KRV's GQE+ Boot Camp: Notes and Problems

Kevin R. Vixie

November 27, 2019

Contents

Chapter 1

Introduction

1.1 Texts and References

I recommend the following documents and texts:

1. the old test packet, available from Jessica Cross
2. *Linear Algebra*, 4th Edition by Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence
3. *Numbers and functions: steps into analysis* by R.P. Burn
4. *Functions of Several Variables* by Wendell Fleming
5. *Calculus with Analytic Geometry* 1979 edition, by Earl William Swokowski or any book with traditional vector calculus chapters including contour integration, divergence theorem, Stokes theorem, and 3 dimensional integration.
6. *Oxford User's Guide to Mathematics* 2004, edited by E Zeidler. This is an excellent reference book that is quite rich in information, partly because of the editor. (Same Author as the 5 volume nonlinear functional analysis series.)
7. problems that we come up with and hand out.
8. ... there are other references I comment on in Chapter 4.

1.2 Tricks of the trade

Making these (and other) tricks and approximations subconsciously available is an important task. Think of this list as a prompt to write your own similar list.

1. $f(x) = o(x)$ means that $f(x) = xh(x)$ and $h(x) \to 0$ as $x \to 0$; $f(x) = O(x)$ means that $f(x) = xh(x)$ and $h(x) \le C < \infty$ for $x \in (-a, a)$;
2. $\sqrt{1+\epsilon} = 1 + \frac{\epsilon}{2} + o(\epsilon) \approx 1 + \frac{\epsilon}{2}$
3. if $x \in (-1, 1)$ then $1 + x + x^2 + ... = \frac{1}{1-x}$; $1 + x + x^2 + ... + x^n = \frac{1-x^{n+1}}{1-x}$

4. Jensen's inequality; $\alpha_i \geq 0 \; \forall i$ and $\sum \alpha_i = 1$; if f is convex, then $f(\sum \alpha_i x_i) \leq \sum \alpha_i f(x_i)$.

5. Cauchy-Schwarz inequality; for $v, w \in \mathbb{R}^n$ or some Hilbert space, $|v \cdot w| \leq |v| \; |w|$; understanding that this is just $0 \leq (v - w) \cdot (v - w)$ when $|v| = |w| = 1$.

6. Minkowski's inequality in a Hilbert space; where $|x| = \sqrt{x \cdot x}$, using the C-S inequality, we get that
$|x + y|^2 = (x + y) \cdot (x + y) = x \cdot x + 2x \cdot y + y \cdot y \leq |x|^2 + 2|x| \; |y| + |y|^2 = (|x| + |y|)^2$ which is Minkowski's inequality – $|x + y| \leq |x| + |y|$.

7. Holder's inequality; if $\frac{1}{p} + \frac{1}{q} = 1$, then $|fg|_1 \leq |f|_p \; |g|_q$ where $|f|_\alpha = (\int |f|^\alpha dx)^{\frac{1}{\alpha}}$ for $1 \leq \alpha < \infty$.

8. am-gm inequality; $a_1^{p_1} a_2^{p_2} a_3^{p_3} \cdots a_n^{p_n} \leq p_1 a_1 + p_2 a_2 + p_3 a_3 + \cdots + p_n a_n$; special cases are $\sqrt{ab} \leq \frac{1}{2}(a + b)$, and $\sqrt[n]{a_1 a_2 \cdots a_n} \leq \frac{1}{n}(a_1 + a_2 + \cdots a_n)$

9. Triangle inequality: $||x| - |y|| \leq |x \pm y| \leq |x| + |y|$

10. integrating inequalities, remembering that you cannot differentiate inequalities; i.e. $f(x) \leq g(x) \not\Rightarrow f'(x) \leq g'(x)$.

11. $(a - b)^2 > 0$ with equality only when $a = b \Rightarrow$ we have that $ab \leq \frac{a^2 + b^2}{2}$ with equality only when $a = b$.

12. $|(f(x) - f(a)) - df_a(x - a)| \leq o(|x - a|)$; understanding the cone picture that goes with this.

13. $f(x) > 0$, $\alpha(x) > 0$ and $-\alpha(x) f(x) \leq f'(x) \leq \alpha(x) f(x) \Rightarrow$ implies that
$f(a) \; e^{\int_a^b -\alpha(x) \; dx} \leq f(b) \leq f(a) \; e^{\int_a^b \alpha(x) \; dx}$.

14. if $f : \mathbb{R} \to \mathbb{R}$ and f is differentiable everywhere, then $f(b) - f(a) = f'(c)(b - a)$ for some $c \in (a, b)$;

15. (continued) in higher dimensions ($f : \mathbb{R}^n \to \mathbb{R}^m$ and $f \in C^1$) we have that if $\gamma : [a, b] \to \mathbb{R}^n$ and $|\dot{\gamma}(s)| = 1$ for all s, then $\int_a^b Df(\dot{\gamma}(s)) \; ds \leq \int_a^b |Df_{\gamma(s)}| ds \leq (\max_{[a,b]} |Df_{\gamma(s)}|) \; |b - a|$. So there is a $c \in [a, b]$ such that $|f(\gamma(b)) - f(\gamma(a))| \leq |Df_\gamma(c)| \; |b - a|$. (Actually, there will be a $c \in (a, b)$ such that $|f(\gamma(b)) - f(\gamma(a))| \leq |Df_\gamma(c)| \; |b - a|$.)

16. (Continued) Notice that if $x, y \in \mathbb{R}^n$, $|b - a| = |y - x|$ and we choose $\gamma(s) = x + \frac{y - x}{|y - x|}(s - a)$ then $|y - x| = |\gamma(b) - \gamma(a)| = |b - a|$ so that we also have that $|f(\gamma(b)) - f(\gamma(a))| \leq |Df_\gamma(c)| \; |\gamma(b) - \gamma(a)|$.

17. $\sin(\epsilon) = \epsilon + O(\epsilon^3)$; $\cos(\epsilon) = 1 - \epsilon^2 + O(\epsilon^4)$

18. using Taylor series; truncating them, integrating them, etc

19. $| \int f(x) \; dx| \leq \int |f(x)| \; dx$

20. $1 + x \leq e^x$

21. $\lim_{n \to \infty} (1 + \frac{c}{n})^n = e^c$

22. $\dot{x} = \alpha x \to x(t) = x(0) e^{\alpha t}$

23. SVD as the Swiss army knife of L^2 linear algebra; geometry, norms, determinants, etc.

24. $\ln(x+1) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ...$

25. Taylor Series for $\sin(x)$, $\cos(x)$ and e^x.

26. Derivatives:

    (a) $\frac{d}{dx}\sin(x) = \cos(x)$

    (b) $\frac{d}{dx}\cos(x) = -\sin(x)$

    (c) $\frac{d}{dx}\tan(x) = \frac{1}{\cos^2(x)} = \sec^2(x)$

    (d) $\frac{d}{dx}\arcsin(x) = \frac{1}{\sqrt{1-x^2}}$

    (e) $\frac{d}{dx}\arccos(x) = \frac{-1}{\sqrt{1-x^2}}$

    (f) $\frac{d}{dx}\arctan(x) = \frac{1}{1+x^2}$

27. Use $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ and $e^{i\theta_1}e^{i\theta_2} = e^{i(\theta_1+\theta_2)}$ to remember formulas for $\sin(\theta_1+\theta_2)$, $\cos(\theta_1+\theta_2)$: $\sin(\theta_1+\theta_2) = \sin(\theta_1)\cos(\theta_2)+\sin(\theta_2)\cos(\theta_1)$ and $\cos(\theta_1+\theta_2) = \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)$ ... anything else like half angle and double angle and tangent formulas follow ...

28. Quadratic formula and implications: $ax^2 + bx + c = 0 \Rightarrow x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$ and we get that the positivity of $D = b^2 - 4ac$ gives us the number of roots, $\{D < 0 \to$ none, $D > 0 \to 2$ and $D = 0 \to 1\}$

29. Suppose A is an n-by-n matrix. If $A = A^T$ (A is symmetric), then

    (a) A is diagonalizable,

    (b) The eigenvectors form a complete basis for $\mathbb{R}^n$. There can be ambiguity here, but an orthogonal basis can always be chosen, even if there are choices that are not orthogonal.

    (c) The product of all the eigenvalues is the determinant of A: $\Pi_{i=1}^{n}\lambda_i = \det(A)$

    (d) if A is also positive definite, all the eigenvalues of A are positive and the graph of the function $F(x) = x^T Ax$ is a paraboloid opening up (i.e. $F :\to \infty$ as $|x| \to \infty$.

30. $\det(AB) = \det(A)\det(B)$

31. if $x \in \mathbb{R}^n$ then all other $y \in \mathbb{R}^n$ can be written as $y = \alpha x + \beta w_y$ where $x \cdot w_y = 0$. (The subscript on w_y reminds us that w_y can of course depend on y.)

32. If A is an n-by-n matrix and the operator norm of A is less than 1: $|A| < 1$, then we have that $(I - A)^{-1} = I + A + A^2 + A^3 + ...$ For any n-by-n matrix. we have that $e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \frac{A^4}{4!} + \cdots$ and $e^{tA} = I + tA + \frac{t^2A^2}{2!} + \frac{t^3A^3}{3!} + \frac{t^4A^4}{4!} + \cdots$ make sense.

33. If O is a matrix of k orthonormal columns in $\mathbb{R}^n$ (so O is an n-by-k matrix), the matrix $P_O \equiv OO^T$ is the matrix which projects $\mathbb{R}^n$ onto the span of O. As a special case, if $|x| = 1$ and x is a column vector in $\mathbb{R}^n$, then xx^T is the rank-1, n-by-n projection matrix projecting $\mathbb{R}^n$ onto the span of x, and $I - xx^T$ is the projection onto the orthogonal complement of x.

34. When we are using test functions ϕ (or any other function that vanishes, together with its derivatives, on the boundary of our region of integration Ω), then integration by parts gives us: $\int_\Omega f'\phi \; dx = -\int_\Omega f\phi' \; dx$. (We call g **the weak derivative of f** if, for all ϕ compactly supported in the (open) region of integration Ω, we have that $\int_\Omega g\phi \; dx = -\int_\Omega f\phi' \; dx$. A function is **compact supported** in the open set U if the closure of $\{x|f(x) \neq 0\}$ is compact and contained in U.)

1.3 Comments

1. Your aim should be to figure out how to prove, how to calculate, how to solve by understanding the underlying why's. Very often early experiences with mathematics boil down to students being handed how to's and then practicing doing without ever understanding the why's. This is very, very undesirable and not very useful.

Chapter 2

Analysis

2.1 Metric spaces

Metric spaces are ubiquitous in analysis – you are already well acquainted with important examples of them: the real line, the spaces $\mathbb{R}^n$, and even spaces of functions, though you may have not thought of collections of functions as spaces in the same way that you think of $\mathbb{R}^n$ as a space of points.

2.1.1 Metrics

A metric space is a set of point X and a metric ρ which defines distances between points in the space and satisfies three simple properties:

1. $\rho(x, y) \geq 0$; $\rho(x, y) = 0 \Leftrightarrow x = y$
2. $\rho(x, y) = \rho(y, x)$
3. $\rho(x, z) \leq \rho(x, z) + \rho(y, z)$

Exercise 2.1.1. Let $X = \mathbb{R}$, $x, y \in X$, and $|x - y|$ denote the absolute value of the difference $x - y$. Prove that defining $\rho(x, y) = |x - y|$ turns the real line $\{\rho, X\} = \{|\cdot|, \mathbb{R}\}$ into a metric space.

Exercise 2.1.2. Let $X = \mathbb{R}^n$, $x, y \in X$, and $\rho(x, y) = |x - y| \equiv |x_1 - y_1| + \cdots + |x_n - y_n|$. Prove that $\{\rho, X\}$ is a metric space.

Exercise 2.1.3. (*) Let $X = \mathbb{R}^n$, $x, y \in X$, and $\rho(x, y) = |x - y| \equiv \sqrt{|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2}$. Prove that $\{\rho, X\}$ is a metric space. Hint: you need to show that $|x \cdot y| \leq |x||y|$.

Exercise 2.1.4. Let $X = \mathbb{R}$ and $\rho(x, y) = 1$ if $x \neq y$ and $\rho(x, y) = 0$ when $x = y$. Show that $\{\rho, X\}$ is a metric space.

Exercise 2.1.5. Let $\{\rho, X\}$ be a metric space and $Y \Subset X$. Show that $\{\rho, Y\}$ is a metric space: i.e. the metric, restricted to the subset, is a metric on that subset.

Exercise 2.1.6. Let X be a set of three points x,y, and z. Give a precise description of all the metrics on this space of points. Hint: think of it as a subset on $\mathbb{R}^3$.

Exercise 2.1.7. Let $|v| = \sqrt{|v_1|^2 + |v_2|^2}$ for $v \in \mathbb{R}^2$. Suppose that we define $\rho_w(x, y) \equiv \inf_\gamma \int_a^b w(\gamma(s))|\gamma'(s)|\, ds$, where we minimize over Lipschitz $\gamma : \mathbb{R} \to \mathbb{R}^2$, $w(x) > 0$ and is continuous for all $x \in \mathbb{R}^2$. Show that $\{\rho_w, \mathbb{R}^2\}$ is always a metric space

Exercise 2.1.8. Define X to be the space of continuous functions on the closed unit interval. We denote this space by $C([0, 1])$. Define $\rho(x, y) \equiv \max_{x\in[0,1]} |x(t) - y(t)|$. Show that $\{\rho, C([0, 1])\}$ is a metric space.

2.1.2 Open Balls, Closed Balls

Let $\epsilon > 0$. Define $B(x, \epsilon) = \{y \in X \mid \rho(x, y) < \epsilon\}$ and $\bar{B}(x, \epsilon) = \{y \in X \mid \rho(x, y) \leq \epsilon\}$. We call $B(x, \epsilon)$ the open ball of radius ϵ centered at x and $\bar{B}(x, \epsilon)$ the closed ball of radius ϵ centered at x.

Exercise 2.1.9. Let D be an open ball. Suppose that $x \in D$. Show there is a $\delta_x > 0$ small enough that $B(x, \delta_x) \subset D$.

Exercise 2.1.10. What are the open balls in the discrete metric space described in Exercise 2.1.4? Which δ's imply $B(x, \delta) \neq \bar{B}(x, \delta)$?

Exercise 2.1.11. (*) Recall the path based metric from Exercise 2.1.7 and suppose that $w(x) = 1$ for $x_2 \geq 0$ and $w(x) = 2$ for $x_2 < 0$. Find the shortest path from the point $(0, 1)$ to $(2, -1)$. Find the open ball of radius 1 centered at the origin $(0, 0)$

Exercise 2.1.12. Recall the metric space in Exercise 2.1.8. What functions are in the unit ball around the function $x(t) = 1 \forall t \in [0, 1]$?

Exercise 2.1.13. Continuing Exercise 2.1.12, suppose that $x_k(t) \equiv t^k$ what does $B(x_1, 1) \cap B(x_2, 1)$ look like? Can you give a precise, geometric description: i.e. can you draw it?.

Exercise 2.1.14. Continuing Exercise 2.1.13, what can you say about $B(x_1, \frac{1}{8}) \cap B(x_2, \frac{1}{8})$ look like?

Exercise 2.1.15. What do the unit balls look like in the metric space introduced in Exercise 2.1.2 when n = 2 and n=3?

2.1.3 Open sets, limit points, closed sets, closure, interior, exterior, boundary

In a metric space, we will say that a set U is **open** if, for every $x \in U$, there is an open ball $B(x, \epsilon) \subset U$. We will say that x is a **limit point** of a set C if there is a sequence $\{y_i\}_{i=1}^{\infty}$ such that $\lim_{i\to\infty} \rho(x, x_i) = 0$. We will say that C is **closed** if it contains all it's limit points. The intersection of all closed sets containing A is called the **closure** of A and is denoted by $cl(A)$. The **interior** of A, denoted A^o, is the union of all open sets contained in A. The **exterior** of A, $ext(A)$, is the interior of the complement of A. I.e $ext(A) = (A^c)^o$. (Recall that the **complement** of A are all the points in X that are not in A, i.e $A \setminus A$.) The **boundary** of A is the points that not in the exterior or the interior of A: $bdy(A) = X \setminus (A^o \cup ext(A))$.

Exercise 2.1.16. Prove that an arbitrary union of open sets is open. (This proves that the interior of a set is open.)

Exercise 2.1.17. Prove that a finite intersection of open sets is open.

Exercise 2.1.18. Prove that a finite intersection of closed sets is close.

Exercise 2.1.19. Prove that an arbitrary intersection of closed sets is closed.(This proves that the closure of a set is closed.)

Exercise 2.1.20. Prove that A is open if and only if A^c is closed.

Exercise 2.1.21. Prove that X and $\emptyset$ are both open.

Exercise 2.1.22. Prove that a set is closed if and only is its complement is open.

Exercise 2.1.23. Let $lp(A)$ be the set of limit points of A. Prove that $lp(A) \cup A = cl(A)$.

Exercise 2.1.24. Prove that $cl(A \cap B) \subset cl(A) \cap cl(B)$.

Exercise 2.1.25. Prove that $A^o \cap B^o = (A \cap B)^o$.

Exercise 2.1.26. Why is the bdy(A) always closed?

Exercise 2.1.27. Draw pictures to illustrate Exercises 2.1.16-2.1.26

2.1.4 Limits, subsequences, completeness, compactness

We say a sequence $\{x_i\}_{i=1}^{\infty} \subset X$ **converges** to x – $x_i \to x$ – if $\lim_{i\to\infty} \rho(x, x_i) = 0$. Let $k : \mathbb{N} \to \mathbb{N}$ be any strictly increasing function from the natural number to the natural numbers. The sequence $\{x_{k(i)} = x_{k_i}\}_{i=1}^{\infty} \subset X$ is called a **subsequence** of $\{x_i\}_{i=1}^{\infty} \subset X$. A **Cauchy sequence** is any sequence such that for all $\epsilon > 0$, there is an $N_\epsilon < \infty$ such that if $i, j > N_\epsilon$, then $\rho(x_i, x_j) \leq \epsilon$. A metric space is **complete** if every Cauchy sequence converges: that is, if for every Cauchy sequence $\{x_i\}_{i=1}^{\infty} \subset X$ there is an $x \in X$ such that $x_i \to x$, then we say that X is complete. We say that X is **totally bounded** if for every $\epsilon > 0$ there is a finite collection of ϵ balls $\{B(x_i, \epsilon)\}_{i=1}^{N_\epsilon}$ such that $X \subset \cup_{i=1}^{N_\epsilon} B(x_i, \epsilon)$. Such a collection of centers $\{x_i\}_{i=1}^{N_\epsilon}$ is called an **ϵ-net**. We say that $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ **covers** A if $A \subset \cup_\alpha U_\alpha$. A finite subcollection $\{U_{\alpha_i}\}_{i=1}^{N}$ of a cover of A, $\mathcal{U} = \{U_\alpha\}_{\alpha\in\mathcal{A}}$, is called a **finite subcover** if $A \subset \cup_i^N U_{\alpha_i}$. A set or space is **compact** if every open cover has a finite subcover. In general, compactness allows you to reduce countably infinite collections and arguments to finite collections and arguments.

Exercise 2.1.28. Prove that $S \equiv \{x_i\}_{i=1}^{\infty}$ converges if and only if every subsequence of S converges.

Exercise 2.1.29. Prove that if $S \equiv \{x_i\}_{i=1}^{\infty}$ converges, then every subsequence converges to the same limit.

Exercise 2.1.30. Let $X =$ the open unit ball in $\mathbb{R}^2$. Let ρ be the standard euclidean distance on $\mathbb{R}^2$ (see Exercise 2.1.3 with n = 2). Show that $\{\rho, X\}$ is not complete. Find a Cauchy sequence that does not converge to an $x \in X$

Exercise 2.1.31. Prove that any metric space X can be embedded isometrically into another metric space $\hat{X}$ that is complete, such that every point in $\hat{X} \setminus X$ is the limit of some Cauchy sequence in X. Hint: let $\hat{X}$ be the space of all Cauchy sequences of X ...

Exercise 2.1.32. Prove that a subset Y of a complete metric space X is complete as a metric subspace of X if and only if Y is closed in X.

Exercise 2.1.33. Prove that if X is totally bounded and complete, then every infinite subset S has a limit point; I.e. there is a sequence $\{x_i\}_{i=1}^{\infty}$ in S and a point $x \in X$ such that $x_i \to x$.

Exercise 2.1.34. Assume that X is complete. Prove that if for all $i \in \mathbb{N}$, $C_i \subset X$ is closed and the sets are nested, $C_{i+1} \subset C_i$, then $\cap_{i=1}^{\infty} C_i \neq \emptyset$.

Exercise 2.1.35. Show that X is totally bounded and complete if and only if every open cover has a finite subcover.

Exercise 2.1.36. Show that in $\mathbb{R}^n$ a subset is compact if and only if it is closed and bounded. Hint: The only thing we need that is not in the previous exercises is that $\mathbb{R}$ is complete. You can take this as an axiom. Next show that $\mathbb{R}^n$ is complete.

Exercise 2.1.37. Suppose that A is compact. Suppose that for every $x \in A$, there is a δ_x and an ϵ_x such that $f(y) > \delta_x$ for all $x \in B(x, \epsilon_x)$. Show that there is a $\delta > 0$ such that $A > \delta$ for all $x \in A$.

2.1.5 Connectedness

A set A is **disconnected** if there are two disjoint open sets U_1 and U_2 such that $A \cap U_1 \neq \emptyset \neq A \cap U_2$ and $A \subset U_1 \cup U_2$. A set that is not disconnected is **connected**. The arbitrary union of connected sets all containing a common point is connected (exercise below) and so we can defined the the **connected component** containing a point x to be the union of the connected sets containing x.

Suppose A has the property that for every pair of points $a \neq b$ in A, we have a pair of disjoint open sets U_a and U_b such that $A \subset U_a \cup U_b$ and $a \in U_a$ and $b \in U_b$. We say that such a set A is **totally disconnected**.

Exercise 2.1.38. Show that any finite subset of the real line with the usual metric $|\cdot|$, is totally disconnected.

Exercise 2.1.39. Show that rational numbers $\mathbb{Q} \subset \mathbb{R}$ is a totally disconnected set.

Exercise 2.1.40. Show that an arbitrary union of connected sets that all have at least one point in common is also connected.

Exercise 2.1.41. Show the intervals $(0, 1)$ and $[0, 1]$ are connected. Hint: Assume they are disconnected and use the fact that least upper bounds (supremums) exist on the real line, after you choose two points which are each in different open sets with are disjoint yet together cover the interval.

2.1.6 Continuity

We say that f is **continuous at x** $\in X$ if for every $\epsilon > 0$ there is a $\delta > 0$ such that $y \in B(x, \delta)$ implies that $f(y) \in B(f(x), \epsilon)$. We define the **limit of f at x** to be $\lim_x f = \hat{f}$ if, for every $\epsilon > 0$, there is a $\delta > 0$ such that $y \in (B(x, \delta) \setminus x)$ implies that $f(y) \in B(\hat{f}, \epsilon)$. In these terms, f is continuous at x if $\lim_x f = f(x)$. We say f is **continuous** if it is continuous for every $x \in X$. Equivalently a function from one metric space to another, $f : X \to Y$, is **continuous** if the inverse image of any open set on Y is open in X. We say that f is **uniformly continuous** if δ is a function only of ϵ and not of x.

Exercise 2.1.42. Prove that in metric spaces, the two definitions of continuity, given in the first two sentences of this subsection above, are equivalent. (The first definition works in general topological spaces whereas the second does not.)

Exercise 2.1.43. Suppose X and Y are metric space. Prove that if $f : X \to Y$ is continuous, then $f^{-1}(y) \subset X$ is closed for any point $y \in Y$. More generally, prove that the inverse image of any closed set in Y, is closed in X.

Exercise 2.1.44. Prove that continuous functions map connected sets to connected sets.

Exercise 2.1.45. Use 2.1.44 to prove the intermediate value theorem: Assume $f : \mathbb{R} \to \mathbb{R}$ is continuous. Assume also that $f(a) \leq f(b)$. Then for every $f(a) \leq y \leq f(b)$ there is at least one $x \in [a, b]$ such that $f(x) = y$. Comment why the case of $f(a) > f(b)$ is completely analogous.

Exercise 2.1.46. Prove that a continuous function on a compact set is uniformly continuous.

Exercise 2.1.47. Prove that a continuous function on a compact set attains a maximum and a minimum.

Exercise 2.1.48. Let $A \subset \mathbb{R}^n$ and define $d(x, A) = \inf_{a \in A} \rho(x, a)$. Prove that d(x,A) is continuous as a function from $x \in \mathbb{R}^n$ to $\mathbb{R}$.

Exercise 2.1.49. If A is closed, prove that there is a point in $a^* \in A$ such that $\rho(x, a^*) = d(x, A)$.

Exercise 2.1.50. Give examples of closed sets where there is more than one point $a^* \in A$ that minimizes $\rho(x, a)$ (Remember, we are fixing x and varying a). Give examples of open sets A and point x such that there is no point $a^* \in A$ minimizing $\rho(x, a)$.

Exercise 2.1.51. Define the mapping $f : c([0, 1]) \to \mathbb{R}$ by $f(x) = \int_0^1 x(t)\ dt$. Prove that this function is continuous on the metric space introduced in Exercise 2.1.8.

Exercise 2.1.52. Continuing exercise 2.1.51, what can you say about $f^{-1}(1)$? Is this a bounded set? Is it a closed set?

Exercise 2.1.53. Suppose that $U \subset X$ is open and $A \subset U$ is compact. Use distance functions to create a continuous function from X to $\mathbb{R}$ such that $0 \leq f \leq 1$ on X and $f = 1$ on A and $f = 0$ on U^c.

Exercise 2.1.54. Suppose that $f : [a, b] \to [a, b]$ is continuous. Prove that there is an $x \in [a, b]$ such that $f(x) = x$.

