New Notes, Exercises and Problems for 402, Spring 2020

Kevin R. Vixie

April 13, 2020

Contents

1 Differentiation **3**

1.1 The Derivative 3

1.2 Variational Derivative for $\int_\Omega \nabla u \cdot \nabla u \, dx$ 3

1.3 Jacobian Matrices 6

1.3.1 If f is differentiable, then the derivative is the matrix of partial derivatives, ... 7

1.3.2 If $f \in C^1$, then f is differentiable. 9

1.3.3 Some More Exercises 13

1.4 Derivatives and Intersections 13

1.4.1 Warm up Exercises 14

1.4.2 The Theory 15

1.5 Pause for an overview of where we are going 16

1.6 Taylor Series 18

1.6.1 Mean Value Theorem Approach I 18

1.6.2 Mean Value Theorem Approach II 20

1.6.3 Derivative Definition Approach 21

1.7 Index Theory and Sard's Theorem 23

1.7.1 A special case of Sard's Theorem via the 5R covering theorem 24

1.7.2 A special case of Sard's theorem via more smoothness and Compactness 26

1.7.3 Another Exercise 27

1.8 Norms of Operators 28

1.9 Introduction to using Derivative Approximations 28

1.9.1 Inverse Function Theorem: $f : \mathbb{R} \to \mathbb{R}$ 29

1.9.2 Implicit Function Theorem: $f : \mathbb{R}^2 \to \mathbb{R}$ 29

1.10 Inverse and Implicit Function Theorems 30

1.10.1 Review: $\mathbb{R}^n$ and why we like it. 30
1.10.2 k-Manifolds in $\mathbb{R}^n$ are locally like $\mathbb{R}^k$ 32
1.10.3 Reminder: Derivatives as linear approximations 33
1.10.4 Review: Full rank maps 34
1.10.5 Finally: Inverse and Implicit function theorem in higher dimensions 35
1.10.6 Implicit Function Theorem, Intuitively, Again 37
1.10.7 Using the Implicit Function Theorem: Co-Dimension 1 38

2 Integration **46**
2.1 Riemann vs Lebesgue 46
2.2 Outer Measures 50
2.2.1 Lebesgue and Hausdorff Measures 52
2.2.2 Caratheodory Criterion 57
2.2.3 Radon Measures and Approximation 59
2.3 Measurable Functions and Integration 59
2.3.1 Lebesgue Integration 60
2.3.2 Simple Functions 62
2.3.3 Integrating Simple Functions 62
2.3.4 Integrating Arbitrary Functions 62
2.3.5 Properties of Integrals and Measurable Functions 64
2.4 Modes of Convergence and Three Theorems 64
2.4.1 Examples 65
2.4.2 Types or Modes of Convergence 66
2.4.3 The Three Theorems 67
2.4.4 Proofs and Discussion of the Three Theorems 69
2.5 Area, Co-Area, and Sard's Theorem 71
2.5.1 Lipschitz Functions 71
2.5.2 Area and Coarea formulas 71
2.5.3 Sard's Theorem 73

3 Pause: A return to the Three Integrals in Section (1.5) **74**

4 Other Topics **74**
4.1 Fixed point theorems: Banach Fixed Point Theorem 74
4.2 Transversality is Generic 76

5 Problems for the Course **77**
5.1 Geometry of Differentiation 77
5.2 Higher Order Derivative Implications 78
5.3 Lebesgue Measure 78
5.4 Hausdorff Measure 79
5.5 Integration and Differentiation 79

1 Differentiation

1.1 The Derivative

Definition 1.1. *Suppose that X and Y are normed vector spaces, with norms $|\cdot|_X$ and $|\cdot|_Y$, and $f : X \to Y$. Then we say that f has the derivative $D_x f \equiv A$ at $x \in X$, if there is a linear operator $A : X \to Y$ such that:*

$$f(x+h) - f(x) = A(h) + g(h)$$

where

$$\frac{|g(h)|_Y}{|h|_X} \underset{|h|_X \to 0}{\to} 0.$$

We will usually suppress the X and Y on the norms so that the last condition becomes

$$\frac{|g(h)|}{|h|} \underset{|h| \to 0}{\to} 0$$

where we understand from the context that the norm is the correct one for the vector it is measuring. I.e. since $g(h) \in Y$, then $|g(h)|$ must actually be $|g(h)|_Y$.

In a nutshell: a function f is differentiable at x, if it is arbitrarily well approximated by a fixed linear transformation near x.

Remark 1.1. *Of any $g(h)$ satisfying this last condition, we would say "$g(h)$ is in $o(h)$", which read literally as "$g(h)$ is in little o of h".*

1.2 Variational Derivative for $\int_\Omega \nabla u \cdot \nabla u \, dx$

Suppose that

1. For any twice differentiable $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ we define the operator:

$$F(u) \equiv \int_\Omega \nabla u \cdot \nabla u \;\; dx$$

2. and we consider a perturbation to u, $h : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ and $h|_{\partial\Omega} = 0$. (Think of this as a direction in the function space we want to move and see how F changes. $h|_{\partial\Omega} = 0$ means that h is 0 on the boundary of the domain Ω.)

3. We also recall that in this function space $|h| = (\int_\Omega h^2 \, dx)^{\frac{1}{2}}$.

4. Now we restrict ourselves to h's of the form $h = \alpha_h g$ where $|g| = 1$, where α_h is some real number. (Notice that this is really no restriction since for any h, we can define $\alpha_h \equiv |h|$, note that $|\frac{h}{|h|}| = 1$ and get that $h = \alpha_h \frac{h}{|h|}$.)

We want to show that $h \to \int_\Omega \Delta u\ h\ \ dx$ is a linear approximation to the derivative to F at u. That is, that it is the derivative operator for F at u.

Recalling the definition of derivative as the linear operator L_u (if it exists) that satisfies:

$$F(u+h) - F(u) = L_u(h) + r(h)$$

where $\frac{|r(h)|}{|h|} \to 0$ as $|h| \to 0$, we begin computing and rearranging terms:

$$\begin{aligned}
F(u+h) - F(u) &= \int_\Omega \nabla u \cdot \nabla u\ dx + 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h\ dx - \int_\Omega \nabla u \cdot \nabla u\ dx \\
&= 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h\ dx \\
&= 2\alpha_h \int_\Omega \nabla u \cdot \nabla g + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx \\
&= -2\int_\Omega \Delta u\ (\alpha_h g) + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx \text{ (Divergence Theorem)} \\
&= -2\int_\Omega \Delta u\ h + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx
\end{aligned}$$

where we have used the vector calculus version of integration by parts to get the last equation.

(To see the step labeled "Divergence Theorem" above, notice that by the divergence theorem, we get

$$\int_\Omega \nabla \cdot (g\nabla u)\ dx = \int_{\partial\Omega} (g\nabla u) \cdot \vec{n}\ d\sigma$$

where $\vec{n}$ is the outward normal vector to $\partial\Omega$, and since $g = 0$ on $\partial\Omega$, we get that the right hand side is 0. That is

$$\int_\Omega \nabla \cdot (g\nabla u) = 0$$

Evaluating the left hand side, we get:

$$\begin{aligned}
\int_\Omega \nabla \cdot (g\nabla u)\ dx &= \int_\Omega \nabla g \cdot \nabla u + g(\nabla \cdot \nabla u)\ dx \\
&= \int_\Omega \nabla g \cdot \nabla u + g\Delta u\ dx
\end{aligned}$$

)

Restating where we are:

$$F(u+h) - F(u) = -2\int_\Omega \Delta u\ h + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx$$

we first recognize that

$$-2\int_\Omega \Delta u\ h\ dx$$

is linear in h so we define

$$L_u(h) \equiv -2\int_\Omega \Delta u\ h\ dx$$

which lets us conclude that:

$$F(u+h) - F(u) = L_u(h) + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx$$

Using the fact that $\alpha_h = |h|$, we define

$$\begin{aligned} r(h) &\equiv |h|^2 \int_\Omega \nabla\left(\frac{h}{|h|}\right) \cdot \nabla\left(\frac{h}{|h|}\right)\ dx \\ &= \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx \end{aligned}$$

and we get that

$$F(u+h) - F(u) = L_u(h) + r(h).$$

All we need to do now is show that $r(h) \sim o(h)$ and we are done.

$$\begin{aligned} \frac{|r(h)|}{|h|} &= \frac{|\alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx|}{|h|} \\ &= \frac{|\alpha_h^2 \int_\Omega \nabla g \cdot \nabla g\ dx|}{\alpha_h} \\ &= \alpha_h |\int_\Omega \nabla g \cdot \nabla g\ dx| \\ &= |h| \left|\int_\Omega \nabla g \cdot \nabla g\ dx\right| \\ &\to 0 \text{ (as } |h| \to 0) \end{aligned}$$

Note that when we fixed g and varied α_h in order to change h, this resulted in us using $\alpha_h \to 0$ to get $|h| \to 0$. And in doing this, we chose one, 1-dimensional path to 0. (That is, we ended up calculating a directional derivative.) In order to get the more general limit that is path independent, we actually need to choose our norm on the space of functions and perturbations more carefully. Given that we are assuming that the functions are twice differentiable, a norm of a function $w\Omega\mathbb{R}^n \to \mathbb{R}$, $|w|$, that works for this task is given by:

$$|w| \equiv \int_\Omega |w(x)|\, dx + \int_\Omega |\nabla w(x)|\, dx \int_\Omega |\Delta w(x)|\, dx$$

Exercise 1.1. Show that **if**

1. we define L_u as before,
2. use this new norm,
3. recall that we just calculated:

$$\begin{aligned} F(u+h) - F(u) &= \int_\Omega \nabla u \cdot \nabla u\, dx + 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h\, dx - \int_\Omega \nabla u \cdot \nabla u\, dx \\ &= 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h\, dx \\ &= -2\int_\Omega \Delta u\, h + \int_\Omega \nabla h \cdot \nabla h\, dx \\ &= L_u(h) + \int_\Omega \nabla h \cdot \nabla h\, dx \end{aligned}$$

4. and define $r(h) \equiv \int_\Omega \nabla h \cdot \nabla h\, dx$,

we can conclude that

1. $|r(h)| < |h|^2$
2. $F(u+h) - F(u) = L_u(h) + r(h)$ and
3. $r(h) \sim o(h)$

The point of this exercise is that using a harder to understand norm, leads to an easier proof of a nicer limit (the limit is path independent, whereas the first limit we found was actually a directional derivative.

1.3 Jacobian Matrices

While we know (by definition) that f is differentiable at x if there is an L_x such that:

$$f(x+h) - f(x) = L_x(h) + r(h)$$

where

1. $f : \mathbb{R}^n \to \mathbb{R}^m$,
2. $r(h) \sim o(h)$ and
3. $L_x : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function,

a practically important question is "**How do we compute L_x from $f(x)$?**"

Answer: L_x is the matrix of partial derivatives of f:

$$D_x f = L_x = \partial_x f = \begin{pmatrix} \partial_{x_1} f_1 & \partial_{x_2} f_1 & \cdots & \partial_{x_n} f_1 \\ \partial_{x_1} f_2 & \partial_{x_2} f_2 & \cdots & \partial_{x_n} f_2 \\ \vdots & \vdots & & \vdots \\ \partial_{x_1} f_m & \partial_{x_2} f_m & \cdots & \partial_{x_n} f_m \end{pmatrix}$$

How we go about showing this is true: We will first show that if there is a linear function satisfying the definition of derivative, it must be the matrix of partial derivatives, and then we show that if f has continuous partial derivatives, then f is differentiable.

1.3.1 If f is differentiable, then the derivative is the matrix of partial derivatives,

We will show this in the case that $n = m = 2$ and note that the proof in the case of general n and m is completely analogous. In that case, the equation for the derivative is given by:

$$\begin{bmatrix} f_1(x_1 + h_1, x_2 + h_2) \\ f_2(x_1 + h_1, x_2 + h_2) \end{bmatrix} - \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} r_1(h) \\ r_2(h) \end{bmatrix}$$

where we have used a completely general form for the derivative matrix:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Notice first that this is really two equations:

$$f_1(x_1 + h_1, x_2 + h_2) - f_1(x_1, x_2) = ah_1 + bh_2 + r_1(h)$$

and

$$f_2(x_1 + h_1, x_2 + h_2) - f_2(x_1, x_2) = ch_1 + dh_2 + r_2(h).$$

and that each equation is true for all h. Suppose we set $h_2 = 0$. This gives us that:

$$f_1(x_1 + h_1, x_2) - f_1(x_1, x_2) = ah_1 + r_1((h_1, 0))$$

and if we divide by h_1, we get:

$$\begin{aligned}
\frac{f_1(x_1+h_1,x_2)-f_1(x_1,x_2)}{h_1}-a &= \frac{r_1((h_1,0))}{h_1} && (1)\\
&\le \frac{|r_1((h_1,0))|}{|h_1|} && (2)\\
&= \frac{|r_1(h)|}{|h|} && (3)\\
&\to 0 \text{ (as } |h|\to 0 \text{)} && (4)
\end{aligned}$$

But this is just saying that

$$(\partial_{x_1} f_1)(x) = a$$

i.e. $\partial_{x_1} f_1$ evaluated at x equals a. We will suppress the point at which we are evaluating the partial derivative if it is clear from the context where that point is.

Now setting $h_1 = 0$, we get

$$\begin{aligned}
\frac{f_1(x_1,x_2+h_2)-f_1(x_1,x_2)}{h_2}-b &= \frac{r_1((0,h_2))}{h_2} && (5)\\
&\le \frac{|r_1((0,h_2))|}{|h_2|} && (6)\\
&= \frac{|r_1(h)|}{|h|} && (7)\\
&\to 0 \text{ (as } |h|\to 0 \text{)} && (8)
\end{aligned}$$

and conclude that

$$(\partial_{x_2} f_1)(x) = a$$

i.e. $\partial_{x_2} f_1$ evaluated at x equals b.

In a completely analogous way, we get that

$$(\partial_{x_1} f_2)(x) = c$$

and

$$(\partial_{x_2} f_2)(x) = d$$

so that we have:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \partial_{x_1} f_1 & \partial_{x_2} f_1 \\ \partial_{x_1} f_2 & \partial_{x_2} f_2 \end{bmatrix}.$$

And that is what we set out to show.

1.3.2 If $f \in C^1$, then f is differentiable.

The four ingredients we need for this part are:

(1) The mean value theorem in 1 dimension: if $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ and f is differentiable everywhere in (a, b), then there is a $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{(b - a)} = f'(c)$$

which can be rewritten in the equivalent form

$$f(a + h) - f(a) = f'(c) \cdot h$$

where $b - a = h$.

(2) A function s that bounds the convergence of a collection of functions, each continuous at x: if

(a) $g_j : \mathbb{R}^n \to \mathbb{R}$ (for $j = 1, 2, ...k$)

(b) $\lim_{y \to x} g_j = g_j(x)$ (for $j = 1, 2, ..., k$),

then there is a function $s : [0, \infty] \to [0, \infty]$ such that:

(a) s is monotonically increasing: $\omega_1 < \omega_2 \Rightarrow s(\omega_1) \leq s(\omega_2)$.

(b) $\lim_{\omega \to 0} s(\omega) = 0$

(c) and

$$\sup_{j \in \{1,2,...,k\},\ y \in B_x(\omega)} |g_j(y) - g_j(x)| \leq s(\omega)$$

where $B_x(\omega)$ is the ball centered at x with radius.

Exercise 1.2. Prove that such an $s(\omega)$ exists.

(3) The realization that we can go from x to $x + h$ in n dimensions in a series of n steps that each change only one coordinate:

$$\begin{aligned}
f(x_1 + h_1, \quad ... \quad &, x_n + h_n) - f(x_1, ..., x_n) \\
\textbf{step n} \quad = \quad & f(x_1 + h_1, ..., x_n + h_n) - f(x_1 + h_1, ..., x_{n-1} + h_{n-1}, x_n) \\
\textbf{step n-1} \quad + \quad & f(x_1 + h_1, ..., x_{n-1} + h_{n-1}, x_n) - f(x_1 + h_1, ..., , x_{n-2} + h_{n-2}, x_{n-1}, x_n) \\
\textbf{step n-2} \quad + \quad & f(x_1 + h_1, ..., x_{n-2} + h_{n-2}, x_{n-1}, x_n) - f(x_1 + h_1, ..., x_{n-3} + h_{n-3}, x_{n-2}, ...) \\
+ \quad & \vdots \\
\textbf{step 2} \quad + \quad & f(x_1 + h_1, x_2 + h_2, x_3, ..., x_n) - f(x_1 + h_1, x_2, x_3 ..., x_n) \\
\textbf{step 1} \quad + \quad & f(x_1 + h_1, x_2, x_3, ..., x_n) - f(x_1, ..., x_n)
\end{aligned}$$

(4) The realization that we only need to prove the assertion for a function $f : \mathbb{R}^n \to \mathbb{R}$: because the general case of $f : \mathbb{R}^n \to \mathbb{R}^m$ is just a collection of m functions $f_i : \mathbb{R}^n \to \mathbb{R}$. That is, the assertion that there is an L_x such that:

$$f(x+h) - f(x) = L_x(h) + r(h)$$

where

(a) $f : \mathbb{R}^n \to \mathbb{R}^m$,

(b) $r(h) \sim o(h)$ and

(c) $L_x : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function,

is completely equivalent to the assertion that, for $i = 1, 2, ...m$ there is an L_x^i such that:

$$f_i(x+h) - f_i(x) = L_x(h) + r_i(h)$$

where

(a) $f_i : \mathbb{R}^n \to \mathbb{R}$,

(b) $r_i(h) \sim o(h)$ and

(c) $L_x^i : \mathbb{R}^n \to \mathbb{R}$ is a linear function,

Now, putting these together, we start by writing what we want to show: Ingredient **(4)** implies that what we want to prove is:

$$\begin{aligned} f(x_1+h_1, ..., x_n+h_n) - f(x_1, ..., x_n) &= \partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n + r(h) \\ &= \nabla f \cdot \vec{h} + r(h) \end{aligned}$$

with the constraint that $r(h) \sim o(h)$ or, equivalently

$$f(x_1+h_1, ..., x_n+h_n) - f(x_1, ..., x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n) = r(h) \qquad (9)$$

for some $r(h)$ such that $r(h) \sim o(h)$.

Now, using (3) we get that: the left hand side of the last equation

$$f(x_1+h_1, ..., x_n+h_n) - f(x_1, ..., x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n)$$

is the sum of n pieces:

$$
\begin{aligned}
& f(x_1+h_1,...,x_n+h_n) - f(x_1,...,x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n) \\
= \; & f(x_1+h_1,...,x_n+h_n) - f(x_1+h_1,...,x_{n-1}+h_{n-1},x_n) - \partial_{x_n} f \cdot h_n \\
+ \; & f(x_1+h_1,...,x_{n-1}+h_{n-1},x_n) - f(x_1+h_1,...,,x_{n-2}+h_{n-2},x_{n-1},x_n) - \partial_{x_{n-1}} f \cdot h_{n-1} \\
+ \; & f(x_1+h_1,...,x_{n-2}+h_{n-2},x_{n-1},x_n) - f(x_1+h_1,...,x_{n-3}+h_{n-3},x_{n-2},...) - \partial_{x_{n-2}} f \cdot h_{n-2} \\
+ \; & \vdots \\
+ \; & f(x_1+h_1,x_2+h_2,x_3,...,x_n) - f(x_1+h_1,x_2,x_3...,x_n) - \partial_{x_2} f \cdot h_2 \\
+ \; & f(x_1+h_1,x_2,x_3,...,x_n) - f(x_1,...,x_n) - \partial_{x_1} f \cdot h_1
\end{aligned}
$$

Now we use (1) to get that the first difference: of each of these n pieces is exactly equal to the partial derivative evaluated at a point

$$\hat{c}_i \equiv (x_1, x_2, ..., x_{i-1}, c_i, x_{i+1}+h_{i+1}, ..., x_n+h_n),$$

where $(x_i < c_i < x_i + h_i)$. That is,

$$
\begin{aligned}
f(x_1+h_1,...,x_n+h_n) - f(x_1+h_1,...,x_{n-1}+h_{n-1},x_n) &= \partial_{x_n} f(\hat{c}_n) \cdot h_n \\
f(x_1+h_1,...,x_{n-1}+h_{n-1},x_n) - f(x_1+h_1,...,,x_{n-2}+h_{n-2},x_{n-1},x_n) &= \partial_{x_{n-1}} f(\hat{c}_{n-1}) \cdot h_{n-1} \\
f(x_1+h_1,...,x_{n-2}+h_{n-2},x_{n-1},x_n) - f(x_1+h_1,...,x_{n-3}+h_{n-3},x_{n-2},...) &= \partial_{x_{n-2}} f(\hat{c}_{n-2}) \cdot h_{n-2} \\
&\vdots \\
f(x_1+h_1,x_2+h_2,x_3,...,x_n) - f(x_1+h_1,x_2,x_3...,x_n) &= \partial_{x_2} f(\hat{c}_2) \cdot h_2 \\
f(x_1+h_1,x_2,x_3,...,x_n) - f(x_1,...,x_n) &= \partial_{x_1} f(\hat{c}_1) \cdot h_1
\end{aligned}
$$

this then allows us to write the left hand side of Equation (9) as

$$
\begin{aligned}
& (\partial_{x_1} f(\hat{c}_1) \cdot h_1 - \partial_{x_1} f(x) \cdot h_1) \\
+ \; & (\partial_{x_2} f(\hat{c}_2) \cdot h_2 - \partial_{x_2} f(x) \cdot h_2) \\
\vdots \; & \\
+ \; & (\partial_{x_n} f(\hat{c}_n) \cdot h_n - \partial_{x_n} f(x) \cdot h_n)
\end{aligned}
$$

or equivalently as

$$
\begin{aligned}
& (\partial_{x_1} f(\hat{c}_1) - \partial_{x_1} f(x)) \cdot h_1 \\
+ \; & (\partial_{x_2} f(\hat{c}_2) - \partial_{x_2} f(x)) \cdot h_2 \\
\vdots \; & \\
+ \; & (\partial_{x_n} f(\hat{c}_n) - \partial_{x_n} f(x)) \cdot h_n
\end{aligned}
$$

But we are assuming that each of the partial derivatives are continuous at x, so by (2) we have that: there is a function $s : [0, \infty] \to [0, \infty]$ such that $s(|h|) \to 0$ when $|h| \to 0$ and

$$|\partial_{x_i} f(\hat{c}_i) - \partial_{x_i} f(x)| \leq s(|h|)$$

due to the fact that, for all i,

$$|\hat{c}_i - x| \leq |h|$$

Now, noting that for all i, $|h_i| \leq |h|$, allows us:, finally, to compute a bound for the left hand side of Equation (9):

$$\begin{aligned} f(x_1 + h_1, ..., x_n + h_n) &- f(x_1, ..., x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n) \\ &= (\partial_{x_1} f(\hat{c}_1) - \partial_{x_1} f(x)) \cdot h_1 \\ &+ (\partial_{x_2} f(\hat{c}_2) - \partial_{x_2} f(x)) \cdot h_2 \\ &\quad \vdots \\ &+ (\partial_{x_n} f(\hat{c}_n) - \partial_{x_n} f(x)) \cdot h_n \\ &\leq s(|h|) \cdot |h_1| + s(|h|) \cdot |h_2| + ... + s(|h|) \cdot |h_n| \\ &\leq n \, s(|h|)|h| \end{aligned}$$

Which implies that

$$\begin{aligned} |r(h)| &\equiv |f(x_1 + h_1, ..., x_n + h_n) - f(x_1, ..., x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n)| \\ &\leq n \, s(|h|)|h| \end{aligned}$$

which implies that

$$\begin{aligned} \frac{|r(h)|}{|h|} &\leq n \, s(|h|) \\ &\to 0 \text{ (as } |h| \to 0) \end{aligned}$$

which means that

$$r(h) \sim o(h)$$

and **that concludes our proof that $f \in C^1$ implies f is differentiable (i.e. there are linear approximations where f is C^1)**.

