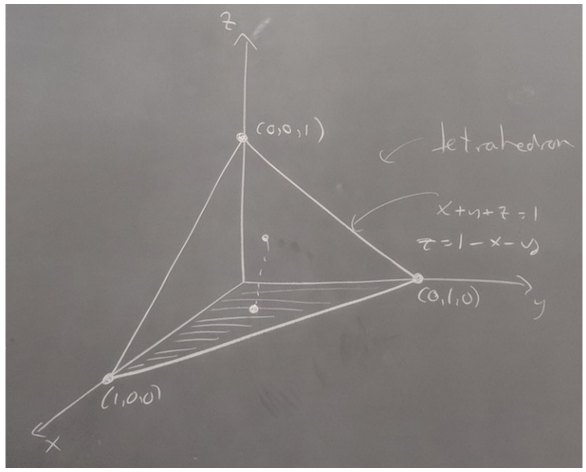

Figure 34.1: The tetrahedron is our domain we want to integrate over.

Note that the domain of our variables for the region D is as follows, we we first fix an x value then derive the other variables:

$$\begin{aligned} 0 &\leq x \leq 1 \\ 0 &\leq y \leq 1-x \\ 0 &\leq z \leq 1-x-y \end{aligned}$$

(there are actually 3! ways to specify these domains depending on the order we express the variables).

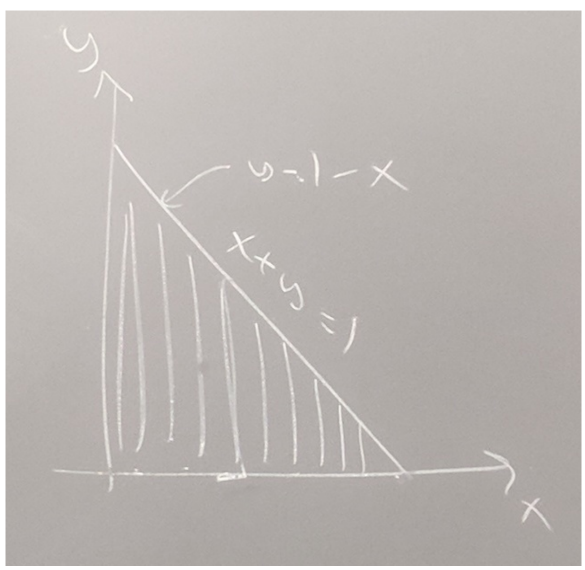

Figure 34.2: The bounds between x and y.

By Fubini

$$\int_D f = \int_0^1 \int_0^{1-x} \int_0^{1-x-y} y \quad \mathrm{d}z \, \mathrm{d}y \, \mathrm{d}x$$

Remark 34.1. A quick check to see if the order of integrals make sense is to see if the variables in the integral bounds "disappear" by the last integral.

$$\begin{aligned}
\int_D f &= \int_0^1 \int_0^{1-x} \left(yz \Big|_{z=0}^{z=1-x-y} \right) \mathrm{d}y\, \mathrm{d}x \\
&= \int_0^1 \int_0^{1-x} (1-x)y - y^2 \,\mathrm{d}y\, \mathrm{d}x \\
&= \int_0^1 \left((1-x)\frac{y^2}{2} - \frac{y^3}{3} \right) \Big|_{y=0}^{y=1-x} \mathrm{d}x \\
&= \int_0^1 \frac{1}{6}(1-x)^3 \,\mathrm{d}x \\
&= \frac{-(1-x)^4}{24} \Big|_0^1 \\
&= \frac{1}{24}
\end{aligned}$$

Example 34.2. Find the volume of the region D lying inside the "elliptic" cylinder $x^2 + 4y^2 = 4$ above the x-y plane and below the plane $z = 2 + x$.
We want to find $Vol(D) = \int_D 1$.

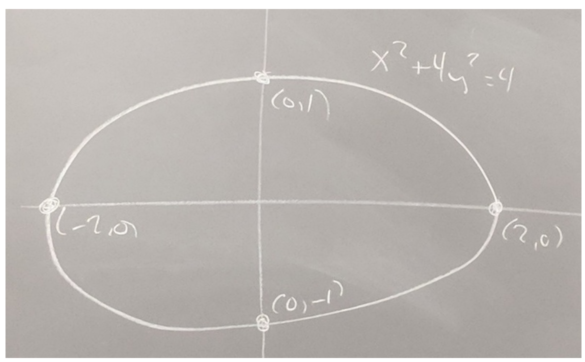

Figure 34.3: The region $x^2 + 4y^2 = 4$ on the x-y plane.

Note that the base is an ellipse since $x^2 + 4y^2 = 4 \Rightarrow \left(\frac{x}{2}\right)^2 + y^2 + 1$. We then extend this ellipse along the z axis.