2.2 More Limits and $\liminf$, $\limsup$

We already spent some time thinking about limits in the section on metric spaces, but we will spend more time now looking at limits in $\mathbb{R}$ and $\mathbb{R}^n$, as well as in spaces of functions.

If we have a sequence $\{a_i\}_{i=1}^{\infty}$ it may not converge to anything. but we always have the following two derived sequences: $L_k \equiv \sup\{a_i | i \geq k\}$ and $l_k \equiv \inf\{a_i | i \geq k\}$. We now define the $\limsup(\{a_i\}) = \lim_{k\to\infty} L_k$ and $\liminf(\{a_i\}) = \lim_{k\to\infty} l_k$.

If $f : X \to \mathbb{R}$ and X is a metric space, we can define $\liminf$ and $\limsup$ in terms of epsilon balls. We define $\limsup\limits_x f \equiv \lim_{r\to 0}(\sup_{y\in(B(x,r)\setminus x)} f(y))$ and $\liminf\limits_x f \equiv \lim_{r\to 0}(\inf_{y\in(B(x,r)\setminus x)} f(y))$.

Exercise 2.2.1. Prove that if a_i is non-decreasing then there is an $a \leq \infty$ such that $\lim_{i\to\infty} a_i = a$.

Exercise 2.2.2. Prove that if a_i is non-increasing then there is an $a \geq -\infty$ such that $\lim_{i\to\infty} a_i = a$.

Exercise 2.2.3. Prove that for any sequence $\{a_i\}$, $\limsup(\{a_i\})$ and $\liminf(\{a_i\})$ exist and satisfy $\limsup(\{a_i\}) \geq \liminf(\{a_i\})$. Prove also that $\{a_i\}$ converges to a limit if and only if $\limsup(\{a_i\}) = \liminf(\{a_i\})$.

Exercise 2.2.4. Prove that for any $f : X \to \mathbb{R}$, we have that $\limsup\limits_x f$ and $\liminf\limits_x f$ exist. Prove that $\lim\limits_x f$ exists if and only if $\limsup\limits_x f = \liminf\limits_x f \neq \pm\infty$, in which case $\lim\limits_x f = \limsup\limits_x f = \liminf\limits_x f$

Exercise 2.2.5. Draw illustrations of the ideas behind $\liminf$, $\limsup$ and $\lim$.

2.3 Infinite sequences and Series

An **infinite sequence** is an indexed sequence of numbers, $\{a_i\}_{i=1}^{\infty} \subset \mathbb{R}$ though, strictly speaking, it is a function from the natural numbers $\mathbb{N}$ into $\mathbb{R}$. A **infinite series** is an infinite sum or numbers or functions: for example, $\sum_{i=1}^{\infty} \frac{1}{i^2}$ or $\sum_{i=0}^{\infty} x^i$. On the GQE there is always at least one question about series. Asking whether or not a series converges is, more precisely, asking if the sequence $S_k \equiv \sum_{i=1}^{k} a_i$ or $S_k \equiv \sum_{i=1}^{k} a_i(x)$ converges. In the second case, the answer may vary with x and if it does converge the speed it converges may vary with x. In particular, the case of **power series** where we are studying $\sum_{i=0}^{\infty} a_i x^i$, takes a prominent role in the study of series, both for historical reasons (e.g. newton used power series to solve differential equations) and the fact that **Taylor series** – a special case of power series – are a very important family of local approximation approximations to functions. We will study Taylor series later in these notes. We say a series $\sum_i a_i$ is **absolutely convergent** if $\sum_i |a_i|$ converges and is **conditionally convergent** if $\sum_i a_i$ is convergent but $\sum_i |a_i|$ is not.

Generally speaking, to answer whether something converges or not, we have to know what metric we are measuring proximity in: $a_i \to \hat{a}$ if $\rho(a_i, \hat{a}) \to 0$. In the next few sections, we will often be using the euclidean norm on $\mathbb{R}^n$, with this being simply the absolute value function on $\mathbb{R}$. We denote them all by $|a|$ or $|a - b|$ whenever $a, b \in \mathbb{R}^n$ for any n.

Exercise 2.3.1. Suppose that $a_i \to a$ and $b_i \to b$ where $-\infty < a, b < \infty$. Prove that:

1. $a_i + b_i \to a + b$,
2. $a_i b_i \to ab$,
3. for the case that $b \neq 0$ we also have $a_i / b_i \to a/b$, and
4. $\frac{1}{n} \sum_{I=1}^{n} a_i \to a$.

Exercise 2.3.2. Suppose that:

1. $a_i^N > 0$ for all i and N
2. $\sum_{i=1}^{N} a_i^N = 1$ for all N
3. For any fixed $M > 0$, we have that $\lim_{N\to\infty}(\sum_{i=1}^{M} a_i^N) = 0$

Prove that if $b_i \to b$, then $\lim_{N\to\infty}(\sum_{i=1}^{N} a_i^N b_i) = b$

Exercise 2.3.3. Suppose that $-\infty < \sum_i^{\infty} a_i < \infty$, but that $\sum_i^{\infty} |a_i| = \infty$. Prove that by rearranging the order of the summation, we can make this series converge to any real number!

Exercise 2.3.4. Prove that if $\sum_i |a_i|$ converges then $\sum_i |a_i|^p$ also converges as long as $p \geq 1$. Find a counterexample for the claim that this is also true when $0 < p < 1$.

Exercise 2.3.5. Assume that $|a_i| < 1$ for all i. Prove that $\sum_i a_i$ converges if and only if $\sum_i \ln(1+a_i)$ converges.

Exercise 2.3.6. Suppose that $\lim_{i\to\infty} a_i = 0$. Use exercise 2.3.5 to prove that $\Pi(1+a_i)$ converges to $p \neq 0$ if and only if $\sum a_i$ converges.

Exercise 2.3.7. Use what you know about the integral $\int_1^\infty \frac{1}{x^\alpha}dx$ to establish the convergence properties of $\sum_i \frac{1}{i^\alpha}$ for $\alpha > 0$.

Exercise 2.3.8. Show that if $\sum_{i=0}^\infty a_i x^i$ converges for $x = x^*$, then this power series converges absolutely for all x such that $|x|$ ¡ $|x^*|$.

Exercise 2.3.9. The previous Exercise (2.3.8) explains the term **Radius of convergence, R** which is defined to be the supremum of the r such that $\sum_{i=0}^\infty a_i x^i$ when $|x| \leq r$. What is the radius of convergence R of the series $\sum_{i=0}^\infty x^i$? What can you say about convergence at $x = \pm R$

Exercise 2.3.10. What is the radius of convergence R of the series $\sum_{i=1}^\infty \frac{(-x)^i}{i}$? What can you say about convergence at $x = \pm R$

Exercise 2.3.11. The various ways of finding the radius of convergence of a power series are explored in Burn's book: look them up and do the corresponding exercises there.

2.4 Pointwise and Uniform Convergence

We now look at a zoo of limit problems, including those in which there are multiple versions of convergence happening (or not happening) simultaneously. We say that f_i **converge pointwise** to f if, for $x \in X$ we have that $|f_i(x) - f(x)| \underset{i\to\infty}{\to} 0$. The key point here is that the rates at which this convergence happens, can vary from point to point in X. We say that f_i **converge uniformly** to f if $\sup_{x\in X}|f_i(x) - f(x)| \underset{i\to\infty}{\to} 0$. Recalling Exercise 2.1.8, we see that the metric in that spaces was the metric of uniform convergence. The famous **Weierstrass M-test** is another example of uniform convergence – see Exercise 2.4.2. (There are topologies more complicated than metric space topologies that can be used to study pointwise convergence, but we will not study these in this course.)

When we get to our more in depth look at integration, we will encounter additional modes of convergence, some of which are hinted at in the exercises.

Exercise 2.4.1. Find a sequence of continuous functions which converge pointwise to another continuous function, but the convergence is not uniform. Hint: you can chose $\{x_i\}_{i=1}^\infty \in C([0,1])$ all of which have values between 0 and 1, with $x_i(0) = x_i(1) = 0$ for all i.

Exercise 2.4.2. (**Weierstrass M-Test**) Suppose that

1. $\sup_{x\in X}|f_i(x)| \leq M_i$ and
2. $\sum_i M_i < \infty$.

Prove that there is an f such that $f_i \to f$ uniformly.

Exercise 2.4.3. Suppose that $f_i \to f$ uniformly and that each of the f_i are continuous. Prove that f is continuous.

Exercise 2.4.4. Prove that if $f_i \to f$ uniformly and $[a,b]$ is a bounded interval, then $\int_a^b f_i dx \to \int_a^b f dx$. Find an example of $f_i \to f$ that converge pointwise in $[a,b]$ and yet $\int_a^b f_i dx \not\to \int_a^b f dx$.

Exercise 2.4.5. We define $f_i : [0,1] \to \mathbb{R}$ for any i: for all odd $0 < k < 2^i$, define $f_i : [0,1] \to \mathbb{R}$ to be -1 when $\frac{k-1}{2^i} \leq x < \frac{k}{2^i}$, and 1, when $\frac{k}{2^i} \leq x < \frac{k+1}{2^i}$ $\frac{k}{2^i}$. Define $f_i(1) = 0$ for all i. Show that f_i converge pointwise nowhere except at $x = 1$, yet $\int_0^1 f_i(y)h(y)\, dy \to \int_0^1 0 \cdot h(y) dy = 0$ for any fixed h which is continuous on $[0,1]$. (In this case we say that f_i **converge weakly** to f

Exercise 2.4.6. We have been thinking about $f : [a,b] \to \mathbb{R}$, which are functions on spaces with a uncountably infinite number of points. Consider instead $f : \mathbb{N} \to \mathbb{R}$ or even $f : (1,2,3,...,n) \to \mathbb{R}$ – the infinite or finite sequence spaces. In the second case, we can see these functions as simply points in $\mathbb{R}^n$. In the second case, we often think of these functions as infinite sequences $(a_1, a_2, ..., a_n, a_{n+1}, ...)$ and use the metrics $\rho_p(a,b) = \sqrt[p]{\sum_i |a_i - b_i|^p}$ with some fixed $1 \leq p < \infty$ or even $\rho_\infty(a,b) = \sup_i |a_i - b_i|$. **Prove** that in the case of finite sequences, pointwise convergence is equivalent to convergence in any of the usual metrics in $\mathbb{R}^n$ – $\rho_p(x,y) = |x-y|_p = \sqrt[p]{\sum_i |x_i - y_i|^p}$.

Exercise 2.4.7. Prove that pointwise convergence in the space of infinite sequences is not the same as convergence in any of the ρ_p metrics.

Exercise 2.4.8. Let us designate the space of sequences x such that $|x|_p < \infty$ by l_p. Suppose that $f_n = (f_1^n, f_2^n, f_3^n, ...)$ is a sequence of infinite sequences in l_1. Suppose further that $|f_i^n| \leq b_i$ for all i and all n and $|b|_1 < \infty$. **Prove** that in this case, pointwise convergence is equivalent to convergence in the ρ_1 metric.

2.5 Norms and Inner Products

When we are working in vector spaces, we will often be working in spaces that have extra structure that turns the space into a metric space or even a metric space that flows from a generalized dot product called an **inner product**. When these vector spaces are infinite dimensional – for example, when the space is a space functions like all functions $f : [0,1] \to \mathbb{R}$ such that $\int_{[0,1]} |f| dx < \infty$ – life gets much more interesting, and complicated. We will look at that in a little bit of detail later in these notes. A **vector norm** or simply **norm** is a length measure of vectors satisfying (1) $|x| \geq 0$, with $|X| = 0$ if and only if $x = 0$, (2) $|\alpha x| = |\alpha||X|$ for all (real or complex) scalars, and (3) $|x+y| \leq |x| + |y|$. An **inner product** on a vector space is a generalized dot product. An **inner product over the real numbers** is generated by any bi-linear, symmetric, positive definite function $F : V \times V \to \mathbb{R}$ is an inner product – recall that F(x,y) is (1) bi-linear if $F(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 F(x_1, y) + \alpha_2 F(x_2, y)$ and $F(x, \alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 F(x, y_1) + \alpha_2 F(x, y_2)$, (2) symmetric if $F(x,y) = F(y,x)$ and (3) positive definite if $F(x,x) > 0$ whenever $x \neq 0$. The definition of an **inner product of the complex numbers** $(F : V \times V \to \mathbb{C})$ is a bit different: we have that F is an inner product on a vector space of the complex numbers if it is conjugate symmetric, linear in the first term and positive definite. Using $\overline{\alpha}$ to indicate the complex conjugate of α, this reduces to

(1) $F(\alpha_1 x_1 + \alpha_2 x_2, y) = \alpha_1 F(x_1, y) + \alpha_2 F(x_2, y)$ and $F(x, \alpha_1 y_1 + \alpha_2 y_2) = \overline{\alpha_1} F(x, y_1) + \overline{\alpha_2} F(x, y_2)$, (2) $F(x, y) = \overline{F(y, x)}$ and (3) $F(x, x) > 0$ whenever $x \neq 0$. **It is traditional to denote $F(x, y)$ by $< x, y >$.** An inner product generates a norm: if $\langle \cdot, \cdot \rangle$ is an inner product, then $|x| \equiv \sqrt{|\langle x, x \rangle|}$ is a norm. If the normed vector space $(V, |\cdot|)$, is complete as a metric space with the metric $\rho(x, y) = |x - y|$, then V is a **Banach Space**. When the Banach space norm is generated by an inner product, V is a **Hilbert Space**.

Exercise 2.5.1. Prove that if $\langle \cdot, \cdot \rangle$ is an inner product, then $|x| \equiv \sqrt{|\langle x, x \rangle|}$ is a norm.

Exercise 2.5.2. Prove that if $E \subset \mathbb{R}^n$ is any bounded convex set with nonempty interior, that is symmetric about the origin, defines a norm. In a bit more detail, suppose that α is the largest scalar such that $\alpha x \in E$. Define $F(x) = \frac{1}{\alpha}$: prove that if $|x| \equiv F(x)$ is a norm. **Suggested steps**.

1. Prove that E contains some epsilon ball containing the origin.
2. Prove that $F(x) \geq 0$ if and only if $x = 0$.
3. Prove that $F(\alpha x) = |\alpha| F(x)$
4. Prove that the F is a convex function.
5. Prove that $F(x + y) \leq F(x) + F(y)$.

Exercise 2.5.3. Prove that $E \subset \mathbb{R}^n$, $E \equiv x||x| \leq 1$ is a bounded convex set with non-empty interior and is symmetric about the origin. Conclude that the procedure outlined in Exercise 2.5.2 using E gives us back the $|\cdot|$ used to define E.

Exercise 2.5.4. Suppose we are in a Hilbert space H with inner product $\langle \cdot, \cdot \rangle$. Choose some fixed $x \in H$. Prove that any other $y \in H$ can be written as $\alpha_1 x + \alpha_2 w$ where w (which can vary with y) satisfies $\langle x, w \rangle = 0$ and α_1 and α_2 are scalars.

Exercise 2.5.5. Recall that in a Hilbert space H, $|x| \equiv \sqrt{\langle x, x \rangle}$. Suppose $x, y \in H$, $|x| = |y| = 1$ and $x \neq y$. Prove that there are nonzero scalars α_1 and α_2, and a vector w, such that $|w| = 1$, $\langle x, w \rangle = 0$, $y = \alpha_1 x + \alpha_2 w$ and $\alpha_1^2 + \alpha_2^2 = 1$.

Exercise 2.5.6. A linear subspace S of a vector space V is any subset which is closed under vector addition and scalar multiplication. That is, if α is a scalar and $x, y \in S$, then $x + y \in S$ and $\alpha x \in S$. Suppose that $E \subset V$ and $x \in E$. We define $E - x \equiv \{y \in V | y = e - x \text{ for some } e \in E\}$. Prove that $S - x = S$ if and only if $x \in S$

Exercise 2.5.7. An **affine subspace of a vector space** is a subset of the vector space such that for any $x \in A$, then $A - x$ is a linear subspace of V. Prove that every affine subspace is of the form $S + x$ where $x \in V$ and S is a subspace of V.

Exercise 2.5.8. Suppose that $\mathcal{H}$ is n-dimensional. Prove that $H_{w,c} \equiv \{x \in \mathcal{H} | \langle w, x \rangle = c\}$ is an affine subspace of dimension $n - 1$.

Exercise 2.5.9. Define the **half spaces** $H^+_{w,c} \equiv \{x \in \mathcal{H} | \langle w, x\rangle \geq c\}$ and $H^-_{w,c} \equiv \{x \in \mathcal{H} | \langle w, x\rangle \leq c\}$. We will sometimes simplify $H_{w,c}$, $H^+_{w,c}$, and $H^-_{w,c}$ to H, H^+, and H^- when w and c are understood from the context.

Exercise 2.5.10. **Assume** the fact that E **is closed and strictly convex** if and only if every point x in the boundary of E is contained in an affine subspace H of V, intersecting E only at x and $E \subset H^-$. Prove that $E = \{x | \, |x| = \sqrt{\langle x, x\rangle} \leq 1\}$ is strictly convex. **Suggestion**: consider $x \in E$ such that $|x| = 1$ and $H^-_x \equiv \{y | \langle y, x\rangle \leq 1\}$

2.6 Vector Calculus

In this section we deal with integration of various quantities when the domain or range (or even both the domain and range) are higher dimensional. This adds many interesting, intriguing aspects to the study of calculus. What we have here will be somewhat sparse and will instead expect you to peruse the calculus book by Swokowski (or something equivalent) fairly carefully.

2.6.1 Contour integration and Stokes Theorem

Integration of functions of various kinds along 1-dimensional paths in an n-dimensional domain is called **contour integration**. One often studied version computes the integral of the dot product of a vector field with the tangent vector of a path $\gamma : [a, b] \to \mathbb{R}^n$: $\int_\gamma v \cdot \dot\gamma(s)ds$. If $v = \nabla f$, where $f : \mathbb{R}^n \to \mathbb{R}$, then the result will be **the change in f** from one endpoint of γ to the next: $f(\gamma(b)) - f(\gamma(a)) = \int_a^b \nabla f \cdot \dot\gamma(s)ds$. If instead F represents the force exerted on an object moving around γ, then $\int_\gamma F \cdot \dot\gamma(s)ds$ gives us the **energy gained or lost** by an object moving along γ. Sometimes, one might want instead to compute $\int_\gamma F(\gamma(s))|\gamma(s)|ds$, in order to know the **total force exerted on an object** represented by the curve γ. (The units for F would be force per unit length.)

We are also sometimes interested in knowing **how much stuff flows through a curve** γ when the flow velocity is given by $\vec{F}(x)$. In this case, we are assuming that the domain is $\mathbb{R}^2$ and we want to compute $\int_\gamma \vec{F} \cdot n_\gamma |\dot\gamma(s)|ds$, where n_γ is the norm vector to the curve pointed in the outward direction. (For example on can consider the outward direction to be the direction pointing to the right of the direction ahead is defined to be $\dot\gamma(s)$.)

Note that we will often write $\int_{\partial\Omega} v \cdot dl$, which, when we parameterize the curve $\partial\Omega$ (with $\gamma : [a, b] \to \mathbb{R}^2$ and $\gamma(a) = \gamma(b)$), becomes $\int_\gamma v \cdot \dot\gamma(s)ds$. If the curve we are integrating over is not a boundary, then instead of $\int_{\partial\Omega} v \cdot dl$ we have $\int_\Gamma v \cdot dl$ which, in parameterized form, is again $\int_\gamma v \cdot \dot\gamma(s)ds$ except that now $\gamma(a) \neq \gamma(b)$. Notice that the answer depends on the direction you integrate. If the curve is a boundary, then you can either integrate in the direction that keeps the inside to your left as you move ahead on the boundary (the standard choice) or the direction that keeps the inside on your right. The answers differ by their sign. Likewise with Γ - you will get two possible answers differing by their sign, depending on which direction you integrate.

The classical **Stokes Theorem** – $\int_\Omega \nabla \times v \; d\sigma = \int_{\partial\Omega} v \cdot dl$ – often helps us compute contour integrals. This is demonstrated in the exercises that follow. The curl of a vector field measures the "twistyness" of the vector field and is a special case of something called an exterior derivative. What we are calling Stokes theorem above is actually a special case of a general Stokes theorem that relates the integral of the exterior derivative of a form on a region to the integral of the form over the boundary of the region. For those that know what forms are, suppose that ω is a k-form, Ω is a $k+1$-dimensional set that has sufficient regularity (smoothness), and $d\omega$ is the exterior derivative of ω. Then the general Stokes theorem says that $\int_{\partial\Omega} \omega = \int_\Omega d\omega$.

Exercise 2.6.1. Look up the definition of $\nabla \times v$, the curl of v if you do not already know it. Note that the definition should work for the case in which is a vector field in 3 dimensional space.

Exercise 2.6.2. Let $F(x) = x_1^2+2x_1x_2+x_2^2$ and suppose that $v(x) = \nabla F$. Compute $\int_\gamma v(x)\cdot\dot{\gamma}(s)ds$ where γ is the closed contour that traces out the set $|x_1| + |x_2| = 1$. Do not use Stokes Theorem (i.e.$\int_\Omega \nabla \times v \, d\sigma = \int_{\partial\Omega} v \cdot dl$). Do this problem with **and** without the help of Stokes Theorem.

Exercise 2.6.3. Prove that $\nabla \times \nabla f = 0$ for $f \in C^2$.

Exercise 2.6.4. Define $v(x) = (x_1^2 + x_2^4, x_1 + x_1x_2)$. Suppose that $\Omega =$ the square defined by the points four point $(0, 0), (1, 0), (0, 1), (1, 1)$: i.e. Ω is the convex hull of those four points. Compute $\int_{\partial\Omega} v \cdot dl$ integrating in the counterclockwise direction. Does Stokes Theorem help you do this problem?

Exercise 2.6.5. Define $v(x) = (-x_2, x_1)$. Compute $\int_{\partial\Omega} v \cdot dl$ for the squares in the positive quadrant with one corner at $(0, 0)$ and (1) side length 1, (2) side length 3, and (3) side length 5. Do this problem with **and** without the help of Stokes Theorem.

Exercise 2.6.6. Prove that if $\nabla \times v = 0$ in a region $\Omega \subset \mathbb{R}^2$, then there is a function F such that in Ω, $v = \nabla F$.

Exercise 2.6.7. The classical Stokes theorem works on 2-dimensional surfaces in $\mathbb{R}^3$ as well. (This is of course clear if one knows the general Stokes theorem that applies to k-dimensional sets in $\mathbb{R}^n$.) In this case though, the statement is little a bit different: $\int_\Omega \nabla \times v \cdot n_\Omega \, d\sigma = \int_{\partial\Omega} v \cdot dl$ where n_Ω is the normal vector to Ω. Suppose that v is a vector field in $\mathbb{R}^3$, i.e. $v : \mathbb{R}^3 \to \mathbb{R}^3$. Now suppose that you consider the vector field $w = \nabla \times v$ and its integral curves, i.e. the solution curves generated by the vector field $\nabla \times v$. If γ is a closed curve that is everywhere transverse to the vector field w prove that $\int_\gamma v \cdot dl$ is invariant under the flow generated by w.

Exercise 2.6.8. Prove Stokes Theorem ($\int_\Omega \nabla \times v \, d\sigma = \int_{\partial\Omega} v \cdot dl$) for Ω with C^1 boundaries by (1) proving it first for rectangles and right triangles and then (2) decomposing Ω (assumed to have a C^1 boundary!) into rectangles and right triangles and a very small difference region.

Exercise 2.6.9. Explain geometrically why Stokes theorem works for arbitrary shapes in $\mathbb{R}^2$ and the vectorfield $v = (-x_2, 0)$. use a non-convex region with smooth boundary to illustrate this.

2.6.2 The Divergence Theorem and other Vector Calculus Theorems

Another special case of the general Stokes theorem is the **divergence theorem** that relates the flux of a vector field through the boundary of a region to the integral of the divergence of the vector field over the region. More precisely, $\int \nabla \cdot v \; dx = \int_{\partial\Omega} v \cdot \vec{n} \; d\sigma$. In $\mathbb{R}^2$, you can prove that the divergence theorem is just Stokes and vice versa. But in higher dimensions, we have to either use the general Stokes or prove the theorem directly.

The **product rule** from the first course in calculus – $(fg)' = f'g + fg'$, is not always taught as the key to remembering **integration by parts** – $\int_a^b f'g dx = \int_a^b (fg)'dx - \int_a^b fg'dx$, even though it is. After you learn the divergence theorem, you have at your disposal another theorem based on the same idea. Now we have that $\nabla \cdot hv$, where h is a scalar function, v is a vector field – $h : \mathbb{R}^n \to \mathbb{R}$ and $v\mathbb{R}^n \to \mathbb{R}^n$, can be written as $\nabla \cdot hv = \nabla h \cdot v + h\nabla \cdot v$ and we end up (after an application of the divergence theorem) with $\int_\Omega \nabla h \cdot v \; dx = \int_{\partial\Omega} hv \cdot n_{\partial\Omega} \; d\sigma - \int_\Omega h\nabla \cdot v \; dx$, where $n_{\partial\Omega}$ is the outward normal to $\partial\Omega$ and integration with respect to σ is integration over the n-1-dimensional surface $\partial\Omega$.

Exercise 2.6.10. How is the fundamental theorem of calculus ($F(x) - F(a) = \int_a^x \frac{dF(y)}{dy} dy$) just the divergence theorem?

Exercise 2.6.11. Use the divergence theorem and the classical Stokes theorem to prove that $\nabla \cdot (\nabla \times v) = 0$. Hint: use two hemisphere to enclose any region of space where $\nabla \cdot (\nabla \times v) \neq 0$.

Exercise 2.6.12. Use the fact that $f \cdot dl$ is same as the dot product of f rotated clockwise by $\pi/2$ and and the normal of the boundary of Ω to show that (1) the classical Stokes and (2) the divergence theorem are the same theorem for 2-dimensional Ω!

Exercise 2.6.13. Prove Green's first identity: $\int_\Omega f\Delta g + \nabla f \cdot \nabla g \; dx = \int_{\partial\Omega} f\nabla g \cdot n_{\partial\Omega} \; d\sigma$.