1.3.3 Some More Exercises

Note: exercises are not always directly related to what has just been covered. They are meant to encourage exploration and discovery in the same general vicinity as what we are covering, but you should not necessarily try to see some close connection to the section we just studied. IN this case, none of these problems are about derivatives of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. Nevertheless, these exercises do give you a facility that is very useful in your quest for mastery of (a non-boring version) of analysis.

Exercise 1.3. Find a function $f : \mathbb{R} \to \mathbb{R}$ that is

1. discontinuous everywhere except at $x = 0$
2. is not only continuous at $x = 0$ but is actually also differentiable $x = 0$.

Hint: use the region between the graphs of $f(x) = x^2$ and $f(x) = -x^2$ to guide your thinking.

Exercise 1.4. **(Harder)** Find a function $f : [0, 1] \subset \mathbb{R} \to [0, 1]$ that is:

1. Monotonically increasing
2. Discontinuous at every rational point in (0,1)
3. Continuous at every irrational point in (0,1)

Hints: (a) enumerate the rationals in $\mathbb{Q} \cap (0, 1)$ to get $q_1, q_2, ...$ and (b) notice that $\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$.

Exercise 1.5. **(Harder) Suppose:** that we denote the number of points in a set E by $|E|$ and we have that

1. $f : [0, 1] \to \mathbb{R}$ is differentiable everywhere,
2. $X_c \equiv \{x \mid f(x) = c\}$, and
3. $|\frac{df}{dx}(y)| \geq \alpha > 0$ for all $y \in X_c$.

Prove: that $|X_c|$ is finite. Hint: use the cone property you have been asked to prove in part 1 of **Problem (5.1)** .

1.4 Derivatives and Intersections

We now take Exercise (1.5) and run with it. Here are three exercises to get us started (you also encountered them in the class on January 29, 2020).

1.4.1 Warm up Exercises

Exercise 1.6. **(without hints = Harder) Suppose:** that we denote the number of points in a set E by $|E|$ and we have that

1. $f : [0,1] \to \mathbb{R}$ is differentiable everywhere,
2. $X_c \equiv \{x \mid f(x) = c\}$, and
3. $|\frac{df}{dx}(y)| > 0$ for all $y \in X_c$.

Prove: that $|X_c|$ is finite.
Hint: use (1) an assumption that all derivatives are non-zero **and** $|X_c| = \infty$ (in an effort to get a contradiction) , (2) the compactness of [0,1], (3) the continuity of f, (4) the cone property you have been asked to prove in part 1 of **Problem (5.1)** .

Exercise 1.7. **Again:** suppose that we denote the number of points in a set E by $|E|$ and we have that

1. $f : [0,1] \to \mathbb{R}$ is differentiable everywhere,
2. $|\{x \mid df dx(x) = 0\}| < \infty$
3. $X_c \equiv \{x \mid f(x) = c\}$

Prove: that $|X_c|$ is finite. Hint: Suppose that $N \equiv |\{x \mid df dx(x) = 0\}|$

and that $|X_c| \geq N + 2$. Find a contradiction.

Exercise 1.8. **(Harder) Let's:** see if we can bound the number of points in $|X_c|$:

1. $f : [0,1] \to \mathbb{R}$ is in fact twice differentiable everywhere, i.e. $f \in C^2$,
2. $\left|\frac{d^2f}{dx^2}\right| < \beta$,
3. $|\frac{df}{dx}(y)| \geq \alpha > 0$ for all $y \in X_c$.
4. $X_c \equiv \{x \mid f(x) = c\}$, and

Prove: that $|X_c| \leq \frac{1}{\frac{2\alpha}{\beta}} = \frac{\beta}{2\alpha}$. Hint: what if $f(x) = \frac{1}{2}\beta x^2$?

1.4.2 The Theory

Now we look a little more deeply at level sets on which the derivative is non-zero. We begin with three definitions.

Definition 1.1 (**Level Sets**). *A level set of* $f : E \subset X \to Y$ *is any set of the form* $X_c \equiv \{x \mid f(x) = c \in Y\}$. *The set* X_c *is sometimes called the c-level set of* f *and is also denoted by* $f^{-1}(c)$, *the inverse image, under* f, *of the point* $c \in Y$

Definition 1.2 (**Regular Level Sets for functions** $f : E \subset \mathbb{R} \to \mathbb{R}$). *A level set of a function* $f : \mathbb{R} \to \mathbb{R}$, $X_c \equiv \{x \mid f(x) = c\}$, *is called a* **regular level set** *if, for every* $y \in X_c$ *there exists an open ball interval* $(y - \delta_y, y + \delta_y)$ *with* $\delta_y > 0$ *such that* $(y - \delta_y, y + \delta_y) \cap X_c = \{y\}$.

Definition 1.3 (**Regular values:** $f : E \subset \mathbb{R} \to \mathbb{R}$). *Suppose that* $f : E \subset \mathbb{R} \to \mathbb{R}$ *and that* $X_c \equiv \{x \mid f(x) = c\}$. *If every derivative on the level set is non-zero: I.e.* $y \in X_c \Rightarrow |\frac{df}{dx}(y)| \neq 0$, *we say that* c **is a regular value of** f

You have now seen, in the exercises, that if c is a regular value, the X_c is a regular level set. That is, you know that:

Theorem 1.1 (**Regular Level Sets**). *Level sets defined by regular values are regular.*

The question that you might have is how much does this generalize? Is this true in higher dimensions? The answer is that this is true much more generally. In the next section, I outline the entire course and how this question and similar ones are actually central to what we will explore and learn.

Now we give the generalizations to the case in which the spaces X and Y in Definition (1.1) are given by $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.

Definition 1.4 (**Regular Level Sets for functions** $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$). *Define* $k \equiv \max(n - m, 0)$. *A level set of a function* $f : \mathbb{R}^n \to \mathbb{R}^m$, $X_c \equiv \{x \mid f(x) = c\}$, *is called a* **regular level set** *if, for every* $y \in X_c$ *there exists an open ball* $B(y, \epsilon)$, *centered at* y *with radius* ϵ, *such that* $B(y, \epsilon) \cap X_c$ *is well approximated by* $B(y, \epsilon) \cap \{y + V_y\} \cap E$ *where* V_y *is a* k*-dimensional subspace of* R^n. *(Well approximated means that there is a smooth change of coordinates, converging to the identity map as* $\epsilon \to 0$, *mapping these two sets bijectively onto each other.)*

Definition 1.5 (**Regular values:** $f : E \subset \mathbb{R} \to \mathbb{R}$). *Suppose that* $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$ *and that* $X_c \equiv \{x \mid f(x) = c\}$. *If every derivative on the level set is full rank: I.e.* $y \in X_c \Rightarrow rank(D_y f) = \min(m, n)$, *we say that* c **is a regular value of** f. *In that case, for all* $y \in X_c$, *the* V_y *in definition (1.4) equals* $D_y f(0)^{-1}$

Exercise 1.9. See if you can show that Definitions (1.2) and (1.3) are special cases of Definitions (1.4) and (1.5)

1.5 Pause for an overview of where we are going

We now have enough under our belt to motivate a big picture of the course and what we will explore. We begin with a very simple smooth function $f : \mathbb{R} \to \mathbb{R}$:

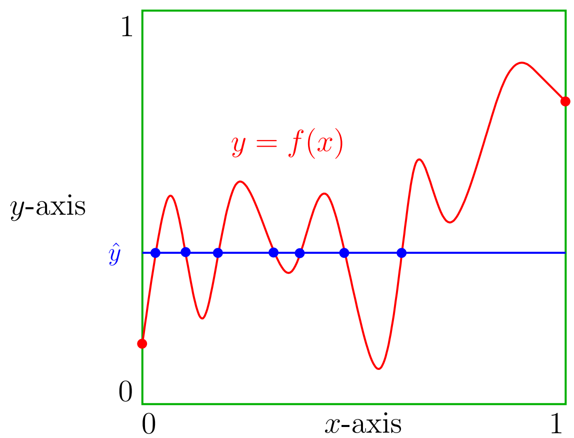


Figure 1: The level set $X_{\hat{y}}$ has 7 elements, shown as 7 blue dots in this figure. Same figure can be used to illustrate each of the three integrals.

Three integrals, their generalizations and the wild and woolly intersection of differentiation and integration:

Index Theory

$$\int_0^1 \sum_{x \in X_y} \text{sign}\left(\frac{df}{dx}(x)\right)\, dy \quad = \quad \text{oriented length of f([0,1]) with cancellation}$$

→ special case of index Theory

→ will bring up Sard's Theorem for us

Area/Coarea

$$\int_0^1 \left|\frac{df}{dx}(y)\right|\, dx \quad = \quad \text{length of f([0,1]) with multiplicities}$$

→ special case of area and coarea formulas

Stokes Theorem

$$\int_0^1 \frac{df}{dx}(y)\, dx \quad = \quad f(1) - f(0) = \text{ oriented length of f([0,1]) with cancellation}$$

→ simple case of divergence theorem

→ which is itself a simple case of Stokes Theorem

The first integral gets us thinking about regular values and regular level sets which leads to a bunch of cool stuff:

Regular Values of Mappings $\mathbb{R}^n \to \mathbb{R}^m$

$\text{rank}(D_y f) = \max(n-m, 0) \;\; \forall \;\; y \in X_c$
$\to$ Sard's Theorem also comes up
$\to$ Which brings up the 5R covering theorem
$\to$ Which becomes a good place to begin looking at outer measures

Regular Level sets $\mathbb{R}^n \to \mathbb{R}^m$

$(B(y,\epsilon) \cap \{y + V_y\} \cap E) \sim (B(y,\epsilon) \cap X_c) \;\; \forall \;\; y \in X_c$
$\to$ Really the same idea as Derivative = linear approximation
$\to$ Introduces Manifolds

Regular Value implies Regular level set $\mathbb{R}^n \to \mathbb{R}^m$

$\big(B(y,\epsilon) \cap \{y + D_y f^{-1}(0)\} \cap E\big) \sim (B(y,\epsilon) \cap X_c) \;\; \forall \;\; y \in X_c$
$\to$ Level sets corresponding to Regular values = manifolds

The second integral formula introduces the area and coarea formulas. These generalize to rather wild functions and sets. The third is a special (and very simple) case of Stokes Theorem.

Area/Coarea Formulas: $f : \mathbb{R}^n \to \mathbb{R}^m$

$$\int_\Omega g(x) J^* f dx = \int_{f(\Omega)} \left(\int_{f^{-1}(w)} g(x) d\mathcal{H}^{\max(n-m,0)}(x) \right) d\mathcal{H}^m(w)$$

$\to$ a very powerful general tool for tracking and computing mapped volumes
$\to$ We encounter Outer Measures and Hausdorff Measures in earnest here!

Stokes Theorem – Briefly

$$\int_{\partial\Omega} \omega = \int_\Omega d\omega \text{ (Stokes Theorem)}$$

$$\to \int_{\partial\Omega} v \cdot \vec{n} \, d\sigma = \int_\Omega \nabla \cdot v \, dx \text{ (Divergence Theorem)}$$

$$\to \oint_{\partial\Omega} \vec{v} \cdot T_{\partial\Omega} = \int_\Omega \nabla \times \vec{v} \, dx \text{ (Little Stokes Theorem)}$$

We will also encounter and use tools that we focus on because of their importance. We have already spent some time with derivatives. We are about to look at Taylor series which are a higher order generalization of the Mean Value Theorem.

Working with Inequalities Cauchy-Schwartz's, Hölder's, Jensen's, AM-GM, Chebeshev, etc.

Understanding Linear Maps and Subspaces The Singular Value Decomposition (SVD): $A = USV^t$ where U and V are orthogonal and S is diagonal with diagonal elements non-negative ordered from greatest to least as we move down the diagonal, QR decomposition $A = QR$ where Q is orthogonal and R is an upper triangular ... this is just Gram-Schmidt, $Ax = \lambda x$ eigenvector/eigenvalue, Hermitian/selfadjoint matrices, linear subspaces, orthogonal matrices and projections, etc

Derivatives = Linear Approximations: As long as $F : X \to Y$, where X and Y are complete normed Linear spaces, we can hope that $F(x + h) - F(x) = A(h) + o(h)$ is true for some linear operator A that depends on x, which we then call $D_x F$ ... i.e. that F is differentiable at x. Infinite dimensions are not a problem!

Mean Value Theorem (and Taylor Series) If f is differentiable in (a, b), then there is always a $c \in (a, b)$ such that $f(b) - f(a) = f'(c)(b - a)$. We can turn this into a route to get Taylor series which are really higher order versions of the mean value theorem.

Measure Theory Facts: the essentials Outer measures, Radon measures, Hausdorff measures, 5R covering theorem, approximation theorems, etc

Misc lots of interesting little detours and wild functions and spaces to explore ...

1.6 Taylor Series

There are a three approaches to proving different version of Taylor series approximations. Two use the mean value theorem and the third, the definition of derivative.

1.6.1 Mean Value Theorem Approach I

We first use the mean value theorem in a very straightforward way to get Taylor Series approximations to a function. In this approach we assume that $f \in C^{n+1}$ and conclude that

$$f(x + h) - \sum_{k=0}^{n} f^k(x) \frac{h^k}{k!} = f^{n+1}(c(h)) \, \frac{h^{n+1}}{(n + 1)!}$$

for some $c(h)$ between x and $x+h$.

We begin by demonstrating how it goes when $n = 1$.

1. Begin with the Mean Value Theorem:

$$g(x+h) - g(x) = g'(c)h \text{ (for some } c \in (x, x+h) \text{ or } (x+h, x))$$

(We will assume that $h > 0$ and note that everything works when $h < 0$ too. But you should convince yourself this is true!)

2. Apply this to $g(x) = f'(x)$ and assume $f \in C^2$ and conclude that

$$f'(x+h) - f'(x) = f''(c(h))h \text{ (for some } c \in (x, x+h))$$

3. Integrate this to get:

$$\int_0^h f'(x+t)\,dt - \int_0^h f'(x)\,dt = \int_0^h f''(c(t))t\,dt \qquad (10)$$

$$\rightarrow f(x+h) - f(x) - f'(x)h = \int_0^h f''(c(t))t\,dt \qquad (11)$$

4. Define

$$f''_m = \min_{s\in[x,x+h]} f''(s)$$

$$f''_M = \max_{s\in[x,x+h]} f''(s)$$

5. Because $x < c(t) < x+t \leq x+h$, we get that

$$\int_0^h f''_m\, t\,dt \leq \int_0^h f''(c(t))\, t\,dt \leq \int_0^h f''_M\, t\,dt$$

$$f''_m \frac{h^2}{2} \leq \int_0^h f''(c(t))t\,dt \leq f''_M \frac{h^2}{2}$$

6. Now define I_h^f by

$$I_h^f\, \frac{h^2}{2} \equiv \int_0^h f''(c(t))\, t\,dt$$

to get

$$f''_m\, \frac{h^2}{2} \leq I_h^f\, \frac{h^2}{2} \leq f''_M \frac{h^2}{2}$$

which implies

$$f''_m \leq I_h^f \leq f''_M.$$

7. Because f'' is continuous on $[x, x+h]$,the intermediate value theorem tells us there is a point $\hat{c} \in [x, x+h]$ such that $f''(\hat{c}) = I_h^f$.

8. We immediately have that Equation (11) can be rewritten:

$$\begin{aligned} f(x+h) - f(x) - f'(x)h &= f''(\hat{c})\,\frac{h^2}{2} \\ &= f''(\hat{c}(h))\,\frac{h^2}{2} \end{aligned}$$

9. In general, we have that

$$f(x+h) - \sum_{k=0}^{n} f^k(x)\frac{h^k}{k!} = f^{n+1}(c(h))\,\frac{h^{n+1}}{(n+1)!}$$

and the proof is completely analogous except that in this case we assume that $f \in C^{n+1}$ and begin with

$$g(x+h) - g(x) = g'(c)h$$

which we apply to $g(x) = f^n(x)$ to get

$$f^n(x+h) - f^n(x) = f^{n+1}(c(h))h \text{ (for some } c \in (x, x+h))$$

which lets us conclude following our steps, exactly, that

$$f^{n-1}(x+h) - f^{n-1}(x) - f^n(x)h = f^{n+1}(\hat{c}(h))\,\frac{h^2}{2}$$

which, in turn, leads by completely analogous steps to

$$f^{n-2}(x+h) - f^{n-2} - f^{n-1}(x)h - f^n(x)\frac{h^2}{2} = f^{n+1}(\hat{\hat{c}}(h))\,\frac{h^3}{3!}$$

10. We can continue this to the desired conclusion, though we usually just let $c(h)$ represent the function, mapping into the interval $[x, x+h]$, that changes from iteration to iteration.

1.6.2 Mean Value Theorem Approach II

There is a shorter proof that I may eventually stick in these notes, but it comes straight from page 386 of Fleming's book. It assumes slightly less: we assume only that the function has an (n+1)'th derivative everywhere in the interval $[x, x+h]$, not that it is continuous. Go ahead and read that proof there.

Exercise 1.10. Write out the Taylor series centered at $x = 0$ for each of these functions:

1. $\sin(x)$
2. $\cos(x)$
3. $\tan(x)$
4. $\arcsin(x)$
5. $\arccos(x)$
6. $\arctan(x)$
7. $\ln(x)$
8. e^x
9. e^{-x^2}

Exercise 1.11. How far out in the series for e^{-100} does one have to go to be guaranteed to be within 10^{-6} of the correct answer? That is, what N makes $\sum_{i=0}^{N} \frac{(-100)^i}{i!}$ differ from e^{-100} by no more than $\frac{1}{1{,}000{,}000}$?

Exercise 1.12. Given the differential equation $y'' - y' + y = 0$, and $y = \sum_{i=0}^{\infty} a_i x^i$, find the a_i's and then find the solutions in terms of functions studied in Exercise 1.10. Confirm these are solutions by direct differentiation and substitution into the differential equations.

1.6.3 Derivative Definition Approach

Define
$$T_f^{a,k}(x) \equiv \sum_{i=0}^{k} f^i(a) \frac{(x-a)^k}{k!}$$
where $f^j = \{\text{the } j\text{th derivative of } f\}$ and $f^0 \equiv f$.

In this subsection, we discuss that cool fact that $|f(x) - T_f^{a,k}(x)| = o(|x-a|^k)$ even if the only thing we know is that $f^i(x)$ exists at $x = a$ for $i = 1, 2, ..., k$. This is a generalization to higher orders of the statement that if f is differentiable at a, then $f(x) - (f(a) + f'(a)(x-a)) = o(|x-a|)$ where we only need that f' exists at a, in order for the approximation to be true. Of course we get existence in a neighborhood of a for lower order derivatives from the existence of higher order derivatives at a. The source for this theorem is Kennan Smith's interesting *A Primer in Analysis*. (Every analyst should have a copy.)

Theorem 1.2. *If $f^i(a)$ exists for $i = 1, 2, ..., k$, then $|f(x) - T_f^{a,k}(x)| = o(|x-a|^k)$ for some interval $|x - a| \leq \delta$.*

Proof of Theorem 1.2.
Suppose that $f^i(a)$ exists for $i = 1, 2, ..., k$. We note that:

1. $(T_f^{a,k})' = T_{f'}^{a,k-1}$.

2. if $k \geq 2$, $f^k(a)$ existing, implies that f^i exists in a neighborhood of $x = a$ for $i = 1, 2, ..., k-1$ and f^i is continuous for in a neighborhood of $x = a$ for $i = 1, 2, ..., k-2$ and f^i. In particular, if $k \geq 3$, then $f(x) - f(a) = \int_a^x f^1(t)dt$.

3. Now a lemma that we will use more than once in the proof and is generally useful in other circumstances:

   Lemma 1.1. *if $f(x) = o(x^k)$ then $\int_0^x f(y)dy = o(x^{k+1})$.*

   Proof of Lemma 1.1.
   Since $f(x) = o(x^k)$, $f(x) = h(x)x^k$, where $h(x) \underset{x\to 0}{\to} 0$. Define $h^+(x) = \sup_{t\in[-x,x]} |h(t)|$. Note that $h^+(x) \underset{x\to 0}{\to} 0$ and $|h^+(x)| \geq |h(x)|$ for all x. Notice that $|\int_0^x h(t)t^k dt| \leq h^+(x) \int_0^x t^k dt = \frac{h^+(x)}{k}|x|^{k+1}$ □

4. using the previous items, if $k \geq 3$, then if $|f'(x) - T_{f'}^{a,k-1}| = o(|x-a|^{k-1})$, we conclude that $\left|\int_a^x \left(f'(t) - T_{f'}^{a,k-1}(t)\right) dt\right| = \left|f(x) - T_f^{a,k}(x)\right| = o(|x|^k)$. So the theorem is true for k if it is true for $k-1$.

5. We note that the case of $k = 1$ is just the definition of derivative. We need only prove the theorem for the case $k = 2$. Because, in the case that $k = 2$, we cannot directly assume that $f(x) - f(a) = \int_a^x f^1(t)\, dt$ $(= \int_a^x f'(t)\, dt)$, we have to put a bit more work into this case.

   (a) As noted above, because $f^2(a)$ exists, $f^1(x) = f'(x)$ exists in some neighborhood of a and we have that $f'(x) - f'(a) - f''(a)(x-a) = h(|x-a|)(x-a)$, where $h(|x-a|) \to 0$ as $x \to a$.

   (b) Suppose that $g'(y)$ exists for all $y \in [a, x]$. Choose $\epsilon > 0$ and note that for each point $y \in [a, x]$, there is a ball $B(y, \delta_y)$ such that $g(z) - g(y) = K(z)(z-y)$ and $g'(y) - \epsilon \leq K(z) \leq g'(y) + \epsilon$. Because $[a, x]$ is compact there are a finite number of these balls (intervals!) that cover $[a, x]$. We can choose y_i such that $a = y_1 < y_2 < \cdots < y_N = x$ and $g(y_{i+1}) - g(y_i) = K_i(y_{i+1} - y_i)$ where $g'(y_i) - \epsilon \leq K_i \leq g'(y_i) + \epsilon$.

(c) Apply the previous step to $g = f - f'(a)(x-a) - \frac{f''(a)}{2}(x-a)^2$. We get that $|g(x) - g(a)| = |\sum_i K_i(y_{i+1} - y_i)| \leq \sum_i |K_i|(y_{i+1} - y_i)$ and since that sum is dominated by $\int_a^x h(|t-a|)(t-a) + \epsilon \, dt$ and ϵ was arbitrary, we are done after a use of the above lemma.

□

Exercise 1.13. Work through the details of step 5 above.

Exercise 1.14. Give an example of a function that is differentiable at $x = 0$ but differentiable anywhere else.