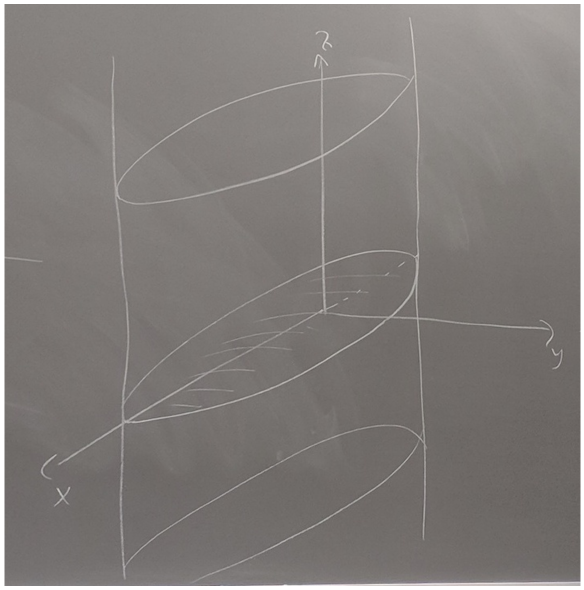

Figure 34.4: The ellipse extended along the z-axis.

Note that the plane is a function of only x. It is also not parallel to the x-y plane so our cylinder has a slanted top. From our ellipse we see that

$$-2 \leq x \leq 2$$
$$-\sqrt{1-\frac{x^2}{4}} \leq y \leq \sqrt{1-\frac{x^2}{4}}$$

And of course from our plane and the x-y plane we have

$$0 \leq z \leq 2+x$$

Thus we have

$$\begin{aligned}
Vol(D) &= \int_{-2}^{2}\int_{-\sqrt{1-\frac{x^2}{4}}}^{\sqrt{1-\frac{x^2}{4}}}\int_{0}^{2+x} 1 \quad \mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-2}^{2}\int_{-\sqrt{1-\frac{x^2}{4}}}^{\sqrt{1-\frac{x^2}{4}}} (2+x)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-2}^{2} 2(2+x)\sqrt{1-\frac{x^2}{4}}\,\mathrm{d}x \\
&= 4\int_{0}^{2} 2\sqrt{1-\frac{x^2}{4}}\,\mathrm{d}x \\
&= 8\int_{0}^{2} \sqrt{1-\frac{x^2}{4}}\,\mathrm{d}x \\
&= 16\int_{0}^{1} \sqrt{1-u^2}du && x = 2u, \mathrm{d}x = 2du \quad 0 \leq x \leq 2, 0 \leq u \leq 1 \\
&= 4\pi
\end{aligned}$$

Example 34.3. Suppose we wanted to find the integral

$$\begin{aligned}\int_0^1\int_z^1\int_0^x e^{x^3}\,\mathrm{d}y\,\mathrm{d}x\,\mathrm{d}z &= \int_0^1\int_z^1 e^{x^3}y\Big|_{y=0}^{y=x}\mathrm{d}x\,\mathrm{d}z\\ &= \int_0^1\int_z^1 xe^{x^3}\,\mathrm{d}x\,\mathrm{d}z\end{aligned}$$

It is hard (i.e. impossible) to find the explicit anti-derivative of xe^{x^3} in terms of x.
Fubini's theorem says we can change the order of integration. Note that we had $0 \leq z \leq 1$ and $z \leq x \leq 1$ from our integral bounds. Let us express x first then find the bounds of z in terms of x, i.e. $0 \leq x \leq 1$ and $0 \leq z \leq x$.
Thus we have

$$\begin{aligned}\int_0^1\int_0^x xe^{x^3}\,\mathrm{d}z\,\mathrm{d}x &= \int_0^1\left(xe^{x^3}z\Big|_{z=0}^{z=x}\right)\mathrm{d}x\\ &= \int_0^1 x^2e^{x^3}\,\mathrm{d}x\\ &= \frac{1}{3}e^{x^3}\Big|_0^1\\ &= \frac{1}{3}(e-1)\end{aligned}$$

34.2 Change of variables theorem

Theorem 34.1 (Change of variables theorem)**.** Let $U \subseteq \mathbb{R}^n$ be open and non-empty. Let $K \subseteq U$ be **compact**, non-empty and **sizeable**. Suppose $\psi : U \to \mathbb{R}^n$ is in $C^1(U)$. Suppose $\exists$ a subset $D \subseteq K$ with $\mu(D) = 0$ such that

1. $\psi\Big|_{K\setminus D}$ is injective
2. $det((D\psi)_x) \neq 0$ for all $x \in K \setminus D$.

Then $\psi(K)$ is **sizeable** and for any $f : \psi(K) \to \mathbb{R}^p$ which is *continuous*, then f is integrable on $\psi(K)$ and

$$\int_{\psi(K)} f = \int_K (f\circ\psi)|det(D\psi)|$$

where $f \circ \psi : K \to \mathbb{R}^p$ and $|det(D\psi)|$ is the "scaling factor".

34.3 Cylindrical coordinates

Example 34.4. Note that the cylindrical coordinates on $\mathbb{R}^3$ are

$$\begin{aligned}x &= r\cos\theta\\ y &= r\sin\theta\\ z &= z\end{aligned}$$

where

$$
\begin{aligned}
&0 \le r \le \infty \\
&0 \le \theta \le 2\pi \\
&-\infty < z < \infty
\end{aligned}
$$

(there is some symmetry around the z-axis).