Exercise 2.6.14. Suppose that Ω is a region in space that is being advected by a flow generated by the vector field v and f is a scalar function that varies over space x and time t. Prove Reynolds Transport Theorem: $\frac{d}{dt}(\int_\Omega f dx) = \int_\Omega \frac{\partial f}{\partial t} \; dx + \int_{\partial\Omega} v \cdot n_{\partial\Omega} \; d\sigma$

Exercise 2.6.15. Choose two points in $\mathbb{R}^2$, x and y. Let S_{xy} be the line segment joining them. Now consider any other smooth curve γ beginning at x and ending at y. Consider the unit vectorfield that is normal to the direction of S_{xy} and oriented to the left of that direction ($\pi/2$ in the counterclockwise direction). Use the divergence theorem to show that the length of γ strictly exceeds the length of S_{xy} unless $\gamma = S_{xy}$. To make the problem simpler you can assume that γ does not intersect S_{xy} except at the endpoints. This is a simple example pf the method of calibrations – the vector field is a **calibration** designed to establish the minimality of S_{xy}.

Exercise 2.6.16. Find a family of vector fields that satisfy $\nabla \cdot v = 0$ by assuming that $v = (g_1(x)h_1(y), g_2(x)h_2(y))$.

Exercise 2.6.17. Assume we are in $\mathbb{R}^n$. Use the Reynolds transport theorem, the divergence theorem and an extension of the unit normal vector field u in the neighborhood of a smooth n-1-manifold or piece of smooth n-1-manifold M, to show that $\int_M \nabla \cdot v d\sigma = \frac{d}{dt} \int_M 1 d\sigma$; i.e. the integral of the divergence of this unit vector field gives us the instantaneous rate of change of the n-1-dimensional volume of this manifold as it is advected by the flow generated by the vectorfield u. Hint: use a very thin neighborhood of the manifold – thin in the normal direction – and let that envelope flow with this normal field flow and think about the change in n-volume of this envelope and its relation to the n-1-volume of the manifold the envelope is centered on.

Exercise 2.6.18. Let Ω be the L shaped region defined by the 6 points $(0,0)$, $(0,2)$, $(1,2)$, $(1,1)$, $(2,1)$, and $(2,0)$. Compute $\int_{\partial\Omega} v \cdot dl$ for the vector field $v = (-y + y^2x, x + yx^2)$.

Exercise 2.6.19. Let Ω be the L shaped region defined by the 6 points $(0,0)$, $(0,2)$, $(1,2)$, $(1,1)$, $(2,1)$, and $(2,0)$. Compute $\int_{\partial\Omega} v \cdot n_{\partial\Omega} d\sigma$ for the vector field $v = (x + yx^2, y - y^2x)$.

Exercise 2.6.20. Suppose that Ω is a region in $\mathbb{R}^2$ with smooth boundary parameterized by $\gamma(s)$. Show that the area of Ω is given by $\frac{1}{2}\int_\gamma \gamma(s) \times \dot{\gamma}(s) ds$. **Suggestion**: try proving it for smooth convex regions first. **Comment**: this works in $\mathbb{R}^n$ as well, where one now works with $n-1$-vectors which are wedge products of $n-1$, n-dimensional vectors and the integrand is the wedge product of the position vector on the surface of $\partial\Omega$ and the n-1 vector tangent orienting $\partial\Omega$.

Exercise 2.6.21. Suppose that a homogeneous substance S has a heat capacity of C_S and a heat conductivity of κ (both constants) and that you know that the flow of heat is in the direction of $-\nabla U$ (the negative of the gradient of the temperature U), with flow magnitude equal to the $k|\nabla U|$. Show that $U_t = \frac{\kappa}{C_S}\Delta U$, where ΔU is the Laplacian of U and is given by $\sum_{i=1}^n \frac{\partial^2}{(\partial x_i)^2} U$. Assume that the temperature field is smooth ($U \in C^2(\mathbb{R}^2)$). **Suggestion**: reason with tiny regions around some arbitrary point x in the domain. **Remark**: $U_t = \frac{\kappa}{C_S}\Delta U$ is called the heat equation and it is often studied assuming that $\kappa = C_S = 1$.

Exercise 2.6.22. How does the heat equation from Exercise 2.6.21 change when both κ and C_S are functions of position x?

2.6.3 Volumes of solids, tricky iterated integrals

Integrals can be used to compute volumes in $\mathbb{R}^3$ but these integrals can be tricky to compute. To compute the integral, we use Fubini's theorem to turn $\mathcal{H}^3(\Omega = Vol(\Omega) = \int_\Omega 1 dx$ into three iterated integrals. The art of making this computation work is often figuring out the right order for these three integrals.

Exercise 2.6.23. Use your calculus book to do $20 - 30$ 3-dimensional volume integrals involving triple integrals.

2.6.4 Tangents and curvature

Curves in $\mathbb{R}^n$ – images of function from $\mathbb{R}$ or segments in $\mathbb{R}$ to $\mathbb{R}^n$ – are of interest for many practical reasons. For example, when some physical process can be described by n state variable, time evolution of the system will be a curve in $\mathbb{R}^n$. For example, if we have a particle evolving in a potential field everything about the system is encoded in the positions in x,y, and z and velocities in the x,y, and z directions, so $n = 6$ in this case. If we have 3 particles interacting with each other and a potential field, $n = 18$. The case of $n = 2$ and $n = 3$ are covered in vector calculus. The **tangent vector to a curve** $f : \mathbb{R} \to \mathbb{R}^n$ is the derivative vector $\dot{f}(t) \in \mathbb{R}^n$. Writing $f(t) = (f_1(t), f_2(t), ..., f_n(t))$, the tangent vector is $\dot{f}(t) = (\frac{df_1(t)}{dt}, \frac{df_2(t)}{dt}, ..., \frac{df_n(t)}{dt})$. If we chose the parameterization so that we have unit speed $|\dot{f}(t)| = 1$ everywhere, we find that $\ddot{f}(t)$ is orthogonal to $\dot{f}(t)$.

In the case of unit speed parameterization, we call the quantity $\kappa \equiv \ddot{f}(t)|$ the **curvature**. In $\mathbb{R}^3$, the unit vector in the third orthogonal direction is called the **binormal**, whose direction is chosen by the right hand rule using $\dot{f}(t)$ and $\ddot{f}(t)$ as the first two directions, in that order.

Exercise 2.6.24. Suppose that $\gamma(s) = (\sin(2s), \cos(3s))$. Compute $\dot{\gamma}(s)$ and then compute $\int_0^{2\pi} |\dot{\gamma}(s)|ds$.

Exercise 2.6.25. Suppose that $\gamma : [0, L] \to \mathbb{R}^2$ is a simple closed curve.

1. What is the difference between the two integrals $\int_\gamma \dot{\gamma}(s)\, ds$ and $\int_\gamma |\dot{\gamma}(s)|\, ds$? What are they computing?

2. Suppose for the remainder of this exercise that $|\dot{\gamma}(s)| = 1$ for all s. Suppose we define $n(s)$ to be the unit vector obtained by rotating γ $\pi/2$ in the clockwise direction. Define κ to be that scalar that makes $\ddot{\gamma}(s) = \kappa(s)n(s)$ true. What do $\int_\gamma \kappa(s)ds$ and$\int_\gamma |\kappa(s)|ds$ compute?

3. Compute κ for a circle if radius r.

4. Compute $\int_\gamma \kappa(s)ds$ and$\int_\gamma |\kappa(s)|ds$ for a circle of radius r.

5. Suppose γ parameterizes the boundary of a fattened U shape region – see figure 2.1. Compute $\int_\gamma \kappa(s)ds$ and$\int_\gamma |\kappa(s)|ds$. You can assume that γ is 4 half circles glued together.

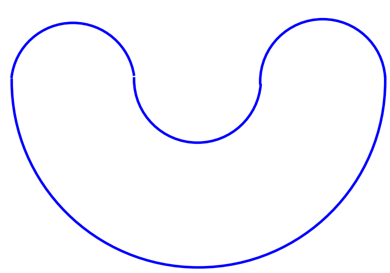

Figure 2.1: The boundary of a fattened U shaped region

6. Can you prove that for a simple close curve γ, $\int_\gamma \kappa(s)ds = 1$? Suggestion: prove that if we define $G : \gamma(s) \to \mathbb{S}^1$ mapping the point $\gamma(s)$ to the outward normal at $\gamma(s)$, κ is the signed Jacobian of G. That is, if we define s to be the length parameter on γ and θ to be the parameter on $\mathbb{S}^1$, $\kappa = \frac{d\theta}{ds}$.

Exercise 2.6.26. In the exercise, we explore parameterizations of curves. We assume that $\gamma : [a, b] \to \mathbb{R}^n$ is smooth and that $\dot{\gamma}(t) \neq 0$ for all $t \in [a, b]$.

1. Define $s = f(t) = \int_a^t |\dot{\gamma}(u)| du$

2. Prove that $s = f(t)$ smooth and that $t = f^{-1}(s)$ is smooth as well. (Use the inverse function theorem!)

3. Show that $\nu(s) \equiv \gamma(f^{-1}(s))$ is smooth and that $|\dot{\nu}(s)| = 1$ for all $s \in [0, f(b)]$.

4. We will say that a parameterization $f : [a, b] \to [a, b]$ is regular if f is smooth and $\dot{f} > 0$ on $[a, b]$. Denote the family of all regular parameterizations by $\mathcal{R}$.Suppose that $h : \mathbb{R}^n \to \mathbb{R}$ is continuous. Show that $\inf_{f \in \mathcal{R}} \int_a^b h(\gamma(f(t))\, dt = \inf_{s \in [a,b]} h(\gamma(s))$ and $\sup_{f \in \mathcal{R}} \int_a^b h(\gamma(f(t))\, dt = \sup_{s \in [a,b]} h(\gamma(s))$.

5. If instead, we evaluate $H_\gamma(f) \equiv \int_a^b h(\gamma(f(t))|\frac{d\gamma(f(t))}{dt}|(t)\, dt$, show that $H_\gamma(f)$ is invariant under changes in f (changes in parameterization). **Hint**: Notice that $H_\gamma(g) = H_\gamma(g \circ g^{-1} \circ f)$ and use the fact that regular parameterizations have smooth inverses.

Exercise 2.6.27. Suppose that $\gamma : \mathbb{S}^1 \to \mathbb{R}^n$ is smooth i.e., γ parameterizes a smooth, closed 1-dimensional curve in $\mathbb{R}^n$. Prove that for any $n-1$ dimensional subspace of $\mathbb{R}^n$, there are at least two points $\gamma(s_1)$ and $\gamma(s_2)$ such that $\dot{\gamma}(s_1) \in S$ and $\dot{\gamma}(s_2) \in S$. Hint: Every such Subspace has a unique (up to sign) normal vector.

Exercise 2.6.28. Suppose that v is a smooth vectorfield in $\mathbb{R}^n$ for which $w_v \equiv \nabla \times v \neq 0$ everywhere. Let Γ be a curve segment in $\mathbb{R}^n$ that is nowhere tangent to w_v. Let M be the 2-dimensional subset of $\mathbb{R}^n$ that is swept out by Γ under the flow generated by w_v. Suppose $N \subset M$ and that ∂N is a simple closed curve. Show that $\int_{\partial N} v \cdot dl = 0$.

Exercise 2.6.29. in Exercise 2.1.11 we found that when the cost per unit length had a discontinuity in it, minimal paths could develop kinks in them – discontinuities of the tangent vector, at the interface between regions of constant cost. (This is precisely what happens with the transmission of light at the interface between regions of different refractive indexes.) Prove that if $w : \mathbb{R}^n \to \mathbb{R}$ is smooth and we minimize $\int_\gamma w(\gamma(s))|\dot{\gamma}(s)|\, ds$ over γ satisfying $\gamma : [a, b] \to \mathbb{R}^n$, $\gamma(a) = x$ and $\gamma(b) = y$, minimal paths cannot have kinks in them.**suggestion**: assume there is a kink at x, zoom into the kink so that $w(x)$ is almost constant in a ball centered at x, and construct a comparison path with smaller cost.

Exercise 2.6.30. Continuing along the lines of Exercise 2.6.27, suppose that $\gamma : [a, b] \to \mathbb{R}^n$, γ is smooth and that $|\dot{\gamma}|(s) = 1$ for all $s \in [a, b]$. Show that if $n \geq 2$, then the $\{\gamma(s) | s \in [a, b]\} \neq \mathbb{S}^n$. **Suggestion**: look up Sard's theorem. **Another Suggestion**: For those of you that have some acquaintance with Hausdorff measures, another proof can be obtained by realizing that $\mathcal{H}^1(\gamma([a, b])) = b - a < \infty \Rightarrow \mathcal{H}^\alpha(\gamma([a, b])) = 0$ for $\alpha > 1$

2.7 Differentiation

Approximation by linear functions and operators, or more generally, by simpler classes of functions, in some Small neighborhood of a point is the idea behind differentiation or generalized differentiation. In this section, we look at some of these ideas. We begin with derivatives as linear approximations. In addition to Fleming's book, I also recommend the chapter *Derivatives, Geometrically* in the book I am writing. I will send this to all of you in case you do not have a copy.

2.7.1 Derivatives as linear approximations

The **classical definition of derivative** for $f : \mathbb{R} \to \mathbb{R}$ is of course $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$. Linear maps from $\mathbb{R}$ to $\mathbb{R}$ are simply lines through the origin, so $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ can be restated as $\lim_{h\to 0} \frac{f(x+h)-(f(x)+L_x(h))}{h} = 0$ where $L_x(h) =$ the linear map $(f'(x))h$. Note that this in turn, can be simply restated as $f(x+h) - (f(x) + L_x h) = o(h)$ or $f(x+h) = f(x) + L_x h + o(h)$. Now we have a definition that generalizes to arbitrary dimensions: $f : B_1 \to B_2$ (and B_1 and B_2 are Banach spaces which are finite or infinite dimensional) is differentiable at $x \in B_1$ if there is a continuous linear operator from B_1 to B_2, L_x, such that $f(x+h) = f(x) + L_x(h) + o(|h|)$. **The linear map $\mathbf{L}_x$ is the derivative of f at x.**

Higher order derivatives for functions $f : \mathbb{R} \to \mathbb{R}$ are pretty simple – simply iterate the classical definition over and over. In higher dimensions, we no longer have such a simple reduction. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ then $Df(x)$ is a linear map that depends on $x \in \mathbb{R}^n$. If f is continuously differentiable, then we have that $Df : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$. The second derivative at any point in $\mathbb{R}^n$ is a Linear map from $\mathbb{R}^n$ to $L(\mathbb{R}^n, \mathbb{R}^m)$, i.e. $D^2 f(x) \in L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$ and $D^2 f : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m))$. Computationally speaking, once we have chosen bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, the first derivative will be an m by n matrix that varies as we move from point to point in $\mathbb{R}^n$, the second derivative m by n by n tensor, the third derivative is a m by n by n by n tensor, etc. So, higher order derivative at a point in the domain can be thought of both as linear map and as multilinear map.

In the following set of exercises, we will use the Hausdorff distance between sets. Suppose that E and F are sets in a metric space X and ρ is the metric. Define $\rho(x, E) = \inf_{y\in E} \rho(x, y)$. Now define $N_\epsilon(E) \equiv \{x|\rho(x, E) \le \epsilon\}$. Define $H(E, F) \equiv \inf\{\epsilon \mid E \subset N_\epsilon(F) \text{ and } F \subset N_\epsilon(E)\}$. We will also use $\rho(E, F) \equiv \inf_{x\in E, y\in F} \rho(x, y)$

Exercise 2.7.1. Suppose that $f(x) = x_1^2 + x_2^2 + x_3^2$. Show that $L_x(h) = 2x_1h_1 + 2x_2h_2 + 2x_3h_3$.

Exercise 2.7.2. Represent $x \in \mathbb{R}^n$ by $x = (x_1, x_2, ..., x_n)$. Suppose that all the partial derivatives of $f : \mathbb{R}^n \to \mathbb{R}$, $f_{x_i} \equiv \frac{\partial f}{\partial x_i}$, exist in $B(\hat{x}, \epsilon)$, and are continuous there. Prove that $L_{\hat{x}}(h) = f_{x_1}(\hat{x})h_1 + f_{x_2}(\hat{x})h_2 + \cdots f_{x_n}(\hat{x})h_n$.

Exercise 2.7.3. Let $f(x) = x^2 \sin(\frac{1}{x})$. Prove that $L_0(h)$, the derivative of $f(x)$ at $x = 0$, exists and equals $0 \cdot h$.

Exercise 2.7.4. Letting $f(x)$ be the function from Exercise 2.7.3, prove that $f'(x)$ is not continuous at $x = 0$.

Exercise 2.7.5. Find an example of a function $f : \mathbb{R}^2 \to \mathbb{R}^1$ and a point $x \in \mathbb{R}^2$, such that f_{x_1} and f_{x_2} exist at x, yet nevertheless, f is not differentiable at x.

Exercise 2.7.6. Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$ is smooth (has derivatives of all orders and all those derivatives are continuous). In this exercise we explore how the set of roots of f changes under simple perturbations. We see why this is relevant for dynamical systems.

1. Suppose that $f(y) = 0 \Rightarrow f'(y) \neq 0$. Prove that the set of roots does not have an accumulation point.

2. Suppose that we focus on a compact subset $K \subset \mathbb{R}$. Show that the set $\{x \mid f(x) = 0, x \in K\}$ is a finite set.

3. Now consider $g(x) \equiv f(x) + \lambda$ and define $R_\lambda \equiv \{x \mid g(x) = 0, x \in K\}$. Assume that $R_0 \subset K^o$ (recall that K^o is the interior of K). $\#R_0 = \#R_\lambda$ for all $|\lambda| \leq \epsilon$ for some $\epsilon > 0$, where $\#S$ is the number of elements of the set S.

4. Suppose that $\inf_{x \in R_0} |f'(x)| > C_1 > 0$ and $\sup_{x \in \mathbb{R}} |f''(x)| < C_2 < \infty$. Find $\delta(C_1, C_2)$ such that $H(R_0, R_\lambda) < \epsilon$ when $\lambda < \delta(C_1, C_2)$.

5. Suppose now that $K = [a, b]$, $f(a) > \alpha$ and $f(b) > \alpha$. Suppose also that $|f''(x)| \neq 0$ whenever $f'(x) = 0$. Again we define $g(x) \equiv f(x) + \lambda$ and we will look at $\lambda \in [-\alpha, \alpha]$. Show that the conditions we have assumed imply that

   (a) the points x in K where $f'(x) = 0$ are isolated, and therefore

   (b) there are a finite number of points x in K where $f'(x) = 0$,

   (c) $\#R_\lambda$ is even except when $\lambda = -f(\hat{x})$ and $f'(\hat{x}) = 0$ in which case it might or might not be even. Give examples of functions in which $\#R_\lambda$ changes and the $\#R_\lambda$ becomes odd for some λ and also examples of functions in which $\#R_\lambda$ changes and $\#R_\lambda$ never becomes odd for any $\lambda \in [-\alpha, \alpha]$.

   (d) Suppose that $\frac{dx}{dt} = g(x) = f(x) + \lambda$. The behavior is determined by the roots and sign of g, as well as the sign of $\frac{dg}{dx}$ at the roots of g. What do you observe about correspondence between the behavior of the solutions of $\frac{dx}{dt} = g(x) = f(x) + \lambda$ and the plot of R_λ versus λ, with the points of R_λ labeled by the sign of g' at those points?

   hint: work this all out for the case in which $f(x) = \frac{1}{6}(x-1)(x-2)(x-3)(x-4)$, $a = 0$ and $b = 5$ and $[-\alpha, \alpha] = [-3.9, 3.9]$. Draw pictures!

Comment: in this problem, we are developing some of the pieces of **bifurcation theory**, the study of the qualitative changes in behavior of dynamical systems as a system parameter or parameters change.

Exercise 2.7.7. Prove that if f is differentiable everywhere, $|f'(x)| \leq C|f(x)|$ everywhere, and $f = 0$ anywhere on the real line, then $f(x) = 0$ everywhere. (This problem was given to me by Yunfeng Hu.)

2.7.2 Cones and Tangent Cones

A **cone**, $\mathbf{C}^{\theta}_{v,\hat{x}}$, around **direction vector** $v \equiv (1, a) \in \mathbb{R}^2)$ and **centered on** $\hat{x} = (\hat{x_1}, \hat{x_2})$ is the set of $x = (x_1, x_2)$ such that $\cos^{-1}(\left|\frac{x-\hat{x}}{|x-\hat{x}|} \cdot \frac{v}{|v|}\right|) \leq \theta$, where we assume that $\theta < \pi/2$. An **equivalent cone-based definition of derivative**, says that $f : \mathbb{R}^1 \to \mathbb{R}^1$ is differentiable at $\hat{x}$ if there is a direction $w_{\hat{x}} = (1, f'(\hat{x}))$ such that for any $\epsilon > 0$, there is a $\delta > 0$ such that $x \in B(\hat{x}, \delta)$, all $(x, f(x)) \in C^{\epsilon}_{v,\hat{x}}$. This definition generalizes to all dimensions.

Define the projection of a set F onto the unit ball to be $P_1(F) \equiv \{\frac{x}{|x|} \mid x \in F\}$. The **tangent cone** of $E \subset \mathbb{R}^n$ at a point $x \in \mathbb{R}^n$, is the set $\{\mathbb{R} \geq 0\} \cdot \{\bigcap_i \text{clos}(P_1(B(x, \frac{1}{i}) \cap E \setminus \{x\}))\}$.

Exercise 2.7.8. Prove that for $f : \mathbb{R}^1 \to \mathbb{R}^1$, the cone based definition of derivative is equivalent to the linear approximation definition.

Exercise 2.7.9. Generalize the cone-based definition of derivative to the case of $f : \mathbb{R}^n \to \mathbb{R}^1$.

Exercise 2.7.10. Generalize the cone-based definition of derivative to the case of $f : \mathbb{R}^n \to \mathbb{R}^m$.

Exercise 2.7.11. Show that f is differentiable at $\hat{x}$ with derivative $f'(\hat{x})$ if and only if the tangent cone of the graph of f in $\mathbb{R}^2$, at $(\hat{x}, f(\hat{x}))$, is the line $x_2 = f'(\hat{x})x_1$.

Exercise 2.7.12. Find a 1-dimensional subset of $\mathbb{R}^2$ that is not the graph of any $f : \mathbb{R}^1 \to \mathbb{R}^1$, yet it still has tangent cones at every point equal to lines in $\mathbb{R}^2$. **Hint**: What kind of curve that does not cross itself cannot be a graph no matter how you rotate the 2-dimensional space the curve live in?

2.7.3 Taylor Series

As far as I know Newton was the first to use power series to solve differential equations, though I suspect, like many other things, others were there or almost there before he moved in this direction. At any rate, the idea of using power series to approximate functions locally and globally has been around for a long time. In this section, we look at Taylor series, which are always better and better approximations for f as the differentiability of f increases. (This is true even when the Taylor series diverges.) The **Taylor series** of a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ at a point a is given by $\sum_{n=0}^{\infty} \frac{f^n(a)(x-a)^n}{n!}$ where f^n denotes $\frac{d^n}{dx^n} f$. We define $T^{a,k}_f(x) \equiv \sum_{n=0}^{k} \frac{f^n(a)(x-a)^n}{n!}$. Note that the Taylor series need not converge even if f is infinitely differentiable or it may converge but not be equal to the function in any neighborhood of a.

If a function has $k+1$ derivatives at $x = a$ and those derivatives are continuous in $B(a, \epsilon) = [a - \epsilon, a + \epsilon]$, then we have that in this neighborhood $f(x) = T^{a,k}_f(x) + \frac{f^{k+1}(c)(x-a)^{k+1}}{(k+1)!}$ for some $c \in [a, x]$ for $x > a$ $c \in [x, a]$ if $x < a$. The **remainder term** $\frac{\mathbf{f^{k+1}(c)(x-a)^{k+1}}}{\mathbf{(k+1)!}}$ is obtained using the mean value theorem.

In higher dimensions, we have a completely analogous result: Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$. Fix $a \in \mathbb{R}^n$. Define for any $x \in \mathbb{R}^n$, define $h = x - a$. Let $D^k F : \mathbb{R}^n \to L(\mathbb{R}^n, L(\mathbb{R}^n, ...L(\mathbb{R}^n, \mathbb{R}^m)...))$ be the nth derivative (represented by an $m \times \underbrace{n \times n \times ... \times n}_{\text{k n's}}$ tensor) taking k input vector increments.

Then **the Taylor series of F at a is** $\sum_{i=0}^{\infty} \frac{D^i f(a) h^i}{i!}$ where $D^i f(a)h^i = D^i f(a)(\underbrace{h, ..., h}_{\text{k h's}})$. Defining $T_F^{a,k} = \sum_{i=0}^{k} \frac{D^i f(a) h^i}{i!}$, we get that if the first k derivatives exist and are continuous in $B(a, \delta)$, and the $k+1$ derivative exists everywhere in $B(a, \delta)$, then for any $x \in B(a, \delta)$, $F(x) = T_F^{a,k} + \frac{D^{k+1} f(c) h^{k+1}}{(k+1)!}$ for some $c \in B(a, \delta)$.

Exercise 2.7.13. Look up and read the proofs of the Taylor series with remainder expressions on pages 386 and 96 of Fleming's book.

Exercise 2.7.14. In the cone based definition of derivative, the width of the cone $C^{\theta}_{w_{\hat{x}}, \hat{x}}$, θ decreases as the ball about $\hat{x}$ decreases in radius. Suppose that f'' ...

Exercise 2.7.15. 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 2.7.16. Prove that if all the coefficients of the Taylor polynomial for $f(x)$ are positive, and the series converges to $f(x)$ for all x, then $f(x) \underset{x \to \infty}{\to} \infty$.

Exercise 2.7.17. 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 2.7.18. 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 2.7.15. Confirm these are solutions by direct differentiation and substitution into the differential equations.