Exercise 1.15. (**Hard**) Find an example of a function $f : [0,1] \to \mathbb{R}_1$ that is both differentiable everywhere and Lipschitz, such that derivative is not continuous on a set with positive measure. (I tried proving this was not possible. That was very hard, for a good reason – it is possible!)

1.7 Index Theory and Sard's Theorem

Recall the simple 1-dimensional example of index theory in section 1.5:

$$\int_0^1 \sum_{x \in X_y} \text{sign}(\frac{df}{dx}(x)) \, dy \quad = \quad \text{oriented length of f([0,1]) with cancellation}$$

$\rightarrow$ special case of index Theory

$\rightarrow$ will bring up Sard's Theorem for us

In this section, we prove that the complement of the set of regular values has measure 0. Using the ideas we developed in section 1.4 allows us to conclude that for almost all $y \in [0,1]$, the level sets X_y = a finite set of points. Because integrals ignore sets of measure zero, we know that this means the above integral is well-defined.

Exercise 1.16. As a follow on to the exercises in section 1.4, show that at regular values y, the sum

$$\sum_{x \in X_y} \text{sign}(\frac{df}{dx}(x))$$

is either -1, 0 or 1.

Exercise 1.17. Use the results of the last exercise to conclude that

$$\begin{aligned}\int_0^1 \sum_{x\in X_y} \text{sign}(\frac{df}{dx}(x))\, dy &= f(1) - f(0)\\ &= \text{the oriented length of the image of } f([0,1])\end{aligned}$$

Theorem 1.3 (Sard's Theorem in $\mathbb{R}^1$)**.** *Suppose that* $f : [0,1] \to [0,1]$ *and* $f'(x)$ *exists for all* $x \in [0,1]$*. Define*

$$D_0 \subset [0,1] \equiv \{x \in [0,1] | f'(x) = 0\}.$$

Then,

$$\mathcal{H}^1(f(D_0)) = 0.$$

That is, the length of the complement of the set of regular values has length zero.

There are two ways we are going to prove this.

1.7.1 A special case of Sard's Theorem via the 5R covering theorem

First proof of Theorem 1.3:

Proof.
Because f is differentiable, for any $\epsilon > 0$, we can do the following:

1. Use the cone result (see Problem 5.1) to choose a small enough $\delta_x^\epsilon > 0$ for every $x \in D_0$, such that

$$|f(x) - f(y)| \leq \epsilon|x - y| \text{ when } y \in \hat{U}_x \equiv (x - 5\delta_x^\epsilon, x + 5\delta_x^\epsilon)$$

2. This last step tells us that that f maps the $\hat{U}_x$ whose lengths are $10\delta_x^\epsilon$, into (not necessarily onto!) intervals that are no longer than $\epsilon 10\delta_x^\epsilon$.

3. Now define $U_x = (x - \delta_x^\epsilon, x + \delta_x^\epsilon)$. Notice that $D_O \subset \cup_x U_x$.

4. Now use the 5R theorem (below) to get a countable disjoint sub-collection of the U_x's, $\{U_{x_i}\}_{i=1}^\infty$ such that

$$D_0 \subset \bigcup_{x\in D_O} U_x \subset \bigcup_{i=1}^{\infty} \hat{U}_{x_i}$$

5. Now we note that because the $\{U_{x_i}\}_{i=1}^{\infty}$ are disjoint,

$$\sum_{i=1}^{\infty} \mathcal{H}^1(U_{x_i}) = \mathcal{H}^1(\bigcup_{i=1}^{\infty} U_{x_i}) \leq \mathcal{H}^1([0,1]) = 1$$

and this implies that

$$\sum_{i=1}^{\infty} \mathcal{H}^1(\hat{U}_{x_i}) \leq 5.$$

6. Now we compute:

$$\begin{aligned} \mathcal{H}^1(f(D_0)) &\leq \sum_{i=1}^{\infty} \mathcal{H}^1(f(\hat{U}_{x_i})) \\ &\leq \sum_{i=1}^{\infty} \epsilon \mathcal{H}^1(\hat{U}_{x_i}) \\ &\leq 5\epsilon \end{aligned}$$

7. Because ϵ was arbitrary, we can conclude that $\mathcal{H}^1(f(D_0)) = 0$.

Now for the 5R theorem.

Theorem 1.4 (5R Covering Theorem)**.** *If E is a ball (open or closed) with center p and radius r, let $\hat{E}$ denote the ball (open or closed) with center p and radius $5r$.*

Suppose $\mathcal{U} = \{U_\beta\}_{\beta \in \mathcal{B}}$ is a (possibly uncountable) collection of balls in $\mathbb{R}^n$ whose radii are bounded above by $C < \infty$. Then there exists a countable subcollection

$$\mathcal{F} = \{U_{\beta_i}\}_{i=1}^{N_\mathcal{B} \leq \infty}$$

such that:

1. *$U_{\beta_i} \cap U_{\beta_j}$ for $i \neq j$ and*

2. *$\{U_\beta\}_{\beta \in \mathcal{B}} \subset \cup_{i=1}^{N_\mathcal{B}} \hat{U}_{\beta_i}$*

Proof.
We break the proof into steps:

1. We partition the balls into subcollections: $\{\mathcal{E}_k\}_{k=1}^{\infty}$, where

$$\mathcal{E}_k = \left\{ U_\beta \,\middle|\, \frac{1}{2^k} C < \text{radius}(U_\beta) \leq \frac{1}{2^{k-1}} C \right\}$$

.

2. Now choose a maximal sets of disjoint balls in E_1: We use Zorn's lemma to get a maximal collection of pairwise disjoint balls in E_1: Zorn's lemma implies that there exists, F_1 a subcollection of balls in E_1 such that (a) every pair of balls $\{U, W\} \in F_1$ are disjoint and (b) if $U \in E_1 \setminus F_1$, then $U \cap W \neq \emptyset$ for some $W \in F_1$.

3. It follows that F_1 is a countable set and can be enumerated, $F_1 = \{U_{\beta_i^1}\}_{i=1}^{N_1 \leq \infty}$

4. It also follows that
$$\bigcup_{U \in E_1} U \subset \bigcup_{U \in F_1} \hat{U}.$$

5. Now construct F_i from E_i by (a) first getting rid of all the balls in E_i that intersect any ball in $\bigcup_{k=1}^{i-1} F_k$ and then (b) finding a maximal pairwise disjoint collection of the balls in E_i that are left. It follows that:
$$\bigcup_{U \in \cup_{k=1}^{i} E_k} U \subset \bigcup_{U \in \cup_{k=1}^{i} F_k} \hat{U}.$$

6. Define $\mathcal{F} \equiv \cup_{i=1}^{\infty} F_i$.

7. By the above construction $\mathcal{F}$ is a pairwise, countable subcollection of $\mathcal{U}$ whose dilation by 5 creates of collection of balls whose union covers the union of the balls in $\mathcal{U}$.

□

Exercise 1.18. Look up Zorn's lemma and make sure you understand how that lemma gives us the maximal subcollections we use.

□

1.7.2 A special case of Sard's theorem via more smoothness and Compactness

Now we prove Theorem 1.3 with the added assumption that $f \in C^1$ – not only is f differentiable, the derivative f' is continuous as well.

1. Because $f' : [0, 1] \to \mathbb{R}$ is now assumed continuous, we know that $D_0 = (f')^{-1}(0)$ is closed since $\{0\}$ is a closed set. Since it is also bounded, D_O is compact.

2. Now use the cone result (see Problem 5.1) to choose a small enough $\delta_x^\epsilon > 0$ for every $x \in D_0$, such that
$$|f(x) - f(y)| \leq \epsilon|x - y| \text{ when } y \in U_x \equiv (x - \delta_x^\epsilon, x + \delta_x^\epsilon)$$

3. These open intervals $\{U_x\}_{x \in D_0}$ cover D_0 and so there is a finite subcover of D_0, $\{U_{x_1}, ..., U_{x_N}\}$. I.e. we have $D_0 \subset \bigcup_{i=1}^N U_{x_i}$.

4. Without loss of generality, we can assume that $x_1 < x_2 < ... < x_N$.

5. We can assume also that if one of the U_{x_i}'s is removed from $\{U_{x_i}\}_{i=1}^N$, the $N-1$ open intervals that remain do not cover D_0.

6. We define l_i and r_i by $U_{x_i} = (l_i, r_i) = (x_i - \delta_{x_i}^\epsilon, x_i + \delta_{x_i}^\epsilon)$.

7. Because we assume none of the intervals can be left out of the cover, we can conclude that $l_1 < l_2 < ... < l_N$ and $r_1 < r_2 < ... < r_N$.

8. Because $l_{i+2} < r_i$ would imply that the $U_{x_{i+1}}$ is covered by $U_{x_i} \cup U_{x_{i+2}}$, we can conclude that every point in $\cup_{i=1}^N U_{x_i}$ is in at most two of the U_{x_i}'s, implying that:
$$\sum_{i=1}^N \mathcal{H}^1(U_{x_i}) \leq 2\mathcal{H}^1(\cup_{i=1}^N U_{x_i}) \leq 2$$
since $\bigcup_{i=1}^N U_{x_i} \subset [0,1]$.

9. Now, as before (except with a 2 instead of a 5), we have
$$\begin{aligned} \mathcal{H}^1(f(D_0)) &\leq \sum_{i=1}^N \mathcal{H}^1(f(U_{x_i})) \\ &\leq \sum_{i=1}^N \epsilon \mathcal{H}^1(U_{x_i}) \\ &\leq 2\epsilon \end{aligned}$$

10. Because ϵ was arbitrary, we can conclude that $\mathcal{H}^1(f(D_0)) = 0$.

Exercise 1.19. Convince yourself that the steps (4-8) above are justified. You should sketch the situation. See Figure (2).

1.7.3 Another Exercise

Exercise 1.20. Show that the conclusion of Exercises (1.16-1.17) need not be correct if f is discontinuous, even if f is differentiable at every point except the points where it is discontinuous and there are only a finite number of discontinuities. Show this by showing, for any $\alpha \in \mathbb{R}$, how to construct a function $f_\alpha : [0,1] \to [0,1]$ for which
$$\alpha = \int_0^1 \sum_{x \in X_y} \text{sign}(\frac{df_\alpha}{dx}(x)) \, dy$$

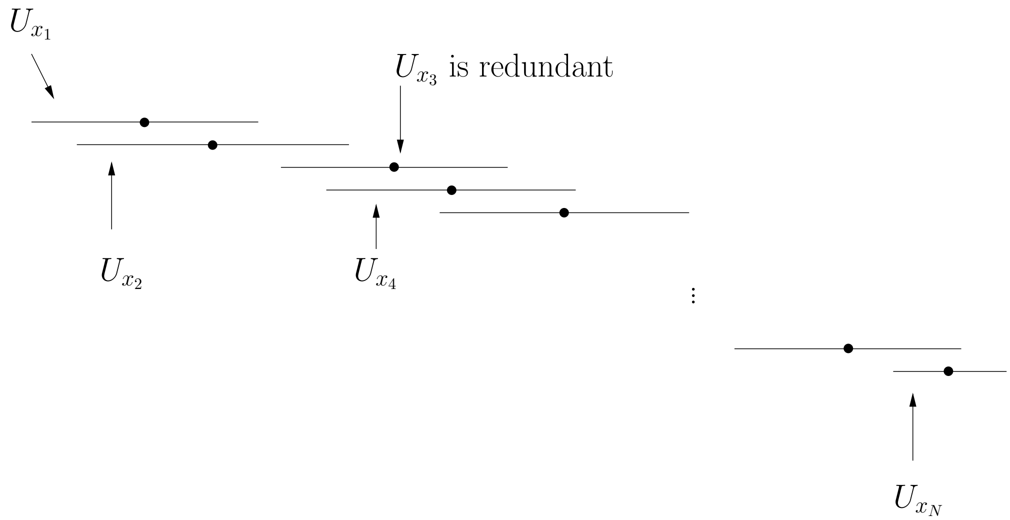


Figure 2: Example sketch to get you thinking. Remember that the intervals are symmetric about the x_i's shown as dots here.

1.8 Norms of Operators

Definition 1.6 (Operator Norm)**.** *Suppose that* $A : x \in B_1 \to y \in B_2$ *where* B_1 *and* B_2 *are linear spaces with norms* $|\cdot|_1$ *and* $|\cdot|_2$*, and* A *is a linear operator. We define the norm of the operator* A *to be:*

$$|A| \equiv \sup_{x \in B(0,1)} |A(x)|_2,$$

or equivalently

$$|A| \equiv \sup_{x \in \partial B(0,1)} |A(x)|_2,$$

or equivalently

$$|A| \equiv \sup_{x \in B_1 \setminus \{0\})} \frac{|A(x)|_2}{|x|_1},$$

where $B(0,1)$ *is the unit ball, centered in the origin in* B_1*, so* $\partial B(0,1)$ *is the boundary of the unit ball, the unit sphere centered on the origin.*

1.9 Introduction to using Derivative Approximations

We now use the fact that the derivative approximates the function locally to (1) get local invertibility and (2) the local parameterization of level sets. We first look at the simplest possible cases to illustrate the ideas.

1.9.1 Inverse Function Theorem: $f : \mathbb{R} \to \mathbb{R}$

We begin with an example of a function $f : \mathbb{R} \to \mathbb{R}$:

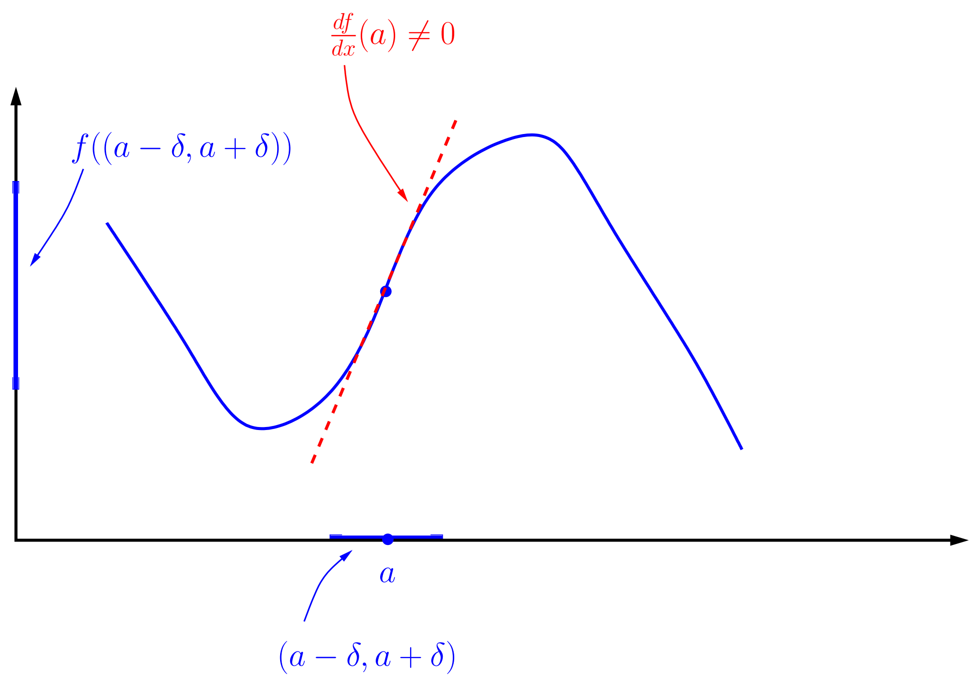


Figure 3: **Inverse Function Theorem - A simple, one dimensional example**: if the derivative of f is invertible at a, then, in a small enough neighborhood of a, $(a-\delta, a+\delta)$, the function itself is invertible. **Technical Details**: We need to assume that not only is the derivative at a invertible – in this case that means the slope = {1-by-1 matrix} is nonzero – we also need the derivative function mapping points in the domain to their slopes to be continuous at a.

Remark 1.2. *We need the requirement that the derivative is continuous since it is not too hard to come up with examples of functions that are differentiable at a point, but not invertible in any neighborhood of that point. See Figure 4.*

1.9.2 Implicit Function Theorem: $f : \mathbb{R}^2 \to \mathbb{R}$

The simplest example of the implicit function theorem is provided by a function from $\mathbb{R}^2$ to $\mathbb{R}$. The assumptions are that the derivative of f is full rank, which in this case, means that at least one of the partial derivatives is non-zero. See Figure 1.9.2.

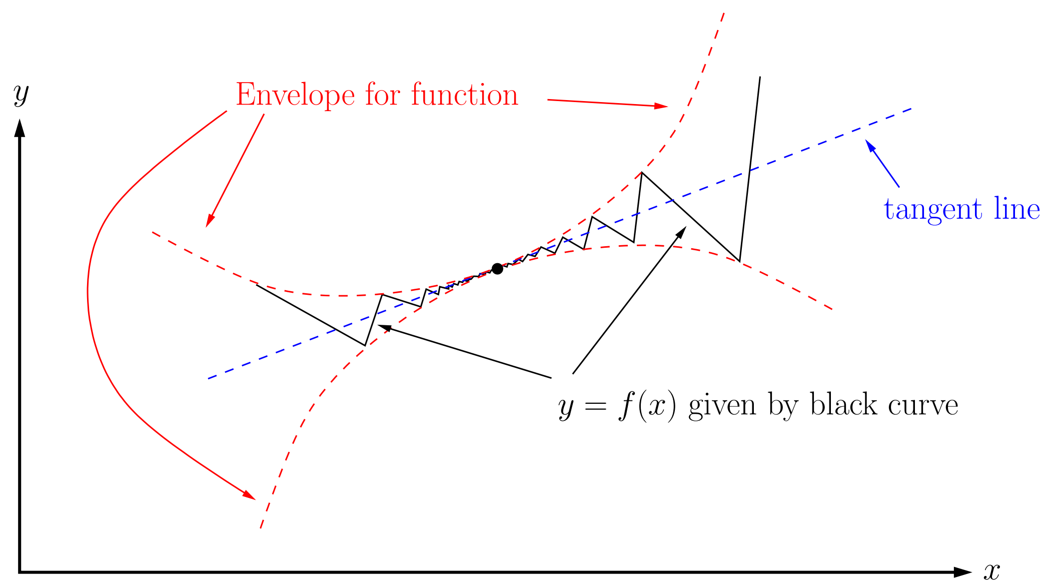


Figure 4: Example of a function whose derivative at a point is invertible but the function is not invertible in any neighborhood of that point, because the derivative is not continuous.

Remark 1.3. *Suppose for example, that $f_x(a) \neq 0$. Then locally we can change the value of the function by changing the value of x: if $f(x^*, y^*) = c$ and we perturb y, from y^* to $y^* + \epsilon$, we will generally find that $f(x^*, y^* + \epsilon) = c + \delta$ but because $f_x \neq 0$, we can just find an $\eta(\epsilon)$ such that $f(x^* + \eta(\epsilon), y^* + \epsilon) = c$. $\eta(\epsilon)$ will be approximately given by $f_x(a)\eta(\epsilon) \approx -\delta$ or $\eta(\epsilon) \approx \frac{-\delta}{f_x(a)}$*

1.10 Inverse and Implicit Function Theorems

In addition to the full versions of the Inverse and Implicit Function Theorems, we give an intuitive overview of manifolds which are central to nonlinear analysis.

1.10.1 Review: $\mathbb{R}^n$ and why we like it.

We are all acquainted with $\mathbb{R}^2$ and $\mathbb{R}^3$. Many of us have worked extensively with $\mathbb{R}^n$, usually by analogy with $\mathbb{R}^2$ and $\mathbb{R}^3$. Here are some familiar properties and things we can do using those properties:

Vector Space: $\mathbb{R}^n$ is a vector space with elements of the form $\mathbf{x} = (x_1, x_2, ..., x_n)$

Inner product: The inner product of $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \cdot \mathbf{y}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$, is given by $\sum_{i=1}^{n} x_i y_i$.

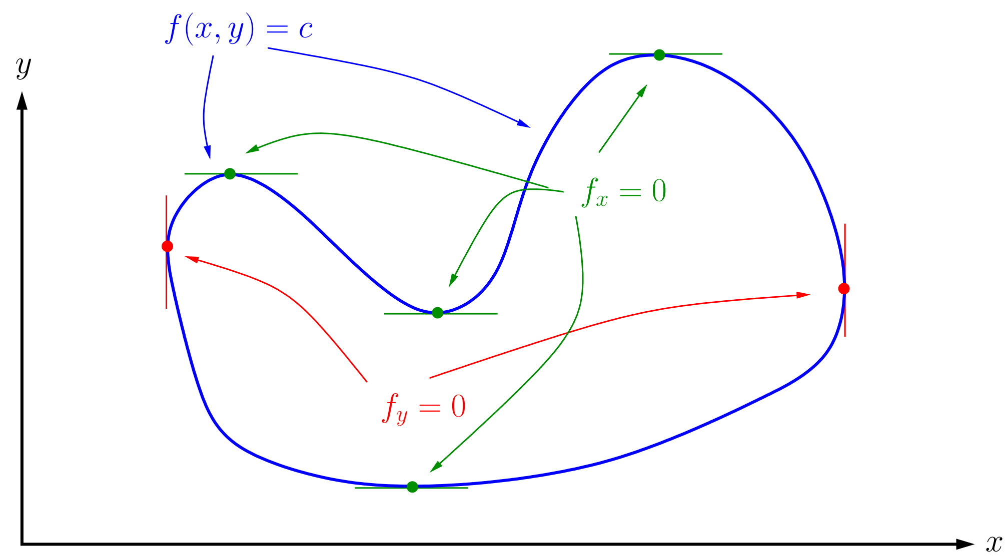


Figure 5: **Implicit Function Theorem - simple example**: if the derivative of f is full rank at some point $a = (x^*, y^*)$ in the $f = c$ level set, then, in a small enough neighborhood of a, then at least one of these (non-exclusive) cases holds: **(Case 1:)** There is a function of y, $g(y)$, and a $\delta > 0$ such that for $y \in (y^* - \delta, y^* + \delta)$ $f(g(y), y) = c$. (This is true if $f_x(a) \neq 0$) **(Case 2:)** There is a function of x, $h(x)$, and a $\delta > 0$ such that for $y \in (y^* - \delta, y^* + \delta)$ $f(x, h(x)) = c$. (This is true if $f_y(a) \neq 0$) **Technical Details**: While the theorem only needs the derivative to be full rank at a, if the derivative of f is full rank on the entire level set, this means that we have local coordinates everywhere, though sometimes only x or only y will work as local coordinates. The derivative in our case is $\nabla f = (f_x, f_y)$ and being full rank means there is at least one nonzero element of this gradient vector. We are also assuming that the derivative is continuous, as we did in the inverse function theorem case, because, in fact we use the inverse function theorem to prove this theorem.

Euclidean distance: The length of a vector $\mathbf{x}$ is given by $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x \cdot x}$, so the distance between two points is simply $||\mathbf{x} - \mathbf{y}||_2$.

Angles between vectors: Angles between vectors are given by $\cos(\theta) = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||}$.

Linear Transformations: A Linear transformation between $\mathbb{R}^n$ and $\mathbb{R}^m$, which is most often represented and computed using matrices $A \in \mathbb{R}^{m \times n}$, makes sense because the $\mathbb{R}^k$ is a linear space for all k.

Calculus: Differentiation also makes sense because of the linear space structure of $\mathbb{R}^n$. We also use the metric structure to define volumes and integration.

