In and Around Geometric Analysis: An Invitation

by

Kevin R. Vixie

Contents

Read This Preface!

Analysis is often first encountered in an undergraduate course that seems like calculus plus rigor. While a few find it interesting, many are turned off from analysis because of the mundane and needlessly dry nature of such a class. This text is intended to change that conception by introducing students to ideas in analysis that are not often encountered until a student is in the second or third year of graduate school.

The Backstory for This Text

I decided to write this text after teaching parts of the contents of this text to three undergraduates during the summer of 2011. We met 2-3 hours a day, 2 days a week. The next year I expanded it a bit and taught it to about 10 students – I believe it was 8 graduate students and 2 undergraduate students. I did not assume a course in analysis as a prerequisite, though students with and without that background took the course.

The text will not dwell on applications though it will make many references to applications and to texts that deal with applications. And though this text is not about applications, the inspiration for this text comes from applications. In particular, geometric measure theory or more broadly, geometric analysis (which I define more organically to be the intersection of analysis and geometry) are loaded with under-exploited insights and tools, begging to be applied to all sorts of interesting data analysis and data modeling problems. Less often seen and discussed, is the fact that these areas of applications are full of inspiration for new mathematics on the purer side of mathematics.

I am very geometrically and intuitively oriented, having become more analytically skilled only through my interest in geometric measure theory and geometric analysis. That bias comes through in the book. In fact, I believe it is what permits me to introduce more advanced ideas long before they would be introduced in the usual progression of courses that a student might take. In particular, I will make efforts to give intuitive arguments for all the ideas introduced, many of which can be developed into full proofs with a bit of work. I will also give detailed references to works in which proofs of results can be found.

One note of caution: I am sure that my way of dealing with some definitions will be irritating to some very precise souls, who will view what I do as too fast and loose for anybody's good. My advice to them is to read other books. There are many other books, though I firmly believe that there is room for a whole host of approaches and the one in this book being too scarce a model in published texts.

Part of the inspiration for the course is Bill Thurston's 1994 Bulletin article [41], where he shows a few entries from a list of the different ways in which the derivative can be understood. He made the statement that "The list continues; there is no reason for it ever to stop". That stuck with me. The course notes that inspired this book were based on the idea that derivatives can lead you (almost) anywhere in analysis. And to prove it, I did just that, throwing in integration and high dimensional geometry as needed.

Another premise of that course was that geometric intuition is a very powerful tool that can be exploited to bring students with a modest background into productive contact with mathematics that is much more sophisticated than they would encounter at their level.

The final premise that formed the basis for these notes is that the first exposure to analysis need not be the dry, calculus + $\{\epsilon$'s and δ's$\}$ that is usually taught. Analysis is a fascinating area, with big surprises and beautiful opportunities for exploration. The typical introductory course in no way makes that clear, though the three other books I have used for this course (books by Kennan Smith, Wendell Fleming, and Tom Lindstrøm [38, 14, 25]) were very pleasant and inspiring exceptions to this pessimistic assessment.

Minimalism

I am now convinced that the path to mastery is characterized by a barehanded immersion, enabled by recognition of a minimalist path, in which the essence is recognized and used to strip away the heavy impediments that accumulate when the minimalistic core is not recognized and exploited.

I think I first became consciously aware of this kind of approach to problems when I sat in mathematics classes taught by Andy Fraser, my PhD advisor. His physicist knack for cutting through the complexity by seeing the simple core of things was profound and inspiring. Moving to the lab, seeing other physicists do the same thing (John Pearson and Chris Morris spring to mind) further influenced my focus on essence.

When I discovered my muse was in geometric measure theory, inspired by David Caraballo – he was the pied piper, I was happy to be led into the field – I found that if I looked for the geometric core, I could penetrate things that otherwise looked very difficult. Sometimes doing that was even easy.

Then I began teaching geometric measure theory and analysis, both at the graduate level and the undergraduate level. I realized that the usual teaching style – more or less writing the textbook on the board – was not only numbing and boring, it was not something I could bring myself to do. I saw it as my duty to highlight nuance, to delve into delicate corners the students missed, to point out the simplicity at the core and otherwise enlighten them to viewpoints that enabled them to see further than if they simply studied the book.

All this convinced me the way to gain mastery is to find and follow the minimalistic, barehanded, immersive path to an instinctive facility – one that insists on mastery of the simple essences that truly unleash the power of the subject.

The writing of this book

I decided to create many of the figures by hand, so that the figures would resemble what might appear on the blackboard or whiteboard if we were having a conversation together, deeply engaged in chasing the nuances, in finding the essence and using that to see more deeply.

When this is released, I will have finished the first edit of the book, but I will continue to revisit the book from time to time. I have a different perspective on writing and editing, including the fact that I am very, very careful about not letting the process of editing remove too much.

In fact, I think that the real art is in letting there be just enough of what some would call raw or rough traces in the writing to lend real authenticity to the writing. I think perhaps a better term would be "idiosyncrasies", instead of the phrase "raw or rough traces".

I know that a trained editor might itch to make those pieces conform to their view of writing, but this is not something I believe is the right thing to do if readers are to see into the writer's experience, into what they might experience in a conversation with the writer.

Prerequisites

As far as prerequisites go, I assume very good courses (and good performance) in the entire calculus sequence, including vector calculus, and both linear algebra and differential equations. A course in proofs, especially one that works on simple analysis proofs and ideas in metric spaces would be helpful.

Ideally, a course that taught the basics of metric spaces would also be a prerequisite, but I have included an early chapter with the results we will need from metric spaces because few seem to have experience with metric spaces before they take their first analysis class.

I would suggest that this book is best when combined with either a mentor or at least someone who can give you pointers now and then, because this is what I do with my students. A little suggestion this way or that can be helpful from time to time. If you do not have such a person, perhaps you can find another student to collaborate with. But if you are reading this and dedicated to learning the subject, you will find that those who are already skilled in the subject are more often than not willing to be generous with their time. Because truly motivated, passionate students are not very common and are a joy to work with!

A Note on Problems and Exercises

One of my favorite geometric analysis books, Evans and Gariepy [12], has no explicit exercises, but I found that studying the book closely required me to fill in between the lines, so to speak, and frequently left me asking why conditions were necessary, how I could vary definitions or theorems and how I could extend ideas in various ways. The result was that these exercises and problems that emerged from my own deep reading, seemed highly natural and organic. I found that they fit with the flow that I would get into when studying the text and did not feel like a distraction or an unwelcome, alien invasion of the flow I was immersed in. The same cannot often be said about the traditional approach to problems in many mathematical texts.

Ideally, you will focus on emergent exercises and problems when you read this text, only some of which you will find in those I have put into the text. While I believe that many of the exercises in the book fit this spirit of exercise and problem generation, such a mode is inescapably personal, so when you read this book that deeply, you may very well end up with different questions you want to answer. You should therefore not be afraid to diverge (at least sometimes!) from the exercises and problems I have selected. It also takes practice to learn how long to work on a problem before giving it a rest or how to pick the problems you want to work on. Not every problem needs to be solved for you to master the text!

This art of learning how much to work on a problem and when to diverge from stated problems that, for you, break the flow you are emerged in, will take time and is made much more enjoyable when there are others you can talk with, if for no other reason but for you to explain to someone else what you have learned. I therefore encourage you to find at least one or two others who are as inspired as you are to learn the subject.

Finally, there are deep benefits that you can reap from merely thinking about a problem carefully, even if you do not solve the problem. You might find that you are inspired to vary the problem, to go off on an inspiring tangent suggested by that problem. I would encourage this! If there is anything that kills the spirit of deep flow, it is the idea that there is a "right way" to get to some goal, that the author or expert should be treated as an oracle or guru, that they know better than the muse of your own spirit in the midst of flow that you find. Down with such ideas! While those who have gone before have useful insights and even wisdom, they are all partial and in the end, often faulty in some (perhaps big, perhaps small) way. This is a journey that you must take a craft, You have to open to the muse, to the flow. Those "experts" and "authorities" know little of your precise path and may have little to say that can help. Learning to filter what others say, to hear when what is said or claimed fits with the flow you have found, and having the courage to move in different direction (not simply to be different, but because your inspired path is in that direction) is a part of becoming a true master.

Notes on Chapters

As made obvious by it's title, I expect everyone to read the Preface and I believe the Preamble, though highly unusual, is perhaps the most important three Chapters of the book.

The next section, *Analysis I*, is an invitation to analysis intended as something between an informal flyover of fascinating territory and teaser to be sampled here and there, hopefully dispelling the notion that Analysis is merely calculus, with ϵ's and δ's.

When I teach from this book, I am usually teaching to students whose preparation in metric spaces and inequalities is very, very weak. They also misunderstand proofs and the central role of creative exploration, of taking the time to think and feel and see. As a result, in the first semester, we focus on the *Analysis I* Section and spend quite a bit of time on Chapter 7, *Just Enough Metric Spaces* and Chapter 8, *The Art of Inequalities*.

The second semester focuses on the *Analysis II* Section and mastery of Chapters 11 and 12, *Derivatives* and *Measures and Integrals*, respectively. This might seem, at first sight, to be a rather limited course, but the path taken in these chapters lead to interesting and surprising directions for students expecting analysis to be just very fussy calculus (i.e. calculus with ϵ's and δ's).

In the Section titled *Analysis I*, Chapter 9, on everything linear, is assumed known, though in practice most have to review this chapter. Likewise, before tackling chapters 11 and 12, it is expected the students will work through Chapter 10.

The final section of the book, *Analysis III*, is an invitation to deeper analysis in general and geometric measure theory in particular.

Have Fun!

I do intend that all 16 Chapters be read, but I suspect that most students will not do this first time through. Nevertheless, those that are aiming at mastery will read more than is required, and in a mode that yields immersion and flow. Taking the time to think and feel an see (see David Levy's Google Tech Talk [23]), they will learn that mastery is both hard work and, at the same time fun! For, as is clear from Chapter 1 in the *Preamble*, playfulness is central to mastery, and, as a result, so is having fun!

In order to spread the fun as widely as possible, the e-copy will always be free and you should also feel free to print a printed copy for personal use.

If you find typos or mistakes, I would be grateful to hear about them from you and will of course acknowledge the fact you did so in the corrected version. While the e-copy will be updated when typos/errata are pointed out, the printed version I have for sale will be updated quite a bit more slowly.

I hope that the readers of the book will find the contents inspiring in the best way – that they will find themselves, time and time again going off on tangents of their own devising, exploring, creating, asking, answering.

Acknowledgments

In a work such as this, tracing the influence of those that supported me is not actually possible – one can try, and I will here, but my own mastery and what I have learned has certainly been the result of a very meandering path to the present, with all sorts of nuances interacting in highly nonlinear ways to produce what I know, what I see, what I feel about the subject.

My first exposure to the subject was the study I made of Lipman Ber's book on calculus, given to me by my father, a polymath musician to whom I owe a great deal in my own intellectual development, as much if not more, for the rich environment he made a reality as for the explicit influence he had.

Early on in my mathematical studies, Justin MacCarthy, an erudite scholar and mathematician who happened to be in Deming NM, where I lived, spent countless hours beguiling me with stories and little bits and pieces of mathematics.

Moving to college from my homeschool (autodidactic) explorations of mathematics, I came under the influence of the mathematics professors Tom Thompson and Ken Wiggins (both of whom also opened their home to me), as well as physics professor Gordon Johnson and Electrical Engineering Professors Carlton Cross and Rodney Heisler. From them, and with great patience for my wildly inconsistent performances due to the trauma of parents dying, I obtained a good foundation in mathematics, physics and even some engineering. (The Melvin Johnson family, in College Place, Washington, was also a place of safety and healing, in the midst of loss and distress I could not even really process at the time. The Thompsons, Wiggins and Melvin Johnsons were very much responsible for the fact I finished undergraduate school.)

At UW, my interactions with Chris Bretherton were deeply inspiring. After pleasant interactions at OSU in Corvallis (Ron Guenther) and Oregon Graduate Institute (several applied physics professors) I landed at Oregon Health Sciences University running Johnny Delashaw's Research Lab in Neurosurgery.

At that time I started thinking about mathematics again. Two truly remarkable events were foundational to my return to mathematics. First, I met Beata and 5 months later we were married. Then about two and a half years later, Levi was born. Two remarkable conversations – one with Johnny Delashaw and one with Tom Thompson – motivated me to return to graduate school, which I did there, in Portland Oregon. I did this, first part time and then, after quitting OHSU, full time. I returned to work with Andrew M. Fraser, who had previously showed patience with my forgetting to tell him about a cancellation of a talk that he needlessly drove over an hour to attend.

Classes in mathematics from John Erdman and Serge Preston in the mathematics department and Andy in Systems Science were highlights of the experience there. I also benefited from classes taught by other faculty, including the systems science philosophy class from Marty Zwick.

Moving to Los Alamos to work on a dissertation in 1998, through connection to Andy's friends there, opened up a huge new universe of people and ideas. (To see more of those names, consult the list in my dissertation.)

The influence of David Caraballo was deeply significant, for our meeting in San Antonio in 1999 was the beginning of a genuine friendship and the start of my immersion in geometric measure theory. Andrea Bertozzi invited me to Duke university in about 2001 or 2002. While I was there, she introduced me to Bill Allard. I invited Bill to the Lab and we quickly became friends and collaborators. I introduced him to new applications in the data analysis and signal/image processing world, while he started teaching me the finer points of geometric measure theory. Through Bill I became friends with Frank Morgan, Bob Hardt, and Jean Taylor.

Moving to WSU in 2008, I began work with my students, that included several papers at the intersection of geometric measure theory and applications or computations. Many of these were collaborations with Bala Krishnamoorthy whose computational expertise made the work possible.

Teaching multiple advanced courses in analysis and geometric measure theory, both before I moved to WSU (at UCLA/IPAM in 2007 and LANL summer 2007) and at WSU, were invaluable in my developing sense for my corner of geometric analysis.

Teaching the undergraduate analysis classes 401-402 over multiple years has also deepened my grasp of many aspects of analysis, as an art with a minimalistic core.

My current and former graduate students include Sharif Ibrahim, Keith Clawson, Ben Van Dyke, Eric Larson, Josh Sackos, Josh Cruz, Abigail Higgins, Heather Van Dyke (now Heather Moon), Hossein Noorazar, Justin Theriot, Yunfeng Hu, Josh Kaminsky, Yusen Zhan, Laramie Paxton, Yufeng Cao, Henry Riely, Vlad Oles, Enrique Alvarado, Katrina Sabochick, Rommel Cortez, Curtis Michels, Jared Brannan, Blake Cecil and Liya Boukhbinder. Some of these were students whose committee I only served on, not led, but whom I spent time mentoring anyway. All of these students influenced me one way or another.

My family, defined organically includes not only Beata Vixie, Levi Vixie, and my brother, Curtis Vixie, (and Oliver and Charlie!), but also Patrick Campbell, Gary Sandine, and Michael Forbes. The positive, inspirational influence of these people is too large to even outline here. Another highly significant influencer in my life is Bala Krishnamoorthy, who has been a constant collaborator and consistent ally the entire time I have been at WSU. They have had influence on how my vision for the subject develops that is hard to describe and would be impossible to overstate.

Michael Forbes not only is my closest collaborator at WSU, but is also my friend and the expert in LaTeX who helped me by creating the style used to write this book.

Acknowledgments are also due to the students who sat through very early versions of pieces these notes. Two students in particular – Josh Sackos and Josh Cruz – paid very close attention to those classes. For that I am grateful.

I am also grateful to Liam Crafton for pointing out numerous errata to me when he took the 2 semester sequence from me 2021-2022, as well as to Hossein Noorazar and Altansuren Tumurbaatar, both former graduate students of mine (and Matt Sottile's), and Curtis Michels, a current graduate student, all of whom are reading through the book and sending me lists of errata. Patrick Campbell and Gary Sandine, longtime friends and collaborators have also submitted detailed lists of corrections.

to my family

Version Notes

Version: 1.0, October 30, 2023
ISBN (print): 978-1-7357988-2-0
ISBN (e-copy): 978-1-7357988-3-7
Author: Kevin R. Vixie

Contact: vixie@speakeasy.net
Mathematics Website: http://geometricanalysis.org
Photo Website: http://viksekrarts.com
Notes Website: http://notesfromkevinrvixie.org

Part I

Preamble

1 Intuition, Playfulness and Proving Things: A Practical Guide

In the next three short chapters, I will grapple briefly with the meta-issues that are rarely talked about in courses on analysis, but are, nevertheless, crucially important for real mastery of the art of analysis. While it is true that some (or much) of this is best learned individually because it has to be learned in a unique way by everyone while actually doing analysis, bringing these subjects into the awareness aids in that (at least semi-solitary) process.

The hardest thing for many students is the transition from crank and grind, black-box problem solving to the flow mode in which they vividly imagine new things, grasp subtle arguments solidly enough so they can modify, extend or completely transform them, and immerse themselves so deeply that the problem comes alive. This does not happen very quickly for most.

Adding to this challenge is the fact that so many students are taught proofs in ways that are, at best, unhelpful and at worst, simply incorrect and misleading.

Formal, written proofs are, very simply, clear arguments that convince careful thinkers of some statement's truth. They are a small part of doing mathematics. Many times students come away from proof classes thinking sentence structure, proof structure, and very particular language are crucial or even central to mathematics. But this is false. Even grammar can be unimportant, because there are things that give grammarians a fit, yet in no way decreases the clarity of a written statement. There are styles of writing proofs that use very short sentences quite effectively. While these sentences are usually complete, they are very far from flowing prose that students just fresh from proof classes seem to write in.

But instead of focusing on what is wrong with the typical proof class experience, I will focus on two things I believe should be taught, before, during and after the topics in analysis are studied.

1.1 Playfulness as a Strategy

A huge piece of making the transition from routine, crank and grind mathematics to creative and original thinking, is the devotion to a playful attitude and mode.

While I do not always succeed in staying in that mode, this is always my goal when I have the quietness and emotional resources this takes.

The goal is child-like wonder, flexibility, and optimism. (It is very important to distinguish between "child-like" and "childish" – there is no advantage and many disadvantages of childish modes.) The child-like states are states of unbounded creativity and originality – states in which the possibilities are endless. A crucial component here is the emotional health that comes from having dealt with past traumas – for trauma, and the fear and closed, disconnected mode of living that results often blocks this state of flow.

1.1.1 Why Trauma Enters This Discussion

Though I know talking about trauma and emotions is unusual in a book on technical mathematics, I am convinced it is necessary because mathematics is an art, and art depends on our emotional resources. One has to know how to carve out places in space and time where safety and flow are possible. Eventually, to reach the highest levels of creativity and consistent, sustainable flow leading to freshly original productivity, past emotions have to be healed.

Fortunately, there is a lot of awareness and methods to heal trauma. A good starting point for those that want to know more are the talks by Tara Brach and Gabor Mate [8, 28]. It is now well understood that this is a whole person mission – you body actually does keeps the score. See the book with that title of a by Bessel van der Kolk [42].

And while it is true that trauma also gives you gifts, those gifts cannot be fully reaped until the trauma is healed. So it makes no sense to ignore these issues.

In my own case, those gifts included mastering all the skills of an introvert and finding solace in deep study and the solitude of wandering the desert with my dog. Opening to healing did not remove those skills or the impulse to reap the benefits I might never have experienced – in fact, it has only increased my ability to plumb the depths of those gifts. As I found healing and places of safety on my path to healing, I was able to begin again finding those child-like qualities that open up the creative flood-gates.

1.1.2 Playfulness

Instead of prescribing a concrete list of answers for those seeking this flow state (I don't think this is possible!), I will instead give the principles that help me find the flow state and some simple exercises I have found useful. You may find a very different path into that state and that will be just fine.

Quietness I find it hard to explain the power of quietness, at least in any deep way. And it is more than a little paradoxical to write and talk about quietness. But it is out of the stillness we find in the pursuit of depth and living inspiration, that our creativity, our fresh originality emerges. My early efforts to explain this to others dates back to my attempts to get my 7th and 8th graders to sit in quietness in nature, listening to what might speak to them. (I taught 7th and 8th grade science and history for one year when I was in graduate school). Eventually, I decided that simply telling the story of my own walkabouts, in person, one on one, was the most effective thing I could do in evangelizing for quietness. I also started recommending David Levy's Google Tech Talk from 2008 [23]. Quietness as a practice is closely related to the skill of deep reading, where one enters an immersive flow state focused on understanding and exploring the ideas in a text. There is a strong correlation between those that have developed this skill and those that master analysis/geometric analysis.

Warm up play Starting a day of mathematics by playing around with elementary questions is something I have found very helpful. It is related to the habit of finding the simplest examples of an idea and getting to know every aspect of those examples, from every angle, as a way to move into a state where the ideas are alive and flowing. I have, for example, frequently recommended Burn's book offering a problems based approach to elementary analysis, [10]

Take things apart When you are presented with an example or a theorem or a definition, look at all the pieces and how they relate to each other.

Build things from scratch After understanding the definitions involved, attempt to get (similar) theorems from scratch, making your own additional definitions as necessary.

Change things Perturb assumptions and see what you can prove. Find out if assumptions are necessary or just convenient. Try to come up with families of definitions/theorems by perturbing the definitions and assumptions.

Rewrite things Rewrite pieces of text that are not crystal clear, in order to put the ideas in your own frame of reference more solidly.

Teach things to others with pictures After you figure something out (new or old), explain it to others.

1.2 Building 1st and 2nd Order Intuitions

As your experience grows, your intuitions grow and in analysis, one of the first intuitions that you learn that turns out to be very, very useful is that of 1st order approximations, i.e. derivatives. That is, smooth things are locally linear. With this sense for local behavior, you can go a long ways. Much of Chapter 11 an immediate consequence of this appealing "first order" intuition.

What do I mean by "first order"? Because I assume you have had calculus, I can use the idea of a Taylor Series to explain the answer. If f is a function that has continuous first, second and third derivatives, we know that the following expansion is true:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + E\,(x - x_0)$$

where $|E((x - x_0))| \leqslant C|x - x_0|^3$ for some constant C.

Because we are evaluating the derivatives at x_0, and the only variable in sight that is not fixed is x, we can rewrite this as:

$$f(x) = a_0(x - x_0)^0 + a_1(x - x_0)^1 + a_2(x - x_0)^2 + E\,((x - x_0))$$

where the constants $\{a_0, a_1, a_2\} = \{f(x_0), f'(x_0), \frac{1}{2}f''(x_0)\}$ respectively.

Now we notice that for small enough $|x - x_0|$ the terms in the equation decrease as we move from the left to the right on the right hand side. That is, there is a $\delta > 0$ such that $|x - x_0| < \delta$ implies

$$a_0(x - x_0)^0 >> a_1(x - x_0)^1 >> a_2(x - x_0)^2 >> E\,((x - x_0))$$

and that this means the first term, which is a constant and denoted the **zeroth order term** (the power that the variable $x - x_0$ is raised to), is the most important value and is at least a roughly correct approximation to f close to x_0. The next term – the **first order term** $a_1(x - x_0)^1$ is the linear correction to the first order term that, when added to $a_0(x - x_0)^0$

gives a correction and results in an even more accurate approximation to f near x_0. Finally, if we add the quadratic term – the **second order term** – we get an approximation to f whose error ($E(x - x_0)$ that is cubic in $x - x_0$ and therefore goes to zero even more quickly as we approach x_0.

In higher dimensions, the notation is more complicated, but the idea is the same. (Except in higher dimensions, the topology of the graphs of the second order terms is a bit more complicated and this leads to pleasant things like Morse Theory.)

It turns out that knowing the zeroth and first order terms is enough to say a great deal about some situation, analytically, geometrically and if you also know the second order correction, you are often able to answer any question you are interested in.

While gaining an instinctive mastery of the use of first order approximations takes some work, pretty much everyone who works in analysis, and in fact anyone who uses it in science (so lots of non-mathematicians too), learns to wield first order approximations intuitively.

> *They have decent or good first order intuitions.*

Developing intuitions that instinctively wield second order corrections is more work, and many fewer do what it takes to really make these insight-generating calculations truly instinctive.

Exercise 1.2.1. play around with polynomials of degree 3 –i.e. $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ – using some tool like Matlab, octave, R or python in order to plot and see the effects of approximating f (all 4 terms) with only the first k terms, $k = 1, 2, 3$ (i.e. the zeroth, first and second order terms). Notice that the biggest effects happen when $a_1 = 0$ or $a_2 = 0$ while some higher order term is not zero. You might begin by, noticing that setting $a_0 = 0$ has no effect on the shape of the graph of f and start by looking at $f(x) = x(x^2 - 1)$

Now, if we are thinking of an object like a smooth 2-dimensional surface $S \subset \mathbb{R}^3$, we can always express it locally at p, perhaps after a rotation and translation in $\mathbb{R}^3$, as the graph of a function which has a horizontal tangent plane at p. We note that in these coordinates, the pesky need for the second order information is there because the tangent plane (the derivative approximation) is horizontal and we have no idea from looking at the derivative if the shape of the surface is like an upward pointing paraboloid, a downward pointing paraboloid or a saddle with p as the saddle point. Assuming the second derivative is nonsingular, we can immediately know the shape. (Of course, it is possible that all the derivatives up to some order k are zero or are singular in which case you have to have even higher order terms to get a grip on the shape of the object of interest. But this is a more advanced.)

The gist of this little story is that we always need first and second order information to understand shape – and sometimes even more. Thus, developing first **and** second order intuitions is a very good thing if you want to master geometric analysis.

I will close this section with a very short story.

At one point I was asking a mentor of mine, Bill Allard, what set apart his colleague Fred Almgren (both Fred and Bill were very famous for their work in geometric measure theory). Bill thought for a moment and responded, "Fred had a second order intuition".

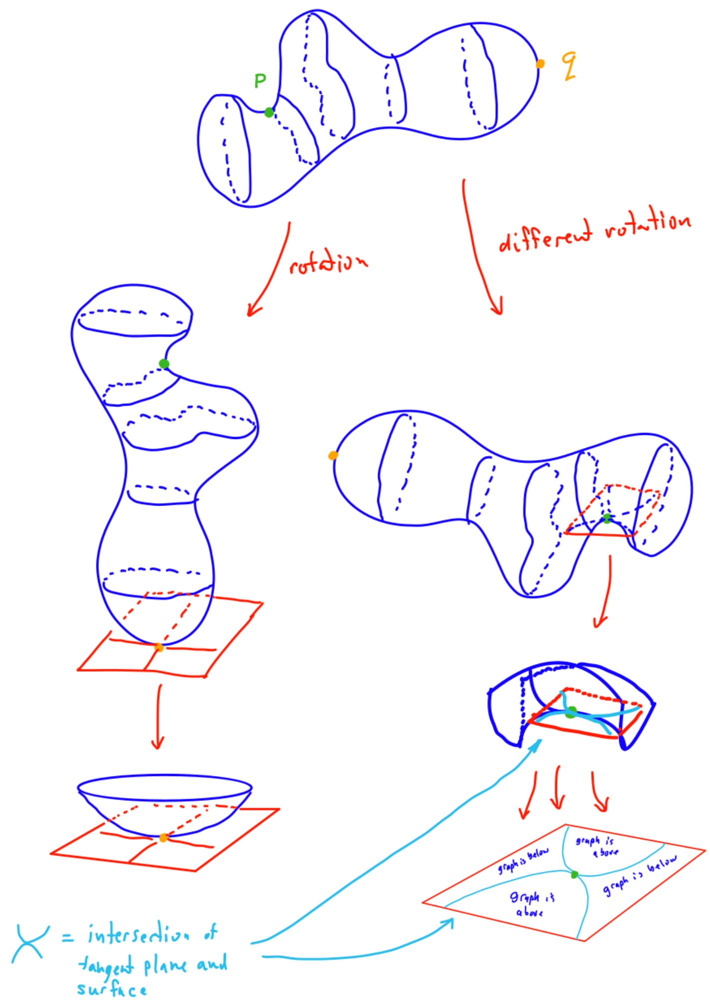


Figure 1: Examples of the rotations of a surface that (1) locally looks like a paraboloid at the point p and (2) a saddle at the point q.

1.3 Proofs

Written proofs play a crucially important role in mathematics. First, this activity helps you find subtleties you missed, errors you overlooked and second, the proof can easily be transmitted to others.

There are least three things crucially important to a well developed ability to create mathematical proofs – (1) the cultivation of child-like perspectives and habits, (2) the restoration of emotional health and (3) the development a deeply instinctive and enlightened intuition.

Why? Because proofs are first and foremost a (hopefully) clear vision of why something is true, first intuitively and then the refined (and often slightly, sometimes greatly, corrected) version that emerges by iteration ... and eventually gets written down. And in turn, this process depends on the creative freedom and deep insight that (1)-(3) open up and support.

Though I have less evidence for this, I believe that (1)-(3) also help you in translating what you see and imagine into writing that is clearer, even though, as noted in the first section, it seems this process very often falls short of capturing the true vision that was imagined.

2 Music as Poetry, Mathematics as Music

I will begin this very short chapter with a piece I wrote on my blog and then included in my book, Verses and Footnotes. The title of the piece was, *Speaking the Language that Cannot be Spoken*.

2.1 Speaking the Language that cannot be Spoken

Though I have gone further in my mathematical career, my past is also filled with music. Violin performance in multiple orchestras and chamber groups, together with several concerto competitions and the intense practice this required, had a large impact in my life. There is also the fact that my father was a musician, that our family was immersed in a musical environment.

I recently listened to a concerto while watching the notes scroll by. The complexity of the notes on the page were, somehow, not matched by the experience of music in that innermost place – that place from which you play when you perform. I found this experience very similar to the experience of creating new mathematics and then writing it down to communicate what I see.

Why?

Because I believe musical notes on paper and detailed proofs in a book or published paper are both misleading.

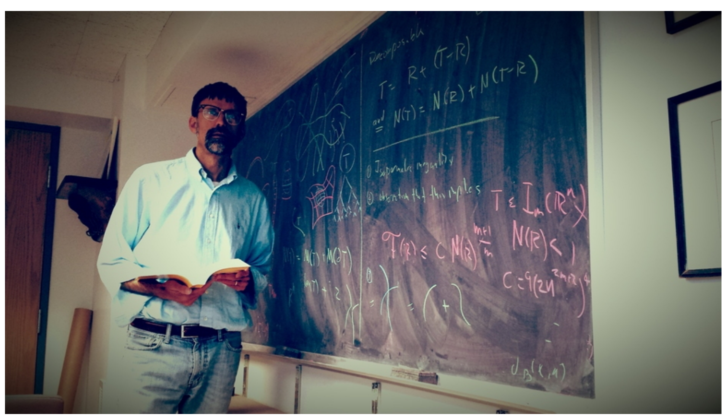
The music and the mathematics are, somehow, much simpler in their pure, newly created form. The complication evident in the written form comes from the unnatural way we have to communicate music and mathematics.

When I slowly recall or relive the creation of a proof, whose written form is non-trivial and may even seem imposing, I find the natural state of the proof in the imagination to be simpler, even minimalistic. Yet when written, expanding to something that looks imposing, it is often hard to read or imagine.

In the natural language of the soul, both mathematical proofs and musical compositions sing and flow. But in the language of things written down, we usually lose this living simplicity and beauty.

When we do harmonize the imagining and the telling, it is through the action of a whole person, in real time, speaking, writing, drawing, adapting, listening, responding ... finding the music that, for us, connects the inner and outer universe.

For it is the human being that integrates the universe, speaking the language that cannot be spoken.



2.2 Poetry, Music and Mathematics

Listening to pieces like Max Richter's *On the Nature of Daylight* or Medtner's *The March of Paladin* or the second movement of Shostakovich's Piano Concerto II, I am drawn into the poetry, the stillness that sings, the flow filled with light.

Surrendering to this instrumental music, I find myself in an altered state, a state that is somehow poetic, though without words. In that transformed state of flow, I am very close to the state of flow that comes when I immerse myself in a problem or deeper exposition of some part of mathematics. Perhaps this is not a common experience, but I suspect that it is at least accessible to anyone willing to invest the stillness, the quietness, the time.

Like the preceding chapter on playfulness and the intuition, this chapter is definitely not the usual fare for a book on analysis.

So why do I include it here?

This book is a door to the art of geometrically inclined analysis. Because the door leads to unfamiliar, very often challenging places, I believe the entire wholistic ecosystem that mathematics is a part of ought to be exploited in the pursuit of mastery of the art. I believe this ecosystem includes music and poetry. This poetry might include formal poetry, but for me at least, this refers to those organic, poetic expressions, ungoverned by strict rules of the more formal examples of written poetry.

As noted in the first section of this chapter, I was raised in a very musical household, exposed at an early age to a wide range of classical music, both recorded and live, at Orchestra Hall in Chicago. I began on the Piano, switching to the violin, eventually practicing and playing with an intensity that prepared me for deeper mathematics. For the quietness required to really hear music, to feel the poetry is the same quietness we need to really see and hear and feel the music and poetry in the mathematics. So, in some deeper sense, these three provide a sort of cross-training for each other.

Is there practical advice to be found in this chapter?

Yes. Seek music and (organically defined) poetry you can connect with. To do this, you must master the art of quietness, of hearing the stillness speak and sing. Unplug, begin walkabouts in nature, read books that move slowly, yet deeply and find that flow opening doors to the infinite everywhere around you.

You will find this strengthens the same faculties necessary to do deeper mathematics.

The closing piece of advice is, do not take the above advice something you **must do**, like a chore or calculations that are necessary but somehow rather dull or scales that you have to play when you are practicing an instrument. What I am encouraging you to do is fall in love with music that inspires stillness, with poetry that illuminates and heals. Then write your poetic reactions, fiddle around with a musical instrument – experiment, explore, improvise.

Find your place in the flow.

3 The Art of Mastery: The Three David's and One Daniel

This final chapter of the preamble is intended to help you understand a bit more about how environments for innovation and creativity are created and maintained. It is also intended to encourage you to reflect, to broaden your exposure to ideas quite far afield of mathematics. To do this I will recommend three books and one video I believe should be much better known by those in STEM.

The path to real mastery is a path of immersion; this is rarely taught anymore, inside or outside the mathematics classroom. The pernicious influence of social media is rewiring brains to think broadly but only at a very shallow, surface level. Students (and anyone else seduced by these devices of dopamine stimulation) are losing the ability to read deeply, to think deeply, to move patiently towards treasures far below the surface.

The good news is that the plasticity facilitating these negative effects also allows these effects to be completely reversed. It simply requires a decision to strictly limit your interaction with the culture and devices that are designed to exploit you, and a commitment to deeper things.

I am convinced that if more students paid close attention to what I am saying here, many more would master the course I teach in analysis based on this book.

I would encourage you to view this particular chapter as an invitation to new habits that will help you find mastery in analysis (and anything else you want to master).

3.1 Three Books and One Video

The Genius in All of Us, BY DAVID SHENK Shenk explores the implications of epi-genetics and the fact that we each have tremendous potential – Genius is largely learned, not some accident of "genetics". This is both good news – potential genius is as common as dirt – and a call to action – it is a call to finding your groove and then finding flow in that groove, putting in the time to develop your muse, your genius.

Range, BY DAVID EPSTEIN David explores the power of organic diversity. I say "organic diversity" to differentiate it from simply diversity which has become a code-word for an extremely narrow vision of diversity. i.e. racial diversity, gender diversity, etc. He makes a convincing case for the fact that doing lots of different things before you find the place you contribute deeply is not a distraction, but a feature of those personal paths that makes for much deeper and stronger contributions. I strongly recommend reading this.

No Time to Think, BY DAVID LEVY This is a Google Tech Talk from 2008 that you can find on YouTube. In the talk David makes a compelling case for disconnecting, slowing down, understanding the power of quietness and contemplation. I recommend this to all my students and anyone else who will listen.

The Culture Code, BY DANIEL COYLE I have built two industrial labs (think small Bell Labs, or Google Labs) based on the ideas I and my co-founders had learned first at places like Los Alamos National Laboratory and MIT and other places, but were first explained well in something we could hand to others in Daniel Coyle's delightful book. It would be worth more than the purchase price if you bought it solely for the enlightening anecdotes. It has my very highest recommendation. (Another data point here is that I thought so highly of the book that I bought literally dozens of the book and gave them out!)

In the next section, I look much more deeply at the book by Coyle. (This first appeared on my blog and then in my book *Verses and Footnotes*.)

3.2 Cultures of Creativity and Innovation

Books reliably inspiring enthusiastic conversations are books worthy of close attention. When Beata read and recommended Daniel Coyle's book, *The Culture Code*, encouraging me by reading bits and pieces of it to me, it was not long before I knew that I had been introduced to just such a book. Soon I was buying copies and giving them away. Over the course of 2-3 months I gave away a bunch of copies and organized an evening in the top floor of the Monarch Motel in Moscow, Idaho devoted to discussion of the book.

The present article is part of my evolving reaction to the stories and theories in Coyle's book, prompted by an immersive, barehanded engagement with the ideas.

The stories of remarkable environments for creativity and productivity, as well as the stories of studies and research aimed at understanding cultures of creativity and productivity, are brilliantly chosen. For this reason alone, I can, and do, recommend the book to everyone.

If those stories are listened to, and felt and thought about, and experimented with, the effect on the reader is large.

When I get a (non-mathematical) book and read it carefully, it means I have chosen to engage rather deeply. Usually I write in the margins, in a sort of hand to hand combat with the details and nuances.

There is a fair bit now written in the margins of this book.

While I sometimes have issues with the theories used to explain things - mostly nit-picky like the fact that nonlinearity does not equal non-logical, see the story of the Allen Curve - the quality of the inspiration affected by the book completely outweighs any concern about the book's shortcomings.

The core of this book is the threefold cord of (1) safety, (2) vulnerability and (3) purpose which, expanded a bit, becomes:

1 **Safety and belonging** - taxing existential questions are never the lot of individuals in highly creative, productive environments. The growing scarcity of safety and belonging in many workplaces should be a source of deep concern. The gig economy is an indication that we are eating our seed corn and have ceased to pay even lip service to wisdom and a sustainable future.
2 An empathetically evolved environment enabled by **vulnerability powered connection**. A status flat environment in which creative energy flows easily is an environment in which truth and kindness (together!) are common, even foundational. Empathy, in its nuanced and expanded incarnations, is at the root of all highly effective, sustainable environments.
3 **Purpose and vision** - a bold, omnipresent clarity on the deeper foundational laws of being as well as the aims, the goals and the lofty visions that drive everything. An environment filled with signals keeping these principles and visions in constant view, is an environment whose vision is sustainable. Opposing the natural trend towards higher organizational entropy, these signals are an energy that enables the culture to remain inspired and organized for innovation and collaboration.

These three threads are the pillars of environments that have no trouble retaining those entering their influence. We visit and never want to leave - quite literally. In fact, Daniel himself admitted that when he was doing the research for the book, he found himself making excuses why he needed to stay in the environments he was investigating, even after he had the information he needed for his book.

Of course, some of the research was historical, visible only through the stories of those who were lucky enough to be part of those past places. Take for instance, Bell Labs in its heyday and Harry Nyquist.

In trying to understand the smaller group of super-innovators at Bell Labs, every possible factor was eliminated until it was discovered that all of these super-innovators ate lunch with Harry. He would draw out and listen to his lunch-mates with interest and curiosity, quietly giving them inspiring ideas and questions to go away and think about. Though Harry was also well known and influential because of his own research and innovation, neither this fact, nor his ability to spark innovation in others, seemed to effect his gentle, fatherly demeanor or tranquil reliability. In fact, these characteristics seemed to be significant part of the reason for his power. Disarmed by his demeanor, they opened up to his relentless curiosity.

At IDEO, the design company responsible for a large number of design innovations, Roshi Givechi plays a similar role, roaming from one design group to another, helping them to overcome obstacles and find new creative grooves through a powerful ability to listen and ask questions. In fact, when Daniel Coyle told her the title of the book he was doing the research for, it was not long before he had a new subtitle after she asked a question about his choice of subtitle.

The other stories and anecdotes are very well selected and wide ranging. Some illustrate principles of collaboration. The Allen curve, showing that effectiveness of collaboration is inversely proportional to the distance between desks of those collaborating, is another striking story of discovery that is both surprising when you hear it for the first time and sensible, even intuitively reasonable, when you take it in and think about it for awhile. While it is not an illogical relationship, as Coyle asserts, it is a non-linear one that will nonetheless make sense to anyone whose intuitions include some instincts for physics and chemistry and interactions and reactions.

Other stories are rich with insight, a sort of living book waiting to be read more and more deeply. One of my favorites is the story Coyle starts the book with. In that story, kindergartners outdo, by a factor of two, groups of business students and professionals in a challenge to build the highest tower with a piece of tape, a string, a few dried spaghetti and a single marshmallow. The reasons for this difference, of course, motivates much of what follows and sets the tone for the rest of the book.

As noted above, I ended up with a book full of marginal notes (in pencil!) and a lot of thoughts that were discussed with others. If I had to select a phrase that captured the influence of the book on me, I think it would be:

> ... brilliantly selected stories and simple principles that were even more compelling because they were validated by my own experiences in trying to build highly effective teams of innovators ...

And the effect of the book does not end with the sharing of the book and discussions.

The histories of places like Bell Labs, Xerox PARC, Los Alamos and the Rad Labs in Boston were already part of my own context, either through direct experience or through careful histories I had read and internalized, but something about the combination of this book and my own struggles with getting groups together that were sometimes partly or mostly successful, and other times were pretty clear failures, created in me a deeper openness and readiness to put these principles into action.

Beata's discovery actually coincided with an invitation from a FinTech company to join them and up their algorithmic game. I did this by starting a research lab – think of a small, updated Bell Labs focused on financial algorithms and education, plus more. I was able to use Coyle's book as a tool for evangelization of the ideas I planned to implement in creating the culture of the Lab. The leadership of the company read and understood the ideas rather deeply. This led to the essentially blank check from, and personal involvement of, the CEO and founder of the company. While the full details of this story will not be told here (and, when the CEO moved on to his next startups, I also left the company), I can say that the discovery that Beata made and passed on to me, was remarkable.

I give the book my highest recommendation.

3.3 Down with Silos!

Sustainable creativity and invention depends on mastery of both (1) deep, solitary work characterized by immersive flow and (2) frequent collisions with others and the serendipity emerging from intentional design of the culture and spaces we work in.

The books and video I am highlighting in this chapter all relate to moving the reader towards a more wholistic, organically healthy immersion into life. Human thriving should always be the first goal, because only this can support the kind of environments that people never want to leave, if they are so lucky to find them. And that instinct – to stay, to create there, with others, is an important key to sustainability of any enterprise.

Though this flies in the face of the siloing tendencies that engulfed the world in the 20th century, this is only a good thing – those silos need to be overthrown.

If you develop a habit of reading, thinking, living as though those silos were not just silly, but very negative, you will find a richer life, one where disembodied expertise, disintegrated world views, and our culture's current dearth of wisdom are foreign to your experience and the experience of those in your circle of influence.

Part II

Preview

Analysis: A Motivating Exploration 4

Analysis is the careful, creative exploration of sets, functions, and measures. Nature and integrative, creative human intelligence collaborate to generate or discover useful tools and insights.

While mathematics suggested and inspired by nature is useful in understanding aspects of nature, it can by no means contain nature. This is a good thing, for it suggests that nature is an infinite source of ideas in mathematics, or at least I am convinced of that.

I think that it is often best to begin with the end, or goal, of a mission very clearly in view. We will therefore begin by looking at the sets, functions and measures we are interested in exploring and understanding.

Many of the figures in the book will be hand drawn, some free hand and others using xfig, a gnu tool. Ideally, the figures would be on a white board that I and a few students are gathered around, engaged in a lively back and forth conversation. While I realize that hand drawn figures are not to everyone's taste, I do believe that they lend a little bit of informality which in turn, will encourage some (hopefully quite a few!) to get into the bare-handed exploration mode as they interact with the book.

4.1 Examples of Sets and Functions

We begin with some examples: (1) and (2) illustrate graphs of functions from $\mathbb{R}$ to $\mathbb{R}$, (3) illustrates the boundary of a 3-dimensional set in $\mathbb{R}^3$, (4)-(7) illustrate subsets (possibly supports of measures) in $\mathbb{R}^2$. The

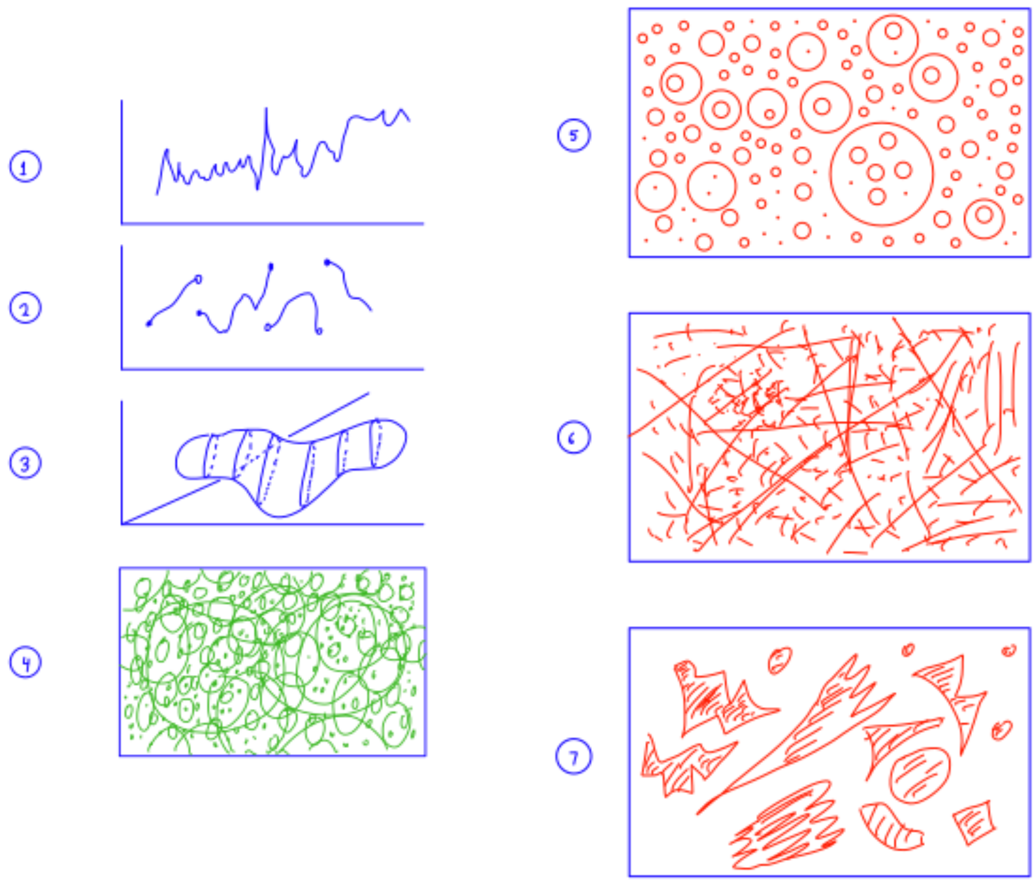


Figure 2: Examples of sets and functions in 2 and 3 dimensions.

questions that occur pretty naturally include:

How Big Is It?

1 How long are the sets shown in (1), (2), (4), (5), and (6)?
2 How much area does the set in (7) cover?
3 How much volume does the set in (3) enclose?
4 How long is the boundary of the set in (7)? (What exactly do we mean by boundary?)

Zoomed In, What Does It Look Like?

1 Are there tangent lines (i.e do derivatives exist) at every point of the graph shown in (1)?
2 What are the tangent cones of all the points in the set shown in (4)? (What do I mean by tangent cone?)
3 If the function shown in (1) has a derivative in the neighborhood of x, does that derivative have a derivative at x?

How close is one thing to another?

1 If we have a metric (for example, the Euclidean norm) in $\mathbb{R}^2$, can we use that to create a metric (measure of distance) between subsets of $\mathbb{R}^2$?
2 How many different (and useful) ways can we do this?
3 What are the implications of the choice of a particular metric, on the solutions we obtain?
4 For example, we might consider (5) to be a bunch of elements of the space of all circles in $\mathbb{R}^2$. Picking any two circles and defining a distance so that a distance of zero means that they are the same circle, which two circles would be the closest?
5 What would a convergent sequence of circles look like?

What does the space of all possible sets look like?

1 Is there a reason to restrict the sets we allow into our space of sets? (Hint: yes, we will want only those for which reasonable measures of length and area and volume can be constructed. It turns out this does not include all possible sets, but does include all the sets you might be able to imagine drawing.)
2 Is there a way to define spaces of sets that constrain the wildness enough to allow analysis, but not so much that we cannot model intricate, even crazy behavior?

3 For example, what sorts of spaces could we construct in which (4), (5), (6) and (7) are considered nice sets? Note that these sets are certainly not graphs of functions, even discontinuous ones. (Can you get these sets as images of discontinuous functions – I.e. = f([0,1]) for some mapping from $\mathbb{R}^1$ into $\mathbb{R}^2$?

This is pretty open ended and not every question will be very productive, at least in the short term, so let's look at 3 examples of the kinds of questions we will examine.

1 **How long is it?**
2 **How much does the area of a set E $\subset \mathbb{R}^2$ change when I map the set from one place to another?**
3 **How wiggly can the graph of a function be?**

4.2 How Long Is It?

We begin at the beginning, with subsets of the real line, $\mathbb{R}$.

4.2.1 An Interval in $\mathbb{R}$

If $E \equiv [a, b] \subset \mathbb{R}$ then we will agree that the length of E is just $\mathcal{L}^1(E) = b - a$. This certainly matches what many would identify as the length.

But it is hard to make a case against defining the length of E to be $\mathcal{E}_g \equiv g(b) - g(a)$, where g is some fixed strictly increasing or even simply non-decreasing function on the real line. In fact, it will turn out that this is a perfectly acceptable generalized length or *measure*.

For now, we will stick to lengths that are typically used, which correspond to the *1-dimensional Lebesgue measure* of sets in $\mathbb{R}^1$: $\mathcal{L}(E) = b - a$.

4.2.2 *An Interval in* $\mathbb{R}^2$

If we have a line segment in $\mathbb{R}^2$, which we denote by $[a, b]$ where $a, b \in \mathbb{R}^2$, we would expect that the length is given by*

$$|b - a| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

If we translate and rigidly rotate the coordinate system so that a and b are in the x-axis, with $b_1 \geqslant a_1$, we get

$$\begin{aligned} |b - a| &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \\ &= \sqrt{(b_1 - a_1)^2 + (0 - 0)^2} \\ &= b_1 - a_1 \end{aligned}$$

Exercise 4.2.1. Sketch examples of intervals in $\mathbb{R}^1$ and $\mathbb{R}^2$ and explain what you have just read to somebody who has forgotten all the mathematics they ever learned.

4.2.3 *Length of Graph of a Nice Function*

You might remember from calculus how to compute the length of a graph of a continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$. If so, at least the formula in Equation (1) will be familiar. First, suppose that

$$E \equiv \{\text{graph of } f \text{ from } x = a \text{ to } x = b\}$$

then

$$\text{length}(E) = \int_a^b \sqrt{1 + \left(\frac{df}{dx}\right)^2}\, dx \tag{1}$$

Even if you remembered this formula, perhaps you don't remember the reasoning behind it†.

*In this book, $|\cdot|$ will correspond to the usual euclidean norm in $\mathbb{R}^n$. Note that in $\mathbb{R}^1$ this is equivalent to the absolute value function.

†Actually, Equation 1 is true even if f is merely Lipschitz.

Exercise 4.2.2. Before reading the rest of this section, take a pencil and paper and think about why it might be true. Hint: first figure this out when $y = f(x) = sx + e$, the line with slope s and y-intercept e

Intuitive Proof. The intuitive proof, which essentially assumes that the function is differentiable everywhere and that this derivative is continuous – the derivative itself does not change too quickly, can be seen by observing that:

1 If we zoom way in, the little piece of the graph above the interval $[x, x + \Delta x]$ is the length of straight line from $(x, f(x))$ to $(x + \Delta x, f(x + \Delta x))$. But $f(x + \Delta x) = f(x) + \frac{df}{dx}(x)\,\Delta x$. So the length of that little piece of the graph is

$$\begin{aligned}
\text{Length of } f([x, \Delta x]) &= |(x, f(x)) - (x + \Delta x, f(x + \Delta x))| \\
&= |(x, f(x)) - (x + \Delta x, f(x) + \frac{df}{dx}(x)\,\Delta x)| \\
&= \sqrt{(\Delta x)^2 + (\frac{df}{dx}(x)\,\Delta x)^2} \\
&= \left(\sqrt{1 + (\frac{df}{dx}(x))^2}\right)\Delta x
\end{aligned}$$

2 This implies that if define $\Delta x = \frac{(b-a)}{M}$, we have a partitioning sequence of the interval $[a, b]$, $\{x_i\}_{i=0}^{M}$ given by $x_i = a + i\Delta x$, $i = 0, 1, 2, ...M$ and the length of the graph of f from a to b is:

$$\begin{aligned}
\text{length}(E) &= \lim_{M \to \infty} \sum_{i=0}^{M-1} \sqrt{1 + (\frac{df}{dx}(x_i))^2}\Delta x \\
&= \int_a^b \sqrt{1 + \left(\frac{df}{dx}\right)^2}\, dx
\end{aligned}$$

$\square$

To actually prove this result carefully, first we have to define what we mean by length of a curve in $\mathbb{R}^2$.

We suppose that a curve Γ is parameterized by some interval in $\mathbb{R}$. I.e. $\gamma : [\alpha, \beta] \to \Gamma \subset \mathbb{R}^2$. Next we take any ordered set of points $P \equiv \{y_i\}_{i=1}^{N_P} \subset [\alpha, \beta]$ such that $y_1 < y_2 < ... < y_{N_P}$. Define $\mathcal{P}$ to be the set of all such finite sequences of points in $[\alpha, \beta]$. Now we are ready to define the length of Γ. (See Figure (3) for a Sketch of the ideas.)

Definition 4.2.1 (Length of a curve in $\mathbb{R}^2$). *We define the length of a curve Γ by*

$$\text{Length}(\Gamma) = \sup_{P \in \mathcal{P}} \left(\text{Length}_P(\gamma)\right) \equiv \sup_{P \in \mathcal{P}} \sum_{i=1}^{N_P - 1} |\gamma(y_{i+1}) - \gamma(y_i)| \qquad (2)$$

Note that this definition does not depend on the differentiability of γ.

Exercise 4.2.3. Convince yourself that if P and Q are ordered sequences, then the ordered sequence $P \cup Q$ generates a length sum bigger than or equal to either P or Q.

Now the careful proof.

Careful Proof. We divide the proof into steps to make it easier to parse the ideas:

1. Since the derivative of f is continuous on the compact interval $[a, b]$, there is a δ_ϵ small enough that $|x - y| < \delta_\epsilon$ implies $|\frac{df}{dx}(x) - \frac{df}{dx}(y)| < \epsilon$.
2. Because $\frac{df}{dx}$ is continuous on the compact interval $[a, b]$, there is a bound on the magnitude of the derivative on the $[a, b]$: $\frac{df}{dx}(x) < K$ for all $x \in [a, b]$.
3. Choose M large enough that $h \equiv \frac{|b-a|}{M} < \delta_\epsilon$.
4. Define $x_i = a + i\,h$ for $i = 0, 1, 2, ..., M$ and note that $x_M = b$.
5. The mean value theorem that tells us for $x, y \in [a, b]$, $f(y) - f(x) = (x - y)\frac{df}{dx}(w)$ for some $x < w < y$

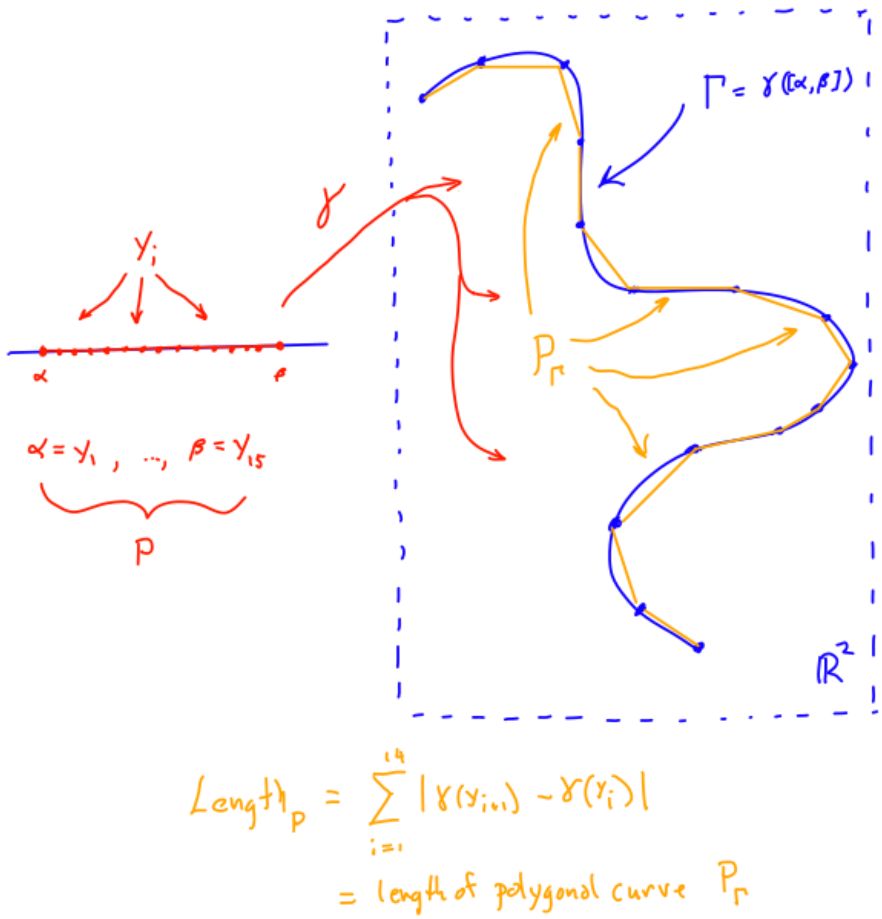


Figure 3: Sketch illustrating the idea of the curve length calculation.