2.7.4 The Trickier (cooler) Approximation

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 2.7.1. *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 2.7.1.
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 2.7.1. *if $f(x) = o(x^k)$ then $\int_0^x f(y)dy = o(x^{k+1})$.*

   Proof of Lemma 2.7.1.
   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 2.7.19. Work through the details of step 5 above.

Exercise 2.7.20. explain why this theorem is a generalization of the definition of derivative to high dimensions and exactly how the result varies from the result on remainders of Taylor series from Section 2.7.3.

Exercise 2.7.21. Give an example of a function that is differentiable at $x = 0$ but differentiable anywhere else.

Exercise 2.7.22. Prove that if a function $f(x)$ has a second derivative at a point $x = a$, then in some neighborhood of a, $f(x)$ is Lipschitz.

Exercise 2.7.23. (*) 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!)

2.8 Inverse and Implicit Function Theorems

The **inverse function theorem** tells us that if $F : \mathbb{R}^n \to \mathbb{R}^n$ is C_1, then that function is invertible locally at x if it's linear approximation at x, $DF(x)$ is invertible at x. And in that case, $D(F^{-1})(x) = ((DF)(x))^{-1}$.

Suppose that $m < n$. The **implicit function theorem** says that we can represent a level set of a function $F : \mathbb{R}^n \to \mathbb{R}^m$ passing through $\hat{x}$ as a graph of a function $\phi : \mathbb{R}^{n-m} \to \mathbb{R}^m$ when F is C^1 and there is some set of m columns of $DF(x)$ that are linearly independent. That is, if there is an invertible $m \times m$ submatrix of $DF(\hat{x})$, we can find such a ϕ. Recall that an $m \times n$ matrix M, $m \leq n$, is **full rank** if it has m independent columns.

Now suppose, without loss of generality, that the first m columns are independent and that we represent $x \in \mathbb{R}^n$ as $x = (y, z)$ where $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^{n-m}$. Therefore, $\hat{x} = (\hat{x_1}, \hat{x_2})$, $x_1 \in \mathbb{R}^m$ and $x_2 \in \mathbb{R}^{n-m}$. For the ϕ the implicit function theorem gives us, for some $\epsilon > 0$, we have that for $z \in B(\hat{x}_2, \epsilon) \subset \mathbb{R}^{n-m}$, $F(\phi(z), z) = F(x)$.

We can prove the implicit function theorem using the inverse function theorem.

Exercise 2.8.1. Look up the proof of the inverse function and implicit function theorems in Fleming's book and study them. You also might like the two chapters from the notes I wrote for another course – I have included them here as Appendix A and B.

Exercise 2.8.2. Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$ and $f \in C^1(\mathbb{R}^1, \mathbb{R}^1)$, the space of C^1 functions from $\mathbb{R}^1$ to $\mathbb{R}^1$. At what points in the domain of f, is there a neighborhood such that, when restricted to that neighborhood, f is invertible?

Exercise 2.8.3. Suppose that $f(x) = x^3 - 4x + 1$. Where is f locally invertible? For any given $x \in \mathbb{R}$ where f is locally invertible, what is the maximal set of points E_x such that $x \in E_x$ and $f|_{E_x}$ is invertible?

Exercise 2.8.4. Suppose that $f(x) = x^2 - 4x + \frac{7}{4}$. Where is f locally invertible? For any given $x \in \mathbb{R}$ where f is locally invertible, what is the maximal set of points E_x such that $x \in E_x$ and $f|_{E_x}$ is invertible? What is the inverse function in those regions?

Exercise 2.8.5. Define $f : \mathbb{R}^2 \to \mathbb{R}^2$ by $f_1(x_1, x_2) = x_1^2 + x_2^2$ and $f_1(x_1, x_2) = x_1 + x_2$. Where is f locally invertible? For each x where x is locally invertible, what is the maximal E_x such that $x \in E_x$ and $f|_{E_x}$ is invertible?

Exercise 2.8.6. Define $f : \mathbb{R}^2 \to \mathbb{R}^1$ by $f(x_1, x_2) = x_1^2 + x_2^2$. At what points in $\mathbb{R}^2$ can the level sets of f be expressed locally as a graph of a function $\phi : \mathbb{R}^1 \to \mathbb{R}^1$?

Exercise 2.8.7. Define $f : \mathbb{R}^3 \to \mathbb{R}^2$ by $f_1(x_1, x_2, x_3) = x_1^2 - x_2^2 + x_3$ and $f_1(x_1, x_2) = x_1 - x_2$. At what points in $\mathbb{R}^3$ can the level sets of f be expressed locally as a graph of a function $\phi : \mathbb{R}^1 \to \mathbb{R}^2$? What is ϕ at those points?

Exercise 2.8.8. Define $f : \mathbb{R}^3 \to \mathbb{R}^1$ by $f(x_1, x_2, x_3) \equiv P_1(x_1) + P_2(x_2) + P_3(x_3)$ where The P_i are polynomials of degree 2. How would you go about finding the points where you can express the level sets of f as graphs of functions $\phi : \mathbb{R}^2 \to \mathbb{R}^1$. Choose specific second order polynomials and find all the points where you can find a locally valid ϕ.

Exercise 2.8.9. For any $f : \mathbb{R}^n \to \mathbb{R}^1$ what is the criterion for determining if the level set at x is locally an n-1-dimensional graph? (I.e. a graph of a function from $\mathbb{R}^{n-1}$ dimensions to $\mathbb{R}^1$.)

Exercise 2.8.10. Assume that $m < n$, $F : \mathbb{R}^n \to \mathbb{R}^m$ and $DF(x^*)$ is full rank. Use the facts that (1) DF is full rank when the all the rows of DF are linearly independent and (2) the gradient of a function is orthogonal to (normal to) the level sets of that function, to prove that the level set a C^1 $F : \mathbb{R}^n \to \mathbb{R}^m$ is locally n-m-dimensional at x if and only if the linear approximation to the level set of F at x is n-m-dimensional.

Exercise 2.8.11. Construct another geometric understanding of the implicit function theorem by proving that: If there is an $m \times m$ invertible submatrix of the $m \times n$ DF, the m-dimensional coordinate plane associated with those m columns intersects the level set of F at one point only, for each choice of the other n-m coordinates, as we stay in an ϵ neighborhood of x.

Exercise 2.8.12. The fact that the ϕ obtained by the implicit function theorem is smooth allows us to conclude that it has a n-m-dimensional tangent space at every point. Prove that all unit vectors tangent to the graph of ϕ at x do not lie in the span of the m linearly independent columns (i.e. those selected in Exercise 2.8.11). Prove that, in fact, the minimum angle between those tangent vectors and the m-dimensional coordinate plane corresponding to those m columns is bounded away from 0.

2.9 Convexity

Let $f : \mathcal{H} \to \mathbb{R}$ and $E \subset \mathcal{H}$, where $\mathcal{B}$ is any complete inner product vector space (also called a **Hilbert space**).

A **function f is convex** if $f(\alpha_1 x+\alpha_2 y) \leq \alpha_1 f(x)+\alpha_2 f(y)$ for all $\alpha_i \geq 0$ satisfying $\alpha_1+\alpha_2 = 1$ and a **function f is strictly convex** if $f(\alpha_1 x + \alpha_2 y) < \alpha_1 f(x) + \alpha_2 f(y)$ for all $\alpha_i > 0$ satisfying $\alpha_1 + \alpha_2 = 1$.

A **set** E **is convex** if for any two points x and y in the set, $\alpha_1 x + \alpha_2 y$ is also in the set for all $\alpha_i \geq 0$ satisfying $\alpha_1 + \alpha_2 = 1$. A **set** E **is strictly convex** if, for $x, y \in E$, $\alpha_1 x + \alpha_2 y$ is in the interior of E when $\alpha_i > 0$ and $\alpha_1 + \alpha_2 = 1$.

The theory of convex sets and functions is a very rich subject. In nonlinear analysis, these are the nicest sets and functions where everything in sight behaves as it should. You will encounter some of this good behavior in the exercises below.

In all the exercises in this section, (1) we assume E is closed and convex and (2) the notation carries over from one exercise to the next.

Exercise 2.9.1. Define $d(x, E) \equiv \inf_{y\in E} |x - y|$ where $|\cdot|$ is the usual 2-norm in $\mathbb{R}^n$. Prove that if x is a point not in E, then there is a unique closest point $y_x \in E$.

Exercise 2.9.2. Denote the hyperplane through y_x, orthogonal to $x-y_x$ by $h_{y_x,x-y_x}$. Let $H_{y_x,x-y_x}$ denote the closed halfspace defined by $h_{y_x,x-y_x}$ such that for which $x - y_x$ is an outward pointing normal vector. Show that E lies entirely in the closed halfspace $H_{y_x,x-y_x}$. A hyperplane that intersects the boundary of E and contains E in one of the halfspaces it defines is called a **supporting hyperplane**. Hint: see if you can prove that $\langle y - y_x, x - y_x\rangle$ is always non-positive.

Exercise 2.9.3. Prove that $E = \bigcap_{x\in E^c} H_{y_x,x-y_x}$.

Exercise 2.9.4. (*) Show that the level sets of the distance function, $L_E(c) \equiv \{x|d(x, E) = c > 0\}$, have tangent planes at every point of the level set. Show that those tangent planes are continuous with respect to variation along the level set. Hint: if $x \in L_E(c)$ show that there is a δ small enough that for $w \in B(x, \delta) \cap L_E(c)$, $\epsilon \leq \langle x - y_x, w - x\rangle \leq 0$.

Exercise 2.9.5. (*) Use Exercise 2.9.4 to show that every point on the boundary of E has a supporting hyperplane through it. Hint if $y \in \text{bdy}(E)$ and it is not the nearest point for some $x \in L_E(1)$, then $d(y, L_E(1)) > 1$ and since $L_E(1)$ is closed and $\{y\}$ is compact, there is a point $x \in L_E(1)$ such that $d(y, L_E(1)) = |x - y| > 1$. But there must also be a closer point $y_x \in \text{bdy}(E)$ such that $d(y_x, x) = 1$.

Exercise 2.9.6. Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$ and f is convex. Show that the left derivatives and right derivatives, $f'_L(x) \equiv \lim\limits_{y\uparrow x} \frac{f(x)-f(y)}{x-y}$ and $f'_R(x) \equiv \lim\limits_{y\downarrow x} \frac{f(x)-f(y)}{x-y}$, exist at each point in the domain and that $f'_L(x) = f'_R(x) = f'(x)$ except when $x \in J \subset \mathbb{R}^1$, where J is at most countably infinite.

Exercise 2.9.7. Suppose that $f : D \subset \mathbb{R}^1 \to \mathbb{R}^1$, where D is a closed (possibly infinite) interval in $\mathbb{R}$. Show that **epigraphs** $E_f \equiv \{(x, y) \mid f(x) \leq y\}$ are convex and closed in $\mathbb{R}^2$. Show that f is a convex function if and only if the epigraph E_f is convex.

Exercise 2.9.8. Assume that $f'(x^*)$ exists. Show that the tangent line to f at x^*, $\{(x, y) \mid f'(x^*)x + (f(x^*) - f'(x^*)x)\}$, is a supporting (1-dimensional) hyperplane of the epigraph E_f at $(x, f(x))$.

Exercise 2.9.9. Suppose that E_α for all $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is an arbitrary index set (not necessarily countable). Prove that $E \equiv \bigcap_{\alpha\in\mathcal{A}} E_\alpha$ is convex.

Exercise 2.9.10. Define $f_M(x) = \sup_{f\in\mathcal{F}} f(x)$ where $\mathcal{F}$ is a class of uniformly bounded, convex functions, $f : [a, b] \to \mathbb{R}$ and $[a, b]$ is a bounded interval. Show that F_M is a convex function. Hint: What is the relationship between the epigraphs if the f's in $\mathcal{F}$ and the epigraph of F_M.

Exercise 2.9.11. The *uniformly bounded* condition in Exercise 2.9.10 is not actually necessary, but we assumed it to avoid dealing with functions that take on the value $+\infty$. Now we allow infinite values. Such functions take on values in the extended reals, $f : \mathbb{R} \to \bar{\mathbb{R}}$ where $\bar{\mathbb{R}} = \{\mathbb{R} \cup \{-\infty, \infty\}\}$. Prove that f is convex if and only if the epigraph is convex in $\mathbb{R}^2$. Note: the epigraph is still $\{(x, y) \mid f(x) \leq y < \infty\} \subset \mathbb{R}^2$. Define $f_M(x) = \sup_{f\in\mathcal{F}} f(x)$, where $\mathcal{F}$ is any class of convex functions $f : \mathbb{R}^1 \to \mathbb{R}^1$. Prove that F_M is convex.

Exercise 2.9.12. The definition of epigraph $E_f \subset \mathcal{H} \times \mathbb{R}$ is identical for any $f : \mathcal{H} \to \mathbb{R}$ where $\mathcal{H}$ is a complete inner product space. Go through the arguments proving f is convex if and only if E_f is convex for the case $f : \mathbb{R}^1 \to \bar{\mathbb{R}}$ to see that they translate without change to the case $f : \mathcal{H} \to \bar{\mathbb{R}}$.

Exercise 2.9.13. A function f is said to be concave if -f is convex and is said to be strictly concave if -f is strictly convex. Prove that f is concave if $f(\alpha_1 x + \alpha_2 y) \geq \alpha_1 f(x) + \alpha_2 f(y)$ for all $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 = 1$.

Exercise 2.9.14. Let E be a bounded, closed, convex subset of $\mathbb{R}^2$. Let D be the projection of E onto the x-axis. Define $f_E : D \to \mathbb{R}$ by $f_E(x) = \mathcal{H}^1(\{\{x\} \times \mathbb{R}\} \cap E)$. Show that f_E is concave.

Exercise 2.9.15. Prove that every line through $(x, f(x))$ with slopes ranging from $f_L(x)$ to $f_R(x)$ are supporting lines for f at $(x, f(x))$.

Exercise 2.9.16. Let H_f be the collection of supporting lines of the convex function f. Show that $f(x) = g(x) \equiv \sup\limits_{h\in H_f} h(x)$. Consequently, the epigraph of f is the intersection of the upper halfplanes defined by the supporting lines.

Exercise 2.9.17. Let $f \in C^2(\mathbb{R}^1, \mathbb{R}^1)$ and suppose that $f''(x) \geq 0$ for all $x \in \mathbb{R}$. Show that f is convex. Hint: consider $g(x) = f(x) - f(x^*) - f'(x^*)(x - x^*)$ and use what you know about Taylor series to compute $g(x)$.

Exercise 2.9.18. Suppose that $f \in C^2(\mathbb{R}^1, \mathbb{R}^1)$ and f is convex. show that $f''(x) \geq 0$ for all $x \in \mathbb{R}$

Exercise 2.9.19. Let $f, g \in C^2(\mathbb{R}^1, \mathbb{R}^1)$ be convex. Assume also that f and g are (a) non-negative and (b) have derivatives whose signs always agree. Prove that $w \equiv fg$ is also convex. Give examples to demonstrate why conditions (a) and (b) are both necessary.

Exercise 2.9.20. Suppose $f : \mathbb{R}^n \to \mathbb{R}^1$ is convex. Show that the sublevel sets $S_f(c) \equiv \{x \in \mathbb{R}^n \mid f(x) \le c\}$ are convex. Give an example of a non-convex function g whose sublevel sets $S_g(c)$ are all convex.

Exercise 2.9.21. If $f(x) \to \infty$ as $|x| \to \infty$, then we say **f is coercive**. To be completely clear, we mean that for every $c > 0$, there exists an $N > 0$ such that $|x| > N$ implies $f(x) > c$. Prove that a coercive, convex function f has a minimal value f_m and that the set $M \equiv \{x \mid f(x) = f_m\}$ is convex. Hint: choose a ball $B(0, C) = \{x \mid |x| \le C\}$ big enough that $f(x) > 2f(0)$ for $x \in B(0, C)^c$ and use the fact that $\{x \quad f(x) \le C\}$ is a compact set.

Exercise 2.9.22. Give an example of a convex function that does not have a minimal value.

Exercise 2.9.23. We will say that **f is directionally coercive** if, for all $v \in \partial B(0, 1) \subset \mathbb{R}^n$, $f(sv) \underset{s\to\infty}{\to} \infty$. Prove that when f is convex, directionally coercive implies coercive. Hint: suppose that $f(0) = \alpha$. Define $R(v) = \sup\{r|f(sv) < 2|\alpha| \;\forall s < r\}$: i.e. $r = R(v)$ is the smallest radius for which $f(rv) = 2|\alpha|$. Now suppose that $\sup\limits_{v\in\partial B(0,1)} R(v) = \infty$. Because $\partial B(0, 1)$ is compact, there is a v^* that is the limit of v_i's such that $R(v_i)$ diverges as $i \to \infty$. By taking points on the rays in the directions v_i, we can prove that $f(sv^*) \le 2|\alpha|$ for all $s > 0$. This is a problem. Also, because f is convex, for $s > R(v)$, $f(sv) \ge 2|\alpha|$.

Exercise 2.9.24. (*) Give an example $f : \mathbb{R}^2 \to \mathbb{R}^1$ that is continuous and directionally coercive but not coercive.

Exercise 2.9.25. Show that a coercive, strictly convex function has a unique minimizer x^* such that $f(x^*) < f(x)$ for all $x \ne x^*$.

Exercise 2.9.26. (*) Suppose that $f : x \in \mathbb{R}^n \to y \in \mathbb{R}^1$, $f \in C^1(\mathbb{R}^n, \mathbb{R}^1)$ and that f is minimal at x^*. Prove that the hyperplane $y = \langle 0, x\rangle + f(x^*)$ is a supporting hyperplane of the function at $(x, f(x^*))$. Show that $y = h(x) = \langle \nabla f(z), x - z\rangle + f(z)$ is a supporting hyperplane at $(z, f(z))$. Use the fact that f is convex to conclude that if x is not (globally!) minimal, then $\nabla f \ne 0$.

Exercise 2.9.27. Even though Exercise 2.9.26 implies that gradient descent cannot converge unless we are converging to a minimizer, we are not guaranteed we are converging very fast.

1. Construct a convex function $f : \mathbb{R}^1 \to \mathbb{R}^1$, for $x = 0$ is the unique minimizer, such that if $\frac{dx}{dt} = -f'(x)$, where t is in seconds, it still takes 10^{100} seconds to travel from -1 to 0 unit of distance. Hint: play with $f(x) = |x|$.

2. (*) Create a smooth, strictly convex f with (unique) minimizer at $x = 0$, such that the time it takes to descend the gradient (i.e. follow evolution in the domain specified by the differential equation $\dot{x}(t) = -f'(x)$) from $x = 1$ to $x = 0$ is (a) $T < \infty$ or (b) $T = \infty$. Hint: consider $f(x) = |x|^\alpha$.

Exercise 2.9.28. Define f^*, the Legendre-Fenchel transform of f, by

$$f^*(k) \equiv \sup_{x \in \mathbb{R}^n} (\langle k, x\rangle - f(x))$$

, where k is in the dual space to $\mathbb{R}^n$ which we have identified, via the inner product with $\mathbb{R}^n$. In other words, k lives in the space of gradients. Transforming again,

$$f^{**}(x) \equiv \sup_{k \in \mathbb{R}^n} (\langle k, x\rangle - f^*(k))$$

where now x is in the double dual to R^n which is just R^n. **Prove that** if $f : \mathbb{R}^n \to \mathbb{R}^1$ is convex, then $f^{**} = f$. Hint: note that $h(x) \equiv \langle k, x\rangle - f^*(k)$ is a supporting plane for the function f. **Note:** f* frequently attains infinite values.

Exercise 2.9.29. Assume $f : \mathbb{R}^1 \to \mathbb{R}^1$. Compute f^* and f^{**} when:

1. $f(x) = |x|$
2. $f(x) = x^2$
3. $f(x) = \begin{cases} \infty & x < -1 \\ 0 & -1 \le x \le 1 \\ \infty & 1 < x \end{cases}$

Exercise 2.9.30. Prove that f^{**} is always convex even if f is not.

2.10 Inequalities

Inequalities are at the center of analysis: pick up any advanced book on analysis, and that will stand out. You have already encountered this in these problems/notes, but now we dive into the classical inequalities.

When we are working in an inner product space, inequalities of the form $\langle w, x\rangle < c$ or $\langle w, x\rangle \le c$ pop up all the time. Understanding that there are the open and closed half spaces bounded by $\langle w, x\rangle < c$, with outward pointing normal w, allows us to reason with these inequalities much more intuitively and efficiently. In general, inequalities of the form $f(x) < c$ or $f(x) \le c$ are **sublevel** sets of f, and in the case that f is continuous, are open and closed, respectively.

Of course, even the most basic definitions in analysis depend on inequalities: f is continuous at some fixed x^* if for every $\epsilon > 0$ there is a $\delta > 0$ such that $|y - x^*| < \delta$ implies that $|f(y) - f(x^*)| < \epsilon$.

Probably the simplest inequality that is nevertheless central to the study of inner product spaces is the Cauchy-Schwarz inequality.

2.10.1 Cauchy-Schwarz

Suppose that $|x|$ is the 2-norm, $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$. Then

$$|\mathbf{x} \cdot \mathbf{y}| \le |\mathbf{x}||\mathbf{y}|.$$

Note that the left hand side is the 2-norm of a scalar i.e. the absolute value of a scalar. Actually, if we are in an inner product space H (e.g a Hilbert space) and $\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{H}}$ is the inner product of f and g in H, then we have that

$$|\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{H}}| \leq |\mathbf{f}||\mathbf{g}|$$

where $|\mathbf{f}| = |\langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{H}}|^{\frac{1}{2}}$.

Exercise 2.10.1. Prove **Cauchy's inequality** which states that for all $a, b \in \mathbb{R}$, $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$

Exercise 2.10.2. Prove **Cauchy's inequality with** ϵ which states that for all $a, b \in \mathbb{R}$, $ab \leq \frac{\epsilon a^2}{2} + \frac{b^2}{4\epsilon}$.

Exercise 2.10.3. Prove the Cauchy-Schwarz inequality in an real inner product space. Hint: Prove it for the case in which $|f| = |g| = 1$ and notice that the general case reduces to that case very quickly. Now consider the fact that $0 \leq \langle f + \epsilon g, f + \epsilon g \rangle$ for all ϵ including $\epsilon = \pm 1$.

Exercise 2.10.4. Prove the Cauchy Schwarz inequality in the case you are working in an inner product space over the complex numbers. (Recall that $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, $\langle x, \alpha y \rangle = \bar{\alpha} \langle x, y \rangle$ and $\langle x, y \rangle = \overline{\langle y, x \rangle}$.) Hint: You can get this by using the fact that $0 \leq \langle x + \alpha y, +\alpha y \rangle$ for all complex α. Computing, we get

$$\langle x + \alpha y, +\alpha y \rangle = \langle x, x \rangle + \alpha \langle y, x \rangle + \bar{\alpha} \langle x, y \rangle + |\alpha|^2 \langle y, y \rangle .$$

Now let $be^{i\theta} = \langle y, x \rangle$ and $\alpha = te^{-i\theta}$. Notice that the resulting expression is a positive quadratic in t ... (source: Conway's *A course in Functional Analysis*).

Exercise 2.10.5. Notice that the CS-inequality in $\mathbb{R}^2$ has been known to you for a long time; in fact you have a more informative form of it from pre-calculus, $x \cdot y = \cos(\theta)|x||y|$. In arbitrary real inner product spaces, we *define* the angle between two unit vectors to be $\arccos\left(\frac{\langle x,y \rangle}{|x|\ |y|}\right)$. **Prove** $x \cdot y = \cos(\theta)|x||y|$ in $\mathbb{R}^2$ using only facts about trig functions and rotations in $\mathbb{R}^2$. Hint: prove that rotations of the coordinate system leave dot products unchanged.

Exercise 2.10.6. Let H and K be two linear subspaces of $\mathbb{R}^n$ and suppose that $H \not\subset K$ and $K \not\subset H$. Define $P_S(x)$ to be the projection of x onto the subspace S. We define the angle $\angle(H, K) = \arccos\left(\min_{x \in B(0,1) \cap H} \left(\max_{y \in B(0,1) \cap K} \langle x, y \rangle\right)\right)$. **Find example** spaces H and K demonstrating that $\angle(H, K) \neq \angle(K, H)$ can happen. Hint: show that this angle measure tells us the maximum rotation needed to rotate any vector in H so that it is contained in K.

2.10.2 Jensen's

The most elementary form of Jensen's inequality is really just the inequality used to define convex functions: $\mathbf{f} : \mathbb{R}^{\mathbf{n}} \to \mathbb{R}^{\mathbf{1}}$ is convex if

$$\mathbf{f}(\alpha \mathbf{x} + (\mathbf{1} - \alpha)\mathbf{y}) \leq \alpha \mathbf{f}(\mathbf{x}) + (\mathbf{1} - \alpha)\mathbf{f}(\mathbf{y})$$

Now suppose that $\sum_i \alpha_i = 1$, $\alpha_i \geq 0$ for $i = 1, ..., n$, $\alpha(x) \geq 0$ and $\int \alpha(x)\ dx = 1$. **Jensen's inequality** generalizes this to

$$\mathbf{f}\left(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \alpha_{\mathbf{i}} \mathbf{x}_{\mathbf{i}}\right) \leq \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \alpha_{\mathbf{i}} \mathbf{f}(\mathbf{x}_{\mathbf{i}})$$

and even to

$$\mathbf{f}\left(\int \alpha(\mathbf{x})\mathbf{x}\ \mathbf{dx}\right) \leq \int \alpha(\mathbf{x})\mathbf{f}(\mathbf{x})\ \mathbf{dx};$$

i.e. if f is convex, f satisfies both of these generalizations of the inequality satisfied by definition. Of course the opposite inequalities hold for concave functions.

Exercise 2.10.7. Prove both forms of Jensen's inequality. Hint: prove the second (integral) form for nice smooth compactly supported functions α that satisfy $\int \alpha(x)\ dx$, then approximate arbitrary positive α satisfying $\int \alpha(x)\ dx$ with smooth α's to get the desired results.

Exercise 2.10.8. Use Jensen's inequality to prove **Young's Inequality**, which states that $a, b \in \mathbb{R}$, $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ for $\frac{1}{p} + \frac{1}{q} = 1$ and $a, b > 0$. Hint: e^x is convex.