All this makes life in $\mathbb{R}^n$ beautiful. Calculations are easy, shortest distances between points are straight lines, and our experience with 2 and 3 dimensions, which $\mathbb{R}^n$ is meant to mimic and extend, makes it all very accessible, intuitively speaking.

But the subsets of $\mathbb{R}^n$ we work with are often curved and contorted. k-dimensional surfaces are everywhere, from graphs of functions to parameterized sets in $\mathbb{R}^n$, from level sets of mappings to sets in $\mathbb{R}^n$ that contain all possible samples of some data set we are trying to model. On top of that, there are spaces of points that we find natural to use and possess $\mathbb{R}^k$-like properties, yet are not subsets of any $\mathbb{R}^n$

The structure that comes to our rescue is the k-manifold.

1.10.2 k-Manifolds in $\mathbb{R}^n$ are locally like $\mathbb{R}^k$

Definition 1.2 **(k-manifold in $\mathbb{R}^n$).** *Define L_k to be the k-dimensional subspace of $\mathbb{R}^n$ defined by holding the last $n-k$ coordinates equal to 0, i.e. all points in $\mathbb{R}^n$ of the form $(x_1, x_2, ..., x_k, 0, ..., 0)$. A k-dimensional manifold M_k is a subset that is locally like $\mathbb{R}^k$. At every point $x \in M_k$, there is*

1. *a neighborhood $U \subset \mathbb{R}^n$ containing x and*
2. *a diffeomorphism $\phi_x : U \to W \subset \mathbb{R}^n$*

such that

1. *W is a neighborhood of 0 in $\mathbb{R}^n$,*
2. *$\phi_x(x) = 0$*
3. *$\phi_x(U \cap M_k) = W \cap L_k$*

This definition is far from as general as possible, but for our purposes it will work quite well. In fact, one can take this definition a long ways, and understanding it thoroughly equips one to work with the other more general definitions out there.

The idea is that we will want to use the ϕ's to enable ourselves to do calculus on the manifold. Care must be taken, but everything works out pretty much as one would expect. One tool that is used over and over is the use of local approximations to the manifolds and mappings between manifolds. The first is called the *tangent space at x*, the second is DF_x, the derivative or differential of F at x.

The tangent space of M_k at x is the k-plane T_x that is tangent to M_k at x. As we zoom into M_k at x, it looks more and more like T_x: this is really just a higher dimensional analog of the tangent line you are acquainted with from the idea of derivatives in Calculus 1. To be a bit more precise,

Definition 1.3 (**Tangent Space at x**). *If M_k is a k-manifold, then T_k is the unique k-dimensional subspace of $\mathbb{R}^n$ such that for every $\epsilon > 0$ there is an r_ϵ such that for every point $y \in M_k \cap B(x, r_\epsilon)$*

$$||P_{T_x}(y - x)|| \geq (1 - \epsilon)||y - x||$$

where $P_{T_x}(u)$ is the orthogonal projection of u onto T_x.

This definition says that given any ϵ and a sufficiently small ball around x, the piece of the manifold inside that ball, $M_k \cap B(x, r_\epsilon)$, lives in a cone about T_x whose apical half angle is $\cos^{-1}(1 - \epsilon)$. Thus, by making ϵ sufficiently small, the tangent plane approximates M_k as well, provided we zoom in far enough.

In the next section, we review derivatives as approximations to mappings.

1.10.3 Reminder: Derivatives as linear approximations

Ordinarily, one thinks of derivatives as slopes of tangent lines or even the limit of the ratio $\frac{f(x+h)-f(x)}{h}$ as $h \to 0$. While this is correct for maps from $\mathbb{R}$ to $\mathbb{R}$, another equivalent definition turns out to be very useful. First we define $o(h)$

Definition 1.4. *We say $f(h) = g(h) + o(h)$ if $\frac{|f(h)-g(h)|}{|h|} \to 0$ as $h \to 0$. $o(h)$ is pronounced "little o of h".*

Now we can define derivatives, approximation style:

Definition 1.5 (**Derivative of a map $F : \mathbb{R}^n \to \mathbb{R}^m$**). *Given $F : \mathbb{R}^n \to \mathbb{R}^m$, we will say that F is differentiable at $x \in \mathbb{R}^n$ if there is a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$F(x + h) - F(x) = A(h) + o(h)$$

We denote this linear operator A by DF_x.

In other words, DF_x is the local, linear approximation of $(\Delta_x F)(h) = F(x + h) - F(x)$, the change or increment of F at x.

If $F(x) = (F_1(x), F_2(x), ..., F_m(x))$ is differentiable, the linear map that gives us this approximation turns out to be the matrix of partial derivatives of F:

$$DF_x = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(x) & \frac{\partial F_1}{\partial x_2}(x) & \dots & \frac{\partial F_1}{\partial x_n}(x) \\ \frac{\partial F_2}{\partial x_1}(x) & \frac{\partial F_2}{\partial x_2}(x) & \dots & \frac{\partial F_2}{\partial x_n}(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(x) & \frac{\partial F_m}{\partial x_2}(x) & \dots & \frac{\partial F_m}{\partial x_n}(x) \end{bmatrix}$$

Example 1.1 ($F : \mathbb{R}^n \to \mathbb{R}$)**.** *In the case of a function mapping $\mathbb{R}^n$ to the real numbers, we get $DF_x = \nabla F|_x$: the derivative of F at x is the gradient of F at x, a row vector made up of the partial derivatives of F.*

Remark 1.4. *The tangent plane of M_k at x can now be expressed quite simply. If ϕ_x is the coordinate map of M_k at x, then $T_x + x = D(\phi_x^{-1})_x(L_k)$, where L_k is defined as in Definition 1.2.*

When F is differentiable, it is natural to ask, "How differentiable?"

Definition 1.6. *If the derivative of F exists and is continuous, then we will say F* **is** *C^1. When that derivative has a derivative that is continuous, it is C^2. Likewise* **when F is k-times continuously differentiable, it is** *C^k.*

1.10.4 Review: Full rank maps

Definition 1.7 (**Full Rank**)**.** *Let A be an $m \times n$ matrix. Then A is full rank if any of the following equivalent conditions are true:*

1. *dimension of the null space of A is $\max(0, n - m)$*
2. *there are $\min(m, n)$ independent columns*
3. *there are $\min(m, n)$ independent rows*

Remark 1.5. *If a matrix is full rank, then a sufficiently small perturbation will not change that fact.*

Definition 1.8 (**Level sets**)**.** *The* **level sets** *of a mapping $F : R^n \to R^m$ are the collection of sets $F^{-1}(y) \subset \mathbb{R}^n$ for all $y \in \mathbb{R}^m$.*

Definition 1.9 (**Full Rank Mapping**)**.** *A mapping $F : R^n \to R^m$ is full rank on a level set $F^{-1}(y)$, if DF_x is full rank for all $x \in F^{-1}(y)$.*

Define $W_y = F^{-1}(y)$. When DF_x is full rank on W_y, properties of the level sets of the derivative at points in W_y translate into properties of the nonlinear set W_y.

Definition 1.10. *When the coordinate diffeomorphisms in the definition of a k-manifold are of C^p, then we say that the manifold is of class C^p.*

Theorem 1.5 (Full Rank Theorem). *Suppose that F is C^p with $p \geq 1$. When DF_x is full rank on $W_y = F^{-1}(y)$, W_y is a C^p, k-manifold in R^n, with $k = \max(0, n-m)$.*

The Inverse and Implicit Functions Theorems (general versions in the next section) are in fact the deeper explanation of this last theorem.

1.10.5 Finally: Inverse and Implicit function theorem in higher dimensions

For smooth maps, the derivative gives us complete local information about the structure of the level sets of F.

Theorem 1.6 (Inverse Function Theorem). *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$, $x \in \mathbb{R}^n$, F is C^k, $k \geq 1$ and DF_x is invertible. Then there is some $\epsilon > 0$ such that $F : B(x, \epsilon) \to F(B(x, \epsilon))$ is invertible and the inverse function $G : F(B(x, \epsilon)) \to B(x, \epsilon)$ is also C^k.*

The basic idea is that when the map is full rank (in this case, the derivative is invertible) the derivative's invertibility, the fact that the derivative approximates the nonlinear function locally, and the fact that being full rank is stable to small perturbations all translate into the nonlinear map being invertible.

Proof.
We outline the proof: Assume without loss of generality (WLOG) that F(0) = 0. (That is, we really are interested in showing that $\Delta F(h) \equiv F(x+h) - F(x)$ is invertible in h if $D_x(\Delta F) = D_x F$ is invertible. So, we simplly replace F with ΔF, noting that $\Delta F(0) = 0$.)

1. Choose $0 < \epsilon < 1/2$
2. Define $G = I - DF_0^{-1} \circ F$.
3. Using the fact that F is C^1 we notice that the norm of DG, $|DG|$, is less than ϵ if we stay in some small neighborhood of the origin $U = B(0, \delta(\epsilon))$: I.e. $\frac{||DG(h)||}{||h||} < \epsilon$ for all $h \in U$.
4. Define $W = B(0, \frac{\delta(\epsilon)}{2})$.
5. Using the mean value theorem in vector spaces, we get that restricted to W, G is a contraction mapping with contraction constant ϵ.

6. Define $H = (I + G + G^2 + G^3 + ...)$. Notice that H is differentiable and $DH = I + DG + DG \circ DG + ...$

7. At this point we note that we differentiate to get linear a power series of linear operators because we want to use this fact: if A is linear, and $|A| < 1$, then

$$\begin{aligned} I &= (I - A)(I + A + A^2 + A^3...) \\ &= (I + A + A^2 + A^3...)(I - A) \end{aligned}$$

and this implies that

$$(I - A)^{-1} = (I + A + A^2 + A^3...).$$

8. Notice that $D(H(I - G)) = DH \circ D(I - G) = I$.

9. Now choose $y \in W$. Integrating, we get:

$$\begin{aligned} H(I - G)(y) &= H(I - G)(y) - H(I - G)(0) \\ &= \int_0^1 \left(D(H(I - G))_{ty}\right)(y)dt \\ &= \int_0^1 I \cdot ydt \\ &= y \end{aligned}$$

so that $H(I - G)|_W = H \circ DF_0^{-1} \circ F = I_W$.

10. defining $\hat{F} = H \circ DF_0^{-1}$, we get that $\hat{F} \circ F = I_W$.

11. Note that on $H(W) \subset 2W = U$.

12. Likewise $D((I-G)H) = D((I-G) \circ DH = I$ implying that $(I-G)H|_W = I_W$ or $DF_0^{-1} \circ F \circ H = I_W$. multiplying the last equation on the left by DF_0 and on the right by DF_0^{-1}, we get that $F \circ \hat{F} = I_W$.

13. The C^k differentiability of $\hat{F}$ follows from the C^k differentiability of F.

□

Theorem 1.7 (Implicit function Theorem). *Suppose that F is C^k, $F : \mathbb{R}^n \to \mathbb{R}^m$, $m < n$, and DF is full rank at $x^* \in \mathbb{R}^n$. We will denote the first m coordinates by x' and the last $n - m$ by x'' so that $x = (x', x'')$. Suppose further, without loss of generality, that the first m columns of DF are independent. Then there is an $\epsilon > 0$ and a C^k mapping $g : \mathbb{R}^{n-m} \to \mathbb{R}^m$ such that $F(g(x''), x'') = F(x^*)$ for all $x'' \in \mathbb{R}^{n-m}$ such that $||x'' - (x^*)''|| < \epsilon$.*

Proof.
The idea of the proof is simple: we augment F to get an invertible transformation and then fiddle with it. Define $\hat{F} : \mathbb{R}^n \to \mathbb{R}^n$ by $\hat{F}(x) = (F(x), x'')$. Now we note that $D\hat{F}_{x^*}$ is invertible so that there is an inverse of $\hat{F}$, $G(y) = (g(y', y''), y'')$. Computing $\hat{F} \circ G(y)$ $(= y)$ we have $\hat{F}(G(y)) = (F(g(y', y''), y''), y'') = (y', y'')$ for all $y = (y', y'')$ in some neighborhood of $(F(x^*), x^{*''})$. Looking at the first component only, we have $F(g(y', y''), y'') = y'$. Fixing $\hat{g}(y'') = g(F(x^*), y'')$, we get that $F(\hat{g}(y''), y'') = F(x^*)$ for all $||y'' - x^{*''}|| < \epsilon$ for some sufficiently small $\epsilon > 0$. □

Example 1.2. *Consider some function f mapping $\mathbb{R}^n$ to $\mathbb{R}$. Then in order to apply the implicit function theorem at some point x^*, we need $Df = \nabla f$ to be full rank at x^*. Since $\min(m, n) = 1$, at least one component of the gradient needs to be non-zero at x^* in order to conclude that locally, the level set through x^* is an $(n-1)$-manifold.*

1.10.6 Implicit Function Theorem, Intuitively, Again

The idea behind the implicit function theorem is that:

Full Rank if I am at some point x^* on the c-level set of $f : \mathbb{R}^{n+k} \to \mathbb{R}^n$ and $D_{x^*}f$ is full rank, I know that, after possibly relabeling the coordinates, the first n columns of the derivative matrix for $D_{x^*}f$ form a non-singular n-by-n submatrix.

Invertible + C^1 means ... If we let $x \in \mathbb{R}^{n+k}$ be represented by $x = (x', x'')$ where $x' \in \mathbb{R}^n$ and $x'' \in \mathbb{R}^k$, we have that

$$D_x f(x^*) = [D_{x'} f(x^*) \;\; D_{x''} f(x^*)]$$

where $D_{x'} f(x^*)$ is an n-by-n matrix and $D_{x''} f(x^*)$ is an n-by-k matrix, and $D_{x'} f(x^*)$ maps the first n variables in $\mathbb{R}^{n+k}$ invertibly onto the range of f: we can get anywhere in the range by putting the correct input into $D_{x'} f(x^*)$. Because f is C^1, we know that the derivative $D_{x'} f(x^* + h)$ is also non-singular for small enough h: $|h| < \epsilon$ for some $\epsilon > 0$.

What is boils down to: if we know that $f(x^*) = f(x^{*'}, x^{*''}) = c$ and we now that $D_{x'} f(x^*)$ is invertible (because Df is full rank at x^*), then we know that

1. $f(x^{*'}, x^{*''}) = c$
2. If we perturb (i.e. change) $x^{*''}$ by a small $\eta'' \in \mathbb{R}^k$ to get $x^{*''} + \eta''$, f will change from c to $c + \delta$ for some small δ.
3. That is: $f(x^{*'}, x^{*''} + \eta'') = c + \delta$.
4. Now, because $D_{x'} f(x^{*'}, x^{*''} + \eta'')$ is non singular, the inverse function theorem says that for any small enough δ in the range, there is a unique small η', such that $f(x^{*'} + \eta', x^{*''} + \eta'') - f(x^{*'}, x^{*''} + \eta'') = -\delta$.

5. Since $\eta^{'}$ depends on $\eta^{''}$, we write $\eta^{'} = g(\eta^{''})$

6. We arrive at

$$\begin{aligned} f(x^{*'} + g(\eta^{''}), x^{*''} + \eta^{''}) &= f(x^{*'} + g(\eta^{''}), x^{*''} + \eta^{''}) - f(x^{*'}, x^{*''} + \eta^{''}) \\ &+ f(x^{*'}, x^{*''} + \eta^{''}) \\ &= -\delta + (c + \delta) \\ &= c \end{aligned}$$

7. Because $g : \mathbb{R}^k \to \mathbb{R}^n$, we see that for some small enough $\epsilon > 0$, the set, $B(x^*, \epsilon) \cap \{x = (x', x'') \mid c = f(x', x'')\}$ is actually the set $B(x^*, \epsilon) \cap \{x = (g(x''), x'') \mid c = f(g(x''), x'')$. But because $x'' \in \mathbb{R}^k$ this implies that the set is k-dimensional. **The crucial fact is that $x'' \to (g(x''), x'')$ is an invertible map.**

Details The smoothness of the function $g : \mathbb{R}^k \to \mathbb{R}^n$ follows from the properties of the inverse function theorem.

1.10.7 Using the Implicit Function Theorem: Co-Dimension 1

How would you use the implicit function theorem? Here is a pseudo-computational explanation, by which I mean that it leans towards computation, but is actually intended to give a deeper idea of what the theorem means and give someone a path to investigate for computational purposes. We explore the co-dimension 1 case.

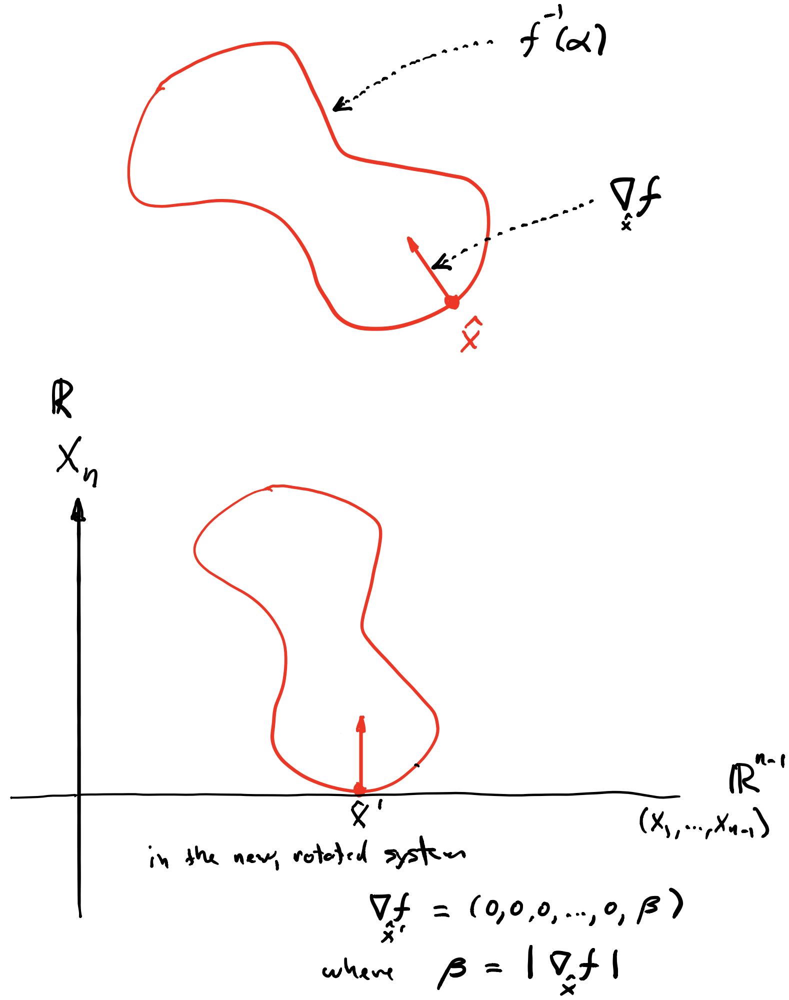
1. So you are given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a level α and a point $\hat{x} \in f^{-1}(\alpha)$.

2. You need to know that $D_{\hat{x}} f = \nabla_{\hat{x}} f$ is full rank, which in the case of 1 by n matrices (i.e. row vectors, which is what a gradient is) we simply want to know that one of the partial derivatives of f at $\hat{x}$ is **not** equal to zero.

3. (Actually, if α is a regular value of f, then we know this since this means, by definition, that the derivative of f is full rank at **every** point in $f^{-1}(\alpha)$. But we actually only need f differentiable at $\hat{x}$)

4. We also need that $f \in C^1$ – the derivative exists and is continuous.

5. Now, rotate the coordinates so that the gradient vector of f points in the direction of the nth coordinate axis. I.e. rotate $\mathbb{R}^n$ so that $\nabla_{\hat{x}} f = \beta(0, 0, ..., 0, 1)$ where $\beta = |\nabla_{\hat{x}} f|$.



6. Now we note that the gradient is normal to the α-level set $f^{-1}(\alpha)$ which is

the same statement as "the planes defined by the gradient vector are tangent to the level surface". This tangent plane is determined by the gradient vector as follows.

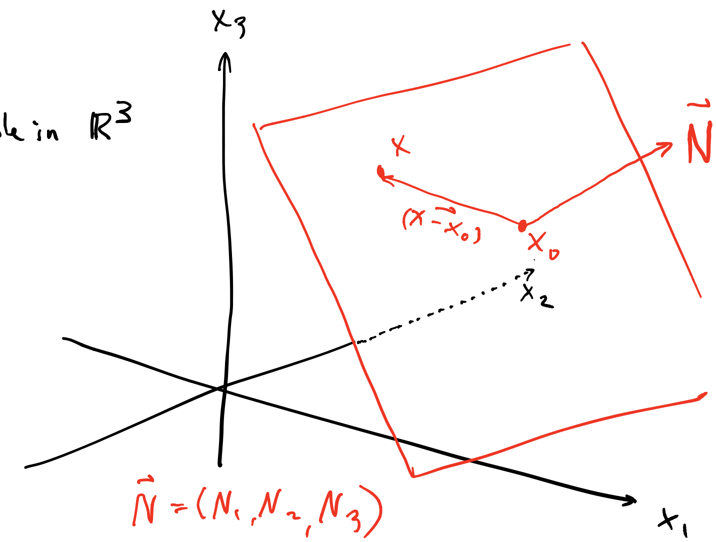
7. Define
$$N = \frac{\nabla_{x_0} f}{|\nabla_{x_0} f|}.$$

8. As long as the gradient is not horizontal – by which we mean that it has an nth coordinate $= 0$ – then we can write the nth coordinate as a function the first n-1 coordinates:

$$\begin{aligned} N &= (N_1, N_2, ..., N_n) \\ N \cdot (x - x_0) &= 0 \\ N_1 x_1 + N_2 x_2 + \cdots + N_n x_n &= N \cdot X_0 \\ &= C \end{aligned}$$

See the example in 3 dimensions ...

Example in $\mathbb{R}^3$



$$\vec{N} = (N_1, N_2, N_3)$$

$$\vec{N} \cdot (\overrightarrow{X - X_0}) = 0$$

$$N_1 x_1 + N_2 x_2 + N_3 x_3 = c \quad (c = \vec{N} \cdot \vec{x}_0)$$

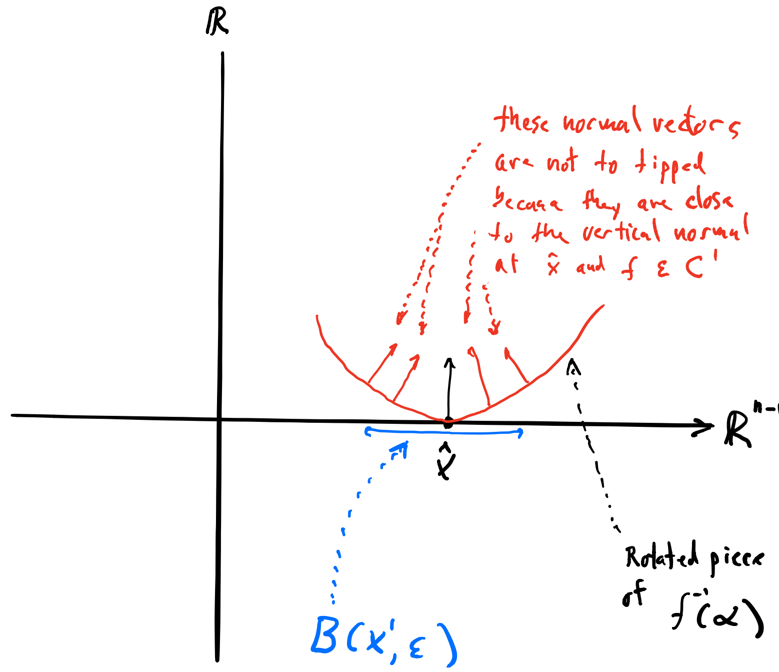
$$\Rightarrow \; x_3 = -\frac{N_1}{N_3} x_1 - \frac{N_2}{N_3} x_2 + \frac{c}{N_3}$$

9. Now we apply these insights to rotated level set and the now vertical gradient

vector, which is therefore normal to a horizontal tangent plane at the rotated point that we again denote by $\hat{x}$.