6 We can therefore conclude that:

$$\begin{aligned}
|(x_i, f(x_i)) - (x_{i+1}, f(x_{i+1}))| &= \sqrt{h^2 + h^2\left(\frac{df}{dx}(w_i)\right)^2} \\
&= h\sqrt{1 + \left(\frac{df}{dx}(w_i)\right)^2}
\end{aligned}$$

for some $x_i < w_i < x_{i+1}$.

7 Notice that Steps (1) and (3) imply that as long as $|u - w| < h$, $|\frac{df}{dx}(u) - \frac{df}{dx}(w)| < \epsilon$. Assuming without losing any generality, that $|\frac{df}{dx}(u)| > F_1 \equiv |\frac{df}{dx}(w)|$,

$$\begin{aligned}
\left|\sqrt{1 + \left(\frac{df}{dx}(u)\right)^2} - \sqrt{1 + \left(\frac{df}{dx}(w)\right)^2}\right| &\leqslant \sqrt{1 + (F_1 + \epsilon)^2} - \sqrt{1 + F_1^2} \\
&= \sqrt{1 + F_1^2}\left(\sqrt{1 + \frac{2F_1 + \epsilon}{1 + F_1^2}\epsilon} - 1\right) \\
&\leqslant \sqrt{1 + F_1^2}\left(\frac{1}{\sqrt{1 + \frac{2F_1 + \epsilon}{1 + F_1^2}}}\right)\left(\frac{2F_1 + \epsilon}{1 + F_1^2}\epsilon\right) \\
&\leqslant \frac{2F_1 + \epsilon}{\sqrt{1 + F_1^2}}\epsilon \\
&\leqslant (2F_1 + \epsilon)\epsilon \\
&< (2K + \epsilon)\epsilon
\end{aligned}$$

8 Claim: given $h < \delta_\epsilon$, $P = \{a, a + h, a + 2h, ..., a + Mh = b\}$, and any other $Q \in \mathcal{P}$, we have that

$$\text{Length}_P(E) = \sum_{i=0}^{M-1} h\sqrt{1 + \left(\frac{df}{dx}(w_i)\right)^2},$$

$$|\text{Length}_P(E) - \text{Length}_{P \cup Q}(E)| < (b - a)(2K + \epsilon)\epsilon$$

and therefore, we can conclude that

$$\text{Length}(E) = \lim_{M\to\infty} \text{Length}_P(E) = \int_a^b \sqrt{1 + \left(\frac{df}{dx}(w)\right)^2}\, dw$$

The crucial insight we are using is the uniform continuity of the derivative of f, as noted in Step (1). The previous (step (7)) gives us the exact thing we need – when h is small enough the thing we are integrating ($\sqrt{1 + \left(\frac{df}{dx}(w)\right)^2}$) changes very little within intervals of length h.

9 We are done!

$\square$

Exercise 4.2.4. Fill in the rest of the details to establish the claim in the last step just above, Step (8).

4.2.4 *Length of Nice Curves in $\mathbb{R}^n$*

Suppose that $\gamma : [a, b] \to \Gamma = \gamma([a, b]) \subset \mathbb{R}^n$ is continuously differentiable. Then an argument very similar to the above argument – but now using the linear approximation definition of derivative, more specifically the "little $o(h)$" property of the approximation – we get that the length exists and equals:

$$\text{Length}(\gamma([a, b])) = \int_a^b |\dot{\gamma}(t)| \, dt \tag{3}$$

The ingredients in the proof are:

1 Uniform continuity of the derivative on $[a, b]$: given any $\epsilon > 0$, there exists a $\delta_1(\epsilon) > 0$ such that if $|s - t| < \delta_1(\epsilon)$ then $|\dot{\gamma}(t) - \dot{\gamma}(s)| < \epsilon$.

2 Uniform approximation of increments by the derivative: Given any $\epsilon > 0$, there exists a $\delta_2(\epsilon) > 0$ such that

$$|(\gamma(t + h) - \gamma(t)) - \dot{\gamma}(t)h| < \epsilon h \text{ when } h < \delta_2(\epsilon).$$

(This takes a bit to prove because this result is claiming that δ_2 does not depend on t.)

Exercise 4.2.5. Using the facts immediately above, see what you can do in outlining a proof of Equation (3)

4.2.5 From Lebesgue to Hausdorff: Measuring crazy lengths

For curve like sets in the plane that do not have nice parameterizations, γ, we resort to the generalization of Lebesgue measure – the measure you intuitively already understand in $\mathbb{R}$ – to 1-dimensional Hausdorff Measure. We will dwell on this much more in the text, but here is an explanation that transmits most of the ideas.

The basic idea behind 1-dimensional Hausdorff measure is to use covers of the set we want to measure and then simply add up the diameters of the cover to get an estimate of the 1-dimensional measure of the set. Figure (4) illustrates the basic step. To get the actual 1-dimensional measure of Γ, we have to take an infimum and then a limit (or equivalently, a supremum):

Definition 4.2.2 (1-Dimensional Hausdorff Measure)**.**

$$
\begin{aligned}
\mathcal{E}_\delta(\Gamma) &\equiv \{\textit{countable covers } E \textit{ of } \Gamma, \textit{ such that } \operatorname{diam} E_i < \delta \textit{ for all } E_i \in E\} \\
H^1_\delta(\Gamma) &\equiv \inf_{E \in \mathcal{E}_\delta(\Gamma)} \sum_{E_i \in \mathcal{E}} \operatorname{diam}(E_i) \\
&= \inf_{E \in \mathcal{E}_\delta(\Gamma)} H^1_E(\Gamma) \textit{ (See Figure (4))} \\
H^1(\Gamma) &\equiv \lim_{\delta \to 0} H^1_\delta(\Gamma) \\
&= \sup_\delta H^1_\delta(\Gamma)
\end{aligned}
$$

Exercise 4.2.6. Spend some time looking at this definition and the Figure 4, and verify that the basic idea stated just above Definition 4.2.2, is precisely what the definition is focusing on.

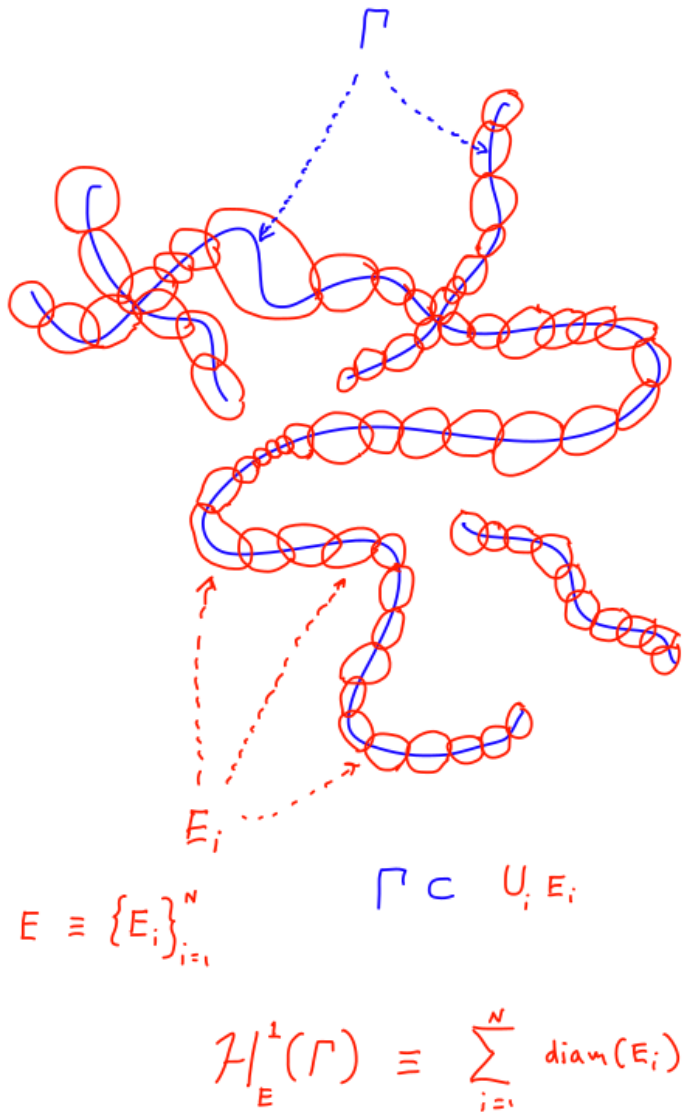


Figure 4: A sketch illustrating 1-dimensional Hausdorff Measure. See Definition (4.2.2) for the complete details, stated fairly concisely.

4.3 Mapping Areas?

For this question, we assume that the function or mapping F, maps $\mathbb{R}^2$ into $\mathbb{R}^n$ where $n \geqslant 2$: more succinctly, $F : \mathbb{R}^2 \to \mathbb{R}^n$. We will further restrict ourselves to functions having derivative matrices, DF, everywhere. This is not necessary, but for the idea in this initial exploration, this is as much generality as we need.

4.3.1 Tiny Preamble on Measure Theory

We will get into measure theory in later chapters, but for now, the following intuitive definitions should make sense from the discussion above, and suffice for understanding the rest of this chapter:

$\mathcal{L}^k$: Lebesgue measure formalizes what you already know intuitively in 1,2 and 3 dimensions and extends it to all positive integral dimensions. So $\mathcal{L}^1([a, b])$ is the length of the interval $[a, b]$ and equals $b - a$, $\mathcal{L}^2([a, b] \times [c, d])$ is the area of the rectangle $[a, b] \times [c, d] \subset \mathbb{R}^2$ and equals $(b - a)(d - c)$, and $\mathcal{L}^3([a, b] \times [c, d] \times [e, f])$ is the 3-volume of the rectangular parallelepiped (or simply 3-rectangle) $[a, b] \times [c, d] \times [e, f] \subset \mathbb{R}^3$ and equals $(b - a)(d - c)(f - e)$. You can see by analogy there is no reason to stop at 3 dimensions. To get the volume of things that are not rectangles, imagine tiling the set of interest with tiny rectangles and adding up the volumes.

$\mathcal{H}^k$: Hausdorff Measure is what we get when we generalize Lebesgue measure – the measure you already really understand intuitively from navigating the world from the time you are born until now – to the case when the set is not a piece of a k-dimensional linear subspace. You have already seen the definition in the case of 1-dimensional Hausdorff measure. We will dive into this in more detail later – and there are a lot more details – but you will not be misguided, for the purposes of the first chapter, if you simply think of cutting the set into pieces that are so small they are essentially pieces of $\mathbb{R}^k$, measuring them with $\mathcal{L}^k$ and summing. Of course, this is not sufficient in the long run, though it is actually essentially correct in the case that (1) the set you are measuring is a piece of a smooth, k-dimensional subset of $\mathbb{R}^n$ called a k-manifold, and (2) the limit of the above process of cutting and summing is taken as the size of the pieces goes to 0.

4.3.2 *When* $F: \mathbb{R}^2 \to \mathbb{R}^2$ *is a linear Map*

When F is linear – i.e. $F(\alpha x + \beta y) = \alpha F(x) + \beta F(y)$ for all scalars α and β and all vectors x and y – the area expands by the magnitude of the determinant of F:

$$\mathcal{L}^2(F(E)) = |\det(F)| \mathcal{L}^2(E)$$

In our case, F is a 2 by 2 matrix:

$$F = \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix}$$

Exercise 4.3.1. Show that F changes areas when it maps regions from $\mathbb{R}^2$ to $\mathbb{R}^2$ by a factor of $|\det(F)|$. **Hints:**

1 Convince yourself that the image of the unit square in $\mathbb{R}^2$, under F, is a parallelogram with sides equal to the vectors:

$$u = \begin{bmatrix} F_1 \\ F_3 \end{bmatrix} \quad v = \begin{bmatrix} F_2 \\ F_4 \end{bmatrix}$$

2 Show that $\det(F) = \det(R_\theta \cdot F) =$ where R is a rotation matrix given by

$$R_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

3 Use the fact that we can use a rotation matrix to rotate the image of the square – the parallelogram with sides u and v – so that u lies on the positive x axis. At this point, we can easily get the area of the parallelogram. It is simply the length of u times the absolute value of the y coordinate of the rotated v.

4 Compute this y value:
   a define $u^* \equiv (u_1^*, u_2^*) = (\frac{F_1}{|u|}, \frac{F_3}{|u|})$: the unit vector in the direction of u
   b Define define $u^p \equiv (-u_2^*, u_1^*) = (\frac{-F_3}{|u|}, \frac{F_1}{|u|})$: the unit vector orthogonal to the direction of u, rotated $\frac{\pi}{2}$ counterclockwise.
   c $|y|$ then is $|v \cdot u^p| = |\frac{-F_2F_3+F_4F_1}{|u|}|$
5 This implies that the area is $|y||u| = |F_4F_1 - F_2F_3|$ which is the absolute value of the determinant of F, as claimed.

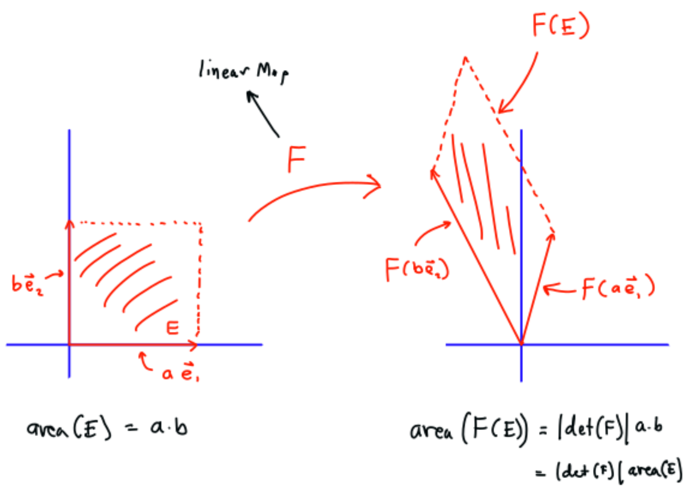


Figure 5: Illustration of the change in area under a Linear mapping.

4.3.3 *When F maps $\mathbb{R}^2$ to $\mathbb{R}^2$, one-to-one.*

When F is nonlinear, but it is also differentiable, we can cut up the set into tiny little bits that we map with the derivative since the derivative is the local linear approximation. Showing this in detail is a fair bit of work, but it will be easier to see and organize after you have a more instinctive feel for derivatives and their precise approximation properties. But the intuitively appealing statement made in the first sentence is correct and can be stated more precisely this way:

$$\int_E |\det(D_x F)| \, d\mathcal{L}^2(x) = \int_{f(E)} d\mathcal{L}^2(y)$$

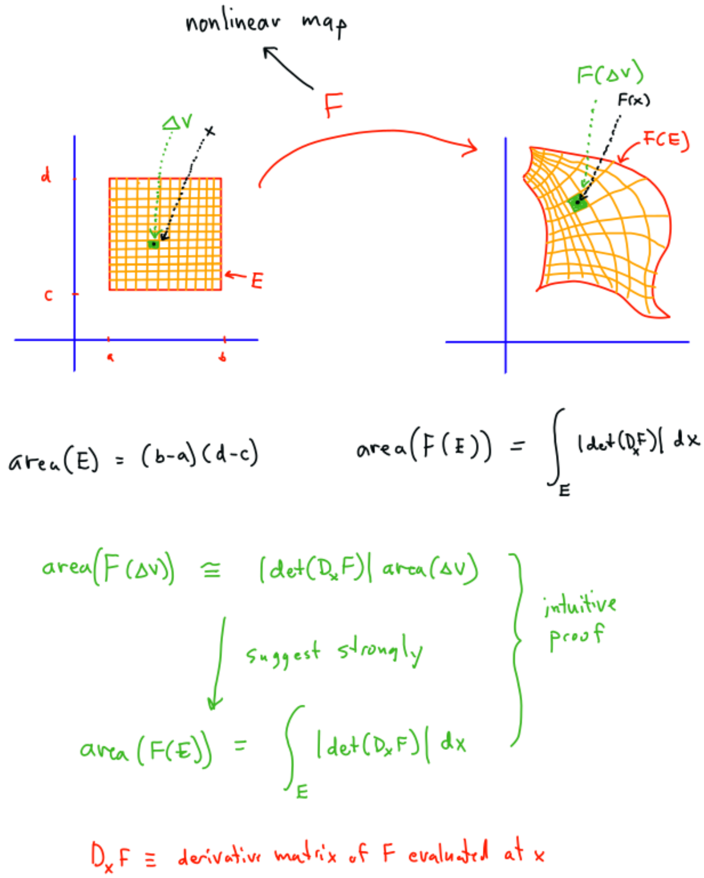


Figure 6: Illustration of change in area under nonlinear, differentiable, 1-to-1 mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$.

4.3.4 *When* F *maps* $\mathbb{R}^2$ *to* $\mathbb{R}^3$*, one-to-one.*

We come to the case in which we are mapping from a lower to a higher dimensional space. This makes things a bit more tricky, because we can no longer compute the determinant of DF. Instead, we have to define something we will call the Jacobian.

More precisely, define $JF = \sqrt{\det(DF^T \cdot DF)}$ where DF^T is the transpose or adjoint of DF. (Note: when DF is a square matrix, $JF = |\det(DF)|$, the absolute value of the determinant.) See Figure (7) for a sketch illustrating this.

Key idea: Infinitesimal squares map to infinitesimal parallelograms:

1 Decompose E into little squares lined up with the coordinate axes, with side lengths equal to some small h.
2 Let e_1 be the unit vector in the x_1 direction and e_2 be the unit vector in the x_2 direction. So the little squares have side vectors he_1 and he_2.
3 Define $hu_x \equiv D_xF(he_1) = h(D_xF(e_1))$ and $hv_x \equiv D_xF(he_2) = h(D_xF(e_2))$. (Note that u_x is the first column of D_xF and v_x is the second column of D_xF.)
4 So each tiny square in the set $E \subset \mathbb{R}^2$ is mapped to a little parallelogram in $\mathbb{R}^3$ and we know that the area of the parallelogram is given by the length of the cross product of the two vectors $hu_x, hv_x \in \mathbb{R}^3$ defining the parallelogram, $|hu_x \times hv_x|$.
5 Because F maps from $\mathbb{R}^2$ to $\mathbb{R}^3$, $h^2 JF = h^2|u_x \times v_x| = |hu_x \times hv_x|$. See the next exercise!
6 But

$$\begin{aligned} \sum_E |hu_x \times hv_x| &\to \int_E |u_x \times v_x| d\mathcal{L}^2(x) \\ &= \int_E JF \, d\mathcal{L}^2(x) \end{aligned}$$

This is the reasoning (with a little more care) that allows us to conclude that

$$\int_E JF \, d\mathcal{L}^2 = \int_{F(E)} d\mathcal{H}^2(y)$$

Exercise 4.3.2. Calculate to show that for if $u, v \in \mathbb{R}^3$ are the first and second columns of the 3 by 2 matrix A, we have $|u \times v| = \sqrt{\det(A^t \cdot A)}$

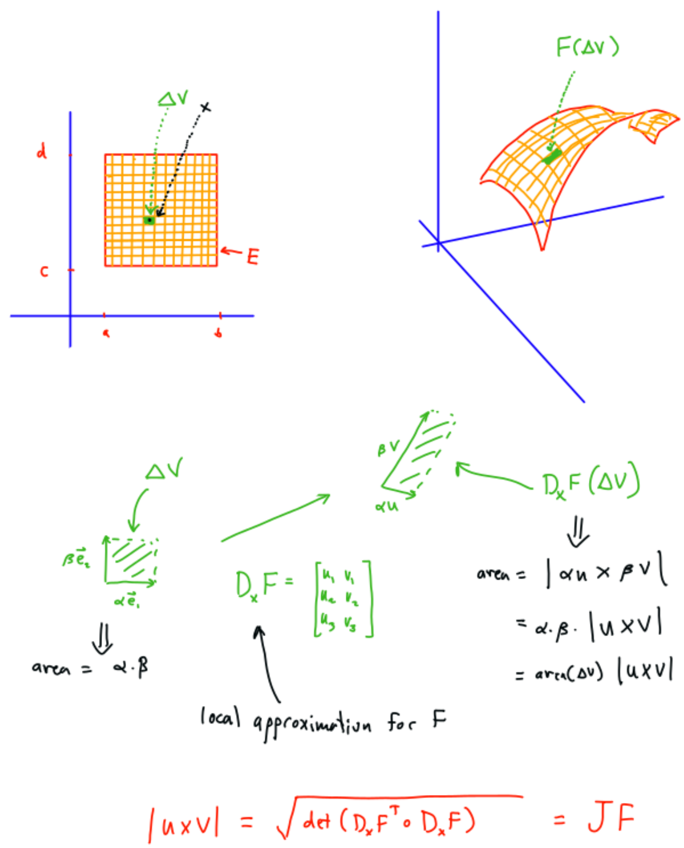


Figure 7: Illustration of the change in area under 1-to-1 mappings from $\mathbb{R}^2$ to $\mathbb{R}^3$.

4.3.5 *When* F *maps* $\mathbb{R}^2$ *to* $\mathbb{R}^3$*, but not one-to-one.*

These assumptions, in the title of the subsection, lead to what is known as the *area formula* which not only holds for differentiable functions, but also for the more general *Lipschitz functions* (Section 4.4). ‡

The new thing in this section is the fact we do not assume the mapping is 1-to-1. This means we have to think about the cases when things overlap. See Figure (8).

The result becomes:

$$\begin{aligned} \int_E JF \, d\mathcal{L}^2 &= \int_{F(E)} \left(\int_{F^{-1}(y)} d\mathcal{H}^0 \right) d\mathcal{H}^2(y) \\ &= \int_{F(E)} \mathcal{H}^0(F^{-1}(y)) \, d\mathcal{H}^2(y) \end{aligned}$$

where we note that $\mathcal{H}^0$ is just the counting measure, so that $\mathcal{H}^0(\Sigma)$ is just the number of points in Σ.

‡It also holds when the sets we are mapping between are subsets of general 2 and 3-dimensional *rectifiable sets* ... see Section 15.1.2.

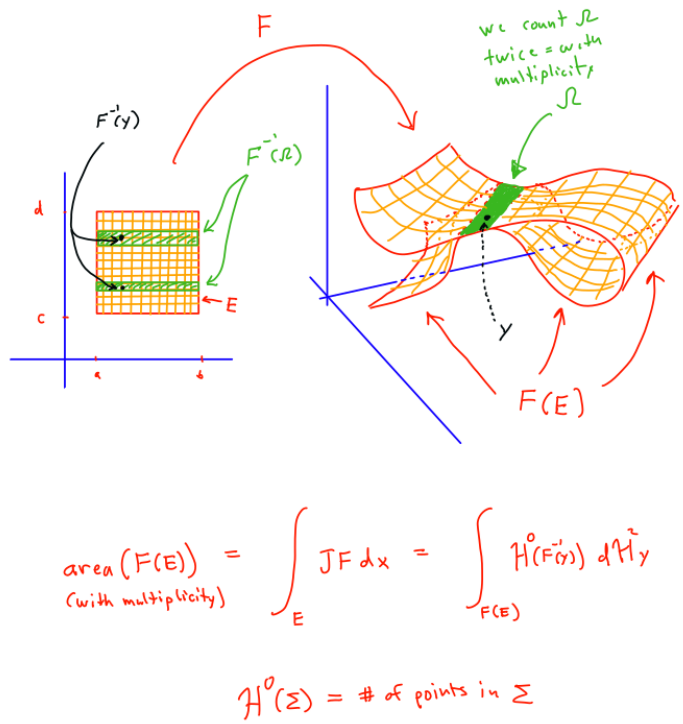


Figure 8: Illustration of the change in area under mapping from $\mathbb{R}^2$ to $\mathbb{R}^3$ that are not necessarily 1-to-1.

4.4 How Wiggly is the Graph of f?

In this question, we will focus on $f : [0,1] \subset \mathbb{R} \to \mathbb{R}$ that are also *Lipschitz*:

$$|f(x) - f(y)| \leqslant K|x - y| \text{ for some } K < \infty \text{ and all } x, y \in [0,1].$$

Such functions are general enough to include all functions which are continuously differentiable on $[0,1]$. But it also includes functions that don't even have a derivative $\frac{df}{dx}$ at lots of points in $[0,1]$. (Think about the graph of the function $y = f(x) = |x - \frac{1}{2}|$ to see that this function is (1) Lipschitz and (2) the derivative jumps from -1 to 1 at $x = 0.5$.)

What we will mean by *wiggly* will be how fast derivatives are changing, and more specifically, we will measure the wiggliness by *the size of the set* $\Sigma \subset [0,1]$ *where the derivative is discontinuous (where it jumps).*

4.4.1 *Constructing* f *so that* Σ *contains 2 points*

This task is very simple. Figure (9) shows an example function where Σ contains exactly two points:

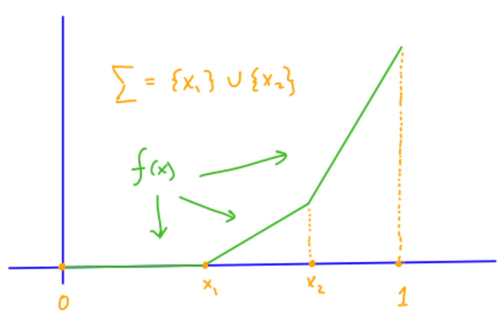


Figure 9: An example of a Lipschitz function with $\mathcal{H}^0(\Sigma) = 2$.

4.4.2 *Constructing* f *so that* Σ *contains an* ∞ *number of points*

This task is a little more involved, but not hard once you see what to do. We construct the function by first constructing a step function h which we will integrate to get f, building into f the discontinuities in the derivative at each step of h. As long as we control the sum of the heights of the steps, the derivative is bounded and the function remains Lipschitz.

In a bit more detail:

1 Define $h(x) = 1 - \frac{1}{2^{i-1}}$ for $x \in [1 - \frac{1}{2^{i-1}}, 1 - \frac{1}{2^i})$, for $i = 1, 2, ...$.
2 h(1) = 1. (But, because we are integrating to get f, this one point does not matter.)
3 Define $f(x) = \int_0^x h(y)dy$ where the function h was defined in the previous step.
4 This function f contains discontinuities in the derivative at every point of the form $x = 1 - \frac{1}{2^i}$ $i = 1, 2, 3, ...$.
5 Thus, we have a $\Sigma \subset [0, 1]$ with an infinite number of points in it.

See Figure (10) to see an illustration of h and f.

4.4.3 *Constructing* f *so that* Σ *is dense in* $[0, 1]$

In this example, we construct a function which has a dense set of points in $[0, 1]$ where the derivative is not defined because the derivative has a discontinuity there. Remarkably, the derivative exists and is continuous at every point not in Σ!

See Figure (11)

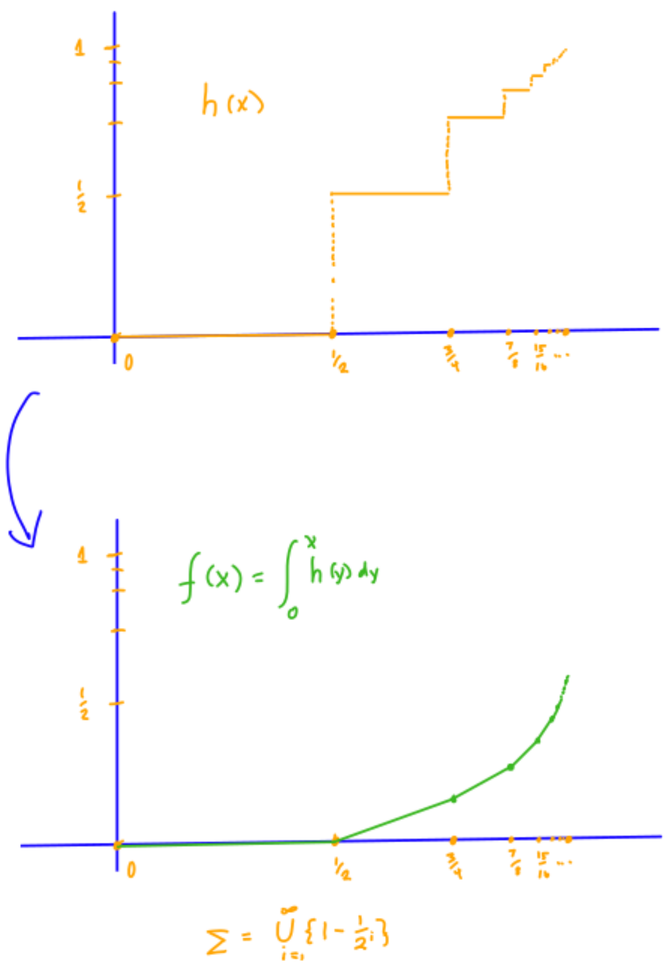


Figure 10: An example of a Lipschitz function with $\mathcal{H}^0(\Sigma) = \infty$.

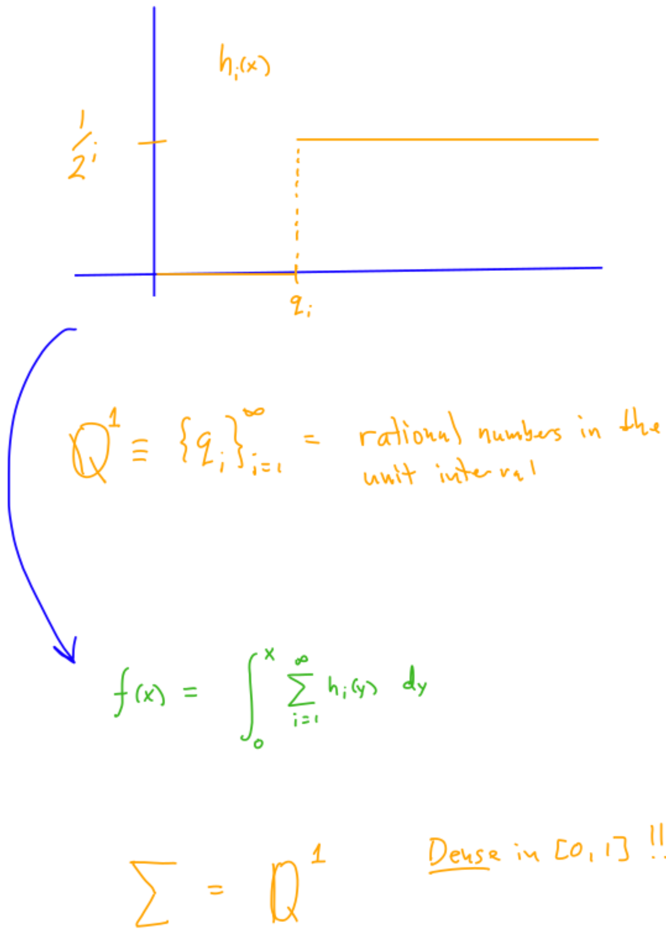


Figure 11: An example of a Lipschitz function with Σ dense in $[0, 1]$.

4.4.4 *Rademacher Implies Σ is not too big*

How big can the set Σ be – that is, how big can $\mathcal{H}^1(\Sigma)$ be? Since $\Sigma \subset [0,1]$ it is obvious that $\mathcal{H}^1(\Sigma) \leqslant 1$.

In fact, a famous theorem of Rademacher's – unsurprisingly known as Rademacher's theorem (See Theorem 15.1.1) – tells us that $\mathcal{H}^1(\Sigma) = 0$!

Exercise 4.4.1. Show that $\mathcal{H}^1(Q^1) = 0$. Hint: assume the following fact – if $E \subset \cup_i E_i$, then $\mathcal{H}^1(E) \leqslant \sum_i \mathcal{H}^1(E_i)$. Apply this to the case that $E = Q^1$ and $E_i = (q_i - \frac{\epsilon}{2^i}, q_i + \frac{\epsilon}{2^i})$. Notice we can choose ϵ as small as we like.

4.4.5 *Discussion of Wiggliness*

We have chosen a narrow definition of wiggliness in this section in order to ask more precise questions. There are of course many different measures of irregularity, oscillation, roughness or wildness. An example would be the integral of the absolute value of some derivative of a function. Or it might be the size of the minimal K in the definition of Lipschitz. Or the dimension of the set of points where the function is not differentiable or continuous.

This general area of study is called regularity theory and is often very challenging. A famous example is the paper that Fred Almgren wrote on the regularity of minimal surfaces of dimension k in R^{k+m} where $m \geqslant 2$. The paper was famously difficult and long. Very long. In fact for many years it circulated as a set of mimeographed notes about 1600 pages long! Later his widow (and first student), Jean Taylor, and his fourth student, Vladimir Scheffer, collaborated on creating a published book containing the paper. The result was (and is) 972 pages long.

We will return to the discussion of regularity from time to time, but this tiny taste does give a sense of the intricacy that can arise when studying *regularity*.

5 Analysis in a Nutshell: 20 Short Pieces

Mastery of anything involves practice founded on understanding which components are central, foundational and which components are derivative, produced by the action of those other fundamental, foundational components.

This leads to a minimalistic set of tools – the collection of a few things with which a master can either (1) solve any problem directly or (2) quickly craft the particular tool for the particular task at hand. And this minimalism strongly supports the experience of creative flow.

Though this succinct set of tools varies from master to master, each will invariably have a few key tools and insights they use over and over, more or less as extensions of themselves when they are in the state of flow.

The next section contains my own (current) list of 20 things: the details might escape the less experienced readers, but that is OK because those details are the goal of this course! The list is intended to give you an idea of the landscape the book will explore, as well as what the approximate minimal list I have looks like.

After these 20 pieces, I will list the ideas these pieces are founded on, in an attempt to reduce the ideas to the true core, the essence. This reduced list is always under development, in that my own understanding and perspective evolves and shifts over time.

Note: Because the geometric/analytic universe is big and impossible to capture in a linear story – first a, then b, then c, etc – the next level up in descriptional complexity is used. I see what I am transmitting here as the centers of an ϵ-net* of the space of insights and tools I have explored well, where ϵ is not too big.

5.1 20 Pieces

Convergence, continuity, connectedness, and compactness
... and the little bit of metric space theory needed for everything else:

$$\rho(x,z) \leqslant \rho(x,y) + \rho(y,z) \text{ where } \rho \text{ is a metric}$$

Understanding of finite dimensions, linearly
I.e. Linear maps and subspaces

$$Ax = b \text{ where } A : \mathbb{R}^n \to \mathbb{R}^m \text{ is a linear map,}$$

and we often think about A in terms of a matrix representation of A.
Derivatives as linear approximations

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

where $A(x) = D_x F$, the derivative at x, $o(h) =$ "little o of h" and means that $g(h) \equiv F(x+h) - F(x) - A(x)(h)$ satisfies $\frac{|g(h)|}{|h|} \to 0$ as $|h| \to 0$.
Inverse and implicit function theorems

$$D_x F \text{ continuous and invertible implies } F : \mathbb{R}^n \to \mathbb{R}^n \text{ invertible at } x$$

The implicit function theorem follows from the inverse function theorem.
Outer measures, in a nutshell
The outer measure theory approach is both very simple and very powerful.

$$\begin{aligned} \mu : 2^X &\to [0, \infty] \\ \mu(\emptyset) &= 0 \\ E \subset \cup_i E_i &\Rightarrow \mu(E) \leqslant \sum_i \mu(E_i) \\ \mu(F) &= \mu(F \cap E) + \mu(F \cap E^c) \;\; \forall\, F \in 2^X \Rightarrow \text{ E measurable} \end{aligned}$$

*An ϵ-net of a set, is a set of points, so that no point in the set is more than a distance ϵ from one of the points. Putting it another way, if that set is used as centers of closed balls of radius ϵ, the union of those balls will cover the set.

Rectifiable sets and the measures they support

$$g\mathcal{H}^k \llcorner W^k$$

where g is some density function, $\mathcal{H}^k$ is k-dimensional Hausdorff measure and W^k is a k-rectifiable set in $\mathbb{R}^n$, $k \leqslant n$.

Weak Convergence

$$\text{(functions)} \quad \int \phi_i \, d\mu \underset{i\to\infty}{\to} \int \tilde{\phi} \, d\mu \qquad (\text{written } \phi_i \rightharpoonup \tilde{\phi})$$

$$\text{(measures)} \quad \int \phi \, d\mu_i \underset{i\to\infty}{\to} \int \phi \, d\tilde{\mu} \qquad (\text{written } \mu_i \rightharpoonup \tilde{\mu})$$

This notion of convergence depends on integration, so it permits much wilder behavior to "converge".

Basic convergence results for integrals

$$\text{(Fatou's Lemma)} \int \liminf_k f_k \, d\mu \leqslant \liminf_k \int f_k \, d\mu$$

... and the other two theorems (bounded and monotone convergence theorems) telling us how sequences of functions and integration interact..

Holder's Inequality

$$\int |fg| \, d\mu \leqslant \left(\int |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int |g|^q \, d\mu \right)^{\frac{1}{q}}$$

which includes the famous Schwarz inequality as a special case, which in turn is an inner product on function spaces, leading to Hilbert spaces.

Covering Theorems

Theorem 5.1.1 (5R-covering). *Suppose that $\mathcal{B}$ is a collection of balls in $\mathbb{R}^n$ whose radii are uniformly bounded above. Then there exists a countable subcollection of pairwise disjoint balls, $\{B_i\}_i^N$, $N \leqslant \infty$ whose dilation by a factor of 5, $\{5B_i\}_i^N$ covers the union of the original collection of balls:*

$$\bigcup_{B\in\mathcal{B}} B \subset \bigcup_i 5B_i$$

This is an example of a very useful covering theorem. Another, much more complicated one (to prove) is the *Besicovitch Covering Theorem*.

Stokes Formula

$$\int_{\partial E} \omega = \int_E d\omega$$

Area/Coarea Formula for $F : E \subset M^m \to N^n$

$$\int_E g\ JF\ d\mathcal{H}^m = \int_{N^n} \left(\int_{F^{-1}(y)} g\ d\mathcal{H}^{\max(0,m-n)} \right) d\mathcal{H}^{\min(m,n)}$$

where M^m is an m-rectifiable set, N^n is an n-rectifiable set, and F is a Lipschitz mapping from M^m to N^n.

Gauss-Bonnet

$$\int_M \kappa(x)\ d\mathcal{H}^m(x) = \mathcal{H}^m(\partial B^{m+1}(0,1))$$

Where M is an m-manifold diffeomorphic to $\partial\left(B^{m+1}(0,1)\right)$, κ is the Gauss curvature on M and $\mathcal{H}^m$ is m-dimensional Hausdorff measure. This is an example of how integration and local properties interact.

Legendre-Fenchel transform

$$f^*(k) \equiv \sup_x \{\langle k, x\rangle - f(x)\}$$

where $x \in \mathbb{R}^n$, $k \in \{\text{dual space of } \mathbb{R}^n\}$ (row vectors), $f(x)$ is the function $f : \mathbb{R}^n \to \mathbb{R}$ we are computing the Legendre-Fenchel transform of, and $f^*(k)$ is the Legendre-Fenchel transform of f.

Fourier transform

$$\hat{f}(k) = \int_{\mathbb{R}^n} f(x) e^{-i2\pi kx}\ dx \quad \text{(Fourier transform)}$$

$$f(x) = \int_{(\mathbb{R}^n)^*} \hat{f}(x) e^{i2\pi kx}\ dk \quad \text{(inverse Fourier transform)}$$

The areas of pure and applied harmonic analysis flow from this transform. These areas, in turn, have had a huge impact in the pure and applied sciences. ($(\mathbb{R}^n)^*$ is the dual space to $\mathbb{R}^n$ which we can usually just think of as $\mathbb{R}^n$ because a Riesz-type representation theorem tells us they are essentially identical. See Theorem (5.1.2) below.)

Banach and Brouwer Fixed Point Theorems

Assuming $F : X \to X$ and X is a Banach space, we have

$$|F(x) - F(y)| \leqslant k|x-y| \text{ with } k < 1 \quad \Rightarrow \quad \exists! x \text{ such that } F(x) = x$$

where $\exists! x$ should be read "there exists a unique x".

The *Brouwer Fixed Point Theorem* says the closed unit ball in $\mathbb{R}^n$, mapped continuously to itself – $F : \bar{B}^n(0,1) \to \bar{B}^n(0,1)$ – must contain a fixed point x: again, $\exists x$ such that $F(x) = x$. In this case though, the fixed point might not be unique.

Linear Flows in $\mathbb{R}^n$

$$\dot{x} = Ax \quad \Rightarrow \quad x(t) = e^{At}x_0$$

where $x \in R^n$, A is an n by n matrix and $x_0 = x(0)$

Isoperimetric inequality

$$\mathcal{H}^n(E) \leqslant C(n) \left(\mathcal{H}^{n-1}(\partial E)\right)^{\frac{n}{n-1}}$$

where $E \subset \mathbb{R}^n$. There are many variations of this inequality that come up over and over in analysis.

Weak Differentiation

A function g such that:

$$\int \phi g \, dx = - \int \frac{\partial \phi}{\partial x} f \, dx$$

holds for all smooth, compactly supported ϕ is called the *weak derivative of* f.

This theme of generalization takes in a great deal of territory. This particular case – generalizing derivatives using the product rule – is just one example of many. Sobolev spaces – spaces of functions having weak derivatives in L^p – are used heavily in the analysis of partial differential equations.

Riesz representation

There are many different analogs of the following theorem, varying as the linear spaces of interest vary.

Theorem 5.1.2 (Reisz Representation Theorem)**.** *Suppose we define* $H =$ *set of all functions* $f : \mathbb{R} \to \mathbb{R}$ *such that* $\int_{\mathbb{R}} |f|^2 \, dx < \infty$. *Define* $||f|| = \left(\int_{\mathbb{R}} |f|^2 \, dx\right)^{\frac{1}{2}}$. *Suppose that* $F : H \to \mathbb{R}$ *is linear and* $\sup_{f:||f||\leqslant 1} F(f) < \infty$. *Then:*

$$\exists g_F \in H \textit{ such that } F(f) = \int g_F f \, dx$$

5.2 The Essence

A roughly multiscale understanding of geometric analysis and its applications makes sense due to the structure of the area. Geometric analysis is not a linearly organized thing, but rather a complex system of ideas, needing at least a graph or network structure, but more likely a hypergraph structure to represent the relationships and dependencies in the subject. The above 20 items follows roughly from the assumption of a graph like structure. As mentioned above, they form my current ϵ-net for a not too big ϵ.

What follows here is an even simpler, minimalist synopsis. While it is a bit harder to grasp – it is really a cryptic synopsis for those with more experience – it is still useful before you reach mastery, as a map to the place you are aiming for.

5.2.1 The synopsis

> **Geometric analysis: the nuanced study of sets, mappings and measures involving linear and non-linear spaces.**

The Three

1 **Linear Spaces, Subspaces and Mappings**: mastery of this foundation for analysis precedes mastery of geometric analysis

2 **Derivatives**: linear approximation and a great deal more ... extremely rich set of ideas and tools.
3 **Measures**: sets, mapping, coverings, interaction with spaces and function spaces through integration.

Three More

1 **Approximation**: understanding crazy things by sequences of nice things, built on a minimalist set of tools from metric spaces.
2 **Transformation**: Fourier transform, Legendre-Fenchel transform, local transformations of all sorts
3 **Topology and integration**: invariants and global/local connections.

Why Nuance?

1 **Linear and nonlinear, linear vs nonlinear**: nice nonlinear = locally linear, not nice nonlinear = singular = not locally linear.
2 **Finite dimensional vs infinite dimensional**: investigations into what stays the same and what changes when you move to infinite dimensions, opens a rich and fascinating wilderness.

Concrete Examples and Problems!

1 **Barehanded study of crazy sets, functions, measures**: putting what you know, right now, at this moment, to work is crucial to the development of mastery. This can take many different forms, from something that is very applied to something that is very esoteric. But work on very concrete questions of one sort or another is a central occupation of those working for mastery.
2 **Immersion in applications, at least some of the time**: even if you are not interested in applications like physics or economics or chemistry or biology, grounded contact with applications in those areas is invaluable as a source of completely new and unexpected ideas for even very pure, esoteric development.

6 Some Other Things: Samples from Geometric Analysis

Though to some, geometric analysis is the rather narrow area of analysis on manifolds, I take the other viewpoint that it is anything that lives at the extraordinarily rich, even infinite, interface between geometry and analysis, often inspired by broadly ranging applications.

In this short chapter, I will sample this rather vast region of the mathematical universe – sometimes mentioning little more than the name of the area or set of problems studied in the area. The purpose of this list is to give a sense for the expanse and to stimulate curiosity how the comparatively limited list of ideas in the previous chapter can really form the foundation for the prodigious range of ideas, tools, and problems.

6.1 Samples

Stochastic Geometry Suppose you have some fixed figure – perhaps a circle of fixed radius – that you randomly drop on a grid of parallel lines separated by twice the diameter of the object. What is the probability that the object will intersect one of the lines? What is the expected number of intersections if you average over all possible random drops? This is a simple example of the type of problem studied in stochastic geometry. The subject gets very involved, complicated and is therefore quite fascinating.

Analysis on Graphs/Networks Instead of having the sets, functions and measures defined in or on R^n, we look at the case in which the set of points are vertices of a graph with a prescribed set of edges and edge weights. This is a rich arena for development and much newer than the classical analog on $\mathbb{R}^n$. Examples would include dif-

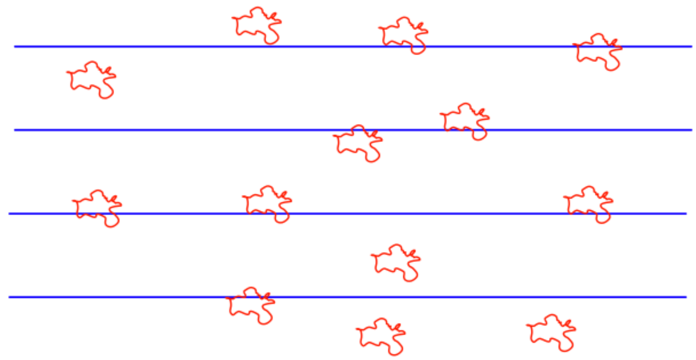


Figure 12: Random intersections with a horizontal grid, showing multiple throws without rotation.

fusion on graphs, which has actually been around for awhile for the very special case of graphs that arise from the discretization of $\mathbb{R}^2$ and $\mathbb{R}^3$.

Pseudo-differential operators in $\mathbb{R}^n$ Let us recast the differential equation

$$\frac{d^2}{dx^2}f(x) - 3\frac{d}{dx}f(x) + 4f(x) = g(x)$$

as an operator equation:

$$\mathcal{D}f = g$$

where we define the differential operator $\mathcal{D}$ by

$$\mathcal{D} \equiv \frac{d^2}{dx^2} - 3\frac{d}{dx} + 4.$$

Now, suppose we take the Fourier transform. We denote the Fourier transform of a function $h(x)$ by $\hat{h}(k)$. We find that the original differential equation transforms into:

$$(i2\pi)^2k^2\hat{f}(k) - 3(i2\pi)k\hat{f}(k) + 4\hat{f}(k) = \hat{g}(k)$$

and we see that, remarkably, the *differential operator in* x has been transformed into *multiplication by a polynomial in* k *in Fourier space*! So we have:

$$\mathcal{D}f = g \quad \Rightarrow \quad (-4\pi^2k^2 - 6i\pi k + 4)\hat{f}(k) = \hat{g}(k)$$

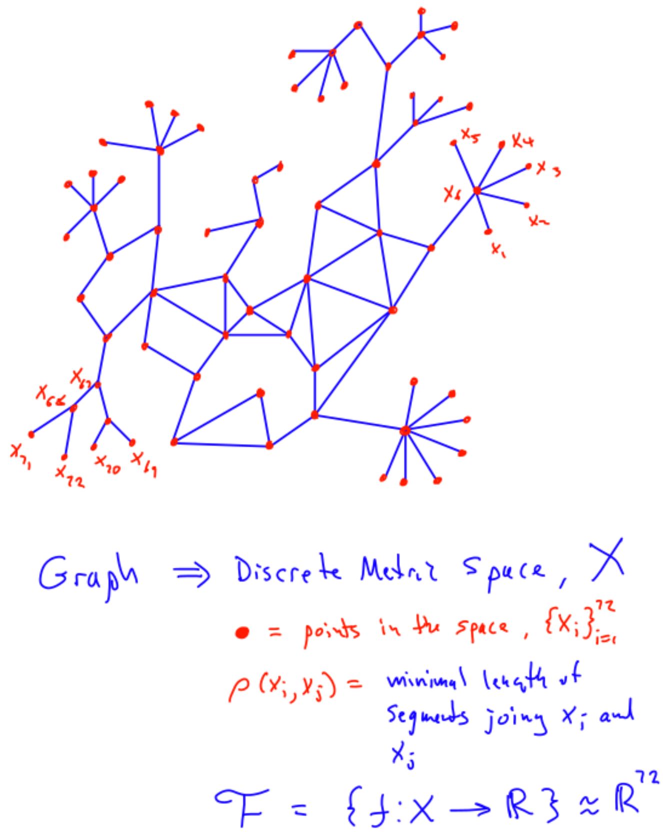


Figure 13: Suggestive cartoon of a graph in $\mathbb{R}^2$, the metric space it generates and resulting function space. This works whether or not the graph can be embedded in a vector space.

While this opens up immediate ideas for solving differential equations, it also generalizes the space of differential operators. Noticing that taking m derivatives in space is the same as multiplying the Fourier transform by $(i2\pi k)^m$, we can ask what differential operator corresponds to multiplication, in Fourier space, by $(i2\pi k)^{1.5}$? The answer is, the pseudo-differential operator $\frac{d^{1.5}}{dx^{1.5}}$. Why this is cool, other than the obvious coolness of generalization, is a longer story.

Vector Fields and Flows on $\mathbb{R}^n$ You may have been introduced to the dynamical systems point of view when you took the first course on differential equations. The basic idea is that at every point $x \in \mathbb{R}^n$, you are given a vector field $\mathbf{v}(x)$. The solution to such a vector field is a family of curves that are everywhere tangent to those vectors. In fact, there are time parameterizations on those curves, γ, so that $\dot{\gamma}(t) = \mathbf{v}(\gamma(t))$. Often, we put all these curve together to get a time dependent diffeomorphism of $\phi(\cdot, t) : \mathbb{R}^n \to \mathbb{R}^n$, called the flow generated by the vector field. This is a very large, very active area of research that is such because so many real phenomena lend themselves to being modeled by dynamical systems of one sort of another. For beginners, I recommend Strogatz [40]. The next step up includes many books, but James Meiss' book, *Differential Dynamical Systems* is a very good next step up, [30].

Ergodic Theory on $\mathbb{R}^n$ (or n-manifolds) Recalling your exposure to vector fields in differential equations (or the previous paragraph), imagine the case in which the flow that gets generated, $\phi(x, t)$, is confined to some bounded region in space, M. Or perhaps, the vector field is on a compact manifold, M, like the unit n-sphere, or a torus, or some bounded energy surface in a conservative mechanical system. Suppose $x \in M$ represents the state of the system and suppose there is some quantity f that you can observe in the state space. When is this asymptotic time average of the observable f

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\phi(x_0, t))dt$$

equal to space average

$$\int_M f(x) d\mu_\phi$$

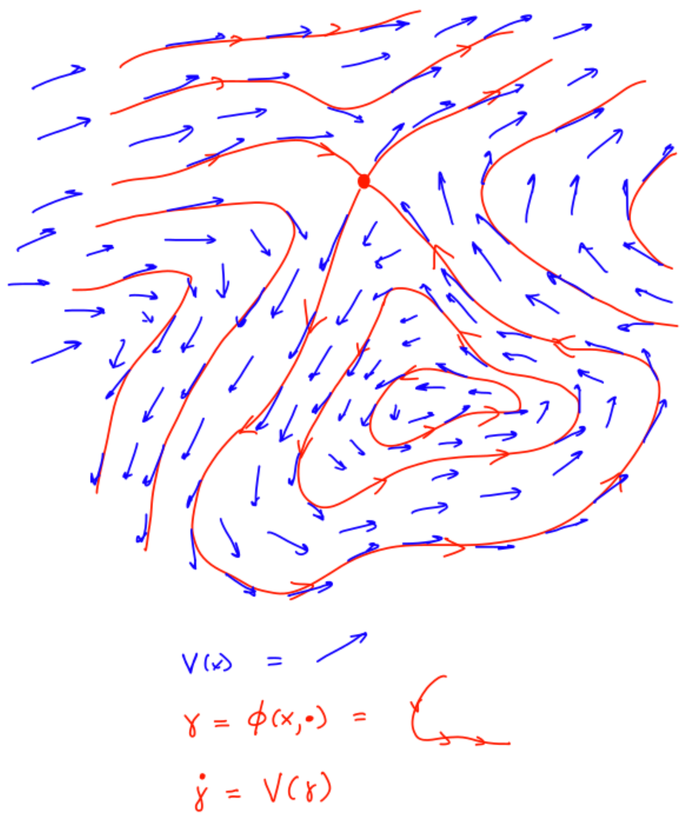


Figure 14: Vector fields and the flows they generate: an example in $\mathbb{R}^2$.

for some flow invariant measure* μ_ϕ? This is a rather old, physics inspired question that gave rise to an entire subfield of analysis, called ergodic theory. It is very geometric, filled with interesting ideas.

Wavelets Inspired by Fourier analysis, other families of orthogonal functions – but now with bounded sets on which they are non-zero – have been invented for the purposes of efficient representation of functions and processes. This entire area is a huge industry actually, driven forward by the large number of practical applications in image and signal processing and the fact that both the pure and the applied aspects of wavelet analysis yeild fascinating questions.

The initial challenge was to invent these things called wavelets which were smooth **and** compactly supported†. One wanted a way to have descriptions of functions which were bounded in both the spatial domain and the frequency domain. In the well known Fourier case, this is impossible. In fact, the precise statement of this is equivalent to the uncertainty principle in quantum mechanics – the smaller the support of the spatial representation, the bigger the support of the frequency representation and vice versa – the smaller the support of the frequency representation, the bigger the support of the spatial representation.

Partial Differential Equations on Manifolds PDEs are a truly vast area of analysis, inspired and driven forward by problems in physics, engineering, chemistry, biology and just about every other science. Deriving their name from the fact that these equations involve partial derivatives of functions with at least two independent variables, the resulting study has generated an enormous amount of creativity and computation. Even the linear equations like

$$u_{tt} = c(u_{xx} + u_{yy}) \qquad \text{(Wave Equation in } \mathbb{R}^2\text{)}$$

*To explain flow invariant, define $F_t(x) = \phi(x, t)$ – the map defined on M for a fixed t. We say that ν is invariant under the flow if $\nu(F_t^{-1}(E)) = \nu(E)$ for all E and all t.

†The support of a function is the closure of the points where the function is nonzero. Compactly supported means that this support set is bounded.