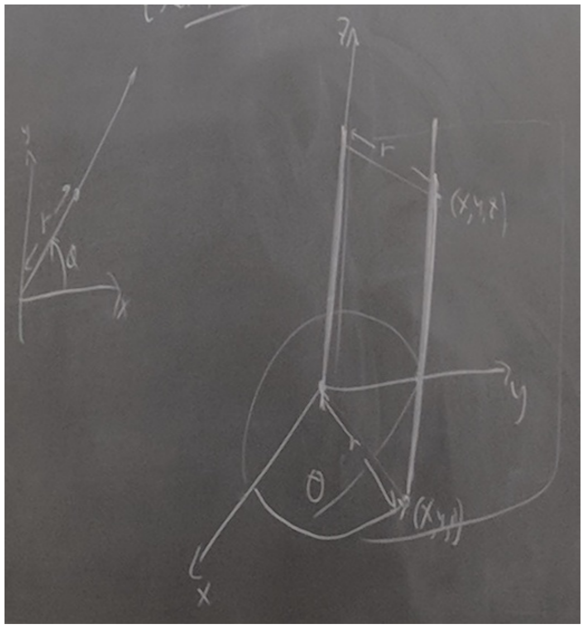

Figure 34.5: Cylindrical coordinates in the xyz $\mathbb{R}^3$ space.

Suppose we have

$$(x, y, z) = \psi(r, \theta, z) = (r\cos\theta, r\sin\theta, z)$$

where $\psi \in C^\infty(\mathbb{R}^3)$. ψ fails to be injective on half-plane $\theta = 0.$, which has size zero in $\mathbb{R}^3$ (when restricted to any compact subset).
We have

$$
\begin{aligned}
D\psi &= \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial z} \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
$$

where $|det(D\psi)| = r \neq 0$ except on set of size zero when $r = 0$ (when intersected with a compact subset).
The Change of variables theorem says if $f : \psi(K) \to \mathbb{R}$ is **continuous**, then

$$\int_{\psi(K)} f = \int_K (f \circ \psi) r$$

So we use cylindrical coordinates if the region and/or the function being integrated becomes simpler.

For example: suppose we want to find $\int_D f$ where $f(x,y,z) = x^2 + y^2$ and D is bounded by

$$\begin{aligned}
&x^2 + y^2 = 1 \\
&x^2 + y^2 = 4 \\
&z = 0 \\
&z = 1 \\
&y = 0 \\
&y = x
\end{aligned}$$

Note that $x^2 + y^2 = r^2$ in cylindrical coordinates.

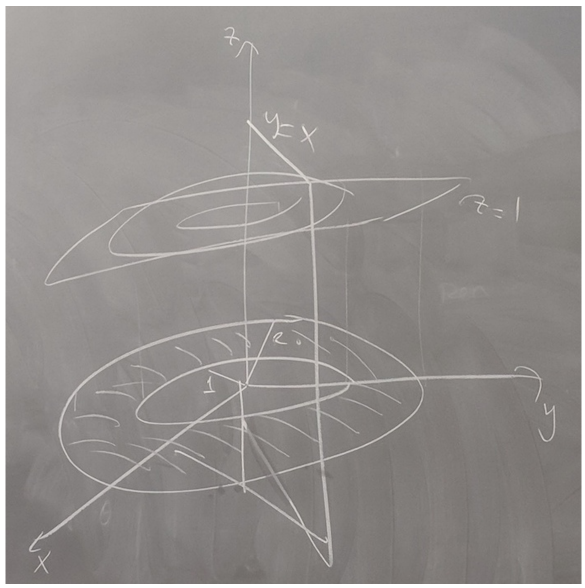

Figure 34.6: Domain of our example.

So in cylindrical coordinates we have

$$\begin{aligned}
1 \leq r \leq 2 \\
0 \leq \theta \leq \frac{\pi}{4} \\
0 \leq z \leq 1
\end{aligned}$$

Thus we find (where $f = r^2$ and $|det(D\psi)| = r$)

$$\begin{aligned}
\int_D f &= \int_0^1 \int_0^{\frac{\pi}{4}} \int_1^2 r^2 r \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}z \\
&= \frac{\pi}{4} \int_1^2 r^3 \, \mathrm{d}r \\
&= \frac{\pi}{16}(2^4 - 1) \\
&= \frac{15\pi}{16}
\end{aligned}$$

35 April 2, 2018

35.1 Geomtric heurisitic of cylindrical coordinates

If we imagine that we change r, θ, z by a small delta in each direction for a given cylindrical coordinate point, we see that it forms a small cube.