Exercise 2.10.9. Prove that $2e^{\frac{y}{2}} \leq 1 + e^y$ and find the only point where equality holds. Prove that this is the only point. Hint: e^x is strictly convex and $e^0 = 1$.

Exercise 2.10.10. Choose any $\alpha > 1$ and assume that $y \geq 0$. Prove that $(1+y)^\alpha \leq 2^{\alpha-1}(1+y^\alpha)$. Again find the one value of y where equality holds and prove that this is the only value where equality is obtained. Hint: x^α (we assume $x \geq 0$) is strictly convex for $\alpha > 1$

2.10.3 am-gm

The **arithmetic mean** and the **geometric mean** are monotonically related to one another: the arithmetic mean is always greater than the geometric mean. But in fact, this is even true with the generalized arithmetic and geometric means. Suppose that $\sum p_i = 1$ and $p_i \geq 0$ then we have that

$$\Pi_{i=1}^n a_i^{p_i} \leq \sum_{i=1}^n p_i a_i.$$

Choosing $p_i = \frac{1}{n}$ for $i = 1, ..., n$ we obtain the **classic am-gm inequality**:

$$\sqrt[n]{\Pi_{i=1}^n a_i} \leq \frac{1}{n}\sum_{i=1}^n a_i.$$

Exercise 2.10.11. Prove the generalized am-gm inequality. Hint: use Jensen's inequality.

2.10.4 Hölder

Suppose that $\frac{1}{p}+\frac{1}{q}=1$ and $\int |f|^p\ dx < \infty$ and $\int |g|^q\ dx < \infty$. Then we have **Hölder's Inequality**

$$\int \mathbf{|fg|\ dx} \leq \left(\int \mathbf{|f|^p\ dx}\right)^{\frac{1}{p}} \left(\int \mathbf{|g|^q\ dx}\right)^{\frac{1}{q}}$$

. Restating this using norm notation, we get:

$$\int \mathbf{|fg|\ dx} \leq \mathbf{|f|_p|g|_q}$$

Exercise 2.10.12. Prove Hölder's inequality using Young's inequality from Exercise 2.10.8. Hint: assume that $|f|_p = |g|_q = 1$ and then get the general case.

Exercise 2.10.13. Prove Hölder's inequality using the am-gm inequality. Hint: this route actually proves Young's using am-gm and then Holder's follows from that as in Exercise 2.10.12.

2.10.5 Minkowski's

Minkowski's inequality is the triangle inequality for L^p spaces. It states that
$\mathbf{|f+g|_p \leq |f|_p + |g|_p}$, where $|h|_p \equiv \left(\int |h|^p\ dx\right)^{\frac{1}{p}}$.

Exercise 2.10.14. Prove Minkowski's inequality. Hint:

$$\begin{aligned} |f+g|^p &= |f+g|^{p-1}|f+g| \\ &\leq |f+g|^{p-1}(|f|+|g|) \end{aligned}$$

2.10.6 Other Inequalities

The most famous, overtly geometric, inequality is the isoperimetric inequality which states that, given a fixed $C > 0$ the n-dimensional body E that maximizes $\mathcal{H}^n(E)$ given the constraint that $\mathcal{H}^{n-1}(\partial E) = 1$ is the ball of radius $r = \left(\frac{1}{n\omega(n)}\right)^{n-1}$. This is the same answer that you obtain in the dual problem of minimizing surface area with fixed volume. Balls are optimal – they have the lowest $\frac{\text{surface area}}{\text{volume}}$ ratio. We can restate this as

$$\frac{\mathcal{H}^{n-1}(\partial E)}{(\mathcal{H}^n(E))^{\frac{n-1}{n}}} \geq n\omega(n)^{\frac{1}{n}}.$$

More often we see it stated as

$$\mathcal{H}^{n-1}(\partial E) \geq n\omega(n)^{\frac{1}{n}}(\mathcal{H}^n(E))^{\frac{n-1}{n}}$$

which is call the **isoperimetric inequality**.

Exercise 2.10.15. Show that minimizing surface area for a fixed volume gives the solution pairs as does maximizing volume with a fixed surface area constraint. Note: volume here is n-dimensional volume while surface area is n-1-dimensional volume of the boundary of the set in question.

Exercise 2.10.16. Look up a proof of the isoperimetric inequality in the case that $n = 2$ and you are trying to prove that disks minimize the boundary length for regions with fixed area.

The solution of every variational problem leads to an inequality: if (1) $\alpha = \inf_x f(x)$ or (2) $\alpha = \inf_x \frac{f(x)}{g(x)}$, then (1') $f(x) \geq \alpha$ or (2') (assuming $g(x) > 0$ for all x) $f(x) \geq \alpha g(x)$. Simple inequalities like this are used all the time. For example, since $K(A) \equiv |A|$ – the norm of the linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$ – is the solution to $\sup_{x\neq 0} \frac{|Ax|}{|x|}$, we know that $|Ax| \leq K(A)|x|$ for all x. For example, this immediately implies that volumes are expanded by A by a factor of at most $K(A)^n$, which is sometimes all we need. (While $\det(A)$ is the exact expansion factor, computing the largest eigenvector of a positive definite linear operator is easier than calculating the determinant of the same operator, so this is actually practically useful as well.)

In probability, a central inequality is the **Markov's inequality** Assume that $f(x) \geq 0$ for all x. Then:

$$\mu\{x|f(x) > \delta\} \leq \frac{\int f(x)\, d\mu}{\delta}$$

Chebyshev's inequality is just an application of Markov's inequality to the function $|x - \lambda|^2$, where $\lambda \equiv \int x\, d\mu$ and μ is a probability measure:

$$\mu\{x||x - \lambda|^2 > \delta^2\} \leq \frac{\int |x - \lambda|^2\, d\mu}{\delta^2}.$$

Using probability notation we get:

$$\mathbb{P}\{|x - \lambda|^2 > \delta^2\} \leq \frac{\mathbb{E}[|x - \lambda|^2]}{\delta^2}.$$

Exercise 2.10.17. Prove Markov's inequality.

Rounding off this section on inequalities, we mention that there are of course an endless parade of inequalities we can study. In the last few exercises we take a look at a few more.

Exercise 2.10.18. Observe that $g(x) \leq h(x)$ for all x implies that $f(g(x)) \leq f(h(x))$ for all x as long as f is monotonically increasing. Use this to show that for $x \geq 0$, we have $\sqrt{3x} \leq \sqrt{x^2 - x + 4}$.

Exercise 2.10.19. Prove that $\frac{(e^x)^2}{e} \leq e^{x^2}$ find the single point where equality holds and show that equality only holds at that point. Hint: $(x - 1)^2 \geq 0$ for all x.

Exercise 2.10.20. Show that

1. $f(a) = g(a)$,
2. $f'(a) = g'(a)$,
3. $f'(x) < g'(x)$ for $x > a$ and
4. $f'(x) > g'(x)$ for $x < a$,

imply that $f(x) \leq g(x)$ for all x.

Exercise 2.10.21. Give an example that shows why $f(x) \leq g(x)$ implies very little about the relationship between $f'(x)$ and $g'(x)$. In particular, $f(x) \leq g(x)$ does not imply that $f'(x) \leq g'(x)$

Exercise 2.10.22. Prove that $1 - \frac{x^2}{2} \leq \cos(x)$ for all x.

Exercise 2.10.23. Prove that $x + 1 \leq e^x$ for all x.

Exercise 2.10.24. Use Exercise 2.10.23 to show that $\lim_{n\to\infty}(1 + \frac{1}{n})^n \leq e$

Exercise 2.10.25. Continue Exercise 2.10.24, show that $e \leq \lim_{n\to\infty}(1 + \frac{1}{n})^n$. Hint: show that for small enough x, $e^x < 1 + x + x^2$.

Exercise 2.10.26. Suppose that $P = \{p_i\}_{i=1}^n$, $\sum_i^n p_i = 1$ and $p_i \geq 0$ for all i. Define the **entropy H(P)** of the probability distribution P to be $H(P) = \sum_i -p_i \ln(p_i)$. **Show that** $0 \leq H(P) \leq \ln(n)$. Hint: use Jensen's, then the fact that e^x is strictly monotonic, and a little bit of geometry.

If $|f(x) - f(y))| \leq K|x - y|$, $K < \infty$ for all x and y, then we say that f **is Lipschitz** with Lipschitz constant $\text{Lip}(f) = K$. Lipschitz functions are less regular than C^1 functions, but the have enough regularity that a great deal can be shown about them. In fact, it is reasonable to think of them as the most general set functions for which calculus still works. We will study them in more detail in section 2.15. Here are a few inequalities involving Lipschitz functions. We will write $f \in \text{Lip}(X, Y)$ to mean that f is Lipschitz with domain X and co-domain Y

Exercise 2.10.27. Suppose that f is Lipschitz with $Lip(f) = K$. Show that wherever f is differentiable, $-K \leq f'(x) \leq K$.

Exercise 2.10.28. Define the **diameter of a set** E**,** $\text{diam}(E)$**,** to be $\sup_{x,y\in E} |x - y|$. Show that if $f \in \text{Lip}(\mathbb{R}^n, \mathbb{R}^m)$ with $\text{Lip}(f) = K$, then $\text{diam}(f(E)) \leq K \,\text{diam}(E)$.

Exercise 2.10.29. We will say that **gauge of** $\mathcal{C}$ **is** δ ($\text{ga}(\mathcal{C}) = \delta$), if $\sup_{C\in\mathcal{C}} \text{diam}(C) = \delta$ and that $\mathcal{C}$ **covers** E**,** $E \subset \bigcup \mathcal{C}$**,** if $E \subset \bigcup_{C\in\mathcal{C}} C$. The length of a set E in $\mathbb{R}^n$, $\mathcal{H}^1(E)$, is defined to be the

$$\lim_{\delta\to 0}\left(\inf_{\{\mathcal{C}\ |\ \text{ga}(\mathcal{C})\leq\delta, E\subset\bigcup\mathcal{C}\}} \sum_{C\in\mathcal{C}} \text{diam}(C)\right).$$

Show that if $f : \mathbb{R}^1 \to \mathbb{R}^1$ is in $\text{Lip}(\mathbb{R}^1, \mathbb{R}^1)$ and $\text{Lip}(f) = K$, then

$$\mathcal{H}^1(\{(x, f(x)) \in \mathbb{R}^2 \mid x \in E\}) \leq \sqrt{1 + K^2}\mathcal{H}^1(E).$$

Exercise 2.10.30. Suppose that $\mathcal{F} \equiv \{f \mid f(a) = c,\ f(b) = d,\ f \in \text{Lip}(\mathbb{R}^1, \mathbb{R}^1),\ \text{Lip}(f) = K\}$ Show that for any two $f, g \in \mathcal{F}$, we have that $\frac{-K}{2}(b - a)^2 \leq \int_a^b f(x)\,dx - \int_a^b g(x)\,dx \leq \frac{K}{2}(b - a)^2$. Show that if you include the c and d in the bounds, you can usually get better (tighter) bounds.

2.11 Riemannian Integration

Integration is usually introduced either as the inverse to differentiation or as the area under curves. Very soon after that, the fundamental theorem of calculus is at least stated. The version of integration is almost always Riemann's version, even though there are a host of other versions. In this section, we review Riemannian integration and look at some of it's properties. While the simplicity, and practical applicability, of this version of integration has led to widespread use, there are theoretical reasons why other versions were introduced and are used in analysis.

We now assume that $f : [a, b] \to \mathbb{R}^1$ and is bounded. A **partition** P **of** $[a, b]$ is a finite, ordered set of points $P = \{x_i\}_{i=0}^n \subset [a, b]$ such that $a < x_0 < x_1 < \cdots < x_n = b$ and $x_i < x_{i+1}$ for all i. The **gauge of the partition** is the length of the longest of the n subintervals generated by the $n + 1$ points. If we have two partitions of P_1 and P_2, we define $P = P_1 \vee P_2$ to be the partition generated by $P_1 \cup P_2$. We will say that $P < Q$ if $Q \subset P$. Define The gauge of a partition P, $\mathrm{ga}(P)$, to be the length of the longest subinterval $x_{i+1} - x_i$, $i = 0, 1, ..., n-1$. Define $\overline{P}(f) = \sum_{i=0}^{n-1}(x_{i+1}-x_i)(\sup_{x\in(x_i,x_{i+1}]} f(x))$ and $\underline{P}(f) = \sum_{i=0}^{n-1}(x_{i+1}-x_i)(\inf_{x\in(x_i,x_{i+1}]} f(x))$. Define $\overline{P}_\delta(f) = \inf_{\{P \mid \mathrm{ga}(P)\geq\delta\}} \overline{P}(f)$ and $\underline{P}_\delta(f) = \sup_{\{P \mid \mathrm{ga}(P)\geq\delta\}} \underline{P}(f)$. Define $\overline{\int}_a^b f(x)\,dx = \lim_{\delta\to 0} \overline{P}_\delta(f)$ and $\underline{\int}_a^b f(x)\,dx = \lim_{\delta\to 0} \underline{P}_\delta(f)$. We say that f **is Riemann integrable** if $\overline{\int}_a^b f(x)\,dx = \underline{\int}_a^b f(x)\,dx$. In that case we define $\int_a^b f(x)\,dx \equiv \overline{\int}_a^b f(x)\,dx = \underline{\int}_a^b f(x)\,dx$.

Exercise 2.11.1. Prove that $P < Q$ implies that $\overline{P}(f) \leq \overline{Q}(f)$ and $\underline{P}(f) \geq \underline{Q}(f)$.

Exercise 2.11.2. Suppose that you have a sequence of partitions $\{P_i\}_{i=1}^\infty$ such that $\mathrm{ga}(P_i) \to 0$. Show that there exists another sequence of partitions Q_i such that $Q_i < P_i$, $Q_{i+1} < Q_i$, and $\mathrm{ga}(Q_i) \leq \mathrm{ga}(P_i)$ for all i.

Exercise 2.11.3. Suppose that $P_{i+1} < P_i$, $Q_{i+1} < Q_i$, $\lim_{i\to\infty} \mathrm{ga}(P_i) = 0$, and $\lim_{i\to\infty} \mathrm{ga}(Q_i) = 0$. Prove that $\lim_{i\to\infty} \overline{P}_i(f) = \overline{Q}_i(f)$ and $\lim_{i\to\infty} \overline{P}_i(f) = \overline{Q}_i(f)$.

Exercise 2.11.4. Use Exercise 2.11.3 to show that we could just have well defined $\overline{\int}_a^b f(x)\,dx = \lim_{i\to\infty} \overline{P}_i(f)$ and $\underline{\int}_a^b f(x)\,dx = \lim_{i\to\infty} \underline{P}_i(f)$ for any $P_{i+1} < P_i$ such that $\lim_{i\to\infty} \mathrm{ga}(P_i) = 0$.

Exercise 2.11.5. Show that if f is continuous, then f is Riemann integrable.

Exercise 2.11.6. Find an example of a function $f : [0, 1] \to \mathbb{R}^1$ that is not Riemann integrable.

Exercise 2.11.7. Show that a function that is continuous at all but a finite number of points is Riemann integrable.

Exercise 2.11.8. Suppose that f is monotonically increasing or decreasing on $[a, b]$. Prove that f is Riemann integrable.

Exercise 2.11.9. Prove that if f and g are Riemann integrable then so is $\alpha f + \beta g$ for any $\alpha, \beta \in \mathbb{R}^1$.

Exercise 2.11.10. Prove that $\int_0^x t\,dt = \frac{x^2}{2}$.

Exercise 2.11.11. Prove that $\int_0^x t^k\,dt = \frac{x^{k+1}}{k}$.

Exercise 2.11.12. Prove that if f is continuous, then $\frac{d}{dx}\int_0^x f(t)\,dt = f(x)$.

Exercise 2.11.13. Suppose that $f : [a, b] \to \mathbb{R}^1$ is continuous at all but a countable number of points d_i and that at those points f has a jump discontinuity with jump size j_i (can be positive or negative!). Assume further that $\sum |j_i| < \infty$. Show that f is Riemann integrable.

Exercise 2.11.14. Find a function $f : [0, 1] \to \mathbb{R}^1$ with an infinite number of jump discontinuities with size j_i such that $\sum |j_i|$ diverges, but is Riemann integrable nonetheless.

2.12 Lebesgue Integration

This is an optional section that I nevertheless recommend covering, if not before the GQE, immediately after, in order to give yourself an initial exposure to measure theory.

Lebesgue measures and integration that were developed around the turn of the 19th century have become the standard integration used in most situations in analysis. There are other forms of integration that are used as well, some of them quite important, but Lebesgue integration is the foundation for modern analysis. Lebesgue measures are special cases of the now standard Hausdorff measures pervasive in geometric analysis. (You have actually already met Hausdorff measures in Exercise 2.10.29.) We will briefly talk a bit more about Hausdorff measures at the end of this section.

2.12.1 Outer measures

A function $\mu : 2^X \to \mathbb{R}^+ \cup \{0\}$, where $\mathbb{R}^+$ denotes the positive real numbers, is called a measure (or more commonly an outer measure) if (1) $\mu(\emptyset) = 0$ and (2) $\mu(E) \leq \sum_i \mu(E_i)$ whenever $E \subset \cup_i E_i$. Such measures are easy to construct. In this section, we will construct two such measures. But unless we restrict which sets we consider to be "good" sets, we end up with paradoxes like the Banach-Tarski paradox that says that, for example, we can cut apart 2 unit balls of radius 1 into a finite number of pieces, rotate and translate them and put them back together into a single ball of radius 1, thus making the idea of volume useless.

The approach we take to eliminating this is to allow only sets which cut other sets in reasonable ways: we will say that F is measurable if and only if $\mu(E) = \mu(E \cap F) + \mu(E \cap F^c)$ is true for all sets E. Such sets make up the set of μ-measurable sets in X. If the set of μ-measurable sets contains all open sets, we say μ is a **Borel measure**.

A measure μ is called a **Regular measure** if for each $E \subset X$, there is a μ-measurable set F such that $E \subset F$ and $\mu(E) = \mu(F)$. A measure μ on $\mathbb{R}^n$ is a **Borel regular measure** if it is Borel and regular and for each $E \subset \mathbb{R}^n$, there is a Borel set F such that $E \subset F$ and $\mu(E) = \mu(F)$. A measure μ on $\mathbb{R}^n$ is a **Radon measure** if it is Borel regular and $\mu(K) < \infty$ for all compact $K \subset \mathbb{R}^n$.

For these sets, we have many useful facts/theorems that we can establish. I refer you to Evans and Gariepy's book for the proof of these facts.

1. Both $\emptyset$ and X are measurable.
2. if $\mu(E) = 0$ then E is measurable.
3. E is measurable if and only if $E^c = X \setminus E$ is measurable.
4. $E \subset F \subset X$ implies that $\mu(E) \leq \mu(F)$.
5. The set of measure sets is a σ-algebra: A collection of sets $\mathcal{S}$ is a σ-algebra if (1) it contains both $\emptyset$ and X (2) if $E \in \mathcal{S}$, then $X \setminus E \in \mathcal{S}$ (3) if each of $\{E_i\}_{i=1}^{\infty}$ are in $\mathcal{S}$, then $\bigcup_i E_i \in \mathcal{S}$, and (4) if each of $\{E_i\}_{i=1}^{\infty}$ are in $\mathcal{S}$, then $\bigcap_i E_i \in \mathcal{S}$ too.
6. if the sets $\{E_i\}_{i=1}^{\infty}$ are pairwise disjoint and measurable and $E = \bigcup_i E_i$, then $\mu(E) = \sum_i \mu(E_i)$.
7. If $E_i \subset E_{i+1}$, then $\lim_{i\to\infty} \mu(E_i) = \mu(\bigcup_i E_i)$.
8. If $E_{i+1} \subset E_i$, then $\lim_{i\to\infty} \mu(E_i) = \mu(\bigcap_i E_i)$.
9. If μ is a Radon measure on $\mathbb{R}^n$. Then for any $A \subset \mathbb{R}^n$, $\mu(A) = \inf_{O \text{ open, } A\subset O} \mu(O)$.
10. If μ is a Radon measure on $\mathbb{R}^n$. Then for any μ-measurable $A \subset \mathbb{R}^n$, $\mu(A) = \sup_{C \text{ compact, } C\subset A} \mu(C)$.

See Chapter 1 of Evans and Gariepy for much more on measures – I follow this chapter closely at various places and I recommend this book very highly.

2.12.2 Lebesgue measure

We will focus on Lebesgue measure which is an outer measure generated by covers with rectangles. A rectangle in $\mathbb{R}^n$ is any region $R = I_1 \times I_2 \times \cdots \times I_n$ where the I_k are either all closed or all open, bounded intervals i.e. $I_k = [a_k, b_k]$ and $-\infty < a_k < b_k < \infty$ for $k = 1, ..., n$ or $I_k = (a_k, b_k)$ and $-\infty < a_k < b_k < \infty$ for $k = 1, ..., n$. We call these closed or open rectangles, respectively.

The volume of a rectangle R, $v(R)$, is simply the product of the lengths of the sides of the rectangle $v(R) = \Pi_k(b_k - a_k)$. Now we define $\mathcal{L}^n(E)$, the n-dimensional Lebesgue measure of a set $E \in \mathbb{R}^n$, to be the $\mathcal{L}^n(E) \equiv \inf_{\mathcal{R}} \left(\sum_{R\in\mathcal{R}} v(R)\right)$ where $\mathcal{R}$ ranges over all possible countable collections of rectangles covering E.

We say that $f : \mathbb{R}^n \to \mathbb{R}^1$ is $\mathcal{L}^n$-measurable if $f^{-1}(U)$ is measurable when U is open. More generally, we say that $f : \mathbb{R}^n \to \mathbb{R}^m$ is $f^{-1}(E)$ is measurable whenever E is measurable.

Exercise 2.12.1. Prove that $\mathcal{L}^n$ is an outer measure.

Exercise 2.12.2. Prove that $\mathcal{L}^n(R) = v(R)$.

Exercise 2.12.3. Prove that rectangles are measurable.

Exercise 2.12.4. Use the facts about outer measures from Section 2.12.1 and Exercise 2.12.3 to prove that open sets and closed sets are measurable.

Exercise 2.12.5. Show that any continuous function $f : \mathbb{R}^n \to \mathbb{R}^1$ is $\mathcal{L}^n$-measurable.

Exercise 2.12.6. Suppose that each $f_i : \mathbb{R}^n \to \mathbb{R}^1 \in \{f_k\}_{k=1}^{\infty}$ is $\mathcal{L}^n$-measurable. Show that:

1. $f^k(x) = \sum_{i=1}^k f_i(x)$ is $\mathcal{L}^n$ measurable for every k.
2. $f(x) = \sup_i f_i(x)$ is $\mathcal{L}^n$-measurable.
3. $f(x) = \inf_i f_i(x)$ is $\mathcal{L}^n$-measurable.
4. $f(x) = \limsup_i f_i(x)$ is $\mathcal{L}^n$-measurable.
5. $f(x) = \liminf_i f_i(x)$ is $\mathcal{L}^n$-measurable.
6. $f(x) = \lim_i f_i(x)$ is $\mathcal{L}^n$-measurable whenever this limit exists for $\mathcal{L}^n$ a.e. every x.

Exercise 2.12.7. Show that $f(x) = \sup_{\alpha\in\mathcal{A}} f_\alpha(x)$ need not be $\mathcal{L}^n$-measurable even if each of the f_α if the index set $\mathcal{A}$ is not countable.

Exercise 2.12.8. Show that $\mathcal{L}^n$ is a Radon measure.

Exercise 2.12.9. (Approximation by step functions) Suppose that $f : \mathbb{R}^n \to [0,\infty]$. For each $x \in \mathbb{R}^n$, define $b(x) = 0.b_1^x b_2^x b_3^x \cdots b_n^x b_{n+1}^x \cdots$ where each b_i^x is either 0 or 1 and $\sum_{i=1}^m \frac{b_i^x}{i} \le f(x)$ for all m, with the additional constraint that if we flip any of the b_i^x's from 0 to 1, this inequality fails for some m. Define $E_i = \{x \text{ such that } b_i^x = 1\}$. Show that $f = \sum_{i=1}^{\infty} \chi_{E_i}$, where $\chi_E(x) = 1 if x \in E$ and $\chi_E(x) = 0 if x \in E^c$. Hint: (1) to show that each of the E_i are measurable show that $E_k = \{x \mid f(x) \ge \frac{1}{k} + \sum_{i=1}^{k-1} \chi_{E_i}\}$ and (2) show that if $f(x)$ is finite, then an infinite number of the b_i^x's are 0. (See E&G Theorem 1.12 on page 19.)

2.12.3 Lebesgue integration

Lebesgue integration is built on top of Lebesgue measure through the use of simple functions. A measurable function $h : \mathbb{R}^n \to \mathbb{R}^1$ is a **simple function** if h takes on only a countable number of values: i.e. $h(\mathbb{R}^n)$ is countable. Let $h(\mathbb{R}^n) = \{\alpha_i\}_{i=1}^{\infty}$ and define $E_i = h^{-1}(\alpha_i)$. We have that $h(x) = \sum_{i=1}^{\infty} \alpha_i \chi_{E_i}$.

We will sometimes write functions, simple or otherwise as $f = f^+ - f^-$, where $f^+ = \max(f,0)$and $f^- = -\min(f,0)$. For a positive simple function h, we define $\int h(x)\ dx = \sum_i \alpha_i \mathcal{L}^n(E_i)$. (We will use dx interchangeably with $d\mathcal{L}^n$.) If $\int h^+(x)\ dx < \infty$ or $\int h^-(x)\ dx < \infty$ we define $\int h(x)\ dx = \int h^+(x)\ dx - \int h^-(x)\ dx$. Define

$$\int_l f(x)\ dx = \sup_{\text{simple } h, h\le f} \left(\int h(x)\ dx \right)$$

and

$$\int^u f(x)\ dx = \inf_{\text{simple } h, h\ge f} \left(\int h(x)\ dx \right).$$

If $\int^u f(x)\ dx = \int_l f(x)\ dx$, then we say that **f is integrable**, even if the value is infinite. A function **f is summable** if f is integrable and $\int |f(x)|\ dx$ finite.