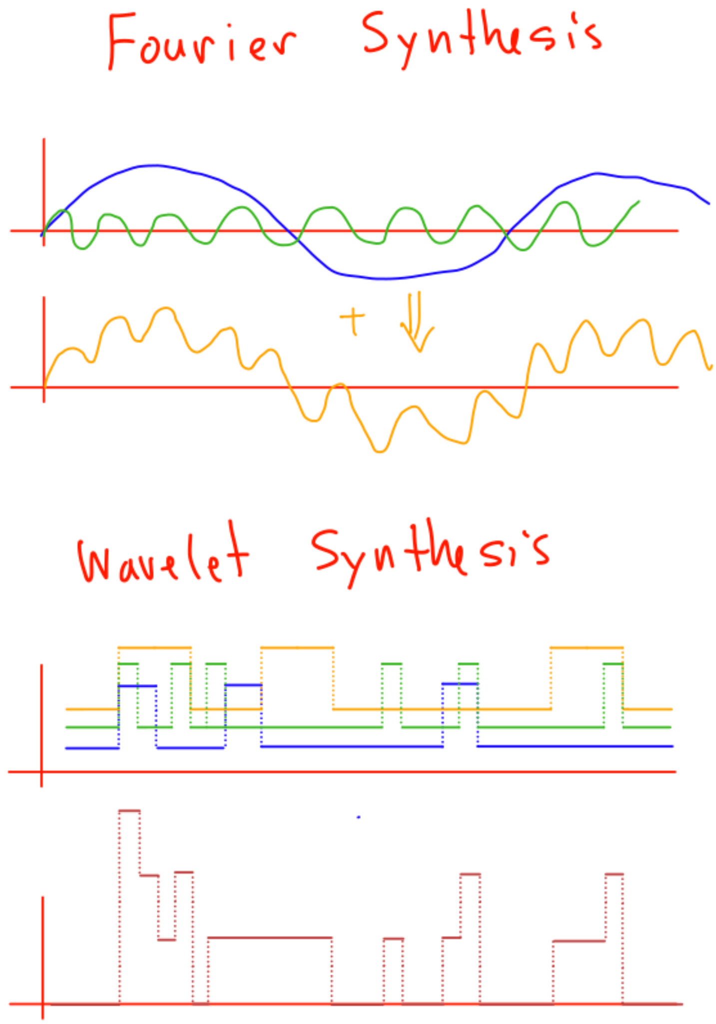


Figure 15: Fourier (with its sinusoidal basis functions) versus Wavelet (with its compactly supported basis functions). In this case I am using the Haar wavelets with are not smooth and have been known for a long time. They are nevertheless quite useful for some signal and image tasks.

and

$$u_t = c(u_{xx} + u_{yy}) \qquad \text{(Diffusion (or Heat) Equation in } \mathbb{R}^2\text{)}$$

are interesting because of the dependence of their solutions on the shape of the region in which we are solving the equations. After completing the present course (or another like it), I cannot recommend a better book that Craig Evans' *Partial Differential Equations* published by the AMS. PDEs on manifolds – what some people call geometric analysis – is all of the usual PDEs, but now on manifolds. The properties of the solutions are enriched by the interaction of the geometry and global topology of the manifold and local properties of the differential operator.

Properties of the Navier-Stokes PDEs in $\mathbb{R}^3$ Here I single out one system of PDEs – the Navier Stokes equations modeling fluid flow in $\mathbb{R}^3$ – because this single system of PDEs has generated an enormous amount of work in analysis. In fact settling the properties of this equation is one of the Clay Mathematical Institute prize problems. The equations are really just conservation of material (mass) and conservation of momentum – $F = ma$ – translated into fluid terms:

$$\begin{aligned} \frac{\partial}{\partial t}(\rho) + \nabla \cdot (\rho \mathbf{u}) &= 0 \\ \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} &= -\frac{1}{\rho}\nabla p + \nu \Delta \mathbf{u} + \mathbf{f}(\mathbf{u}, t) \end{aligned}$$

Harmonic Measures and their structure Returning to a simple, linear PDE given by

$$\nabla \cdot \nabla \phi = 0 \quad \text{(also written } \Delta\phi = 0\text{)}$$

we note that this is incomplete as an equation until we specify (1) where we want to solve this problem and (2) what the boundary conditions are:

$$\begin{aligned} \nabla \cdot \nabla \phi &= 0 \quad (\text{ for all } x \in E) \\ \phi(x) &= f(x) \quad (\text{ for all } x \in \partial E) \end{aligned}$$

It turns out that the values on the boundary, given by f, determine the solution ϕ uniquely. This gets interesting when we realize that for each point $x \in E$, there is a probability measure $\mu_{x,E}$ such that

$$\phi(x) = \int_{\partial E} f(y)\, d\mu_{x,E}(y).$$

We notice that $\mu_{x,E}$ does **not** depend on f. The properties of $\mu_{x,E}$ and their dependence on E is a large, fascinating area of study. Example of a cool property: if you set a particle free to take a random walk at x, then the probability of it hitting $P \subset \partial E$ first is exactly $\mu_{x,E}(P)$! I can recommend no better book than *Harmonic Measure* by John Garnett and Don Marshall.

Concentration in High Dimensions Concentration of measure phenomena is part of the strange and wonderful properties of probabilities and measures in very high dimensions. Here are a couple of interesting, initially surprising facts:

- A Lipschitz function on the unit Sphere in $\mathbb{R}^n$, for n very large (for example n = 10^6, 10^9, or 10^{12}) is almost constant – that is, if you choose a point at random, the probability that it will be very close in value to the median value of the function, $\hat{f}$, is very, very high:

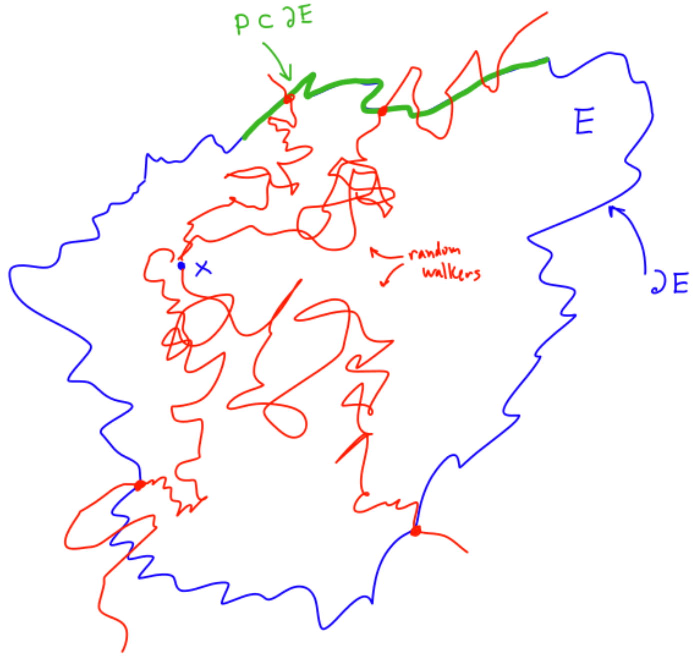
  $$P(\{x : |f(x) - \hat{f}| > \epsilon\}) < \delta$$

  where, as $n \to \infty$ we can make both ϵ and δ go to zero, together.
- As the dimension diverges, the probability of a strip of the unit sphere within epsilon of any given great circle, C, goes to 1:

  $$P(\{x : d(x, C) < \epsilon\}) \to 1$$

  we note that the probability does not depend on which great circle we pick.

The phenomena of concentration of measure and probability in high dimensions show up all over the place in statistical learning theory (the foundation for, and mathematical core of, machine learning) and the growing industry created by sparse representation and sparse inverse problems.



$$\mathrm{Prob}(\{\text{random walker hits } P \text{ first}\}) = \mu_{x,E}(P)$$

Figure 16: Harmonic measures are a rich source of questions for geometric analysts!

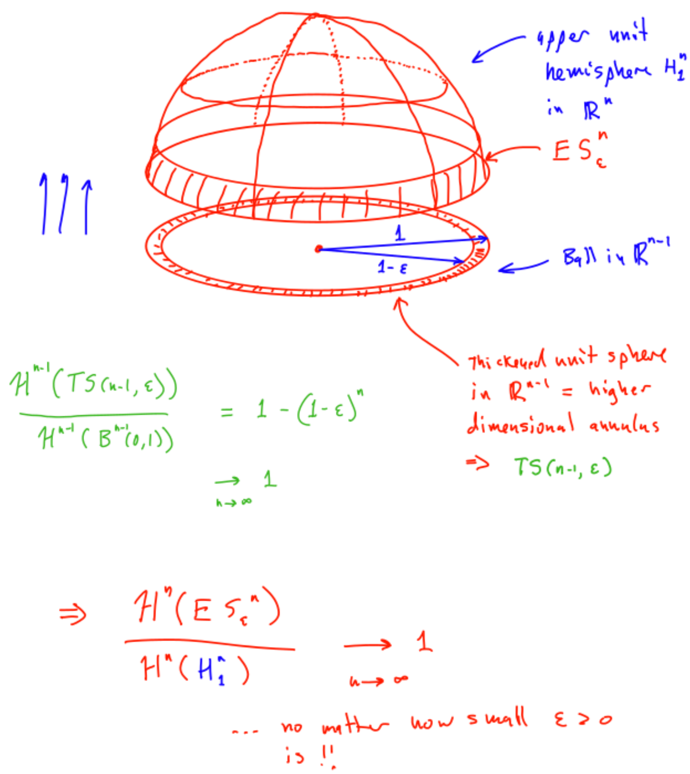


Figure 17: Somewhat cryptic explanation of concentration of measure in the unit sphere in high dimensions.

Stochastic flows on rectifiable sets Flows in which there are stochastic driving terms, or stochastic coefficients are useful in the modeling of all sorts of physical phenomena. While this area seems wide open – I am having difficulty finding papers addressing these ideas – there is work on nonsmooth analysis on smooth manifolds, which is at least something of a cousin to the investigations of stochastic flows on non-smooth sets. See the papers by Ledyaev and Zhu that can be found here: `http://homepages.wmich.edu/~zhu/papers.htm`. To imagine what I see here, think of the 2-dimensional graph of a Lipschitz function $f : \mathbb{R}^2 \to \mathbb{R}$, in the 3-dimensional graph space, $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ and imagine a stochastic vector field on that set. Perhaps a simpler case, containing all the essential features, is a random walk in the graph of the same Lipschitz function f in $\mathbb{R}^3$. This leads to diffusion on the graph. Here is an example of a paper addressing diffusion on a nonsmooth set `https://arxiv.org/pdf/1312.5882.pdf`. While this seems to be somewhat of a boutique area of work, I am convinced that it is very promising especially when the approach becomes a hybrid computational/experimental and theoretical approach.

Complex Dynamics and Conformal Mappings A very large area of work arises from complex analysis and the richness of analytic function theory. In this book we will not deal with complex analysis, and in fact the special nature of analytic functions makes them a bit disjoint from the version of geometric analysis we will study. Yet the beauty of this area and the richness of the results, compels me to include this rich garden in this list. In fact the study of Harmonic measures, mentioned above, is inseparable from the study of complex analysis. Conformal maps – maps that locally preserve angles – have generated a great deal of work. To mention only one very remarkable and useful result, there is the Riemann Mapping theorem:

Theorem 6.1.1 (Riemann Mapping Theorem)**.** *Let Ω be a simply connected region in $\mathbb{C}$ that is not equal to the entire complex plane. Then there is a conformal map ϕ, mapping Ω onto $\mathbb{D}$, the unit disk in the complex plane.*

The ramifications of this result are rather large. See for example Pommerenke's book *Boundary Behavior of Conformal Maps* [33].

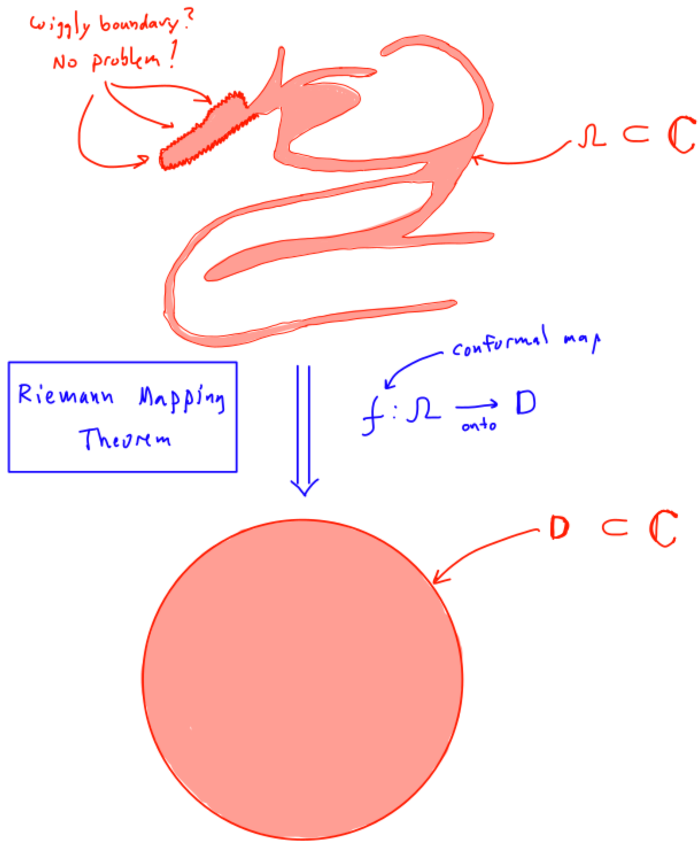


Figure 18: The Riemann Mapping Theorem: even crazy simply connected sets are conformally equivalent to the unit disk!

6.2 Other Things

So what is geometric analysis? That is, what is it beyond "whatever lies at the intersection of analysis and geometry"? If forced to try to make it more precise than that, I would say that it is the attitude, the state of mind, the perspective of the mathematician exploring and creating and illuminating what he or she finds for others. Perhaps it is the state in which you see and create, the one where language seems to just get in the way.

Whatever it is, perhaps it cannot withstand too much dissection – like some sort of uncertainty principle, when you have taken it apart to see the pieces, you find you have a pile of pieces and no understanding of the living, creating thing those pieces comprised.

Part III

Analysis I

7 Just Enough Metric Spaces

In this short chapter, I cover the essentials from metric spaces we will need for the rest of the book. While some readers may have been fortunate enough to have had a “proofs” course using metric spaces as a vehicle for teaching how to create proofs, it is likely that some have not.

The basic idea in studying metric spaces is to see what we can gather about sets of points when all we know is the distance between points. It turns out, we can create very important tools that will be useful all over the geometric analysis landscape.

At the end of the chapter we devote a few sections to the extra things we get when the metric space is also a vector space and the distance is the vector norm. Sometimes this norm will come from a dot product (called an *inner product*) and sometimes not.

7.1 What is a Metric Space?

A metric is a function on $X \times X$ – you stick in two points, and out pops the distance between them. Anything that satisfies three simple axioms for points in X, is a metric:

Definition 7.1.1 (Metric). *A function* $\rho : X \times X \to [0, \infty)$ *satisfying:*

1. $\rho(x, y) \geqslant 0$, *with* $\rho(x, y) = 0 \Leftrightarrow x = y$
2. $\rho(x, y) = \rho(y, x)$
3. $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$ *(triangle inequality)*

is a metric on X.

The space X together with ρ form a *metric space*.

7.2 Open and closed Sets

The basic entity in metric spaces are the open (and closed balls) with some fixed radius.

7.2.1 Balls in a metric space

Definition 7.2.1 (Metric Balls). *We define* ***open and closed balls*** *of radius* $r > 0$*, and center* $y \in X$ *to be all the points* $x \in X$ *whose distance to* y *is, respectively, (1) less than* ***or*** *(2) less than or equal to* r*:*

Open Ball: $B(y, r) \equiv \{x : \rho(y, x) < r\}$
Closed Ball: $\bar{B}(y, r) \equiv \{x : \rho(y, x) \leqslant r\}$

7.2.2 Open Sets and Closed Sets

Definition 7.2.2 (Open and Closed Sets). *We say a set* E *is an* ***open set*** *if for every point* $x \in E$*, there is a radius* $r > 0$ *small enough that* $B(x, r) \subset E$*. A set* E *is a* ***closed set*** *if its complement,* $X \setminus E = E^c$*, is open.*

Exercise 7.2.1. Prove that the union of an arbitrary (possibly uncountable!) collection of open sets is open.

Exercise 7.2.2. Prove that the intersection of a finite collection of open sets is open.

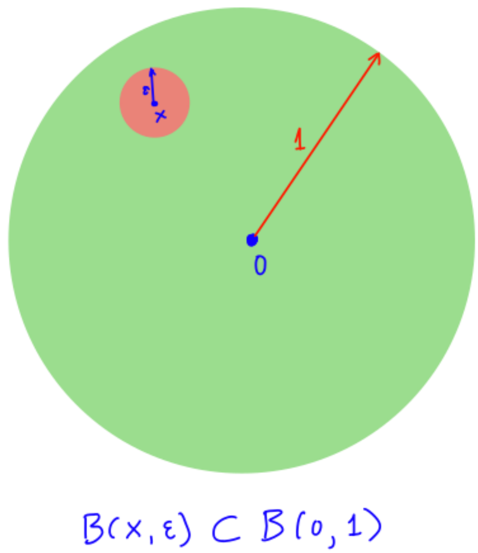


Figure 19: The Open ball is an open set: see Exercise 7.7.3.

7.3 Convergence, Continuity and Completeness

Definition 7.3.1 (Convergence). *We say that the sequence* $\{x_i\}_{i=1}^{\infty} \subset X$ *converges to* x^* *if*

$$\lim_{i \to \infty} \rho(x_i, x^*) = 0$$

We sometimes write this as $x_i \to x^*$

Definition 7.3.2 (Inverse Image). *For* $f : X \to Y$*, we define* $f^{-1}(A) \equiv \{x \mid f(x) \in A \subset Y\}$ *and we refer to* $f^{-1}(A)$ *as the "inverse image of* A *under* f*", even if there is no function such that* $g(y) = f^{-1}(y)$ *for all* $y \in Y$*, i.e. even if* f *is not invertible.*

Definition 7.3.3 (Continuity I). *A function mapping from one metric space* X *to another metric space* Y *(*X *could equal* Y*) is said to be continuous if, for every open set* $E \subset Y$*, the inverse image* $f^{-1}(E)$ *is open in* X*.*

Definition 7.3.4 (Continuity II). *A function mapping from one metric space X to another metric space Y (X could equal Y) is said to be continuous if:*

$$x_i \to x^* \quad \Rightarrow \quad f(x_i) \to f(x^*)$$

Exercise 7.3.1. Prove that the two definitions are equivalent. That is that

$$\{\text{Continuity I} \Rightarrow \text{Continuity II}\}$$

and

$$\{\text{Continuity II} \Rightarrow \text{Continuity I}\}$$

If we want to talk about the continuity of a function or mapping on a set that is not the entire space – that is, we define f on a subset of the space X – then we have to modify the definitions a little bit in the case that the set E is not open.

Definition 7.3.5 (Continuity I*). *A function mapping $\Omega \subset X$ to another metric space Y (X could equal Y) is said to be continuous if, for every open set $E \subset Y$, the inverse image $f^{-1}(E) = \Omega \cap F$ for some open $F \subset X$.*

Exercise 7.3.2. Show that the starred version of Continuity I is equivalent to the unstarred one when Ω is an open subset of X

Remark 7.3.1 (Continuity II*?). *Note: when $f : \Omega \subset X \to Y$, the definition of continuity using sequences is unchanged, so there is no new "Continuity Definition II*".*

Definition 7.3.6 (Cauchy Sequences). *A sequence $\{x_i\}_{i=1}^{\infty} \subset X$ is a Cauchy sequence if, for every $\epsilon > 0$, there an $N(\epsilon)$ such that $i, j > N(\epsilon) \Rightarrow \rho(x_i, x_j) < \epsilon$.*

Definition 7.3.7 (Completeness). *We say that a metric space (ρ, X) is complete if every Cauchy sequence has a limit in X. That is, if $\{x_i\}_{i=1}^{\infty} \subset X$ is a Cauchy sequence, there exists a point $x^* \in X$ such that $x_i \to x^*$.*

Definition 7.3.8 (Continuity III, ϵ, δ-style). *A function mapping from $\Omega \subset X$ to another metric space Y (X could equal Y) is said to be continuous if for every point $x \in \Omega$ and $\epsilon > 0$, there is a $\delta(x, \epsilon) > 0$ such that*

$$f(B(x, \delta(x, \epsilon)) \cap \Omega) \subset B(f(x), \epsilon)$$

Exercise 7.3.3. Show that Continuity III is equivalent to Continuity I*

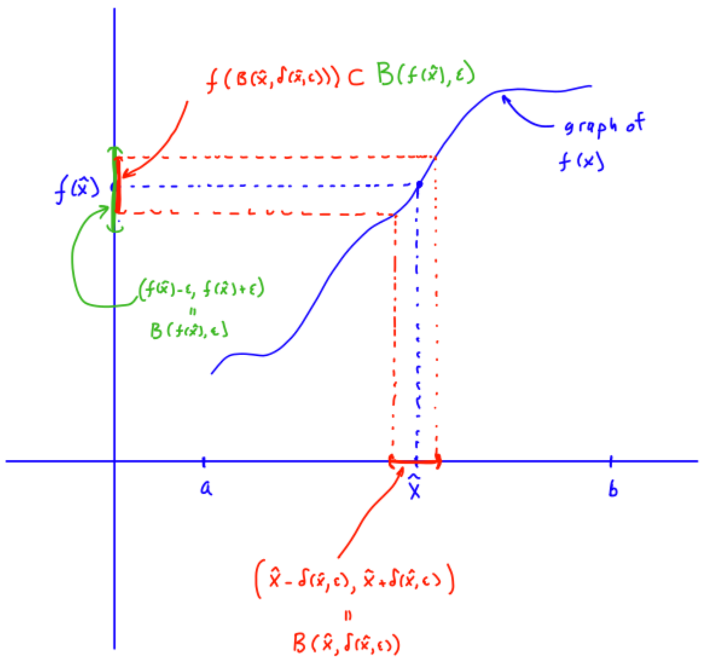


Figure 20: Continuity, ϵ, δ-style.

7.4 Compactness

A fundamental concept in metric spaces is the idea of compactness. In $\mathbb{R}^n$ it will turn out that a set is compact if and only if it is closed and bounded.

Definition 7.4.1 (Compact Sets $K \subset X$). *A set $K \subset X$ is compact if every open cover of K contains a finite subcover. That is, if a collection of open sets $\mathcal{O} \equiv \{O_\alpha\}_{\alpha \in \mathcal{A}}$ covers K – or succinctly, $K \subset \bigcup_{\alpha \in \mathcal{A}} O_\alpha$ – there is a finite subcollection $\{O_{\alpha_i}\}_{i=1}^N \subset \{O_\alpha\}_{\alpha \in \mathcal{A}}$ such that $K \subset \bigcup_{i=1}^N O_{\alpha_i}$.*

We now want to talk about maximizing and minimizing functions which map points in a metric space to the real numbers. To do that we need to define max, min, sup, and inf.

To begin with, we consider a non-empty subset E of the real numbers which is bounded above and below:

$$-\infty < a < b < \infty \text{ and } a < x < b \text{ for all } x \in E$$

Definition 7.4.2 (Supremum, sup). *Suppose we have a subset of the real numbers, $E \subset \mathbb{R}$. An upper bound u is any real number so that $x \in E \Rightarrow x \leqslant u$. We define the set of upper bounds for E to be U_E. It is clear that U_E is non-empty and is bounded below by elements of E. It is a property of the real numbers that every such set has a least element: there is a $u^* \in U_E$ such that $u^* \leqslant u$ for all $u \in U_E$. We define the* ***supremum of E to be*** *u^*. More succinctly, the $\sup(E)$ = the smallest upper bound.*

Definition 7.4.3 (Infimum, inf). *Suppose we have a subset of the real numbers, $E \subset \mathbb{R}$. A lower bound l is any real number so that $x \in E \Rightarrow x \geqslant l$. We define the set of lower bounds for E to be L_E. Again, L_E is non-empty and bounded above. the same property of the real numbers implies that L_E has a greatest element $l_* \in L_E$ such that $l_* \geqslant l$ for all $l \in L_E$. We define the* ***infimum of E to be*** *l_*. More succinctly, the $\inf(E)$ = the greatest lower bound.*

If E is a set that is not bounded above, we say that $\sup(E) = \infty$ and if E is not bounded below, $\inf(E) = -\infty$.

Exercise 7.4.1. Show by example that not every set of real numbers E, with $\sup(E) < \infty$ has a maximum element e^*. That is there might not be an $e^* \in E$ such that $e \leqslant e^*$ for all $e \in E$. Define a minimum by analogy and show that it might not exist, even when $\inf(E) > -\infty$.

Definition 7.4.4 (**ϵ-net**)**.** *Given a set E in a metric space X, we say that $A_E^\epsilon \equiv \{x_\alpha\}_{\alpha \in \mathcal{A}}$ is an ϵ-net for E if*

$$E \subset \bigcup_{\alpha \in \mathcal{A}} B(x_\alpha, \epsilon).$$

I.e. the union of the ϵ-balls, centered on the points in A_E^ϵ, covers E. In other words, every point in E is less than epsilon away from some point in A_E^ϵ

Definition 7.4.5 (**Totally Bounded**)**.** *A set $E \subset X$ is totally bounded if, for every $\epsilon > 0$, there exists a finite ϵ-net for E, A_E^ϵ – i.e. the number of points in A_E^ϵ is finite.*

Theorem 7.4.1 (Compact = Closed and Totaly Bounded)**.** *Suppose that X is a complete metric space. A subset $K \subset X$ is compact if and only if it is closed and totally bounded.*

Exercise 7.4.2. Prove the "only if" part of Theorem (7.4.1). I.e. prove that compactness of K implies K totally bounded and closed.

Theorem 7.4.2 (Compactness implies Convergence)**.** *Suppose that we have a sequence $\{x_i\}_{i=1}^\infty \subset K$, where K is a compact subset of the complete metric space X. Then there is a convergent subsequence: There exists $x^* \in K$ and a monotonically increasing $k : \mathbb{N} \to \mathbb{N}$ such that*

$$x_{k(i)} \underset{i \to \infty}{\to} x^*.$$

Exercise 7.4.3. Prove Theorem (7.4.2). Hint: use the total boundedness of K to generate a Cauchy sequence. Use the completeness to get a limit in X use the closedness of K to get that this limit is in K.

Theorem 7.4.3 (Compactness implies Accumulation)**.** *Suppose that $A \subset K$, where K is a compact subset of the metric space X and A is an infinite set (*

i.e. it has an infinite number of points in it). Then there exists $x^* \in K$ *and a one-to-one function* $i : \mathbb{N} \to a_i \in A$ *such that*

$$a_i \underset{i\to\infty}{\to} x^*.$$

Such a point x^* *is called an accumulation point of* A*.*

Exercise 7.4.4. Prove Theorem (7.4.3).

Definition 7.4.6 (Separable). *A metric space* X *is separable if it contains a dense countable subset. I.e. it contains a set* $F = \{f_i\}_{i=1}^{\infty} \subset X$ *such that, for any* $x \in X$ *and any* $\epsilon > 0$*, there is a point* $f_k \in F$ *such that* $\rho(f_k, x) < \epsilon$*.*

Exercise 7.4.5. Show that any open cover of a set in a separable metric space has a countable subcover.

Exercise 7.4.6. **(Challenge)** Prove the "if" part of Theorem (7.4.1). I.e. prove that K totally bounded and closed implies compact. **Hint**: First show that any open cover $\mathcal{E} = \{E_\alpha\}_{\alpha\in\mathcal{A}}$ of a totally bounded set in a metric space has a countable subcover, $\{E_{\alpha_i}\}_{i=1}^{\infty}$. Now assume K is totally bounded and closed but there is a open cover with no finite subcover. Use the fact there is a countable subcover to construct a sequence $\{x_i\}_{i=1}^{\infty} \subset K$ such that $x_j \notin \cup_{i=1}^{j} E_{\alpha_i}$ for all j. Now use the total boundedness and closure to get that there is an $x^* \in K$ such that a subsequence of $\{x_i\}_{i=1}^{\infty}$ converges to x^*. Find a contradiction.

Definition 7.4.7 (Sequentially Compact). *A set* K *is sequentially compact if every sequence contained by* K *has a convergent subsequence converging to a point in* K*.*

Theorem 7.4.4 (Compact $\Leftrightarrow$ Sequentially Compact, (in metric spaces))**.** *A set* K*, in a metric space* X*, is closed and totally bounded if and only if it is sequentially compact.*

Exercise 7.4.7. Prove Theorem (7.4.4). Hint: the only if part can be proven using the results above. The if part can be shown by assuming that it is not true, that it is sequentially compact, but that there is some $\epsilon > 0$ for which there is no finite ϵ-net and this allows us to pick a sequence in K with no convergent subsequence.

7.4.1 *Max, Min and Uniform Continuity*

Theorem 7.4.5. *If $f : X \to Y$ is continuous and $K \subset X$ is compact, then $f(K)$ is also compact.*

Theorem 7.4.6 (Max's and Min's on Compact Sets)**.** *If a function $f : X \to \mathbb{R}$ is continuous on a compact set K, then there exists $x_*, x^* \in K$ such that $f(x_*) = \inf(f(K))$ and $f(x^*) = \sup(f(K))$.*

Theorem 7.4.7 (Compactness gives Uniform Continuity)**.** *If a function $f : X \to \mathbb{R}$ is continuous on a compact set K, then a $\delta(x, \epsilon)$ (from Definition 7.3.8) can be found that does not depend on x. I.e. the size of the δ-balls is uniform in x.*

7.5 Connectedness

Definition 7.5.1 (Connectedness). *A set $D \subset X$ is connected **if**:*

$$\text{(1) } O_1 \text{ and } O_2 \text{ open, , (2) } O_1 \cap O_2 = \emptyset, \text{ and (3) } D \subset O_1 \cup O_2$$

implies

$$O_1 \cap D = \emptyset \textbf{ or } O_2 \cap D = \emptyset.$$

Putting it succinctly, if we intersect a connected subset with two disjoint open sets, one of the intersections must be empty. We say that D cannot be separated by two disjoint open sets.

Theorem 7.5.1 (Continuity Preserves Connectedness)**.** *If $f : X \to Y$ is continuous and $D \subset X$ is connected, then $f(D)$ is connected in Y.*

Exercise 7.5.1. Prove Theorem (7.5.1). Hint: suppose $f(D)$ is not connected and use two disjoint open sets to separate $f(D)$.

7.6 Liminf and Limsup

Definition 7.6.1 ($\limsup$, $\liminf$ **for Sequences in** $\mathbb{R}$)**.** *Suppose that* $\{x_i\}_{i=1}^{\infty} \subset \mathbb{R}$. *We define:*

$$\limsup_{i\to\infty} x_i \equiv \lim_{m\to\infty} \sup(\{x_i\}_{i=m}^{\infty}) \text{ and}$$
$$\liminf_{i\to\infty} x_i \equiv \lim_{m\to\infty} \inf(\{x_i\}_{i=m}^{\infty}).$$

Definition 7.6.2 (limsup, liminf for Functions). *Suppose that* $f : X \to \mathbb{R}$. *Then*

$$\limsup_{x\to x^*} f \equiv \lim_{r\to 0}\left(\sup_{x\in B(x^*,r)} f(x)\right) \text{ and}$$
$$\liminf_{x\to x^*} f \equiv \lim_{r\to 0}\left(\inf_{x\in B(x^*,r)} f(x)\right).$$

Remark 7.6.1. *Note that some analysts prefer the definition that ignores the value of the function at* x^*. *Thus, in the definition above, instead of* $x \in B(x^*, r)$ *they use* $x \in B(x^*, r) \setminus \{x^*\}$.

Exercise 7.6.1. Prove that $\limsup_{x\to x^*} f = \liminf_{x\to x^*} f = f(x^*)$ for all $x^* \in X$ if and only if $f : X \to \mathbb{R}$ is continuous.

7.6.1 Upper Semi-continuity and Lower Semi-continuity

Definition 7.6.3 (Upper and Lower Semi-continuity). *Suppose* $f : X \to \mathbb{R}$.

If $\limsup_{x\to x^*} f \leqslant f(x^*)$, *we say that* f *is upper semi-continuous at* x^*. *If this is true at every* $x^* \in X$, *then we simply say that* f *is upper semicontinuous on* X.

If $\liminf_{x\to x^*} f \geqslant f(x^*)$, *we say that* f *is lower semi-continuous at* x^*. *If this is true at every* $x^* \in X$, *then we simply say that* f *is lower semicontinuous on* X.

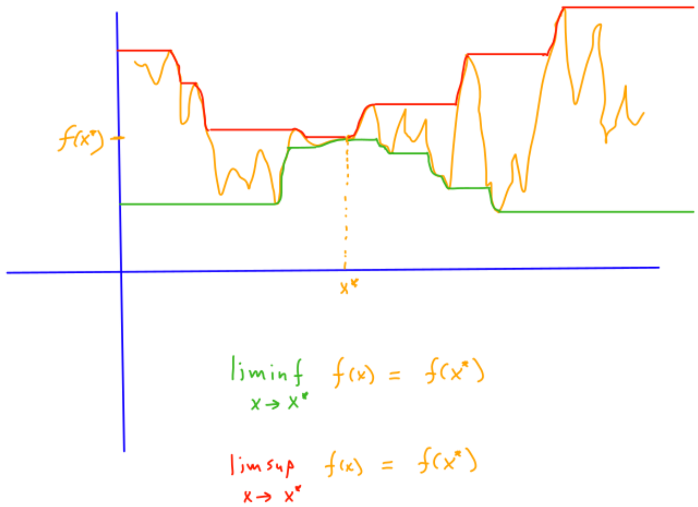


Figure 21: Upper and lower envelopes as x^* is approached, illustrating how, for continuous functions, the envelopes pinch together at the limit.

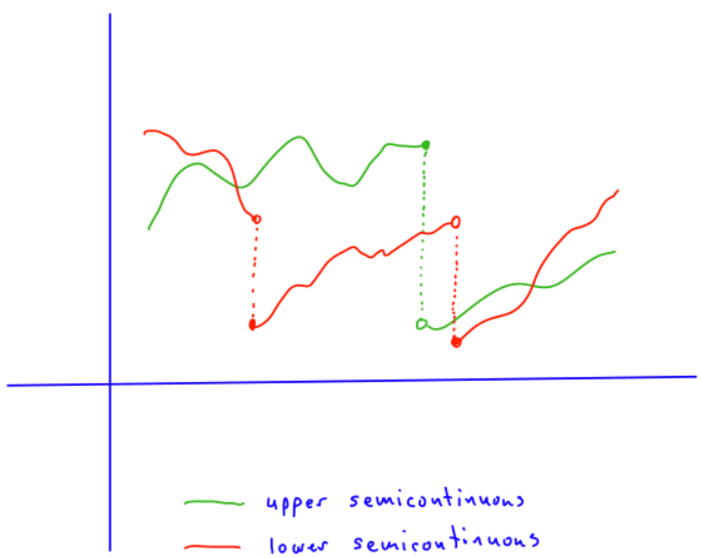


Figure 22: Semicontinuity illustrations.

7.7 Examples of Metric Spaces + Many More Exercises

7.7.1 *Thinking about $\mathbb{R}^2$ as a Metric Space*

Note: I often use 0 to denote the origin in $\mathbb{R}^n$ as well as the number $0 \in \mathbb{R}$.

If we define $|x| \equiv \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2}$ we can verify that $\rho(x, y) \equiv |x - y|$ satisfies:

1. $|x - y| \geqslant 0$ and $|x - y| = 0 \Rightarrow x = y$
2. $|x - y| = |y - x|$
3. $|x - z| \leqslant |x - y| + |y - z|$

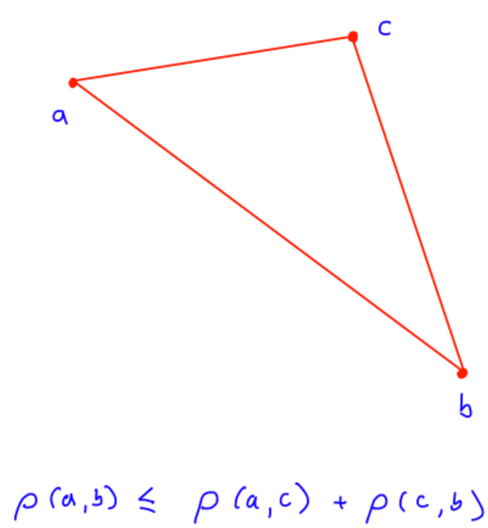


Figure 23: The Triangle Inequality.

Exercise 7.7.1. Show that $x \cdot y \leqslant |x|\,|y|$ for all $x, y \in \mathbb{R}^2$. Hint: assume $|x| = |y| = 1$ and compute $|x - y|$.

Exercise 7.7.2. Show that $\rho(x, y) \equiv |x - y|$ satisfies the triangle inequality/ Hint: Defining $u = x - y$ and $v = y - z$, note that we are trying to prove that $|u + v| \leqslant |u| + |v|$. Now recall that $|u| = \sqrt{u \cdot u}$.

We get that $(\rho(x, y) = |x - y|, \mathbb{R}^2)$ is a metric space.

Exercise 7.7.3. Show that any open ball in $(|x - y|, \mathbb{R}^2)$ is actually an open set. This takes a little bit of work and the triangle inequality. See Figure (19).

Exercise 7.7.4. Show that the closed unit ball $\bar{B}(0, 1) \subset \mathbb{R}^2$ is not open by finding a point $x \in \bar{B}(0, 1)$ such that $B(x, r) \not\subset \bar{B}(0, 1)$ for all $r > 0$.

Definition 7.7.1 (Interior, Exterior). *The interior of a set are all the points in $x \in E$ such that, for some $r > 0$, $B(x, r) \subset E$. The exterior of E is the set of interior points of E^c. We denote the interior by* $\text{int}(E)$ *or* E^o. *We denote the exterior of E by* $\text{ext}(E)$ *or* $(E^c)^0$.

Definition 7.7.2 (Topological Boundary). *The boundary of a set E is the set of points $x \in \mathbb{R}^2$ such that for every $r > 0$, $B(x, r) \cap E \neq \emptyset$ and $B(x, r) \cap E^c \neq \emptyset$. In other words, x is neither an interior point of E nor an exterior point of E. We denote the boundary of E by ∂E.*

Exercise 7.7.5. Find all the boundary points in $\bar{B}(0, 1) \subset \mathbb{R}^2$.

Exercise 7.7.6. Define $\mathbb{Q}$ to be all the points in $\mathbb{R}^2$ with rational coordinates. Find (1) the interior points, (2) the exterior points, and (3) the boundary points of $\mathbb{Q}$.

Exercise 7.7.7. Define D to be the closed unit ball in $\mathbb{R}^2$, centered at the origin 0, with the origin removed: $D = \bar{B}(0, 1) \setminus \{(0, 0)\}$. Find an infinite open cover of D that has no finite subcover.

Exercise 7.7.8. Show that the intersection of an arbitrary family of closed sets in a metric space is closed.

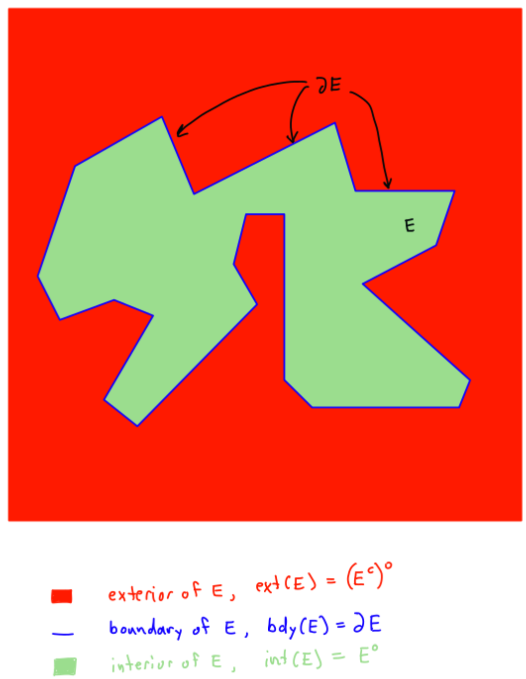


Figure 24: Illustration of interior, boundary and exterior.

Definition 7.7.3 (Closure). *The closure of a set* E *is the intersection of all the closed sets containing* E *and is denoted* $\operatorname{clos}(E)$.

Exercise 7.7.9. Find a set $E \subset \mathbb{R}^2$ such that the $\partial E = \operatorname{clos}(E)$, but $\partial \operatorname{clos}(E) = \emptyset$.

7.7.2 *Path Length Spaces:* $\mathbb{R}^2$

We will define the space $X = \mathbb{R}^2$ and choose a bounded function $g : X \to [\epsilon, C]$ where $0 < \epsilon < C$. First recall that a Lipschitz function $f : \mathbb{R}^k \to \mathbb{R}^m$ is one for which $|f(x) - f(y)| \leqslant K|x - y|$ for all x and y and some (fixed) $K < \infty$. Now define a path to be any image of a Lipschitz function $\gamma : [0, 1] \to X$. Then the length of the path γ from $\gamma(0)$ to $\gamma(1)$ is defined to be:

$$l_g(\gamma) \equiv \int_0^1 g(\gamma(t))|\dot{\gamma}(t)|dt$$

Define

$$\Gamma(x, y) = \{\gamma : [0, 1] \to X \mid \gamma(0) = x \text{ and } \gamma(1) = y\}.$$

Now we define a distance on X by:

$$\rho_g(x, y) = \inf_{\gamma \in \Gamma(x,y)} (l_g(\gamma)) = \inf_{\gamma \in \Gamma(x,y)} \int_0^1 g(\gamma(t))|\dot{\gamma}(t)|dt$$

Exercise 7.7.10. Suppose g is constant on the regions shown in figure (26), taking the value 0 inside the disks and 1 outside the disks. Find the shortest paths between the pairs of points shown. Hint: show that if your path goes through one of the disks where $g = 0$, that the path has to enter and exit the disks orthogonally, and that outside of the disks, the paths are straight lines.

7.7.3 *Path Length Spaces: Graphs*

Suppose that $G = (V, E)$ is a graph embedded in $\mathbb{R}^n$, with vertex set $V = \{v_j\}_{j=1}^M$ and the set of undirected edges $E = \{e_i\}_{i=1}^N$. Define

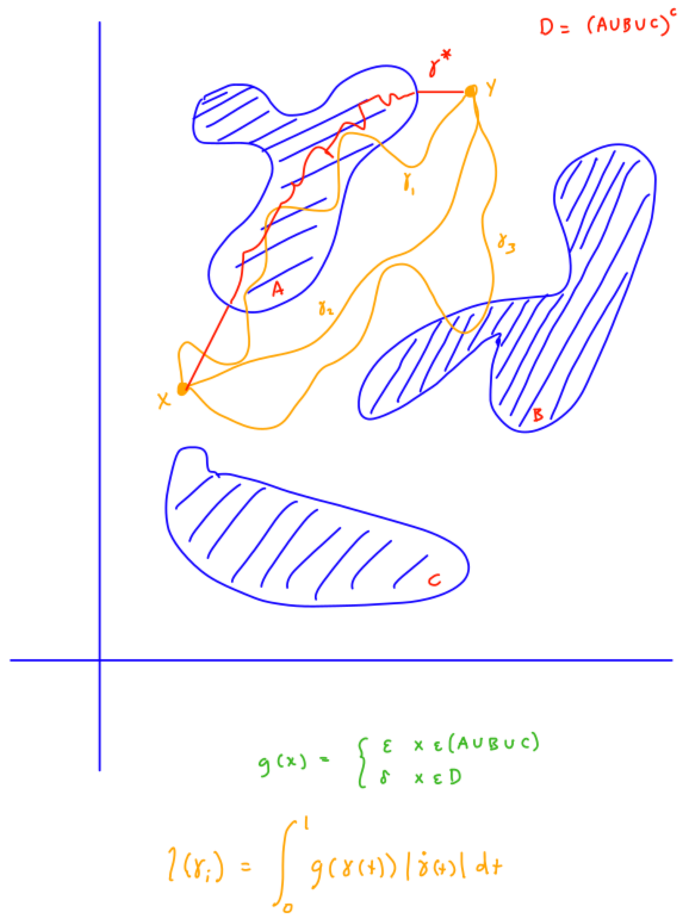


Figure 25: Path Length spaces with simple length kernels, g.

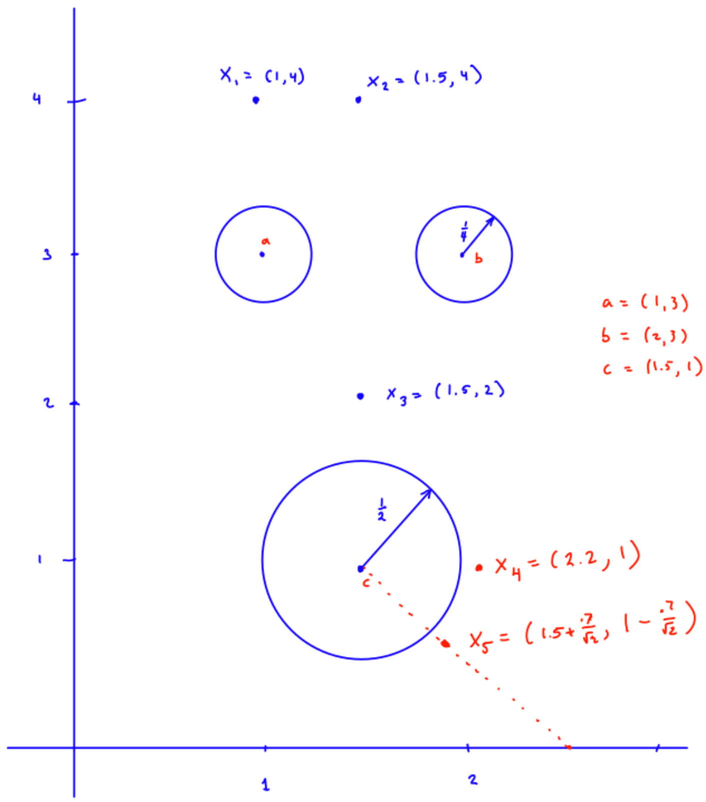


Figure 26: Path-length exercise: See Exercise (7.7.10).

$l_i = |v_{j(i)} - v_{k(i)}|$, the length of the ith edge connecting $v_{j(i)}$ and $v_{k(i)}$. We define $\rho_G(v, w)$ for $v, w \in V$ to be the length of the minimal path from v to w.

More details: if we can move from v to w along Γ, a sequence of connecting edges, $e_{i(1)}, ..., e_{i(J)}$, where v is an endpoint of $e_{i(1)}$ and w is an endpoint of $e_{i(J)}$, the length of Γ is defined to be $l_\Gamma \equiv \sum_{k=1}^{J} l_{i(k)}$.

Define the set of all paths from v to w to be $\Gamma(v, w)$. Now, define

$$\rho_G(v, w) = \min_{\Gamma \in \Gamma(v,w)} l_\Gamma$$

Exercise 7.7.11. Show that (ρ_G, G) is a metric space.

Definition 7.7.4 (Degree of a Vertex, Maximal Degree). *The **degree** of a vertex is the number of edges that connect to that vertex. **Maximal degree** for a graph is the largest degree that any vertex has in the graph.*

Exercise 7.7.12. Find an algorithm for finding minimal paths on a graph. What is the complexity in terms of maximal degree and the number of points in the graph? Hint: Look up dynamic programming.

7.7.4 Normed Vector Spaces: $\mathbb{R}^n$

Definition 7.7.5 (Euclidean Norm). *Suppose that $x \in \mathbb{R}^n$. Define*

$$|x| = \sqrt{x \cdot x}.$$

Exercise 7.7.13. Convince yourself that showing $x \cdot y \leqslant |x|\,|y|$ is as easy in $\mathbb{R}^n$ as it is in $\mathbb{R}^2$. See Exercise 7.7.1.

7.7.5 Normed Vector Spaces: Function Spaces

Definition 7.7.6 (Space of Continuous Functions, $C([a,b])$). *We define* $C([a,b])$ *to be the set of all continuous functions on the closed and bounded interval* $[a,b]$*:* $f:[a,b]\to\mathbb{R}$ *with the distance between any two functions* f,g *given by*

$$|f-g|_{\sup} \equiv \max_{x\in[a,b]} |f(x)-g(x)|$$

.

Exercise 7.7.14. Show that $C([a,b])$ is a metric space.

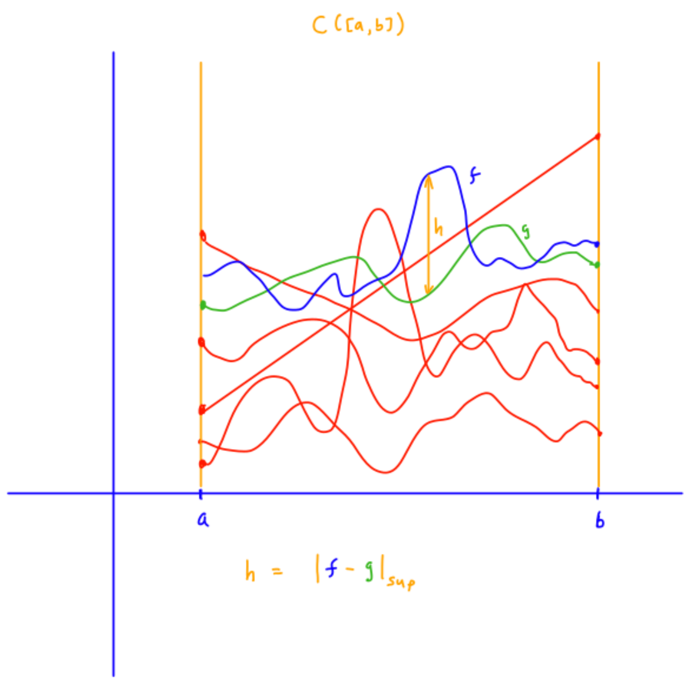


Figure 27: Illustration of $C([a,b])$.

Definition 7.7.7 (Space of Lipschitz Functions, $\mathrm{Lip}([a,b],K,B)$). *We define* $\mathrm{Lip}([a,b],K,B)$ *to be the set of all Lipschitz continuous functions on the closed and bounded interval* $[a,b]$*:* $f:[a,b]\to\mathbb{R}$*, such that the Lipschitz* *constant for the function does not exceed* K *and* $|f(x)| < B$ *for all* $x \in [a, b]$. *Let the distance between any two functions* f, g *again be given by*

$$|f - g|_{\sup} \equiv \max_{x \in [a,b]} |f(x) - g(x)|$$

.

Exercise 7.7.15. **(Challenge)** Show that $\mathrm{Lip}([a, b], K, B)$ is totally bounded.

Definition 7.7.8 (**Space of** L^2 **Functions,** $\mathbf{L^2([a, b])}$). *We define* $L^2([a, b])$ *to be the set of all functions on the interval* $[a, b]$*:* $f : [a, b] \to \mathbb{R}$ *with the distance between any two functions* f, g *given by*

$$|f - g|_2 \equiv \sqrt{\int_a^b |f(x) - g(x)|^2 \, dx}$$

.

Exercise 7.7.16. **(Challenge)** Show that $L^2([a, b])$ is a metric space.

7.8 Remarks about Topological Spaces

Metric spaces are an example of general topological spaces.

Definition 7.8.1 (**Topological Space**). *A topological space is a set of points* X *together with a collection of subsets,* $\mathcal{T}$*, called* ***open sets****, satisfying:*

1. $\emptyset, X \in \mathcal{T}$.
2. *Any arbitrary union of open sets is also open: another way of saying this is that* $\mathcal{T}$ *is closed under arbitrary unions.*
3. *The intersection of any finite collection of open sets is also open:* $\mathcal{T}$ *is closed under finite intersections.*

The universe of topological spaces is an infinite, unimaginably wild zoo of strange beasts. Almost all the spaces we will mention – and all the spaces we work with very much – will be metric spaces. And most of those metric spaces will be normed vector spaces.

Exercise 7.8.1. Show that the following is a topological space, but not a metric space. Let $X = \mathbb{R}$ and $\mathcal{T} = \{$all sets whose complement is a finite number of points$\} \cup \{\emptyset\}$.

7.9 Extra Problems

All of the following are extra credit and are designed to give you extra experience in the exploration of metric spaces.

Exercise 7.9.1. Suppose that $X = \mathbb{R}^2$ and $\rho(x, y) = |x - y|_\infty \equiv \max\{|x_1 - y_1|, |x_2 - y_2|\}$. Show that (ρ_∞, X) is a metric space.

Exercise 7.9.2. Suppose that $X = \mathbb{R}^2$ and $\rho(x, y) = |x - y|_1 \equiv \sum_{i=1}^{2} |x_i - y_i|$. Show that (ρ_1, X) is a metric space.

Exercise 7.9.3. Sometimes the metric space in Exercise (7.9.2) is called the taxicab metric. Why?

Exercise 7.9.4. We again focus on the metric space in Exercise (7.9.2). Suppose that $p = (0, 0)$ and $q = (1, 1)$. Find the set of all paths of minimal length from p to q.

Exercise 7.9.5. The discrete metric is the metric given by (1) $\rho(x, y) = 1$ if $x \neq y$ and (2) $\rho(x, y) = 0$ if $x = y$. Suppose that $X = \mathbb{R}^2$ and ρ is the discrete metric.

1. Describe in detail what open balls of radius .5, 1, 1.1, and 10 look like.
2. Which sets in X are open?
3. Which sets in X are closed?

Exercise 7.9.6. Define the following metric on $\mathbb{R}^2$

$$\rho_{2_T}(x,y) \equiv \begin{cases} |x-y|_2 & \text{if } |x-y|_2 \leqslant 1 \\ 1 & \text{if } |x-y|_2 > 1 \end{cases}$$

where $|\cdot|_2$ is the usual Euclidean norm on $\mathbb{R}^2$. Show that this "metric" is indeed a metric – i.e. show that it satisfies the triangle inequality. We call this the **truncated Euclidean metric** on $\mathbb{R}^2$.

Exercise 7.9.7. **Important fact**: Suppose $\mathbb{R}^n$ is given the usual Euclidean metric. Show that a set $E \subset \mathbb{R}^n$ is compact if and only if it is both closed and bounded.

Exercise 7.9.8. Let M_T be $\mathbb{R}^2$ with the truncated Euclidean metric and M_2 be $\mathbb{R}^2$ with the usual Euclidean metric. Show:

1 The open sets in M_T are open in M_2 and the open sets in M_2 are open in M_T.
2 Which sets are compact in M_2 and which sets are compact in M_T? Does this conflict with Exercise (7.9.7?)

Exercise 7.9.9. Unions of closed sets:

1 Show that a finite union of closed sets is always closed.
2 Show by example that the union of a countably infinite family of closed balls in $\mathbb{R}^2$ need not be closed.
3 Give an example of an infinite union of closed balls in $\mathbb{R}^2$ that is closed.

8 The Art of Inequalities

8.1 Why Inequalities?

The *art of analysis* is tightly bound up with the *art of inequalities*. In this chapter we will explore inequalities, (hopefully) launching you on a lifelong exploration of the endless riches they hold.

The reason for this importance of inequalities has to do with the wider set of questions we ask when we move to more advanced mathematics.

In elementary mathematics – Algebra, calculus beginning linear algebra and differential equations – the focus is often on solving equations, for example:

Solving $\mathbf{Ax} = \mathbf{b}$: Find $x \in \mathbb{R}^n$ that satisfies $Ax = b$, where A is an n by n real matrix and $b \in \mathbb{R}^n$.

Computing $\frac{df}{dx}$: Finding the derivative of f.

Computing $\int_a^b f(x)\, dx$: computing the definite integral of f – the area underneath the curve from a to b.

Solving $\frac{d^2f}{dt^2} + a\frac{df}{dt} + bf = 0$: Finding a function of time, $f : t \in \mathbb{R} \to \mathbb{R}$ which satisfies $\ddot{f} + a\dot{f} + bf = 0$, $f(0) = f_0$ and $\dot{f}(0) = d_0$.

In each of these cases, we are looking for *the* solution, though, in the $Ax = b$ case, we might have a nontrivial null space in which case there will be an entire affine subspace of solutions.

When we move beyond the relatively elementary questions, there are questions added to the exploration, making inequalities both necessary and natural. Here are important examples of those questions. In these examples, we will be considering metric spaces which are vector spaces, with a metric d given by the vector space norm $|\cdot|$,

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

CONTINUITY: One version of continuity of mappings is that a mapping $f : X \to Y$ is continuous at x^*, if for any $\epsilon > 0$, there is a $\delta(\epsilon) > 0$ so that:

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

Equivalently, there is a $\delta(\epsilon) > 0$ so that

$$f\left(B(x^*, \delta(\epsilon))\right) \subset B(f(x^*), \epsilon)$$

for all $\epsilon > 0$. This arises repeatedly – proving something is continuous is often an important part of inferring some property you want to infer.