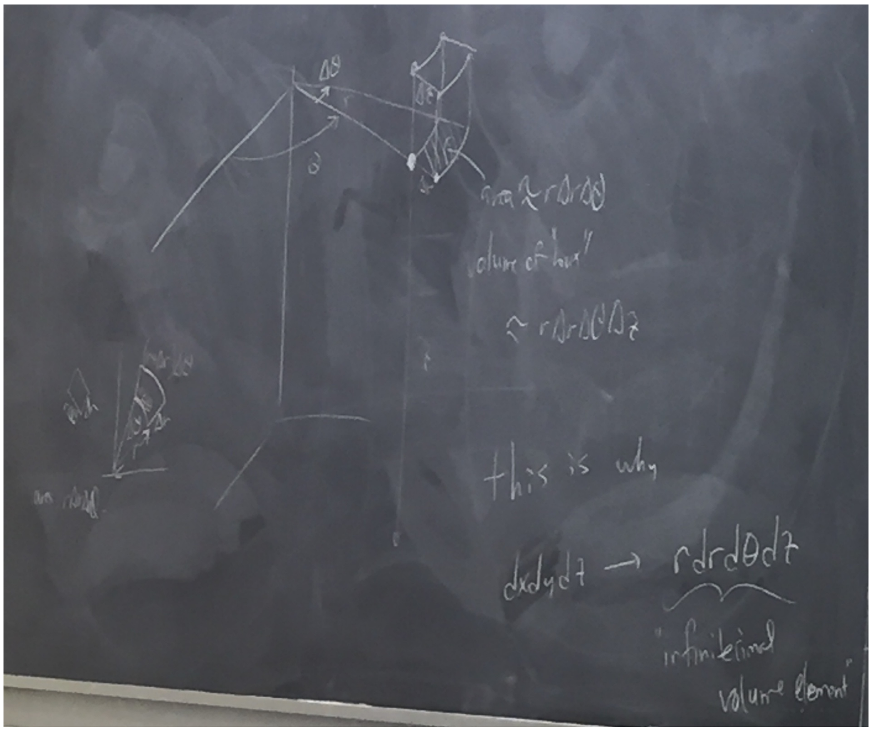

Figure 35.1: We take small deltas in each direction which forms small cubes we sum over.

When we integrate, we integrate over these small boxes to find our total volume.
Note that the base of the triangle is a rectangle with sides $r\Delta\theta$ $(r + \Delta r)\Delta\theta$, and Δr on the two other sides. Since Δr and $\Delta\theta$ is small, then we have a base with area $r\Delta r\Delta\theta$.
This is why $\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \to r\Delta r\Delta\theta\Delta z$ ("infintesimal volume element").

35.2 More examples with cylindrical coordinate

Example 35.1. Find the volume of the region **above** the paraboloid $z = x^2 + y^2 = r^2$. and inside the sphere $x^2 + y^2 + z^2 = 12$.

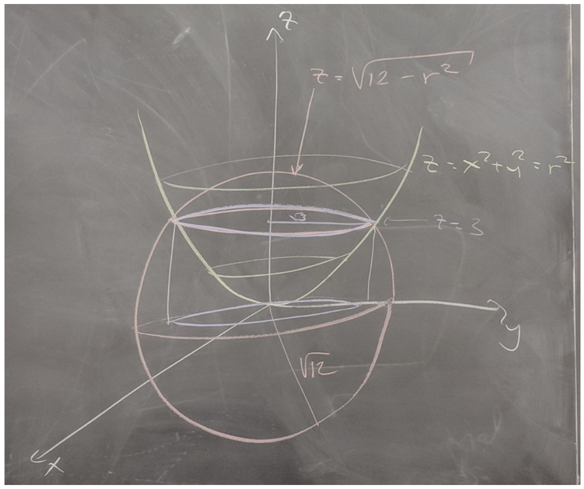

Figure 35.2: The region is bounded by the paraboloid and sphere shown.

These two surfaces intersect when $z = x^2 + y^2 \geq 0$ and $x^2 + y^2 + z^2 = 12$. So

$$\begin{aligned}
&z^2 + z - 12 = 0 \\
&(z-3)(z+4) = 0 \\
&z = 3 && z \geq 0
\end{aligned}$$

Intersect is the circle of radius $\sqrt{3}$ in the plane $z = 3$.
Thus the region D has bounds

$$\begin{aligned}
&0 \leq \theta \leq 2\pi \\
&0 \leq r \leq \sqrt{3} \\
&r^2 \leq z \leq \sqrt{12 - r^2}
\end{aligned}$$

So the volume of D is

$$\begin{aligned}
Vol(D) &= \int_0^{2\pi} \int_0^{\sqrt{3}} \int_{r^2}^{\sqrt{12-r^2}} 1r \,\mathrm{d}z \,\mathrm{d}r \,\mathrm{d}\theta \\
&= \int_0^{2\pi} \int_0^{\sqrt{3}} r(\sqrt{12-r^2} - r^2) \,\mathrm{d}r \,\mathrm{d}\theta \\
&= 2\pi \int_0^{\sqrt{3}} r(\sqrt{12-r^2}) - r^3 \,\mathrm{d}r \\
&= 2\pi \left(\frac{-1}{3}(12-r^2)^{\frac{3}{2}} - \frac{r^4}{4} \right) \Bigg|_0^{\sqrt{3}} \\
&= 2\pi \left(\frac{(12)^{\frac{3}{2}}}{3} - 9 - \frac{9}{4} \right)
\end{aligned}$$