10. Let $\hat{x} = (\hat{x}', \hat{x}_n) \equiv (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{n-1}, \hat{x}_n)$ and note that $\hat{x}' \in \mathbb{R}^{n-1}$. We also represent any x in the rotated frame by $x = (x', x_n)$, $x' \in \mathbb{R}^{n-1}$ and we let x_0 be any point on $f^{-1}(\alpha)$.

11. Note that because $f \in C^1$ we know that for some little ball in the space of the first n-1 coordinates, $B(\hat{x}', \epsilon)$ centered at $\hat{x}'$ in the horizontal $\mathbb{R}^{n-1}$, the gradient is not too far from vertical.



12. Here is a short argument: we know that the surface tangent planes $H_{x_0}(x)$,

thought of as functions from $\mathbb{R}^{n-1} \to \mathbb{R}$, have small gradients eveyrwhere because from the example above, for $x_0' \in B(x', \epsilon)$

$$\nabla H_{x_0}(x') = (\frac{f_1(x_0)}{f_n(x_0)}, \frac{f_2(x_0)}{f_n(x_0)}, \ldots, \frac{f_{n-1}(x_0)}{f_n(x_0)})$$

and at $\hat{x}$ we have that

$$0 = f_1(\hat{x}) = f_2(\hat{x}) = f_3(\hat{x}) = \cdots = f_{n-1}(\hat{x})$$

and

$$\beta = f_n(\hat{x}) \neq 0.$$

Because the first n-1 partial derivatives are continuous, they remain small in a small ball about $\hat{x}'$.

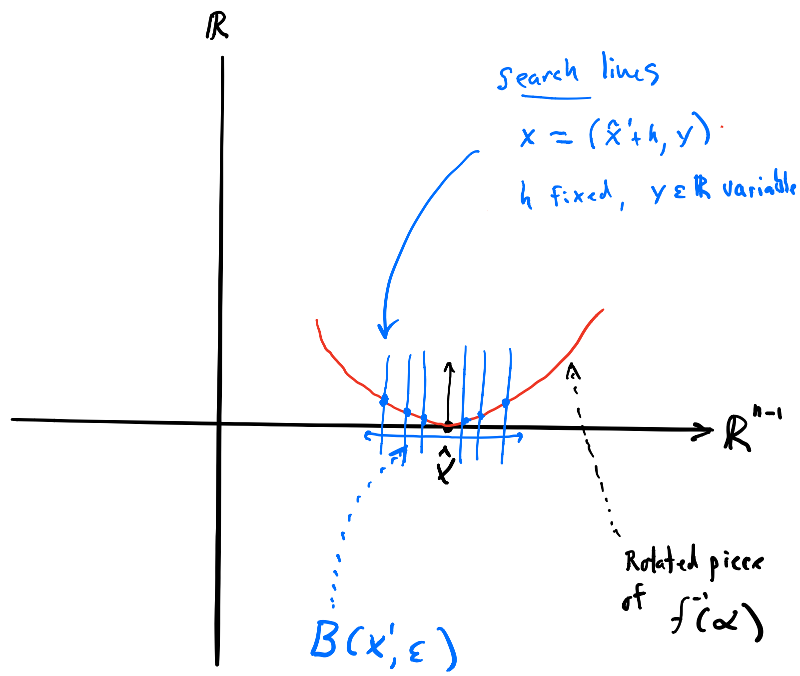
13. because of this, we know that g, which is the function the implicit function theorem gives us

$$f(x_1, x_2, \cdots, x_{n-1}, g(x_1, x_2, \cdots, x_{n-1})) = \alpha$$

is Lipschitz with small Lipshitz constant $K = |(k_1, k_2, ..., k_{n-1})|$ where $|f_1(x_0)| \leq k_1$ and $|f_2(x_0)| \leq k_2$ and $|f_3(x_0)| \leq k_3$ and etc.

14. ... and we can solve for g at any point $x' + h$, for any $h \in \mathbb{R}^{n-1}$ and $|h| \leq \epsilon$, there is a small y such that $f(x' + h, y) = \alpha$.

15. Thus we can shoot vertically at $x' + h$ to find the point where $f = \alpha$.



16. We can actually **use this argument to prove that a g exists that is**

Lipschitz and satisfies

$$f(x_1, x_2, \cdots, x_{n-1}, g(x_1, x_2, \cdots, x_{n-1})) = \alpha$$

in a neighborhood of $\hat{x}$.

Exercise 1.21. Convince yourself that there is a Lipschitz function g, using only the fact that the normals are close to vertical near $\hat{x}$, such that

$$f(x_1, x_2, \cdots, x_{n-1}, g(x_1, x_2, \cdots, x_{n-1})) = \alpha$$

for $x' \in B(0, \epsilon)$ Hint: think about it geometrically ... go ahead and do this when $n = 2$ so you can draw it easily and think about the drawings.

2 Integration

On this section, we dive into the integration of functions by motivating the introduction of the Lebesgue integral (and measure) using a function which is 1 on the rationals and 0 on the irrationals.

2.1 Riemann vs Lebesgue

We begin with an observation that there are functions we would like to integrate (at least for theoretical purposes) that do not have Riemannian integrals.

Define the function $f_Q(x)$ by:

$$f_Q : [0, 1] \to [0, 1] \equiv \begin{cases} 1 \text{ (when } x \in \mathbb{Q} \cap [0, 1]) \\ 0 \text{ (when } x \in [0, 1] \setminus \mathbb{Q}) \end{cases}$$

Now recall that, given a partition P of the domain $[0, 1]$ into sequential intervals by the points

$$0 = p_0 < p_1 < p_2 < p_3 < ... < p_m = 1,$$

the Riemann upper and lower integrals are defined to be:

$$\int_* f(x)\, dx \equiv \sup_{P \in \mathcal{P}} \left(\sum_{i=0}^{m-1} (p_{i+1} - p_i) \inf_{y \in [p_i, p_{i+1})} f(y) \right) \tag{12}$$

$$\int^* f(x)\, dx \equiv \inf_{P \in \mathcal{P}} \left(\sum_{i=0}^{m-1} (p_{i+1} - p_i) \sup_{y \in [p_i, p_{i+1})} f(y) \right) \tag{13}$$

where P is the family of all finite partitions of $[0, 1]$. **We say that f is Riemann integrable if $\int^* f(x)\, dx = \int_* f(x)\, dx$.**

Exercise 2.1. Show that for any two partitions of $[0,1]$, $P = \{p_i\}_{i=0}^m$ and $Q = \{q_i\}_{i=0}^k$,

$$\left(\sum_{i=0}^{m-1}(p_{i+1}-p_i)\sup_{y\in[p_i,p_{i+1})} f(y)\right) \geq \left(\sum_{i=0}^{k-1}(q_{i+1}-q_i)\inf_{y\in[p_i,p_{i+1})} f(y)\right)$$

Exercise 2.2. Show that $1 = \int^* f_Q(x) > \int_* f_Q(x) = 0$.

But it seems completely sensible to say that $\int_0^1 f_Q = 0$. The solution to this problem turns out to be simple: we simply partition the range, instead of the domain.

For now we will assume an intuitive grasp of the idea of $\mu(E)$, the measure of a set E, for now – it really is just what you think it should be, a fact that will be made clear in the next section when we define outer measures and a couple of specific families of measures on R^n.

Suppose that we partition a set $A \subset \mathbb{R}^n$ into $A = \cup_{i=1}^N E_i$, $E_i \cap E_j = \emptyset$ for all $i \neq j$, where $\hat{N} = \infty$ is a possibility. Suppose further that $\chi_E : \mathbb{R}^n \to \mathbb{R}$ is the characteristic function on E: i.e. $\chi_E(x) = 1$ when $x \in E$ and $\chi_E(x) = 0$ when $x \in E^c$. Now, for any non-negative sequence $\{\alpha_i\}_{i=1}^{\hat{N}}$ ($\alpha_i \geq 0$ for all i) we define a *simple function* $s(x)$ by

$$s(x) = \sum_{i=1}^{\hat{N}} \alpha_i \chi_{E_i}(x).$$

We now define the integral of $f(x)$ to be

$$\int s(x)\, d\mu x \equiv \sum_{i=1}^{\hat{N}} \alpha_i \mu(E_i).$$

Now letting $s(x)$ denote a simple function, we define

$$\int^* f d\mu = \inf_{s:s(x)\geq f(x)\forall x} \int s(x)\, d\mu$$

and

$$\int_* f d\mu = \sup_{s:s(x)\leq f(x)\forall x} \int s(x)\, d\mu.$$

Definition 2.1 (Lebesgue Integrable Functions)**.** *Define* $f^+(x) \equiv \max\{0, f(x)\}$ *and* $f^-(x) \equiv \max\{0, -f(x)\}$*. Then*

1. *We say that $f \geq 0$ is Lebesgue integrable if $\int^* f d\mu = \int_* f d\mu$. See Figure (2.1)*

2. *We say that f is Lebesgue integrable if both $f^+(x)\ d\mu$ and $\int f^-(x)\ d\mu$ are. In this case, we define $\int f\ d\mu = \int f^+\ d\mu - \int f^-\ d\mu$ with the convention that $\infty - \infty = \infty$.*

Returning to the integration of any function $f : [0,1] \to [0,1]$, pick a positive integer $M < \infty$ and notice that if we define $E_i = f^{-1}([\frac{i-1}{M}, \frac{i}{M}))$ for $i = 1, ..., M$ and then define

$$s^u(x) \equiv \sum_{i=1}^{M} \frac{i}{M} \chi_{E_i}(x)$$

and

$$s^l(x) \equiv \sum_{i=1}^{M} \frac{i-1}{M} \chi_{E_i}(x)$$

we have that

1. $s^u(x) \geq f(x)$ for all $x \in [0,1]$ and

2. $s^l(x) \leq f(x)$ for all $x \in [0,1]$,

3. which allows us to conclude that

$$\int s^u(x)\ d\mu - \int s^l(x)\ d\mu = \frac{1}{M} \underset{M\to\infty}{\to} 0$$

implying that f is integrable.

Exercise 2.3. Convince yourself (i.e. prove) that if $s(x)$ and $r(x)$ are two simple functions such that $s(x) \leq f(x) \leq r(x)$ for all x, then $\int_A s\ d\mu \leq \int_A r(x)\ d\mu$. Note that we are defining the integrals here, so you have to be careful to not assume what you are trying to prove.

Exercise 2.4. Show that $\int f_Q(x)\ d\mu$ exists and equals 0.

It turns out that the one thing we have assumed – that $\mu(E_i)$ makes sense for any $E_i \equiv f^{-1}([a,b))$ – opens up an important subject for us to look into more carefully. The reason for that is, if we assume that μ makes sense for all subsets of $\mathbb{R}^n$ we run smack dab into the Banach-Tarski paradox implying that we cannot let all sets be "measurable" if we want to have those measurements mean something.

Exercise 2.5. Look up the Banach-Tarski Paradox on Wikipedia and read about it.

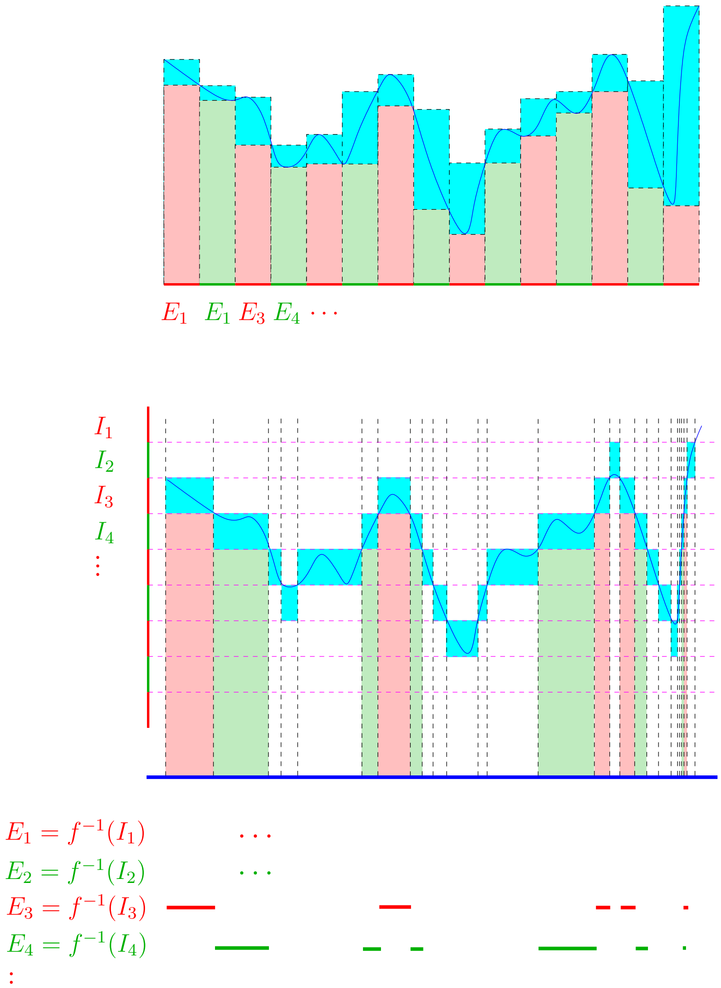


Figure 6: **Riemann versus Lebesgue Integration:** the upper figure illustrates the partition of the domain dictated by the Riemannian approach. The green and red rectangles live completely below the graph of f. Call the area they sum to $A_{\text{lower}}(P)$ where P is the partition. The red and green plus the cyan rectangles live completely above the graph. Call their area A_{upper}. If $\sup_P A_{\text{lower}}(P) = \inf_P A_{\text{upper}}$ then f is Riemann integrable. The lower figure illustrates that key difference for the Lebesgue case: we partition the range and pull that back by f^{-1} to a partition of the domain. It turns out that this is exactly what is needed to make all reasonable functions integrable. Now $A_{\text{lower}}(P) = \sum_i a_i \mu(E_i)$ and $A_{\text{upper}}(P) = \sum_i b_i \mu(E_i)$ where P is a partition of the range into the intervals $I_i = [a_i, b_i)$.

Remark 2.1 (Summable versus Integrable)**.** *We will say a function f is summable if it is integrable **and** the integral of the function is finite.*

Exercise 2.6. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *measurable* if the set $f^{-1}(I)$ is $\mathcal{L}^n$ measurable whenever I is a (possibly infinite) interval. Suppose that the support of f is bounded. Show that the Riemann integral $\int f(x)\, dx$ exists when f is continuous, but that it is even easier to show that the Lebesgue integral $\int f(x)\, d\mathcal{L}^1 x$ exists when f is merely measurable. Hint: A continuous function on a compact set is uniformly continuous.

Remark 2.2. *Note that we are motivating the next section with this last exercise: what does it mean for E to be measurable, other than that $\mu(E)$ "makes sense"? To find the answer we will use, read on ...*

2.2 Outer Measures

The approach to measure theory I like closely follows the approach used by Evans' and Gariepy in their *Measure Theory and Fine Properties of Functions* – a book I very highly recommend for anyone interested in analysis.

Definition 2.2 (Outer Measure)**.** *Any function, μ mapping subsets of a space X to $[0, \infty]$ – $\mu : 2^X \to [0, \infty]$ – satisfying the following two rules is called an **Outer Measure**:*

1. $\mu(\emptyset) = 0$
2. $\mu(E) \leq \sum_{i=1}^{N} \mu(F_i)$ *where* $E \subset \cup_i F_i$ *and* $0 < N \leq \infty$

Both families of measures we use in this class – the Lebesgue measures and the Hausdorff measures – are outer measures. Because of the Banach-Tarski Paradox, we know that we cannot just let every set into the club of sets whose outer measure is meaningful.

Definition 2.3 (Measurable Sets)**.** *If a set $E \subset X$ has the property that for all $A \subset 2^X$:*

$$\mu(A) = \mu(A \cap E) + \mu(A \cap E^c)$$

we say that E is μ-measurable or simply measurable if the μ is clear from the context.

The idea is that E slices every set up in a sensible way.

Exercise 2.7. **(Easy consequences of the definition of measurability)** Show that the definition of measurability easily gives us (1) that E measurable $\Rightarrow E^c$; and (2) X and $\emptyset$ are measurable.

Remark 2.3. *Note that we always have that*

$$\mu(A) \leq \mu(A \cap E) + \mu(A \cap E^c)$$

so we only need to show

$$\mu(A) \geq \mu(A \cap E) + \mu(A \cap E^c)$$

to prove that

$$\mu(A) = \mu(A \cap E) + \mu(A \cap E^c)$$

Exercise 2.8. Show that if $\mu(E) = 0$, then E is measurable.

Definition 2.4 (σ-algebra of sets)**.** *a collection of sets $\mathcal{A}$ is a σ-algebra if:*

1. $\emptyset,\ X \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow X \setminus A = A^c \in \mathcal{A}$
3. *Every set in the sequence* $\{A_i\}_{i=1}^{\infty}$ *are in* $\mathcal{A}$ *implies that* $\cup_i^{\infty} A_i \in \mathcal{A}$.

Theorem 2.1 (Properties of Measures)**.** *Suppose that* $\{E_i\}_{i=1}^{\infty}$ *is a sequence of measurable sets. Then we have that:*

1. $\cup_i^{\infty} E_i$ *and* $\cap_i^{\infty} E_i$ *are measurable.*
2. *The collection of sets which are measurable form a σ-algebra.*
3. *If* $\{E_i\}_{i=1}^{\infty}$ *are pairwise disjoint –* $E_i \cap E_j \emptyset$ *when* $i \neq j$ *– then*

$$\mu\left(\cup_i^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i).$$

   This property is called countable additivity.

4. *If* $E_1 \subset E_2 \subset \cdots \subset E_k \subset E_{k+1} \subset \cdots$ *then*

$$\lim_{i \to \infty} \mu(E_i) = \mu\left(\cup_i^{\infty} E_i\right).$$

5. *If if* $\mu(E_1) < \infty$ *and* $E_1 \supset E_2 \supset \cdots \supset E_k \supset E_{k+1} \supset \cdots$ *then*

$$\lim_{i \to \infty} \mu(E_i) = \mu\left(\cup_i^{\infty} E_i\right).$$

2.2.1 Lebesgue and Hausdorff Measures

The d-dimensional Lebesgue measures are constructed by using d-dimensional rectangles to cover $E \subset \mathbb{R}^d$ and taking an infimum. We define an open rectangle R to be

$$R = R(x^*, \epsilon*) = \{(x_1, x_2, ...x_d) \in (x_1^*, x_1^* + \epsilon_1) \times (x_2^*, x_2^* + \epsilon_2) \times \cdots \times (x_d^*, x_d^* + \epsilon_d)\}$$

and its **content** $c(R)$ to be the product of the side-lengths of the rectangle = its usual d-volume:

$$c(R) = \epsilon_1 \epsilon_2 \cdots \epsilon_d = \Pi_{k=1}^d \epsilon_k.$$

We can now define:

Definition 2.5 (Lebesgue Measure)**.**

$$\mathcal{L}^d(E) \equiv \inf_{\{\mathcal{R} | E \subset \cup_i R_i\}} \sum c(R_i)$$

where we are minimizing over all $\mathcal{R}$, the countable covers of E by open rectangles: $\mathcal{R} = \{R_i\}_{i=1}^\infty$. The following four exercises are **also** given as the third problem for the course:

Exercise 2.9. Show that $\mathcal{L}^d(S) = 0$ for any $S = x \in \mathbb{R}^d$ such that $x_i = c$ – the d-1-dimensional plane obtained by holding the i-th coordinate constant.

Exercise 2.10. Show that if $\mu(D) = 0$, then $\mu(C \cup D) = \mu(C)$

Exercise 2.11. Show that the measure of any rectangle when some or all of the intervals defining it are not open is the same as the corresponding open rectangle.

Exercise 2.12. Show that for any rectangle $R \subset \mathbb{R}^d$,

$$\mathcal{L}^d(R) = c(R).$$

Hint: take the closed rectangle $\bar{R}$ corresponding to R and notice that any cover with open rectangles has a finite subcover also covering $\bar{R}$.

Now we define the family of Hausdorff outer measures. Now we can take any countable cover of a set E and then measure each by the volume of the ball having he same diameter. For any real number $\eta \in [0, \infty)$, we define:

Definition 2.6 (Hausdorff Measures, $\mathcal{H}^\eta$)**.**

$$\mathcal{H}^\eta_\delta(E) \equiv \inf_{\{F_i\}_{i=1}^\infty | E \subset \cup_i F_i \text{ and } \sup_i(\operatorname{diam} F_i) \leq \delta} \alpha(\eta) \sum_i \left(\frac{\operatorname{diam} F_i}{2} \right)^\eta$$

and then

$$\mathcal{H}^\eta(E) = \lim_{\delta \to 0} \mathcal{H}^\eta_\delta(E).$$

Note that $\alpha(\eta) = \frac{\pi^\eta}{2^\eta \Gamma(\frac{\eta}{2}+1)}$*, the* η*-volume of the "*η*-dimensional" unit ball. This number agrees with the usual volume when* η *is an integer.*

It turns out that when η is a non-negative integer, $\mathcal{L}^\eta = \mathcal{H}^\eta$. (See Evans and Gariepy's book for a proof of this fact.) The first thing we will prove is that for any fixed $E \subset \mathbb{R}^n$, the graph of $\mathcal{H}^\eta(E)$ versus η looks like the graph in Figure (7).

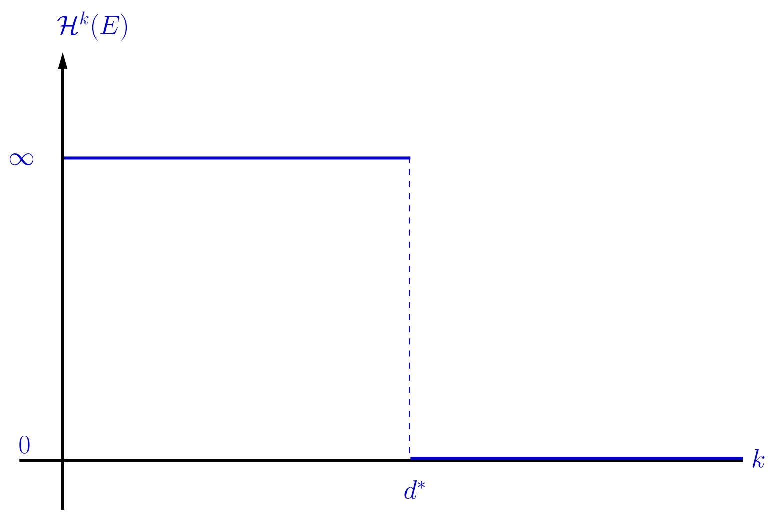


Figure 7: Graph of the Hausdorff measure $\mathcal{H}^k(E)$ of a set E as we vary k, the dimension of the measure. We define d^*, where the measure switches from ∞ to 0 to be the dimension of the set E.