APPROXIMATION AND CONVERGENCE: Here is an example. Suppose that the equation $Ax = b$ is given to us, but that the b is only known to some tolerance ϵ. That is, the true $\hat{b}$ is only known to be a distance at most ϵ from the given b: $\hat{b} \in B(b, \epsilon)$ or equivalently, $|b - \hat{b}| < \epsilon$. Then, as we will see in the next chapter on linear maps, if A is invertible, the set of possible solutions will be a solid ellipsoid, centered on $A^{-1}(b)$, whose longest axis is $\frac{\epsilon}{\sigma_n(A)}$ where $\sigma_n(A)$ is the smallest singular value of A. Such an ellipsoid can be expressed as

$$\{x : x^t B x \leqslant c\}$$

for some positive definite symmetric matrix B. Another example comes from statistical learning theory (SLT) where we often encounter something like this:

Theorem 8.1.1 (Caricature of SLT type result)**.** *Suppose that*

1. $X \equiv \{x_i\}_{i=1}^N$ *are independent samples from some (often unknown) probability distribution in* $\mathbb{R}^n$*,*
2. *we are given a function that has some sort of regularity condition (for example,* f *is Lipschitz with Lipschitz constant* k*),*
3. *we know the values of* f *on* X*,*

4 *and that we have constructed an empirical estimator of* f, $\bar{f}_N$ *based on those samples and values.*

Then, the probability P *of finding points where the estimator is off by more than* ϵ *goes to* 0 *as* $N \to \infty$ *because something like:*

$$P\{y : |f(y) - \bar{f}_N(y)| > \epsilon\} < pl(N, \epsilon)e^{-g(N,\epsilon)}$$

is true for a fixed polynomial $pl(N, \epsilon)$ *and function* $g(N, \epsilon)$ *such that*

$$g(N, \epsilon) \underset{N\to\infty}{\to} +\infty$$

for any fixed ϵ*.*

Thus, inequalities allow us to make the level of uncertainty in approximation error precise. Yet another example are the kinds of statements you can prove about convergence of discrete approximations to definite integrals. Suppose that $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ is continuous. Define $\Delta y = \frac{b-a}{N}$

$$\left| \int_a^b f(y)\, dy - \sum_{i=0}^{N-1} f(i\Delta y + a)\Delta y \right| < C_f(b-a)\Delta y = C_f \frac{(b-a)^2}{N} \underset{N\to\infty}{\to} 0$$

where C_f is a fixed constant that depends on f but not on N.

CONVERGENCE TO 0: Often we would like to prove that some set is small. Perhaps this is the set of points where a function is discontinuous or otherwise badly behaved. Very often we are trying to prove that the measure of such a set is zero by proving a sequence of statements like this:

$$\mu(E) \leqslant h(\epsilon_i)$$

where E is the set in question, $\mu(E)$ is the measure of E, $h(x)$ is a positive function converging to 0 as $x \to 0$, and $\epsilon_i > 0$, with $\epsilon_i \underset{i\to\infty}{\to} 0$. (The work required to prove the inequalities can sometimes be considerable.)

Very generally, showing that a discrete approximation G_N converges to some true G (G can be a set or function or measure) aims at proving statements like

$$\rho(G, G_N) \underset{N\to\infty}{\to} 0$$

for some appropriate metric ρ, or equivalently, there is a function $H : \mathbb{N} \to \mathbb{R}$ such that

$$\rho(G, G_N) < H(N) \underset{N \to \infty}{\to} 0.$$

So we accomplish our result by proving a sequence of inequalities.

BOUNDEDNESS: Sometimes everything we want to know about a function or mapping f follows from knowing

$$\sup_{x \in X} |f(x)| < C < \infty$$

for some norm $|\cdot|$ and constant C. Or we might know that $E \subset \mathbb{R}^n$ is closed and proving that E is bounded:

$$\sup_{x \in E} |x| < C < \infty$$

is all we need to show to get that E is compact, after which we can conclude a sequence of convenient properties about continuous functions on E. Or perhaps we have a set of functions F and we want to know that

$$\sup_{f \in F} |f| < C < \infty$$

where $|\cdot|$ is a norm on the space of functions.

CONSTRAINTS AND SUBSET DEFINITION: Very often, in the case in which we are given a function $F : X \to \mathbb{R}$ that we want to maximize or minimize (i.e. we want to optimize), we are only interested in a subset of X, called the feasible set E. Very often this set is defined by an inequality or collection of inequalities:

$$E \equiv \{x \in X : g(x) \leqslant c\}$$

or

$$E \equiv \{x \in X : g_i(x) \leqslant c_i \text{ for } i = 1, ..., N\}.$$

In the case of the classic linear programming case, the g and g_i are linear functions of $x \in \mathbb{R}^n$.

INCLUSION IN SUBSETS: Showing a function is in some space or subset of a space is often equivalent to showing the function or derived property of the function satisfies some inequality. For example, the Banach fixed point theorem (see Remark 8.1.1), stating:

Theorem 8.1.2 (Banach fixed Point)**.** *If (1) $f : \mathcal{B} \to \mathcal{B}$ for some Banach space $\mathcal{B}$ and (2) $|f(x) - f(y)| \leqslant k|x - y|$ for all $x, y \in \mathcal{B}$ and some fixed $0 \leqslant k < 1$ (i.e. f is a contraction mapping), we can conclude that there exists a unique $x^* \in X$ such that*

$$f(x^*) = x^*.$$

can be used to solve various nonlinear problems. And to apply this theorem, the work is often focused on showing that f belongs to the family of contraction mappings, i.e. that

$$|f(x) - f(y)| \leqslant k|x - y|$$

for all $x, y \in \mathcal{B}$ and some fixed $0 \leqslant k < 1$.

Remark 8.1.1. *Any complete, normed, vector space is called a Banach space. For example $\mathbb{R}^n$, for all n, is a Banach space when we use the usual Euclidean norm $|x| = \sqrt{x \cdot x}$ (or actually, when we use any vector norm as the metric).*

EMBEDDING: This is a special version of "Inclusion": here we want to know in what cases one space of functions is actually completely included in another space of functions. We remind ourselves of the definition of L^p functions on a space E with measure μ,

Definition 8.1.1 (L^p **spaces**)**.** *A function $f : E \to \mathbb{R}$ is said to be in $L_p(E)$ if $L_p(f) \equiv \left(\int_E |f|^p\, dx\right)^{\frac{1}{p}} < \infty$. Here we assume $1 \leqslant p \leqslant \infty$ and we get the usable definition of L^∞*

$$L^\infty(f) = \inf\{a \in \mathbb{R} \mid \mu(\{x||f(x)| > a\}) = 0\}$$

by a limiting process.

Now an example of embedding. When the domain of the family of functions is a compact subset $K \subset \mathbb{R}^n$, we know that the family of L_p functions on K, $L_p(K)$, is also in the family of L_1 functions on K, $L_1(K)$. I.e. $L_p(K) \subset L_1(K)$. This inclusion follows immediately from Holder's inequality:

$$\int_E |fg|\, dx \leqslant |f|_p |g|_q$$

where $1 \leqslant p, q \leqslant \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. To see this, we define $g = \chi_E$, the function that is 1 on E and 0 everywhere else. Supposing that $E = K$ is compact and therefore has finite n-dimensional volume, the inequality turns into:

$$\begin{aligned} \int_K |fg| \, dx &= \int_K |f| \, dx \\ &= |f|_1 \\ &\leqslant |f|_p |1|_q \\ &= \text{Vol}(K) |f|_p \end{aligned}$$

So

$$|f|_p < \infty \Rightarrow |f|_1 < \infty$$

i.e.

$$L_p(K) \subset L_1(K).$$

UNCERTAINTY: The area of uncertainty quantification – figuring out how errors in measurement of data effect inferences made from that data – has always been important, but challenging to the point that only after the advent of considerable computing power have many of these questions become at least partly approachable. In general, we measure data like initial temperatures and pressures and displacement and then we run a simulation to figure out some future quantity. Very often we are simulating some partial differential equation, so the discrete approximation is a very high-dimensional problem. In essence, we want to know how the uncertainty – we know our initial point is in some (hopefully small) ϵ-ball in $\mathbb{R}^N$ where N is quite large – evolves under the influence of nonlinear mappings F_k. I.e. denoting composition of G and H by $G \circ H$, what does

$$F_k \circ F_{k-1} \circ \cdots \circ F_1(B(x_0, \epsilon))$$

look like? In particular what is the smallest Δ such that

$$F_k \circ F_{k-1} \circ \cdots \circ F_1(B(x_0, \epsilon)) \subset B(F_k \circ F_{k-1} \circ \cdots \circ F_1(x_0), \Delta)?$$

Equivalently, what is the smallest Δ such that

$$|x_0 - x| < \epsilon \Rightarrow |F_k \circ F_{k-1} \circ \cdots \circ F_1(x_0) - F_k \circ F_{k-1} \circ \cdots \circ F_1(x)| < \Delta$$

All of the above questions involve some metric or metric like measure mapping the sets, functions or measures or pairs of the same, to the real numbers. While it is the case that we compare things that are not real numbers – for example we say that the function f is less than the function g, $f < g$, if for all $x \in \mathbb{R}^n$, $f(x) < g(x)$ – when the objects we are comparing are real numbers, a and b, we know that one of three things is true: (1) $a < b$, (2) $a = b$, or (3) $a > b$. That is, the real numbers are *totally ordered*.

As a result, mastery of inequalities begins with mastery of some simple inequalities in $\mathbb{R}$.

8.1.1 Beginning Facts

Suppose that $a, b, c, d \in \mathbb{R}$. Then:

1. $a < b$ and $b < c \Rightarrow a < c$
2. $a < b \Rightarrow a + c < b + c$
3. $a < b$ and $c > 0 \Rightarrow ac < bc$.
4. $a < b$ and $c < 0 \Rightarrow ac > bc$. In particular, $a < b \Rightarrow -a > -b$.
5. $0 < a < b$ $0 < c < d \Rightarrow ac < bd$.
6. In each of the above lines, replacing all the $<$'s and $>$'s with $\leqslant$'s and $\geqslant$'s generates statements that remain true.

Exercise 8.1.1. Prove that $0 < a < 1 \Rightarrow 1 > a > a^2 > a^3 > ....$

Exercise 8.1.2. Prove that $0 < 1 < a \Rightarrow 1 < a < a^2 < a^3 < ....$

Exercise 8.1.3. Prove that $0 < a < b \Rightarrow a^2 < b^2$.

Exercise 8.1.4. Give an example showing that $a < b \not\Rightarrow a^2 < b^2$.

Exercise 8.1.5. Prove that $0 < a < b \Rightarrow 0 < \frac{1}{b} < \frac{1}{a}$

Exercise 8.1.6. Suppose that $a \leqslant b + \epsilon$ for all $\epsilon > 0$. Prove that $a \leqslant b$.

Exercise 8.1.7. Suppose that $a \geqslant b - \epsilon$ for all $\epsilon > 0$. Prove that $a \geqslant b$.

It is sometimes helpful to think about a particular inequality or collection of inequalities as statements about subsets of $\mathbb{R}$ or conditions defining a subset of the real numbers. Suppose that we know that $a < f(x) < b$ for all $x \in \mathbb{R}$. This is equivalent to the statement that $f(\mathbb{R}) \subset (a, b)$. In the case that $a < f(x) < b$ is not true for all $\mathbb{R}$, $a < f(x) < b$ can be seen as the definition of the set $\{x \mid a < f(x) < b\}$ While this is not always the best way to make sense of a particular inequality, it can be helpful.

There are a few standard ways of proving simple inequalities:

MONOTONICITY $g(x) < f(x) \Rightarrow \int_a^x g(s)\, ds < \int_a^x f(s)\, ds$

NON-NEGATIVITY $(\text{whatever})^2 \geqslant 0$ (as long as whatever is real, i.e. not complex).

SPLITTING INTO PIECES The idea here is that something we want to bound can be split into two or more pieces, each of which can be bounded or controlled, usually by different means. The first subsection of the next section demonstrates this using a simple proof of openness.

LEVERAGING SIMPLE Use a simple fundamental inequality (like the Cauchy-Schwarz inequality or Jensen's inequality) to prove a trickier inequality. We will discuss both of those inequalities in more detail in the next section, but Cauchy-Schwarz (CS) is the one that says $x \cdot y \leqslant |x|\, |y|$ and Jensen's, in it's simplest form, say that $f(\alpha x + (1 - \alpha)y) \leqslant \alpha f(x) + (1 - \alpha) f(y)$ for all convex functions f. In fact it is usually used as the definition of convex functions – a function is convex if it satisfies Jensen's inequality.

SPECIAL CASES Prove the inequality for special cases and then extend: for example, the standard way to prove the CS inequality first assumes that $|x| = |y| = 1$.

On the other hand, it is the *art of inequalities* and art is always a personal, idiosyncratic practice. So you need to collect your own personal bag of tricks through experience, by grappling with a variety of inequalities in your own way. True mastery comes only through a very personal immersion in inequalities.

We begin that process by looking at 10 inequalities.

8.2 A Beginning Repertoire

8.2.1 Triangle Inequality - Open Balls are Open Sets

The previous chapter introduced you to metrics and the triangle inequality. You also had an exercise that asked you to prove that open balls are open. We prove that again here and then discuss the ideas.

We assume we are in a metric space $M = (X, \rho)$. The triangle inequality says that

$$\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$$

for any triple of points $x, y, z \in X$.

Task: Prove again that the open ball $B(x, r)$ – $\{y : \rho(y, x) < r\}$ – is indeed an open set. I.e. prove that for every $y \in B(x, r)$, there is an $s > 0$ such that $B(y, s) \subset B(x, r)$. See Figure 28. This should be an almost instantaneous proof for you by now, but the point of this is to look at it again and notice how inequalities are everything (in this proof).

What we are wanting to prove is that for every point y in the open ball $B(x, r)$, there is a small enough positive number s, so that all the points within a distance of s of y are also within a distance of r of x. I.e there is a small enough s so that:

$$z \in B(y, s) \Rightarrow z \in B(x, r)$$

or

$$\rho(z, y) < s \Rightarrow \rho(z, x) < r$$

or

$$B(y, s) \subset B(x, r)$$

What we have at our disposal – what we control is only one variable – s. We do not have a way of knowing $\rho(z, x)$ directly, but once we have chosen y and s, we know $\rho(y, x)$ and a bound for $\rho(z, y)$.

At this point, we take advantage of the fact that the triangle inequality gives us a bound on the distance of interest – $\rho(z, x)$ – in terms of things we know or have bounds for – $\rho(x, y)$ and $\rho(y, z)$.

$$\rho(z, x) \leqslant \rho(x, y) + \rho(y, z)$$

Having chosen y, we know that $\rho(x, y) < r \to \rho(x, y) = r - \delta$ where $\delta > 0$. Since s is up to us, and can be chosen after we choose y, we choose $0 < s(y) < \delta/2$. This leads immediately to:

$$\rho(z, x) \leqslant \rho(x, y) + \rho(y, z) < r - \delta + \delta/2 = r - \delta/2 < r$$

and this is true for every $z \in B(y, s)$, we get that $B(y, s) \subset B(x, r)$ which is what we were trying to prove.

Reviewing, we took a distance we wanted to bound and split it up into two distances we could control. That theme – splitting things we want to know into two of more pieces, each of which is handled differently, is an very common theme in analysis.

Exercise 8.2.1. Recall Definition (4.2.1) of the length of a curve γ connecting x and y in $\mathbb{R}^2$. Show that the triangle inequality implies that the length of any curve connecting two points is never less than the distance between to the two points. **Hint**: Actually, the exercise following the Definition already solves this problem. See Figure (29).

Exercise 8.2.2. Suppose that we want to minimize the length of a curve in $\mathbb{R}^2$ that connects x and y and intersects a vertical line L, exactly one time. Show that the optimal curve is two line segments whose angles, θ_1 and θ_2, as shown in Figure (30), are equal. **Hint**: keep x fixed and reflect y and the piece of the curve connecting y to L about L to the other side of L. Does this change the length of anything?

Exercise 8.2.3. **(Challenge)** Suppose now that the two points we want to connect with a curve of minimal length are on opposite sides of L, see Figure (31). Suppose that on the left side $\rho(u, w) = a\sqrt{(u-w)\cdot(u-w)}$ and on the right side $\rho(v, z) = b\sqrt{(v-z)\cdot(v-z)}$, where $a > 0$ and $b > 0$ and $a \neq b$.

1 Given x, y and L, find the relationship between θ_1 and θ_2.
2 Use this to derive an equation of the form

$$\frac{c_1}{p^2} = c_2 + \frac{c_3}{(h-p)^2}$$

for the position of the point where the shortest path crosses L. Here $h = |x_2 - y_2|$ is the vertical separation between x and y and $0 \leqslant p \leqslant h$ is the vertical distance of the crossing point from the height of x.

8.2.2 *Cauchy-Schwarz Inequality - $|x|_2$ is a Metric*

The Cauchy-Schwarz inequality* states that in any normed space X with an inner-product (i.e. a Hilbert space, providing it is also complete) we have:

$$x \cdot y \leqslant |x|\,|y|.$$

One proof uses a special case to prove the general case.

First we suppose that $|x| = |y| = 1$. We also use non-negativity of the norm. $|w| \geqslant 0$ for all $w \in X$. The rest is easy:

*While it is true that by now, you are intimately acquainted with this inequality, the Cauchy-Schwarz inequality belongs in every list of 10 initial inequalities!

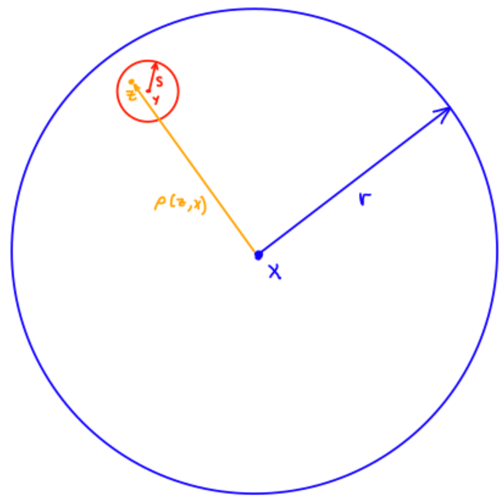


Figure 28: The open ball is open!

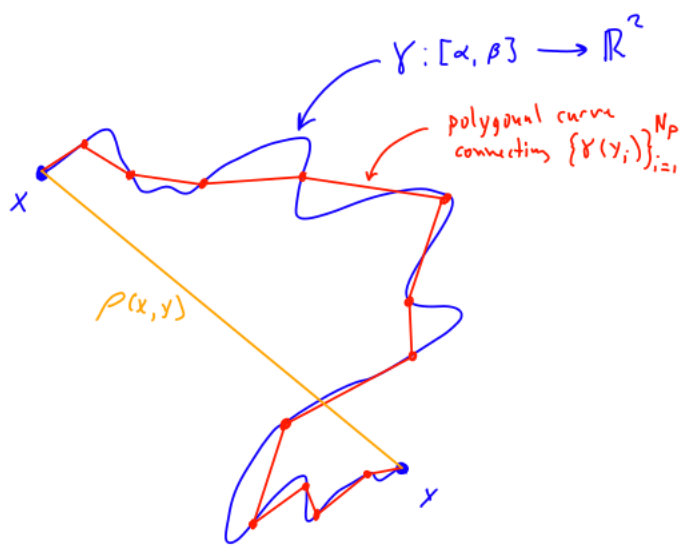


Figure 29: Discrete approximations to the length of a curve. The supremum of such lengths is defined to be the length of the curve.

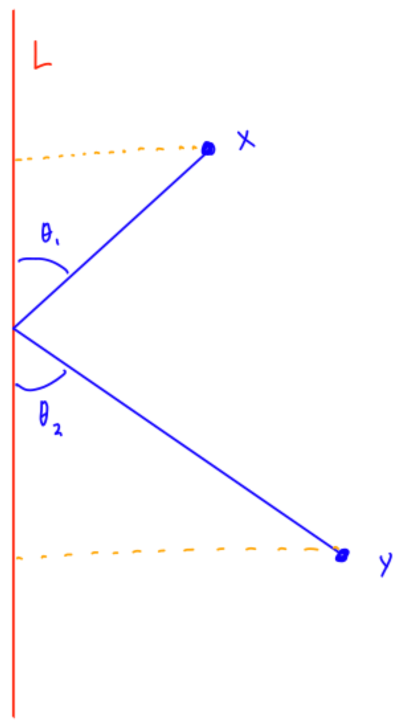


Figure 30: Finding the angle of a reflection – proving that it is in fact the shortest length between two points if you assume it bounces off the surface.

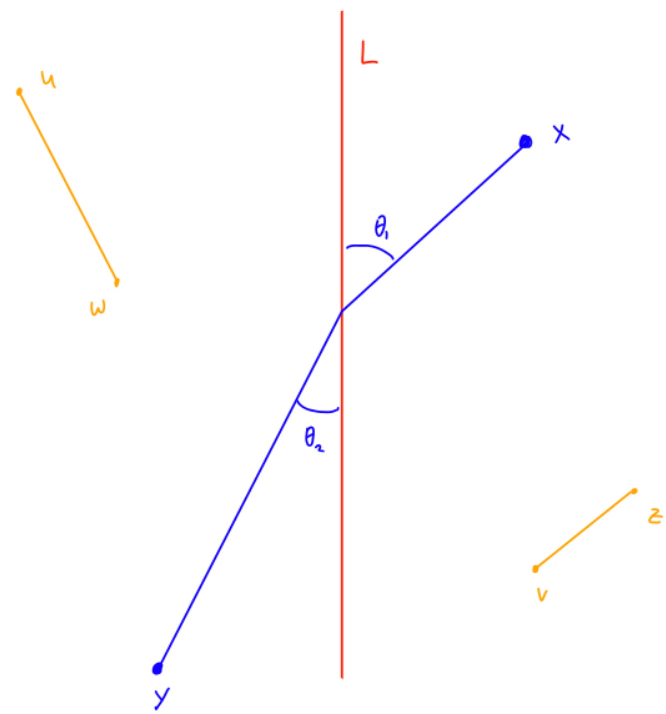


Figure 31: Showing that different metrics on either side leads to bent shortest paths – just like the phenomena of refraction!

$$|x-y|^2 \geqslant 0 \Rightarrow (x-y)\cdot(x-y) = x\cdot x + y\cdot y - 2x\cdot y \geqslant 0$$

from which we conclude that

$$2 \geqslant 2x \cdot y$$

or

$$x \cdot y \leqslant 1.$$

In the general case (assuming that neither x nor y is 0, in which case the result is immediate,) when x and y do not both have norm 1, we apply the result to $u = \frac{x}{|x|}$ and $w = \frac{y}{|y|}$ to get

$$u \cdot w \leqslant 1 \Rightarrow \frac{x}{|x|} \cdot \frac{y}{|y|} \leqslant 1$$

implying the desired result.

Reviewing, we used both a special case and non-negativity to get the desired result.

Exercise 8.2.4. Use the Cauchy-Schwarz inequality to prove the triangle inequality for the Euclidean distance in $\mathbb{R}^n$. Hint: Use this form of the triangle inequality in normed spaces: $|x+y| \leqslant |x| + |y|$ and start by convincing the reader of your proof that this is indeed the triangle inequality in normed spaces, i.e. in spaces where $\rho(u,v) = |u-v|$ for some norm $|\cdot|$.

Exercise 8.2.5. Recall from Exercise (4.3.1) that the rotation matrix R_θ rotates vectors in $\mathbb{R}^2$ by θ. Show that $x \cdot y = x^t R_\theta^T R_\theta y$ and use this fact to show that for $x, y \in \mathbb{R}^2$, $x \cdot y = \cos(\theta)|x|\,|y|$.

Exercise 8.2.6. Prove that the distance, on the unit sphere in $\mathbb{R}^3$, defined by $\rho(x,y) = \arccos(x \cdot y)$, satisfies the triangle inequality for $x, y \in \partial B(0,1) \subset \mathbb{R}^3$. **Hint**: Sketch what this distance is, geometrically on the unit sphere. Recall that arccos takes inputs in $[-1,1]$ and spits out $\theta \in [0,\pi]$.

8.2.3 *Jensen's Inequality - AM-GM Inequality*

A convex function is any function $f : X \to \mathbb{R}$ mapping X to $\mathbb{R}$ that satisfies Jensen's inequality:

Definition 8.2.1 (Convex Function). *A function $f : X \to \mathbb{R}$ is a convex function if*

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y)$$

for all $x, y \in X$ and all $\alpha \in [0, 1]$

Remark 8.2.1. *Note that the domain of a convex function need only be a vector space. We will usually be working in $\mathbb{R}^n$ and from this point on in the subsection, we will be working with functions whose domains are $\mathbb{R}^n$ or subsets of $\mathbb{R}^n$ for some n.*

Definition 8.2.2 (Convex Set). *A set E is convex if, for any two points $x, y \in E$ and all $\alpha \in [0, 1]$, $\alpha x + (1-\alpha)y \in E$.*

The theory of convex functions and convex sets is both very useful and interesting. We will explore a bit of this in the exercises now.

We will need a couple of definitions.

Definition 8.2.3 (Epigraph). *The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set*

$$\text{epi}(f) \equiv \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \ : f(x) \leqslant y\}$$

Exercise 8.2.7. Prove that $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the $\text{epi}(f) \subset \mathbb{R}^{n+1}$ is convex.

Exercise 8.2.8. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then

$$f(\sum_i^N \alpha_i x_i) \leqslant \sum_i^N \alpha_i f(x_i)$$

for all $\alpha \in \mathbb{R}^N$ such that $\alpha_i \geqslant 0$ for all i and $\sum_i \alpha_i = 1$.

Exercise 8.2.9. **(Challenge)** Suppose that $f : \mathbb{R} \to \mathbb{R}$ is convex, $\phi : \mathbb{R}^n \to \mathbb{R}$ is continuous, $\phi(x) \geqslant 0$ for all $x \in \mathbb{R}^n$, $\phi = 0$ for $|x| > R$ for some $R < \infty$, and $\int_{\mathbb{R}^n} \phi(x)dx = 1$. Show that

$$f\left(\int_{\mathbb{R}^n} x\phi(x)\, dx\right) \leqslant \int_{\mathbb{R}^n} f(x)\phi(x)\, dx$$

. Hint use approximation – you can discretize the integral and use Exercise (8.2.8) to get an approximate result. Careful work in this direction gets you the result.

Definition 8.2.4 **(Concave Function).** *A function* $f : X \to \mathbb{R}$ *is a concave function if*

$$f(\alpha x + (1-\alpha)y) \geqslant \alpha f(x) + (1-\alpha)f(y)$$

for all $x, y \in X$ *and all* $\alpha \in [0, 1]$

Exercise 8.2.10. **(AM-GM Inequality)** Use the fact that $\log : \mathbb{R} \to \mathbb{R}$ is a concave function to get that

$$\sum_i^N \alpha_i x_i = \alpha_1 x_1 + \alpha_2 x_2 ... + \alpha_N x_N \geqslant x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_N^{\alpha_N} = \Pi_i^N \alpha_i x_i^{\alpha_i}$$

for any $\{x_i\}_{i=1}^N \subset \mathbb{R}$ and all $\alpha \in \mathbb{R}^N$ such that $\alpha_i \geqslant 0$ for $i = 1, ..., N$ and $\sum_i \alpha_i = 1$.

Remark 8.2.2. *"AM-GM" stands for "Arithmetic Mean - Geometric Mean". The (generalized) arithmetic mean of* $\{x_i\}_{i=1}^N \subset \mathbb{R}$ *is* $\sum_i^N \alpha_i x_i$ *and the (generalized) geometric mean is* $\Pi_i^N x_i^{\alpha_i}$. *They are generalized because all we require is that* $\sum_{i=1}^N \alpha_i = 1$ *and* $\alpha_i \geqslant 0\ \forall i$, *whereas the usual arithmetic and geometric means result when* $\alpha_i = \frac{1}{N}$ *for all* i.

8.2.4 $1 + x \leqslant e^x$ - *learning bounds*

Plotting the functions $y = f(x) = 1 + x$ and $y = g(x) = e^x$ makes it clear that $f(x) \leqslant g(x)$ for all $x \in \mathbb{R}$. To prove this analytically, we make use of the following fact:

Theorem 8.2.1. *Suppose that* $h_1(x_0) \leqslant h_2(x_0)$ *and* $\frac{dh_1(x)}{dx} \leqslant \frac{dh_2(x)}{dx}$ *for all* $x_0 \leqslant x$ *and* $\frac{dh_1(x)}{dx} \geqslant \frac{dh_2(x)}{dx}$ *for all* $x \leqslant x_0$*. Then* $h_1(x) \leqslant h_2(x)$ *for all* $x \in \mathbb{R}$.

Exercise 8.2.11. Prove Theorem (8.2.1). **Hint**: use monotonicity both for $x \geqslant x_0$ and $x \leqslant x_0$.

Exercise 8.2.12. Use Theorem (8.2.1) to prove that

$$1 + x \leqslant e^x \quad \forall x \in \mathbb{R}.$$

See Figure (32).

Exercise 8.2.13. Show that for $x \in (-\infty, 0.5)$ we have:

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

Hint: split the argument into two pieces – one for $x \in (-\infty, 0)$ and one for $x \in [0, 0.5]$. Now reason carefully with inequalities involving the derivatives.

Exercise 8.2.14. Now conclude that for $|x| < 0.5$ (actually $|x| < \ln(2)$ also works), we have:

$$\begin{aligned} &|e^x - (1+x)| \leqslant x^2 \\ &1 \leqslant \frac{e^x}{1+x} \leqslant 1 + \frac{x^2}{1+x} \\ &1 \leqslant \frac{e^{nx}}{(1+x)^n} \leqslant \left(1 + \frac{x^2}{1+x}\right)^n. \end{aligned}$$

Learning a set from samples: Suppose now that we want to recover a set E based on samples of that set.

1 Assume that N is the number of samples and we have divided the set up into m equal probability pieces. That is, we have used the uniform distribution to cut E into m equal size pieces. The idea now is that we want to know what happens as N and m increase. We will assume that both $k \equiv \frac{N}{m} \to \infty$ and $m \to \infty$.

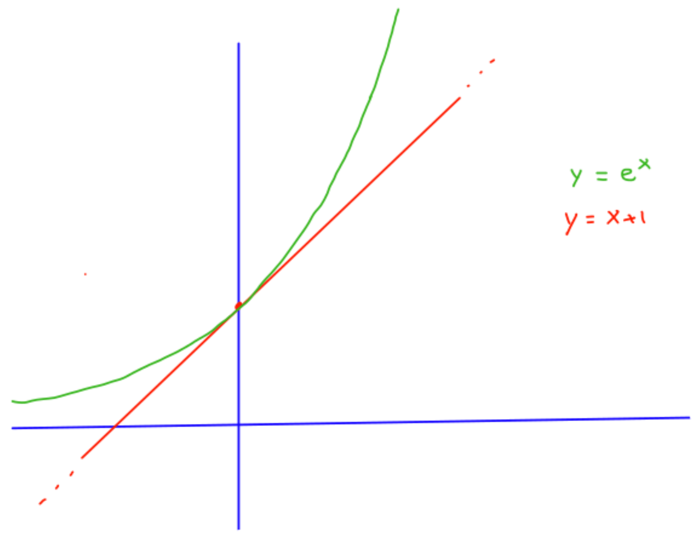


Figure 32: Picturing $1 + x \leqslant e^x$.

2 Notice that the probability of one particular piece never generating a sample is

$$\left(1 - \frac{1}{m}\right)^N = \left(1 - \frac{1}{m}\right)^{mk} \leqslant \left(e^{-\frac{1}{m}}\right)^{mk} = e^{-k}$$

3 if we union over all of the pieces – we get a union bound saying:

$$\text{Prob}(\{\text{missing at least one of the } m \text{ pieces}\}) \leqslant me^{-k}$$

4 Because we can choose any k we want (we can choose how m and N increase!), we choose $k = m$. Put differently, we can sample the set m^2 times when there are m pieces.

5 This leads to the statement that

$$\begin{aligned} \text{Prob}(\{\text{missing at least one of the } m \text{ pieces}\}) &\leqslant me^{-m} \\ &\underset{m \to \infty}{\to} 0 \end{aligned}$$

Exercise 8.2.15. Supposing it is possible to divide up the set E into m equal probability pieces such that the diameter of every piece is bounded by $\delta > 0$. Prove that the balls $B(x_i, \delta)$, centered on the samples $\{x_i\}_{i=1}^N$, will cover the original set with probability at least $1 - me^{-m}$. See Figure (33).

Exercise 8.2.16. **(Exploration)** Suppose that $E \subset \mathbb{R}^2$ is connected, and that the area of E, $A(E)$, is $1 - A(E) = 1$.

SORT OF OPTIMAL? Find an example of a set in $E \subset \mathbb{R}^2$, that we can divide into $10,000$ pieces, each with as area of $\frac{1}{10,000}$ and each contained in a ball of radius $\frac{\sqrt{2}}{200}$. **Hint**: try nice, easy-to-divide-up sets in $\mathbb{R}^2$.

OPTIMAL Suppose now that your set E is the union of $10,000$ closed balls that intersect only on their boundaries, each of radius $\frac{1}{100\sqrt{\pi}}$. Show that the obvious partition into $10,000$ pieces of equal area is an optimally small cover ... i.e., for any connected set of area 1 that you cut into $10,000$ pieces of equal area, you cannot get a cover with $10,000$ covering sets, the maximal diameter of which is smaller than $\frac{2}{100\sqrt{\pi}}$. **Hint:** You will need the fact (called the *isodiametric inequality*) that, of all sets with some fixed diameter D, none can contain more volume (in our case, 2-dimensional area) than a ball with diameter D.

LOW DENSITIES Suppose now that if $x \in E$, then $\frac{A(B(x,r)\cap E}{\pi r^2} < \frac{1}{100}$ for all $r \geqslant \delta$. Show that if $m = 10,000$ and $\delta = \frac{1}{101\sqrt{\pi}}$, then the smallest diameter of a cover made up if $10,000$ balls, is at least $\frac{1}{10\sqrt{\pi}}$.

8.2.5 Isoperimetric Inequality - Concentration of Measure in High Dimensions

You have perhaps heard of the classical isoperimetric ratio – the smallest perimeter of a region in the plane with a fixed area A is the perimeter of a circle that has area equal to A: I.e. if a $E \subset \mathbb{R}^2$ has

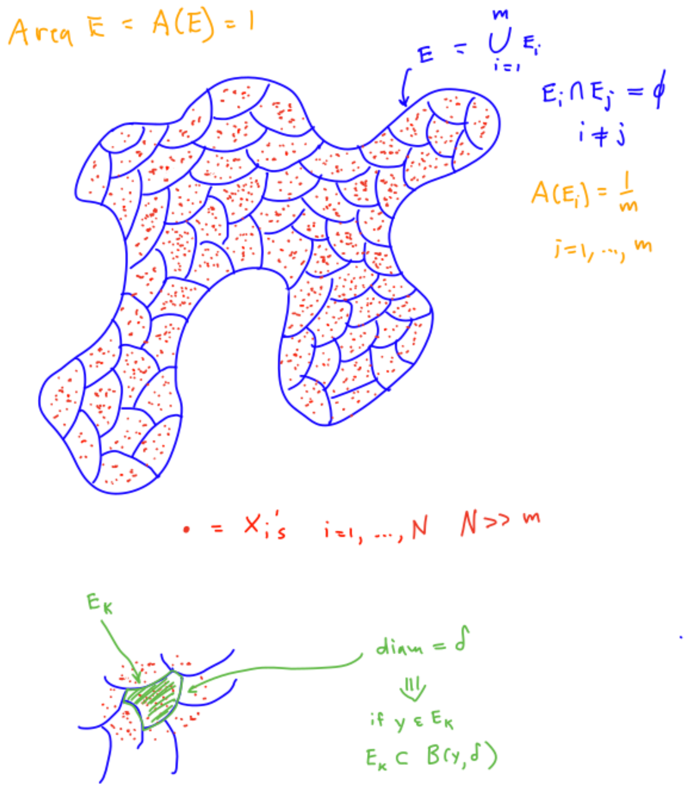


Figure 33: Reconstructions from Samples: If we sample N points from a partition with m elements, $N >> m$, the probability of getting a cell with no samples in it is very small. If the diameter of the partition elements is bounded by δ, then the balls of radius δ centered on samples cover the original set with high probability.

area A, the smallest it's perimeter can be is $P_{\min}(A) = 2\pi\sqrt{\frac{A}{\pi}} = 2\sqrt{\pi A}$. Stating this as an inequality:

$$\frac{P^2(A)}{A} \geqslant 4\pi$$

This is true in great generality, but for our purposes we will just use its generalization to $\mathbb{R}^n$ and to sets on the unit sphere in $\mathbb{R}^n$:

Though we will introduce it in careful detail later, you have already seen Hausdorff measure in Section (4.2.5) in the special case of the dimension being 1. We have also referred to higher dimensional versions $\mathcal{H}^k$. In general, the k-dimensional measure of a set $E \subset \mathbb{R}^n$, $k \leqslant n$ is $\mathcal{H}^k(E)$. It is a theorem that for $E \subset \mathbb{R}^n$, $\mathcal{H}^n(E) = \mathcal{L}^n(E)$ = usual volume in n-dimensions. (Part of the proof of this theorem is quite challenging.)

Suppose that $B_n(r) \equiv B(0, r) \subset \mathbb{R}^n$. Define $\alpha_n \equiv \mathcal{H}^n(B_n(1))$. It follows that $\mathcal{H}^n(B_n(r)) = \alpha_n r^n$.

Exercise 8.2.17. By thinking about how the volumes of cubes change when you scale them from a side length of 1 to a side length of r, convince yourself that $\alpha_n \equiv \mathcal{H}^n(B_n(1))$ implies that $\mathcal{H}^n(B_n(r)) = \alpha_n r^n$. **Hint:** think of using tiny cubes to tile the ball.

Now we compute to get the $n-1$ dimensional volume of the unit sphere:

$$\begin{aligned}
P_n(r)|_{r=1} &\equiv \mathcal{H}^{n-1}(\partial B_n(r))\big|_{r=1} \\
&= \left.\frac{d\mathcal{H}^n(B_n(r))}{dr}\right|_{r=1} \quad \text{(See Exercise 8.2.19)} \\
&= n\alpha_n r^{n-1}\big|_{r=1} \\
&= n\alpha_n
\end{aligned}$$

It follows that for a ball of radius r in $\mathbb{R}^n$, the perimeter $P_n(r) = n\alpha_n r^{n-1}$.

We can state the two generalizations. We define $\mathbb{S}^{n-1} \equiv \partial B_n(1)$

ISOPERIMETRIC INEQUALITY IN $\mathbb{R}^n$

$$\frac{(\mathcal{H}^{n-1}(\partial E))^{\frac{n}{n-1}}}{\mathcal{H}^n(E)} \geqslant \frac{(\mathcal{H}^{n-1}(\partial B_n(r)))^{\frac{n}{n-1}}}{\mathcal{H}^n(B_n(r))} \tag{4}$$

$$= \frac{(n\alpha_n)^{\frac{n}{n-1}}}{\alpha_n} \tag{5}$$

$$= (n^n\alpha_n)^{\frac{1}{n-1}} \tag{6}$$

ISOPERIMETRIC INEQUALITY ON $\mathbb{S}^{n-1}$ For this inequality, we need to define Balls on the unit sphere $\mathbb{S}^{n-1}$.

STEP 1 Choose $x \in \mathbb{S}^{n-1}$. By $B^{\mathbb{S}}(x,r) \subset \mathbb{S}^{n-1}$ we mean all the points $y \in \mathbb{S}^{n-1}$ such that the shortest path from x to y in $\mathbb{S}^{n-1}$ is less than r in length.

STEP 2 Because $\mathcal{H}^{n-1}(B^{\mathbb{S}}(x,2r)) \neq 2^{n-1}\mathcal{H}^{n-1}(B^{\mathbb{S}}(x,r))$ the result we get is similar, but not identical, since the optimal ratio is not independent of the radius of the ball.

RESULT So, the best we can do is this: Suppose that $E \subset \mathbb{S}^{n-1}$ and $B^{\mathbb{S}}(x,r) \subset \mathbb{S}^{n-1}$. Then

$$\mathcal{H}^{n-1}(E) = \mathcal{H}^{n-1}(B^{\mathbb{S}}(x,r)) \tag{7}$$

$$\text{implies} \tag{8}$$

$$\mathcal{H}^{n-2}(\partial E) \geqslant \mathcal{H}^{n-2}(\partial B^{\mathbb{S}}(x,r)) \tag{9}$$

where, of course, the boundaries ∂E and $\partial B^{\mathbb{S}}(x,r)$ are the boundaries in $\mathbb{S}^{n-1}$.

Exercise 8.2.18. Define the distance from a point x to a set $E \subset \mathbb{R}^n$, to be

$$d(x,E) \equiv \inf_{y \in E} |x-y|.$$

Prove that if E is closed, there there is a point $y^* \in E$ such that $|x-y^*| = d(x,E)$, implying that the infimum is actually a minimum.

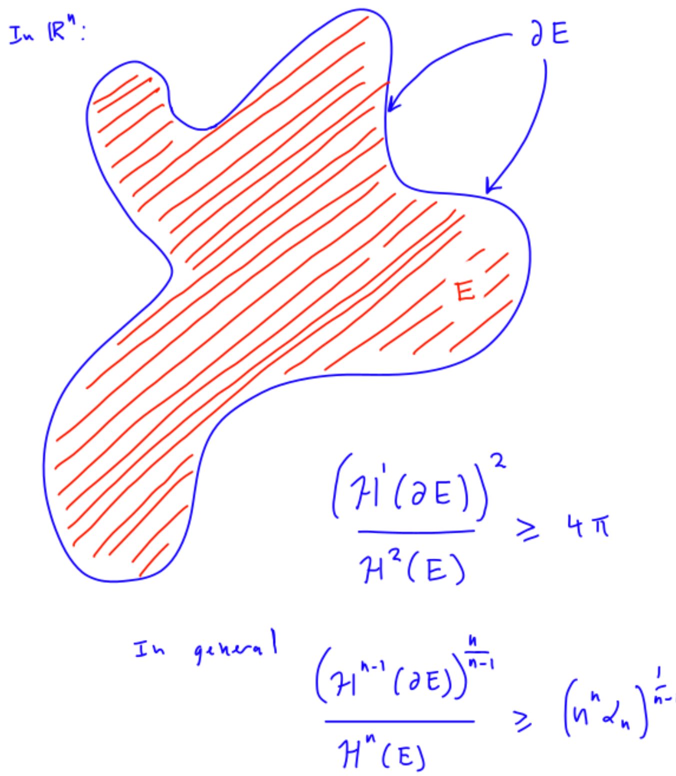


Figure 34: The isoperimetric inequality in $\mathbb{R}^n$.

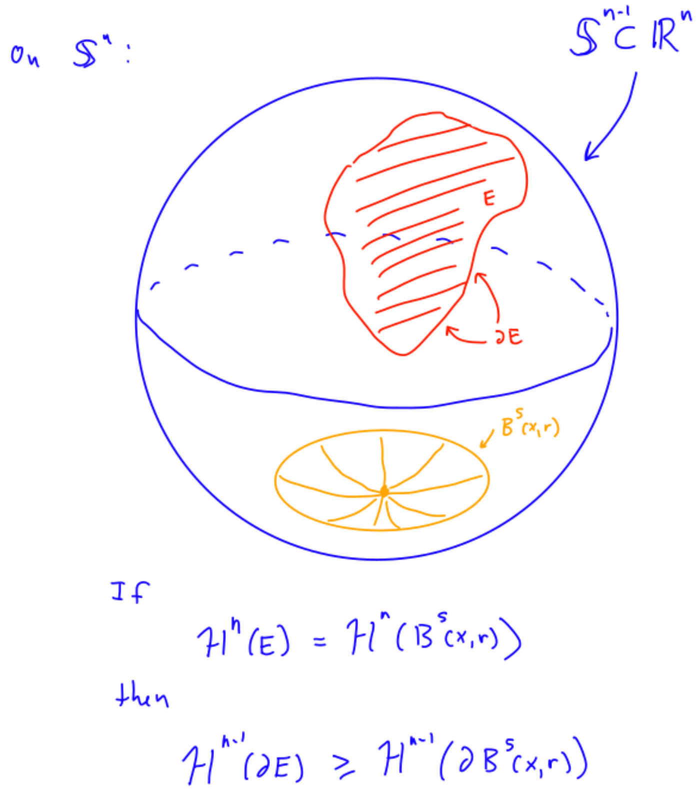


Figure 35: The isoperimetric inequality in $\mathbb{S}^{n-1} \subset \mathbb{R}^n$.

Exercise 8.2.19. Convince yourself, by playing around with sets in the plane, that if

$$E_s \equiv \{x \in \mathbb{R}^2 \ : \ d(x, E) \leqslant s\}$$

then $\frac{d\mathcal{H}^2(E_s)}{ds}\Big|_{s=r} = \mathcal{H}^1(\partial E_r)$ and in particular that $\frac{d\mathcal{H}^2(E_s)}{ds}\Big|_{s=0} = \mathcal{H}^1(\partial E)$. After convincing yourself, write up why you are convinced, in a proof – this can be done geometrically.

Exercise 8.2.20. **(Challenge)** Assume this is true in higher dimensions and on the sphere – i.e. that

$$\frac{d\mathcal{H}^k(E_s)}{ds}\Big|_{s=r} = \mathcal{H}^{k-1}(\partial E_r)$$

for sets $E \subset \mathbb{R}^k$ and $E \subset \mathbb{S}^k$. Now suppose that $E \subset \mathbb{S}^k$ and $\mathcal{H}^k(E) = \mathcal{H}^k(B^{\mathbb{S}}(0, r))$ where 0 is some fixed point in $\mathbb{S}^k$. Use the isoperimetric inequality on the sphere, Equations (7-9), Theorem (8.2.1) and great care(!) to show that $\mathcal{H}^k(E_s) \geqslant \mathcal{H}^k(B^{\mathbb{S}}(0, r+s))$ for all s.

Now we will use the results of the exercises to get a very interesting result on high dimensional spheres. First we note that for any fixed $\epsilon > 0$:

$$\frac{\mathcal{H}^n(B(0, 1-\epsilon))}{\mathcal{H}^n(B(0, 1))} = (1-\epsilon)^n \leqslant e^{-\epsilon n} \underset{n\to\infty}{\to} 0.$$

The geometric interpretation of this is that the outer ϵ-thick shell/layer of the ball contains basically all the volume of the ball as the dimension gets really large.

We can use this to show that the ϵ-neighborhood of any great circle on the unit sphere contains almost all the surface area of the sphere. This is the basic insight, exploited in high dimensional settings in all sort of useful ways and is the fundamental *concentration of measure* result.

The idea is shown in Figures (36-37).

We start by observing that we can "bulge up" a horizontal n-1 dimensional ball (disk) in $\mathbb{R}^n$ to get the upper hemisphere of $\mathbb{S}^{n-1} \subset \mathbb{R}^n$.

We know that the s corresponding to the ϵ (see the figures!) satisfies $1 - \epsilon = \cos(s)$. Using the inequality from Exercise (8.2.21), we get that

$$(1-\epsilon)^{n-1} \leqslant e^{-\frac{s^2}{2}(n-1)} \tag{10}$$

Now we can make the statement we were aiming for:

Theorem 8.2.2 (Concentration of Measure on $\mathbb{S}^{n-1}$)**.** *Suppose that E_s is the s-neighborhood of some great circle C on $\mathbb{S}^{n-1}$ and $\mu_{\mathbb{S}^{n-1}}$ is the uniform probability measure on $\mathbb{S}^{n-1}$. Then:*

$$\mu_{\mathbb{S}^{n-1}}(E_s) \geqslant 1 - e^{-\frac{s^2}{2}(n-1)}$$

Proof. We use the idea in Figures (36-37).

1 Observe that the fraction of the sphere contained in the s-neighborhood of F is the same as the fraction of the upper hemisphere contained in the s-neighborhood of F: $1 - \frac{C}{C+D}$.
2 Geometrically, it is clear that $1 - \frac{C}{C+D} > 1 - \frac{B}{A+B} = 1 - (1-\epsilon)^{n-1}$
3 Using Equation (10) we have that
4 $1 - (1-\epsilon)^{n-1} \geqslant 1 - e^{-\frac{s^2}{2}(n-1)}$.

□

Exercise 8.2.21. Show that for $0 \leqslant u \leqslant \frac{\pi}{2}$, $\cos(u) \leqslant e^{-\frac{u^2}{2}}$. **Hint**: write out the series for each remove the first two that are the same in each and show that, by collecting subsequent pairs, each pair of terms in the series for $e^{-\frac{u^2}{2}}$ is bigger than the corresponding pair of terms in the series for $\cos(u)$.

Exercise 8.2.22. Use Theorem (8.2.2) to show that as the dimension gets very large, the probability that two randomly chosen unit vectors are almost orthogonal is close to 1. What does s getting small in the Theorem mean, geometrically? **Hint**: choose the first unit vector randomly and then focus your attention on the great circle defined by that vector (there is only one this vector defines uniquely!). Now choose the second vector randomly. By choosing randomly, I mean according to the uniform probability measure on the sphere.

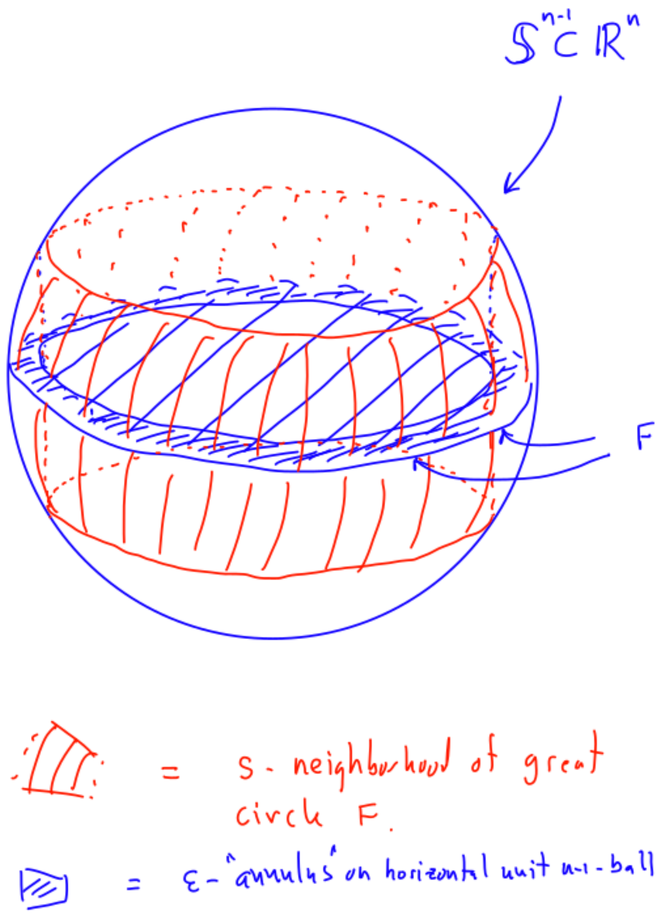


Figure 36: Concentration of measure, illustration 1: the s-neighborhood of a great circle.

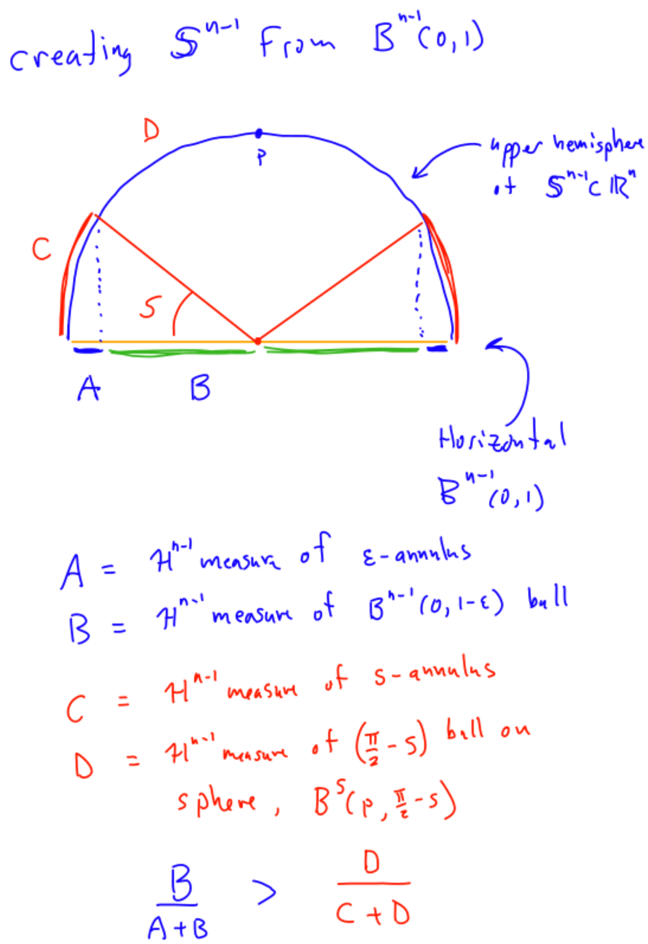


Figure 37: Concentration of measure, illustration 2: aids to understanding the proof of Theorem (8.2.2).

8.2.6 Chebyshev Inequality - Concentration type bounds

We begin with Markov's inequality:

Theorem 8.2.3 (Markov Inequality). *Suppose that* $\int_{\mathbb{R}^n} |f(y)|\, d\mu(y) < \infty$ *and* $t > 0$*. Then we can conclude that*

$$\mu(\{x \,:\, |f| > t\}) \leqslant \frac{\int_{\mathbb{R}^n} |f| d\mu}{t}.$$

the special case in which μ *is a probability measure –* $\mu(\mathbb{R}^n) = 1$ *– is worth pointing out, using the language of probability theory:*

$$\textit{Prob}\{x \,:\, |f| > t\} \leqslant \frac{\mathbb{E}(|f|)}{t}$$

Proof. The proof is very simple. See Figure (38).

1 Notice that on $\{x \,:\, |f|\, d\mu > t\}$, which mean that, well, $|f| > t$ which in turn implies

$$\int_{\{x\,:\,|f|>t\}} |f|\, d\mu \geqslant t\mu(\{x \,:\, |f| > t\}).$$

2 And of course

$$\int_{\mathbb{R}^n} |f|\, d\mu \geqslant \int_{\{x\,:\,|f|>t\}} |f|.$$

3 So we get that $t\mu(\{x \,:\, |f| > t\}) \leqslant \int_{\mathbb{R}^n} |f| d\mu$ which rearranges into our result.

4 The probability result is exactly the same statement, because Expectations are just integrals and measures are probabilities.

□

Recall that for a probability measure ρ on $\mathbb{R}^n$, we define the mean to be

$$\mu \equiv \mathbb{E}(x) = \int_{\mathbb{R}^n} x\, d\rho$$

and the variance to be

$$\sigma^2 = \int_{\mathbb{R}^n} |x - \mu|^2 \, d\rho.$$

Theorem 8.2.4 (Chebyshev Inequality)**.** *Suppose that the probability measure ρ has finite mean μ and variance σ^2. Then*

$$\rho(\{x \,:\, |x - \mu| > k\sigma\}) \leqslant \frac{1}{k^2}$$

Proof. This is an immediate application of Markov's inequality after you note that

$$\{x \,:\, |x - \mu| > k\sigma\} = \{x \,:\, |x - \mu|^2 > k^2\sigma^2\}.$$

□

Exercise 8.2.23. **(Markov Inequality Tests)** For each of the following functions that map $\mathbb{R}$ to $\mathbb{R}$, compute the bounds given by the Markov inequality on sizes of the superlevel sets and then compute the actual size of those sets and compare. Assume the usual measure 1-dimensional Lebesgue is the measure integrated against in the integrals implied in the problem statement.

1. $f(x) = 1$ when $x \in [-1, 1]$ and $f(x) = 0$ elsewhere.
2. $f(x) = \frac{1}{x^2}$ for $|x| > 1$ and $f(x) = 1$ when $|x| \leqslant 1$.
3. $f(x) = \frac{1}{x^2}$ for $|x| > 1$ and $f(x) = \frac{1}{\sqrt{|x|}}$ when $|x| \leqslant 1$.
4. $f(x) = e^{-|x|}$.