35.3 Spherical coordinates

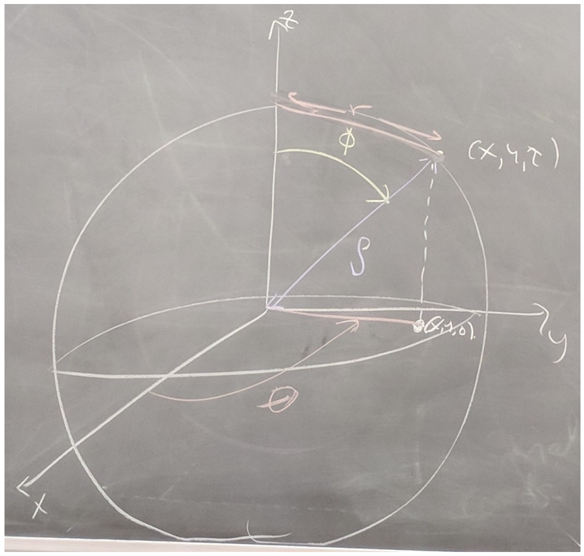

Figure 35.3: Spherical coordinates with ρ, θ, ϕ against the xyz axes.

We have three axes in spherical coordinates:

ρ^2 **- square of distance to origin** $\rho^2 = x^2 + y^2 + z^2$

θ **- "longitude"** Same as in cylindrical, angle from *positive x-axis*, $0 \leq \theta \leq 2\pi$

ϕ **- "latitude"** Angle from *positive z-axis*, $0 \leq \phi \leq \pi$, where

- $\phi = 0$ is the "north pole"
- $\phi = \pi$ is the "south pole"
- $\phi = \frac{\pi}{2}$ is the "equator"

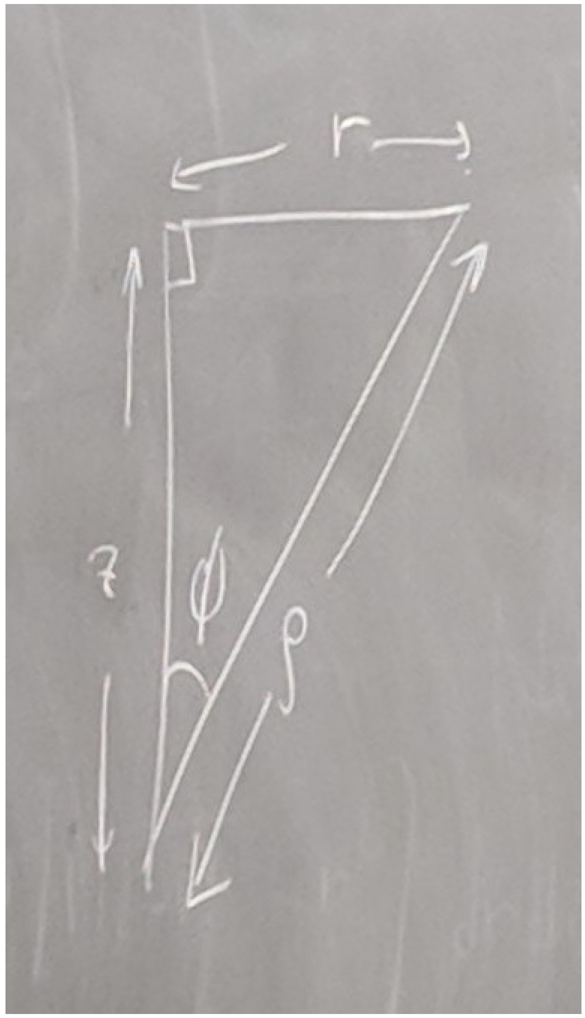

Figure 35.4: The z-axis with respect to ρ and r from cylindrical coordinates.

Note that from the triangle (derived by comparing with cylindrical coordinates), we see that $r = \rho \sin\theta$, thus

$$\begin{aligned} z &= \rho\cos\theta \\ x &= r\cos\theta = \rho\sin\phi\cos\theta \\ y &= r\sin\theta = \rho\sin\phi\sin\theta \end{aligned}$$

thus $\psi(\rho, \theta, \phi) = (\rho\sin\phi\cos\theta, \rho\sin\phi\sin\theta, \rho\cos\theta)$.

So taking the determinant of the derivative

$$\begin{aligned}
det(D\psi) &= det\begin{bmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \phi} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \phi} & \frac{\partial y}{\partial \theta} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \phi} & \frac{\partial z}{\partial \theta} \end{bmatrix} \\
&= det\begin{bmatrix} \sin\phi\cos\theta & \rho\cos\phi\cos\theta & -\rho\sin\phi\sin\theta \\ \sin\phi\sin\theta & \rho\cos\phi\sin\theta & \rho\sin\phi\cos\theta \\ \cos\phi & -\rho\sin\phi & 0 \end{bmatrix} \\
&= 0 + \rho^2\sin\phi\cos^2\phi cos^2\theta + \rho^2\sin^3\phi\sin^2\theta + \rho^2\sin\phi\cos^2\phi\sin^2\theta + 0 + \rho^2\sin^3\phi\cos^2\theta \\
&= \rho^2\sin\phi\cos^2\phi + \rho^2\sin\phi\sin^2\phi \\
&= \rho^2\sin\phi \qquad 0 \le \phi \le \pi \quad \sin\phi \ge 0
\end{aligned}$$

so $|det(D\psi)| = \rho^2\sin\phi$.