Theorem 2.2 (Definition of Hausdorff Dimension)**.** *Suppose that* $E \subset \mathbb{R}^n$ *for some* $n < \infty$*. Then* $\mathcal{H}^k(E) = 0$ *for* $k > d^*$ *for some* $d^* \leq n$ *and, if* $d^* > 0$*,* $\mathcal{H}^k(E) = \infty$ *for* $k < d^*$*. This is illustrated in Figure (7).*

Remark 2.4. *It turns out that there are* E*'s, such that* $\mathcal{H}^{d^*}(E)$ *is any number between and including* 0 *and* ∞*.*

Proof.

1. First we show if $0 \leq \mathcal{H}^k(E) < \infty$ then $\mathcal{H}^{k+\eta}(E) = 0$ for any $\eta > 0$.

   (a) Choose $\epsilon > 0$

   (b) Choose $0 < \delta < 1$ such that $\left(\frac{\delta}{2}\right)^\eta < \epsilon$.

   (c) By the definition of Hausdorff measure, there is a $\delta_\epsilon < \delta$ such that

   $$\left|\mathcal{H}^k_{\delta_\epsilon}(E) - \mathcal{H}^k(E)\right| < \epsilon/2$$

   (d) There is also a cover $\{F_i\}_{i=1}^\infty \in \mathcal{F}_{\delta_\epsilon}$ such that

   $$\left|\mathcal{H}^k_{\delta_\epsilon}(E) - \alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k\right| \leq \epsilon/2$$

   (e) We get that

   $$\left|\mathcal{H}^k(E) - \alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k\right| \leq \epsilon$$

   from which we get that

   $$\alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k \leq \mathcal{H}^k(E) + \epsilon$$

   (f) Now we note that this implies that

   $$\begin{aligned}\alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^{k+\eta} &= \alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k \left(\frac{\operatorname{diam} F_i}{2}\right)^\eta \\ &\leq \alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k \left(\frac{\delta}{2}\right)^\eta \\ &< \alpha(k)\sum_i \left(\frac{\operatorname{diam} F_i}{2}\right)^k \epsilon \\ &\leq (\mathcal{H}^k(E) + \epsilon)\epsilon\end{aligned}$$

   (g) This implies $\mathcal{H}^{k+\eta}_\delta(E) < \epsilon(\mathcal{H}^k(E) + \epsilon)$

   (h) Since $\epsilon > 0$ was arbitrary and $\mathcal{H}^k(E) < \infty$, we conclude that $\mathcal{H}^{k+\eta}_\delta(E) = 0$

   (i) But δ can be chosen arbitrarily small, implying that

   $$\mathcal{H}^{k+\eta}(E) = \lim_{\delta \to 0} \mathcal{H}^{k+\eta}_\delta(E) = 0$$

2. Suppose that $\mathcal{H}^0(E) < \infty$. Then what we just proved shows that we have that $\mathcal{H}^k(E) = 0$ for all $k > 0$ and we are done.

3. **Exercise**: show that $\mathcal{H}^{n+1}(\mathbb{R}^n) = 0$ where we are considering $\mathbb{R}^n$ to be equal to $\{(x_1, ..., x_{n+1} | x_{n+1} = 0\} \subset \mathbb{R}^{n+1}$.

4. Suppose that $\mathcal{H}^0(E) = \infty$. Define $d^* \equiv \sup\{x < n+1 | \mathcal{H}^x(E) = \infty\}$.

5. Using the result we proved first, we get that

$$\mathcal{H}^{d^*+\eta/2} < \infty \Rightarrow \mathcal{H}^{d^*+\eta} = 0.$$

6. For any $\eta > 0$ there must be a $0 < \delta < \eta$ such that $\mathcal{H}^{d^*-\delta} = \infty$, implies that $\mathcal{H}^{d^*-\eta} = \infty$. Otherwise, because $d^* - \eta < d^* - \delta$, if $\mathcal{H}^{d^*-\eta} < \infty$ we would have $\mathcal{H}^{d^*-\delta} = 0$.

7. We conclude that $\mathcal{H}^k(E) = \infty$ for $k < d^*$ and $\mathcal{H}^k(E) = 0$ for $k > d^*$.

□

Exercise 2.13. Show that set

$$E \equiv \{(x_1, x_2, x_3) \mid x_1 \in [0,1], x_2 \in [0,1], x_3 = 0\}$$

– a 2 dimensional square embedded in $\mathbb{R}^3$ – satisfies $\mathcal{H}^3(E) = 0$.

Exercise 2.14. Show that you can assume the sets used to generate the covers in the Hausdorff definition are convex. **Hint:** show that for any $E \subset \mathbb{R}^n$,

$$\text{diam}(E) = \text{diam}(\text{cnv}(E))$$

where $\text{diam}(A)$ denotes diameter of a set A and $\text{cnv}(A)$ denotes the convex hull of a set A. Do this by showing:

- For the set to have finite diameter, it must be bounded
- that because $E \subset \text{cnv}(E)$, $\text{diam}(E) \leq \text{diam}(\text{cnv}(E))$
- that there is a sequence of pairs of of points in $\text{cnv}(E)$, $\{p_i, q_i\}_{i=1}^{\infty}$ such that $|p_i - q_i| \to \text{diam}(\text{cnv}(E))$.
- there is a subsequence $i(k)$ such that $p_{i_k} \to p^*$ and $q_{i_k} \to q^*$ and $|p^* - q^*| = \text{diam}(\text{cnv}(E))$.
- that the projection of E onto the line L_{p^*,q^*} through p^* and q^* has diameter at most $\text{diam}(E)$.

- Define the smallest interval, in L_{p^*,q^*}, containing this projection to be P and show that E lives between the two $n-1$-dimensional planes orthogonal to L_{p^*,q^*} through the endpoints of P.

- Conclude that cnv(E) also must be contained between the two $n-1$-dimensional planes orthogonal to L_{p^*,q^*} through the endpoints of P (since cnv(E) = intersection of all convex sets containing E and the set of points between and including the two planes is convex).

- Therefore, we must have the that

$$\begin{aligned}\text{diameter of cnv}(E) &= |p^* - q^*| \\ &= \text{diameter of projection of cnv}(E) \text{ onto } L_{p^*,q^*} \\ &\leq \text{diam}(P) \\ &\leq \text{diam}(E)\end{aligned}$$

Exercise 2.15. Show that you can assume the sets used to generate the covers in the Hausdorff definition are open. **Hint:** show that ...

1. For any cover of a set E, $\{F_i\}_{i=1}^{\infty}$, the sets

$$\hat{F}_i \equiv B(F_i, \delta_i) \equiv \cup_{x\in F_i} B(x, \delta_i)$$

are open. (Reminder: $B(x, \eta)$ = open ball of radius η centered at x).

2. Now assume that $\{F_i\}_{i=1}^{\infty}$ is a cover of E and show that if we define

$$\delta_i = \left(\left(\frac{\text{diam}(F_i)}{2}\right)^d + \epsilon^i\right)^{\frac{1}{d}} - \frac{\text{diam}(F_i)}{2}$$

then

$$\left(\frac{\text{diam}(F_i)}{2} + \delta_i\right)^d = \left(\frac{\text{diam}(F_i)}{2}\right)^d + \epsilon^i$$

3. Use this to show

$$\alpha(d) \sum_{i=1}^{\infty} \left(\frac{\text{diam}\,\hat{F}_i}{2}\right)^d \leq \alpha(d) \sum_{i=1}^{\infty} \left(\frac{\text{diam}\,F_i}{2}\right)^d + \alpha(d)\frac{\epsilon}{1-\epsilon}.$$

4. Use this to deduce that the infimum over arbitrary covers is the same as the infimum over arbitrary covers.

Exercise 2.16. Show that we may also restrict ourselves to closed sets when covering a set in order to compute Hausdorff measures. **Hint:** If we denote the closure of E by $\bar{E}$, you simply need to show that $\mathrm{diam}(E) = \mathrm{diam}(\bar{E})$.

Exercise 2.17. Suppose that S = a line segment in $\mathbb{R}^n$, $n \geq 2$. You can assume that the line segment is a subset of the x_1 axis. I.e. $S = [a,b] \times \{0\} \times \cdots \times \{0\}$. Use the results in Exercises 2.14 and 2.15 to prove that the $\mathcal{H}^1(S) = b - a$. **Hint:** Use the fact that S is compact to covert every cover to a finite cover.

2.2.2 Caratheodory Criterion

There is a very useful criterion that tells us when the Borel sets are measurable is the Caratheodory Criterion:

Theorem 2.3 (Caratheodory Criterion)**.** *If μ is an outer measure on $\mathbb{R}^n$ and we know that* $\mathrm{dist}(A,B) > 0 \;\; \Rightarrow \;\; \mu(A \cup B) = \mu(A) + \mu(B)$*, then μ is a Borel measure – i.e. all Borel sets are measurable.*

Proof.
If we show that all closed sets are measurable, then, because the class of measurable sets is a σ-algebra, we know all open sets are also measurable. Therefore the measurable sets contains the smallest σ-algebra containing the open sets – the Borel σ-algebra.

1. Let A be an arbitrary set in X and C be a closed set.

2. The result is immediate if $\mu(A) = \infty$, so assume $\mu(A) < \infty$.

3. Define
$$C_n = \{x \in \mathbb{R}^n \mid \mathrm{dist}(x, C) \leq \frac{1}{n}\}$$
where $\mathrm{dist}(x, C)$ is the distance from x to the set C.

4. Because $\mathrm{dist}(C, C_n^c) = \frac{1}{n} > 0$, we know that
$$\mu(A) \geq \mu(\{A \cap C\} \cup \{A \cap C_n^c\}) = \mu(A \cap C) + \mu(A \cap C_n^c)$$

5. Define
$$A_i = \{x \in A \mid \frac{1}{i+1} < d(x, C) \leq \frac{1}{i}\}.$$
Then
$$A = \{A \cap C\} \cup \{A \cap C_n^c\} \cup \{\cup_{i=n}^{\infty} A_i\}.$$

6. Our aim is to show that
$$\mu(A) \geq \mu(A \cap C) + \mu(A \cap C^c)$$

7. but because of Step 4 above and $\{A \cap C\} \cup \{A \cap C^c\} \subset \{A \cap C\} \cup \{A \cap C_n^c\} \cup \{\cup_{i=n}^{\infty} A_i\}$ we know that

$$\begin{aligned} \mu(A) + \sum_{i=n}^{\infty} \mu(A_i) &\geq \mu(A \cap C) + \mu(A \cap C_n^c) + \sum_{i=n}^{\infty} \mu(A_i) \\ &\geq \mu(A \cap C) + \mu(A \cap C^c) \end{aligned}$$

8. All we need to do now is show that $\sum_{i=1}^{\infty} \mu(A_i) < \infty$ which implies that $\sum_{i=n}^{\infty} \mu(A_i) \to 0$ which then implies that

$$\mu(A) + \epsilon \geq \mu(A \cap C) + \mu(A \cap C^c)$$

for all $\epsilon > 0$.

9. But, defining $A_n' = A_1 \cup A_3 \cup A_5 \cup ... \cup A_{2n+1}$ $A_n'' = A_2 \cup A_4 \cup A_6 \cup ... \cup A_{2n}$ and noting that for all n, we have

   - $\mu(A) \geq \mu(A_n')$
   - and by the hypothesis $\text{dist}(A, B) > 0 \;\; \Rightarrow \;\; \mu(A \cup B) = \mu(A) + \mu(B)$ we have

   $$\begin{aligned} \mu(A_n') &= \mu(A_1 \cup A_3 \cup A_5 \cup ... \cup A_{2n+1}) \\ &= \mu(A_1) + \mu(A_3) + \mu(A_5) + ... + \mu(A_{2n+1}) \end{aligned}$$

   - and $\mu(A) \geq \mu(A_n'')$
   - and by the hypothesis $\text{dist}(A, B) > 0 \;\; \Rightarrow \;\; \mu(A \cup B) = \mu(A) + \mu(B)$ we have

   $$\begin{aligned} \mu(A_n'') &= \mu(A_2 \cup A_4 \cup A_6 \cup ... \cup A_{2n}) \\ &= \mu(A_2) + \mu(A_4) + \mu(A_6) + ... + \mu(A_{2n}) \end{aligned}$$

   - we conclude that $\sum_{i=1}^{\infty} \mu(A_i) \leq 2\mu(A)$.

10. We are done!

□

Exercise 2.18. Show that Lebesgue measure of a set can be found be restricting yourself to covers with rectangles whose side-length is bounded by any $\delta > 0$.

Exercise 2.19. Show that both Lebesgue and Hausdorff measures are Borel Regular.

2.2.3 Radon Measures and Approximation

Recall that the Borel sets are any subset in the smallest σ-algebra of sets containing the open sets.

Definition 2.7 (Regular, Borel, Radon ...)**.** *Suppose that μ is an outer measure on $X = \mathbb{R}^n$.*

Regular *If, for every set $A \subset X$, there is a measurable set B, such that $A \subset B$ and $\mu(A) = \mu(B)$, then we say μ is a* ***regular*** *measure.*

Borel *If every Borel set is measurable by μ, we say that μ is a* ***Borel measure****.*

Borel Regular *If μ is a Borel measure and for every set $A \subset X$, there is a Borel set B, such that $A \subset B$ and $\mu(A) = \mu(B)$, then we say μ is a* ***Borel regular*** *measure.*

Radon *If μ is a Borel Regular measure and $\mu(K) < \infty$ for all compact sets K, we say that μ is a* ***Radon*** *measure.*

The following approximation property of Radon measures is very useful.

Theorem 2.4 (Approximation of Radon Measures)**.** *Suppose that μ is a Radon Measure. Then*

1. *We can approximate from the outside with open sets: For any set $E \subset \mathbb{R}^n$,*

$$\mu(E) = \inf\{\mu(O) \mid E \subset O, \;\; O \textit{ is open}\}$$

2. *We can approximate from the inside with compact sets: For any measurable set $E \subset \mathbb{R}^n$,*

$$\mu(E) = \sup\{\mu(K) \mid K \subset E, \;\; K \textit{ is compact}\}$$

Exercise 2.20. Show that Hausdorff measures satisfy part 1 of Theorem 2.4 ... I.e. the measure of a set is approximated by open sets from outside.

2.3 Measurable Functions and Integration

Lebesgue integration is the typical choice of analysts when they want to think about integrating things. But it is not the only choice. Daniell integrals, Steltjes integrals, and a bunch of others are out there, all with their particular uses and enthusiasts. Our approach here is pragmatic: Lebesgue works for most things and for those things we will use it. When it doesn't quite fit the bill, we use what does work.

So, what is Lebesgue integration and how does it differ from Riemann integration? As you have already seen, in Riemann integration, we partition the *domain* into regular subsets (intervals or rectangles) and take the the largest and smallest functional values attained in each subset, multiply these values by the measure of those subsets and sum these up, after which we take infimums and supremums:

$$\int^* f d\mu \equiv \inf_P \sum_i \sup_{x \in I_i} f(x)\mu(I_i)$$

$$\int_* f d\mu \equiv \sup_P \sum_i \inf_{x \in I_i} f(x)\mu(I_i)$$

where P is the partition of the domain into intervals I_i. If $\int^* f d\mu = \int_* f d\mu$ then we say f is Riemann integrable.

In Lebesgue integration, we parition the *range* into intervals I_i and pull them back to a partition of the domain: $E_i = f^{-1}(I_i)$. (This paritition can be very far from regular!) We get:

$$\int^* f d\mu \equiv \inf_P \sum_i \left(\sup_{y \in I_i} y \right) \mu(f^{-1}(I_i)) = \inf_P \sum_i b_i \mu(f^{-1}(I_i))$$

$$\int_* f d\mu \equiv \sup_P \sum_i \left(\inf_{y \in I_i} y \right) \mu(f^{-1}(I_i)) = \sup_P \sum_i a_i \mu(f^{-1}(I_i))$$

We are rewarded for our change in pespective by the result that now, *every* respectable function is integrable! (By integrable we will mean the upper and lower integrals are equal). As a result, we like the Lebesgue integral and are not so inclined to like the Riemann integral, even though for many practical purposes they are indistinguishable (because for really nice functions, they are the same.) Figure 8 illustrates both versions of integration.

2.3.1 Lebesgue Integration

Now we work through the definition of Lebesgue integration a bit more slowly and carefully. We will do this in three steps:

1. Define step functions caefully (We'll call them *simple functions*).

2. Define integrals of step functions.

3. Approximate general functions using step functions and define the integeal of the function as the limit of the integals of the approximating step functions.

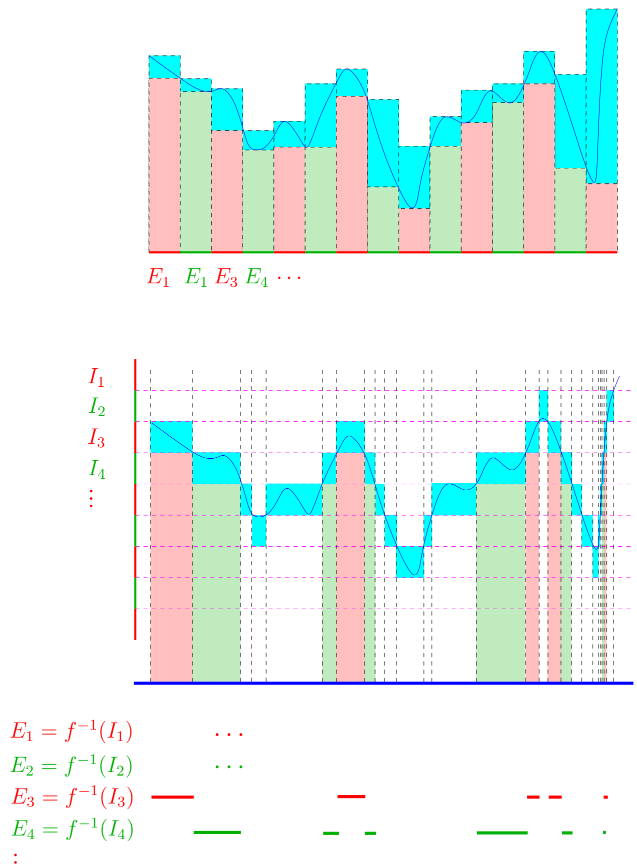


Figure 8: **Riemann versus Lebesgue Integration:** the upper figure illustates the partition of the domain dictated by the Riemannian approach. The green and red rectangles live completely below the graph of f. Call the area they sum to $A_{\text{lower}}(P)$ where P is the paritition. The red and green plus the cyan rectangles live completely above the graph. Call their area A_{upper}. If $\sup_P A_{\text{lower}}(P) = \inf_P A_{\text{upper}}$ then f is Riemann integrable. The lower figure illustrates that key difference for the Lebesgue case: we partition the range and pull that back by f^{-1} to a partition of the domain. It turns out that this is exactly what is needed to make all reasonable functions integrable. Now $A_{\text{lower}}(P) = \sum_i a_i \mu(E_i)$ and $A_{\text{upper}}(P) = \sum_i b_i \mu(E_i)$ where P is a parition of the range into the intervals $I_i = [a_i, b_i)$.

2.3.2 Simple Functions

From one point of view, the simplest function we can define is a function that takes on the value 1 on a some set E and 0 on the complment of E, $E^c \equiv x \in \mathbb{R}^n \setminus E$. We call such a function the ***characteristic function*** of E and we denote it by χ_E.

$$\chi_E(x) \equiv \begin{cases} 1 \text{ if } x \in E \\ 0 \text{ if } x \in E^c \end{cases}$$

Now we can build any step function we might want to build by scaling characteristic functions and adding them together. One way to do this is to partition the domain $\mathbb{R}^n$ into a countable collection of sets $\{E_i\}_{i=1}^N$ where $N \in \{\mathbb{Z}^+ \cup \infty\}$. This yields:

$$g(x) \equiv \sum_i \alpha_i \chi_{E_i}(x).$$

We call any such step function a ***simple function***. An equivalent definition defines simple functions $g : \mathbb{R}^n \to \mathbb{R}$ to be those functions that take on at most a countable number of values.

2.3.3 Integrating Simple Functions

It should seem completely natural to *define* the integral of χ_E to be the measure of E – it agrees with the intuition of area under the graph, supported by our definition of the area of rectangles to be width times height. And so this is what we do:

$$\int \alpha \chi_E \, d\mu \equiv \alpha \mu(E).$$

If g is a simple function with representation $g(x) = \sum_i \alpha_i \chi_{E_i}(x)$, this leads us to define the integral of the simple function g***integral of a simple function*** to be:

$$\int g \, d\mu \equiv \sum_i \alpha_i \mu(E_i).$$

Note: we will require that the E_i's partition the domain and we will define $0 \cdot \infty = \infty \cdot 0 = 0$.

2.3.4 Integrating Arbitrary Functions

Above, we are measuring sets like $E_i = g^{-1}(\alpha_i)$, the inverse image of a point in the range of g. More generally, we will work with inverse images of Borel sets and we would like the f's we work with to have the property that such subsets of the domain are always measurable. If they are, we say f is a ***μ-measurable function***:

Definition 2.1 (measurable functions)**.** *If $E = f^{-1}(B)$ is a μ-measurable subset of $\mathbb{R}^n$ for all Borel $B \subset \mathbb{R}$, then $f : \mathbb{R}^n \to \mathbb{R}$ is said to be* **measurable***.*

Unless otherwise indicated, all functions will be assumed μ-measurable.

If $f : \mathbb{R}^n \to [0, \infty]$, we define

$$\int^* f d\mu \equiv \inf_{\text{simple } g \geq f} \int g d\mu$$

$$\int_* f d\mu \equiv \sup_{\text{simple } g \leq f} \int g d\mu$$

Notice that $\int^* f d\mu \geq \int_* f d\mu$. If these two values are equal, then we say f is ***integrable*** with respect to μ and we define the ***integral of a non-negative function*** f to be that common value. Finally, if $f : \mathbb{R}^n \to [-\infty, \infty]$, we say that f is integrable if both it's positive and negative parts – f^+ and f^- – are integrable and one of the values is not infinite. That is, we define:

Definition 2.2 (***integral of an arbitrary function***)**.** *Define* $f^+ = \max\{f, 0\}$ *and* $f^- = \max\{-f, 0\}$. *If* f^+ *and* f^- *are integrable and one of the values is not infinite, then*

$$\int f \, d\mu = \int f^+ d\mu - \int f^- \, d\mu$$

Theorem 2.5. *Any μ-measurable, non-negative function is integrable.*

Exercise 2.21. Prove Theorem 2.5. Hint: define the $E_i \equiv f^{-1}([\alpha^i, \alpha^{i+1})$ where $\alpha > 1$. Define $F \equiv \mathbb{R}^n \setminus \bigcup_{i \in \mathbb{Z}} E_i$. Consider simple functions based on the partition of $\mathbb{R}^n$ $F \cup \bigcup_i E_i$.