8.2.7 Hölder's Inequality - $L_1 \Rightarrow L_p$ for bounded domains

Definition 8.2.5 (Conjugate Exponents). *Real $1 \leqslant p, q \leqslant \infty$ are conjugate exponents or simply Conjugate if*

$$\frac{1}{p} + \frac{1}{q} = 1$$

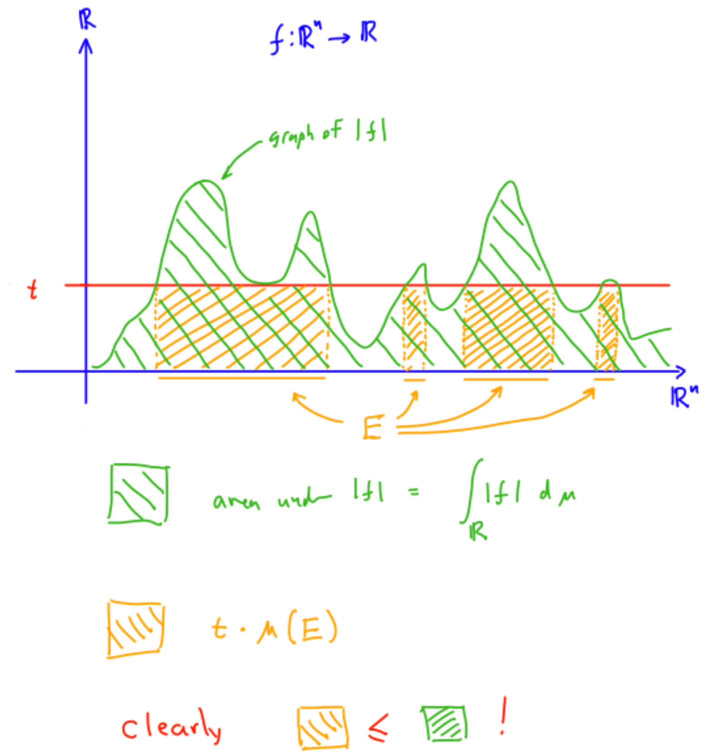


Figure 38: Geometric proof of Theorem (8.2.3).

Lemma 8.2.1 (Young's Inequality)**.** *Suppose that* $a, b \in \mathbb{R}$ *are non-negative and* p *and* q *are conjugate. Then*

$$\frac{a^p}{p} + \frac{b^q}{q} \geqslant ab$$

Proof. Apply the AM-GM inequality – $\frac{x}{p} + \frac{y}{q} \geqslant x^{\frac{1}{p}} y^{\frac{1}{q}}$ – to the quantities $x = a^p$ and $y = b^q$. □

Theorem 8.2.5 (Hölder's Inequality)**.** *Suppose that* $f, g : \mathbb{R}^n \to \mathbb{R}$, p *and* q *are conjugate and both* $f \in L^p$ *and* $g \in L^q$, *i.e. that*

$$|f|_p \equiv \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}} < \infty \text{ and } |g|_q \equiv \left(\int |g|^q \, d\mu\right)^{\frac{1}{q}} < \infty$$

Then

$$\int |fg| d\mu \leqslant |f|_p |g|_q$$

Remark 8.2.3 (p-norm)**.** $|f|_p$ *where* $1 \leqslant p \leqslant \infty$ *is called the* p *norm of* f. *Note that in the case of* $p = \infty$ *we have to be careful. The result is that* $|f|_\infty$ *is the essential supremum – the infimum of values* U, *such that* $\mu(\{x : f(x) > U\}) = 0$.

Proof. We use Young's inequality and begin by assuming that $|f|_p = |g|_q = 1$:

1 Observe that Young's inequality gives us that $|f(x)g(x)| \leqslant \frac{|f(x)|^p}{p} + \frac{|g(x)|^q}{q}$ for all $x \in \mathbb{R}^n$.
2 Now integrate the inequality to get:

$$\begin{aligned} \int |fg| d\mu &\leqslant \frac{1}{p}\int |f|^p \, d\mu + \frac{1}{q}\int |g|^q \, d\mu \\ &= \frac{1}{p} + \frac{1}{q} \\ &= 1 (= |f|_p |g|_q) \end{aligned}$$

3 In the case in which at least one of $|f|_p$ or $|g|_q$ is not equal to one, we apply the result we just obtained to $v = \frac{f}{|f|_p}$ and $u = \frac{g}{|g|_q}$ yielding:

$$\frac{1}{|f|_p |g|_q} \int |fg| d\mu \leqslant 1$$

which in turn is a rearrangement of the result we were aiming for.

$\square$

Exercise 8.2.24. Use the AM-GM inequality to prove Schwarz' inequality for functions.

Exercise 8.2.25. Notice that Young's inequality only makes sense when $p, q < \infty$. But Hölder's inequality is true if one of p or q is infinity and the other is 1. **Show Hölder's in the $p = 1$, $q = \infty$ case.** To simplify this exercise, consider $|f|_\infty$ to simply be the maximum value that f attains in $\mathbb{R}^n$. (**Note**: Reminder, as mentioned above, $|f|_\infty$ is actually the essential supremum – the infimum of values U, such that $\mu(\{x : f(x) > U\}) = 0$.)

Exercise 8.2.26. Suppose that for $x \in \mathbb{R}^n$ we define

$$|x|_p \equiv \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

Show that if p and q are conjugate these discrete versions of the p and q norms also satisfy the Hölder inequality.

Exercise 8.2.27. **(Challenge)** Show that for $x \in \mathbb{R}^n$

$$|x|_\infty \equiv \max_{1 \leqslant i \leqslant n} |x_i| = \lim_{p \to \infty} |x|_p.$$

Exercise 8.2.28. Draw the unit balls in $\mathbb{R}^2$ for each of these discrete norms: (1) $|x|_1$, (2) $|x|_2$, (3) $|x|_\infty$.

8.2.8 *Little o, Big O, and Derivatives - the cone condition for derivatives*

We will dive into derivatives in much more detail in a later chapter, but we begin here by giving an alternate definition of derivative that is much more powerful than the usual definition you usually see in the first course in calculus, because it generalizes very easily. First we recall the notion of "little o" (first seen in Chapter 7) and introduce the notion of "big O".

Definition 8.2.6 (**Recalling little** o**, introducing big** O)**.** *We say say that* $g : h \in \mathbb{R}^n \to \mathbb{R}^m$ *is in "little o(h)" if*

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

We say that $g : h \in \mathbb{R}^n \to \mathbb{R}^m$ *is in "big O(h)" if for some* $\delta > 0$

$$\frac{|g(h)|}{|h|} \leqslant C < \infty \text{ for } |h| < \delta$$

Remark 8.2.4. *Of course the* $|\cdot|$ *indicate whatever norm is being used on* $\mathbb{R}^n$ *and* $\mathbb{R}^m$*. And these definitions are exactly the same when* $g : X \to Y$ *and* X *and* Y *are any two normed vector spaces, even infinite dimensional spaces. We will also write* $g \in o(h)$ *to indicate that "g(h) is in little o(h)".*

Definition 8.2.7 (**Linear Maps**)**.** *Recall that a map from one vector space to another,* $f : X \to Y$ *(for example the* X *and* Y*'s can be* $\mathbb{R}^n$ $n = 1, 2, ...$*) is linear if*

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \ \ \forall x, y \in X \text{ and } \forall \text{ scalars } \alpha \text{ and } \beta.$$

We now give the definition of a derivative in the simplest case: when $f : \mathbb{R} \to \mathbb{R}$.

Definition 8.2.8 (**Derivative**)**.** *Suppose that* $f : \mathbb{R} \to \mathbb{R}$*. We say that* f *is differentiable at* x *if there is a linear map* $A_x : \mathbb{R} \to \mathbb{R}$ *such that*

$$f(x + h) - f(x) - A_x(h) = g(h) \text{ and } g \in o(h)$$

If f is differentiable, we call A_x the derivative of f at x

Exercise 8.2.29. Prove that the only linear maps $L : \mathbb{R} \to \mathbb{R}$ are maps of this form $y = L(x) = ax$ for some real a

Exercise 8.2.30. Show that the above definition of derivative is actually the definition you learned in Calculus I, in disguise.

Exercise 8.2.31. Show that $g \in o(h)$ if and only if, for any $\epsilon > 0$, there is a $\delta(\epsilon) > 0$ such that

$$|h| < \delta(\epsilon) \Rightarrow |g(h)| < \epsilon|h|$$

Exercise 8.2.32. Show that we can restate the definition of derivative above as follows:

Definition 8.2.9 (Derivative Restated). *Suppose that $f : \mathbb{R} \to \mathbb{R}$. We say that f is differentiable at x and A_x is the derivative of f at x, if there is a linear map $A_x : \mathbb{R} \to \mathbb{R}$ such that for any $\epsilon > 0$ there is a $\delta > 0$ and*

$$|f(x+h) - f(x) - A_x(h)| < \epsilon|h| \textit{ whenever } |h| < \delta$$

See what you can say about the geometry of the derivative given this result. (The next exercise makes that concrete!)

Exercise 8.2.33. Exploration: Think about the following challenge enough to make it precise and then see if you can prove it. This problem is explored very carefully, with more hints, in Exercise 11.3.4 in Chapter 11. **The Challenge**: Show that exercise (8.2.31) combined with Definition (8.2.8) implies that the graph of f can be found in narrower and narrower cones about the tangent line at x^*, as we restrict ourselves to smaller and smaller neighborhoods of x. **Hint**: look at the case of $f(x) = x^2$ at the point $x^* = 0$ and see Figure (39).

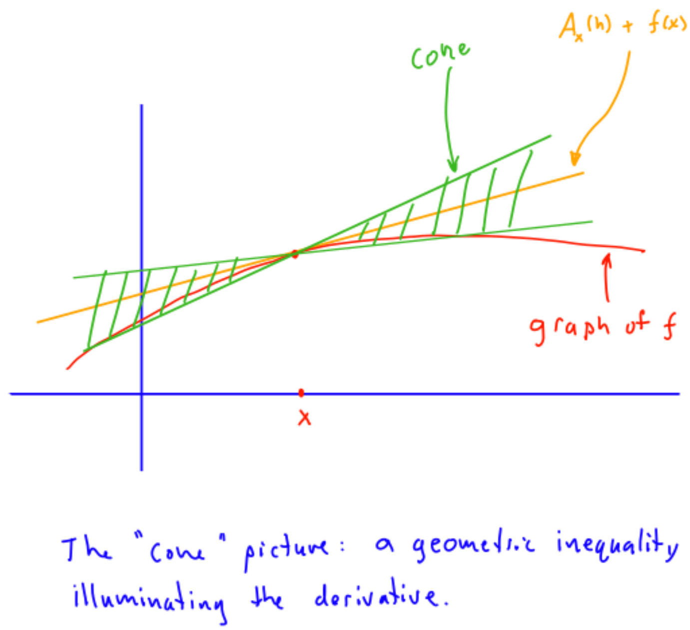


Figure 39: Assistance for Exercise (8.2.33).

8.2.9 $1+\frac{\epsilon}{2}-\frac{\epsilon^2}{8} < \sqrt{1+\epsilon} < 1+\frac{\epsilon}{2}$ *for* $0 < \epsilon$

You will often see this inequality in its use as an approximation:

$$\sqrt{1+\epsilon} \approx 1+\frac{\epsilon}{2}$$

or perhaps

$$\sqrt{1+\epsilon} = 1+\frac{\epsilon}{2}+o(\epsilon)$$

where "+ $o(\epsilon)$" means that the correction term is "in little o of ϵ".

Remark 8.2.5. *The inequality we focus on in this example is rooted in the approximation for small* ϵ *of* $\sqrt{1+\epsilon} \approx 1+\frac{\epsilon}{2}$ *which is itself a restatement of the fact that* $\sqrt{1+\epsilon} = 1+\frac{\epsilon}{2}+o(\epsilon)$ *is equivalent to the statement that* $\sqrt{1+\epsilon}$ *is differentiable at* $\epsilon = 0$ *with derivative there or* $\frac{1}{2}$. *So of course, every differentiable function can be approximated with upper and lower bounds being generated using the derivative in the form of the first two terms of the Taylor series:* $f(x_0)+f'(x_0)(x-x_0)$.

Exercise 8.2.34. Compute the Taylor Series for $\sqrt{1+\epsilon}$ around the point $\epsilon = 0$.

Exercise 8.2.35. Show that if we approximate $\sqrt{1+\delta} \approx 1+\frac{\delta}{2}$, then the magnitude of the error in squares does not exceed $\frac{\delta^2}{4}$, I.e. that:

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

Exercise 8.2.36. We deal with $\delta > 0$:

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

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

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

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

Exercise 8.2.37. Now we want to deal with $\delta < 0$.

1 Use the mean value theorem to prove that:

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

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

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

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

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

3 Show that, in fact:

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

4 Now use this last inequality to show that:

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

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

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

Exercise 8.2.38. Collect the results from the Exercises (8.2.35 - 8.2.37) to conclude that if

$$|\delta| \leqslant \frac{1}{2},$$

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

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

Remark 8.2.6. ***Question***: *why is this approximation valuable?* ***Answer***: *because, very often,* $1 + \frac{\delta}{2}$ *is (algebraically) much easier to work with than* $\sqrt{1+\delta}$.

Exercise 8.2.39. One final exercise on the approximation of $\sqrt{1+\delta}$: Use the Taylor series for $\sqrt{1+\delta}$ with a second derivative error term to conclude that for $\delta \in (-\frac{1}{2}, \frac{1}{2})$:

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

Exercise 8.2.40. See if you can find inequality that approximates $y = f(x) = x^5$ near $x = 1$. I.e. approximate $(1+\epsilon)^5$ for small ϵ and give explicit bounds in the form of inequalities.

8.2.10 Gronwall's Inequality –

In this subsection, we look at an example of Gronwall's differential inequality:

Theorem 8.2.6 (Gronwall). *Suppose that* $x(t) \geqslant 0$ *is smooth,* $a \geqslant 0$ *and* $x(t)$ *satisfies*

$$\frac{dx}{dt}(t) \leqslant ax(t) \ \ \forall t \in [0, T]. \tag{11}$$

Then

$$x(t) \leqslant x(0)e^{at} \ \forall \ t \in [0, T]. \tag{12}$$

Proof. **(Geometric)** The basic idea is that the slope of the curves $x(t)$ must never exceed the slope of the solutions of the equation

$$\frac{dy}{dt}(t) = ay(t) \tag{13}$$

which means $x(t)$ can only cross from above to below any curve $y(t)$ which is a solution curve of equation 13. That is the geometric proof in a nutshell.

In order to be a bit more careful, we introduce a perturbed version of the inequality:

$$G(\epsilon): \quad \frac{dx}{dt}(t) \leqslant (a+\epsilon)x(t) \ \ \forall t \in [0,T] \tag{14}$$

and a perturbed version of the equation:

$$GE(\epsilon): \quad \frac{dw}{dt}(t) = (a+\epsilon)w(t) \ \ \forall t \in [0,T] \tag{15}$$

We will use $x(t)$ for curves satisfying Inequality (11), $y(t)$ to indicate integral curves of Equation (13), $w(t)$ for integral curves of 15.

See Figure (40)

1 We assume that we are above the horizontal axis, since the case of $x(t) = 0 \ \forall \ t \in [0,T]$ is settled – it is a common solution for all the inequalities and equations. I.e. $x = y = w = 0$ for all $t \in [0,T]$ is a solution. We also assume $a > 0$ because the case of $a = 0$ is immediate.
2 Since $x(t) > 0$ (Step (1)), we note $x(s) = w(s)$ for some $s \in [0,T]$ implies $\frac{dx}{dt}(s) < \frac{dw}{dt}(s)$ implying that (locally) $x(t)$ crosses $w(t)$ once at s, from above to below as t increases.
3 Now suppose that $x(r) > x(0)e^{ar}$ for some $r > 0$, contradicting 12.
4 We can choose $y^* = y(0) > x(0)$ such that $y(r) = x(r)$, because we know that the various $y(t)$ curves we get by changing $y(0)$ and solving 13 fill the first quadrant. Note that because $a > 0$ (Step (1)), the vertical distance between the $y(t) = x(0)e^{at}$ and $y(t) = y^*e^{at}$ is never less than $\delta \equiv y^* - x(0) > 0$.
5 By the continuity of the solutions $w(t)$ with respect to ϵ, we get that we can choose ϵ small enough $w(t) = (x(0) + \frac{\delta}{2})e^{(a+\epsilon)t}$ lies strictly between $y(t) = x(0)e^{at}$ and $y(t) = y^*e^{at}$ for $t \in [0,T]$.
6 This forces $x(t)$ to cross from below $w(t)$ at $t = 0$ to above $w(t)$ at $t = r$. Because the curves are continuous, the curves must intersect, but by Step (2), whenever they do, x goes from above w to below w as we move from left to right. By continuity, the set of crossings has a right most point $u \in (0,r)$, such that $x(t) > w(t)$ for $u < t \leqslant r$. But Step (2) again implies this is impossible.

□

Remark 8.2.7. *Another proof is available if you have more sophisticated tools at your disposal. For example, it is a theorem that you can choose coordinates for the graph space of the family of solutions to* $\dot{x} = ax$ *so that those solutions (the integral curves) are horizontal lines. Then, a curve* γ *that starts on the left below one of those integral curves and ends up above that horizontal line on the right must have a positive slope at some point. But that violates the assumption that the derivative of* γ*, in these transformed coordinates, must never be positive.*

Here is another much shorter proof. (Actually the geometric proof is very short in conception but long in writing – which is a characteristic of geometric approaches.)

Proof. **Analytic Proof:** Here we can get the proof by noting:

1

$$\begin{aligned} \frac{d}{ds}\left(x(s)e^{-as}\right) &= \dot{x}(s)e^{-as} - ax(s)e^{-as} \\ &\leqslant 0 \end{aligned}$$

2 By integrating both sides, we get:

$$x(t)e^{-at} - x(0) \leqslant 0$$

which rearranges to the result we wanted.

□

Exercise 8.2.41. See if you can modify the second proof of Theorem (8.2.6) to prove that if

$$\frac{dx}{dt}(t) \leqslant ax(t) + b \quad \forall t \in [0, T].$$

Then

$$x(t) \leqslant e^{at}(x(0) + bt)\forall t \in [0, T].$$

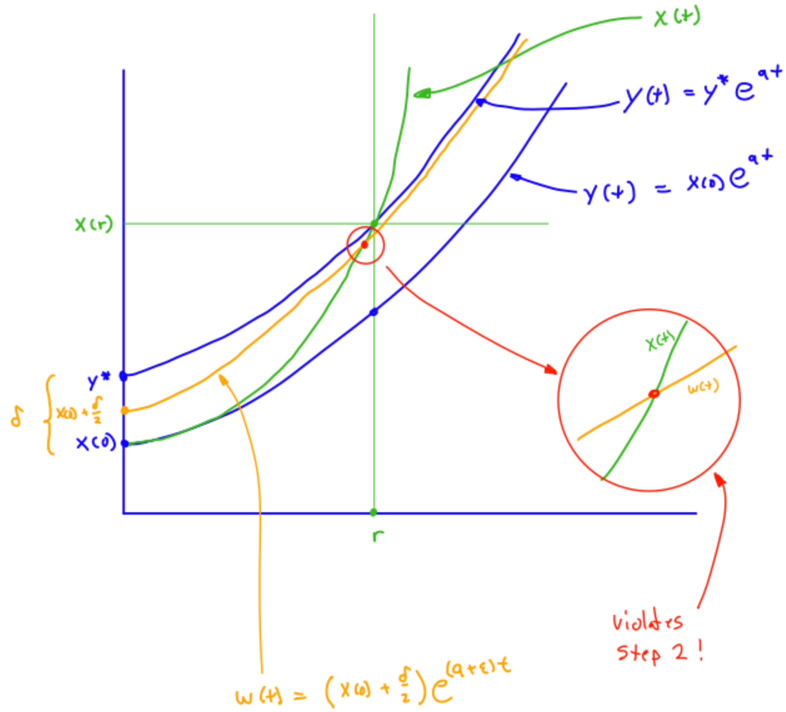


Figure 40: Visual assistance for the geometric proof of Gronwall's inequality.

8.3 Play and the Art of Inequalities

As stated before the 10 previous subsections, mastery of the art of inequalities requires immersion. Training the instincts – making the grasp instinctive – is a much deeper task than shooting for mere acquaintance. In the course of working in any area of analysis and geometric analysis, that practice with inequalities will happen quite naturally, but there are also nice references I recommend that focus explicitly on inequalities.

HARDY, LITTLEWOOD AND POLYA [19]: The classic *Inequalities*, published in 1934 and written by GH Hardy, JE Littlewood and G Polya and is still used.

BURAGO AND ZALGALLER [9]: With the title *Geometric Inequalities* it is no surprise that this is a book devoted to inequalities that are geometric in nature. The authors are Yurii D. Burago and Viktor A. Zalgaller. I recommend it highly. It was first published in Russian in 1980 and then translated by Springer in 1988.

STEELE [39]: *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities* is a 2004 book by Michael J Steele, with the flavor of a masterclass in book form. It will not be to everyone's taste, but it is definitely worth a look to see if it fits your learning mode.

BECKENBACH AND BELLMAN [7]: *Inequalities* by Edwin F Beckenbach and Richard Bellman was first published in 1965. I do not have this book, but plan on getting a copy. (I ran into it when using Google Scholar to get the citations for the other three books here.)

Relatively simple exercises with inequalities make an almost ideal warmup for hours devoted to analysis research. In the same way that you warm up before full-on sprints, playing with inequalities can be a fun and effective way to get yourself in the flow, the creative groove.

9 Linearity: Linear Spaces, Linear Maps, Linear Intuitions

While Chapter 5 made it clear that linearity is a crucially important foundation for analysis, it was not explored or explained in that chapter.

While I am sure that others might debate this – and that the debate would be very interesting – I believe that the fundamental reason for the importance of linearity is the existence of momentum and the usefulness of smooth functions and objects. The idea that acceleration requires force means that big objects move rather smoothly.

Intuitively, smoothness is the property of zooming in and finding that the thing (path of movement, curve, surface, etc.) you are zooming in on, looks flat. The more you zoom in, the flatter it looks. Of course real objects are typically only approximated by smooth things – zoom in far enough and you find details and roughness, even discreteness.

Simply put: Linear functions are useful because their level sets and graphs are flat. And linear functions have very nice properties.

This chapter contains a simple sequence of definitions and properties with examples and figures throughout, as well a section (Section 9.2) that discusses why we care. While I am assuming you have seen linear algebra before, linear algebra courses vary in content and perspective. This chapter should serve to synchronize the readers with the particular way we will look at and use linear spaces and maps. **In particular note that this review is speedy and would be a challenging way to first learn the linear algebra and geometry it contains**.

Note: we will often use the terms *matrix* and (linear) *operator* interchangeably even though a matrix is merely one, particular representation of a finite dimensional linear operator given a specific choice of basis for the domain and for the range. While it is true that the operator does not depend on the matrix (which typically changes when we change the basis for either the domain or the range or both), since we will usually be dealing with finite dimensional vector spaces, often with particular bases in mind, the interchangeability will be a rather harmless glossing over of details.

9.1 The Basic Ideas and Tools

9.1.1 *Linear/Vector Spaces*

A linear space X is another name for a vector space, and is (1) a collection of points called vectors, together with (2) an operation of addition "+" combining vectors to get new vectors and (3) a scalar field S (either the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$) that we can multiply vectors by and get a vector as a result. More precisely, if $x, y, z \in X$ and $\alpha, \beta \in S$, 0 is the additive identity in S, 1 is the multiplicative identity of S and $\vec{0}$ is the additive identity of X:

1. $x + y \in X$
2. $x + y = y + x$
3. $x + (y + z) = (x + y) + z$
4. $x + \vec{0} = x$
5. For every $x \in X$ there is a vector $w \in X$ such that $x + w = \vec{0}$. This is also referred to as $-x$, so we have $x + (-x) = \vec{0}$
6. $\alpha(\beta v) = (\alpha\beta)v$
7. $(\alpha v) + (\beta v) = (\alpha + \beta)v$
8. $1v = v$
9. $0v = \vec{0}$

Remark 9.1.1. *We will usually use* 0 *to represent both* $\vec{0}$ *and* 0 *if the intention is clear from the context.*

The span of a set of vectors $\{w_i\}_{i=1}^k$ is the set of all weighted linear combinations: $W \equiv \{x \in X : x = \sum_{i=1}^k \alpha_i w_i$ for some set of scalars $\{\alpha_i\}_{i=1}^k\}$. If $W \neq X$, W is called a *linear subspace* of X.

Any subset of a vector space X that can be represented as the sum of a linear subspace W and an element $y \in X$,

$$W + y \equiv \{x \in X : x = w + y \text{ for some } w \in W\}$$

is called an *affine subspace* of X.

Exercise 9.1.1. Show that if W is a linear subspace of X, an affine subspace $W + y$ is also a linear subspace of X if and only if $y \in W$.

Definition 9.1.1 (Linearly Independent). *A set of vectors* $\{w_i\}_{i=1}^k$ *is linearly independent if* $\sum_{i=1}^k \alpha_i w_i = \sum_{i=1}^k \hat{\alpha}_i w_i$ *implies* $\alpha_i = \hat{\alpha}_i$ *for all* i.

Definition 9.1.2 (Independent Basis). *A set of vectors* $\{v_i\}_{i=1}^N \subset X$ *is an independent basis for* X *if every* $x \in X$ *has a unique representation* $x = \sum_{i=1}^N \alpha_i v_i$.

In many important cases, we need an infinite number of basis elements, but we will mostly stick to the case in which $n < \infty$. The number n is the dimension of X.

9.1.2 Linear Maps

Definition 9.1.3 (Linear Map). *A map from one vector space to another,* $F : X \to Y$ *is a linear map if*

$$F(\alpha x + \beta y) = \alpha F(x) + \beta F(y)$$

for all $x, y \in X$ *and all* $\alpha, \beta \in S$.

Notice that we can decompose this into two pieces:

$$F(x+y) = F(x) + F(y)$$

and

$$F(\alpha x) = \alpha F(x)$$

for all $x, y \in X$ and all $\alpha, \in S$. In other words scaling and addition can be done before or after the mapping and you will get the same result. These properties are very useful. For example, the scaling property immediately implies that if for any $\epsilon > 0$, we know F on $B(0, \epsilon)$, we know it everywhere. Or, even if we know if only on $\partial B(0, \epsilon)$, the sphere of radius ϵ, we know it everywhere. But we know even more. Because it is linear, then for any basis $\{v_i\}_{i=1}^N$ of X, knowing $\{F(v_i)\}_{i=1}^N$ completely specifies F on all points in X!

Exercise 9.1.2. Show that the scaling property immediately implies that if for any $\epsilon > 0$, we know F on $B(0, \epsilon)$, we know it everywhere.

Exercise 9.1.3. Show that $\{w_i\}_{i=1}^k$ is linearly independent if and only if $\sum_{i=1}^k \hat{\alpha}_i w_i = 0$ implies that $\hat{\alpha}_i = 0$ for all i.

Exercise 9.1.4. Show that if $\{v_i\}_{i=1}^N$ is a basis for the X, then knowing $\{F(v_i)\}_{i=1}^N$ completely specifies a linear map F on all points in X. Hint: write an arbitrary $x \in X$ in terms of the basis $\{v_i\}_{i=1}^n$ and then compute $F(x)$

Exercise 9.1.5. Show that $F(\text{span}(\{w_i\}_{i=1}^k) = \text{span}(\{F(w_i)\}_{i=1}^k)$.

An *affine map* is any map $F : X \to Y$, X and Y vector spaces, such that $G(X) \equiv F(x) - F(0)$ is a linear map.

9.1.3 Null Space

Suppose that $A : X \to Y$ is a linear map. We define the *Null Space* of A to be $A^{-1}(0)$ ... the null space of A is the set of $x \in X$ which map to the 0 element of Y.

Exercise 9.1.6. Show that for linear $A : X \to Y$, $A^{-1}(0)$ is either all of X or a linear subspace of X.

9.1.4 Matrices

As noted in the linear map section, if $\{v_i\}_{i=1}^n$ is a basis for the linear space X and $F : X \to Y$ is a linear map to the linear space Y, all we need to know $F(x)$ for all $x \in X$, is $\{F(v_i)\}_{i=1}^n$.

Let $\{v_i\}_{i=1}^n$ be a basis for X and $\{w_i\}_{i=1}^m$ be a basis for Y. Suppose that:

$$\begin{aligned}
F(v_1) &= F_{1,1}w_1 + F_{2,1}w_2 + ... + F_{m,1}w_m \\
F(v_2) &= F_{1,2}w_1 + F_{2,2}w_2 + ... + F_{m,2}w_m \\
\vdots \quad & \quad \vdots \\
F(v_n) &= F_{1,n}w_1 + F_{2,n}w_2 + ... + F_{m,n}w_m
\end{aligned}$$

and let

$$M_F \equiv \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,n} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,n} \\ \vdots & \vdots & & \vdots \\ F_{m,1} & F_{m,2} & \cdots & F_{m,n} \end{bmatrix}.$$

We note that if we represent an arbitrary $x = x_1v_1 + x_2v_2 + ... + x_nv_n$ where the x_i are *scalar coordinates* of x in the basis $\{v_i\}_{i=1}^n$ and $y = y_1w_1 + y_2w_2 + ... + y_mw_m$ where the y_i are again *scalar coordinates* of y in the basis $\{w_i\}_{i=1}^m$, then $F(x)$ is the $y \in Y$ with coordinates given by:

$$M_F \cdot x \equiv \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,n} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,n} \\ \vdots & \vdots & & \vdots \\ F_{m,1} & F_{m,2} & \cdots & F_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

i.e. letting

$$y \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

we get that

$$F(x) = y = \begin{bmatrix} F_{1,1}x_1 + F_{1,2}x_2 + \cdots + F_{1,n}x_n \\ F_{2,1}x_1 + F_{2,2}x_2 + \cdots + F_{2,n}x_n \\ \vdots \\ F_{m,1}x_1 + F_{m,2}x_2 + \cdots + F_{m,n}x_n \end{bmatrix}$$

The *matrix representation of F* under the choice of basis of $\{v_i\}_{i=1}^{n}$ for X and $\{w_i\}_{i=1}^{m}$ for Y is M_F.

Remark 9.1.2. *For the rest of the text, we will use $x \in X$ to represent both the point in the n-dimensional vector space X and the coordinate representation, in column vector form. I.e. we will also mean:*

$$x \equiv \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where these coordinates are with respect to some basis. By x^T we will mean the transpose of x, the row vector:

$$x^T \equiv \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

If we define:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

then the transpose of the matrix of A is given by:

$$A^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{bmatrix}$$

9.1.5 Norms

A norm on a vector space is a function $||\cdot|| : X \to [0, \infty)$ that satisfies:

1. $||x|| = 0 \Leftrightarrow x = 0$
2. $||\alpha x|| = |\alpha|\, ||x||\ \forall \alpha \in S$, where S is the scalar field for X.
3. $||x + y|| \leqslant ||x|| + ||y||$

We have already encountered norms in the chapter on metric spaces.

Exercise 9.1.7. Prove that norms are convex functions.

9.1.6 Norms of Maps

We define the operator norm of the linear map (or equivalently, operator) F to be

$$\max_{\{x:||x||\leqslant 1\}} ||F(x)||.$$

Exercise 9.1.8. Prove that linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ are continuous.

Exercise 9.1.9. Prove that what you proved in Exercise 9.1.8 implies that the maximum in the definition of operator norm, above, actually exists.

9.1.7 Inner Products = dot products

You are already acquainted with dot products in $\mathbb{R}^n$: $x \cdot y \equiv x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$. The dot product is just an example of an *inner product*: We will now focus vector spaces where the scalar field is the real numbers $\mathbb{R}$. Any map from $X \times X \to \mathbb{R}$, where X is a real vector space, satisfying

1. $\langle x, y \rangle \in \mathbb{R}$
2. $\langle x, x \rangle \geqslant 0$ and $\langle x, x \rangle = 0 \Leftrightarrow x = 0$
3. $\langle x, y \rangle = \langle y, x \rangle$
4. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$ (and so of course $\langle z, \alpha x + \beta y \rangle = \alpha \langle z, x \rangle + \beta \langle z, y \rangle$ as well)

is called an *inner product* on X

Exercise 9.1.10. Show that the usual dot product satisfies these requirements and is therefore an inner product.

Exercise 9.1.11. (Challenge): Use matrix representations of linear maps to show that any inner product $\langle x, y \rangle$, expressed in a chosen basis, can be represented by $x^T A y$, where A is a symmetric matrix.

An n-dimensional vector space X, with an inner product $\langle x, y \rangle = x^T A y$ for some n-by-n matrix A and a norm $|\cdot|$ defined by

$$|x| \equiv \sqrt{\langle x, x \rangle}$$

is called an *inner product space*. Note that $\mathbb{R}^n$ with the usual dot product is an inner product space.

Remark 9.1.3. *If we are working in a vector space over the complex numbers, while it is still true that $\langle x, x \rangle > 0$ if $x \neq 0$, the inner product also satisfies $\langle x, y \rangle = \langle y, x \rangle^*$ where the $*$ indicates complex conjugation. As a result, while $\langle x, y \rangle$ is linear in the first term, it satisfies $\langle x, \alpha y + \beta z \rangle = \alpha^* \langle x, y \rangle + \beta^* \langle x, z \rangle$ in the second term. One simple inner product in $\mathbb{C}^n$ is just the usual dot* *product between the first vector and the complex conjugate of the second vector:* $\langle x, y\rangle \equiv x \cdot y^*$

9.1.8 Orthogonality, Orthogonal Subspaces and Projections

Suppose that X is an inner product space. Two non-zero vectors $x, y \in X$ are said to be *orthogonal* if $\langle x, y\rangle = 0$.

The *angle between two non-zero vectors* in an inner product space is defined to be

$$\theta \equiv \arccos\left(\frac{\langle x, y\rangle}{|x|\ |y|}\right)$$

.

A set of vectors are said to be orthogonal if the inner product of any two distinct vectors in the set is equal to 0. Given any k-dimensional subspace W, one can always find a set of k vectors which are orthogonal that span W. We will look at a method for finding this spanning set in Section 9.1.10.

A basis (i.e. an independent basis) for X that is also orthogonal is said to be an *orthogonal basis*.

Suppose that $\{v_i\}_{i=1}^k$ is an orthogonal set of vectors in an n-dimensional space. Then, defining

$$V \equiv \begin{bmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & \cdots & | \end{bmatrix}$$

we define the *orthogonal projection onto* $\text{span}(\{v_i\}_{i=1}^k)$, P_V to be

$$P_V \equiv VV^T.$$

See Figure 41.

We say that two linear subspaces of X, V and W, are *orthogonal subspaces* to each other if $\langle v, w\rangle = 0$ for all $v \in V$ and $w \in W$. If $V + W = X$ and V

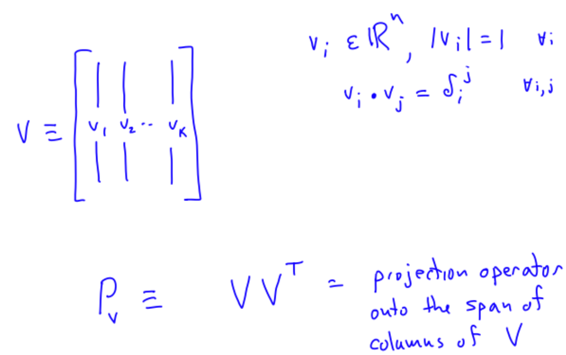


Figure 41

and W are orthogonal to each other, we say that V is the orthogonal complement of W and W is the orthogonal complement of V.

We say that P is a projection operator if $P \circ P = P$. A projection is an orthogonal projection if $\langle x - P(x), P(x) \rangle = 0$. That is, if after projecting an x onto the range of P, the residual $x - P(x)$ is orthogonal to x.

Exercise 9.1.12. Show that a projection operator satisfies $P(x) = x$ for any x in the range of P.

Exercise 9.1.13. Show that if $x \in \text{span}(\{v_i\}_{i=1}^k)$, then $P_V(x) = x$.

Exercise 9.1.14. Show that the range of $I - P_V$ and the range of P_V are orthogonal complements.

Exercise 9.1.15. Show that the "orthogonal projection onto $\text{span}(\{v_i\}_{i=1}^k)$, P_V" is indeed, an orthogonal projection according to the definition just given.

Exercise 9.1.16. Suppose that A is matrix representing a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, $m < n$. Suppose that $\{w_i\}_{i=1}^m \subset \mathbb{R}^n$ are the m rows of A. Show that the null space is precisely the orthogonal complement of the $\text{span}(\{w_i\}_{i=1}^m)$.

9.1.9 *Singular Value Decomposition (SVD)*

Every m by n matrix $A \in \mathbb{R}^{mn}$ has a decomposition:

$$A = U\Sigma V^T$$

where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and Σ is an $m \times n$ diagonal matrix, with ordered, non-negative diagonal entries. See Figure (42).

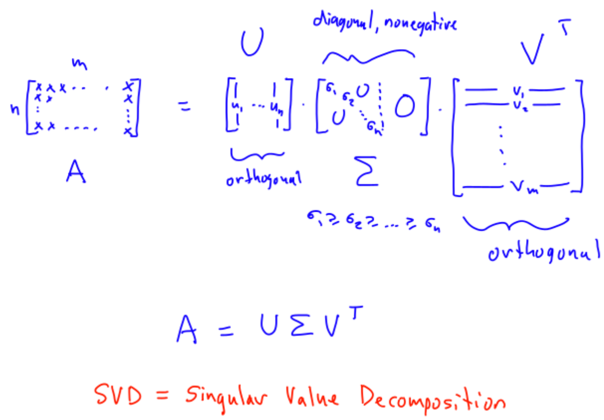


Figure 42

Exercise 9.1.17. Show how to use the SVD of a matrix A to find the null space and range of A.

Exercise 9.1.18. Assume that A is an $n \times n$, invertible matrix. Show that errors in the solutions to the linear equation $Ax = b$, that occur when b is perturbed by noise to $b + \eta$ instead – i.e. that the difference between the solution of $Ax = b$ and $Ax = b + \eta$ – is bounded by $\frac{|\eta|}{\sigma_{min}}$ where σ_{min} is the smallest singular value of A. Since we are assuming A is nonsingular, $\sigma_{min} > 0$. Hint: Use the SVD to (1) compute A^{-1} and (2) examine the effects of A^{-1} on η. Alternatively write x and b in the basis created by the rows of V^T and the columns of U, the left and right orthogonal matrices from the SVD.

9.1.10 QR Decomposition

Given k independent vectors in $\mathbb{R}^n$, $\{v_i\}_{i=1}^k$, $k \leqslant n$, we can find an orthonormal basis of $\mathbb{R}^n$, $\{x_i\}_{i=1}^n$ (a set of orthogonal vectors, each of length 1, that span $\mathbb{R}^n$), such that:

Condition 9.1.1 (GS-condition). *We have:*

$$\begin{array}{rcl} \operatorname{span}(v_1) & \subset & \operatorname{span}(x_1) \\ \operatorname{span}(\{v_i\}_{i=1}^2) & \subset & \operatorname{span}(\{x_i\}_{i=1}^2) \\ \operatorname{span}(\{v_i\}_{i=1}^3) & \subset & \operatorname{span}(\{x_i\}_{i=1}^3) \\ & \vdots & \\ \operatorname{span}(\{v_i\}_{i=1}^k) & \subset & \operatorname{span}(\{x_i\}_{i=1}^k) \end{array}$$

This can be done using the Gram-Schmidt procedure.

A little more generally, if we are given given m (not necessarily independent) vectors in $\mathbb{R}^n$, $\{v_i\}_{i=1}^m$, (with $m > n$ possible), an orthogonal basis can be found that satisfies Condition 9.1.1 for $k \leqslant n$ and, of course satisfies $\operatorname{span}(\{v_i\}_{i=1}^k) \subset \operatorname{span}(\{x_i\}_{i=1}^n) = \mathbb{R}^n$ for $k > n$ (in the case that $m > n$)

Remark 9.1.4. *The GS condition is chosen so that it applies to both to independent and non-independent $\{v_i\}_{i=1}^k$. If we were only concerned with independent $\{v_i\}_{i=1}^k$, all the subset symbols (i.e. $\subset$) in the condition could be replaced by equal symbols (i.e $=$).*

The *QR decomposition* is the matrix decomposition that accomplishes this, giving us this basis and the coefficients that allow you to get the v_i's from the x_i's. More concretely, suppose that we define V to be the matrix whose columns are the vectors $\{v_i\}_{i=1}^m$. Then:

Theorem 9.1.1 (QR Decomposition). *There is an orthogonal, $n \times n$ matrix Q and an upper triangular, $n \times m$ matrix R such that $V = QR$. See Figure (43)*

Remark 9.1.5. *Notice that the fact* R *is an upper triangular matrix implies that the columns of* Q *give us the* x_i*'s, that together with the* v_i*'s give us Condition 9.1.1.*

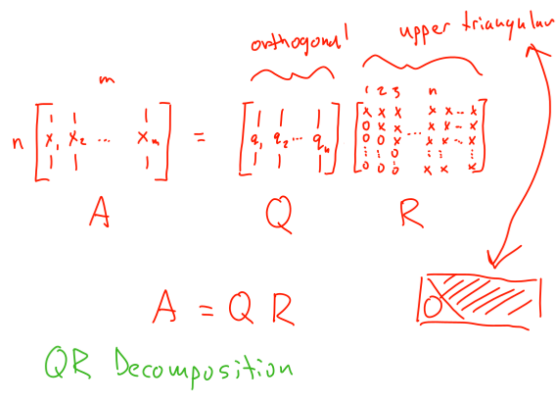


Figure 43

Exercise 9.1.19. **(Challenge)** Suppose that $\{v_i\}_{i=1}^{k} \subset \mathbb{R}^n$ are linearly independent. Show how to find orthonormal vectors $\{x_i\}_{i=1}^{k}$ such that Condition 9.1.1 is satisfied. **Hint**: Define $x_1 = \frac{v_1}{|v_1|}$. Define P_{x_1} to be the orthogonal projection onto the span of x_1 and consider $P_{x_1}(v_2)$ and $(I - P_{x_1})(v_2)$ ...

9.1.11 Symmetric, Normal

Linear operators represented by symmetric matrices play a important role in mathematics and physics. Because we sometimes work with vector spaces over the real numbers and sometimes over the complex numbers, we differentiate between the two in this section. One important reason we care about complex matrices is that the eigenvalues of a real matrix can come in complex conjugate pairs. Another important reason is that mathematical physics in general, and quantum mechanics in particular, use complex operators and complex states to represent highly validated, everyday (though microscopic) physics.

A real matrix is said to be *symmetric* if $A = A^T$. An operator or matrix A is *symmetric and positive definite* if $A = A^T$ and $x^T Ax > 0$ whenever $|x| > 0$.

In the case that the matrix is complex, we call the matrix that satisfies $A = A^\dagger$, *Hermitian*, where the $\dagger$ indicates we have taken both the transpose and the complex conjugate of the matrix.

We note that in the case that A is a real matrix, $A = A^T \Leftrightarrow A = A^\dagger$. (Which means if we were most concerned about brevity, we could dispense with the definitions focused only on real matrices, but this is not the choice here – sometimes brevity does not clarify.)

A real matrix (or operator) A, is a *normal matrix (or operator)* if $A^T A = AA^T$. A complex matrix (or operator) A, is a *normal matrix (or operator)* if $A^\dagger A = AA^\dagger$.

A term that is used for both symmetric and Hermitian matrices and operators is *self adjoint*. It comes from the fact that the *adjoint* of an operator A on an inner product space X with inner product $\langle , \rangle$ is the operator A^* defined to be the operator that satisfies $\langle A^* x, y\rangle = \langle x, Ay\rangle$ for all $x, y \in X$. Self adjoint means that $A = A^*$. in the case of matrices, self-adjoint means symmetric (in the case of real matrices) and Hermitian (in the case of complex matrices).

9.1.12 Determinants

In the courses (or even courses) you have taken in linear algebra, you have certainly encountered the determinant of a matrix: for the simple two by two matrix,

$$A \equiv \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

you certainly know that

$$\det(A) = a_{11}a_{22} - a_{21}a_{12}.$$

It is less certain that you have encountered the geometric interpretations or the connections to exterior vector algebras.

In this short section, I will guide you to an understanding that the determinant of a matrix is a signed volume of the parallelogram defined by the columns of the matrix.

9.1.12.1 *Determinants are Signed Volumes*

We will show this by a sequence of definitions and exercises with hints.

1 We define the determinant of an $n x n$ matrix

$$\begin{aligned} A &= \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \\ &= \begin{bmatrix} | & | & \cdots & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & \cdots & | \end{bmatrix} \end{aligned}$$

to be:

$$\det(A) = \sum_{\sigma \in \mathrm{Perm}(n)} sign(\sigma) \Pi_{i=1}^{n} a_{i,\sigma(i)}$$

Note: $sign(\sigma)$ is 1 or -1 depending on whether the permutation σ is even or odd.

Exercise 9.1.20. Show that the determinant is the sum of all products defined by paths through the matrix hitting each row and column exactly once (with sign determined by the sign of the path viewed as a permutation).

2 Connection to k-vectors: The wedge product of k vectors in $\mathbb{R}^n$, $a_1 \wedge a_2 \wedge \cdots \wedge a_k$ is an element in a vector space – **the space of k-vectors** – with dimension $\frac{n!}{k!(n-k)!}$. The wedge product is linear in each factor

$$\begin{aligned} a_1 \wedge \cdots \wedge (\alpha a_i + \beta b_i) \quad \wedge & \quad \cdots \wedge a_k \\ = & \quad \alpha(a_1 \wedge \cdots \wedge a_i \wedge \cdots \wedge a_k) \\ + & \quad \beta(a_1 \wedge \cdots \wedge b_i \wedge \cdots \wedge a_k) \end{aligned}$$

and it is alternating – i.e. if you switch any pair of terms, the sign of the result changes.

Exercise 9.1.21. Prove that:

a The space of n-vectors created wedging together vectors from $\mathbb{R}^n$ has dimension 1.

b If $\{e_i\}_{i=1}^n$ are orthogonal basis vectors for $\mathbb{R}^n$, show that

$$a_1 \wedge a_2 \wedge \cdots \wedge a_n = \det(A) \cdot e_1 \wedge e_2 \wedge \cdots \wedge e_n$$

.

3 Properties of Determinants:

Exercise 9.1.22. Prove the following properties of the determinant:

a Switching two columns of a matrix changes the sign of the determinant

b The determinant is linear in each column.

c $\det(A) = \det(A^T)$. Hint: $sign(\sigma^{-1}) = sign(\sigma)$ for all σ.

d Part (a) implies the determinant of a matrix with two identical columns is 0.

4 The matrix $R^{\theta}_{i,j}$ rotates space by rotating everything in the 2-dimensional subspace spanned by the ith and jth basis vectors by θ:

$$R^{\theta}_{i,j} = \begin{array}{c|c} & \begin{array}{cccccccccc} 1 & 2 & \cdots & i & & \cdots & j & & \cdots & n \end{array} \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ i \\ \vdots \\ j \\ \vdots \\ n \end{array} & \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \cos(\theta) & \cdots & -\sin(\theta) & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \sin(\theta) & \cdots & \cos(\theta) & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \end{array}$$

Exercise 9.1.23. Prove that $\det(R^{\theta}_{i,j}) = 1$ and that $\det(R^{\theta}_{i,j} A) = \det(R^{\theta}_{i,j}) \det(A)$ using the linearity in each column and the fact that switching columns changes the sign.

5 Creation of orthogonal matrices:

Exercise 9.1.24. Show that you can create any orthogonal matrix $\hat{O}$ by a product of $R^{\theta}_{i,j}$'s and perhaps one coordinate flip (reflection about one $n-1$-plane), more specifically either:

$$\hat{O} = I \circ \left(\Pi_{i=1}^{n-1} R^{\theta_i}_{i,i+1} \right)$$

or

$$\hat{O} = \hat{I} \circ \left(\Pi_{i=1}^{n-1} R^{\theta_i}_{i,i+1} \right)$$

where I is the identity matrix and $\hat{I}$ is the matrix with ones down the diagonal except that $I_{n,n} = -1$. **Hint**: an orthogonal matrix is a matrix whose columns are orthogonal to each other and have unit norm. Turn the identity matrix I into $\hat{O}$, through a sequence of rotations that rotate each column (each of which start out as a single 1 in some row), from left to right, into the columns of $\hat{O}$ through pairwise rotations. **Note**: $\Pi_{i=1}^{k} A_i = A_k \circ A_{k-1} \circ A_{k-2} \circ \cdots \circ A_1$, i.e. we are defining a product in which the multiplication is on the left.

6 Odds and ends:

Exercise 9.1.25. Prove:

a $\det(\text{diagonal matrix}) = \text{product of the diagonal elements}$

b If Σ is a diagonal matrix, $\det(\Sigma \circ A) = \det(\Sigma)\det(A)$

c $\det(\hat{I}) = -1$

d Use the fact that, for any n by n matrix A, $A = U \circ \Sigma \circ V^T$ where U and V are orthogonal matrices and Σ is a diagonal matrix with non-negative elements to get that:

$$\begin{aligned} \det(A) &= \pm|\text{volume of image of } [0,1]^n \text{ under } A| \\ &= \pm\Pi_{i=1}^n \sigma_i \end{aligned}$$

where $\{\sigma_i\}_{i=1}^n$ are the singular values of A and the sign depends on the orientation of A

7 Products, in general:

Exercise 9.1.26. Prove that $\det(AB) = \det(A)\det(B)$. Hint: use the SVD to write $A = U \circ \Sigma \circ V^T$ and Exercises 9.1.23 and 9.1.24.

The fact that $\det(A) =$ oriented volume of the parallelepiped generated by the columns of A is used all over geometric analysis.

9.1.13 *Eigenstuff*

A very important decomposition of a linear operator is the *eigenvalue-eigenvector decomposition*. Not every linear operator has such a decomposition, but when it does, this gives us a very simple representation of the operator. The eigenvalue equation is:

$$Ax = \lambda x$$

and the components of the solution pairs $\{x, \lambda\}$, are called *eigenvectors* and *eigenvalues*, respectively. The eigenvalues are scalars that can be complex – they can have nonzero imaginary parts.

We will deal mostly with the case in which A is real. If $A : \mathbb{R}^n \to \mathbb{R}^n$ is also symmetric, we automatically get a complete set of real, orthogonal eigenvectors with real eigenvalues.

In this case, if we use the eigenvectors of A as the basis for the space, and we define V = the n by n matrix in which each of the columns is a different eigenvector and Λ is the diagonal matrix with the eigenvalues along the diagonal, we get

$$AV = V\Lambda$$

implying that

$$A = V\Lambda V^{-1} = V\Lambda V^{T}$$

which in turn implies

$$V^{T}AV = \Lambda$$

that is, if we use the (orthogonal) columns of V as a basis for $\mathbb{R}^n$, then the A is a diagonal matrix.

In the case that A is real and normal, we may get complex eigenvalues and the closest we can get to a real decomposition is one in which pairs of orthogonal vectors span an invariant space that is rotated by A.

Exercise 9.1.27. Show that $\det(A - \lambda I) = 0$ if λ is an eigenvalue of A.

Exercise 9.1.28. By computing the determinant, $\det(A - \lambda I)$, show that for every distinct eigenvalue of A, there is at least one eigenvector.

Exercise 9.1.29. **(Challenge)** Show that if $A : \mathbb{R}^n \to \mathbb{R}^n$ and $A = A^{\dagger}$, then (1) the eigenvalues are real and (2) the eigenvectors form an orthogonal basis for the $\mathbb{R}^n$.

Exercise 9.1.30. Look up and read about the Jordan normal Form for real, $n \times n$ matrices. This form is the best we can do in the case that A is not normal. If we work over the complex numbers, then the Jordan normal form has all its nonzero entries on either the diagonal or first super-diagonal. If we work over the reals, we get a matrix whose non-zero entries are restricted to the main diagonal and the first super and sub-diagonals.

Remark 9.1.6. *If* $A : \mathbb{R}^n \to \mathbb{R}^n$ *is normal, i.e.* $AA^{\dagger} = A^{\dagger}A$, *then there is a complete basis of orthogonal (possibly complex) eigenvectors. One can* *diagonalize A with a real, orthogonal basis where there are either real eigenvalues or 2 by 2 blocks along the diagonal. The 2 by 2 blocks correspond to 2 dimensional invariant subspaces which* A *rotates by some angle* θ *and the resulting 2 by 2 block is*

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

9.1.14 Extra Exercises

Exercise 9.1.31. Given k vectors, $\{v_i\}_{i=1}^k \subset \mathbb{R}^n$, $k < n$, show how to use the QR decomposition to find the dimension of the span$(\{v_i\}_{i=1}^k)$.

Exercise 9.1.32. Given k independent vectors, $\{v_i\}_{i=1}^k \subset \mathbb{R}^n$, $k < n$, show how

1. to use the QR decomposition to compute the orthogonal projection operator onto span$(\{v_i\}_{i=1}^k)$.
2. to use the QR decomposition to create a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^{n-k}$ with null space equal to span$(\{v_i\}_{i=1}^k)$.

9.2 Remarks on Linear and Nonlinear Spaces and Maps

9.2.1 Linear vs Nonlinear

Minimalistic, almost knee-jerk, pictures of linearity and non-linearity are:

LINEAR flat planes of various dimensions, very simple descriptions, local = global

SMOOTH, NON-LINEAR wild behavior/shape possible, complex descriptions, local structure $\neq$ global structure, **but(!)** locally linear everywhere (because of everywhere differentiability!)

Nonsmooth (and of course, Nonlinear!) completely wild behavior/shape possible, arbitrarily complex descriptions, local structure $\neq$ global structure and **not(!)** locally linear everywhere (i.e. not differentiable everywhere!)

Because many nonlinear things are mostly locally linear – differentiable everywhere or almost everywhere – we can leverage our mastery of things linear to begin to get a real grasp on things non-linear. And many phenomena/dynamics in nature are approximately or locally linear between singularities, those discrete events completely outside of the realm of linear behavior.

A great deal of insight into nonlinear sets, maps and evolution can be gained by understanding linear and quadratic approximations (i.e. the first and second order derivatives). Furthermore, the more intricate approaches to nonlinear problems often mimic the ideas that come from these first and second order approximations.

9.2.2 *Some Simple Examples*

Figures 44-49 give simple examples that are helpful in creating visual bookmarks to the nature of linearity (and of non-linearity):

9.2.3 *Maps and linearity (or lack of it)*

There is a hierarchy of complexity in the world of maps. Both maps and spaces can be:

1 Linear
2 Smooth and Nonlinear
3 Nonsmooth (and therefore also Nonlinear)

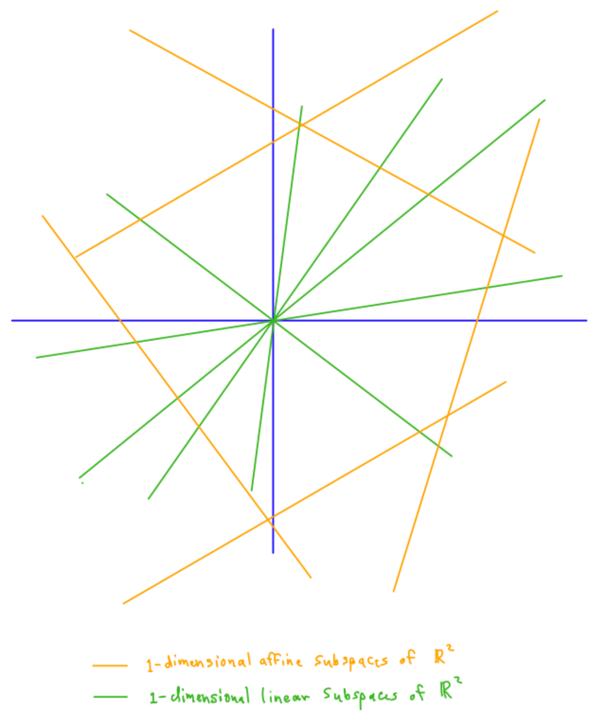


Figure 44: The difference between affine and linear subspaces.

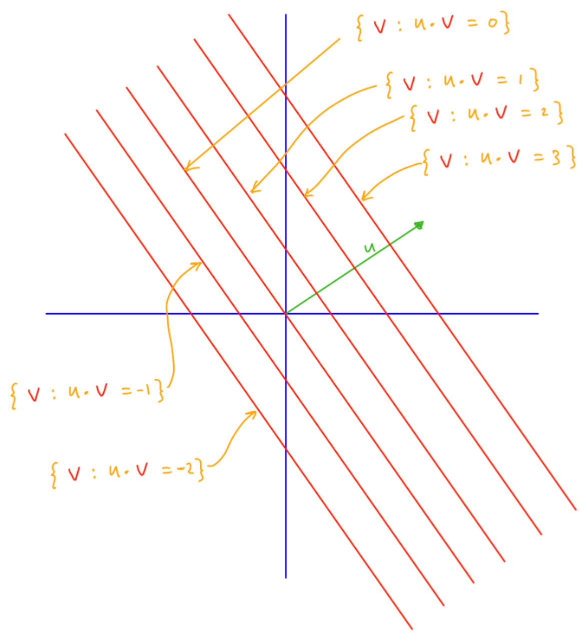


Figure 45: Level sets of linear functions are linear (or affine) subspaces of the domain.