35.4 Geometric heuristic of spherical coordinates

We can perform the same geometric analysis of spherical coordinates by taking small deltas in each direction of ρ, θ, ϕ and note that we end up with a small cube.

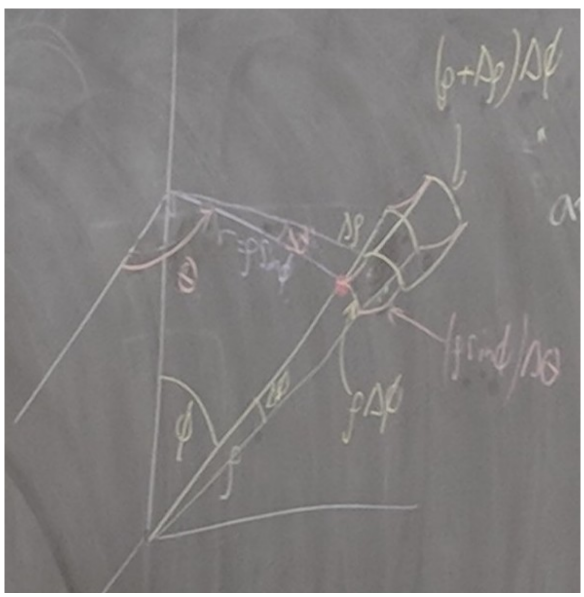

Figure 35.5: We take small deltas in each direction which forms small cubes we sum over.

The lengths of the base consists of a side of length $\rho\Delta\phi$ and $(\rho+\Delta\rho)\Delta\phi$, so we have a side $\rho\Delta\phi$.
Another side of the cube is simply $\Delta\rho$.
Finally the last sides has length $r\Delta\theta$ where $r = |\rho\sin\phi|$ and $\rho\sin(\phi+\Delta\phi)$. Note $\sin(\phi+\Delta\phi) = \sin\phi\cos(\Delta\phi) + \cos\phi\sin(\Delta\phi) = \sin\phi$, thus we get for the volume to be approximatley $\rho^2\sin\phi\Delta\rho\Delta\phi\Delta\theta$.

35.5 Canonical examples of spherical coordinates

Note that for ρ constant, it is a **sphere** centred at the origin of that radius.
Note that for ϕ constant, it is a **cone** centred at the origin where $0 < \phi < \frac{\pi}{2}$.

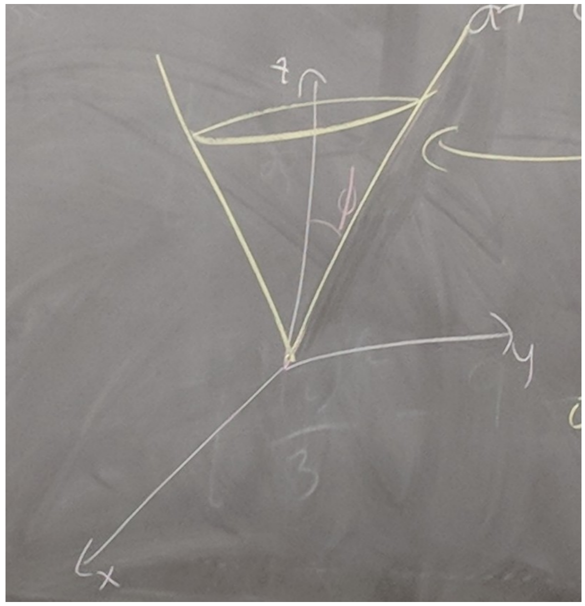

Figure 35.6: When $0 < \phi < \frac{\pi}{2}$ constant, we have the cone centred at the origin
.

For $\phi = \frac{\pi}{2}$ we have the circl on the x-y plane.

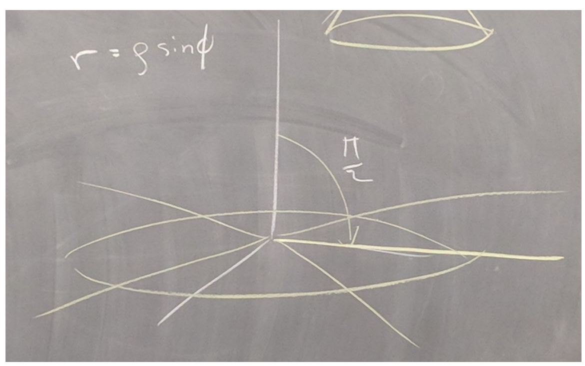

Figure 35.7: When $\phi = \frac{\pi}{2}$ constant, we have a circle on the x-y plane.

For $\frac{\pi}{2} < \phi < \pi$ we have an upside down cone.