Exercise 2.12.10. Prove that every $\mathcal{L}^n$-measurable function is integrable. (Remember, this just means that $\int^u f(x)\ dx = \int_l f(x)\ dx$, not that the common value is finite).

Exercise 2.12.11. Suppose that we define $I(f) = \int f(x)\, dx$. Prove that: $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$ for all $\alpha, \beta \in \mathbb{R}^1$

Exercise 2.12.12. Let $f_\lambda(x) = f(x - \lambda)$. Prove that $I(f_\lambda) = I(f)$ for all $\lambda \in \mathbb{R}^1$.

Exercise 2.12.13. Suppose that $J : \mathcal{F} \to \mathbb{R}^1$ is a linear functional on measurable functions $\mathcal{F}$, that satisfies:

1. $J(\alpha f + \beta g) = \alpha J(f) + \beta J(g)$ for all $\alpha, \beta \in \mathbb{R}^1$
2. $J(f_\lambda) = J(f)$ for all $\lambda \in \mathbb{R}^1$.
3. If $0 \le f \le g$, $J(f) \le J(g)$.

Prove that $J(f) = cI(f)$ for some real constant c.

2.12.4 The three main theorems

We very often need to know the the limit of integrals of a sequence is the integral of the limit of the sequence. We have three standard theorems.

1. (Fatou's Lemma)
$$\int \liminf_{i\to\infty} f_i\, dx \le \liminf_{i\to\infty} \int f_i\, dx$$
2. (Monotone Convergence) Suppose that $\{f_i\}$ are all measurable and that $0 \le f_1 \le ... \le f_i \le f_{i+1} \le ....$ Then we have that
$$\lim_{i\to\infty} \int f_i\, dx = \int \lim_{i\to\infty} f_i\, dx.$$
3. (Dominated Convergence Theorem) If $f_i \to f$ μ a.e., $|f_i|, |f| < g$ and $\int g\, dx < \infty$, then
$$\int |f_i - f|\, dx \to 0 \text{ as } i \to \infty.$$

Exercise 2.12.14. Read the proofs of the three theorems I have included in Appendix C. (These notes are an excerpt from an incomplete draft of book I am writing.)

2.12.5 Lebesgue Measure is a special case of Hausdorff Measures

The family of outer measures we will most often use are Hausdorff Measures. Lebesgue measures are in fact special cases of these more general measures. (Because the definitions are different, this fact is a theorem that is not terribly easy to prove.) We will reiterate the Caratheodory construction for this case and define the Hausdorff measures in a series of steps:

1. A cover of a set E is a family of sets $\mathcal{F}$ such that
$$E \subset \bigcup_{F_i \in \mathcal{F}} F_i$$

2. We will say that a cover is a δ-cover is $\operatorname{diam}(F) < \delta$ for all $F \in \mathcal{F}$.
3. We will denote the family of all δ-covers of E by $\mathbb{F}_\delta(E)$ i.e.

$$\mathbb{F}_\delta(E) \equiv \{\text{All } \mathcal{F} \text{ such that } E \subset \bigcup_{F_i \in \mathcal{F}} F_i \text{ and } \operatorname{diam}(F_i) < \delta \ \ \forall F_i \in \mathcal{F}\}$$

4. Now we define the Hausdorff δ-measures:

$$\mathcal{H}^s_\delta(E) \equiv \inf_{\mathcal{F} \in \mathbb{F}_\delta(E)} \sum_{F_i \in \mathcal{F}} \alpha(s) \frac{(\operatorname{diam} F_i)^s}{2^s}$$

   where $\alpha(s)$ is the volume of the s-dimensional unit ball, where we have used the Γ-function to extend the definition in the case of nonintegral dimensions.

5. Finally, we can define the Hausdorff measures:

$$\mathcal{H}^s(E) = \lim_{\delta \downarrow 0} \mathcal{H}^s_\delta(E)$$

Note that we get a different Hausdorff measure for each $s \in [0, \infty)$. Figure 2.2 illustrates the idea of Hausdorff measure.

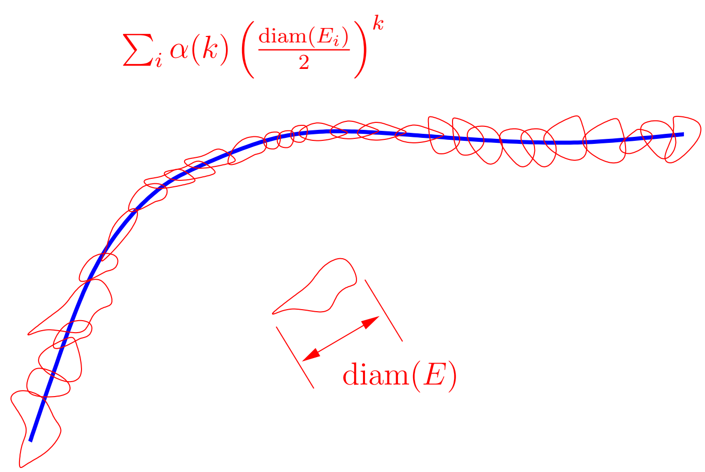

Figure 2.2: Visualizing the Hausdorff measure of a set: the actual computation for complicated sets is often difficult. Finding a sequence of covers that provably get the defined infimum with $\delta \to 0$ can be very challenging. And of course, the brute force computation is impossible.

Exercise 2.12.15. Show that if the s-dimensional Hausdorff measure of E is greater than 0, $\mathcal{H}^s(E) > 0$, then $\mathcal{H}^t(E) = \infty$ for all $t < s$. Show that $\mathcal{H}^s(E) = 0$ implies that $\mathcal{H}^t(E) = 0$ for all $t > s$.

Exercise 2.12.16. Show that $\mathcal{H}^s_\delta(E)$ is a monotonically decreasing function of δ.

Exercise 2.12.17. Show that any countable set of points (in any $\mathbb{R}^n$) has dimension equal to 0 by showing that the dimension is less than β for any $\beta > 0$. The main point here is that this includes countably infinite sets of points.

2.12.6 Some Approximation

We define $(E\Delta F)$, the symmetric difference between two sets E and F, to be $(E\Delta F) \equiv (E \cap F^c) \cup (F \cap E^c)$.

Exercise 2.12.18. Prove that if E is measurable and $\mathcal{L}^n(E) < \infty$, then for any $\epsilon > 0$, there is a finite union of closed rectangles, call this set F_ϵ, such that $\mathcal{L}^n(E\Delta F) < \epsilon$. Show that the same thing holds if you use finite unions of open rectangles.

Exercise 2.12.19. Use Exercise 2.12.18 to show that for any characteristic function χ_E, where $E \subset \mathbb{R}^n$ is measurable, we can find a positive continuous $f_E^\epsilon : \mathbb{R}^n \to \mathbb{R}^1$ such that $\int |\chi_E - f_E^\epsilon(x)|\, dx$.

Exercise 2.12.20. Prove that in Exercise 2.12.19, you can require $f_E^\epsilon \le \chi_E$.

Exercise 2.12.21. Use the previous exercises and the theorems in Section 2.12.4, to prove that for $f > 0$ such that $\int f\ dx < \infty$, there exists a sequence of continuous functions f_i such that $\int |f - f_i|\, dx \to 0$ as $i \to \infty$.

Exercise 2.12.22. Prove that there are bounded, $\mathcal{L}^n$-measurable sets $E \subset B(0,1)$, such that if we demand the finite union F in Exercise 2.12.18 **contains** E, then we cannot get a small difference between E and F. More concretely, show that there is a set E with $\mathcal{L}^n(E)$, such that every finite union of rectangles including E has measure exceeding $\mathcal{L}^n(B(0,1))$.

2.13 Mean value theorem $f : \mathbb{R}^1 \to \mathbb{R}^1$

The mean value theorem says that if $f : [a,b] \to \mathbb{R}$ is differentiable everywhere, then there is point $c \in [a,b]$ such that $f(b) = f(a) + f'(c)(b-a)$. We will call a line $y = L(x) = ax + b$ a **supporting line** of the graph of $f : [a,b] \to \mathbb{R}^1$ if the graph of L and the graph of f have at least one point strictly between a and b, in common, and the graph of f is everywhere above the graph of L or the graph of f is everywhere below the graph of L. Stated differently, L is a **supporting line** of the graph of f if (1) there is some point $c \in (a,b)$ such that $f(c) = L(c)$, and (2) either $f(x) \le L(x)$ for all $x \in [a,b]$ or $f(x) \ge L(x)$ for all $x \in [a,b]$.

Exercise 2.13.1. Prove that if $f(a) = f(b)$, f is continuous on $[a,b]$, and f is differentiable for all $x \in (a,b)$, there is a point $c \in (a,b)$ such that $f'(c) = 0$.

Exercise 2.13.2. Prove that if f is differentiable on (a,b) and continuous on $[a,b]$, there is a point $c \in (a,b)$ such that $f(b) = f(a) + f'(c)(b-a)$.

Exercise 2.13.3. Prove that prove that $f(a) = f(b)$, f is continuous on $[a,b]$ then there is a horizontal supporting $L(x) = 0x + b$ for the graph of f.

Exercise 2.13.4. Prove that if f is continuous on $[a,b]$, then there is a point $c \in (a,b)$ such that either (1) $f(x) \le \frac{f(b)-f(a)}{b-a}(x-c) + f(c)$ for all $x \in [a,b]$ or (2) $f(x) \ge \frac{f(b)-f(a)}{b-a}(x-c) + f(c)$ for all $x \in [a,b]$. I.e. there is a supporting line for the graph of f that intersects the graph at some point between a and b and have slope equal to $\frac{f(b)-f(a)}{b-a}$.

Exercise 2.13.5. Give an example of a function $f : [a, b] \to \mathbb{R}^1$ differentiable everywhere in (a, b), that does not satisfy the mean value theorem.

Exercise 2.13.6. Consider the function $f(x) = |x|$ on the interval $[-1, 1]$. What are the supporting lines of the graph of f on that interval? If $\mathcal{L}$ is the family of all supporting lines, find the function $g(x) = sup_{L\in\mathcal{L}} L(x)$.

Exercise 2.13.7. Suppose that $f(x)$ is continuously differentiable, and we know that $|f'(y) - f'(x)| \le \epsilon$ when $|x - y| \le \delta(\epsilon)$. Find bounds $|f(y) - f(x) - f'(x)(y - x)|$ in terms of ϵ and $\delta(\epsilon)$ using the mean value theorem.

2.14 Mean value theorem in higher dimensions

In higher dimensions, we have the **generalized mean value theorem**, a simpler result that is nevertheless useful: suppose $\gamma : [a, b] \to \mathbb{R}^n$ is continuously differentiable everywhere. Then for some $c \in [a, b]$, we have

$$|\gamma(b) - \gamma(a)| \le |\dot{\gamma}(c)|\, |b - a|.$$

A bit more generally, suppose that E is a 1-dimensional set in $\mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz. Then for some $c \in E$,

$$\mathcal{H}^1(F(E)) \le |DF(c)|\, \mathcal{H}^1(E).$$

The same statements hold if $\gamma : [a, b] \to B_1$ and $F : B_1 \to B_2$. where B_1 and B_2 are finite or infinite dimensional Banach spaces. In the next section we will see the area and co-area formulas which have the generalized mean value theorems as simple corollaries

Exercise 2.14.1. Prove the first version of the generalized mean value theorem: $|\gamma(b) - \gamma(a)| \le |\dot{\gamma}(c)|\, |b - a|$ for differentiable γ.

Exercise 2.14.2. Assume the fact that when γ is Lipschitz $\int_a^b \dot{\gamma}(t)\, dt = \gamma(b) - \gamma(a)$, to get the generalized mean value theorem in the case that γ is merely Lipschitz.

Exercise 2.14.3. Can you find an example $\gamma : [-1, 1] \to \mathbb{R}^1$ that is differentiable everywhere on $[-1, 1]$, and yet the graph of γ on $[-1, 1]$ is infinitely long? In other words, defining the map $g : t \to (t, \gamma(t))$, we get that $\mathcal{H}^1(g([-1, 1])) = \infty$. Hint: prove that such a γ cannot have a continuous derivative.

Exercise 2.14.4. Find an example of a function $\gamma : [a, b] \subset\to \mathbb{R}^3$ such that there is no point $c \in [a, b]$ where $\gamma(b) - \gamma(a) = \dot{\gamma}(c)(b - a)$. Hint: find a curve whose tangents are never parallel to $\gamma(b) - \gamma(a)$.

Exercise 2.14.5. Find an example of a function $\gamma : [a, b] \subset \mathbb{R}^1 \to \mathbb{R}^2$ such that there is no point $c \in [a, b]$ where $\gamma(b) - \gamma(a) = \dot{\gamma}(c)(b - a)$. Hint: the hint on the previous problem **will not** work here.

Exercise 2.14.6. (*) One can view the various mean value theorems as theorems that differ because averages over connected point sets in $\mathbb{R}^1$ versus $\mathbb{R}^n$, $n > 1$ are fundamentally the different. **Prove that** a probability distribution on a connected set in $\mathbb{R}^1$ must has its mean value in that connected set, but that there are examples in $\mathbb{R}^n$, $n > 1$ where the mean value is not in the connected set.

Exercise 2.14.7. (*) Use Exercise 2.14.6 to explain the the usual mean value theorem versus Exercise 2.14.5.

2.15 Lipschitz functions

Suppose that $E \subset \mathbb{R}^n$. We define **the distance function from** x **to** E**,** $d(x, E)$**,** by $d(x, E) = inf_{y \in E}|x - y|$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **Lipschitz with Lipschitz constant K** if $|f(x) - f(y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}^n$ and some fixed $K < \infty$. **Define** $\text{Lip}(f)$ to be the smallest constant K for which $|f(x) - f(y)| \leq K|x - y|$ for all $x, y \in \mathbb{R}^n$ is true.

It turns out that Lipschitz functions are the right class of functions to consider in geometric analysis – many things that are true about smooth functions are true about Lipschitz functions, but not functions with even less regularity. **Rademacher's Theorem** tells us that Lipschitz functions are differentiable almost everywhere. A **set is Rectifiable** if it is contained in at most countably many images of Lipschitz maps. More precisely, $E \subset \mathbb{R}^n$ is said to be **k-rectifiable** if $E \subset \left\{\bigcup_i f_i(\mathbb{R}^k)\right\} \bigcup E_0$ where $k \leq n$, $\text{Lip}(f_i) = K_i < \infty$ for all i, and $\mathcal{H}^k(E_0) = 0$. These sets are in some sense the most general, least regular sets on which one can do calculus.

Two very important theorems that work for rectifiable sets and Lipschitz mappings between those sets are the **area formula** and the **coarea formula**. While the proofs of these formula are quite involved (see Evans and Gariepy's book [5] for the proofs), we can state the formulas after we define the Jacobian of a matrix. Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ and f is differentiable at x. Then the Jacobian of f at x is $Jf \equiv \sqrt{\det(Df^* \cdot Df)}$ if $m \geq n$ or $Jf \equiv \sqrt{\det(Df \cdot Df^*)}$ if $m \leq n$, where A^* is the transpose of A.

If $n \leq m$ and $f : \mathbb{R}^n \to \mathbb{R}^m$, we have the **area formula:**

$$\int_E g(x)Jf(x)\, dx = \int_{f(E)} \left(\sum_{x \in f^{-1}(y)} g(x) \right) d\mathcal{H}^n y.$$

If $m \leq n$ and $f : \mathbb{R}^n \to \mathbb{R}^m$, we have the **coarea formula:**

$$\int_E g(x)Jf(x)\, dx = \int_{f(E)} \left(\int_{x \in E \cap f^{-1}(y)} g(x)\, d\mathcal{H}^{n-m} x \right)\, d\mathcal{H}^m y.$$

In the important case in which $g(x) = 1$, these reduce to:

$$\int_E Jf(x)\, dx = \int_{f(E)} \mathcal{H}^0(f^{-1}(y)) d\mathcal{H}^n y.$$

and

$$\int_E Jf(x)\, dx = \int_{f(E)} \mathcal{H}^{n-m}(E \cap f^{-1}(y))\, d\mathcal{H}^m y.$$

Exercise 2.15.1. (*) Suppose that E is closed. Prove that $|\nabla d(x,E)| = 1$ for all x such that $d(x,E) > 0$ and $d(x,E)$ is differentiable. Hint: study the intersection of the closed ball of radius $d(x,E)$ and E.

Exercise 2.15.2. Let $(W)^o$ be the interior of the set W. Use Exercise 2.15.1 and Rademacher's Theorem to show that if $F \subset (E^c)^o \subset \mathbb{R}^n$ then $\mathcal{H}^n(F) = \int_0^\infty \mathcal{H}^{n-1}(d_E^{-1}(r))\, dr$

Exercise 2.15.3. Use the area formula to prove that if $E \subset \mathbb{R}^1$, $\text{Lip}(f) = K$ where $f : \mathbb{R}^1 \to \mathbb{R}^1$, and the graph of f over E is $G \equiv \{(x,y) \in \mathbb{R}^2 | x \in E,\ y = f(x)\}$, then $\mathcal{H}^1(E) \leq \mathcal{H}^1(G) \leq (\sqrt{1+K^2})\mathcal{H}^1(E)$

Exercise 2.15.4. Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$, $\text{Lip}(f) = K$ and the graph of f over E is $G \equiv \{(x,y) \in \mathbb{R}^2 | x \in E,\ y = f(x)\}$. Define $C^K_{(x_0,y_0)}$ to be the set $\{(x,y) \in \mathbb{R}^2 || \frac{y-y_0}{x-x_0}| \leq K\}$.

Exercise 2.15.5. Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$, $\text{Lip}(f) = K$ and the graph of f over E is $G \equiv \{(x,y) \in \mathbb{R}^2 | x \in E,\ y = f(x)\}$. Prove that G is a rectifiable set.

Exercise 2.15.6. (*) Prove that if $G \subset \mathbb{R}^n$ is k-rectifiable, $k < n$, then $G \subset \bigcup_i G_i \bigcup G_0$, where the G_i are graphs of Lipschitz $f_i : \mathbb{R}^k \to \mathbb{R}^{n-k}$.

Exercise 2.15.7. Suppose that $E \subset \mathbb{R}^1$, $f : E \to R^1$ and for all $x, y \in E$, $f(x) - f(y) \leq K|x-y|$. for any $z \in E$, Let $p_z(x) \equiv f(z) + |x-z|$. Define $g : \mathbb{R}^1 \to \mathbb{R}^1$ by $g(x) = \inf_{z\in E}\ p_z(x)$. Prove that G is Lipschitz, $\text{Lip}(G) = \text{Lip}(f)$, and $g(x) = f(x)$ when $x \in E$.

Exercise 2.15.8. Draw pictures illustrating the proof in Exercise 2.15.7.

Exercise 2.15.9. (*) Suppose that $f : \mathbb{R}^1 \to \mathbb{R}^1$ is continuous and differentiable on $\mathbb{R}^1 \setminus S$, with S being a countable set, on which the function can fail to be continuous or differentiable. Prove that graph of f, $G \equiv \{(x,y) \in \mathbb{R}^2 | x \in E,\ y = f(x)\}$, is rectifiable. While this seems at first to be wrong, because f can be non-Lipschitz, we are **not** saying that there is one mapping $f_1 : \mathbb{R}^1 \to \mathbb{R}^2$ such that $G \subset f_1(\mathbb{R}^1)$.

Exercise 2.15.10. A function is called **bi-Lipschitz** if $k|x-y| \leq |f(x) - f(y)| \leq K|x-y|$ with $0 < k \leq K < \infty$. Note that this means f is invertible. (Why?) Prove that this means that either (1) the graph of G is contained in every cone $C^{+,k,K}_{(z,f(z))} \equiv \{(x,y) \in \mathbb{R}^2 | k \leq \frac{y-f(z)}{x-z} \leq K\}$ or (2) the graph of G is contained in every cone $C^{-,k,K}_{(z,f(z))} \equiv \{(x,y) \in \mathbb{R}^2 | -K \leq \frac{y-f(z)}{x-z} \leq -k\}$

Exercise 2.15.11. Prove that if f if Lipschitz and $\text{Lip}(f) = K$, then $g(x) \equiv f(x) + (K+1)x$ is bi-Lipschitz. Can you draw a picture (or 2 or 3) that proves this?

Exercise 2.15.12. Prove that if f is Lipschitz with $\text{Lip}(f) = K$, then $f^m(x) = \underset{1}{f} \circ \underset{2}{f} \circ \cdots \circ \underset{m}{f}(x)$ is Lipschitz with $\text{Lip}(f^k(x)) \leq K^m$.

Exercise 2.15.13. Find an example of a function $f : \mathbb{R}^1 \to \mathbb{R}^1$, with $\text{Lip}(f) = K$, such that $\text{Lip}(f^2) < K^2$.

Exercise 2.15.14. (*) Suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\text{Lip}(f) = K$. Prove that $Jf \leq K^n$. Hint: think about the singular value decomposition (SVD) of the matrix Df.

Exercise 2.15.15. Find an example of $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\text{Lip}(f) = K$ such that $Jf = K^n$.

Exercise 2.15.16. (*) We will say that $f : \mathbb{R}^n \to \mathbb{R}^1$ is **locally convex** at x if there is a neighborhood of x, U_x, such that $f|_{U_x}$ is convex (I.e f restricted to a domain of U_x is convex.) The definition of **local concave** is completely analogous. **Create a Lipschitz function** $f : \mathbb{R}^1 \to \mathbb{R}^1$ which is not locally convex or concave at any point in $\mathbb{R}^1$. Hint: start with the sawtooth function and work with scaled and dilated versions of that function. (**Sawtooth function**: $f(x) = x$ from $x = 0$ to $x = 1$ and $f(x) = -x + 2$ from $x = 1$ to $x = 2$, now repeat periodically to define f on $\mathbb{R} \setminus [0, 2]$)

Exercise 2.15.17. Prove that a convex function $f : \mathbb{R}^1 \to \mathbb{R}^1$ is Lipschitz on any compact set. Hint: (1) for all x, $f(x) < \infty$ and (2) f convex implies $f'(x)$ is nondecreasing.

Exercise 2.15.18. Prove that if $f \in C^1(\mathbb{R}, \mathbb{R})$, then f is Lipschitz on any compact subset of $\mathbb{R}$. (Recall that $C^k(\mathbb{R}^n, \mathbb{R}^m)$ is the set of all functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ which whose first k derivatives exist and are continuous on $\mathbb{R}^n$. When $k = 0$, this simply means that the function is continuous. Often we write $C(\mathbb{R}^n, \mathbb{R}^m)$ instead of $C^0(\mathbb{R}^n, \mathbb{R}^m)$.

Exercise 2.15.19. Suppose that $f \in C(\mathbb{R}, \mathbb{R})$. Define $F(x) = \int_0^x f(t)\, dt$. Prove that F is Lipschitz on any compact subset of $\mathbb{R}$.

Exercise 2.15.20. Study Exercise 2.7.23 to see that the function constructed in the Chapter 6, in my solution to Exercise 2.7.23, is in fact Lipschitz, so that even though Lipschitz functions are differentiable almost everywhere, the derivative can be discontinuous on all but a set of measure ϵ.

2.16 When does the fundamental theorem of calculus apply?

In calculus we first learn that $F(x) = \int_a^x f(t)\, dt$ if and only if $F'(x) = f(x)$. Of course there are conditions f or F must satisfy. That is the point of this section: how generally does this theorem hold? When is $F(x) = \int_a^x F(t)\, dt + F(a)$ a true statement? How wild can F be?

It is not difficult to show that if $f : \mathbb{R}^1 \to \mathbb{R}^1$ is continuous, then $F'(x) = f(x)$ for all x. In fact, that is one of the exercises below. What about the case in which f has a finite number of jump discontinuities? Does $F'(x) = f(x)$ all most everywhere, or to reverse the question, suppose that G is differentiable almost everywhere. In which cases is it true that $G(x) = \int_a^x G'(t)\, dt + G(a)$ for all x?

$f(x)$ is said to be **absolutely continuous** if, for any $\epsilon > 0$ there is δ such that if $\sum_{i=1}^{n}(b_i - a_i) \leq \delta$, then $\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \epsilon$ where the intervals (a_i, b_i) are disjoint. It is a theorem in analysis that if f is absolutely continuous, then $f(x) = f(a) + \int_a^x f'(t)\, dt$ for all x. See for example Folland's text [7].

In the vector calculus section, we mentioned that the vector calculus theorems were all special cases of the general Stokes theorem involving differential forms, which we do not cover here. We also mentioned in one of the problems that the vector calculus theorems were generalizations of the fundamental theorem of calculus. But the problem with **higher dimensional analogs of the fundamental theorem of calculus** is that the boundary of a region we integrate the derivative over in higher dimensions does not correspond to a point in the space, but rather a codimension 1 subset. So that integral cannot correspond to a function evaluation, but rather the integral of the function over a codimension 1 set. See Exercise 2.16.3 to explore this a little bit more.

Exercise 2.16.1. Prove that if f is continuous, then $F(x) = \int_a^x f(t)\, dt$ if and only if $F'(x) = f(x)$.

Exercise 2.16.2. Prove that if $f : \mathbb{R}^1 \to \mathbb{R}^1$ is Lebesgue integrable, then $\int f(t)dt$ is absolutely continuous. Hint: define $g_n = \min\{|f|, n\}$. Consider $\int_{\mathbb{R}} g_n(t) + (|f|(t) - g_n(t))\, dt$. Define N_ϵ so that $n > N_\epsilon$ implies $\int_{\mathbb{R}} |f|(t) - g_n(t))\, dt \leq \frac{\epsilon}{2}$.

Exercise 2.16.3. In order to recover the value of a function through integrals over a sufficiently rich set of codimension 1 sets, we need the help of an area of study called **integral geometry**. The **Radon Transform** is the most famous tool in that area. **Look up an exposition** of this transform and **read the proof** that you can reconstruct a function from integrals of the function. I recommend either Bracewell [1] or Kak and Slaney [8].