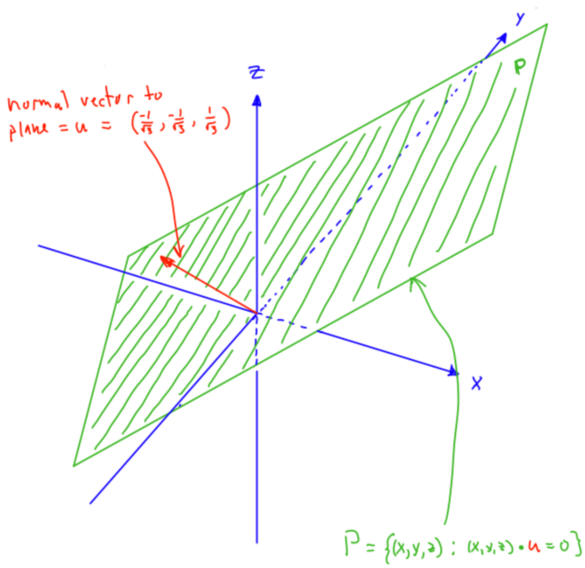


Figure 46: Example of a hyperplane.

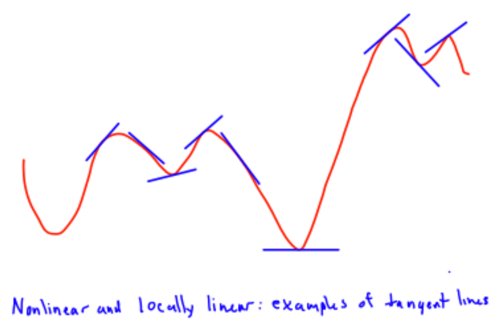


Figure 47: Example of the graph of a smooth (locally linear approximations everywhere) function from $\mathbb{R}$ to $\mathbb{R}$.

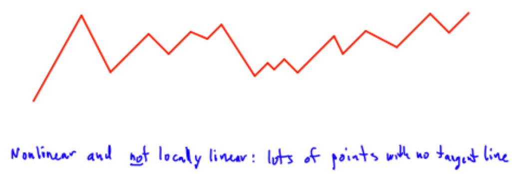


Figure 48: Example of a graph of a function with many points of nonsmoothness.

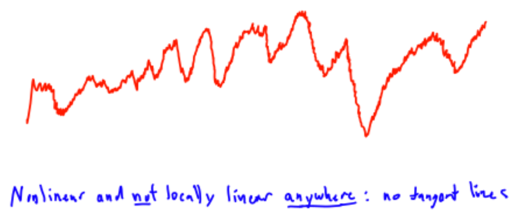


Figure 49: Example of a graph of a function with no smoothness.

For example, we can have a smooth, nonlinear map between a linear space and a nonlinear space:

$$\{\text{linear space}\} \xrightarrow{\text{smooth, nonlinear map}} \{\text{nonsmooth, nonlinear space}\}$$

There are two restrictions: (1) the map can be linear if and only if the domain and range are linear spaces and (2) The map can be smooth if and only if both domain and range have locally linear approximations.

Remark 9.2.1. *There is a subtle detail here – a nonsmooth set can have locally linear approximations almost everywhere, in which case we can (in many cases) ignore the places where there is no locally linear approximation. We will see this with the very useful nonsmooth sets known as rectifiable sets (see chapter 15).*

The combination with the potential for the most complexity is the case in which the domain, map and range are all nonsmooth and nonlinear.

9.2.3.1 *More Examples*

EVERYTHING LINEAR The usual equation studied in linear algebra courses,

$$Ax = b$$

where $b \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ and A is an m by n matrix defining a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ is studied for a reason – it is exceedingly useful and very general. It is not an exaggeration to say that almost all applied mathematics problems, at one point or another, either reduce to solving $Ax = b$ or to a sequence of such problems. Here is a simple example in which the domain is infinite dimensional – the space of continuous functions on the unit interval:

$$F : C([0,1]) \to \mathbb{R} \text{ defined by } F(g) \equiv \int_0^1 g(x)dx$$

where $C([0,1])$ is the infinite dimensional space of continuous maps from the unit interval to the real numbers, with the sup norm defined by $|f - g| \equiv \sup_{x\in[0,1]} |f(x) - g(x)|$. See Figure 50.

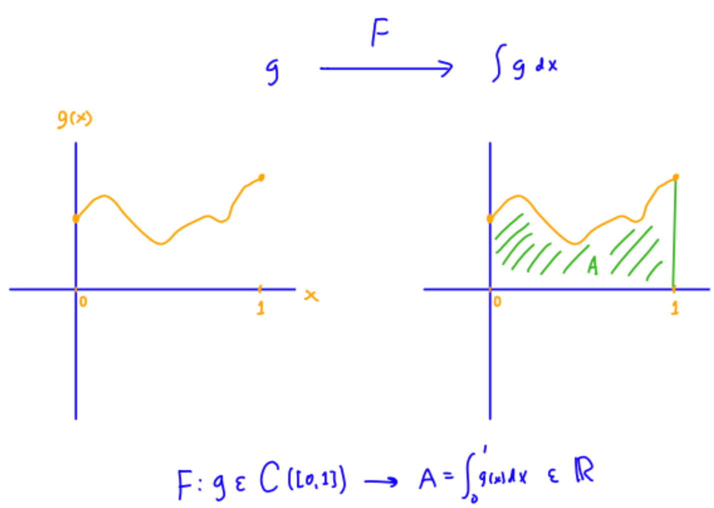


Figure 50: A linear map from $C([0,1])$ to $\mathbb{R}$.

LINEAR SPACES The space $\mathbb{R}^n$ is the most commonly used linear space. Linear and nonlinear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the setting for a great deal of geometric analysis, as is the study of sets and measures in and on $\mathbb{R}^n$. Other common normed linear spaces are usually either Hilbert spaces or Banach spaces. (Recall that *Banach spaces* are complete, normed linear spaces and *Hilbert spaces* are complete, normed linear spaces with an inner product that generates the norm, $|x| = \sqrt{\langle x, x\rangle}$). The following map is a nonlinear map from the Banach space $C([0,1])$ to the Hilbert space $\mathbb{R}$ For example:

$$F : C([0,1]) \to \mathbb{R} \text{ defined by } F(g) \equiv \int_0^1 |g(x)|^3 + |g(x) - 1|^2 dx.$$

NONLINEAR SPACES When the spaces are nonlinear, the map must also be, because the definition of a linear map implicitly assumes both the domain and the range are linear spaces. The most common example in this category are mappings between manifolds – think of surfaces in $\mathbb{R}^n$, like maps between the torus and the sphere in $\mathbb{R}^3$. See Figure 51.

EVERYTHING NONSMOOTH Suppose that the spaces we are interested in are fractal sets in $\mathbb{R}^2$ or if we are interested in slightly less exotic cases, we might pick as domain and range, sets in $\mathbb{R}^2$ which are themselves images of Lipschitz maps, and study Lipschitz maps between those spaces. An example of a mapping between a non-smooth, nonlinear domain and a piece of the extended reals, would be that given by a measure: See Figure 52. Another example of a map between the $\mathbb{R}$ and a nonsmooth, nonlinear space is that induced by the graph of a nonsmooth map. See Figure 53.

Exercise 9.2.1. Think about how nonsmoothness might arise. What processes lead to nonsmooth functions, shapes or dynamics? This is really an invitation to dig around, explore, even speculate. If you need a place to start, you can begin by looking at where, in nature, you would find shapes that seem to be modeled pretty well by fractals. There is another well studied phenomena called **Diffusion-limited aggregation** you might want to look at as well. But there are many

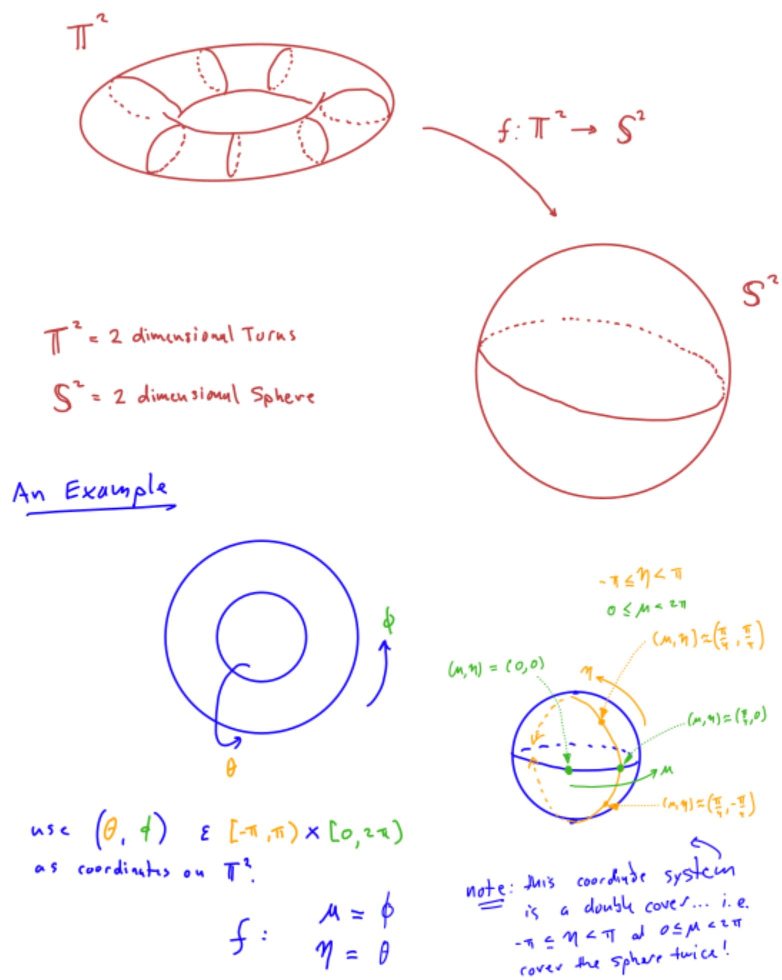


Figure 51: Smooth map between the 2-torus and the 2-sphere.

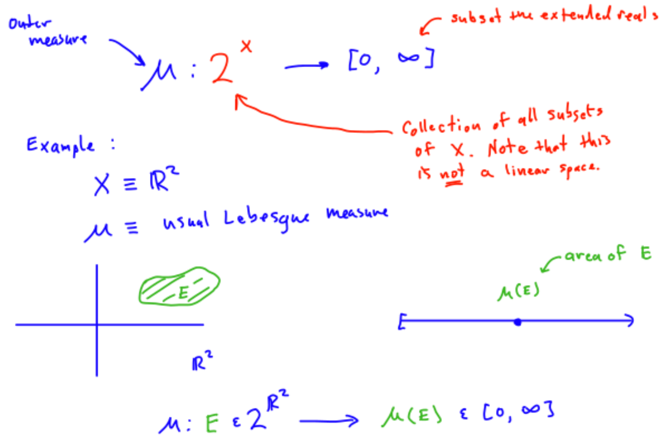


Figure 52

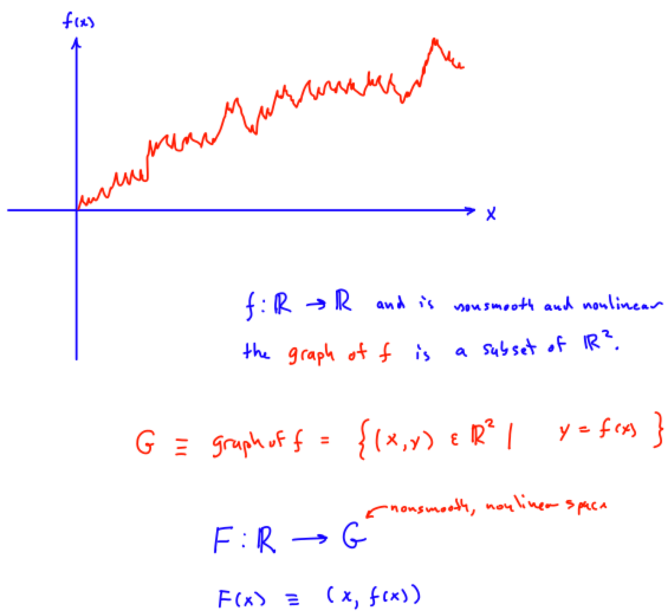


Figure 53: Example of a nonsmooth range.

other places to look, and simply trying to imagine how something could end up being non-smooth is not a bad place to start.

9.2.3.2 *Practical Considerations*

Linear maps are easily and simply specified once we have chosen a basis for the domain and a basis for the range. If the dimension of the domain is n and the dimension of the range is m, then the mapping is specifies by an m-by-n matrix of real numbers. On the other hand, a nonlinear map might be so complicated that we must simply define it by what it does on a very dense set of points in the domain. Suppose that $n = 100$ and $m = 100$, then while a linear map requires a specification of a 100-by-100 matrix, we might need to simply record the values of a nonlinear map on $\left(\frac{1}{\Delta}\right)^{100}$ points, where Δ is the lengthscale at which the map is regular enough to allow samples spaced this far apart to be a good representation of the map. Supposing that the domain is the unit cube in $\mathbb{R}^{100}$ and that $\Delta = .01$. This implies that we must record the value of the map on $100^{100} = 10^{200}$ points – a truly ridiculous number of points since this is greater than the estimated number of particles in the universe, or even the number of edges and vertices in the complete graph whose vertices are the particles in the universe!

Exercise 9.2.2. **(Challenge)** Construct a continuous function $f : \mathbb{R} \to \mathbb{R}$ such that, for any continuous function $g : [0, k] \to \mathbb{R}$ and any $\epsilon > 0$, there is an integer m such that $\max_{x \in [0,k]} |f(x - m) - g(x)| < \epsilon$. Hint: Use the fact that any function on a compact interval $[a, b]$ can be arbitrarily well approximated by a polynomial on that interval. Then use the fact that we can assume the coefficients of the polynomial are rational.

Part IV

Analysis II

Approximations: Metrics, Norms, and Convergence 10

10.1 Why Approximation?

The heart of this part of the book is a detailed first look at differentiation, measure, and integration. The key idea in differentiation is *local approximation*. That idea is also key to understanding nuanced ideas in measure and integration. So we will prepare for the next three chapters by spending a little bit of time looking at approximation, as well as the metrics and norms that measure approximation and convergence.

Of course there are reasons other than the use of approximation in differentiation to understand and use approximation. Here are some examples. In these examples, we are measuring closeness or distance with some norm indicated by $|\cdot|$.

Solving $Ax = b$ When we set about solving $Ax = b$ for $x \in \mathbb{R}^n$, given $b \in \mathbb{R}^k$ and a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^k$, it is often the case that the given b is a noisy or perturbed b, $\hat{b} \equiv b + \epsilon$ for some (hopefully small) $\epsilon \in \mathbb{R}^k$. In this case we will find $\hat{x}$, the solution to $Ax = \hat{b}$. A very reasonable question is, how close is $\hat{x}$ to x? I.e. how big is $|x - \hat{x}|$? where $|\cdot|$ is some norm, for example the Euclidean norm. Things might be more interesting: The matrix we are using could also be a noisy version of A, which we will call $\hat{A}$. Now we are actually solving $\hat{A}x = \hat{b}$ to get $\hat{\hat{x}}$ – i.e. $\hat{A}\hat{\hat{x}} = \hat{b}$. Again, how big is $|x - \hat{\hat{x}}|$? Can we bound it in terms of the sizes of $|A - \hat{A}|$ and $|b - \hat{B}|$?

Finite Dimensional Approximations Very often the sets, functions, and measures we are interested in live in some infinite dimensional space X. To do any computing, we use finite dimensional approximations to X, F_k. That is, we try to find a k-dimensional subspace $F_k \subset X$ such that for the points $x \in X$ of interest or importance, there is a point $f \in F_k$ such that $|x - f| < \delta$.

APPROXIMATION AS A TECHNICAL TOOL We can sometimes prove results using approximation. We will see that locally linear approximation

$$F(x+h) - F(x) \approx D_x F(h)$$

means that results easily shown for linear maps, are also provable for differentiable nonlinear maps. Two important examples to be encountered in the next chapter are the inverse and implicit function theorems.

DEFINITION BY APPROXIMATION Suppose we have a set $Y \subset X$ that is dense in X, i.e.

$$\forall \{x \in X,\ \epsilon > 0\} \quad \exists y \in Y \text{ such that } |x - y| \leqslant \epsilon.$$

and on Y we can define a functional $F : Y \to \mathbb{R}$ that is also continuous on Y, but undefined on X. Then we can define F on X by approximation. Because Y is dense in X, for any $x \in X \setminus Y$, define $F(x) = \lim_{i\to\infty} F(y_i)$ where $\{y_i\}_{i=1}^{\infty}$ is any sequence such that $x = \lim_{i\to\infty} y_i$. Because F is continuous on Y, it does not matter what sequence converging to x that we use.

10.2 Metric Spaces, Again, and What They Give Us

In some ways, the study of metric spaces tell us how much we can learn about spaces and maps between them if all we have is the notion of approximation in the form of the metric. As a result, the entire previous chapter on metric spaces has already demonstrated that approximation can be used quite fruitfully!

Metric spaces also demonstrate what we lose by not having other, additional structures or properties.

For example, in a general metric space X, for any $x, y \in X$, there is no easy way to make sense of $x + y$, let alone $\langle x, y \rangle$ or $\alpha x + (1 - \alpha) y$.

(But see the relatively new area of analysis in metric spaces. See for example [4, 5, 2, 6, 21, 20], the last three of which are most introductory, though they assume a course in analysis.)

We can use an even weaker notion of closeness not derived from a distance to get some of these notions. All we really need are open sets, i.e. a topology. Using only open sets, we can get continuity, connectedness, compactness and convergence.

But a metric encodes the notion of closeness and approximation in a very straightforward way. The fact that we can use topologies (without metrics) to define open and closed sets and then use those to define things like continuity, connectedness, compactness and convergence does not imply that we should dispense with the distance. In many cases, it is more natural to work with the distance and to use it as needed.

In fact, instead of less structure (because we can) we will require more structure, because there are things we want to use that are most naturally obtained when the metric space is also a complete vector space with an inner product – such a space is called a *Hilbert Space*.

More concretely, while the concepts/tools of

1. approximation or closeness – directly encoded in the metric
2. continuity of maps from a space to itself or between two spaces
3. connectedness of a space or subset of a space
4. compactness of a space or subset of a space
5. convergence of a sequence of points

are given to us by the metric, the concepts and tools that require/imply extra structure include:

1. addition of elements in the space (vector space structure)
2. multiplication of points in the space by a scalar (vector space structure)

3 existence of a basis for the space (vector space structure)
4 differentiation (vector space structure)
5 angles (vector space structure and an inner product)
6 spatially invariant metric (e.g. vector space norm)

We next look at the increased structure we will use, after which we will look at an example of the geometric implications of the choice of metric. A penultimate section on the geometry of the particular metrics we will most often be using – vector norms on $\mathbb{R}^n$, is followed by a final section on convergence.

10.3 Finding the Sweet Spot between Generality and Structure

In mathematics, there is almost always a tradeoff between (1) generality and (2) richness of results we can prove and objects (sets, measures) and mappings that can be carefully explored.

In this text we will focus in $\mathbb{R}^n$ and what flows from considering sets, measures and functions/mappings in and on $\mathbb{R}^n$. See Figure 54:

The key point: a vector space structure that comes from an inner product on a space that is both complete and finite dimensional, is enough structure to open the door to a very rich collection of subsets, measures and mappings that can be explored and understood with great precision.

You have already encountered many examples of this wild menagerie in Chapters 4-8

We now look at one particular family of metrics on the space of functions from $\mathbb{R}$ to $\mathbb{R}$.

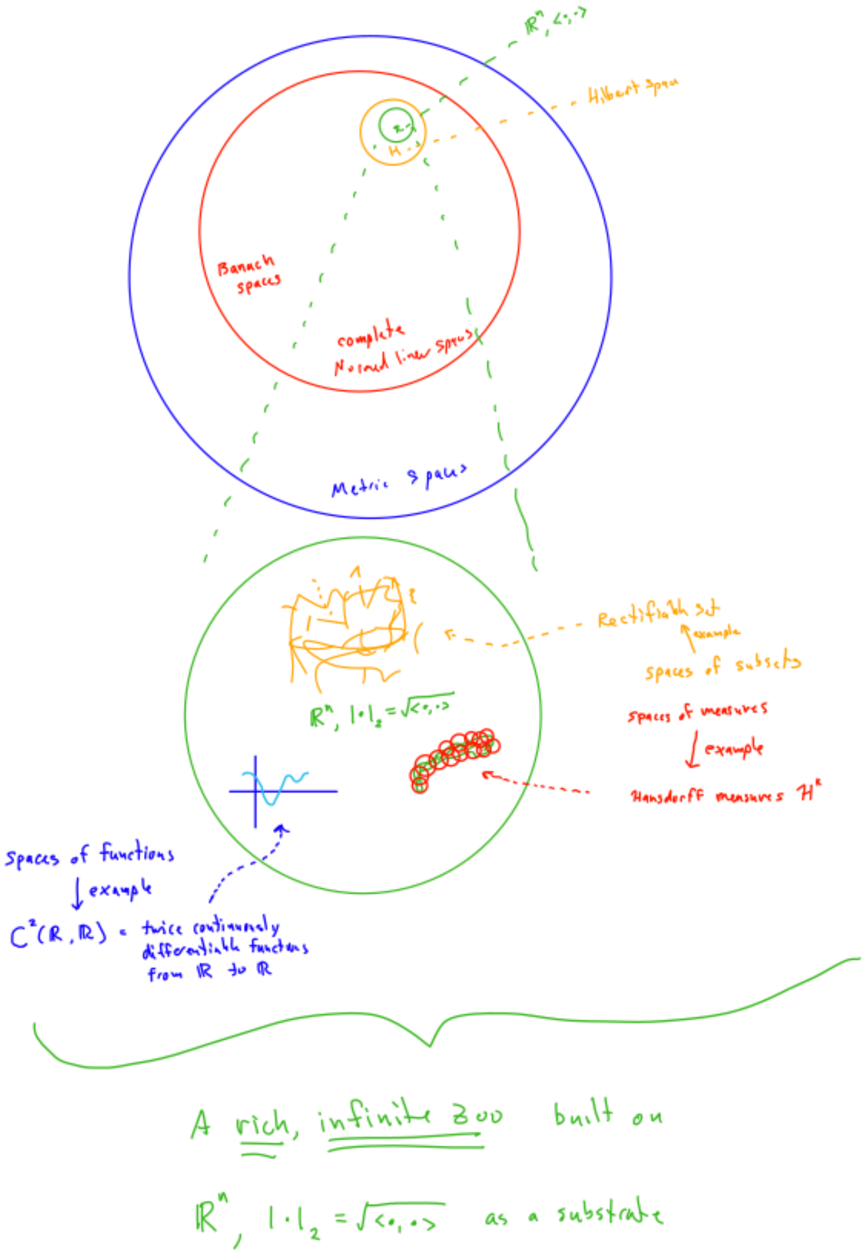


Figure 54: A cartoon of the choice of $\mathbb{R}^n$ as the sandbox we will focus on.

10.4 Geometry imposed by metrics: A (deceptively) simple example

Define the following family of metric-like functionals on the space of differentiable functions $f : \mathbb{R} \to \mathbb{R}$ such that ∇f has compact support (i.e. there is a compact set $K \subset \mathbb{R}$ such that $\nabla f(x) = 0$ for all $x \in \mathbb{R} \setminus K$):

$$|f|_{BV(p)} \equiv \int_{\mathbb{R}} |\nabla f|^p \, dx$$

where $p \geqslant 0$. (It turns out that the case of $p = 1$ is very interesting and filled with nuances, especially when you consider the analog for $f : \mathbb{R}^n \to \mathbb{R}$. We will not get into those details now, but the curious can get a sense for what is involved by consulting chapter 5 of Evans and Gariepy's *Measure Theory and Fine Properties of Functions* [12].)

We will sometimes be interested in using this functional to encourage smoothness. For example, if $d(x)$ is a measured function that we know is noisy, then we could minimize:

$$F_p(f) \equiv \int_{\mathbb{R}} |\nabla f|^p \, dx + \lambda \int_{\mathbb{R}} |f - d| dx$$

so that we are trying to find a function that is not to far from the data (i.e. the second term is not too big) but it is also not too noisy (i.e. the first term, measuring oscillation, is not too big).

This is, in fact, a family of functionals that appears in the signal processing literature, at least for some values of p. We now make a simple observation that $F_1(f)$, taken to its full generality on functions $f : \mathbb{R}^n \to \mathbb{R}$, encounters deep waters (in fact it is the focus of the famous theorem 4.5.9 in Federer's *Geometric Measure theory* [13]).

Consider the family of functions f_δ defined by:

$$f_\delta \equiv \begin{cases} 0 & \forall x < 1 - \delta \\ \frac{1}{\delta}(x - 1 + \delta) & \text{for } x \in [1 - \delta, 1] \\ 1 & x > 1 \end{cases}$$

Computing, we get:

$$|f_\delta|_{BV(p)} = \delta^{1-p}$$

from which we conclude that, while for $p > 1$ we are biased towards smaller slopes (we want big δ's) – smoother functions – and $p < 1$ we are biased towards discontinuities ($\delta = 0$), for $p = 1$ we don't care about how the function gets from 0 to 1 as long as the function is monotonic. Edges (discontinuities – in the limit) are fine. That turns out to have deep ramifications. See Figure 55.

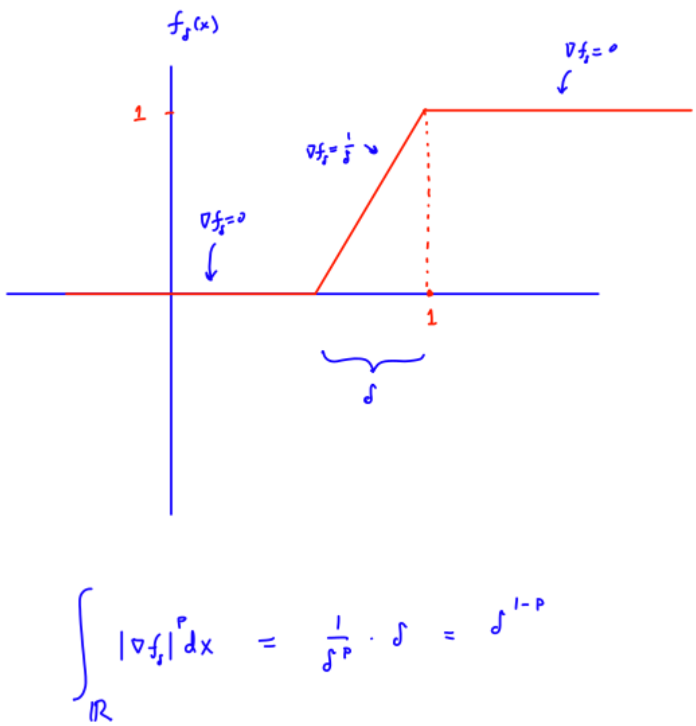


Figure 55: The choice of p determines whether we are biased for jumps, biased against jumps or are agnostic about jumps.

Remark 10.4.1. *The nuances and depth in the study of $F_1(f)$ come from the first term in the function – the BV seminorm, which is defined for functions with discontinuities. Functions for which the first integral is finite are called BV functions. The theory of BV functions is the focus of that famous theorem* *in Federer's book – Theorem 4.5.9 mentioned above. The theorem statement alone takes up three and a half pages!*

Exercise 10.4.1. **(Challenge)** In this exercise we consider $|f|_{BV(p)}$ in the case that $p = 1$, which we refer to as $|f|_{BV}$ – the BV seminorm. The exercise invites you to explore, perhaps resorting to Chapter 5 of Evans and Gariepy's book, [12], after you have spent some significant time thinking about the exercise.

1 Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a piecewise smooth function with a finite number of jump discontinuities, such that $K \equiv \text{clos}(\{x \mid f(x) \neq 0\})$ (the closure of the set where $f \neq 0$) is a compact subset of $\mathbb{R}$. See if you can define ∇f $(= \frac{df}{dx})$ by thinking of it as a the sum of a function $g(x)$ and a finite sum of weighted Dirac measures (i.e. Dirac δ "functions"), $\sum_{i=1}^{M} m_i \delta(x - a_i)$:

$$\nabla f = g(x) + \sum_{i=1}^{M} m_i \delta(x - a_i)$$

such that the integral

$$\int_{\mathbb{R}} |\nabla f| \; dx = \int_{\mathbb{R}} |g(x)| dx + \sum_{i=1}^{M} |m_i| = \lim_{i \to \infty} \int_{\mathbb{R}} |\nabla f_i| dx$$

where $f_i(x)$ is a sequence of Lipschitz functions converging to $f(x)$:

$$\int_{\mathbb{R}} |f(x) - f_i(x)| dx \to 0.$$

Note: the weights m_i are real values.

2 Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$. See if you can't define ∇f as a sum of a smooth, compactly supported vector field and a sum of vectorfields ν_i, multiplying 1-dimensional Hausdorff measure restricted to the jump set (the curves where $f(x)$ has jump discontinuities) in the case that $f(x)$ is the sum of a smooth function and a "cartoon"* – a piecewise

*This terminology comes from the area of image analysis/processing, where a "cartoon" is a piecewise constant image.

constant function with compact support that attains a finite number of positive values $\{\nu_i\}_{i=1}^M$ – and $E_i \equiv f^{-1}(m_i)$ each have smooth, simple closed boundaries ∂E_i and $\partial E_i \cap \partial E_j = \emptyset$ when $i \neq j$. I.e. you are aiming to prove that, for your definition of ∇f:

$$\nabla f \equiv \vec{g}(x) + \sum_{i=1}^{M} \vec{\nu_i}(x) \left(\mathcal{H}^1 \llcorner \partial E_i\right)$$

you have that

$$|f_i|_{BV} \underset{i\to\infty}{\rightarrow} |\nabla f|_{BV} = \int |\vec{g}(x)|dx + \sum_{i=1}^{M} \int_{\partial E_i} |\vec{\nu_i}(x)| d\mathcal{H}^1$$

for smooth approximations of f, f_i. **Remarks:** Note:

a for a measure μ and a set E, $\mu \llcorner E$ means μ restricted to E and is defined by

$$\mu \llcorner E(\Omega) \equiv \mu(\Omega \cap E)$$

b Likewise $f \llcorner E$ is defined to be $f \cdot \chi_E$, where f is a function or a vectorfield, and $f \llcorner \mu$, where f is a function and μ is a measure is defined by:

$$f \llcorner \mu(\Omega) \equiv \int_\Omega f d\mu$$

c You are welcome to assume that the ν_i's are the inward pointing vector fields with constant magnitude ν_i. But to actually figure out what the vectorfields ν_i should be, you need to find the

$$\nu_i \llcorner \partial E_i$$

such that

$$\int_{\mathbb{R}^2} \chi_{E_i} \nabla \cdot \vec{\varphi} = \int_{\mathbb{R}^2} \nabla \chi_{E_i} \cdot \vec{\varphi}.$$

You are essentially figuring out what makes the divergence theorem work for the generalized gradient of a characteristic function χ_E.

10.5 Approximation Theorems: Two Examples

While there is an entire area of mathematics – approximation theory – that deals with approximation questions, we will illustrate the kinds of results possible with two examples: (a) approximation of functions in L^p spaces with smooth functions and (b) the Weierstrass polynomial approximation theorem.

For both of these results, I state the theorems and explain them, but will refer you to my favorite references for proofs and more details. If you are a student of mine, reading this, there is a good chance you already have the references I will recommend.

10.5.1 Approximation in L^p Spaces

For many more details, see Evan's and Gariepy [12], Chapter 4.

Recall from 8.2.7 that for any $1 \leqslant p < \infty$, a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is in $L^p(U, \mathbb{R})$ if

$$\int_U |f|^p \, d\mu < \infty$$

in which case we define the p-norm $|\cdot|_p$ on $L^p(U, \mathbb{R})$ by

$$|f|_p \equiv \left(\int_U |f|^p \, d\mu \right)^{\frac{1}{p}}$$

where μ is often taken to be the usual n-dimensional Lebesgue measure.

Now we (somewhat informally) define a bump function of arbitrary width – See figure 15.3.4.

1 define η_0 = any smooth (i.e. C^∞), non-negative function supported in the unit ball in $\mathbb{R}^n$ that increases as $|x|$ decreases and is symmetric about the origin – i.e. $\eta_0(-x) = \eta_0(x)$.

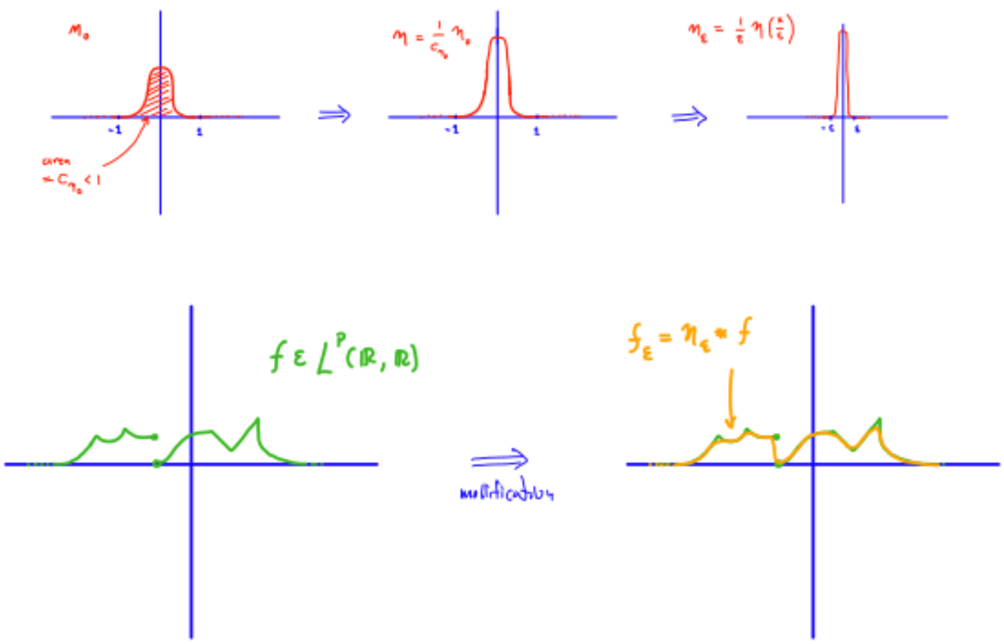


Figure 56: Example of mollification (smoothing) in 1-dimension.

2 Because η_0 is continuous on a compact set, it is bounded, so we know that

$$c_{\eta_0} \equiv \int_{\mathbb{R}^n} \eta_0 \, d\mu = \int_{B(0,1)} \eta_0 \, d\mu < \infty.$$

3 Define

$$\eta(x) = \frac{1}{c_{\eta_0}} \eta_0(x)$$

immediately giving us that

$$\int_{\mathbb{R}^n} \eta(x) d\mu(x) = \int_{B(0,1)} \eta(x) d\mu(x) = 1$$

4 Define the mollifier of radius ϵ to be

$$\eta_\epsilon(x) \equiv \frac{1}{\epsilon^n} \eta\left(\frac{x}{\epsilon}\right)$$

and note that

$$\int_{\mathbb{R}^n} \eta_\epsilon(x) d\mu(x) = \int_{B(0,\epsilon)} \eta_\epsilon(x) d\mu(x) = 1$$

5 The ϵ-mollified version of $f : U \subset \mathbb{R}^n \to \mathbb{R}$ to be:

$$f_\epsilon \equiv \eta_\epsilon * f$$

where the convolution $g * h$ is defined by:

$$g * h(x) \equiv \int_{\mathbb{R}^n} g(x-y) h(y) d\mu(y)$$

Theorem 10.5.1 (Approximation with C^∞ functions; See [12]). *Suppose that $f \in L^p(\mathbb{R}^n, \mathbb{R})$ and that V is any bounded open subset of $\mathbb{R}^n$.*

1 *For all $\epsilon > 0$, $f_\epsilon \in C^\infty(V)$ – that is, f_ϵ is smooth.*
2 *The f_ϵ's converge to f in L_p:*

$$\int_U |f - f_\epsilon|^p \, d\mu \underset{\epsilon \to 0}{\to} 0$$

3 *Suppose that* f *is also continuous. Then, on every compact subset* $K \subset U$, *we have that*

$$\sup_{x \in K} |f - f_\epsilon| \underset{\epsilon \to 0}{\to} 0$$

See Evans and Gariepy's chapter four for full details on this and a great deal more. **Note:** I carefully required f to be in $f \in L^p(U, \mathbb{R})$ with $U = \mathbb{R}^n$ to avoid having to worry about the boundary of U. What [12] has is more general – and there is a lot more to explore there.

Exercise 10.5.1. Prove part 3 of the Theorem 10.5.1 for the case in which $f : \mathbb{R} \to \mathbb{R}$. **Hint:** Use the uniform continuity of f on compact sets.

10.5.2 Weierstrass Approximation Theorem

Theorem 10.5.2 (Weierstrass Approximation Theorem)**.** *Suppose* $f : [a, b] \to \mathbb{R}$ *is continuous. For every* $\epsilon > 0$, *there exists a polynomial* $p_{f,\epsilon}$ *such that*

$$\max_{x \in [a,b]} |f(x) - p_{f,\epsilon}(x)| < \epsilon.$$

Another way to say this is that, using the sup-norm as a metric, the space of polynomials on $[a, b]$ is dense in the space of continuous functions on $[a, b]$:

$$P([a, b], \mathbb{R}) \underset{\text{dense in sup norm}}{\subset} C([a, b], \mathbb{R})$$

There is a beautiful generalization of this theorem for functions on compact Hausdorff spaces, the Stone–Weierstrass Theorem. The Wikipedia page on this topic is good, but my favorite reference is George F. Simmons book *An Introduction to Topology and Modern Analysis*, [35], an exquisitely written reference I recommend all students of analysis own and use.

In the next section we examine the structure of vector norms on $\mathbb{R}^n$.

10.6 Norms and Symmetric, Convex Subsets in $\mathbb{R}^n$

Definition 10.6.1. *A vector norm, or norm $|\cdot|$, on a vector space X, is a function that maps vectors to the non-negative real numbers – $|\cdot| : X \to [0, \infty)$ – that satisfies the following conditions.*

1. $|x| = 0 \Leftrightarrow x = 0$
2. $|x + y| \leqslant |x| + |y|$
3. $|\alpha x| = |\alpha|\,|x|$ *where $|\alpha|$ is the absolute value of α*

Remark 10.6.1. *We will often use $|\cdot|$ to indicate the usual Euclidean norm on $\mathbb{R}^n$. Therefore $|\langle \text{some scalar} \rangle|$ is often the usual absolute value, because absolute value is the Euclidean norm on R^1. We will sometimes label the norms with a subscript – e.g. $|x|_1 \equiv \sum_{i=1}^{n} |x_i|$, the 1-norm on $\mathbb{R}^n$ (this is not the Euclidean norm!). But $|\cdot|$ can refer to any norm we happen to be using, as long as which norm we are using is clear form the context.*

Theorem 10.6.1. *There is a one to one correspondence between vector norms on $\mathbb{R}^n$ and subsets $K \subset \mathbb{R}^n$ such that:*

1. K *is convex*
2. $0 \in K$
3. *if $x \in K$, then $-x \in K$ (i.e. K is symmetric about the origin.)*
4. K *has non-empty interior in the Euclidean metric (in the metric space induced by the Euclidean norm).*
5. K *is compact in the Euclidean metric.*

Proof.
Suppose that we have a norm $\|\cdot\|$ on $\mathbb{R}^n$. Define $K \equiv \{x \mid \|x\| \leqslant 1\}$. We now prove that K satisfies the 5 conditions in the theorem.

K is Convex: if $0 \leqslant \alpha \leqslant 1$ and $||x|| \leqslant 1$ and $||y|| \leqslant 1$, then by norm properties 2 and 3,

$$\begin{aligned} 1 &\geqslant \alpha||x|| + (1-\alpha)||y|| \\ &= ||\alpha x|| + ||(1-\alpha)y|| \\ &\geqslant ||\alpha x + (1-\alpha)y|| \end{aligned}$$

implying that $x, y \in K \Rightarrow \alpha x + (1-\alpha)y \in K$.

$0 \in K$: By property one of norms $||0|| = 0 < 1$.

$x \in K$ implies $-x \in K$: $||-1 \cdot x|| = |-1|\,||x|| = ||x||$

The Interior of K is not empty: We prove this in steps:

1 Define $e_1 = (1, 0, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$, ... , $e_n = (0, 0, ..., 0, 1)$

2 Because, for any $x \in \mathbb{R}^n$, $x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$, we know that

$$||x|| \leqslant |x_1|\,||e_1|| + |x_2|\,||e_2|| + \cdots + |x_n|\,||e_n||.$$

3 Define $M = \max_i ||e_i||$ and let $x \in B(0, \frac{1}{nM})$, the ball of radius $\frac{1}{nM}$ in the Euclidean metric (norm).

4 from Step 2, we get that for all $x \in B(0, \frac{1}{nM})$,

$$\begin{aligned} ||x|| &\leqslant \frac{1}{nM}M + ...\frac{1}{nM}M \\ &= n\frac{1}{nM}M \\ &= 1 \end{aligned}$$

implying that $B(0, \frac{1}{nM}) \subset K$.

K compact in the Euclidean metric: We use continuity to prove this.

1 Claim: $||\cdot|| : \mathbb{R}^n \to \mathbb{R}$ is continuous in the Euclidean metric.

(a) $||x+h|| \leqslant ||x|| + ||h||$ and $||x+h-h|| \leqslant ||x+h|| + ||-h|| = ||x+h|| + ||h||$ which together give $|\,||x+h|| - ||x||\,| \leqslant ||h||$.

(b) Now suppose that $h \in B(x, \epsilon)$, the Euclidean epsilon ball centered at x. The previous part proves that $||h|| \leqslant nM\epsilon$.

(c) thus

$$\begin{aligned} |\,||x+h|| - ||x||\,| &\leqslant ||h|| \\ &\leqslant nM|h|_{\text{Euclidean}}. \end{aligned}$$

(d) Thus, $||\cdot||$ is continuous in the Euclidean metric.

2 This immediately gives us that $K = ||\cdot||^{-1}([0,1])$ is closed in the Euclidean norm.

3 Now we show that K is bounded

(a) Because $\partial B(0,1)$ is compact (in the Euclidean norm), $0 \notin \partial B(0,1)$ and $||\cdot||$ is continuous by Step 1, we have

$$m \equiv min_{x \in \partial B(0,1)} ||x|| > 0.$$

(b) Now Define $K_m = \{x \mid ||x|| \leqslant m\}$.

(c) By the previous line and the fact that $||\alpha x|| = \alpha ||x||$ for all $0 < \alpha$, we have that $K_m \subset \bar{B}(0,1)$.

(d) This implies that $K \subset \bar{B}(0, \frac{1}{m})$ and so K is bounded in the Euclidean metric.

We conclude that K is compact in the Euclidean metric.

Now suppose that K satisfies the 5 properties in the theorem. You are asked, in the next exercise, to show that $||x||_K \equiv \frac{1}{k(x)}$ where $k(x) \equiv max\{\alpha > 0 \mid \alpha x \in K\}$, is a norm. This concludes the proof! □

Exercise 10.6.1. Show that $||x||_K \equiv \frac{1}{k(x)}$ where $k(x) \equiv max\{\alpha > 0 \mid \alpha x \in K\}$, is a norm.

Hints for analytic approach: In Steps ...

1 The only difficult piece is showing that $||x+y||_K \leqslant ||x||_K + ||y||_K$, so begin by showing that (a) $||x||_K = 0 \Leftrightarrow x = 0$ and (b) $||\alpha x||_K = |\alpha|\, ||x||$

2 choose $\beta > 0$ and $\mu > 0$ so that $||\beta x||_K = 1$ and $||\mu y||_K = 1$

3 Now use the fact that K is convex to play with

$$||\alpha\beta x + (1-\alpha)\mu y||_K$$

4 Further hints ...

5 the facts that $||\beta x||_K = 1$ and $||\mu y||_K = 1$ and $\alpha\beta x + (1-\alpha)\mu y \in K$ (prove these facts!) implies that

$$\begin{aligned} ||\alpha\beta x + (1-\alpha)\mu y||_K \leqslant 1 &= \alpha||\beta x||_K + (1-\alpha)||\mu y||_K \\ &= \alpha\beta||x||_K + (1-\alpha)\mu||y||_K \end{aligned}$$

6 Now choose α such that $\alpha\beta = (1-\alpha)\mu$ and use this to get that $\|x+y\|_K \leqslant \|x\|_K + \|y\|_K$.

Hints for geometric approach: In Steps ...

1 Again, the only difficult piece is showing that $\|x+y\|_K \leqslant \|x\|_K + \|y\|_K$, so begin by showing that (a) $\|x\|_K = 0 \Leftrightarrow x = 0$ and (b) $\|\alpha x\|_K = |\alpha|\,\|x\|_K$.

2 The approach is to show that the epigraph of $\|\cdot\|_K : \mathbb{R}^n \to \mathbb{R}$ is convex by finding supporting hyperplanes for every point in the graph of $\|\cdot\|_K$.

3 Since K is convex and has a nonempty interior, show that for any point $k \in \partial K$, the supporting hyperplane at k, H_k generates a supporting hyperplane for the graph of $\|\cdot\|_K$ in $\mathbb{R}^{n+1}$ when combined with the vector $(k, 1) \in \mathbb{R}^{n+1}$. I.e. $\mathrm{span}(H_k, (k, 1))$ is a supporting hyperplane of the graph of $\|\cdot\|_K$ in $\mathbb{R}^{n+1}$.

4 Because this proves that $\|\cdot\|_K$ is a convex function, it immediately follows that $\|\frac{1}{2}x + \frac{1}{2}y\|_K \leqslant \frac{1}{2}\|x\|_K + \frac{1}{2}\|y\|_K$ from which the desired result follows!

Exercise 10.6.2. Use the theorem to prove that for any two vector norms on $\mathbb{R}^n$, $|\cdot|_a$ and $|\cdot|_b$, we have that there are constants $0 < c < C < \infty$ such that

$$c|\cdot|_b \leqslant |\cdot|_a \leqslant C|\cdot|_b$$

which implies that the open sets induced by the norms are the same and convergence in one metric is the same as convergence in the other:

$$\{|x^* - x_i|_a \to 0\} \Leftrightarrow \{|x^* - x_i|_b \to 0\}.$$

Note: This result is not true when we are working in infinite dimensional spaces and is a source for many interesting nuances in functional analysis – the study of infinite dimensional spaces of functions.

10.7 The Many Ways of Measuring Convergence

As mentioned in exercise 10.6.2, different metrics (often, these are vector space norms), on a fixed space X, can lead to distinct notions of convergence when the space is not finite dimensional. That is, if $\{x_i\}_{i=1}^{\infty} \subset X$, $x* \in X$, and two different space-norm combinations, $(X, |\cdot|_a)$ $(X, |\cdot|_b)$, we can have that

$$\{|x^* - x_i|_a \to 0\} \not\Leftrightarrow \{|x^* - x_i|_b \to 0\}$$

Exercise 10.7.1. Consider the space of functions $f : [0, 1] \to \mathbb{R}$ with norms

$$|f|_1 \equiv \int_{[0,1]} |f| d\mu \text{ and } |f|_2 \equiv \left(\int_{[0,1]} |f|^2 d\mu \right)^{\frac{1}{2}}$$

Define $f^* \equiv 0$ and $f_i \equiv \sqrt{\frac{2}{4^i}} \frac{1}{x^{\frac{1}{2} - \frac{1}{4^i}}}$. Show that:

1 $|f^* - f_i|_1 \underset{i \to \infty}{\to} 0$
2 $|f^* - f_i|_2 = 1 \ \forall i$

In the case when X is the space of functions $F : \mathbb{R}^n \to \mathbb{R}$ with some norm $|\cdot|_X$ and a measure μ that measures $\mathbb{R}^n$, i.e. $\mu(E)$ is the size of $E \subset \mathbb{R}^n$, we have a variety of commonly used measures of convergence. **Reminder**: $|\cdot|$ is the Euclidean norm on $\mathbb{R}^n$, unless otherwise specified or implied by the context. In the case of $n = 1$, the Euclidean norm is the usual absolute value function.

NORM CONVERGENCE: We say that $\{f_i\}_{i=1}^{\infty} \subset F$ *converges in norm* to $f^* \in F$ if

$$|f^* - f_i|_X \underset{i \to \infty}{\to} 0.$$

POINTWISE CONVERGENCE: We say that $\{f_i\}_{i=1}^{\infty} \subset F$ *converges pointwise* to $f^* \in F$ if, for every $x \in \mathbb{R}^n$,

$$|f^*(x) - f_i(x)| \underset{i \to \infty}{\to} 0.$$

CONVERGENCE IN MEASURE: We say that $\{f_i\}_{i=1}^{\infty} \subset F$ *converges in measure* to $f^* \in F$ if, for all $\alpha > 0$,

$$\mu(\{x \mid |f^*(x) - f_i(x)| > \alpha\}) \underset{i \to \infty}{\to} 0.$$

WEAK CONVERGENCE: Suppose that $F \equiv \{f \mid \int_{\mathbb{R}^n} |f(x)| d\mu < \infty\}$, and that $\{f_i\}_{i=1}^{\infty} \cup \{f^*\} \subset F$. Then we say that $\{f_i\}_{i=1}^{\infty}$ *converges weakly* to f^* if

$$\int_{\mathbb{R}^n} f_i \phi d\mu \quad \underset{i\to\infty}{\to} \quad \int_{\mathbb{R}^n} f^* \phi du$$

for every compactly supported $\phi \in C^{\infty}(\mathbb{R}^n; \mathbb{R})$.

At this point, simple awareness of these different measures of convergence is the goal – more exposure, deeper exposure will come later in the book. Here is an exercise to begin the exploration of these ideas:

Exercise 10.7.2. For each part of this exercise, find a sequence of functions $\{f_i\}_{i=1}^{\infty}$, $f_i : [0, 1] \subset \mathbb{R} \to \mathbb{R}$, such that:

1. $|f_i|_X \equiv \int_{[0,1]} |f_i| dx \underset{i\to\infty}{\to} \infty$ but $\{f_i\}_{i=1}^{\infty}$ converges pointwise $f^* \equiv 0$.
2. $\{f_i\}_{i=1}^{\infty}$ converges in norm to $f^* \equiv 0$, but $\{f_i\}_{i=1}^{\infty}$ does not converge pointwise to $f^* \equiv 0$ anywhere: I.e.

$$\forall\, x \in [0, 1], \qquad |f^*(x) - f_i(x)| \underset{i\to\infty}{\not\to} 0.$$

11 Derivatives: A Path into Geometric Analysis

11.1 The Derivative

Definition 11.1.1 (Derivative). *Suppose that* X *and* Y *are normed vector spaces, with norms* $|\cdot|_X$ *and* $|\cdot|_Y$, *and* $f : X \to Y$. *Then we say that* f *has the derivative* $D_x f \equiv A$ *at* $x \in X$, *if there is a linear operator* $A : X \to Y$ *such that:*

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

where

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

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

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

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

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

For any $w(h)$ satisfying this last condition, we say "$w(h)$ is in $o(h)$", which read literally as "$w(h)$ is in little o of h". Here is a reminder of that definition (which we first saw on page 159).

Definition 11.1.2 (little o of h, $o(h)$). *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".*

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

Suppose that

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

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

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

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

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

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

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

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

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

$$\begin{aligned}
F(u+h) - F(u) &= \int_\Omega \nabla u \cdot \nabla u \; dx + 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h \; dx - \int_\Omega \nabla u \cdot \nabla u \; dx \\
&= 2\int_\Omega \nabla u \cdot \nabla h + \int_\Omega \nabla h \cdot \nabla h \; dx \\
&= 2\alpha_h \int_\Omega \nabla u \cdot \nabla g + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g \; dx \text{ (using 4)} \\
&= -2\int_\Omega \Delta u \; (\alpha_h g) + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g \; dx \text{ (Divergence Theorem)} \\
&= -2\int_\Omega \Delta u \; h + \alpha_h^2 \int_\Omega \nabla g \cdot \nabla g \; dx
\end{aligned}$$

where the note "(Divergence Theorem)" indicates we have used the vector calculus version of integration by parts (which uses the divergence theorem) for that step. To see this, notice that by the divergence theorem, we get

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

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

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

Evaluating the left hand side in a different way, we get:

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

which, combining the two different evaluations of the left hand side, gives us

$$\int_\Omega \nabla g \cdot \nabla u \; dx = -\int_\Omega g\Delta u \; dx.$$

Restating where we are:

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

we first recognize that

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

is linear in h so we define

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

which lets us conclude that:

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

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

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

and we get that

$$\mathbf{F}(\mathbf{u}+\mathbf{h}) - \mathbf{F}(\mathbf{u}) = \mathbf{L_u}(\mathbf{h}) + \mathbf{r}(\mathbf{h}).$$

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

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

Note that when we fixed g and varied α_h in order to change h, this resulted in us using $\alpha_h \to 0$ to get $|h| \to 0$. And in doing this, we chose one, 1-dimensional path to 0. (That is, we ended up calculating a directional derivative.) We note though, that the derivative at u, depends only on u, not the perturbation. And that in fact, we are assuming u and h live in the normed space with norm given by

$$|w|_* \equiv \left(\int_\Omega |w(x)|^2 \, dx + \int_\Omega |\nabla w(x)|^2 \, dx + \int_\Omega |\Delta w(x)|^2 \, dx \right)^{\frac{1}{2}}$$

for any function $w : \Omega \subset \mathbb{R}^n \to \mathbb{R}$.

Exercise 11.2.1. Show that **if**

1. we define L_u as before,
2. use this new norm, $|\cdot|_*$
3. recall that we just calculated:

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

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

we can conclude that

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

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

11.3 Jacobian Matrices

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

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

where

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

A practically important question is "**How do we compute $\mathbf{L_x}$ from $\mathbf{f(x)}$?**"

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

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

How we go about showing this is true: We will first show that (1) if there is a linear function L_x such that $f(x+h) - f(x) - L_x(h) \sim o(h)$ – i.e. f is differentiable at x, L_x must be the matrix of partial derivatives of f at x, and then we show that (2) if f has continuous partial derivatives, then f is differentiable.

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

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

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

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

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

Notice first that this is really two equations:

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

and

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

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

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

and if we divide by h_1, we get:

$$\begin{aligned} \frac{f_1(x_1+h_1, x_2) - f_1(x_1, x_2)}{h_1} - a &= \frac{r_1((h_1,0))}{h_1} && (16) \\ &\leqslant \frac{|r_1((h_1,0))|}{|h_1|} && (17) \\ &= \frac{|r_1(h)|}{|h|} && (18) \\ &\to 0 \text{ (as } |h| \to 0 \text{).} && (19) \end{aligned}$$

But this is just saying that

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

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

Now setting $h_1 = 0$, we get

$$\begin{aligned}
\frac{f_1(x_1, x_2 + h_2) - f_1(x_1, x_2)}{h_2} - b &= \frac{r_1((0, h_2))}{h_2} && (20)\\
&\leqslant \frac{|r_1((0, h_2))|}{|h_2|} && (21)\\
&= \frac{|r_1(h)|}{|h|} && (22)\\
&\to 0 \text{ (as } |h| \to 0 \text{)} && (23)
\end{aligned}$$

and conclude that

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

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

In a completely analogous way, we get that

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

and

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

so that we have:

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

And that is what we set out to show.

11.3.2 *If f has continuous partial derivatives, f is differentiable.*

The four ingredients we need for this part are:

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

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

which can be rewritten in the equivalent form

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

where $b - a = h$.

2 **A function s, bounding the convergence of a collection of functions, each continuous at x.** If
 a $g_j : \mathbb{R}^n \to \mathbb{R}$ (for $j = 1, 2, ...k$)
 b $\lim_{y \to x} g_j(y) = g_j(x)$ (for $j = 1, 2, ..., k$),
then there is a function $s : [0, \infty] \to [0, \infty]$ such that:
 a s is monotonically increasing: $\omega_1 < \omega_2 \Rightarrow s(\omega_1) \leqslant s(\omega_2)$.
 b $\lim_{\omega \to 0} s(\omega) = 0$
 c and

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

 where $B(x, \omega)$ is the ball centered at x with radius ω.

Exercise 11.3.1. Prove that such an $s(\omega)$ exists.