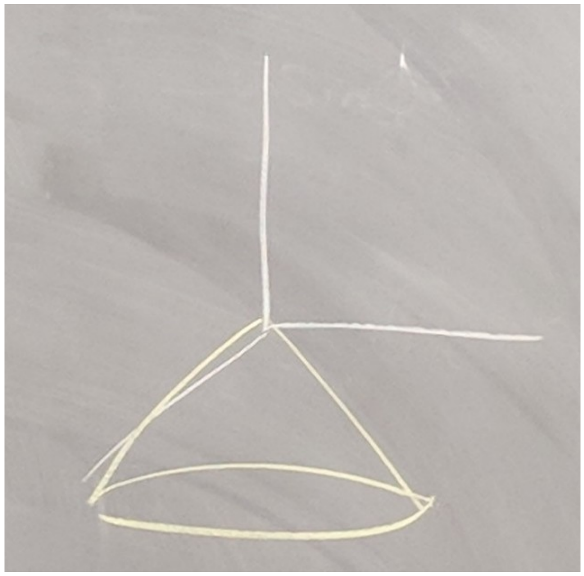

Figure 35.8: When $\frac{\pi}{2} < \phi < \pi$ constant, we have an upside-down cone.

35.6 Examples of spherical coordinates

Example 35.2. Find $\int_D g$ where $g(x, y, z) = 1 - \sqrt{x^2 + y^2 + z^2} = 1 - \rho$ and D is the region above the cone $z = \frac{1}{\sqrt{3}}\sqrt{x^2 + y^2}$ and inside the sphere $x^2 + y^2 + z^2 = 1$.

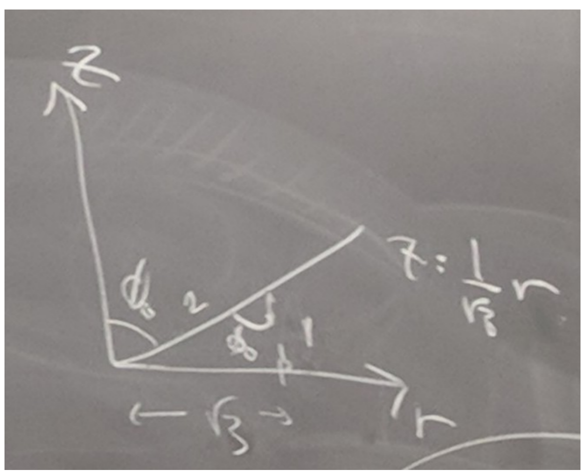

Figure 35.9: We compare the z value to r to find a bound for ϕ.

Note that $z = \frac{1}{\sqrt{3}}r$ so $\cos\phi_0 = \frac{1}{2}$ thus $\phi_0 = \frac{\pi}{3}$ (the angle from the z-axis of the cone).
The region of D is thus bounded by the cone and the sphere

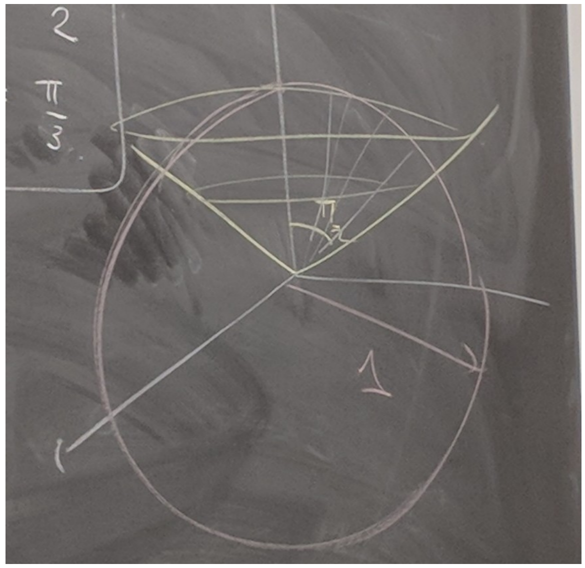

Figure 35.10: We want to find the bounds for the region bounded by the cone and the sphere.

Thus we have for the bounds of D

$$\begin{aligned} 0 &\le \theta \le 2\pi \\ 0 &\le \phi \le \frac{\pi}{3} \\ 0 &\le \rho \le 1 \end{aligned}$$

Thus we have

$$\begin{aligned} \int_D g &= \int_0^{2\pi}\int_0^{\frac{\pi}{3}}\int_0^1 (1-\rho)\rho^2 \sin\phi \,\mathrm{d}\rho\,\mathrm{d}\phi\,\mathrm{d}\theta \\ &= 2\pi \int_0^{\frac{\pi}{3}}\int_0^1 (\rho^2-\rho^3)\sin\phi\,\mathrm{d}\rho\,\mathrm{d}\phi \\ &= 2\left(\frac{1}{3}-\frac{1}{4}\right)\int_0^{\frac{\pi}{3}} \sin\phi\,\mathrm{d}\phi \\ &= \frac{\pi}{6}(-\cos\phi)\Big|_0^{\frac{\pi}{3}} \\ &= \frac{\pi}{6}\left(1-\frac{1}{2}\right) \\ &= \frac{\pi}{12} \end{aligned}$$