Remark 2.5. *Note that many other authors use the term integrable to mean what we mean by integrable* and *that* $|\int f d\mu| < \infty$.

Definition 2.3. *We will say that f is* μ-summable *if f is integrable and* $|\int f d\mu| < \infty$.

Remark 2.6. *We notice immediately that sets of measure zero have no impact on the value of the integral: we may redefine the function on a set of measure zero and the integral remains unchanged. Notice also that a countable number of measure zero sets has a union that also has measure zero. This is a handy fact to keep in mind.*

Remark 2.7. *Notice that the hint in exercise 2.21 implies that in fact, we can focus on paritions of the domain that are pullbacks of partitions of the range* $\mathbb{R}$ *into intervals:* $E_i = f^{-1}(I_i)$.

Any intuitive idea you already have of integration will work if you allow for the fact that the measure we are integrating against may measure the sets in the domain quite differently than the usual Lebesgue measure, though we will usuaully be using either Lebesgue or Hausdorff measures and these do what you think they should (possibly after studying a few examples). What takes longer to grasp are the exotic sets that one can define. In fact, from one point of view, that is the whole point of a large part of geometric measure theory.

2.3.5 Properties of Integrals and Measurable Functions

Exercise 2.22. **(Linearity)** Show that if the integral of one of f or g is finite, the Lebesgue integral is linear:

$$\int (\alpha f + \beta g)\ d\mu = \alpha \int f\ d\mu + \beta \int g\ d\mu$$

Exercise 2.23. You might like to try to prove the following theorem that appears on page 5 of Evans and Gariepy: at least think about it before you look up the proof. *Let μ be a Borel regular measure on $\mathbb{R}^n$. Define $\mu \llcorner A(E) \equiv \mu(E \cap A)$. Suppose that $A \subset \mathbb{R}^n$ is μ-measurable and $\mu(A) < \infty$. Then $\mu \llcorner A$ is a Radon measure.*

Exercise 2.24. Suppose that $f : X \to Y$ and suppose that $(f^{-1}(A)|A \in \mathcal{A})$ is measurable in X. Prove that $(f^{-1}(B)|A \in \mathcal{B})$ is also measurable where $\mathcal{B}$ is the σ-algebra in Y generated by $\mathcal{A}$.

Exercise 2.25. **(Properties of Measurable Functions I)** Use the results of exercise 2.24 to show that if $f : \mathbb{R}^n \to \mathbb{R}$, then showing that all sets in $\{f^{-1}((-\infty, a)) \mid a \in \mathbb{R}\}$ are measurable is enough to show that f is measurable. Do the same for the collection of sets $\{f^{-1}((-\infty, a]) \mid a \in \mathbb{R}\}$.

Exercise 2.26. **(Properties of Measurable Functions II)** Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$, and $\{f_k : \mathbb{R}^n \to [-\infty, \infty]\}_{i=1}^{\infty}$ are all μ-measurable. Prove:

1. $f + g$, fg, $|f|$, $\min(f, g)$ and $\max(f, g)$ are all measurable. f/g is also μ-measurable provided $g \neq 0$ on $\mathbb{R}^n$.
2. $\inf_k f_k$, $\sup_k f_k$, $\liminf_{k\to\infty} f_k$ and $\limsup_{k\to\infty} f_k$ are all μ-measurable.

Hint: see Evans and Gariepy, Theorem 6 in section 1.1 (page 11).

2.4 Modes of Convergence and Three Theorems

If $\{f_i\}_{i=1}^{\infty}$ is a sequence of functions from our measure space to $\mathbb{R}$, $f_i : X \to \mathbb{R}$, we would like to know how the integral behaves in relation to convergence of the sequence. That is when is it true that:

$$\lim_{i\to\infty} \left(\int f_i\ dx \right) = \int \left(\lim_{i\to\infty} f_i \right)\ dx? \tag{14}$$

This is actually a motivating question that leads us to try to understand the differences between the different modes of convergence or closeness that can be defined. We begin by exploring some examples a bit.

2.4.1 Examples

Reminder – Uniform Convergence: we say that f_i converges uniformly to f if

$$\sup_{x\in X} |f_i(x) - f(x)| \underset{i\to\infty}{\to} 0.$$

When the **measure** and **convergence of f_i to f** are

Finite and Uniform: i.e. $\mu(X) < \infty$, and $\sup_{x\in X} |f_i(x)-f(x)| \underset{i\to\infty}{\to} 0$, the answer to the question in Equation 14 is yes!

Non-finite Measure, Uniform Convergence: The same question is answered no, and

Finite Measure, Non-uniform Convergence: no in this case too.

Exercise 2.27. Show that finite measure and uniform convergence implies we can switch limits with integration, in other words that the answer to the question in Equation 14 is yes.

Exercise 2.28. Give an example of a sequence of functions f_i approaching f uniformly, on a measure space X for which $\mu(X)$ is infinite, where the answer to 14 is no. Hint: look at constant functions on the real line.

Exercise 2.29. Give and example of a uniformly convergent sequence f_i on an infinite measure space X, such that

$$\lim_{i\to\infty} \left(\int f_i \, dx \right) = 2$$

and

$$\int \left(\lim_{i\to\infty} f_i \right) \, dx = 0$$

Exercise 2.30. Give an example of a non-uniformly convergent sequence f_i on a finite measure space X where again the answer to 14 is no. Hint: on the unit interval, with the usual Lebesgue measure, try to construct a sequence $f_i \to f \equiv 0$ such that $\int f_i \, dx = 1$ for all i.

2.4.2 Types or Modes of Convergence

The above examples look at the question of the connection between pointwise convergence and congergence in norm. But convergence in norm (i.e. $\int |f_i - f|\, dx \to 0$) is not the only alternative to pointwise convergence. Here are the five modes of convergence that are important to know about.

Uniform Convergence We say that f_i converges uniformly to f if

$$\sup_{x \in X} |f_i(x) - f(x)| \underset{i\to\infty}{\rightarrow} 0.$$

Convergence AE If $f_i \to f$ as $i \to \infty$ for all but a measure 0 set of points, we say that f_i is converes to f **almost everywhere (a.e.)**. This is somtimes refered to as poitwise convergence.

Convergence in measure If, for any $\epsilon > 0$ we have that

$$\lim_{i\to\infty} \mu(\{x||f_i(x) - f(x)| \geq \epsilon\}) = 0$$

then we sat that f_i converges to f in measure.

Convergence in norm If $\lim_{i\to\infty} ||f_i - f|| = 0$, where $|| \cdot ||$ is a norm on the function space containing The seqeunce f_i and limit f, then we sat that the f_i converge in norm to f. This is also refered to as **strong Convergence**.

Weak Convergence To define weak convergence, we need the notion of a family of **test functions**. Typically, test functions are functions that are nice or even very nice, like positive C^∞ functions with compact support. We will denote the family of functions by Φ and an individual test function my ϕ.)

We will say that f_i **converges weakly to** f if

$$\lim_{i\to\infty} \int \phi f_i \, dx = \int \phi f \, dx$$

for all test functions $\phi \in \Phi$.

Exercise 2.31. Find an example of a sequence of functions f_i that converges to $f \equiv 0$ in norm even though $f_i(x)$ does not converge to $0 (= f(x))$ for any $x \in X$

Exercise 2.32. Find an example of a sequence of functions f_i that converges pointwise to $f \equiv 0$ (everywhere, not just a.e.), even though $||f_i(x) - f(x)|| = ||f_i(x)|| = \int |f_i|\, dx$ does not converge to 0. (I.e. f_i does not converge in norm to f

Exercise 2.33. Find an example to show that convergence in measure does not imply convergence in norm. Hint: the f_i need not be bounded.

Exercise 2.34. Suppose we choose the norm given by $||g|| = \int |g|\ dx$. Show that if the f_i and f are uniformly bounded (i.e. $-C \leq f_i, f \leq C$ for some $C > 0$), then convergence in measure implies convergence in norm and convergence a.e.

Exercise 2.35. Find an example of a sequence of functions f_i which converge to 0 nowhere, but which do converge weakly to $f \equiv 0$.

Exercise 2.36. **Look at all the posibilities!** Suppose we identify each of the 5 bit binary numbers with a set of convergence types: $f_i \to_{01101} (f \equiv 0)$ would be shorthand for the fact that f_i converges to the zero function a.e., in measure and weakly but not uniformly or in norm. Is it possible to find sequences converging to zero for each of the binary numbers? If not which ones are possible?

2.4.3 The Three Theorems

The next three theorems and the examples that follow tell us that we have to be a bit careful, but that in many useful cases, things go well – we can switch the order of integration and limit taking! First though, we need to introduce the notion of $\liminf f$ and $\limsup f$.

Definition of limsup and liminf Suppose that $f : \mathbb{N} \to \mathbb{R}$. Then the behavior of f as its argument approaches infinity can be complicated. In particular, it might not approach a limit. If we think visually about the sets $F_n \equiv \{f(i)|i \geq n\}$, we could imagine the smallest closed inteval containing F_n – call it I_n – and ask how I_n varies as $n \to \infty$. Then $\liminf f$ and $\limsup f$ are the left and right endpoints of the smallest interval in the range that "eventually" contains f. This is made precise in the following exercise.

Exercise 2.37.

1. Show that $I_i \supset I_{i+1}$ for all i
2. Choose l_i and r_i such that $I_i = [l_i, r_i]$. Show that $l^* \equiv \lim_{i\to\infty} l_i$ and $r^* \equiv \lim_{i\to\infty} r_i$ both exist and that $l^* \leq r^*$.
3. Suppose that $l^* = r^*$. Show that $\lim_{i\to\infty} f(i)$ exists and is equal to $l^* = r^*$.
4. Suppose that $l^* < r^*$. Show that if $l^* < \alpha < r^*$, then for every n there exists $i > n$ such that $f(i) > \alpha$ and a $j > n$ such that $f(j) < \alpha$.

We call the l^* the $\liminf$ and r^* the $\limsup$. By working through the excercise, it becomes clear that the $\liminf_{i\to\infty} f$ and $\limsup_{i\to\infty} f$ define the eventual envelope which contains f's oscillations "at infinity".

We now define $\liminf f$ and $\limsup f$ more concisely:

Definition 2.4 ($\limsup_{i\to\infty} f(i)$ and $\liminf_{i\to\infty} f(i)$)**.** *Suppose that* $f:\mathbb{N}\to\mathbb{R}$.

$$\limsup_{i\to\infty} f \equiv \lim_{n\to\infty}\left(\sup_{i>n} f(i)\right)$$

$$\liminf_{i\to\infty} f \equiv \lim_{n\to\infty}\left(\inf_{i>n} f(i)\right)$$

Exercise 2.38. Rework Exercise 2.37 for functions $f:\mathbb{R}\to\mathbb{R}$, to get the analog notions, $\liminf_{x\to\infty} f$ and $\limsup_{x\to\infty} f$.

Definition 2.5 ($\limsup_{i\to\infty} f_i(x)$ and $\liminf_{i\to\infty} f_i(x)$)**.** *Suppose that* $f_i : X\to\mathbb{R}$ *for some measure space* X*. For a sequence of functions* $f_i(x)$ *we define*

$$\liminf_{i\to\infty} f_i$$

to be the pointwise limit,

$$l(x) = \liminf_{i\to\infty} f_i(x),$$

and we define

$$\limsup_{i\to\infty} f_i$$

to be the pointwise limit,

$$u(x) = \limsup_{i\to\infty} f_i(x).$$

Now we can state the three theorems:

Theorem 2.6 (Fatou's Lemma)**.**

$$\int \liminf_{i\to\infty} f_i \, dx \le \liminf_{i\to\infty} \int f_i \, dx$$

Theorem 2.7 (Monotone Convergence)**.** *Suppose that* $\{f_i\}$ *are all measureable and that* $0\le f_1\le \ldots \le f_i \le f_{i+1}\le \ldots$*. Then we have that*

$$\lim_{i\to\infty}\int f_i\, dx = \int \lim_{i\to\infty} f_i \, dx.$$

Theorem 2.8 (Dominated Convergence Theorem)**.** *If* $f_i\to f$ μ *a.e.,* $|f_i|, |f| < g$ *and* $\int g\, dx < \infty$*, then*

$$\int |f_i - f|\, dx \to 0 \text{ as } i\to\infty.$$

2.4.4 Proofs and Discussion of the Three Theorems

Traditionally, the monotone convergence theorem is shown and then used to prove Fatou's lemma, which is used to prove the dominated convergence theorem. One can also prove Fatou and use that to prove both the monotone convergence and dominated convergence theorems (See Evans and Gariepy's first chapter). We will prove the three theorems by first proving the dominated convergence theorem and then use that theorem to prove the monotone convergence theorem, which in turn will be used to prove Fatou's lemma.

Proof of the Dominated Convergence Theorem.

(i) First we define a new measure $\mu_g(E) \equiv \int_E g\ dx$ whenever E is μ-measurable. For non-measurable F, we define $\mu_g(F) = \inf_{\{E|F\subset E\}} \int_E g\ dx$ where the E are of course measurable. Since g is μ-summable, we have that $\mu_g(X) < \infty$. One can show that every μ-measurable set E is also μ_g-measurable (See exercise 2.39).

(ii) Choose an $\epsilon > 0$. define $E_n = \{x||f(x) - f(x_i)| < \epsilon 2g\ \forall\ i \geq n\}$. We have that the E_i is μ and therefore μ_g measureable for all i. We also have that $...E_{i-1} \subset E_i \subset E_{i+1}$ for all $i \geq 2$. Since $\mu_g(X) < \infty$, we have that $\lim_{i\to\infty} \mu_g(X \setminus E_i) = 0$.

(iii) Choose n big enough that $\mu_g(X \setminus E_i) \leq \epsilon$ and conclude that

$$\begin{aligned}
\int |f - f_i|\ dx &= \int_{X\setminus E_n} |f - f_i|\ dx + \int_{E_n} |f - f_i|\ dx \\
&\leq 2\int_{X\setminus E_n} g\ dx + \int_{E_n} \epsilon 2g\ dx \\
&\leq 2\mu_g(X \setminus E_n) + \epsilon \int 2g\ dx \\
&\leq 2\epsilon + \epsilon 2 \int g\ dx \\
&\leq 2\epsilon(1 + \int g\ dx)
\end{aligned}$$

Because ϵ is arbitrary, we have that $\int |f - f_i|\ dx \to 0$ as $i \to \infty$.

□

Exercise 2.39. Weighted Measures: μ_g from μ

(a) If μ is an outer measure, with measurability determined using Carathrodory's criterion, $g \geq 0$ and $\int g\ d\mu < \infty$, then we can define

$$\mu_g(F) \equiv \inf_{(E\ \mu\text{-measurable}\ ,\ F\subset E)} \int_E g\ d\mu.$$

Prove that μ_g is an outer measure and that μ-measurability implies μ_g-measurability.

(b) Give an example illustratiing why μ_g-measurability does not imply μ-measurability.

(note) The notation $\mu \llcorner g$ is also used to denote μ_g.

Proof of Monotone Convergence Theorem.

(i) If $\int g\ dx < \infty$, use the dominated convergence theorem to get the result.

(ii) If $\int g\ dx = \infty$, then we can find a simple function g_C such that $g_C \leq g$ everywhere and $\int g_C\ dx > C$.

(iii) Define $E_n = \{x | g_i > (1-\epsilon) g_C\ \forall\ i \geq n\}$. Choose n big enough that $\mu_{g_C}(X \setminus E_n) \leq \epsilon$.

(iv) Note that we have

$$\begin{aligned} \int g_i\ dx &\geq \int_{E_n} g_i\ dx \\ &\geq \int_{E_n} (1-\epsilon) g_C\ dx \\ &\geq (1-\epsilon)(C-\epsilon) \end{aligned}$$

Since ϵ is arbitrary and C is a big as we like, we have that $\int g_i\ dx \to \int g\ dx$.

□

Proof of Fatou's Lemma.

(i) Define $h_n(x) = \inf_{i \geq n} f_i(x)$. Note that $\liminf_{i \to \infty} f_i = \lim_{i \to \infty} h_i$.

(ii) Note that $\int h_n\ dx \leq \int f_i\ dx$ for all $i \geq n$. We conclude that $\int h_n\ dx \leq \liminf_{i \to \infty} \int f_i\ dx$ for all n.

(iii) This implies that

$$\begin{aligned} \liminf_{i \to \infty} \int f_i\ dx &\geq \lim_{n \to \infty} \int h_n\ dx \\ &= \int \lim_{n \to \infty} h_n\ dx \text{ (by the monotone convergence theorem)} \\ &= \int \liminf_{n \to \infty} f_n\ dx \end{aligned}$$

□

Remark 2.8. *Using the fact that these three theorems can be proven in the reverse order so that Fatou implies monotone implies dominated, we see that they are in fact equivalent. In the usual path to the proofs of these theorems, we do not need the fact that*

Remark 2.9. *The dominated convergence theorem is really a finite measure "upstairs" thing. Let me explain. First, one can work in the domain of f (the measure space) or the product space of the measure space and the range (the real line), also known as the graph space. By working upstairs, I mean working in the graph space, in the region above (or upstairs) the domain. If we do that, we see that the region of the graph space between $-g$ and g is finite in measure and the dominated convergence theorem is really saying that if all your messing around is done in a constrained, finite measure set, essentially no misbehavior can result.*

Remark 2.10. *Dominated Convergence is used to get other switching theorems: switching order of differentiation and summation or differentiation and integration or integraion and summation.*

2.5 Area, Co-Area, and Sard's Theorem

2.5.1 Lipschitz Functions

Definition 2.6 (**Lipschitz Mappings**)**.** *$F : X \to Y$ is* **Lipschitz** *if there is a positive number $K \geq 0$ such that $|F(x) - F(y)| \leq K|x - y|$ for all $x, y \in X$.*

Radamacher's theorem tells us that a Lipschitz function is differentiable almost everywhere!

Theorem 2.9 (**Radamacher's Theorem**)**.** *If $F : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz, then the set of points at which it fails to be differentiable has measure zero. I.e.* F Lipschitz $\Rightarrow$ F is differentiable almost everywhere.

It turns out that Lipschitz functions are nice enough for many purposes. While differentiability everywhere generally makes proofs easier, often having only Lipschitz smoothness does not stand in the way of various useful (smooth) theorems being true for them as well.

2.5.2 Area and Coarea formulas

The behavior of integrals and volumes under mappings is the focus of the next two highly useful results.

First we consider Lipschitz maps $F : \mathbb{R}^n \to \mathbb{R}^m$ when $n \leq m$. Define $|JF| = \sqrt{\det(DF^T \circ DF)}$, where the T superscript indicates transpose.

In this case we have:

Theorem 2.10 (**Area Formula**).

$$\int_\Omega |JF| d\mathcal{H}^n = \int_{F(\Omega)} \mathcal{H}^0(F^{-1}(y) \cap \Omega) d\mathcal{H}^n y$$

When a Lipschitz $F : \mathbb{R}^n \to \mathbb{R}^m$ when $n \geq m$. Define $|JF| = \sqrt{\det(DF \circ DF^t)}$. Now we have:

Theorem 2.11 (**Coarea Formula**).

$$\int_\Omega |JF| d\mathcal{H}^n = \int_{F(\Omega)} \mathcal{H}^{n-m}(F^{-1}(y) \cap \Omega) d\mathcal{H}^m y$$

We can add functions to get more general results:

Theorem 2.12 (**Area Formula, version 2**).

$$\int_\Omega g(x)|JF| d\mathcal{H}^n x = \int_{F(\Omega)} \left(\int_{F^{-1}(y)\cap\Omega} g(x) d\mathcal{H}^0 x \right) d\mathcal{H}^n y$$

and:

Theorem 2.13 (**Coarea Formula, version 2**).

$$\int_\Omega g(x)|JF| d\mathcal{H}^n = \int_{F(\Omega)} \left(\int_{F^{-1}(y)\cap\Omega} g(x) d\mathcal{H}^{n-m} x \right) d\mathcal{H}^m y$$

While it is not hard to combine both version 2's to get a general area-coarea formula, there is not much advantage to that.

Remark 2.11. *Integrating over $F(\Omega)$ in each of the RHS's of the above formulas is redundant since we are always taking the intersection $F^{-1}(y) \cap \Omega$.*

At first these two results seem rather abstract, but in fact, you have already used them before since they generalize the change of variables formula you have seen for integrals in calculus. To really understand these two formulas, we need to look at simple examples.

Example 2.1 (**Integrating over spheres and then radii**). *Suppose that we want to integrate a function over $\mathbb{R}^n$ by first integrating it over a sphere centered on the origin and then integrating those results over the various radii. Then we can use version 2 of the Coarea Formula and $F = ||x||$ together with the facts that $\nabla F = \frac{x}{||x||}$ and $|JF| = \frac{x}{||x||} \cdot \frac{x}{||x||} = 1$ for all $x \neq 0$ to get*

$$\int_\Omega g(x) d\mathcal{H}^n = \int_0^\infty \left(\int_{\partial B(0,r)\cap\Omega} g(x) d\mathcal{H}^{n-1} x \right) d\mathcal{H}^1 r$$

Example 2.2 (**A Nonlinear Fubini's Theorem**). *The example above of integrating over spheres and then over radii is a special case of integration over distance functions. If we let $h(x) = d(x, K)$ where $d(x, K)$ is the distance from x to the set K, we have that the gradient of d is is a unit vector everywhere except on the interior of K so the Jacobian $|Jd| = 1$ almost everywhere. Our result is then:*

$$\int_\Omega g(x) d\mathcal{H}^n = \int_0^\infty \left(\int_{\{x|d(x,K)=r\}\cap\Omega} g(x) d\mathcal{H}^{n-1}x \right) d\mathcal{H}^1 r$$

Example 2.3 (**Area of graphs**). *If we want to know the n-area (or n-volume) of a graph of $F : \mathbb{R}^n \to \mathbb{R}^1$ over $\Omega \in \mathbb{R}^n$, then we are asking for the n-volume of the set $\{(x, F(x))|x \in \Omega\} \subset \mathbb{R}^{n+1}$. We define the mapping $\hat{F} : \mathbb{R}^n \to \mathbb{R}^{n+1}$ by $\hat{F}(x) = (x, F(x))$. We get that*

$$D\hat{F} = \begin{bmatrix} I_n \\ \nabla F \end{bmatrix},$$

Where ∇F is the row vector of partial derivatives of F. We could compute $\sqrt{\det(D\hat{F}^t \circ D\hat{F})}$ or we can use the fact that this is simply the n-volume of the n columns and use the generalized Pythagorean theorem to compute this from $D\hat{F}$. That theorem says that the square of the n volume of this matrix is equal to the sum of the squared determinates of the $n+1$, $n \times n$ submatrices. When we compute this we get $\sqrt{1 + \nabla F \cdot \nabla F}$. Another way to get this is to change coordinates so that the the gradient only has an x_n component. Then

$$D\hat{F}^t \circ D\hat{F} = \begin{bmatrix} I_{n-1} & v_1 \\ v_1^t & 1 + \nabla F \cdot \nabla F \end{bmatrix}.$$

where v_1 is a column of $n-1$, 0's, and we get the same result. Finally, looking at this purely geometrically, we can also get this result by noticing that the area of a little piece of the graph is increased by exactly the ratio between the hypotenuse of a triangle with horizontal 1, vertical side $||\nabla F||$ and the horizontal side length.