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

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

4 **The realization that we only need to prove the assertion for a function $f : \mathbb{R}^n \to \mathbb{R}$:** because the general case of $f : \mathbb{R}^n \to \mathbb{R}^m$ is just a

collection of m functions $f_i : \mathbb{R}^n \to \mathbb{R}$. That is, the assertion that there is an L_x such that:

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

where
a $f : \mathbb{R}^n \to \mathbb{R}^m$,
b $r(h) \sim o(h)$ and
c $L_x : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function,
is completely equivalent to the assertion that, for $i = 1, 2, ...m$ there is an L_x^i such that:

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

where
a $f_i : \mathbb{R}^n \to \mathbb{R}$,
b $r_i(h) \sim o(h)$ and
c $L_x^i : \mathbb{R}^n \to \mathbb{R}$ is a linear function.

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

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

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

$$f(x_1+h_1, ..., x_n+h_n) - f(x_1, ..., x_n) - (\partial_{x_1} f \cdot h_1 + ... + \partial_{x_n} f \cdot h_n) = r(h) \tag{24}$$

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

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

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

is the sum of n pieces:

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

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

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

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

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

this then allows us to write the left hand side of Equation (24) as

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

or equivalently as

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

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

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

due to the fact that, for all i,

$$|\hat{c}_i - x| \leqslant |h|.$$

Now, noting that for all i**,** $|h_i| \leqslant |h|$**, allows us:**, finally, to compute a bound for the left hand side of Equation (24):

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

This implies that

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

which implies that

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

which means that

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

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

11.3.3 Some More Exercises

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

Exercise 11.3.2. Find a function $f : \mathbb{R} \to \mathbb{R}$ that is

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

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

Exercise 11.3.3. **(Challenge)** Find a function $f : [0, 1] \subset \mathbb{R} \to [0, 1]$ that is:

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

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

Exercise 11.3.4. Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that f is differentiable at $x = a$.

1 Show that, given an angle θ, we can choose $\delta(\theta) > 0$ small enough so that for all x such that $|x - a| < \delta(\theta)$ we have that the graph of $f(x)$ lies inside of the cone with angle θ around the tangent line. (See Figure 57.)
2 Can you find explicit formulas for $\delta(\theta)$ for the function $f(x) = c_1 x^2 + c_2 x + c_3$ for any arbitrary a?

Following the hints below is merely one path to solution, a path I happen to like, but it is probably not the easiest way to solve the problem. You should explore for yourself and find your own path or follow the path of the hints if it looks interesting to you. **Hints:** (a) First solve for $g(h) = f(a+h) - f(a) - L_a(h)$ where, of course $L_a(h) = f'(a)h$. (b) Prove that a triangle that is obtained by a base of length h and a constant (horizontal, in the figure) height of L has maximal apex angle when the base is bisected (alternatively, the apical angle is bisected) by the x-axis. (See Figure 58.) (One way to do that is show that the maximal area underneath the curve $y = \frac{d(\arctan(x))}{dx} = \frac{1}{1+x^2}$, over an interval of length h is obtained when that interval is centered on the origin.) (c) See Exercises 11.3.6 to 11.3.8 for help in using the idea in Figure 58 to solve the problem.

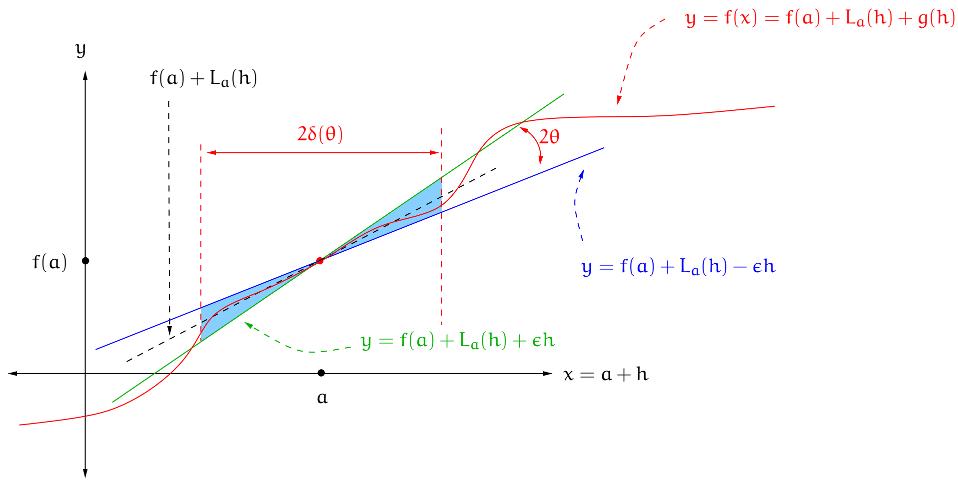


Figure 57: (For Exercise 11.3.4): Here is a picture to stimulate your thoughts and explorations. $L_a(h)$ is a linear function from $\mathbb{R}$ to $\mathbb{R}$ – a line through the origin.

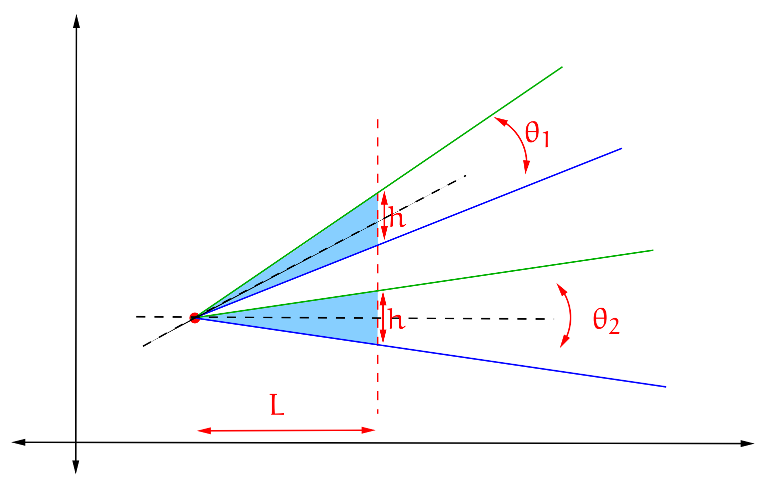


Figure 58: (For Exercise 11.3.4): A triangle whose height (sideways height in this picture) is L and base is a constant h has a maximal angle at the apex (the point furthest to the left) when that apex is bisected by the x-axis. I.e. You are trying to show that $\theta_2 > \theta_1$.

Definition 11.3.1 (Defintion of Lipschitz – Reminder)**.** *If* $f : E \subset X \to Y$ *and*

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

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

Exercise 11.3.5. (Challenge) Suppose that, $f''(a)$, the second derivative of $f : \mathbb{R} \to \mathbb{R}$ at $x = a$, exists. Show that there is some interval around a, $[a - \delta, a + \delta]$ on which f is Lipschitz. **Hints**: first show that in some interval $(a - 2\delta, a + 2\delta)$, the derivative exists and is bounded. Then, for every point in that interval, deduce that there is a narrow cone that works for a possibly tiny interval around it. Get a finite open cover of $[a - \delta, a + \delta]$ using those small intervals and deduce the desired conclusion.

The next three exercises are inspired by Exercise 11.3.4.

Exercise 11.3.6. Suppose that $f(x) > 0$ is continuous for all x and we define $A(d, x^*) = \int_{x^*}^{x^*+d} f(x)dx$. Show that:

1 $A(d, x^*)$, as a function of x^*, is just the area under the curve over a the fixed length interval $[x^*, x^* + d]$ that slides along the x-axis as we change x^*.

2
$$\frac{dA(d,x)}{dx} = f(x+d) - f(x).$$

Exercise 11.3.7. Use the results of Exercise 11.3.6 to show that if we define:

$$\begin{aligned}
\theta_-^d(x) &\equiv \arctan(x) - \arctan(x - d) \\
\theta_+^d(x) &\equiv \arctan(x + d) - \arctan(x)
\end{aligned}$$

and remember that

$$\arctan(x) = \int_0^x \frac{1}{1+s^2} ds$$

we can conclude that

$$\theta_+^d(x) \leqslant \theta_-^d(x)$$

for all $x \geqslant 0$.

Exercise 11.3.8. Continue with Exercise 11.3.7, again using the results of Exercise 11.3.6 to prove that

$$\theta_+^d(-d/2) \geqslant \theta_+^d(x) \ \ \forall x.$$

11.4 Derivatives and Intersections

Here are three exercises to get us started.

11.4.1 Warm up Exercises

Exercise 11.4.1. **(Without Hints = Challenge)** Suppose that we denote the number of points in a set E by $|E|$ and we have that

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

Prove: that $|X_c|$ is finite.
Hint: use (1) an assumption that all derivatives are non-zero **and** $|X_c| = \infty$ (in an effort to get a contradiction) , (2) the compactness of [0,1], (3) the continuity of f, (4) the cone property you have been asked to prove in part 1 of **Exercise (11.3.4)** .

Exercise 11.4.2. Again, suppose that we denote the number of points in a set E by $|E|$ and we have that

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

Prove: that $|X_c|$ is finite. Hint: Suppose that $N \equiv \left|\{x \mid \frac{df}{dx}(x) = 0\}\right|$ and that $|X_c| \geqslant N + 2$. Find a contradiction using the mean value theorem. (See page 229 for a reminder of the mean value theorem.)

Exercise 11.4.3. **(Challenge)** Let's see if we can bound the number of points in $|X_c|$:

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

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

11.4.2 The Theory

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

Recall from the chapter on metric spaces, that we define $f^{-1}(A) \equiv \{x \mid f(x) \in A \subset Y\}$ and we refer to $f^{-1}(A)$ as the "inverse image of A under f", even if there is no function such that $g(y) = f^{-1}(y)$ for all $y \in Y$, i.e. even if f is not invertible.

Definition 11.4.1 (Level Sets). *A level set of* $f : E \subset X \to Y$ *is any set of the form* $X_c \equiv \{x \mid f(x) = c \in Y\}$*. The set* X_c *is sometimes called the* c*-level set of* f *and is also denoted by* $f^{-1}(c)$*, the inverse image, under* f*, of the point* $c \in Y$*.*

Definition 11.4.2 (Regular Level Sets for Functions $f : E \subset \mathbb{R} \to \mathbb{R}$**).** *A level set of a function* $f : \mathbb{R} \to \mathbb{R}$*,* $X_c \equiv \{x \mid f(x) = c\}$*, is called a* **regular level set** *if, for every* $y \in X_c$ *there exists an open interval* $(y - \delta_y, y + \delta_y)$ *with* $\delta_y > 0$ *such that* $(y - \delta_y, y + \delta_y) \cap X_c = \{y\}$*.*

Definition 11.4.3 (Regular Values: $f : E \subset \mathbb{R} \to \mathbb{R}$**).** *Suppose that* $f : E \subset \mathbb{R} \to \mathbb{R}$ *and that* $X_c \equiv \{x \mid f(x) = c\}$*. If every derivative on the level set is non-zero: I.e.* $y \in X_c \Rightarrow |\frac{df}{dx}(y)| \neq 0$*, we say that* c **is a regular value of** f*.*

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

Theorem 11.4.1 (Regular Level Sets). *Level sets defined by regular values are regular.*

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

Now we give the generalizations to the case in which the spaces X and Y in Definition (11.4.1) are given by $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$.

Definition 11.4.4 (Regular Level Sets for Functions $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$**).** *Define* $k \equiv \max(n - m, 0)$*. A level set of a function* $f : \mathbb{R}^n \to \mathbb{R}^m$*,* $X_c \equiv \{x \mid f(x) = c\}$*, is called a* **regular level set** *if, for every* $y \in X_c$ *there exists an open ball* $B(y, \epsilon)$*, centered at* y *with radius* ϵ*, such that* $B(y, \epsilon) \cap X_c$ *is well approximated by* $B(y, \epsilon) \cap \{y + V_y\} \cap E$ *where* V_y *is a* k*-dimensional subspace of* R^n*. (Well approximated means that there is a* *smooth change of coordinates, converging to the identity map as* $\epsilon \to 0$*, mapping these two sets bijectively onto each other.)*

Definition 11.4.5 (Regular Values: $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$). *Suppose that* $f : E \subset \mathbb{R}^n \to \mathbb{R}^m$ *and that* $X_c \equiv \{x \mid f(x) = c\}$*. If every derivative on the level set is full rank: I.e.* $y \in X_c \Rightarrow$ *rank*$(D_y f) = \min(m, n)$*, we say that* c **is a regular value of** f*. In that case, for all* $y \in X_c$*, the* V_y *in definition (11.4.4) equals* $D_y f(0)^{-1}$.

Exercise 11.4.4. See if you can show that Definitions (11.4.2) and (11.4.3) are special cases of Definitions (11.4.4) and (11.4.5).

11.5 Three Integrals of Derivatives

We begin with a very simple smooth function $f : [0,1] \to [0,1]$ (see Figure 59) which we probe with three integrals, the generalizations

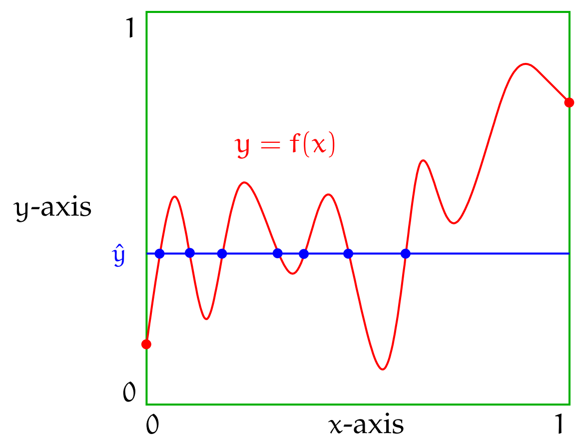


Figure 59: The level set $X_{\hat{y}}$ has 7 elements, shown as 7 blue dots in this figure. The same figure can be used to illustrate each of the three integrals.

of which turn out to be deeply important tools for nonlinear, geometric analysis ...

Degree Theory

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

$\rightarrow$ special case of degree theory

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

Remark 11.5.1. *The integral immediately above need only be over the set* $f([0,1])$ *but because* $f([0,1]) \subset [0,1]$*, integrating from* 0 *to* 1 *works.*

Area/Coarea

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

$\rightarrow$ special case of area and coarea formulas

Stokes Theorem

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

$\rightarrow$ simple case of divergence theorem

$\rightarrow$ which is itself a simple case of Stokes Theorem

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

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

$\operatorname{rank}(D_y f) = \min(n, m) \;\; \forall \;\; y \in X_c$

$\rightarrow$ Sard's Theorem also comes up

$\rightarrow$ Which brings up the 5R covering theorem

$\rightarrow$ Which becomes a good place to begin looking at outer measures

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

$$(B(y,\epsilon) \cap \{y + V_y\} \cap E) \sim (B(y,\epsilon) \cap X_c) \quad \forall \ y \in X_c$$

$\to$ Really the same idea as Derivative = linear approximation

$\to$ Introduces Manifolds

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

$$\left(B(y,\epsilon) \cap \{y + D_y f^{-1}(0)\} \cap E\right) \sim (B(y,\epsilon) \cap X_c) \quad \forall \ y \in X_c$$

$\to$ Level sets corresponding to Regular values = manifolds

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

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

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

... where the Jacobian $J^* f \equiv \begin{cases} \sqrt{Df^t \circ Df} & n < m \\ \sqrt{Df \circ Df^t} & n \geqslant m \end{cases}$

$\to$ a very powerful general tool for tracking and computing mapped volumes

$\to$ We encounter outer measures and Hausdorff measures in earnest here!

Stokes Theorem – Briefly

$$\begin{aligned} \int_{\partial\Omega} \omega \quad &= \quad \int_\Omega d\omega \text{ (Stokes Theorem)} \\ &\to \quad \int_{\partial\Omega} v \cdot \vec{n} \, d\sigma = \int_\Omega \nabla \cdot v \, dx \text{ (Divergence Theorem)} \\ &\to \quad \oint_{\partial\Omega} \vec{v} \cdot T_{\partial\Omega} = \int_\Omega \nabla \times \vec{v} \, dx \text{ (Little Stokes Theorem)} \end{aligned}$$

Remark 11.5.2. *In the statement of the general Stokes Theorem immediately above, ω is an n-1 form, $d\omega$ is the exterior derivative of ω and is therefore an n form. For the definitions of forms, exterior derivatives and integration of forms, see Chapter 7 of Fleming [14].*

We will return to these three integrals later in the text.

11.6 Taylor Series

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

11.6.1 Mean Value Theorem Approach I

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

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

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

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

1 Begin with the Mean Value Theorem:

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

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

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

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

3 Integrate this to get:

$$\begin{aligned} \int_0^h f'(x+t)\ dt - \int_0^h f'(x)\ dt &= \int_0^h f''(c(t))t\ dt & (25) \\ \to\ f(x+h) - f(x) - f'(x)h &= \int_0^h f''(c(t))t\ dt. & (26) \end{aligned}$$

4 Define

$$\begin{aligned} f''_m &= \min_{s\in[x,x+h]} f''(s) \\ f''_M &= \max_{s\in[x,x+h]} f''(s). \end{aligned}$$

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

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

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

6 Now define I_h^f by

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

to get

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

which implies

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

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

8 We immediately have that Equation (26) can be rewritten:

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

9 In general, we have that

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

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

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

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

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

which lets us conclude following our steps, exactly, that

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

which, in turn, leads by steps 2-8 to

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

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

11.6.2 Mean Value Theorem Approach II

Here is a shorter proof following page 386 of Fleming's book. It assumes slightly less: we assume only that the function has an $(n+1)$th derivative everywhere in the interval $[x,x+h]$, not that it is continuous.

Proof.
Define

$$\begin{aligned} G(y) &= f(x+h)-f(y)-f'(y)(x+h-y)-\frac{f''(y)}{2!}(x+h-y)^{2} \\ &\quad -\cdots-\frac{f^{(n)}(y)}{n!}(x+h-y)^{n}-\frac{K}{(n+1)!}(x+h-y)^{n+1} \end{aligned}$$

with K chosen so that $G(x) = 0$. Differentiting G, we get that

$$G'(y) = \frac{(x+h-y)^n}{n!}(-f^{(n+1)}(y) + K)$$

and from the facts that $G(x+h) = G(x) = 0$ and $G'(y)$ exists 0 the interval between x and $x+h$, we have that there is a $c(h) \in (x, x+h)$ such that $G'(c(h)) = 0$ implying that $K = f^{(n+1)}(c(h))$.

Now using the fact that $G(x) = 0$ and $K = f^{(n+1)}(c(h))$, we get the Taylor series result

$$f(x+h) = f(x) + f'(x)h + \frac{f''(x)}{2!}h^2 \cdots + \frac{f^{(n)}(x)}{n!}h^n + \frac{f^{(n+1)}(c(h))}{(n+1)!}h^{n+1}$$

$\square$

Exercise 11.6.1. 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 11.6.2. 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 11.6.3. 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 11.6.1. Confirm these are solutions by direct differentiation and substitution into the differential equations.

11.6.3 *Derivative Definition Approach*

Define

$$T_f^{a,k}(x) \equiv \sum_{i=0}^{k} f^i(a)\frac{(x-a)^k}{k!}$$

where $f^j = \{\text{the jth derivative of } f\}$ and $f^0 \equiv f$.

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

Theorem 11.6.1. *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* $|x-a| \leqslant \delta$.

Proof of Theorem 11.6.1(Challenging).
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 $k \geqslant 2$, $f^k(a)$ existing, implies that f^i exists in a neighborhood of $x = a$ for $i = 1, 2, ..., k-1$ and f^i is continuous for x in a neighborhood of $x = a$ for $i = 1, 2, ..., k-2$. In particular, if $k \geqslant 3$, 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 11.6.1. *if* $f(x) = o(x^k)$ *then* $\int_0^x f(y)dy = o(x^{k+1})$.

Proof of Lemma 11.6.1.
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 $|h^+(x)| \geqslant |h(x)|$ for all x. Notice that $|\int_0^x h(t)t^k dt| \leqslant h^+(x)\int_0^x t^k dt = \frac{h^+(x)}{k}|x|^{k+1}$. (End: Proof of Lemma) $\square$

4 Using the previous items, if $k \geqslant 3$, 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 $f'(x) - f'(a) - f''(a)(x-a) = h(x-a)$, where $|h(x-a)| \sim o(|x-a|)$.
   b We also know that in that neighborhood, $f'(x)$ is bounded, so by the mean value theorem, f is locally Lipschitz. This implies that $f(b) - f(a) = \int_a^b f'(x)\,dx$ for a and b in that neighborhood.
   c Defining $g(x) = f(x) - f(a) - f'(a)(x-a) - \frac{f''(a)}{2}(x-a)^2$, we note that $g'(x) = f'(x) - f'(a) - f''(a)(x-a) = h(x-a)$ and

$$|g(x)| = \left|\int_a^x g'(y)\,dy\right| \leqslant \int_a^x |h(y)|\,dy = o(|x-a|^2) \text{ (by Lemma 11.6.1)}$$

$\square$

Exercise 11.6.4. Give an example of a function that is differentiable at $x = 0$, but is not differentiable anywhere else.

Exercise 11.6.5. **(Challenge)** Find an example of a function $f : [0,1] \to \mathbb{R}_1$ that is both differentiable everywhere and Lipschitz, such that the 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!). **Hint**: Start with the function $f(x) = 0$ and then mess with it on $E \subset [0,1]$, dense in $[0,1]$, with length $\epsilon << 1$.

11.7 Degree Theory and Sard's Theorem

Recall the simple 1-dimensional example of degree theory in section 11.5:

$$\begin{aligned} \int_0^1 \sum_{x \in X_y} \text{sign}(\frac{df}{dx}(x))\, dy &= \text{oriented length of } f([0,1]) \textbf{ with cancellation} \\ &\rightarrow \text{special case of degree Theory} \\ &\rightarrow \text{will bring up Sard's Theorem for us.} \end{aligned}$$

In this section, we prove that the complement of the set of regular values has measure 0. Using the ideas we developed in section 11.4 allows us to conclude that for almost all $y \in [0,1]$, the level sets $X_y =$ a finite set of points. Because integrals ignore sets of measure zero, we know that this means the above integral is well-defined.

Exercise 11.7.1. As a follow on to the exercises in section 11.4, show that at regular values y, the sum

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

is either -1, 0 or 1.

Exercise 11.7.2. Use the results of the last exercise to conclude that

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

Theorem 11.7.1 (Sard's Theorem in $\mathbb{R}^1$). *Suppose that* $f : [0,1] \to [0,1]$ *and* $f'(x)$ *exists for all* $x \in [0,1]$. *Define*

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

Then,

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

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

There are two ways we are going to prove this.

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

First proof of Theorem 11.7.1:

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

1 Use the cone result (see Exercise 11.3.4) to choose a small enough $\delta_x^\epsilon > 0$ for every $x \in D_0$, such that

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

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

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

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

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

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

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

and this implies that
$$\sum_{i=1}^{\infty} \mathcal{H}^1(\hat{U}_{x_i}) \leqslant 5.$$

6 Now we compute:

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

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

□

Now for the $5R$ theorem.

Theorem 11.7.2 (5R Covering Theorem)**.** *If E is a ball (open or closed) with center p and radius r, let $\hat{E}$ denote the ball (open or closed) with center p and radius $5r$.*

Suppose $\mathcal{U} = \{U_\beta\}_{\beta \in \mathcal{B}}$ is a (possibly uncountable) collection of balls in $\mathbb{R}^n$ whose radii are bounded above by $C < \infty$. Then there exists a countable subcollection
$$\mathcal{F} = \{U_{\beta_i}\}_{i=1}^{N_{\mathcal{B}} \leqslant \infty}$$
such that:

1 $U_{\beta_i} \cap U_{\beta_j} = \emptyset$ *for* $i \neq j$ *and*
2 $\{U_\beta\}_{\beta \in \mathcal{B}} \subset \cup_{i=1}^{N_{\mathcal{B}}} \hat{U}_{\beta_i}$.

Proof.
We break the proof into steps:

1 We partition the balls into subcollections: $\{E_k\}_{k=1}^{\infty}$, where

$$E_k = \left\{ U_\beta \,\middle|\, \frac{1}{2^k} C < \text{radius}(U_\beta) \leqslant \frac{1}{2^{k-1}} C \right\}.$$

2 Now choose a maximal sets of disjoint balls in E_1: We use Zorn's lemma to get a maximal collection of pairwise disjoint balls in E_1: Zorn's lemma implies that there exists, F_1 a subcollection of balls in E_1 such that (a) every pair of balls $\{U, W\} \in F_1$ are disjoint and (b) if $U \in E_1 \setminus F_1$, then $U \cap W \neq \emptyset$ for some $W \in F_1$.
3 Because $\mathbb{R}^n$ is separable (i.e. there is a dense countable subset of $\mathbb{R}^n$), it follows that F_1 is a countable set and can be enumerated.
4 It also follows that

$$\bigcup_{U \in E_1} U \subset \bigcup_{U \in F_1} \hat{U}.$$

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

$$\bigcup_{U \in \cup_{k=1}^{i} E_k} U \subset \bigcup_{U \in \cup_{k=1}^{i} F_k} \hat{U}.$$

6 Define $\mathcal{F} \equiv \cup_{i=1}^{\infty} F_i$.
7 By the above construction $\mathcal{F}$ is a pairwise disjoint, countable subcollection of $\mathcal{U}$ whose dilation by 5 creates of collection of balls whose union covers the union of the balls in $\mathcal{U}$.

□

Exercise 11.7.3. Look up Zorn's lemma and make sure you understand how that lemma gives us the maximal subcollections we use.

11.7.2 A special case of Sard's theorem via more smoothness and compactness

Now we prove Theorem 11.7.1 with the added assumption that $f \in C^1$ – not only is f differentiable, the derivative f' is continuous as well.

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

2 Now use the cone result (see Exercise 11.3.4) to choose a small enough $\delta_x^\epsilon > 0$ for every $x \in D_0$, such that

$$|f(x) - f(y)| \leqslant \epsilon |x - y| \text{ when } y \in U_x \equiv (x - \delta_x^\epsilon, x + \delta_x^\epsilon).$$

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

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

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

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

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

8 Because $l_{i+2} < r_i$ would imply that the $U_{x_{i+1}}$ is covered by $U_{x_i} \cup U_{x_{i+2}}$, we can conclude that every point in $\cup_{i=1}^N U_{x_i}$ is in at most two of the U_{x_i}'s, implying that:

$$\sum_{i=1}^N \mathcal{H}^1(U_{x_i}) \leqslant 2\mathcal{H}^1(\cup_{i=1}^N U_{x_i}) \leqslant 2$$

since $\bigcup_{i=1}^N U_{x_i} \subset [0,1]$.

9 Now, as before (except with a 2 instead of a 5), we have

$$\begin{aligned}
\mathcal{H}^1(f(D_0)) &\leqslant \sum_{i=1}^N \mathcal{H}^1(f(U_{x_i})) \\
&\leqslant \sum_{i=1}^N \epsilon \mathcal{H}^1(U_{x_i}) \\
&\leqslant 2\epsilon.
\end{aligned}$$

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

Exercise 11.7.4. Convince yourself that the steps (4-8) above are justified. You should sketch the situation. See Figure (60).

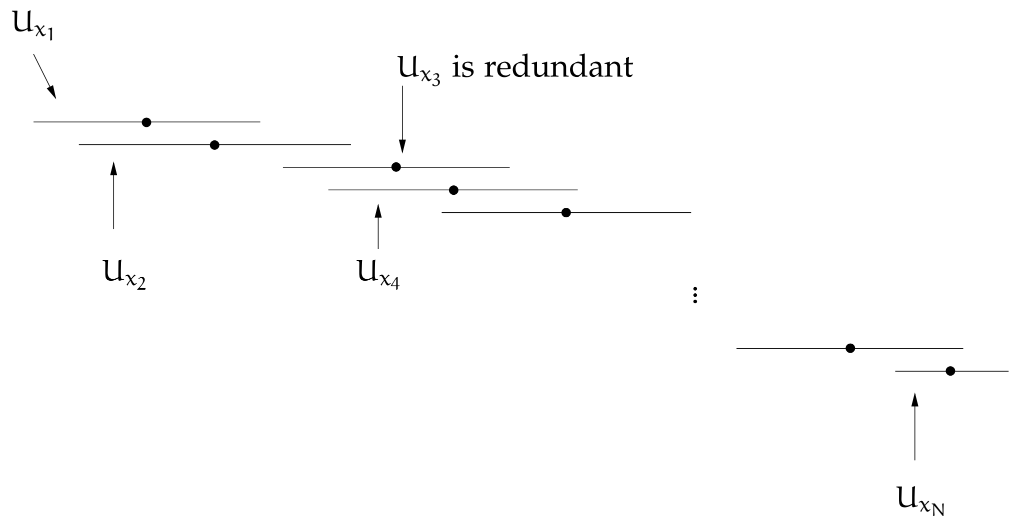


Figure 60: Example sketch to get you thinking. Remember that the intervals are symmetric about the x_i's shown as dots here.

11.7.3 Another exercise

Exercise 11.7.5. Show that the conclusion of Exercises (11.7.1-11.7.2) need not be correct if f is discontinuous, even if f is differentiable at every point except the points where it is discontinuous and there are only a finite number of discontinuities. Show this by showing, for any $\alpha \in \mathbb{R}$, how to construct a function $f_\alpha : [0, 1] \to [0, 1]$ for which

$$\alpha = \int_0^1 \sum_{x \in X_y} \operatorname{sign}\left(\frac{df_\alpha}{dx}(x)\right) dy.$$

11.8 Norms of Operators

Definition 11.8.1 (Operator Norm). *Suppose that* $A : x \in B_1 \to y \in B_2$ *where* B_1 *and* B_2 *are linear spaces with norms* $|\cdot|_1$ *and* $|\cdot|_2$*, and* A *is a linear operator. We define the norm of the operator A to be:*

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

or equivalently

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

or equivalently

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

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

11.8.1 Applicaton: Exponentials of Operators

Exercise 11.8.1. Prove that for $x \in \mathbb{R}$ and $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}.$$

Recall the definition of the norm of an operator given in Definition (11.8.1).

Exercise 11.8.2. Suppose that $A : \mathbb{R}^n \to \mathbb{R}^n$ is linear and $|A| < 1$. Prove that there is an operator B such that

$$B = I + A + A^2 + A^3 + A^4 + \dots$$

and that

$$B = (I - A)^{-1}.$$

I.e. B is the inverse of the operator $I - A$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity map.

Hint: show that for any $x \in \mathbb{R}^n$, the series

$$S_k(x) \equiv \left(I + A + A^2 + ... + A^k\right)(x) = x + Ax + A^2x + ... + A^kx$$

converges to a point in $\mathbb{R}^n$. Now define

$$B(x) = \lim_{k \to \infty} S_k(x).$$

Now compute $(I - A) * S_k(x)$ and see what happens when $k \to \infty$.

Exercise 11.8.3. Now show that

1 for any linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$,

$$e^{tA}(x) \equiv \left(I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \frac{t^4}{4!}A^4 + ...+\right)(x)$$

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

2 Defining:

$$S_A^k(t, x) \equiv \left(I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + ... + \frac{t^k}{k!}A^k\right)(x)$$

use the fact that you know how to compute $\frac{dS_A^k(t,x)}{dt}$ to show that it makes sense to say the solution to:

$$\dot{x}(t) = Ax(t)$$

is $x(t) = e^{tA}x(0)$. One main point to notice is that we do not need a bound on the norm of the operator A.

3 **(Challenge)** There is a detail here that is non-trivial: how can we show that

$$\frac{d}{dt}\left(\lim_{k \to \infty} S_A^k(t, x)\right) = \lim_{k \to \infty}\left(\frac{d}{dt}S_A^k(t, x)\right)?$$

It turns out that this is true in this case and you can go ahead and assume it, but see if you can make progress in figuring out what must be true to get this switch to work.

Hint: when considering whether or not $\frac{d}{dt}\left(\lim_{k\to\infty} f_k(t,x)\right) = \lim_{k\to\infty}\left(\frac{d}{dt} f_k(t,x)\right)$ you care about how the rates of convergence of $\lim_{k\to\infty} f_k(x,t)$ depend on t. You can also stare at

$$\lim_{h\to 0}\frac{1}{h}\left(\lim_{k\to\infty} f_k(x,t+h) - \lim_{k\to\infty} f_k(x,t)\right).$$

11.9 Using Derivative Approximations

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

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

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

Remark 11.9.1. *We need the requirement that the derivative is continuous since it is not too hard to come up with examples of functions that are differentiable at a point, but not invertible in any neighborhood of that point. See Figure 62.*

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

The simplest example of the implicit function theorem is provided by a function from $\mathbb{R}^2$ to $\mathbb{R}$. The assumptions are that the derivative of f is full rank, which in this case, means that at least one of the partial derivatives is non-zero. See Figure 63.

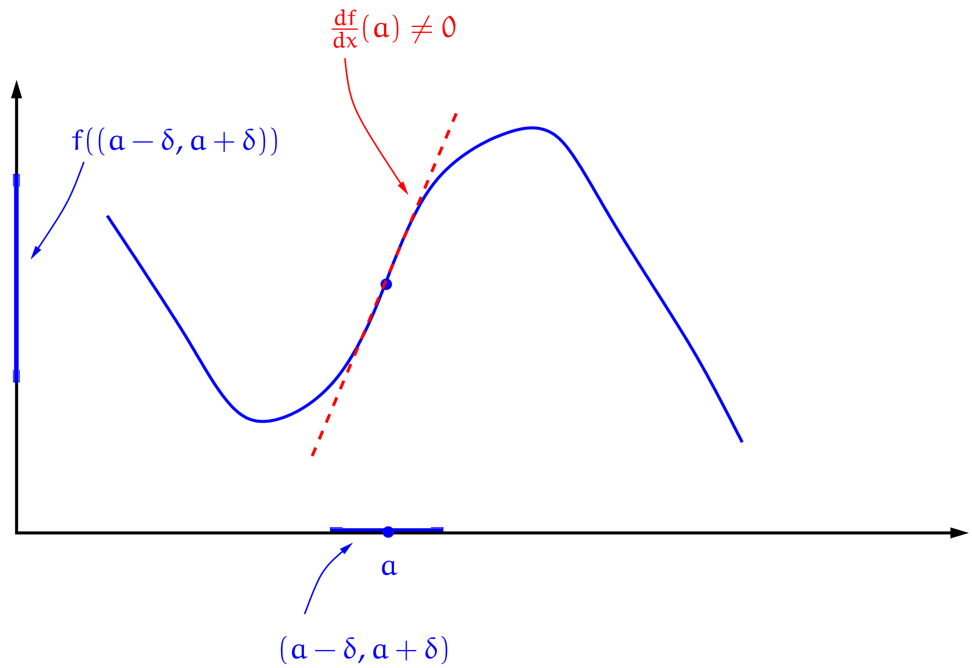


Figure 61: Inverse Function Theorem - A simple, one dimensional example: if the derivative of f is invertible at a, then, in a small enough neighborhood of a, $(a-\delta, a+\delta)$, the function itself is invertible. **Technical Details**: We need to assume that not only is the derivative at a invertible – in this case that means the slope = {1-by-1 matrix} is nonzero – we also need the derivative function mapping points in the domain to their slopes to be continuous at a.

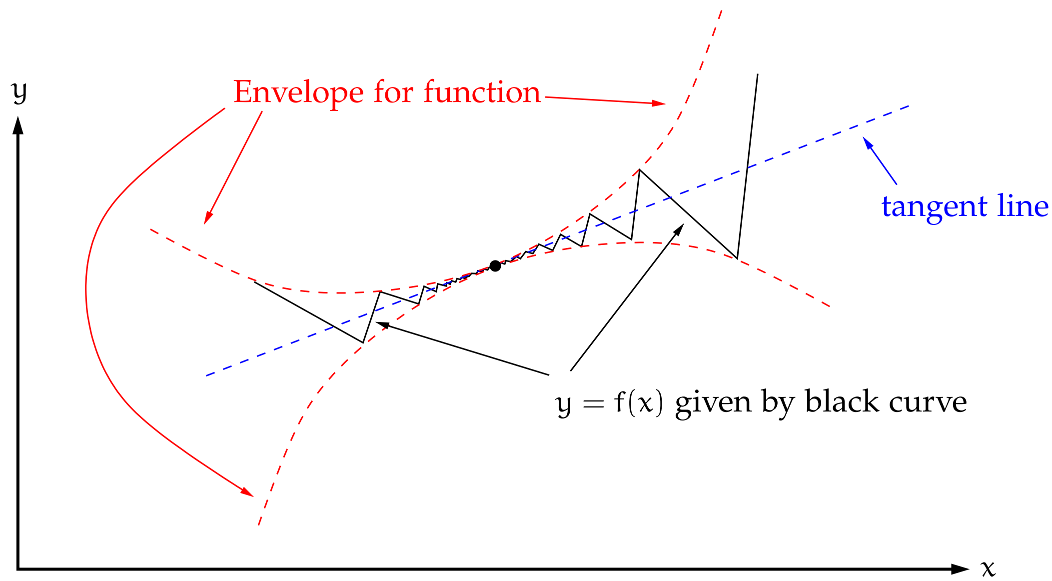


Figure 62: An example of a function whose derivative at a point is invertible but the function is not invertible in any neighborhood of that point, because the derivative is not continuous.

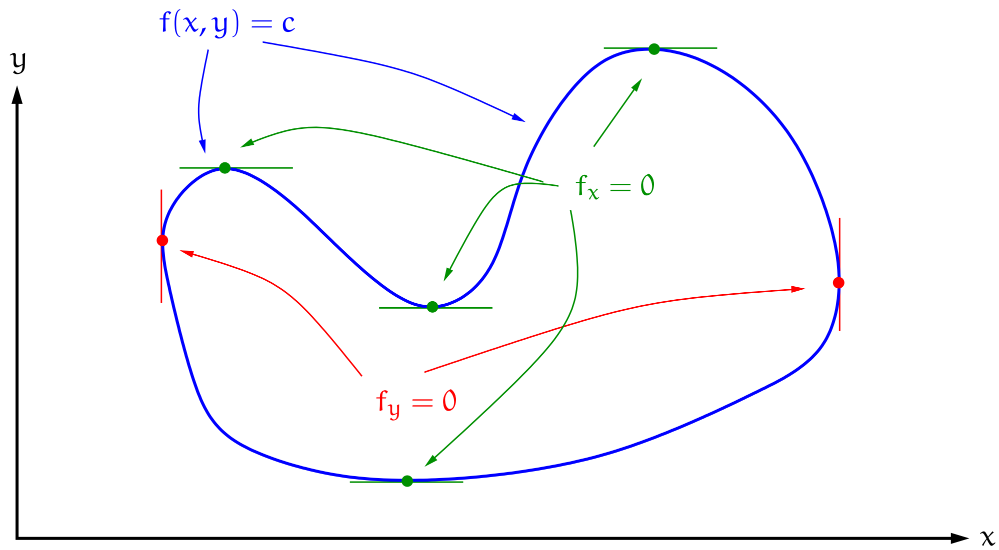


Figure 63: Implicit Function Theorem - simple example: if the derivative of f is full rank at some point $a = (x^*, y^*)$ in the $f = c$ level set, then, in a small enough neighborhood of a, then at least one of these (non-exclusive) cases holds: **(Case 1:)** There is a function of y, $g(y)$, and a $\delta > 0$ such that for $y \in (y^* - \delta, y^* + \delta)$, $f(g(y), y) = c$. (This is true if $f_x(a) \neq 0$.) **(Case 2:)** There is a function of x, $h(x)$, and a $\delta > 0$ such that for $x \in (y^* - \delta, y^* + \delta)$ $f(x, h(x)) = c$. (This is true if $f_y(a) \neq 0$.) **Technical Details**: While the theorem only needs the derivative to be full rank at a, if the derivative of f is full rank on the entire level set, this means that we have local coordinates everywhere, though sometimes only x or only y will work as local coordinates. The derivative in our case is $\nabla f = (f_x, f_y)$ and being full rank means there is at least one nonzero element of this gradient vector. We are also assuming that the derivative is continuous, as we did in the inverse function theorem case, because, in fact we use the inverse function theorem to prove this theorem.

Remark 11.9.2. *Suppose for example, that* $f_x(a) \neq 0$*. Then locally we can change the value of the function by changing the value of x: if* $f(x^*, y^*) = c$ *and we perturb* y*, from* y^* *to* $y^* + \epsilon$*, we will generally find that* $f(x^*, y^* + \epsilon) = c + \delta$ *but because* $f_x \neq 0$*, we can just find an* $\eta(\epsilon)$ *such that* $f(x^* + \eta(\epsilon), y^* + \epsilon) = c$*.* $\eta(\epsilon)$ *will be approximately given by* $f_x(a)\eta(\epsilon) \approx -\delta$ *or* $\eta(\epsilon) \approx \frac{-\delta}{f_x(a)}$.

11.10 Inverse and Implicit Function Theorems

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

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

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

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

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

EUCLIDEAN DISTANCE: The length of a vector $\mathbf{x}$ is given by

$$|\mathbf{x}| \equiv \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x \cdot x},$$

so the distance between two points is simply $|\mathbf{x} - \mathbf{y}|$.

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 $\mathbb{R}^n$ mimics and extends, 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.

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

Definition 11.10.1 (Diffeomorphism). *A mapping* $f : X \to Y$ *is a diffeomorphism if*

(a) f *is a continuous bijection with a continous inverse mapping* $g = f^{-1}$
(b) Both f *anf* g *are continuously differentiable.*

Often we want to specify the differentiability and we say f *is a* C^k *diffeomorphism if* $f, g \in C^k$ *(all* k*-th order derivatives are continuous).*

Definition 11.10.2 (k-manifold in $\mathbb{R}^n$). *Define* L_k *to be the* k*-dimensional subspace of* $\mathbb{R}^n$ *defined by holding the last* $n-k$ *coordinates equal to* 0*, i.e.*

all points in $\mathbb{R}^n$ *of the form* $(x_1, x_2, ..., x_k, 0, ..., 0)$. *A* k*-dimensional manifold* M_k *is a subset that is locally like* $\mathbb{R}^k$. *At every point* $x \in M_k$, *there is*

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

such that

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

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

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

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

Definition 11.10.3 (Tangent Space at x). *If M_k is a k-manifold, then T_k is the unique k-dimensional subspace of $\mathbb{R}^n$ such that for every $\epsilon > 0$ there is an r_ϵ such that for every point $y \in M_k \cap B(x, r_\epsilon)$*

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

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

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

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

11.10.3 Review: 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 recall the definition of $o(h)$

Definition 11.10.4 (Review: little o of h, $o(h)$). *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 11.10.5 (Review: 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 $D_x F$.

In other words, $D_x F$ 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:

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

Example 11.10.1 ($F : \mathbb{R}^n \to \mathbb{R}$)**.** *In the case of a function mapping $\mathbb{R}^n$ to the real numbers, we get $D_x F = \nabla F|_x$: the derivative of F at x is the gradient of F at x, a row vector made up of the partial derivatives of F.*

Remark 11.10.1. *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 $T_x + x = D_0(\phi_x^{-1})(L_k)$, where L_k is defined as in Definition 11.10.2.*

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

Definition 11.10.6. *If the derivative of F exists and is continuous, then we will say F **is** C^1. When that derivative has a derivative that is continuous, it is C^2. Likewise **when** F **is** k**-times continuously differentiable, it is** C^k.*

11.10.4 Review: Full rank maps

Definition 11.10.7 (Full Rank Matrix). *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 11.10.2. *If a matrix* A *is full rank, then a sufficiently small perturbation will not change that fact. I.e. if* $A = \{a_{i,j}\}_{i=1,j=1}^{i=m,j=n}$ *is full rank* $m \times n$ *matrix, there is an* $\epsilon(A) > 0$ *such that if* E *is an* $m \times n$ *matrix such that* $\max_{i,j} e_{i,j} < \epsilon(A)$*, then* $A + E$ *is also full rank.*

Definition 11.10.8 (Level Sets). *The* **level sets** *of a mapping* $F : R^n \to R^m$ *are the collection of sets* $F^{-1}(y) \subset \mathbb{R}^n$ *for all* $y \in \mathbb{R}^m$*.*

Definition 11.10.9 (Full Rank Mapping). *A mapping* $F : R^n \to R^m$ *is full rank on a level set* $F^{-1}(y)$*, if* $D_x F$ *is full rank for all* $x \in F^{-1}(y)$*.*

Define $W_y = F^{-1}(y)$. When $D_x F$ 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 11.10.10. *When the coordinate diffeomorphisms in the definition of a* k*-manifold are of* C^p*, then we say that the manifold is of class* C^p*.*

Theorem 11.10.1 (Full Rank Theorem). *Suppose that* F *is* C^p *with* $p \geqslant 1$*. When* $D_x F$ *is full rank on* $W_y = F^{-1}(y)$*,* W_y *is a* C^p*,* k*-manifold in* R^n*, with* $k = \max(0, n - m)$*.*

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

11.10.5 Finally: Inverse and Implicit function theorem in higher dimensions

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

Theorem 11.10.2 (Inverse Function Theorem). *Suppose that* $F : \mathbb{R}^n \to \mathbb{R}^n$*,* $x \in \mathbb{R}^n$*,* F *is* C^k*,* $k \geqslant 1$ *and* $D_x F$ *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 $F(0) = 0$ (since we can shift the coordinate system to make this true). Now we define $\hat{F} \equiv D_0 F^{-1} \circ F$ so that $\hat{F}(0) = 0$ and $D_0 \hat{F} = I$. Here and below, I denotes the identity on $\mathbb{R}^n$ and I_E denotes the Identity on $\mathbb{R}^n$ restricted to the set E.

1 Choose $0 < \epsilon < 1/4$.
2 Define $G = I - \hat{F}$.
3 Using the fact that $\hat{F}$ is C^k for $k \geqslant 1$, we have that
4 the operator 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. $\frac{||D_x G(h)||}{||h||} < \epsilon$ for $x \in U$ and all $h \neq 0$.
5 Using the mean value theorem in vector spaces, we get that restricted to U, G is a contraction mapping with contraction constant ϵ – i.e. $|G(x) - G(y)| \leqslant \epsilon |x - y|$ if $x, y \in U$.
6 Define $H = (I + G + G^2 + G^3 + ...)$ and note tht $\hat{F} = I - G$.
7 We note that by contruction, for $x, y \in U$,

$$\frac{3}{4}|x - y| \leqslant |\hat{F}(x) - \hat{F}(y)| \leqslant \frac{5}{4}|x - y|$$

$$\frac{1}{2}|x - y| \leqslant |H(x) - H(y)| \leqslant \frac{3}{2}|x - y|$$

and we conclude that both $\hat{F}$ and H are continuous on U, with continuous inverses on $\hat{F}(U)$ and $H(U)$ respectively.

8 Because $\sum_{i=0}^{k} G^i$ converges uniformly to H and $\sum_{i=0}^{k} D_x G^i$ converge uniformly, we know that $DH = I + DG + DG \circ DG + \ldots = \sum_{i=0}^{\infty} D_x G^i$ and that this derivative is continuous. (See pages 189-190 of [25])

9 Define $W = B(0, \frac{\delta(\epsilon)}{2})$.

10 From Step 7 above,

$$\hat{F}(W) \subset \frac{5}{4} W \subset U$$

and

$$H(\hat{F}(W)) \subset 2W = U$$

but we know more because the uniform convergence of the geometric series defining H implies that for all $\in W$

$$(I - G)(I + G + G^2 + \ldots)(x) = x$$

and

$$(I + G + G^2 + \ldots)(I - G)(x) = x$$

so that we have

$$H \circ \hat{F} = I_W$$

and

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

11 Using the fact that $D_y H = [D_{H(y)} \hat{F}]^{-1}$, we deduce that if $\hat{F}$ is C^k differentiable, then so is H. We do this as follows.

a We already have that H is in C^1.

b Suppose that H is in C^r with $r \leqslant k - 1$. We have that $D_y H$ is the composition

$$\{\text{inversion which is } C^\infty\} \circ \{D_* \hat{F} \text{ which is } C^{k-1}\} \circ \{H \text{ which is } C^r\}$$

and so $D_y H \in C^r \to H \in C^{r+1}$.

c We continue this to get $H \in C^k$.

12 So $\hat{F}^{-1}$ exists and is C^k differentiable on $\hat{F}(W)$.

□

Theorem 11.10.3 (Implicit Function Theorem). *Suppose* F *is* C^k, $F : \mathbb{R}^n \to \mathbb{R}^m$, $m < n$, *and* DF *is full rank at* $x^* \in \mathbb{R}^n$. *Denote the first* m *coordinates by* x' *and the last* $n - m$ *by* x'' *so that* $x = (x', x'')$. *Without loss of generality, assume 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 $D_{x^*}\hat{F}$ is invertible so that there is an inverse of $\hat{F}$, $G(y) = (g(y', y''), y'')$. Computing $\hat{F} \circ G(y)$ $(= y)$ we have $\hat{F}(G(y)) = (F(g(y', y''), y''), y'') = (y', y'')$ for all $y = (y', y'')$ in some neighborhood of $(F(x^*), x^{*\prime\prime})$. 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 $\epsilon > 0$. $\square$

Example 11.10.2. *Consider some function* f *mapping* $\mathbb{R}^n$ *to* $\mathbb{R}$. *Then in order to apply the implicit function theorem at some point* x^*, *we need* $Df = \nabla f$ *to be full rank at* x^*. *Since* $\min(m, n) = 1$, *at least one component of the gradient needs to be non-zero at* x^* *in order to conclude that locally, the level set through* x^* *is an* $(n-1)$*-manifold.*

11.10.6 Implicit Function Theorem, intuitively, again

The idea behind the implicit function theorem is that:

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

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

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

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

WHAT IT BOILS DOWN TO: if we know that $f(x^*) = f(x^{*'}, x^{*''}) = c$ and we now that $D_{x'} f(x^*)$ is invertible (because Df is full rank at x^*), then we know that

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

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

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

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

11.10.7 Implicit Function Theorem: Co-Dimension 1

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

1. So you are given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a level α and a point $\hat{x} \in f^{-1}(\alpha)$.
2. You need to know that $D_{\hat{x}} f = \nabla_{\hat{x}} f$ is full rank, which in the case of 1 by n matrices (i.e. row vectors, which is what a gradient is) we simply want to know that one of the partial derivatives of f at $\hat{x}$ is **not** equal to zero.
3. (Actually, if α is a regular value of f, then we know this since this means, by definition, that the derivative of f is full rank at **every** point in $f^{-1}(\alpha)$. But we actually only need f differentiable at $\hat{x}$.)
4. We also need that $f \in C^1$ – the derivative exists and is continuous.
5. Now we note that the gradient is normal to the α-level set $f^{-1}(\alpha)$ which is the same statement as "the planes defined by the gradient vector are tangent to the level surface". This tangent plane is determined by the gradient vector as follows.
6. Define
$$N = \frac{\nabla_{x_0} f}{|\nabla_{x_0} f|}.$$
7. As long as the gradient is not horizontal – by which we mean that it has an nth coordinate $= 0$ – then we can write the nth coordinate as a function the first n-1 coordinates:

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

See the example in 3 dimensions in Figure 64.

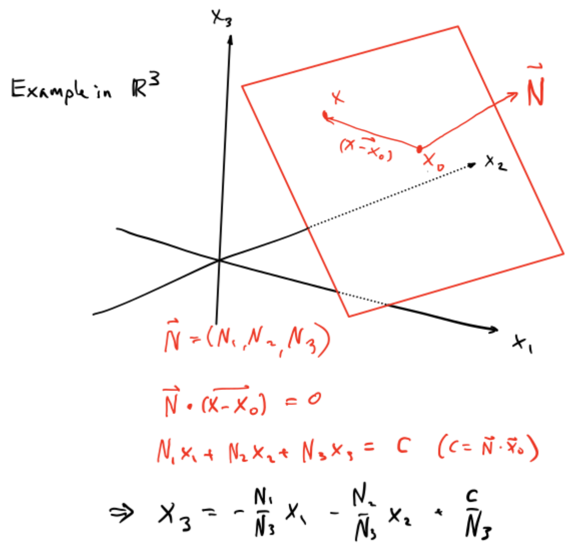


Figure 64: Non-horizontal gradients.

8 Now, rotate the coordinates so that the gradient vector of f points in the direction of the nth coordinate axis. I.e. rotate $\mathbb{R}^n$ so that $\nabla_{\hat{x}} f = \beta(0, 0, ..., 0, 1)$ where $\beta = |\nabla_{\hat{x}} f|$. See Figure 65

9 Now we apply these insights to rotated level set and the now vertical gradient vector, which is therefore normal to a horizontal tangent plane at the rotated point that we again denote by $\hat{x}$.

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

11 Note that because $f \in C^1$ we know that for some little ball in the space of the first n-1 coordinates, $B(\hat{x}', \epsilon)$ centered at $\hat{x}'$ in the horizontal $\mathbb{R}^{n-1}$, the gradient is not too far from vertical. See Figure 66

12 Here is a short argument: we know that the surface tangent planes $H_{x_0}(x)$, thought of as functions from $\mathbb{R}^{n-1} \to \mathbb{R}$, have small gradients everywhere because from the example above, for $x_0' \in B(x', \epsilon)$

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

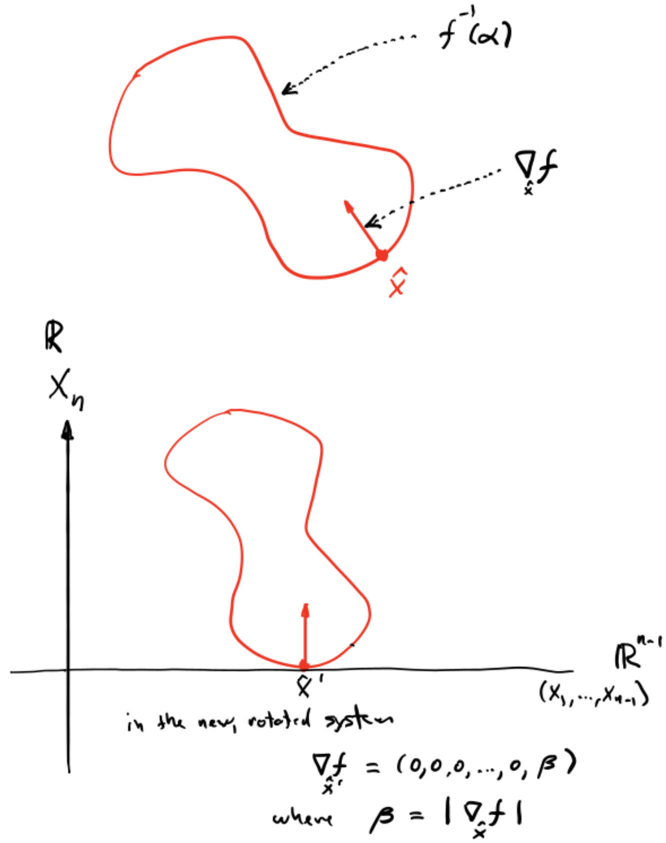


Figure 65: Rotating the coordinate system.

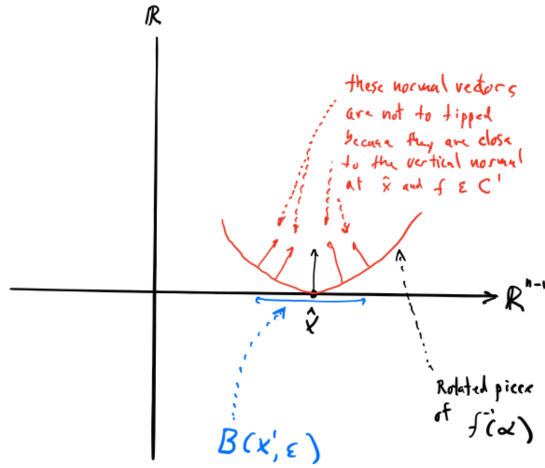


Figure 66: Gradient is continuous and so remains nearly vertical nearby.

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

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

and

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

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

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

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

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

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

15 Thus we can shoot vertically at $x' + h$ to find the point where $f = \alpha$. See Figure 67.

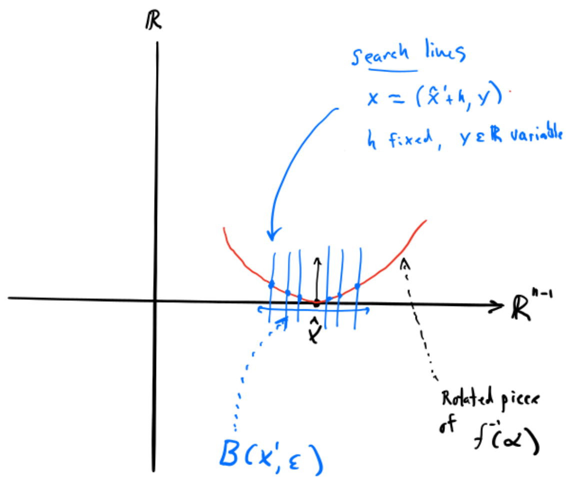


Figure 67: We shoot vertically to find the surface.