2.17 Fubini's Theorem

While it is not that simple to prove, the equivalence of (1) iterated integrals over orthogonal subspaces and (2) the integral over higher-dimensional product space is used as soon as students begin computing higher dimensional integrals. One soon finds out that switching the order of integration can sometimes be a bit tricky due to limits integration depending in non-trivial ways on the shape of the region being integrated over. The theorem that allows us to switch order is **Fubini's Theorem**.

Informally, it says that

$$\int_{\vec{x}\in\mathbb{R}^3} f(x_1, x_2, x_3) d\vec{x} = \int_{x_3\in\mathbb{R}} \left(\int_{x_2\in\mathbb{R}} \left(\int_{x_1\in\mathbb{R}} f(x_1, x_2, x_3) dx_1 \right) dx_2 \right) dx_3$$

with the real challenge coming when we want to do this:

$$\int_{\vec{x}\in\Omega\subset\mathbb{R}^3} f(x_1, x_2, x_3) d\vec{x} = \int_{x_3\in P_{x_2}(P_{x_1}(\Omega))} \left(\int_{x_2\in P_{x_1}(\Omega)\cap l(x_3)} \left(\int_{x_1\in\Omega\cap l(x_2,x_3)} f(x_1, x_2, x_3) dx_1 \right) dx_2 \right) dx_3$$

where $p_y(E)$ is the projection of the set E onto the space orthogonal to the direction of y, and $x \in l(y^*, z^*)$ is the set $\{(x, y, z)|y = y^* \text{ and } z = z^*\}$.

The more precise statement is concerned with the fact that, for example, just because $f : \mathbb{R}^2 \to \mathbb{R}^1$ is $\mathcal{L}^2$ measurable does not imply that, for example, the functions $x \to f(x, 3)$ or $x \to f(x, 17)$ or

$y \to f(6.3, y)$ or $y \to f(5, y)$ are $\mathcal{L}^1$ measurable. Thus, we have to know that these slice functions are measurable almost all the time in order for the theorem to make sense. For the details see Chapter 1 of Evans and Gariepy's book [5]. In the following exercises we explore both Fubini and volume computation problems in general.

Exercise 2.17.1. Define Ω to be the intersection of the positive octant in $\mathbb{R}^3$ and the $\{x | x \cdot (1,1,1) \leq 1\}$. Write $\int_\Omega 1 d\vec{x}$ as an iterated integral and compute it.

Exercise 2.17.2. Look up the precise statement of the theorem in Evans and Gariepy's book [5].

Exercise 2.17.3. How is Fubini's theorem a special case of the coarea formula?

Exercise 2.17.4. Let E be the region in $\mathbb{R}^2$ bounded by the x_2 axis, the line $x_2 = 1$, and the line $x_2 = x_1$. Evaluate $\int_E x_1^2 d\vec{x}$ (1) by integrating first over x_1 and then over x_2 and (2) by integrating first over x_2 and then over x_1.

Exercise 2.17.5. Let E be the region in $\mathbb{R}^2$ bounded by

1. $\{(x_1, x_2) \mid 0 \leq x_1 \leq .35,\ x_2 = 1\}$,
2. $\{(x_1, x_2) \mid \pi/2 \leq x_1 \leq \pi/2 + .35,\ x_2 = 0\}$,
3. $\{(x_1, x_2) \mid 0 \leq x_1 \leq \pi/2,\ x_2 = \cos(x_1)\}$, and
4. $\{(x_1, x_2) \mid .35 \leq x_1 \leq \pi/2 + .35,\ x_2 = \cos(x_1 - .35)\}$.

Evaluate $\int_E 1 d\vec{x}$.

Exercise 2.17.6. Let E be the region in $\mathbb{R}^2$ bounded by

1. $\{(x_1, x_2) \mid 0 \leq x_1 \leq .35,\ x_2 = 1\}$,
2. $\{(x_1, x_2) \mid \pi/2 \leq x_1 \leq \pi/2 + .35,\ x_2 = 0\}$,
3. $\{(x_1, x_2) \mid 0 \leq x_1 \leq \pi/2,\ x_2 = \cos(x_1)\}$, and
4. $\{(x_1, x_2) \mid .35 \leq x_1 \leq \pi/2 + .35,\ x_2 = \cos(x_1 - .35)\}$.

Evaluate $\int_E (x_1 - \arccos(x_2))^2 d\vec{x}$.

Exercise 2.17.7. Define the **symmetric difference between two sets** E and F to be the points that are in one set but not then other: $E \Delta F \equiv (E \cap F^c) \cup (F \cap E^c)$. Let

1. $E_1 = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1,\ |x_1| \leq .2\}$
2. $E_2 = \{(x_1, x_2) \mid x_1^2 + (x_2 - 0.1)^2 \leq 1,\ |x_1| \leq .2\}$

Evaluate $\int_{E_1 \Delta E_2} 1 d\vec{x}$.

Exercise 2.17.8. Prove, using the area formula, that the area of the graph of the function $x_{n+1} = f(x_1, ..., x_n)$ over the set $E \subset \mathbb{R}^n$ equals $\int_E \sqrt{(1 + \nabla f \cdot \nabla f)}\ d\vec{x}$. Hint: use the fact that at any point in the domain, we can choose coordinates $\tilde{x}_i$, $i = 1, ..., n$, which are rotations of the initial coordinates, such that $\nabla f = (\frac{\partial f}{\partial \tilde{x}_1}, ..., \frac{\partial f}{\partial \tilde{x}_n}) = (0, ..., 0, \frac{\partial f}{\partial \tilde{x}_n})$. This makes the computation of $\sqrt{\det(DF^t \circ DF)}$ much easier. Here we are using F to denote the mapping $F : (x_1, ..., x_n) \to (x_1, ..., x_n, f(x_1, ..., x_n))$.

Exercise 2.17.9. Define a pyramid P to be any set in $\mathbb{R}^{n+1}$ such that, possibly after a rotation of the coordinate system and a translation of the set, there is a point $v = (0, ..., 0, x_{n+1})$, $x_{n+1} = h > 0$ and a set $E \subset \{x_1, ..., x_{n+1} \mid x_{n+1} = 0\}$ so that

$$P = \{x_1, ..., x_{n+1} \mid (x_1, ..., x_{n+1}) = \lambda e + (1 - \lambda)v \text{ for some } e \in E \text{ and some } 0 \leq \lambda \leq 1\}.$$

Now:

1. Prove that n-dimensional volume of $\{x_{n+1} = r\} \cap P$ equals $(\frac{h-r}{h})^n$ times the n-dimensional volume of E.

2. Use this to show that that the n+1-volume of the pyramid is $\frac{1}{n+1} h \mathcal{L}^n(E)$.

Hint: after proving the first fact about n-dimensional volume of slices, use the coarea formula and the mapping $g : \mathbb{R}^{n+1} \to \mathbb{R}^1$, given by $g(x) = x_{n+1}$.

Exercise 2.17.10. Use the method used in Exercise 2.17.9 to get a very similar result for spherical pyramids – intersections of $n + 1$-dimensional balls of radius h centered at the origin, with cones also centered at the origin.

Exercise 2.17.11. Assume that $f \in C^2(\mathbb{R}^1, \mathbb{R}^1)$. Use a Riemann integral approach, with the help of a Talyor series argument, to show that the length of the graph of f from $x = a$ to $x = b$ is $\int_a^b \sqrt{(1 + f'(x)^2)}\ dx$.

Exercise 2.17.12. Define E to be the region bounded by (1) the plane passing through $(1, 0, 0)$, $(0, 3, 0)$, and $(0, 0, 10)$, (2) the $x_1 = 0$ plane, (3) the $x_2 = 0$ plane, and (4) the $x_3 = 0$ plane. Compute $\int_E x_3\ d\vec{x}$ in several different ways using the iterated integrals Fubini enables you to use. Which order if iteration is easiest? Hint: think about Exercise 2.17.9.

2.18 Miscellaneous calculus facts

When does $\frac{d}{dx} \sum_{i=1}^{\infty} f_i(x) = \sum_{i=1}^{\infty} \frac{df_i(x)}{dx}$? In other words, **when can commute summation and differentiation?** This is a simple question that comes up often enough when we begin working with power series. A related question asks **when differentiation commutes with integration**: When is it true that $\frac{d}{dt} \int f(x, t)\ dx = \int \frac{\partial}{\partial t} f(x, t)\ dx$?

When we want to find the limit $\lim_{x \to x_0} \frac{g(x)}{f(x)}$ and either $\lim_{x \to x_0} g(x) = \lim_{x \to x_0} f(x) = 0$, we can use **L'Hospital's rule** and differentiate top and bottom and again attempt to take the limit. We can keep repeating this as often as we like.

There is a technique that also arises fairly frequently, that lets us solve constrained optimization problems. When we are trying to minimize $f(x)$ subject to the constraint that $g(x) = c$, then we can use **Lagrange multipliers** to accomplish that. While the full theory is quite involved, in the case in which f and g are smooth, we look for x such that $\nabla f(x) = \lambda \nabla g(x)$ for some real λ. Geometrically, this corresponds to finding the x's where the level sets of f and g meet tangentially.

There are also some odds and ends here concerning the convergence or divergence of series and integrals, but these are mostly prompts to open up R. P. Burn's book and check out the parts on convergence and divergence.

The following exercises explore these questions.

Exercise 2.18.1. Assume that f and $\frac{\partial}{\partial t} f(x,t)$ exist for all x and t. and Prove that $\frac{d}{dt} \int f(x,t)\, dx = \int \frac{\partial}{\partial t} f(x,t)\, dx$. Hint:

1. Assume that there is a $g(x) > 0$ such that $\int g\, dx < \infty$, $|\frac{\partial}{\partial t} f(x,t)| \le g(x)$, and $h_i \downarrow 0$.

2. Use the dominated convergence theorem and the mean value theorem to prove that
$$\frac{1}{h_i}\left(\int f(x,t+h_i)\, dx - \int f(x,t)\, dx\right) = \int \frac{f(x,t+h_i) - f(x,t)}{h_i}\, dx$$
converges to
$$\int \frac{\partial}{\partial t} f(x,t)\, dx$$
as $i \to \infty$.

3. Do the same for the case in which $h_i \uparrow 0$

Exercise 2.18.2. Use Exercise 2.18.1 to show that we can switch summation and differentiation: $\frac{d}{dt} \sum_{i=1}^{\infty} f_i(t) = \sum_{i=1}^{\infty} \frac{df_i(t)}{dt}$ Hint: Note that summation over i can be turned into integration over x by defining $F(x,t) = \sum_i \chi_{[i-1,1)}(x) f_i(t)$ where $\chi_E(x) = 1$ if $x \in E$ and 0 otherwise.

Exercise 2.18.3. Construct an example of an $f(x,t)$ such that $\frac{d}{dt} \int f(x,t)\, dx \neq \int \frac{\partial}{\partial t} f(x,t)\, dx$.

Exercise 2.18.4. Construct an example in which $\frac{d}{dt} \sum_{i=1}^{\infty} f_i(t) \neq \sum_{i=1}^{\infty} \frac{df_i(t)}{dt}$. Hint:

1. Define $f_i(t) = 0$ when $t \in [0, 1 - \frac{1}{2^{i-1}}]$ and 1 when $t \in [1 - \frac{1}{2^i}, 1 - \frac{1}{2^{i+2}}]$, to smoothly go from 0 to 1 when x goes from $1 - \frac{1}{2^{i-1}}$ to $1 - \frac{1}{2^i}$ and to smoothly go from 1 to 0 when x goes from $1 - \frac{1}{2^{i+2}}$ to $1 - \frac{1}{2^{i+3}}$. Following the idea suggested for solving Exercise 2.18.2, define $F(x,t) = \sum_i \chi_{[i-1,1)}(x) f_i(t)$ where $\chi_E(x) = 1$ if $x \in E$ and 0 otherwise.

2. Prove that $\sum_i f_i(1) = 0$ and $\sum_i \frac{df_i}{dt}(1) = 0$.

3. Prove that $\sum_i f_i(t) \ge 1$ for $t \in [\frac{1}{2}, 0)$. Use this to prove that $\frac{d}{dt} \sum_{i=1}^{\infty} f_i(t)$ does not exist.

4. Conclude that $\frac{d}{dt} \sum_{i=1}^{\infty} f_i(t) \neq \sum_{i=1}^{\infty} \frac{df_i(t)}{dt}$

Exercise 2.18.5. Work through chapters 5 and 12 of R.P. Burn's book.

Exercise 2.18.6. Assume that $f : \mathbb{R}^n \to \mathbb{R}^1$ and $g : \mathbb{R}^n \to \mathbb{R}^1$. Prove that if ∇f is not parallel with ∇g at some point $\hat{x} \in L_g(c) \equiv \{x \mid g(x) = c\}$, then there are directions in which you can move from $\hat{x}$, staying in $L_g(c)$, such that $f(x)$ increases and directions in which $f(x)$ deceases.

Exercise 2.18.7. Use Theorem 2.7.1 to explain why L'Hospital's method for finding the limit $\lim_{x\to x_0} \frac{f(x)}{g(x)}$ works when $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = 0$. You can assume that both f and g have k derivatives where at least one of f and g have a non-zero kth derivative at x_0.

A few more, for good measure, to round out the analysis chapter.

Exercise 2.18.8. Construct a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ which is universal in the following sense: for any continuous $g : [0, 1] \to \mathbb{R}$ and any $\epsilon > 0$, there is a triple $(\alpha(\epsilon, g), \beta(\epsilon, g), x_{\epsilon,g}) \in \mathbb{R}^3$ such that

$$\sup_{x\in[0,1]} (\beta(\epsilon, g)f(\alpha(\epsilon, g)(x + x_{\epsilon,g})) - g(x)) \leq \epsilon$$

. In other words a blowup of the function f somewhere in its domain is arbitrarily close to g on $[0, 1]$.

Exercise 2.18.9. Construct a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ which is universal in the following sense: for continuous any $g : [0, 1] \to \mathbb{R}$, any $\epsilon > 0$ and any x_0, there is a quadruple $(\alpha(\epsilon, g), \beta(\epsilon, g), \gamma(\epsilon, g), x_{\epsilon,g}) \in \mathbb{R}^4$ such that

$$\sup_{x\in[0,1]} (\beta(\epsilon, g)(f(\alpha(\epsilon, g)(x + x_{\epsilon,g} + x_0)) - \gamma(\epsilon, g)) - g(x)) \leq \epsilon$$

and $x_{\epsilon,g} < \epsilon$. In other words a blowup of the function arbitrarily near x_0 is arbitrarily close to g on $[0, 1]$.

2.19 Philosophical Comments

Beginning analysis is often taught as though it were just calculus plus ϵ's and δ's, with very few real surprises. This is unfortunate and unnecessary. Because these notes are intended as a review for the graduate qualifying exam at WSU, I have not emphasized the rich diversity that you can encounter if you move off the well beaten paths. But the study of sets, functions and measures is a huge, infinitely rich subject full of diverse surprise and wonder. Deep discipline, combined with intuitive, even wild, creative expeditions, opens to us an enormous wilderness filled with unexpected discoveries and beauty.

2.19.1 First Course in Analysis: Unabridged

Here is a syllabus for what I consider to be an ideal, if intense, first, year long course in analysis. Such a course should be preceded by a rigorous one semester course in proofs, focusing on topics from metric spaces and on inequalities.

Metric Spaces and inequalities metric spaces; lots of examples; metric spaces that are also vector spaces: normed linear spaces; topology-continuity-compactness-connectedness-convergence-etc; Inequalities :Cauchy-Schwarz, Holder, Jensen's, am-gm, etc;

Essentials of integration I Outer measures; measurable functions; two theorems and a Lemma; Fubini

Differentiation I Derivatives as linear approximations; derivatives in infinite dimensions; Differentiable vs C^1 vs $C^{1,1}$; Taylor's Series and the slight improvement on the remainders following Smith; Mean Value Theorems

Integration (and Differentiation) II weak derivatives; densities; Lebesgue differentiation; covering theorems; Hausdorff Measures; Properties of Hausdorff measures;

Properties of differentiable functions Inverse function theorem; implicit function theorem; Manifolds; Constant Rank theorem; Sard's Theorem; Area and Coarea formulas I; Almgren and Lieb's "counterexample";

Convex Sets and Functions Properties; Jensen's inequality and supporting hyperplanes; convex optimization; Subdifferentials and subgradients; Legendre-Fenchel Duality; Tube formulas

Lipschitz Functions: almost differentiable functions (and rectifiable sets) Rademacher; Extensions; Approximation;

Beyond Lipschitz: BV functions Derivatives that are measures; Sets of finite perimeter; Back to Rectifiable sets

Properties of Lipschitz functions and rectifiable sets Area and Coarea formulas II; Divergence Theorem on sets with finite perimeter; C^1 approximation; Crofton's formula; Structure/regularity theorems for rectifiable sets;

Lipschitz Functions in high dimensions Balls and Spheres in high dimensions; Concentration of measure – geometric and analytic; Johnson Lindenstrauss; Lipschitz functions on Spheres in high dimensions;

Probability, data and classes of sets and functions learning sets and functions from data; classification; regression; PAC learning; VC Theory: using data to slice function families differently; Theory of Types; Sanov's Theorem; Kullback-liebler;

2.19.2 First Course in Analysis: Abridged

A minimalist introduction to analysis, that one can teach to motivated students with little background would move less quickly and require less time than the course outlined above. Here is a syllabus for that class.

Part 1: Metric spaces and inequalities Metric spaces and the usual ideas of continuity, compactness, connectedness, topology, etc; normed spaces and inner product spaces; Cauchy Schwarz and Holder and L^p spaces; Pseudo-distances, Kullback-Liebler and theory of types;

Part 2: Measure Theory and Integration I Outer measures and Caratheodory's Criterion; Properties of measures; $5r$-covering theorem; Riemann vs Lebesgue integration; Markov inequality; two theorems and a Lemma;

Part 3: Differentiation Derivatives as linear approximations; derivatives in infinite dimensions: the Laplacian from gradient descent of $f(u) = \int |\nabla u|^2 dx$; the product rule and weak derivatives; tangent places, tangent cones, and measure theoretic tangent planes; Taylor series and Smiths better result; Mean value Theorems;

Part 4: Properties of differentiable functions inverse function theorem and the implicit function theorem; Sard's theorem; understanding and using the area and coarea formulas; Lagrange Multipliers: the reason they work; Tube formulas;

Part 5: Measure Theory and Integration II Hausdorff Measures; Fractals and Cantor Sets and the Cantor function; densities; Lebesgue Differentiation Theorem; Radon-Nikodym decomposition; Extra credit: Besicovitch Covering theorem;

Part 6: Convex, Lipschitz, and BV functions ... and Rectifiable Sets Convex functions; subdifferentials and subgradients; Lipschitz Functions and Rademacher; Area and Coarea still work; Beyond Lipschitz: BV functions; Sets of finite perimeter; Rectifiable sets: intro and inspirational tour;

2.19.3 Supporting Context

There are ideas and tools that should be in the toolbox of anyone in analysis. Of course, there is a wide variety of backgrounds that analysts bring to their work. And I do not believe the "scholarly" approach is the best one: know everything about an area before you do work there. In fact, I favor cultivation of the "ignorant" but energetic/creative approach: dive in and work on a problem, learning what you need as you go along. On the other hand, their are tools that are useful to learn. The benefit from that effort is two fold: (1) the tools themselves are useful in new mathematical exploration and (2) the process of learning the tools helps you learn to create your own tools. Another source of inspiration is bare-handed work on real, data-driven problems and applications. In this work you are confronted with genuinely fresh thoughts and ideas, a fact that is unfortunately usually overlooked or at least rarely emphasized. for this reason, I promote the idea that a mathematician, to be versatile, original, freshly creative, must work on both pure and applied problems.

Here is my own (somewhat idiosyncratic) list of core, supporting and applied subjects that facilitate work in pure and applied analysis:

Real and Variational analysis, Geometric Measure Theory Analysis/geometry of sets, functions and measures in $\mathbb{R}^n$ (and in metric spaces); Forms, currents and varifolds; analysis on rectifiable sets; fractals and exotic sets and functions; analysis on and in non-smooth sets, functions, spaces; Regularity and structure of variational minimizers;

Differential <whatever> smooth and Riemannian manifolds, connections and curvature, differential forms and calculus/analysis on manifolds, intersections and transversality; fixed point theory, degree theory and other pieces of nonlinear functional analysis; dynamical systems: qualitative theory and numerics; partial differential equations;

Harmonic Analysis Singular integrals, harmonic measures, representations of all sorts, etc. This is something that is missing from the above courses, for the most part, and is really a core piece of analysis/geometric analysis. Also included here would be essential pieces of complex analysis. Applied Harmonic analysis, with its diversity of Fourier-like representations lives here as well.

Random, Stochastic things... Uncertainty propagation, ergodic theory, SODEs, SPDEs, Markov processes, random walks, PAC learning, classification/regression/prediction, noisy inverse problems, entropy and information, information geometry, type classes, coding and channels ...

Bits and Pieces Bits of algebraic topology; essential ideas and tools from numerical analysis/linear algebra; ideas from graphs and combinatorics: how to count with cleverness; maybe even a bit from algebraic geometry; key tools and ideas from optimization; algorithms;

Learning from Data Statistical learning Theory: VC theory, SVM's, Neural Nets, graph learning methods, PAC learning; Inverse Problems: image denoising, image segmentation, signal reconstruction, compressed sensing; regularization and priors; data based modeling of natural phenomena;

Chapter 3

Linear Algebra

This chapter focuses on the mathematics of linear transformations, mostly from $\mathbb{R}^n$ to $\mathbb{R}^m$, though we touch on the case of linear transformations in infinite dimensions. The ideas here are actually fundamental to understanding analysis since they are really about the nuts and bolts of linear functions between normed spaces, and that is precisely what derivatives are.

I may expand this later to a chapter more like Chapter 2, but at this point in time, this is a very short chapter with only a list of topics and a bit of explanation. If you simply read other books in order to understand the topics named in the section and subsection titles, you will know enough for most purposes (in beginning analysis).

Here is the brief outline of topics whose mastery will enable you to have a solid, working grasp of linear spaces and subspaces, linear transformations, and the properties of matrices that represent transformations and subspaces:

Vector spaces

1. **Linear Independence and Vector Space bases.** Linear Independence of $\{v_i\}_{i=1}^n$ means that linear combinations of the v_i's, $\sum_{i=1}^n \alpha_i v_i$ are not zero unless $\alpha_i = 0$'s for all i. A basis $H = \{h_i\}_{i=1}^n$ is a set of linear independent vectors such that every vector in $v \in V$ can be written as a linear combination of elements of H: $v = \sum_{i=1}^n \beta_i h_i$.
2. **Subspaces and Affine subspaces.** Subspaces include 0, affine subspaces need not include 0. (Therefore a subspace is an affine subspace, but not vice versa.)
3. **Examples.** A rich diversity: $\mathbb{R}^n$, spaces of polynomials, sequence spaces, other function spaces.

Norms

1. **In $\mathbb{R}^n$.** Norms map vector to the nonnegtive real numbers: $||\cdot|| : V \to \mathbb{R}^+ \cup \{0\}$, satisfying $||\alpha v|| = |\alpha| ||v||$, $||v + w|| \leq ||v|| + ||w||$. Important examples: 1-norm, 2-norm, ∞-norm, p-norm
2. **In functions spaces.** We have the same important examples in function spaces: 1-norm, 2-norm, ∞-norm, p-norm

Linear operators; affine operators

1. **Operator norm.** The definition: $\sup_{x \neq 0} \frac{|L(x)|}{|x|}$

2. **Reduced echelon form and what is tells you about a matrix.** You can directly read off *what the null space is* and therefore the *dimension of the null space*, which also gives you the *dimension of the range*. The reduced echelon form also tells you what columns can be used to span the range of A, so we can read off a *parameterization of the range of A*.

3. **Null space and level sets.** Suppose that N_A is the null space of A. If $A : \mathbb{R}^n \to \mathbb{R}^m$ and $y \in \mathbb{R}^m$, then if $x \in \mathbb{R}^n$ satisfies $A(x) = y$, then $L_y = x + N_A$ is the set of all points in $\mathbb{R}^n$ that map to y.

4. **Span of a set of vectors.** We denote the set of all linear combinations of the columns of A, i.e. the span of the columns of A, by $\text{span}(A)$.

5. **Determinants.** Let $E \subset \mathbb{R}^n$ and $F_A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map represented by a matrix A. The determinant of a matrix is the volume dilation factor: $vol(F_A(E)) = |det(A)|vol(E)$. The sign of $det(A)$ tells you if the orientation of $F_A(E)$ has switched or has stayed the same as the orientation of E.

Inner products and Orthogonality

1. **Orthogonality.** When we have an inner product $\langle x, y\rangle$, we say x is orthogonal to y when $\langle x, y\rangle = 0$. This is sometimes denoted $x \perp y$. If all the columns of O are orthogonal to each other and they each have norm equal to 1, we say that O is an orthogonal matrix and the columns are orthonormal. Then $OO^t = O^tO = I$, the $n \times n$ identity matrix with 1's down the diagonal and zeros everywhere else.

2. **Projections.** let P be a matrix with m orthonormal n-dimensional columns. Let $P^\perp$ be the matrix of $n-m$ orthonormal columns each of which is orthogonal to the columns of P. Then $M_P = PP^t$ is the operator which projects $\mathbb{R}^n$ onto the span of the columns of P and if x is in $\text{span}(P)$ then $M_P(x) = x$, otherwise, we can decompose $x = x_P + x_{P^\perp}$ where $x_P = M_P(x)$ and $x_{P^\perp} = M_{P^\perp}$. This decomposition into an element in P and an element in $P^\perp$ is unique.

3. **Nilpotent operators.** N is nilpotent if $N^p = 0$ for some $p > 1$.

4. **QR decomposition.** Relation to Gram Schmidt orthogonalization: they are basically the same thing. Suppose A is a matrix of m, n-dimensional vectors. Then $A = QR$, where $Q$is upper triangular and Q has orthonormal columns. Thus, $\text{span}(A) = \text{span}(R)$.