36 April 4, 2018

36.1 Idea behind change of variables formula

For the special case: $F : U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ which is $C^1(U)$, one-to-one everywhere and $(DF)_x$ invertible for all $x \in U$. Each component $F_k : U \subseteq \mathbb{R}^n \to \mathbb{R}$ is in $C^1(U)$ hence differentiable. Note that

$$F_k(x) = F_k(a) + (\nabla F_k)(a)\cdot(x-a) + R_{k,a}(x)$$

where $\lim_{x \to a} \frac{R_{k,a}(x)}{\|x-a\|} = 0$ by Taylor's theorem.
Let I be a box in $\mathbb{R}^n$ with a as the "lower left corner", i.e.

$$I = \{a + t_1e_1 + t_2e_2 + \ldots + t_ne_n \mid 0 \leq t_k \leq l_k\}$$

where l_k is the length of the kth edge of box.
Let $\mathbb{R}^n \supseteq F(I) = \{F(a + t_1e_1 + \ldots + t_ke_k\}$ and $x = a + t_1e_1 + \ldots + t_ke_k$ where $t_1e_1 + \ldots + t_ke_k$ is small (small box). Note that the length $\sqrt{t_1^2 + \ldots + t_n^2} \leq \sqrt{l_1^2 + \ldots + {}_n^2}$. We thus have

$$F_k(x) = F_k(a) + (\nabla F_k)(a) \cdot (t_1e_1 + \ldots + t_ne_n) + R_{k,a}(x)$$

In vector form we have

$$\begin{bmatrix} F_1(x) \\ \vdots \\ F_n(x) \end{bmatrix} = \begin{bmatrix} F_1(a) \\ \vdots \\ F_n(a) \end{bmatrix} + \begin{bmatrix} (\nabla F_1)(a) \\ \vdots \\ (\nabla F_n)(a) \end{bmatrix} \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} + R_a(x)$$

Thus we have

$$F(I) = \{F(a) + t_1(DF)_a(e_1) + t_2(DF)_a(e_2) + \ldots + t_n(DF)_a(e_n) + R_a(x)\}$$

and for I small

$$F(I) \approx \{F(a) + \sum_{k=1}^{n} t_k(DF)_a(e_k) \mid 0 \leq t_k \leq l_k\}$$

Note that each $(DF)_a(e_k)$ are linearly independent since $(DF)_x$ invertible for all $x \in U$.

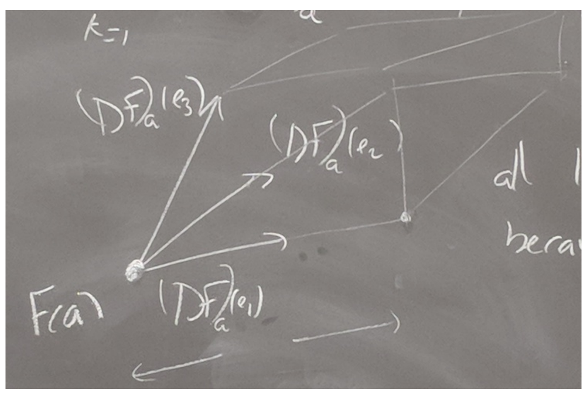

Figure 36.1: The volume of the parallelpiped formed by the vectors of $(DF_k)_a(e_k)$ and $F(a)$ approximates $\mu(F(I))$ for small enough I.

From linear algebra, the volume of this parallelpiped is $l_1 \cdot \ldots \cdot l_n$ multiplied by $|det((DF)a_(e_1) \ldots (DF)_a(e_n))|$. Note that $l_1 \cdot \ldots \cdot l_n$ is $\mu(I)$ thus we have $\mu(I)|det(DF)_a|$ for the volume and thus $\mu(F(I)) \approx |det(DF)_a|\mu(I)$.
So if $h : F(U) \subseteq \mathbb{R}^n \to \mathbb{R}$ where $h \in C^0(U)$ then

$$\int_{F(I)} h = \int_I (h \circ F)|det(DF)|$$

36.2 Fundamental thereom of calculus (FTC) for multiple variables

There is an analogous FTC in $\mathbb{R}^n$ to the FTC in one variable: **Stoke's theorem**.
M^k is a subset of $\mathbb{R}^n$ is called a k**-dimensional submanifold** if locally, it "looks like" an open set in $\mathbb{R}^k$.

The *manifold boundary* (not the same as set boundary) ∂M of M is a $(k-1)$-dimensional submanifold.
On k-dimensional manifolds, the objects that can be "integrated" are called **smooth k-forms** where Ω^k is the space of smooth k-forms.
There exist a natural notion of differentiation $d : \Omega^k \to \Omega^{k+1}$ the **exterior derivative** (takes something integrable on k-dimensional manifolds to something intergrable on $(k+1)$-dimensional manifolds).
Stoke's theorem: Let M be a k-dimensional, ∂M be $(k-1)$-dimensional manifold, $d\omega \in \Omega^k$ and $\omega \in \Omega^{k-1}$. Then

$$\int_{\partial M} \omega = \int_M d\omega$$