Remark 2.12 (**In Class Pictures!**). *I will give an intuitive explanation of both the area and coarea formulas in class. Eventually the pictures and explanations will appear in the notes as well.*

2.5.3 Sard's Theorem

It is clear that the measure of points in the domain where the rank of a mapping is not full can be large. In fact, simply using the 0 mapping gets you a mapping whose rank is never full on the entire domain. This raises the point, what is the measure of the points in the range that come from points in the domain where the rank is not full?

The answer now is not very much: to be more precise, only a set of measure zero.

Theorem 2.14 (**Sard's Theorem**). *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ and that F is C^k with $k \geq n-m+1$. Define $\mathcal{C}$ to be the set of points $x \in \mathbb{R}^n$ such that* $\text{rank}(DF_x) < m$*. Then $\mathcal{H}^m(F(\mathcal{C})) = 0$.*

This theorem is a technical tool extensively used in analysis and geometric analysis. It justifies the intuition that when the rank of the derivative is less than n, so that the derivative is not onto, then the mapping squeezes space down, collapsing at least one dimension, yielding a measure zero set.

Most of the typical proof of this result is not very enlightening, with the exception of the last part in which you show that the measure of the image of $\mathcal{C}_k$, the points where all partial derivatives of order k and below, is zero. The argument uses Taylor's theorem to show that the image of a cover of $\mathcal{C}_k$ must be reduced in volume to a volume that behaves like $\delta^{k+1-\frac{n}{m}}$ where δ is the edge length of a cubical grid that is going to zero as we choose finer and finer discretizations. The first part of the proof is an inductive argument. See chapter 3 of Milnor's little book on differential topology for all the details [1].

3 Pause: A return to the Three Integrals in Section (1.5)

4 Other Topics

4.1 Fixed point theorems: Banach Fixed Point Theorem

Many problems can be written as:

Problem 4.1 (**Finding Fixed Points**). *Given a mapping F from a space X to itself, $F : X \to X$, find x^* such that $F(x^*) = x^*$.*

We will look at one theorem that gives the existence of unique fixed points. First we have to introduce the idea of a Banach space.

Definition 4.1 (**Vector Space Norm**). *Suppose that $\alpha \in \mathbb{R}$ and $x, y \in X$, X a vector space. Then a function from $||\cdot|| : X \to [0, \infty)$ is a norm if is satisfies:*

1. $||x|| > 0$ *when* $x \neq 0$
2. $||\alpha x|| = |\alpha| ||x||$
3. $||x + y|| \leq ||x|| + ||y||$ *(the triangle inequality)*

Definition 4.2 (**Cauchy Sequence**). *Recall that $x_i \in X$ is Cauchy if for any $\epsilon > 0$ there is an $N(\epsilon)$ such that $i, j > N(\epsilon)$ implies that $||x_i - x_j|| < \epsilon$.*

Definition 4.3 (**Complete Space**). *If every Cauchy sequence in X has a limit in x, the X is complete. I.e. if $\{x_i\}_{i=1}^{\infty}$ is Cauchy, then there must also be a point $x^* \in X$ such that $||x_i - x^*|| \to 0$ as $i \to \infty$.*

Definition 4.4 (**Banach Space**). *A complete, normed vector space B is called a* **Banach Space**.

Definition 4.5 (**Contraction Mapping**). *A function from a normed space X to itself is a* **contraction mapping** *if $||F(x) - F(y)|| < k||x - y||$ for some $0 \leq k < 1$.*

Note that a contraction mapping is a special case of a Lipschitz mapping.

Theorem 4.1 (**Banach Fixed Point Theorem**). *Suppose that $F : B \to B$ is a contraction mapping from the Banach space B to itself. Then there is a unique point x^* such that $F(x^*) = x^*$.*

Proof.
First note that if here are two distinct fixed points x^* and y^* then $||x^* - y^*|| = ||F(x^*) - F(y^*)|| < k||x^* - y^*||$ with $k < 1$ which is a contradiction. so there cannot be more than one fixed point. To prove that there is a fixed point

1. choose any $x_0 \in B$ and define $x_1 = F(x_0)$, $x_2 = F(x_1) = F(F(x_0)) = F^2(x_0)$ and likewise $x_n = F^n(x_0)$.

2. We note that x_i is a Cauchy sequence:

   (a) $||F^{i+1}(x_0) - F^i(x_0)|| \leq k^i||F(x_0) - x_0||$

   (b) for $n > m$

$$\begin{aligned}
||x_n - x_m|| &= ||F^n(x_0) - F^m(x_0)|| \\
&\leq \left(\sum_{i=m}^{n-1} k^i\right)||F(x_0) - x_0|| \\
&= k^m(\sum_{i=0}^{n-m-1} k^i)||F(x_0) - x_0|| \\
&\leq k^m(\sum_{i=0}^{\infty} k^i)||F(x_0) - x_0|| \\
&= \frac{k^m}{1-k}||F(x_0) - x_0||.
\end{aligned}$$

   So, as long as $n, m > N$ we have that

$$||F^n(x_0) - F^m(x_0)|| \leq \frac{k^N}{1-k}||F(x_0) - x_0|| \underset{N\to\infty}{\to} 0$$

(c) Thus, $\{x_i\}_{i=1}^{\infty}$ is a Cauchy sequence.

3. Therefore, there is a point x^* in B such that $x_i \to x^*$ as $i \to \infty$.

4. Since F is continuous, we have that $\lim_{i\to\infty} F(x_i) = F(\lim_{i\to\infty} x_i) = F(x^*)$. But $F(x_i) = x_{i+1}$ so $\lim_{i\to\infty} F(x_i) = \lim_{i\to\infty} x_{i+1} = x^*$. Thus $F(x^*) = x^*$.

□

4.2 Transversality is Generic

Intersections of submanifolds of various dimensions are encountered all the time; one can, for instance, look at $Ax = b$ where A is an $m \times n$ matrix, as a statement of a problem of finding a point (or all points) in the intersection of m, n-1-dimensional subspaces of $\mathbb{R}^n$. We are also often interested in how stable our problem is to perturbations. What can we say about some problem if we add a bit of noise, or jiggle some parameters a tiny bit?

For these questions, the key concept is the idea of **transverse intersection of subspaces**.

Definition 4.6 (Transverse Intersection of Subspaces). *Two subspaces of* $\mathbb{R}^n$, U_k *and* W_m *of dimension* k *and* m *respectively, are said to intersect transversely if the* $\text{span}(U_k, W_m) = \mathbb{R}^n$.

This leads directly to the idea of transverse intersections of submanifolds:

Definition 4.7 (Transverse Intersection of Submanifolds). *Two submanifolds of* $\mathbb{R}^n$, M *and* N, *intersecting at* x *are said to intersect transversely at* x *if the tangent spaces* T_xM *and* T_xN *intersect transversely as subspaces of* $\mathbb{R}^n$, *I.e. if* $\text{span}(T_xM, T_xN) = \mathbb{R}^n$.

Example 4.1 (2 Curves in $\mathbb{R}^3$). *In* $\mathbb{R}^3$, *an intersection between 2, 1-manifolds is never transverse.*

Example 4.2 (A 1-Curve and a 2-Surface in $\mathbb{R}^3$). *In* $\mathbb{R}^3$, *an intersection between a 2-dimensional surface and a 1-dimensional curve is transverse if and only if the curve is not tangent to the surface at the point of intersection.*

Example 4.3 (2 arbitrary submanifolds). *If* M_k *and* N_p *are* k *and* p *dimensional submanifolds of* $H = \mathbb{R}^n$, *then they intersect transversely if in a neighborhood of the intersection point* $x \in M_k \cap N_p$, *we have that* $\dim(M_k \cap N_p) = \dim(M_k) + \dim(N_p) - \dim(H) = p + k - n$.

Transverse intersections are stable: if we take an arbitrary intersection between arbitrary compact submanifolds, then if it is not transverse it can be made transverse using an arbitrarily small perturbation. If on the other hand the intersection is transverse, then any perturbation of small enough magnitude will not change that fact.

5 Problems for the Course

5.1 Geometry of Differentiation

Problem 5.1. Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that f is differentiable at $x = a$.

1. Show that, given an angle θ, we can choose $\delta(\theta) > 0$ small enough so that for all x such that $|x - a| < \delta(\theta)$ we have that the graph of $f(x)$ lies inside of the cone with angle θ around the tangent line. (See Figure 9.)

2. Can you find explicit formulas for $\delta(\theta)$ for the function $f(x) = c_1 x^2 + c_2 x + c_3$ for any arbitrary a?
   Hints: (a) first solve for $g(h) = f(a + h) - f(a) - L_a(h)$ where, of course $L_a(h) = f'(a)h$

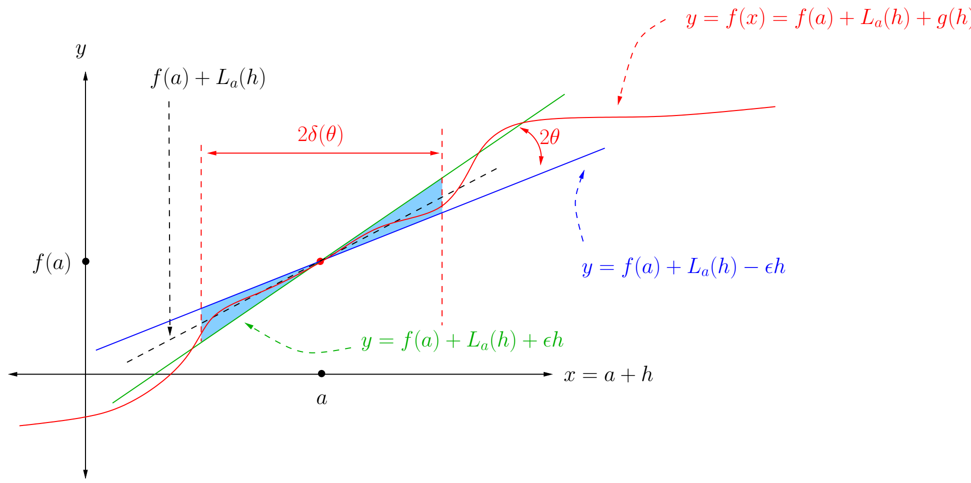


Figure 9: **(For Problem 5.1)** Here is a picture to stimulate your thoughts and explorations. $L_a(h)$ is a linear function from $\mathbb{R}$ to $\mathbb{R}$ – a line through the origin.

(b) Prove that a triangle that is obtained by a base of length h and a constant (horizontal, in the figure) height of L has maximal apex angle when the base is bisected (alternatively, the apical angle is bisected) by the x-axis. (See Figure 10.) (One way to do that is show that the maximal area underneath the curve $y = \frac{d(\arctan(x))}{dx} = \frac{1}{1+x^2}$, over an interval of length h is obtained when that interval is centered on the origin.)

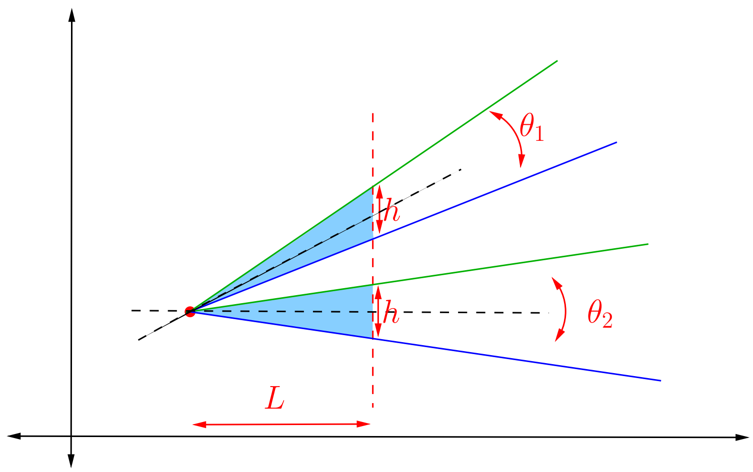


Figure 10: **(For Problem 5.1)** A triangle whose height (sideways height in this picture) is L and base is a constant h has a maximal angle as the apex (the point furthest to the left) when that apex is bisected by the x-axis. I.e. You are trying to show that $\theta_2 > \theta_1$.

5.2 Higher Order Derivative Implications

Definition 5.1 (Defintion of Lipschitz – Reminder). *If $f : E \subset X \to Y$ and*

$$|f(x_2) - f(x_1)| \leq C \ |x_1 - x_2|$$

for all $x_1, x_2 \in E$ and some $0 \leq C < \infty$ then we say f is Lipschitz continuous *or simply* Lipschitz*, with Lipschitz constant C.*

Problem 5.2. (Harder): Suppose that, $f''(a)$, the second derivative of $f : \mathbb{R} \to \mathbb{R}$ at $x = a$, exists. Show that there is some interval around a, $[a - \delta, a + \delta]$ on which f is Lipschitz. **Hints**: first show that in some interval $(a - 2\delta, a + 2\delta)$, the derivative exists and is bounded. Then, for every point in that interval, deduce that there is a narrow cone that works for a possibly tiny interval around it. Get a finite open cover of $[a - \delta, a + \delta]$ using those small intervals and deduce the desired conclusion.

5.3 Lebesgue Measure

Problem 5.3. (Lebesgue Measure of Rectangles = Their Content)

1. Show that $\mathcal{L}^d(S) = 0$ for any $S = x \in \mathbb{R}^d$ such that $x_i = c$ – the d-1-dimensional plane obtained by holding the i-th coordinate constant.

2. Show that if $\mu(D) = 0$, then $\mu(C \cup D) = \mu(C)$

3. Show that the measure of any rectangle when some or all of the intervals defining it are not open is the same as the corresponding open rectangle.

4. Show that for any rectangle $R \subset \mathbb{R}^d$,

$$\mathcal{L}^d(R) = c(R).$$

Hint: take the closed rectangle $\bar{R}$ corresponding to R and notice that any cover with open rectangles has a finite subcover also covering $\bar{R}$. This is not a trivial exercise, so beware of trivial arguments.

5.4 Hausdorff Measure

Problem 5.4. suppose that $S = [a, b] \times \{0\}$ is a closed line segment in $\mathbb{R}^2$. Show that $\mathcal{H}^1(S) = b - a$. **Hint:** See Exercises 2.14, 2.15 and 2.17.

5.5 Integration and Differentiation

Problem 5.5. Read Section 5.12 of Fleming's book and do problem 4 in the same section which asks you to prove Leibniz's rule for differentiation of definite integrals. This will give you some experience with when it is OK to switch the the order in which you do differentiation and integration.

6 Warm up Problems

In this section, I will create a growing list of *relatively* simple problems you can use to get yourself moving, to build momentum you can bring to the harder problems and exercises in the course. Some may require calculation and messing around – if algebraic manipulations are not easy for you, this will help you get better at those too!

Exercise 6.1. Show that if we approximate $\sqrt{1+\delta} \approx 1 + \frac{\delta}{2}$, then the magnitude of the error in squares does not exceed $\frac{\delta^2}{4}$, I.e. that:

$$\left| \left(\sqrt{1+\delta}\right)^2 - \left(1 + \frac{\delta}{2}\right)^2 \right| \le \frac{\delta^2}{4}.$$

Exercise 6.2. We deal with $\delta > 0$:

1. Show that for $0 \leq \delta \leq 2$, we actually have:

$$1 + \frac{\delta}{2} - \frac{\delta^2}{8} \leq \sqrt{1+\delta} \leq 1 + \frac{\delta}{2}.$$

2. Show that this implies that $\left|\sqrt{1+\delta} - \left(1 + \frac{\delta}{2}\right)\right| = \{$the error in assuming $\sqrt{1+\delta} \approx 1 + \frac{\delta}{2}\}$, is at most $\frac{\delta^2}{8}$ (as long as $0 \leq \delta \leq 2$).

Note: this implies approximation is very good as long as $0 \leq \delta << 1$.

Exercise 6.3. Now we want to deal with $\delta < 0$.

1. Use the mean value theorem to prove that:

$$\sqrt{1+\delta} - 1 = \frac{1}{2\sqrt{1+c}}\delta$$

   for some $c \in (\delta, 0)$. Rearranged this says:

$$1 + \frac{1}{2\sqrt{1+c}}\delta = \sqrt{1+\delta}.$$

2. Now prove that because $\delta < 0$ we have that

$$1 + \frac{1}{\sqrt{1+\delta}}\frac{\delta}{2} \leq \sqrt{1+\delta}.$$

3. Show that, in fact:

$$1 + \frac{1}{1+\delta}\frac{\delta}{2} \leq 1 + \frac{1}{\sqrt{1+\delta}}\frac{\delta}{2} \leq \sqrt{1+\delta} \leq 1 + \frac{\delta}{2}$$

4. Now use this last inequality to show that:

$$\left|\sqrt{1+\delta} - \left(1 + \frac{\delta}{2}\right)\right| \leq \frac{1}{2}\frac{\delta^2}{1+\delta}$$

5. Conclude that as long as $-\frac{1}{2} \leq \delta \leq 0$, we have

$$\left|\sqrt{1+\delta} - \left(1 + \frac{\delta}{2}\right)\right| \leq \delta^2$$

Exercise 6.4. Collect the results from the Exercises (6.1 - 6.3) to conclude that:

$$\text{If } |\delta| \le \frac{1}{2}$$

then the error in assuming $\sqrt{1+\delta} \approx 1 + \frac{\delta}{2}$ is at most δ^2:

$$\left| \sqrt{1+\delta} - \left(1 + \frac{\delta}{2}\right) \right| \le \delta^2$$

Remark 6.1. Question*: why is this approximation valuable?* **Answer***: because, very often, $1 + \frac{\delta}{2}$ is (algebraically) much easier to work with than $\sqrt{1+\delta}$.*

Exercise 6.5. One final exercise on the approximation of $\sqrt{1+\delta}$: Use the Taylor series for $\sqrt{1+\delta}$ with a second derivative error term to conclude that for $\delta \in (-\frac{1}{2}, \frac{1}{2})$:

$$\left| \sqrt{1+\delta} - \left(1 + \frac{\delta}{2}\right) \right| \le \frac{1}{4}\delta^2.$$

Exercise 6.6. Show that

$$1 + x \le e^x$$

for all $x \in \mathbb{R}$.

Exercise 6.7. Show that for $x \in (\infty, 0.5)$ we have:

$$1 + x \le e^x \le 1 + x + x^2.$$

Hint: Split the argument into two pieces – one for $x \in (\infty, 0)$ and one for $x \in [0, 0.5]$. Now reason carefully with inequalities involving the derivatives.

Exercise 6.8. Now conclude that for $|x| < 0.5$ (Actually $|x| < \ln(2)$ works) we have:

$$|e^x - (1+x)| < x^2$$
$$1 \le \frac{e^x}{1+x} \le 1 + \frac{x^2}{1+x}$$
$$1 \le \frac{e^{nx}}{(1+x)^n} \le \left(1 + \frac{x^2}{1+x}\right)^n$$

Remark 6.2. *The fact that*

$$1 + x \le e^x$$

is often used for $x = -\epsilon$, where $0 < \epsilon \ll 1$, to get

$$(1-\epsilon)^n \le e^{-n\epsilon}$$

where $1 - \epsilon$ is the probability of an almost sure event.

Exercise 6.9. Prove that for $x \in \mathbb{R}$ and $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Recall the definition of the norm of an operator given in Definition (1.6).

Exercise 6.10. Suppose that $A : \mathbb{R}^n \to \mathbb{R}^n$ is linear and $|A| < 1$. Prove that there is an operator B (It is no longer linear) such that

$$B = I + A + A^2 + A^3 + A^4 + ...$$

and that

$$B = (I - A)^{-1}$$

I.e. B is the inverse of the operator $I - A$, where $I : R^n \to \mathbb{R}^n$ is the identity map.

Hint: show that for any $x \in \mathbb{R}^n$, the series

$$S_k(x) \equiv \left(I + A + A^2 + ... + A^k\right)(x) = x + Ax + A^2x + ... + A^kx$$

converges to a point in $\mathbb{R}^n$. Now define

$$B(x) = \lim_{k\to\infty} S_k(x).$$

Now compute $(I - A) * S_k(x)$ and see what happens when $k \to \infty$.

Exercise 6.11. Now show that

1. for any linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$,

$$e^{tA}(x) \equiv \left(I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \frac{t^4}{4!}A^4 + ...+\right)(x)$$

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

2. Defining:

$$S_A^k(t, x) \equiv \left(I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + ... + \frac{t^k}{k!}A^k\right)(x)$$

   use the fact that you know how to compute $\frac{dS_A^k(t,x)}{dt}$ to show that it makes sense to say the solution to:

$$\dot{x}(t) = Ax(t)$$

   is $x(t) = e^{tA}x(0)$. One main point to notice is that we do not need a bound on the norm of the operator A.

3. **(hard)** There is a detail here that is non-trivial: how can we show that

$$\frac{d}{dt}\left(\lim_{k\to\infty} S_A^k(t,x)\right) = \lim_{k\to\infty}\left(\frac{d}{dt}S_A^k(t,x)\right)?$$

   It turns out that this is true in this case and you can go ahead and assume it, but see if you can make progress in figuring out what must be true to get this switch to work.

   Hint: when considering whether or not $\frac{d}{dt}\left(\lim_{k\to\infty} f_k(t,x)\right) = \lim_{k\to\infty}\left(\frac{d}{dt}f_k(t,x)\right)$ you care about how the rates of convergence of $\lim_{k\to\infty} f_k(x,t)$ depend on t. You can also stare at

$$\lim_{h\to 0}\frac{1}{h}\left(\lim_{k\to\infty} f_k(x,t+h) - \lim_{k\to\infty} f_k(x,t)\right).$$

Exercise 6.12. Suppose that $f(x) > 0$ is continuous for all x and we define $A(d,x^*) = \int_{x^*}^{x^*+d} f(x)dx$. Show:

1. $A(d,x^*)$, as a function of x^*, is just the area under the curve over a the fixed length interval $[x^*, x^*+d]$ that slides along the x-axis as we change x^*,
2. That:

$$\frac{dA(d,x)}{dx} = f(x+d) - f(x),$$

Exercise 6.13. Use the results of Exercise 6.12 to show that if we define:

$$\begin{aligned}\theta_-^d(x) &\equiv \arctan(x) - \arctan(x-d)\\ \theta_+^d(x) &\equiv \arctan(x+d) - \arctan(x)\end{aligned}$$

and remember that

$$\arctan(x) = \int_0^x \frac{1}{1+s^2}ds$$

we can conclude that

$$\theta_+^d(x) \le \theta_-^d(x)$$

for all $x \ge 0$.

Exercise 6.14. Continue with Exercise 6.13, again using the results of Exercise 6.12 to prove that

$$\theta_+^d(-d/2) \ge \theta_+^d(x) \ \ \forall x.$$

Exercise 6.15. (Implicit Function Theorem Exercise) Define $z = f(x,y) = x^2 - y^2$. For what values of c is the c-level set $L_c = \{(x,y) \mid f(x,y) = c\}$ not regular? Find the points (x^*,y^*) in each regular level set L_c such that **either** $f(x,h(x)) = c$ **or** $f(g(y),y) = c$ does not hold near (x^*,y^*). See Section 1.9.2.

References

[1] John W. Milnor. *Topology From The Differentiable Viewpoint.* University Press of Virginia, Charlottesville, 1965. Eighth printing, 1990.