5. **Convex functions and supporting hyperplanes.** Convex functions are to optimization what linear systems of equations are to differential equations: the "easy" case (which is not so easy all the time). A closed convex subset $E \subset \mathbb{R}^n$ equals the intersection of closed half spaces containing E. If $x \in E^c$ and E is convex, then there exists a $v \in \mathbb{R}^n$ such that $\langle y - x, v\rangle < 0$ for all $y \in E$. If E is closed and $x \in E^c$, there there is a closest point in E x^*, such that $operatornamedis(x, E) = ||x^* - x|| > 0$. We have that if we define $v = x - x^*$, $\langle y - x^*, v\rangle \leq 0$ for all $y \in E$.

Symmetric operators; normal operators

1. **Eigenvectors and eigenvalues.** $(A - \lambda I)v = 0 \to v$ is an eigenvector corresponding to the eigenvalue λ. Eigenvectors can be complex numbers.

2. **Diagonalization.** A is diagonalizable if there is an invertible matrix Q such that $A = QDQ^{-1}$ where D is a diagonal matrix. Some matrices are diagonalizable if and

only if we allow Q and D to be complex. If $A = A^t$, where A^t is the transpose of A, then Q can be taken to be an orthogonal matrix: $Q^{-1} = Q^t$

3. **Jordan Normal Form** Generalization of diagonalization that works or all square matrices: allowing complex values, we are able to get that any square matrix A can be decomposed – $A = Q^{-1}JQ$ where the J is an upper triangular matrix with the eigenvalues of A appearing on the diagonal of J.

4. **Relation to operator norm.** The Jordan normal form tells us that the determinant of A equals the product of the eigenvalues of A.

Singular value decomposition (SVD)

1. **All matrices have an SVD.** It is not necessarily unique, but non-uniqueness harmless
2. **Relation to operator norm.** If $||A||$ denotes the operator norm of A, then $||A|| = \sup_i \sigma_i$.
3. **Relation to the determinant.** $\Pi_{i=1}^n \sigma_i = |\det(A)|$ – when determinant is defined. When A not square, $\Pi_{i=1}^n$ is the correct Extension of the determinant since it measure the expansion or contraction of the subspace normal to the null space of A.
4. **How it illuminates the geometry. Either:** {rotation/reflection $\rightarrow$ orthogonal projection $\rightarrow$ dilation along coordinate axes $\rightarrow$ rotation/reflection} **Or:** {rotation/reflection $\rightarrow$ dilation along coordinate axes $\rightarrow$ embedding in higher dimensional space $\rightarrow$ rotation/reflection}

What is different about infinite dimensions?

1. **Hamel Bases versus Schauder Bases.** finite combinations get everything versus infinite sums of a countable basis gets everything. (These exist if and only if the space is separable.)
2. **Subspaces need not be closed.** For example, take any Schauder basis $S \equiv \{s_i\}_{i=1}^{\infty}$ and consider all finite linear combinations of elements of S. The result is a subspace but it is not closed.
3. **All norms are not equivalent.** For example: the 1-norm of the function $\frac{1}{\sqrt{x}}$ on the unit interval is finite but the 2-norm is infinite.
4. **The unit ball is not compact.** Using the topology induced by the norm, the unit ball is not compact.
5. **Not all linear operators are continuous.** bounded = continuous.
6. **The spectrum is complicated.** There are multiple ways that $A - \lambda I$ can fail to be non-invertible. Each way generates different types of elements of the spectrum.
7. **Proving the spectral theorem.** Proving the spectral theorem for normal operators in Banach spaces is *very* involved. Proof of this statement: See Conway's book [3] on functional analysis and his proof of the spectral theorem for normal operators. For example, it involves measures on the complex plane which take values in the space of projection operators.
8. **Hilbert spaces are easier than Banach spaces.** Having an inner product and a notion of orthogonality makes many things easier/possible.

9. **Continuous, self-adjoint operators on Hilbert spaces are nice.** ... things are fairly similar to finite dimensions

10. **Recommendation.** Read through Chapters 1 and 2 of Cheney's book [2], mentioned below in the "Further Reading" Chapter, to get a sense for the main results in infinite dimensional linear theory.

Chapter 4

Further Reading

Here are some recommendations for further reading. This Chapter will grow from time to time.

1. A good book with a pretty good view of what is essential in functional analysis for somebody who is not specifically doing research in functional analysis, is Ward Cheney's *Analysis for Applied Mathematics* [2], and that is the case even if you are not really doing applied mathematics, but merely applying functional analysis to geometric analysis.

2. My favorite first book in graduate analysis is Evans and Gariepy's *Measure Theory and Fine Properties of Functions* [5]. It is appropriate for those that have had a good undergraduate course out of Fleming's *Functions of Several Variables* [6].

3. In addition to Evans and Gariepy, one should always have Folland's *Real Analysis: Modern Techniques and their Applications* [7] on hand as a reference.

4. George F. Simmons *Introduction to Topology and Modern Analysis* [10] is a favorite of mine and is extremely well written. Every young mathematician should own a copy.

5. Another favorite analysis references of mine are the appendices of Evans' PDE book [4].

Chapter 5

More Problems

Here I collect problems I have made up and those that I have found other places, like Yunfeng Hu's large collection of problems. This chapter will grow! Note that there are repetitions of exercises in the previous chapters. If you have already solved the problem, you can of course skip it, but you could also try to find a different proof of the same fact, or modify the problem in some way and solve that modification.

5.1 More Analysis Problems

5.1.1 l_p spaces

Note: I say that $\{a_i\}$ is *monotonically decreasing* when $a_i \geq a_{i+1}$ for $i = 1, 2, ...$; *Nonincreasing* is another, equivalent term.

Exercise 5.1.1. Suppose that $\sum_{i=1}^{\infty} a_i < \infty$ and b_i decreases monotonically to 0. Prove that $\sum_{i=1}^{\infty} a_i b_i < \infty$ as well. Hint: rewrite $\sum_{i=1}^{\infty} a_i b_i$ as $b_1(a_1 + d_2(a_2 + d_3(a_3 + d_4(a_4 + ...))))$ where each of the $d_i \leq 1$. Observe that $b_1 \sum_{i=1}^{\infty} a_i < \infty$ and this implies that $A_k \equiv b_1 \sum_{i=1}^{k} a_i$ is a Cauchy sequence. Use the representation suggested above to show that $B_k \equiv b_1 \sum_{i=1}^{k} a_i b_i$ is also a Cauchy sequence. More precisely, if $|A_k - A_m| \leq \epsilon$ when $k, m > N_\epsilon$, then $|B_k - B_m| \leq \epsilon$ when $k, m > N_\epsilon$ too.

Exercise 5.1.2. Find an example where $\sum_{i=1}^{\infty} a_i < \infty$ and b_i converges to 0, but $\sum_{i=1}^{\infty} a_i b_i = \infty$.

Exercise 5.1.3. Show that if $\sum_{i=1}^{\infty} |a_i| < \infty$ and b_i converges to $b \neq \infty$, then $\sum_{i=1}^{\infty} a_i b_i < \infty$.

Exercise 5.1.4. Suppose that the partial sums in Exercise 5.1.1, $\sum_{i=1}^{n} a_i$ take on both negative and positive values. Show that for any real number C, one can pick the non-increasing (and therefore, a monotonic) sequence b_i's such that $\sum_{i=1}^{\infty} a_i b_i = C$. A bit harder: Show that the b_i's can be chosen to be strictly decreasing.

Exercise 5.1.5. Define $||a||_p \equiv (\sum_{i=1}^{n} |b_i|^p)^{\frac{1}{p}}$ where n can be any positive integer. Prove $\sum_{i=1}^{n} |a_i b_i| \leq ||a||_p ||b||_q$ where $\frac{1}{p} + \frac{1}{q} = 1$. Hint: assume that $||a||_p = ||b||_q = 1$ and use log's and Jensen's Inequality.

Exercise 5.1.6. The sequence space l_p, $1 \leq p < \infty$ is defined to be all the sequences $\{a_i\}_{i=1}^{\infty}$ such that $||a||_p \equiv (\sum_{i=1}^{\infty} |a_i|^p)^{\frac{1}{p}} < \infty$. In the case that $p = \infty$, $||a||_\infty \equiv sup_i\{a_i\}_{i=1}^{\infty}$. Prove Holder's inequality for l_p spaces.

Exercise 5.1.7. Use Holder's inequality to prove that $\sum_{i=1}^{\infty} |a_i b_i| \leq \sum_{i=1}^{\infty} |a_i| \sum_{i=1}^{\infty} |b_i|$. Hint: Prove that $||a||_p < ||a||_1$ when $a \in \mathbb{R}^n$, i.e. when a is a finite sequence: $a = \{a_i\}_{i=1}^{\infty}$.

Exercise 5.1.8. Note that l_2 has copies of each of the spaces $l_2^n \equiv \mathbb{R}^n$, $\forall n < \infty$, embedded in it isomorphically. More specifically, identify sequences which have all bu the first k components set identically to 0, with the space $\mathbb{R}^k$ and name this subspace of l_2, l_2^k. Prove that $\bigcup_n l_2^n$ is a subspace of l_2 that is not closed in l_2, but is dense in l_2.

Exercise 5.1.9. In finite dimensional vector spaces V, (1) a linear map from V to V that is onto is one-to-one and (2) a linear map that is one-to-one is onto. (Show that!) Define S_r and S_l by $S_r : (a_1, a_2, a_3, ...) \to (0, a_1, a_2, a_3, ...)$ and $S_l : (a_1, a_2, a_3, ...) \to (a_2, a_3, a_4, ...)$. Show that these provide counterexamples in infinite dimensions to (1) an (2) above.

Exercise 5.1.10. Find the matrix representation of S_l in the case that we are working with sequences of finite length n, and show that $(S_l)^n \equiv \underset{1}{S_l} \circ \underset{2}{S_l} \circ \cdots \circ \underset{n}{S_l} = 0$. What similar kind of statement can we make about S_l when we are working in infinite sequence spaces? Hint: for a **fixed** $a \in l_p$, think about $(S_l)^k(a)$ as $k \to \infty$.

Exercise 5.1.11. Show that if $A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator, then it has both left and right inverses. Hint: Show that if the columns of A are independent (which implies A is right invertible) then, if there is an $x \in \mathbb{R}^n$ such that $x \neq 0$ and $x^t A = 0$ (which must be the case if A is not left invertible), you can find a $y \in \mathbb{R}^n$ such that $Ay = x$ and that is a contradiction.

Exercise 5.1.12. Continuing the line of thought in Exercise 5.1.11, show that if $A : \mathbb{R}^n \to \mathbb{R}^n$ is linear, and $\hat{A}_l A = I = A\hat{A}_r$, $\hat{A}_l = \hat{A}_r$. Hint: we can work with any matrix representing A, and $I =$ the matrix with ones down the diagonal, commutes with all matrices ...

Exercise 5.1.13. Continuing along the lines of Exercise 5.1.10, show that Exercises 5.1.11- 5.1.12 do not work in infinite dimensions by showing that $S_l \circ S_r = I$ but that $S_r \circ S_l \neq I$.

Exercise 5.1.14. Suppose that $a \in l_2$ and $||a||_2 < \infty$. Define $A_a : x \in l_2 \to y \in l_2$ by $y_i = x_i a_i$. Is A a one-to-one mapping? Is A_a onto? Does the answer depend on a

5.1.2 Representation and Approximation

Exercise 5.1.15. If $f : [0, 1] \to [0, 1]$ is continuous, then Weierstrass-Stone approximation theorem says that there is a sequence of polynomials that converge in the uniform norm to f on $[0, 1]$. How does this imply that only a countable infinity of numbers is needed to describe a continuous function? But there is an easier way. Hint: what are the values of f at numbers of the form $\frac{m}{2^k}$ where $m \leq 2^k$.

Exercise 5.1.16. How much information does it take to encode a function? Suppose that $f : [0, 1] \to [0, 1]$ monotonically: $f(x) \leq f(y)$ when $x \leq y$. Show that you can describe f with a countable sequence of numbers. Hint: define $a_1 = x \in [0, 1]$ such that $a_1^l \equiv \lim_{x \uparrow a_1} f(x) \leq \frac{1}{2}$ and $a_1^r \equiv \lim_{x \downarrow a_1} f(x) \geq \frac{1}{2}$ and $b_1 = a_1^r - a_1^l$ and $c_1 = f(a_1)$. Three infinite sequences suggested describe f at all points in $[0, 1]$. Complete the description of the three sequences and prove they do describe f completely.

Exercise 5.1.17. There is another approach to the description of monotonic functions of Exercise 5.1.16 that, like the second method of solution in Exercise 5.1.15 simply samples the values of f at numbers of the form $\frac{m}{2^k}$ where $m \leq 2^k$. But another possibly infinite sequence of pairs of numbers is needed! What are those pairs needed for? Hint: If f has discontinuities, what does the sampling of f possibly not capture.

5.2 More Linear Algebra Problems

While there are linear algebra problems in the analysis problems above in this chapter and there will be analysis type problems in this section, generally speaking the motivation for a thread of problems here will be some question involving linear operators, subspaces, projections, decompositions, matrices, etc. I will not distinguish between a vector and its representation in coordinates, nor between an operator and its matrix representation unless confusion would result.

Exercise 5.2.1. The projection of a vector in $\mathbb{R}^n$ onto some unit vector $w \in \mathbb{R}^n$ is given by the rank 1 $n \times n$ matrix ww^t. Prove that $I - ww^t$ is the projection onto $w^\perp$, the orthogonal complement of w.

Exercise 5.2.2. Suppose that $w_1, ..., w_k$ are k orthonormal vectors in $\mathbb{R}^n$ and $n > k$. Let $W \equiv \text{span}(w_1, ..., w_k)$. Inspired by Exercise 5.2.1, what is the operator projecting $\mathbb{R}^n$ orthogonally onto $W^\perp$?

Exercise 5.2.3. Suppose that $a_1, a_2, ..., a_n$ are n distinct real numbers. Prove that

$$A = \begin{bmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ 1 & a_3 & a_3^2 & \cdots & a_3^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{bmatrix}$$

is invertible. **Hint:** Show the columns are linearly independent. Notice that for any $\alpha \in \mathbb{R}^n$ the first element of $A\alpha = f(a_1)$ for some $f(x)$. What kind of function is $f(x)$?

Chapter 6

Solutions to some of the problems

Solution to Exercise 2.7.23: The key idea is that we can densely sprinkle around tiny spots where $f' = 1$, but limit the total footprint of those spots to ϵ. You should sketch what is going on in the construction below, because what looks complicated is just the fact that writing down something that is not too complicated to draw is not easy.

1. Define
$$\hat{\Lambda}(x) = \begin{cases} 0 & x \leq -1 \\ x+1 & -1 \leq x \leq 0 \\ 1-x & 0 \leq x \leq 1 \\ 0 & 1 \leq x \end{cases}$$

2. Let $\phi : \mathbb{R}^1 \to \mathbb{R}^1$ be an even, positive C^∞, bump function centered at the origin, with support in $[-\frac{1}{16}, \frac{1}{16}]$ and $\int \phi \, dx = 1$.

3. Define $\Lambda(x) = (\hat{\Lambda} \star \phi)(x)$ – Λ is the convolution of $\hat{\Lambda}$ with ϕ. Λ is a C^∞ version of $\hat{\Lambda}$, that equals $\hat{\Lambda}$ except on the intervals $(-\frac{17}{16}, -\frac{15}{16})$, $(-\frac{1}{16}, \frac{1}{16})$, and $(\frac{15}{16}, \frac{17}{16})$ where it has been smoothed/mollified. Note also that $\Lambda(x) \leq \hat{\Lambda}(x)$ for all $-\frac{1}{16} \leq x \leq \frac{1}{16}$.

4. Define $\Lambda_\alpha(x) = \alpha\Lambda(\frac{x}{\alpha})$. For $r < 1$, Define the (r, y)-patch to be the graph of the function Λ_{r^2} inside the interval (r, r), all shifted so that the interval and function are centered on y.

5. Define $\hat{P}_1 = \{\frac{1}{2}\}$ and $\hat{P}_k = \{\frac{1}{2^k}, \frac{2}{2^k}, ..., \frac{2^k-1}{2^k}\} \setminus \bigcup_{i=1}^{k-1} \hat{P}_i$.

6. Now choose $\epsilon << 1$.

7. Choose r_1 such that $2r_1 = \frac{\epsilon}{2}$

8. Define $P_2 = \hat{P}_2 \cap \left(\bar{B}(\frac{1}{2}, r_1)\right)^c$.

9. Define $P_i = \hat{P}_i \cap \left(\bigcup_{p \in P_j, j<i} \bar{B}(p, r_j)\right)^c$ where r_i is defined by:
   (a) $2r_i|P_i| \leq \frac{\epsilon}{2^i}$
   (b) $\bar{B}(p, r_i) \cap \left(\bigcup_{q \in P_j, j<i} \bar{B}(q, r_j)\right)^c$ for all $p \in P_i$.

10. Notice that all the $\bar{B}(p, r_i) \cap \bar{B}(q, r_j) \neq \emptyset$ whenever $p \in P_i$ and $q \in P_j$.

11. Define $E \equiv \left(\bigcup_{p \in P_i, i=1,2,\ldots} \bar{B}(p, r_i)\right)^c$ and notice that $\mathcal{H}^1(E) \geq 1 - \epsilon$. Notice also that $E = \bigcap_{k=1}^{\infty} E_k$ where $E_k \equiv \left(\bigcup_{p \in P_i, i=1,2,\ldots,k} \bar{B}(p, r_i)\right)^c$ and each E_k is open.

12. Define $f(x) = 0$ when $x \in E$. In each of the $\bar{B}(p, r_i)$ we define f to be a (r_i, p)-patch. Note that f is Lipschitz with Lipschitz constant 1 is differentiable everywhere and that f' is not continuous at any point of E. (It takes a little effort to verify this, but depends on the fact that if $x \in E$, then $x \in E_k$ for all k, so we can choose a radius for a neighborhood centered at x such that this neighborhood does not intersect $(E_k)^c$ which means that all of the graph of f in this neighborhood fits in a cone with upper slope $\leq r_{k+1} \leq \frac{\epsilon}{2^{k+1}}$. Thus while there are points y arbitrarily close to x such that $f'(y) = 1$, the derivative exists at x and equals 0.)

Appendix A

Manifolds and Nonlinear Analysis I: Nonlinear spaces and Manifolds

In this lecture and the next, we give an intuitive overview of some ideas in nonlinear analysis. We will deal primarily with manifolds and mappings between manifolds.

A.1 $\mathbb{R}^n$ 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 $\sum_{i=1}^n x_i y_i$.

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.

A.2 k-Manifolds in $\mathbb{R}^n$ are locally like $\mathbb{R}^k$

Definition A.2.1 (**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 neighboorhood 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 deriviative 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 A.2.2 (**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 look at derivatives as approximations to mappings.

A.3 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 A.3.1. *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 A.3.2 (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 A.3.1 ($F : \mathbb{R}^n \to \mathbb{R}$). *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 deratives of F.*

Remark A.3.1. *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 A.2.1.*

When F is differentiable, it is natural to ask, “How differentiable?”

Definition A.3.3. *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.*

A.4 Full rank maps

Definition A.4.1 (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 A.4.1. *If a matrix is full rank, then a sufficiently small perturbation will not change that fact.*

Definition A.4.2 (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 A.4.3 (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 A.4.4. *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 A.4.1 (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 $k = \max(0, n-m)$.*

We will see the reasons for this in detail in the next section.

A.5 Inverse and implicit function theorem

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

Theorem A.5.1 (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. Choose $0 < \epsilon < 1/2$

1. Define $G = I - DF_0^{-1} \circ F$.
2. 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 $h \in U$.
3. Define $W = B(0, \frac{\delta(\epsilon)}{2})$.
4. Using the mean value theorem in vector spaces, we get that restricted to W, G is a contraction mapping with contraction constant ϵ.
5. Define $H = (I + G + G^2 + G^3 + ...)$. Notice that H is differentiable and $DH = I + DG + DG \circ DG + ...$
6. Notice that $D(H(I-G)) = DH \circ D(I-G) = I$. Choose $y \in W$. Integrating, we get:

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

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

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

8. 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 $F \circ \hat{F} = I_{F(W)}$.

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

□

Theorem A.5.2 (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 neighborhod of $(F(x^*), x^{*\prime\prime})$. 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 $||y'' - x^{*\prime\prime}|| < \epsilon$ for some sufficiently small $\epsilon > 0$. □

Example A.5.1. *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.*

Appendix B

Manifolds and Nonlinear Analysis II: Nonlinear Thinking

B.1 Lipschitz Functions

Definition B.1.1 (Lipschitz Mappings). *$F : X \to Y$ is* **Lipschitz** *if there is a positive number $K \geq 0$ such that $|x - y| \leq K|F(x) - F(y)|$ for all $x, y \in X$.*

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

Theorem B.1.1 (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.

B.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 B.2.1 (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 B.2.2 (**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 B.2.3 (**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 B.2.4 (**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 B.2.1. *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 forrmula you have seen for integrals in calculus. To really understand these two formulas, we need to look at simple examples.

Example B.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 B.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 thatthe 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 B.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 tp the sum of the squared determinates of the $n+1$, $n\times n$ submatrices. When we compute this we get $\sqrt{1+\nabla F\cdot\nabla F}$. Anothe 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 noicing that the area of a little pice of the graph is increased by exactly the ratio between the hypotenuse of a triangle with horizontal 1, vertial side $||\nabla F||$ and the horizontal side length.

Remark B.2.2 (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.*

B.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 B.3.1 (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, yelding 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 [9].

B.4 Transversality is Generic

Intersections of submainfolds 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 nof noise, or jiggle some parameters a tiny bit?

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

Definition B.4.1 (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* $\mathrm{span}(U_k, W_m) = \mathbb{R}^n$.

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

Definition B.4.2 (Transverse Intersection of Submanifolds). *Two sumanifolds 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* $\mathrm{span}(T_xM, T_xN) = \mathbb{R}^n$.

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

Example B.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 B.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.

B.5 Fixed point theorems: Banach Fixed Point Theorem

Many problems can be written as:

Problem B.5.1 (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 B.5.1 (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 B.5.2 (Cauchy Sequence). *Recall that $x_i \in X$ is Cauchy if for any $\epsilon > 0$ there is an $N(\epsilon)$ such that $i, j > N(\epsilon)$ imples that $||x_i - x_j|| < \epsilon$.*

Definition B.5.3 (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 B.5.4 (Banach Space). *A complete, normed vector space B is called a* **Banach Space***.*

Definition B.5.5 (Contraction Mapping). *A function from a normed space X to itself is a* **contraction maping** *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 B.5.1 (Banach Fixed Point Theorem). *Suppose that $F : B \to B$ is a contraction maapping 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 continous, 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 $F(x^*) = x^*$.

□

Appendix C

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:

$$\lim_{i\to\infty}\left(\int f_i \, dx\right) = \int \left(\lim_{i\to\infty} f_i\right) \, dx? \tag{C.1}$$

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.

C.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 C.1 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 C.1.1. 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 C.1 is yes.

Exercise C.1.2. 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 C.1 is no. Hint: look at constant functions on the real line.

Exercise C.1.3. 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 C.1.4. Give an example of a non-uniformly convergent sequence f_i on a finite measure space X where again the answer to C.1 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.

C.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

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

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 C.2.1. 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 C.2.2. 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 C.2.3. Find an example to show that convergence in measure does not imply convergence in norm. Hint: the f_i need not be bounded.

Exercise C.2.4. 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 C.2.5. 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 C.2.6. Look at all the posibilities! Suppose we identify each of the 5 bit binary numbers with a set of convergence types: $f_i \rightarrow_{01101} (f \equiv 0)$ 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?

C.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 C.3.1.

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 C.3.1 ($\limsup_{i\to\infty} f(i)$ 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 C.3.2. Rework Exercise C.3.1 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 C.3.2 ($\limsup_{i\to\infty} f_i(x)$ 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 C.3.1 (Fatou's Lemma)**.**

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

Theorem C.3.2 (Monotone Convergence)**.** *Suppose that* $\{f_i\}$ *are all measureable and that* $0 \leq f_1 \leq ... \leq f_i \leq f_{i+1} \leq ....$ *Then we have that*

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

Theorem C.3.3 (Dominated Convergence Theorem)**.** *If* $f_i \to f$ μ *a.e.,* $|f_i|, |f| < g$ *and* $\int g \, dx < \infty$*, then*

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

C.3.1 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 C.3.3).

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

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

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

□

Exercise C.3.3. 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 C.3.1. *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 C.3.2. *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 C.3.3. *Dominated convergence is used to get other switching theorems: switching order of differentiation and summation or differentiation and integration or integraion and summation. We will discuss this later in this chapter.*

Bibliography

[1] Ronald N. Bracewell. *The Fourier Transform and its Applications*. McGraw-Hill Book Company, 1986.

[2] E. Ward Cheney. *Analysis for Applied Mathematics*. Springer, 2001.

[3] John B. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer Verlag, 2nd edition, 1990.

[4] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathmatical Society, 2nd edition, 2010. ISBN 978-0-8218-4974-3.

[5] Lawrence C Evans and Ronald F Gariepy. *Measure Theory and Fine Properties of Functions*, volume 5. CRC press, revised edition, 2015.

[6] Wendell H. Fleming. *Functions of several variables*. Springer, 1977.

[7] Gerald B. Folland. *Real Analysis: Modern Techniques and their Applications*. Wiley-Interscience, 2007. ISBN-10: 0471317160, ISBN-13: 978-0471317166.

[8] AC Kak and M Slaney. *Principles of Computerized Tomographic Imaging (Philadelphia, PA: SIAM)*. SIAM, 2001. http://www.slaney.org/pct/pct-toc.html.

[9] John W. Milnor. *Topology From The Differentiable Viewpoint*. University Press of Virginia, Charlottesville, 1965. Eighth printing, 1990.

[10] George F. Simmons. *Introduction to Topology and Modern Analysis*. McGraw-Hill, 1963.