16 We can actually **use this argument to prove that a g exists that is Lipschitz and satisfies**

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

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

Exercise 11.10.1. Convince yourself that there is a Lipschitz function g, using only the fact that the normals are close to vertical near $\hat{x}$, such that

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

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

Exercise 11.10.2. Define $z = f(x, y) = x^2 - y^2$. For what values of c is the c-level set $L_c = \{(x, y) \mid f(x, y) = c\}$ not regular? Find the points (x^*, y^*) in each regular level set L_c such that **either** $f(x, h(x)) = c$ **or** $f(g(y), y) = c$ does not hold near (x^*, y^*). See Section 11.9.2.

Measures and Integrals: Mechanics and Nuance 12

12.1 Integration

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

12.2 Riemann vs Lebesgue

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

Define the function $f_Q(x)$ by:

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

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

$$0 = p_0 < p_1 < p_2 < p_3 < \ldots < p_m = 1,$$

the Riemann upper and lower integrals are defined to be:

$$\int^* f(x)\,dx \equiv \inf_{P \in \mathcal{P}} \left(\sum_{i=0}^{m-1} (p_{i+1} - p_i) \sup_{y \in [p_i, p_{i+1})} f(y) \right) \tag{27}$$

$$\int_* f(x)\,dx \equiv \sup_{P \in \mathcal{P}} \left(\sum_{i=0}^{m-1} (p_{i+1} - p_i) \inf_{y \in [p_i, p_{i+1})} f(y) \right) \tag{28}$$

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

Exercise 12.2.1. Show that for any two partitions of $[0,1]$, $P = \{p_i\}_{i=0}^m$ and $Q = \{q_i\}_{i=0}^k$,

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

Exercise 12.2.2. Show that $1 = \int^* f_Q(x) > \int_* f_Q(x) = 0$.

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

For now we will assume an intuitive grasp of the idea of $\mu(E)$, the measure of a set E (i.e. appropriately dimensioned volume of E). It really is just what you think it should be – the 1-volume (length), 2-volume (area), ..., k-volume of a k-dimensional set – a fact that will be made clear in the next section when we define outer measures and a couple of specific families of measures on R^n.

Remark 12.2.1. *Though this is where we start – measures are essentially volumes of various dimensions – and this notion takes us a very long way, there are large/wild (and yet useful!) generalizations of the notion of measure. For example, in the theory of infinite dimensional normal operators, measures are constructed that assign a measure to a subset of the complex numbers and that measure of a set is a projection operator! What we need in this course, will be satisfied by generalized volumes and weighted volumes.*

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

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

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

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

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

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

and

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

Definition 12.2.1 (Lebesgue Integrable Functions). *Define* $f^+(x) \equiv \max\{0, f(x)\}$ *and* $f^-(x) \equiv \max\{0, -f(x)\}$*. Then*

1 *We say that* $f \geqslant 0$ *is Lebesgue integrable if* $\int^* f d\mu = \int_* f d\mu$*. See Figure (68).*

2 *We say that* f *is Lebesgue integrable if both* $f^+(x)$ *and* $f^-(x)$ *are and at least one of these two integrals is finite. In this case, we define* $\int f\ d\mu = \int f^+\ d\mu - \int f^-\ d\mu$*.*

Example 12.2.1 (Lebesgue Integrability)**.** *Returning to the integration of any function* $f : [0, 1] \to [0, 1]$*, pick a positive integer* $M < \infty$ *and notice that if we define* $E_i = f^{-1}([\frac{i-1}{M}, \frac{i}{M}))$ *for* $i = 1, ..., M$ *and then define*

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

and

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

we have that

1. $s^u(x) \geqslant f(x)$ *for all* $x \in [0,1]$ *and*
2. $s^l(x) \leqslant f(x)$ *for all* $x \in [0,1]$,
3. *which allows us to conclude that*

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

implying that f *is integrable.*

Exercise 12.2.3. Convince yourself (i.e. prove) that if $s(x)$ and $r(x)$ are two simple functions such that $s(x) \leqslant f(x) \leqslant r(x)$ for all x, then $\int_A s\ d\mu \leqslant \int_A r\ d\mu$. Note that we are defining the integrals here, so you have to be careful to not assume what you are trying to prove.

Exercise 12.2.4. Show that $\int f_Q(x)\ d\mu$ exists and equals 0.

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

Exercise 12.2.5. Look up the Banach-Tarski Paradox on Wikipedia and read about it.

So, when we have a measure μ on some space X, it is always the case that we will also have some collection of sets $\mathcal{X}_\mu$ that are said to be μ-measurable. We get into this in more detail in Section 12.3.

Remark 12.2.2 (Summable versus Integrable)**.** *We will say a function* f *is summable if it is integrable* ***and*** *the integral of the function is finite.*

We will look more carefully and completely at the Lebesgue integral in Section 12.4 and to prepare for that we need to loom more closely at measures. This is what we do next.

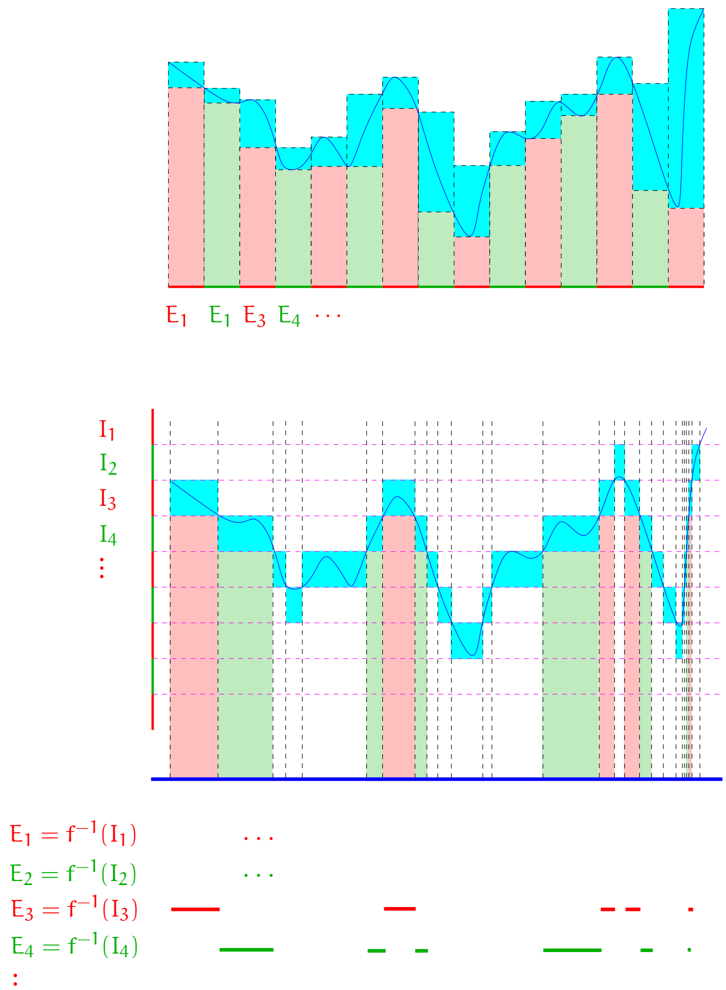


Figure 68: Riemann versus Lebesgue Integration: the upper figure illustrates the partition of the domain dictated by the Riemannian approach. The green and red rectangles live completely below the graph of f. Call the area they sum to $A_{\text{lower}}(P)$ where P is the partition. The red and green plus the cyan rectangles live completely above the graph. Call their area A_{upper}. If $\sup_P A_{\text{lower}}(P) = \inf_P A_{\text{upper}}$ then f is Riemann integrable. The lower figure illustrates that key difference for the Lebesgue case: we partition the range and pull that back by f^{-1} to a partition of the domain. It turns out that this is exactly what is needed to make all reasonable functions integrable. Now $A_{\text{lower}}(P) = \sum_i a_i \mu(E_i)$ and $A_{\text{upper}}(P) = \sum_i b_i \mu(E_i)$ where P is a partition of the range into the intervals $I_i = [a_i, b_i)$.

12.3 Outer Measures

The approach to measure theory I like closely follows the approach used by Evans and Gariepy in their *Measure Theory and Fine Properties of Functions*[12] – a book I very highly recommend for anyone interested in analysis.

Definition 12.3.1 (Power Set)**.** *The collection of all subsets of a space X is denoted 2^X and is called the power set of X.*

Definition 12.3.2 (Outer Measure). *Any function, μ, mapping subsets of a space X to $[0, \infty]$ – $\mu : 2^X \to [0, \infty]$ – satisfying the following two conditions:*

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

is called an ***Outer Measure****.*

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

Definition 12.3.3 (Measurable Sets). *If a set $E \subset X$ has the property that for all $A \in 2^X$:*

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

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

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

Exercise 12.3.1. **(Easy consequences of the definition of measurability)** Show that the definition of measurability easily gives us (1) that E measurable $\Rightarrow E^c$ is measurable; and (2) X and $\emptyset$ are measurable.

Remark 12.3.1. *Note that we always have that*

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

so we only need to show

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

to prove that

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

Exercise 12.3.2. Show that if $\mu(E) = 0$, then E is measurable.

Definition 12.3.4 (σ**-algebra of sets).** *A collection of sets $\mathcal{A}$ is a σ-algebra if:*

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

Theorem 12.3.1 (Properties of Measures)**.** *Suppose that $\{E_i\}_{i=1}^{\infty}$ is a sequence of measurable sets. Then we have that:*

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

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

This property is called countable additivity.

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

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

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

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

Proof.
See the beginning of Chapter 1 of Evans and Gariepy [12] □

Exercise 12.3.3. Show that Exercise 12.3.1 and Theorem 12.3.1 can be used to prove that the collection of measurable sets of any measure is a σ-algebra

12.3.1 *Lebesgue and Hausdorff Measures*

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

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

where $\epsilon_i > 0$ for all i and its **content**, $c(R)$, to be the product of the side-lengths of the rectangle = its usual d-volume:

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

We can now define:

Definition 12.3.5 (Lebesgue Measure).

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

where we are minimizing over all $\mathcal{R}$, the countable covers of E by open rectangles: $\mathcal{R} = \{R_i\}_{i=1}^{\infty}$.

Exercise 12.3.4. Show that the Lebesgue measure of a set E does not change if we use closed rectangles to cover E instead of open rectangles. More precisely, define a closed rectangle $\hat{R}$ to be

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

where $\epsilon_i > 0$ for all i and its **content**, $c(\hat{R})$, to be the product of the side-lengths of the rectangle = its usual d-volume:

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

Now minimize the sums of contents of closed rectangle covers of E. Show that the measure of E remains the same. Thus we can used open or closed rectangles to compute the Lebesgue measure of a set.

Exercise 12.3.5. **(Lebesgue Measure of Rectangles = Their Content)**

1. Show that $\mathcal{L}^d(S) = 0$ for any $S = \{x \in \mathbb{R}^d \mid x_i = c\}$. I.e. The d-dimensional measure of any d-1-dimensional plane in $\mathbb{R}^d$ obtained by holding the i-th coordinate constant, is 0.
2. Show that if $\mu(D) = 0$, then $\mu(C \cup D) = \mu(C)$.
3. Show that the measure of any rectangle when some or all of the intervals defining it are not open is the same as the corresponding open rectangle.
4. Show that for any rectangle $R \subset \mathbb{R}^d$,

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

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

Definition 12.3.6 ($\omega(\eta)$)**.** *We define* $\omega(\eta) \equiv \frac{\pi^{\frac{\eta}{2}}}{\Gamma(\frac{\eta}{2}+1)}$ *to be the* η*-volume of the "*η*-dimensional" unit ball. This number agrees with the usual volume when* η *is an integer.*

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

Definition 12.3.7 (Hausdorff Measures, $\mathcal{H}^\eta$).

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

and then

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

Remark 12.3.2. *It turns out that when η is a non-negative integer and the space we are measuring is $\mathbb{R}^\eta$, $\mathcal{L}^\eta = \mathcal{H}^\eta$. (See Evans and Gariepy's book for a proof of this fact.)*

The first thing we will prove is that for any fixed $E \subset \mathbb{R}^n$, the graph of $\mathcal{H}^k(E)$ versus k looks like the graph in Figure (69).

Theorem 12.3.2 (Definition of Hausdorff Dimension)**.** *Suppose that $E \subset \mathbb{R}^n$ for some $n < \infty$. Then $\mathcal{H}^k(E) = 0$ for $k > d^*$ for some $d^* \leqslant n$ and, if $d^* > 0$, $\mathcal{H}^k(E) = \infty$ for $k < d^*$. This is illustrated in Figure (69).*

Remark 12.3.3. *It turns out that each of the three cases (1) $\dim(E) = d^*$ and $\mathcal{H}^{d^*}(E) = 0$, (2) $\dim(E) = d^*$ and $\mathcal{H}^{d^*}(E) =$, and (3) $\dim(E) = d^*$ and $\mathcal{H}^{d^*}(E) = \infty$ can occur.*

Proof of Theorem 12.3.2.

1 First we show if $0 \leqslant \mathcal{H}^k(E) < \infty$ then $\mathcal{H}^{k+\eta}(E) = 0$ for any $\eta > 0$.
  a Choose $\epsilon > 0$.
  b Choose $0 < \delta < 1$ such that $\left(\frac{\delta}{2}\right)^\eta < \epsilon$.
  c By the definition of Hausdorff measure, there is a $\delta_\epsilon < \delta$ such that

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

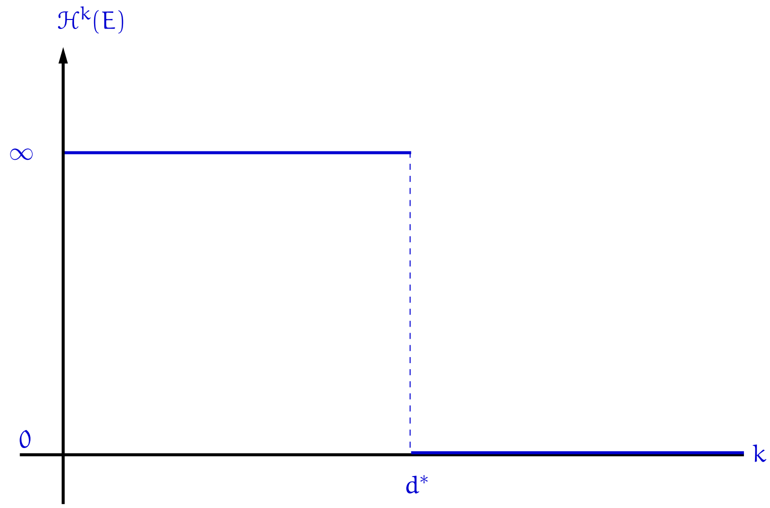


Figure 69: Graph of the Hausdorff measure $\mathcal{H}^k(E)$ of a set E as we vary k, the dimension of the measure. We define d^*, where the measure switches from ∞ to 0 to be the dimension of the set E.

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

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

e We get that

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

from which we get that

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

f Now we note that this implies that

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

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

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

i But δ can be chosen arbitrarily small, implying that

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

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

3 The Hausdorff dimension of a subset of $\mathbb{R}^n$ cannot be bigger than n ...
Exercise 12.3.6. Show that if $\epsilon > 0$ then $\mathcal{H}^{n+\epsilon}(\mathbb{R}^n) = 0$.

4 If we did not end the proof at Step 2, then it must be the case that $\mathcal{H}^{\mathbf{0}}(\mathbf{E}) = \infty$.

5 Define $d^* \equiv \sup\{x \leqslant n \mid \mathcal{H}^x(E) > 0\}$. Then by Step 1 $\mathcal{H}^x(E) = \infty$ for $0 \leqslant x < d^*$

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

□

Exercise 12.3.7. Show that set

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

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

Exercise 12.3.8. Show that you can assume the sets used to generate the covers in the Hausdorff definition are convex. **Hint:** Show that for any $E \subset \mathbb{R}^n$,

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

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

1. For the set to have finite diameter, it must be bounded.
2. $E \subset \operatorname{cnv}(E)$ implies that $\operatorname{diam}(E) \leqslant \operatorname{diam}(\operatorname{cnv}(E))$.
3. There is a sequence of pairs of points in $\operatorname{cnv}(E)$, $\{p_i, q_i\}_{i=1}^{\infty}$ such that $|p_i - q_i| \to \operatorname{diam}(\operatorname{cnv}(E))$.
4. There is a subsequence $i(k)$ such that $p_{i_k} \to p^*$ and $q_{i_k} \to q^*$ and $|p^* - q^*| = \operatorname{diam}(\operatorname{cnv}(E))$.
5. The projection of E onto the line L_{p^*,q^*} through p^* and q^* has diameter at most $\operatorname{diam}(E)$.
6. If the smallest interval in L_{p^*,q^*} containing this projection is defined to be P, the E lives between the two $n-1$-dimensional planes orthogonal to L_{p^*,q^*} through the endpoints of P. (Though not necessarily strictly between!)
7. That $\operatorname{cnv}(E)$ also must be contained between the two $n-1$-dimensional planes orthogonal to L_{p^*,q^*} through the endpoints of P (since $\operatorname{cnv}(E)$ = intersection of all convex sets containing E and the set of points between and including the two planes is convex). This implies that p^* and q^* are the endpoints of P!
8. In conclusion, we must have that

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

Exercise 12.3.9. Show that you can assume the sets used to generate the covers in the Hausdorff definition are open. **Hint:** Show that ...

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

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

are open.

2 Choose $\epsilon > 0$.

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

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

then

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

4 Use this to show

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

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

Exercise 12.3.10. Show that if F is a convex set, then $B(F, \epsilon) \equiv \cup_{x \in F} B(x, \epsilon)$ is an open convex set.

Exercise 12.3.11. Show that we may also restrict ourselves to closed sets when covering a set in order to compute Hausdorff measures. **Hint:** If we denote the closure of E by $\bar{E}$, you simply need to show that $\operatorname{diam}(E) = \operatorname{diam}(\bar{E})$.

Exercise 12.3.12. suppose that $S = [a, b] \times \{0\}$ is a closed line segment in $\mathbb{R}^2$. Show that $\mathcal{H}^1(S) = b - a$. **Hint:** Exercises 12.3.8, 12.3.9, 12.3.10 and the fact that S is compact, allow you use open, convex sets in your cover of S and then throw away all but a finite number of them in your Hausdorff measure computations.

12.3.2 Radon Measures and Approximation

Definition 12.3.8 (Borel Sets)**.** *A Borel set is any subset in the smallest σ-algebra of sets containing the open sets.*

Definition 12.3.9 (**Regular, Borel, Borel Regular, Radon**)**.** *Suppose that μ is an outer measure on $X = \mathbb{R}^n$.*

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

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

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

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

The following approximation property of Radon measures is very useful.

Theorem 12.3.3 (Approximation of Radon Measures)**.** *Suppose that μ is a Radon Measure. Then*

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

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

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

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

Exercise 12.3.13. Show that Hausdorff measures satisfy part 1 of Theorem 12.3.3 ... I.e. the measure of a set is approximated by open sets from outside.

12.3.3 *Caratheodory Criterion*

Here is a very useful criterion telling us when the Borel sets are measurable.

Theorem 12.3.4 (Caratheodory Criterion)**.** *If* μ *is an outer measure on* $\mathbb{R}^n$ *and we know that* $\operatorname{dist}(A, B) > 0 \quad \Rightarrow \quad \mu(A \cup B) = \mu(A) + \mu(B)$*, then* μ *is a Borel measure – i.e. all Borel sets are measurable.*

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

1 Let A be an arbitrary set in X and C be a closed set.
2 The result is immediate if $\mu(A) = \infty$, so assume $\mu(A) < \infty$.
3 Define

$$C_n = \{x \in \mathbb{R}^n \mid \operatorname{dist}(x, C) \leqslant \frac{1}{n}\}$$

where $\operatorname{dist}(x, C)$ is the distance from x to the set C.
4 Because $\operatorname{dist}(C, C_n^c) = \frac{1}{n} > 0$, we know that

$$\mu(A) \geqslant \mu(\{A \cap C\} \cup \{A \cap C_n^c\}) = \mu(A \cap C) + \mu(A \cap C_n^c).$$

5 Define

$$A_i = \{x \in A \mid \frac{1}{i+1} < d(x, C) \leqslant \frac{1}{i}\}.$$

Then

$$A = \{A \cap C\} \cup \{A \cap C_n^c\} \cup \{\cup_{i=n}^{\infty} A_i\}.$$

6 We want to show that

$$\mu(A) \geqslant \mu(A \cap C) + \mu(A \cap C^c).$$

7 Because of Step 4 above and

$$\{A \cap C\} \cup \{A \cap C^c\} \subset \{A \cap C\} \cup \{A \cap C_n^c\} \cup \{\cup_{i=n}^{\infty} A_i\}$$

we know that

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

8 Now all we need is that $\sum_{i=1}^{\infty} \mu(A_i) < \infty$, because this implies

$$\sum_{i=n}^{\infty} \mu(A_i) \to 0$$

as $n \to \infty$ which in turn implies

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

for all $\epsilon > 0$.

9 Now we show that $\sum_{i=1}^{\infty} \mu(A_i) < \infty$.

10 Defining $A'_n = A_1 \cup A_3 \cup A_5 \cup ... \cup A_{2n+1}$ $A''_n = A_2 \cup A_4 \cup A_6 \cup ... \cup A_{2n}$ and noting that for all n, we have

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

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

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

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

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

11 We are done!

□

Exercise 12.3.14. Show that Lebesgue measure of a set can be found be restricting yourself to covers with rectangles whose side-length is bounded by any $\delta > 0$.

Exercise 12.3.15. Show that both Lebesgue and Hausdorff measures are Borel Regular.

12.4 Measurable Functions and Integration

We now look more carefully at the ideas we saw briefly in Section 12.2.

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

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

$$\int^{*} f d\mu \equiv \inf_{P} \sum_{i} \sup_{x \in I_i} f(x)\mu(I_i)$$

$$\int_{*} f d\mu \equiv \sup_{P} \sum_{i} \inf_{x \in I_i} f(x)\mu(I_i)$$

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

As you saw briefly in Section 12.2, the Lebesgue integral partitions the *range* into intervals I_i and pulls them back to a partition of the domain: $E_i = f^{-1}(I_i)$. (This partition can be very far from regular!) We get:

$$\int^{*} f d\mu \equiv \inf_{P} \sum_{i} \left(\sup_{y \in I_i} y \right) \mu(f^{-1}(I_i)) = \inf_{P} \sum_{i} b_i \mu(f^{-1}(I_i))$$

$$\int_{*} f d\mu \equiv \sup_{P} \sum_{i} \left(\inf_{y \in I_i} y \right) \mu(f^{-1}(I_i)) = \sup_{P} \sum_{i} a_i \mu(f^{-1}(I_i)).$$

We are rewarded for our change in perspective by the result that now, *every* respectable non-negative function is integrable! (By integrable we will mean the upper and lower integrals are equal.) As a result, we like the Lebesgue integral and are not so inclined to like the Riemann integral, even though for many practical purposes they are indistinguishable (because for really nice functions, they are the same). Figure 70 illustrates both versions of integration.

12.4.1 *Lebesgue Integration*

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

1. Define step functions carefully (We'll call them *simple functions*).
2. Define integrals of step functions.
3. Approximate general functions using step functions and define the integral of the function as the limit of the integrals of the approximating step functions.

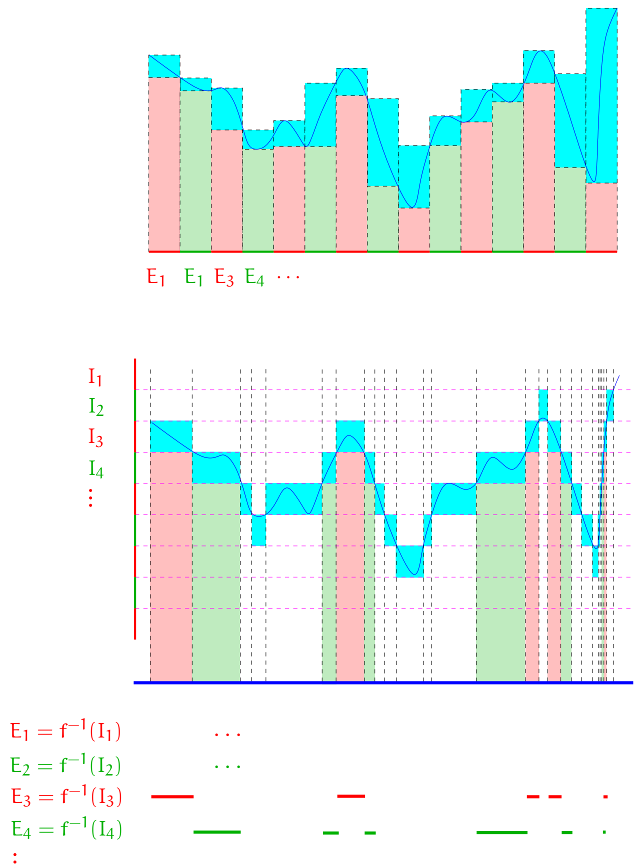


Figure 70: Riemann versus Lebesgue Integration: the upper figure illustrates the partition of the domain dictated by the Riemannian approach. The green and red rectangles live completely below the graph of f. Call the area they sum to $A_{\text{lower}}(P)$ where P is the partition. The red and green plus the cyan rectangles live completely above the graph. Call their area A_{upper}. If $\sup_P A_{\text{lower}}(P) = \inf_P A_{\text{upper}}$ then f is Riemann integrable. The lower figure illustrates that key difference for the Lebesgue case: we partition the range and pull that back by f^{-1} to a partition of the domain. It turns out that this is exactly what is needed to make all reasonable functions integrable. Now $A_{\text{lower}}(P) = \sum_i a_i \mu(E_i)$ and $A_{\text{upper}}(P) = \sum_i b_i \mu(E_i)$ where P is a partition of the range into the intervals $I_i = [a_i, b_i)$.

12.4.2 *Simple Functions*

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

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

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

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

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

Definition 12.4.1 (Simple Functions)**.** *A simple function is any function whose range is a countable set. Suppose that* $\{a_i\}_{i=1}^N$*, with* $N \leqslant \infty$*, are the values that* $f : X \to \mathbb{R}$ *takes. Then defining* $E_i = f^{-1}(a_i)$*, we have that*

$$f(x) = \sum_{i=1}^N a_i \chi_{E_i}(x).$$

We note that $E_i \cap E_j = \emptyset$ *if* $i \neq j$ *and* $\cup_i E_i = X$

12.4.3 *Integrating Simple Functions*

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

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

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

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

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

12.4.4 Integrating Arbitrary Functions

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

Definition 12.4.2 (Measurable Functions). *If $E = f^{-1}(B)$ is a μ-measurable subset of $\mathbb{R}^n$ for all Borel $B \subset \mathbb{R}$, then $f : \mathbb{R}^n \to \mathbb{R}$ is said to be **measurable**.*

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

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

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

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

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

Definition 12.4.3 (Integral of an Arbitrary Function). *Define* $f^+ = \max\{f, 0\}$ *and* $f^- = \max\{-f, 0\}$*. If* f^+ *and* f^- *are integrable and one of the values is not infinite, then*

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

Theorem 12.4.1. *Any* μ*-measurable, non-negative function is integrable.*

Exercise 12.4.1. **(Challenging)** Prove Theorem 12.4.1. Hint: Here are some steps (ignore the hint for a more challenging exercise).

1 Enummerate the unit cubes in the upper half space of $\mathbb{R}^{n+1}$, the graph space of $f : \mathbb{R}^n \to \mathbb{R}$, to get $\{C_i\}_{i=1}^{\infty}$. For example, the first one might be $[0, 1) \times [0, 1) \times \cdots \times [0, 1)$ where the first n factors are in the domain and the $n + 1$st is in the range.

2 Define $\mathrm{ran}(C_i) = [l_i, l_i + 1)$ to be the last interval in the definition of C_i – the interval this cube covers over the n-cube defined by the first n intervals in the definition of C_i.

3 Define $\mathrm{Dom}(C_i)$ to be the n-cube defined by the first n intervals in the definition of C_i.

4 Define, for $k = 1$ to 2^i,

$$E_{i,k} = \left(f^{-1}([l_i + \frac{(k-1)\epsilon}{2^i}, l_i + \frac{k\epsilon}{2^i})) \right) \cap \mathrm{Dom}(C_i).$$

5 Now use very analogous ideas to Example 12.2.1 to construct upper and lower simple functions whose integrals differ by at most ϵ.

Remark 12.4.1. *Note that many other authors use the term integrable to mean what we mean by integrable* and *that* $|\int f d\mu| < \infty$.

Definition 12.4.4 (Summable). *We will say that* f *is* μ-summable *if* f *is integrable and* $|\int f d\mu| < \infty$.

Remark 12.4.2. *We notice immediately that sets of measure zero have no impact on the value of the integral: we may redefine the function on a set of measure zero and the integral remains unchanged. Notice also that a countable number of measure zero sets has a union that also has measure zero. This is a handy fact to keep in mind.*

Remark 12.4.3. *Notice that the hint in exercise 12.4.1 implies that in fact, we can focus on partitions of the domain that are pullbacks of partitions of the range* $\mathbb{R}$ *into intervals:* $E_i = f^{-1}(I_i)$.

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

12.4.5 Properties of Integrals and Measurable Functions

Exercise 12.4.2. **(Linearity)** Show that if the integral of one of f or g is finite, the Lebesgue integral is linear:

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

Exercise 12.4.3. You might like to try to prove the following theorem that appears on page 5 of Evans and Gariepy: at least think about it before you look up the proof. *Let* μ *be a Borel regular measure on* $\mathbb{R}^n$. *Define* $\mu \llcorner A(E) \equiv \mu(E \cap A)$. *Suppose that* $A \subset \mathbb{R}^n$ *is* μ*-measurable and* $\mu(A) < \infty$. *Then* $\mu \llcorner A$ *is a Radon measure.*

Exercise 12.4.4. Suppose that $f : X \to Y$ and suppose that $(f^{-1}(A)|A \in \mathcal{A})$ is measurable in X. Prove that $(f^{-1}(B)|B \in \mathcal{B})$ is also measurable where $\mathcal{B}$ is the σ-algebra in Y generated by $\mathcal{A}$. Hint: use Exercise 12.3.3.

Exercise 12.4.5. **(Properties of Measurable Functions I)** Use the results of exercise 12.4.4 to show that if $f : \mathbb{R}^n \to \mathbb{R}$, then showing that all sets in $\{f^{-1}((-\infty, a)) \mid a \in \mathbb{R}\}$ are measurable is enough to show that f is measurable. Do the same for the collection of sets $\{f^{-1}((-\infty, a]) \mid a \in \mathbb{R}\}$.

Exercise 12.4.6. **(Properties of Measurable Functions II)** Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$, and $\{f_k : \mathbb{R}^n \to [-\infty, \infty]\}_{i=1}^{\infty}$ are all μ-measurable. Prove:

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

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

12.5 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

$$\lim_{i\to\infty} \left(\int f_i \, dx \right) = \int \left(\lim_{i\to\infty} f_i \right) \, dx? \tag{29}$$

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.

12.5.0.1 *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 29 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 12.5.1. 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 29 is yes.

Exercise 12.5.2. 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 the question in Equation 29 is no. Hint: look at constant functions on the real line.

Exercise 12.5.3. Give an example of a uniformly convergent sequence $f_i : \mathbb{R} \to \mathbb{R}$, such that

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

and

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

Hint: choose $f_i \geqslant 0$ such that $\lim_{i\to\infty} f_i(x) = 0$ for every x, and does so uniformly, and $\int_{\mathbb{R}} f_i dx = 2$ for all i.

Exercise 12.5.4. Give an example of a non-uniformly convergent sequence f_i on a finite measure space X where again the answer to the question in Equation 29 is no. Hint: on the unit interval, with the usual Lebesgue measure, construct a sequence $f_i \to f \equiv 0$ such that $\int f_i \, dx = 1$ for all i.

12.5.1 *Types or Modes of Convergence*

The above examples look at the question of the connection between pointwise convergence and convergence 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 A.E. If $f_i(x) \to f(x)$ as $i \to \infty$ for all x except $x \in E$, a set of measure 0, we say that f_i converges to f **almost everywhere (a.e.)**.

CONVERGENCE IN MEASURE If, for any $\epsilon > 0$ we have that

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

then we say 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 sequence f_i and limit f, then we say that the f_i converge in norm to f. This is also referred 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 test 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 12.5.5. 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 12.5.6. 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 12.5.7. Find an example to show that convergence in measure does not imply convergence in norm. Hint: the f_i need not be bounded.

Exercise 12.5.8. Suppose we choose the norm given by $\|g\| = \int |g| \, dx$. Show that if the f_i and f are uniformly bounded (i.e. $-C \leqslant f_i, f \leqslant C$ for some $C > 0$), then convergence in measure implies convergence in norm and convergence a.e.

Exercise 12.5.9. We will work with functions from the $[0, 1] \subset \mathbb{R}$ to the real numbers $\mathbb{R}$. Test functions ϕ, will be anything in the class of infinitely differentiable functions from the closed unit interval to the real numbers, $\Phi = C^\infty([0, 1])$. Find an example of a sequence of functions $\{f_i\}_{i=1}^\infty$ which converges to 0 nowhere, but which converges weakly to $f \equiv 0$. I.e. $\lim_{i\to\infty} f_i(x) \not\to 0$ for any $x \in [0, 1]$, but $\lim_{i\to\infty} \int_{[0,1]} \phi f_i dx \to 0$ for all $\phi \in C^\infty([0, 1])$ Hint: think traveling waves ...

Exercise 12.5.10. **(Look at all the possibilities!)** Suppose we identify each of the 5 bit binary numbers with a set of convergence types:

$f_i \to_{01101} (f \equiv 0)$ would be shorthand for the fact that f_i converges to the zero function a.e., in measure and weakly but not uniformly or in norm. Is it possible to find sequences converging to zero for each of the binary numbers? If not which ones are possible?

12.5.2 The Three Theorems

In the next three theorems and the examples that follow, we examine the fact that while limits and integrals don't always commute, we can switch the order of integration and limit taking in many useful cases. First though, we recall the definition of lim inf and lim sup from Section 7.6 and then extend the notion to sequences of functions.

Definition 12.5.1 ((**Reminder from Chapter 7**) $\limsup$, $\liminf$ **for Sequences in** $\mathbb{R}$). *Suppose that* $\{x_i\}_{i=1}^{\infty} \subset \mathbb{R}$. *We define:*

$$\limsup_{i\to\infty} x_i \equiv \lim_{m\to\infty} \sup(\{x_i\}_{i=m}^{\infty}) \text{ and}$$

$$\liminf_{i\to\infty} x_i \equiv \lim_{m\to\infty} \inf(\{x_i\}_{i=m}^{\infty}).$$

In Chapter 7 this notion came up and went by pretty quickly. Here is an exercise that has you convince yourself that lim inf and lim sup always exist.

Exercise 12.5.11. (**$\limsup$ and $\liminf$, again**) 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 $F_n \equiv \{f(i) | i \geqslant n\}$, we could imagine the smallest closed interval 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 exercise makes that precise.

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 $l^* \leqslant r^*$.
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 exercise, it becomes clear that the $\liminf_{i\to\infty} f$ and $\limsup_{i\to\infty} f$ define the eventual envelope which contains the oscillations of f "at infinity".

Restating the definition of $\liminf$ and $\limsup$ using the notation used in the last exercise, we get

Definition 12.5.2 ($\limsup_{i\to\infty} f(i)$ and $\liminf_{i\to\infty} f(i)$)**.** *Suppose that* $f : \mathbb{N} \to \mathbb{R}$.

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

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

Now we extend this to sequences of functions.

Definition 12.5.3 ($\limsup_{i\to\infty} f_i(x)$ 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 12.5.1 (Fatou's Lemma)**.**

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

Theorem 12.5.2 (Monotone Convergence Theorem)**.** *Suppose that* $\{f_i\}$ *are all measurable and that* $0 \leqslant f_1 \leqslant ... \leqslant f_i \leqslant f_{i+1} \leqslant ...$. *Then we have that*

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

Theorem 12.5.3 (Dominated Convergence Theorem)**.** *If* $f_i \to f$ μ *a.e.,* $|f_i|, |f| < g$ *and* $\int g \, dx < \infty$*, then*

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

In all these theorems, all the functions are assumed to be measurable.

12.5.3 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, for example, Evans and Gariepy's proofs in Chapter 1 of [12].) 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.

1 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 12.5.12).

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

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

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

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

□

Exercise 12.5.12. **(Challenge) (Weighted Measures: μ_g from μ)**

1 If μ is an outer measure, $g \geqslant 0$ 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.

2 Give an example illustrating 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.

1 Define $g(x) = \lim_{i\to\infty} f_i(x)$
2 If $\int g \; dx < \infty$, use the dominated convergence theorem to get the result.
3 Otherwise, because $\int g \; dx = \infty$, we can find a simple function g_C such that $g_C \leqslant g$ everywhere and $\int g_C \; dx > C$ for any $C < \infty$.
4 Define $g_{C,K} = \min(g_C, K)$. Choose $\epsilon > 0$ and K big enough that $\int g_{C,K} dx > C - \frac{\epsilon}{2}$
5 Define $E_n = \{x | f_i \geqslant (1-\epsilon) g_C \;\forall\; i \geqslant n\}$. Choose n big enough that $\mu_{g_C}(X \setminus E_n) \leqslant \frac{\epsilon}{2K}$.
6 This implies that $\mu_{g_{C,K}}(X \setminus E_n) \leqslant \frac{\epsilon}{2K}$.
7 We have that for $i \geqslant n$

$$\begin{aligned}
\int g_i \; dx &\geqslant \int_{E_n} g_i \; dx \\
&\geqslant \int_{E_n} (1-\epsilon) g_C \; dx \\
&\geqslant \int_{E_n} (1-\epsilon) g_{C,K} \; dx \\
&= (1-\epsilon) \int_{E_n} g_{C,K} \; dx \\
&= (1-\epsilon) \left(\int_X g_{C,K} dx - \int_{X \setminus E_n} g_{C,K} dx \right) \\
&\geqslant (1-\epsilon)(C-\epsilon).
\end{aligned}$$

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.

1 Define $h_n(x) = \inf_{i \geqslant n} f_i(x)$. Note that $\liminf_{i\to\infty} f_i = \lim_{i\to\infty} h_i$.
2 Note that $\int h_n \, dx \leqslant \int f_i \, dx$ for all $i \geqslant n$. We conclude that $\int h_n \, dx \leqslant \liminf_{i\to\infty} \int f_i \, dx$ for all n.
3 This implies that

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

□

Remark 12.5.1. *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 measurable functions can be used to create the weighted measures we study in Exercise (12.5.12).*

Remark 12.5.2. *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 12.5.3. *Dominated convergence is used to get other switching theorems: switching order of differentiation and summation or differentiation and integration or integration and summation.*

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

In this section, we state and discuss the meaning of four theorems that we will not prove. (If you are interested in the proofs, I recommend Evans and Gariepy [12].

12.6.1 Lipschitz Functions

Definition 12.6.1 (Lipschitz Functions). $F : X \to Y$ *is* ***Lipschitz*** *if there is a positive number* $K \geqslant 0$ *such that* $|F(x) - F(y)| \leqslant K|x - y|$ *for all* $x, y \in X$.

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

Theorem 12.6.1 (Rademacher'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 Lebesgue measure zero. I.e.* F Lipschitz ⇒ F is differentiable almost everywhere.

It turns out that Lipschitz functions are nice enough for many purposes. While differentiability everywhere generally makes proofs easier, frequently, mere Lipschitz smoothness does not stand in the way of deeper generalizations of theorems best known in their C^k forms, for some $k \geqslant 1$.

12.6.2 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 $F : \mathbb{R}^n \to \mathbb{R}^m$, with $n \leqslant m$. When $n \leqslant m$ we define $|JF| = \sqrt{\det(DF^t \circ DF)}$, where the t superscript indicates transpose.

In this case we have:

Theorem 12.6.2 (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 $F : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz and $n \geqslant m$, we define $|JF| = \sqrt{\det(DF \circ DF^t)}$. In this case, we have:

Theorem 12.6.3 (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 12.6.4 (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 12.6.5 (Coarea Formula, version 2).

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

While it is not hard to combine both of the second versions to get a general area-coarea formula, there is little conceptual advantage to that, and here we are focusing on the conceptual understanding of these two formulas. (The combined version can be found on page 341 of this book.)

Remark 12.6.1. *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 formula you have seen for integrals in calculus. To really understand these two formulas, we need to look at simple examples.

Example 12.6.1 (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*

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

Example 12.6.2 (A Nonlinear Fubini's Theorem). *The example above of integrating over spheres and then over radii is a special case of integration over distance functions. If we let* $h(x) = d(x, K)$ *where* $d(x, K)$ *is the distance from* x *to the set* K, *we have that the gradient of* d *is is a unit vector everywhere except on the interior of* K *so the Jacobian* $|Jd| = 1$ *almost everywhere in* K^c. *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 12.6.3 (Area of Graphs). *If we want to know the n-area (or n-volume) of a graph of* $F : \mathbb{R}^n \to \mathbb{R}^1$ *over* $\Omega \in \mathbb{R}^n$, *then we are asking for* *the n-volume of the set* $\{(x, F(x))|x \in \Omega\} \subset \mathbb{R}^{n+1}$*. We define the mapping* $\hat{F} : \mathbb{R}^n \to \mathbb{R}^{n+1}$ *by* $\hat{F}(x) = (x, F(x))$*. We get that*

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

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

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

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

Remark 12.6.2 (Draw Some Pictures!). *When I teach this, I always give an intuitive explanation of both the area and coarea formulas using pictures. You should try doing this for yourself! Try messing around with simple (i.e. not very complex) functions mapping* $\mathbb{R}^2$ *to* $\mathbb{R}$ *to explore the co-area formula and maps mapping* $\mathbb{R}^2$ *into* $\mathbb{R}^3$ *to explore the area formula. You can start by working on Exercise 12.6.1.*

Exercise 12.6.1. Imagine a function like a slightly skewed, broad Gaussian in $\mathbb{R}^2$. See Figure 71. See if you can understand why the area of the red strips in the first sub-figure of Figure 72 is the same as the integral of the Jacobian over E, or Jacobian weighted area of E. Hint: see Figure 73

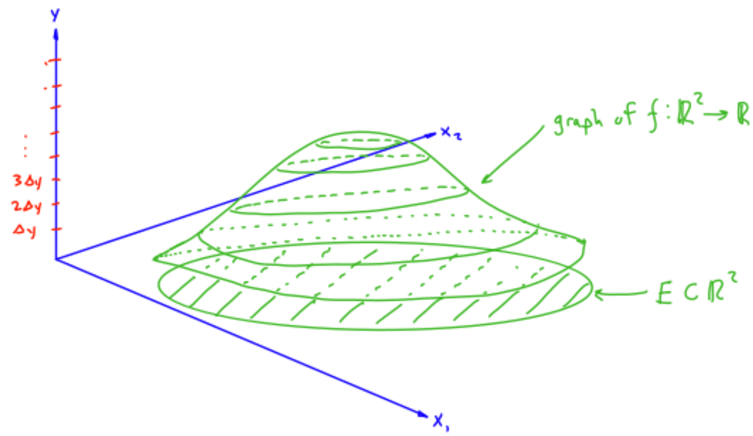


Figure 71: Graph of a slightly skewed, broad Gaussian-like bump function in $\mathbb{R}^2$.

12.6.3 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 (i.e. map the entire domain to the origin in the range) gets you a mapping whose rank is never full on the entire domain, this non-full-rank set is large measured in the domain. But what about in the range? 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 is, not very much: to be more precise, the measure of that set in the range is zero!

Theorem 12.6.6 (Sard's Theorem). *Suppose that* $F : \mathbb{R}^n \to \mathbb{R}^m$ *and that* F *is* C^k *with* $k > \max(0, n-m)$*. Define* $\mathcal{C}$ *to be the set of points* $x \in \mathbb{R}^n$ *such that* $\operatorname{rank}(DF_x) < m$*. Then* $\mathcal{L}^m(F(\mathcal{C})) = 0$*.*

This theorem is a technical tool, extensively used in analysis and geometric analysis. It justifies the intuition that when the rank of the derivative of a map $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \leqslant n$, is less than m, so that

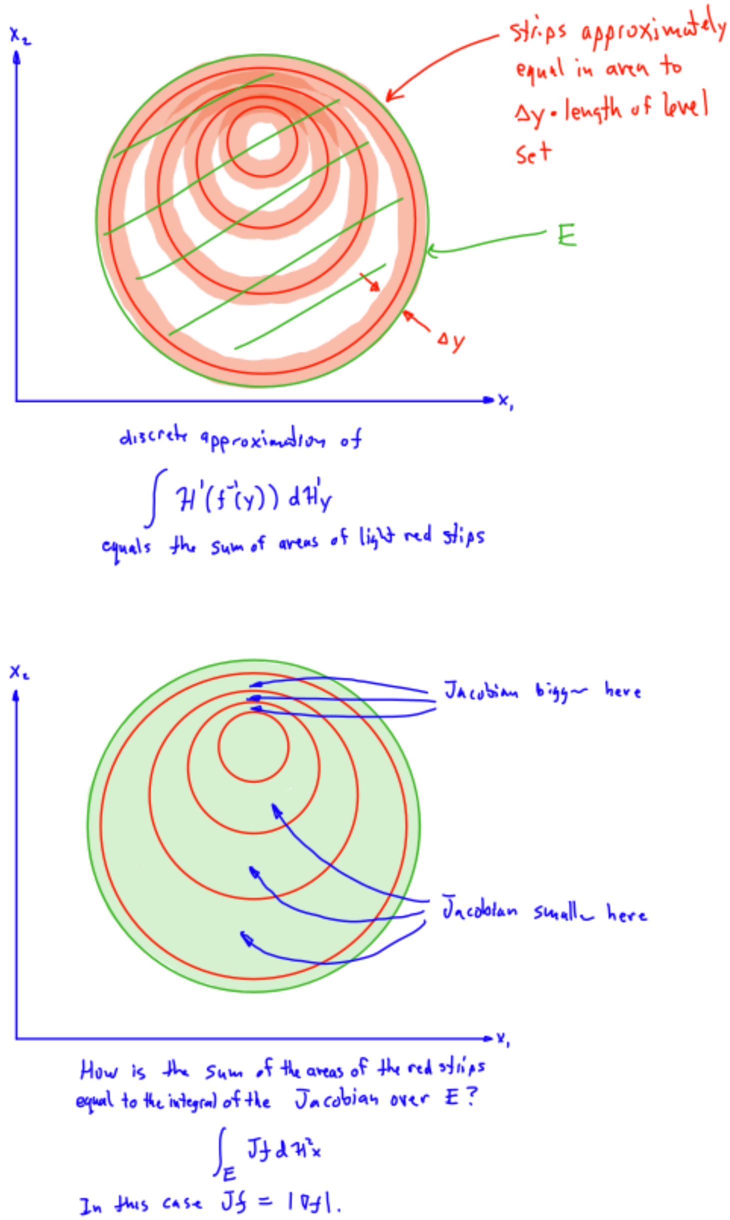


Figure 72: Explain why the area of the red strips, added together, equals that area of E, weighted by the Jacobian – i.e. the integral of Jf over E.

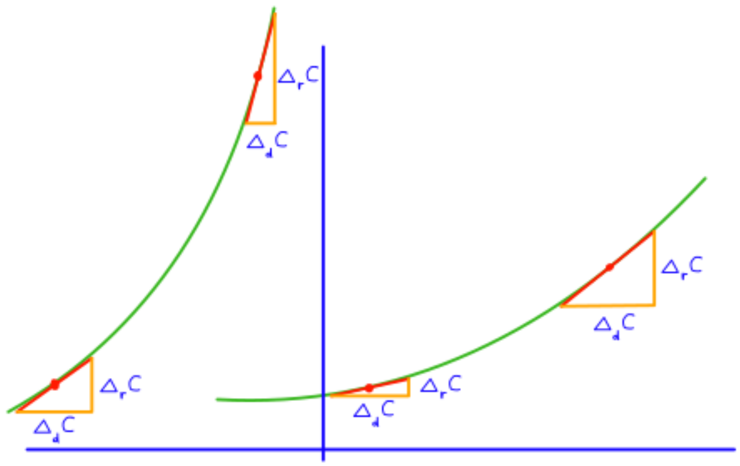


Figure 73: Hint for Exercise 12.6.1: slicing through the graph in a direction orthgonal to the leve sets, we get curves and can interpret slopes using $\Delta_d C$ = {spacing between level sets in domain} and $\Delta_r C$ = {spacing between level sets in the range}.

the derivative is not onto, then the mapping squeezes space down, collapsing at least one dimension, yielding a measure zero set.

Most of the easier proof of this result when $F \in C^\infty$, 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 [31].

Remark 12.6.3. *In the smooth case $F \in C^k$ with $k = \infty$, it would appear that we only need $k+1 > \frac{n}{m}$ while in the case that we do not assume as much smoothness, we need $k > \max(0, n-m)$. We should expect that*

$\max(0, n - m) + 1 \geqslant \frac{n}{m}$. *But when* $n \leqslant m$ *the inequality is obvious. So we assume* $n > m$ *and define* $p = n - m$. *Then* $\max(0, n - m) + 1 \geqslant \frac{n}{m}$ *turns into* $p + 1 \geqslant \frac{p}{m} + 1$ *which is also crearly true for all* $m = 1, 2, 3, \ldots$.

Part V

Analysis III

13 Other Tools for Nonlinear Analysis

In this chapter, we look at three ideas important for nonlinear analysis, after which we revisit the three integrals that connect the local and global properties of nonlinear functions and mappings.

Solving $F(x) = b,\ x, b \in \mathbb{R}^n$: the nonlinear version of $Ax = b$ for A an $n \times n$ matrix. The first step in solving a problem is to ensure that there is a solution – to show existence. We will consider one such tool that also gives us an iterative way to compute a solution: The **Banach Fixed Point Theorem**.

Solving $F(x) = b,\ x \in \mathbb{R}^n, b \in \mathbb{R}^{n-k},\ 1 \leqslant k < n$: the nonlinear version of $Ax = b$ for A an $n \times (n-k)$ matrix. Solutions are level sets of F, with regular level sets being k-dimensional submanifolds of $\mathbb{R}^n$. Solutions of systems of nonlinear equations that define regular level sets is intimately tied up with the notion of **transverse intersections**.

Solving $\min_x F(x)$, **for** $F : \mathbb{R}^n \to \mathbb{R}$: there are no non-trivial linear analogs of this unconstrained minimization problem – the only interesting linear optimization problems are interesting because they are constrained, i.e. the search is restricted to $E \subset \mathbb{R}^n$ – $\min_{x \in E} F(x)$. The simplest class of interesting unconstrained optimization are those in which F is a convex function. Thus we take a detailed look at **convex functions and their properties**.

13.1 Banach Fixed Point Theorem

Many problems can be written as:

Problem 13.1.1 (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 review all the ideas, starting with vector norms, leading to the idea of a Banach space.

Definition 13.1.1 (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 it satisfies:*

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

Definition 13.1.2 (Cauchy Sequence). *Recall that* $x_i \in X$ *is Cauchy if for any* $\epsilon > 0$ *there is an* $N(\epsilon)$ *such that* $i, j > N(\epsilon)$ *implies that* $\|x_i - x_j\| < \epsilon$.

Definition 13.1.3 (Complete Space). *If every Cauchy sequence in* X *has a limit in* X*, then* 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 13.1.4 (Banach Space). *A complete, normed vector space is called a* ***Banach Space****.*

Mappings that are Lipschitz, with a Lipschitz constant strictly less than 1 are called contractions:

Definition 13.1.5 (Contraction Mapping). *A function from a normed space* X *to itself is a* ***contraction mapping*** *if* $\|F(x) - F(y)\| < k\|x - y\|$ *for some* $0 \leqslant k < 1$.

Now we are ready to state the main theorem in this section:

Theorem 13.1.1 (Banach Fixed Point Theorem). *Suppose that* $F : B \to B$ *is a contraction mapping from the Banach space* B *to itself. Then there is a unique point* x^* *such that* $F(x^*) = x^*$.

Proof.
First note that if there 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. $\|F^{i+1}(x_0) - F^i(x_0)\| \leqslant k^i\|F(x_0) - x_0\|$
   b. for $n > m$

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

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

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

   c. Thus, $\{x_i\}_{i=1}^{\infty}$ is a Cauchy sequence.
3. Therefore, there is a point x^* in B such that $x_i \to x^*$ as $i \to \infty$.
4. Since F is continuous, we have that

$$\lim_{i\to\infty} F(x_i) = F(\lim_{i\to\infty} x_i) = F(x^*).$$

But $F(x_i) = x_{i+1}$ so

$$\lim_{i\to\infty} F(x_i) = \lim_{i\to\infty} x_{i+1} = x^*.$$

Thus $F(x^*) = x^*$.

$\square$

Here are some figures illustrating the Banach fixed point theorem in the case that the Banach space is $\mathbb{R}$. The first figure, Figure 74 shows an F such that its derivative is positive and strictly less than one everywhere. The second, Figure 75 shows a contractive iteration converging to the fixed point x^*. The Third figure, Figure 76 shows that we can change the function away from where the iteration happens and preserve iterations converging to the fixed point.

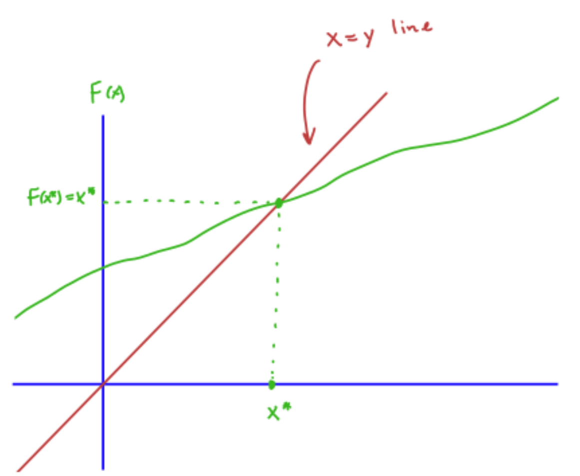


Figure 74: Contraction mapping with a single fixed point, x^*.

Exercise 13.1.1. Show that one can use the $x = y$ line to quickly plot the trajectory of a point under iteration by the formula $x_n = F(x_{n-1})$. (See Figure 77.)

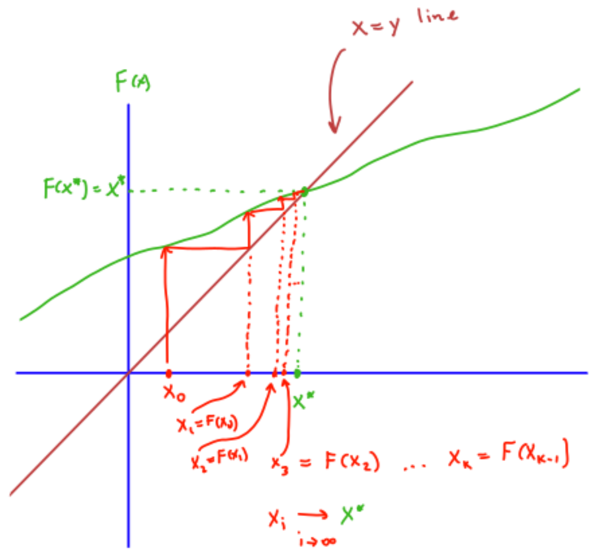


Figure 75: Iteration to the fixed point from a starting point x_0.

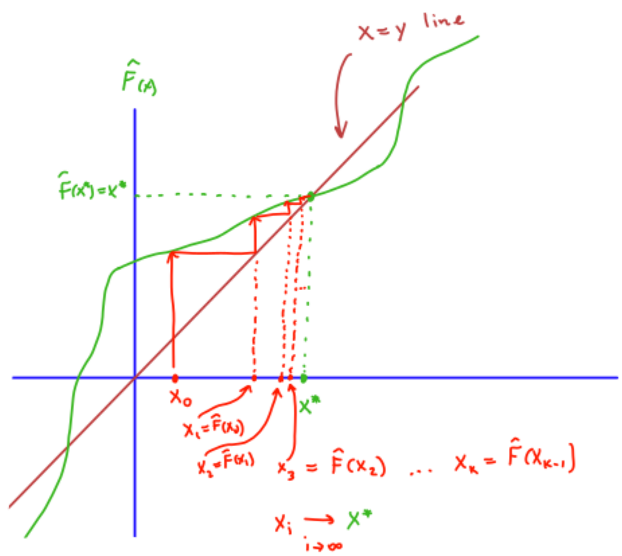


Figure 76: The iteration remains unchanged if we do not change the function locally.

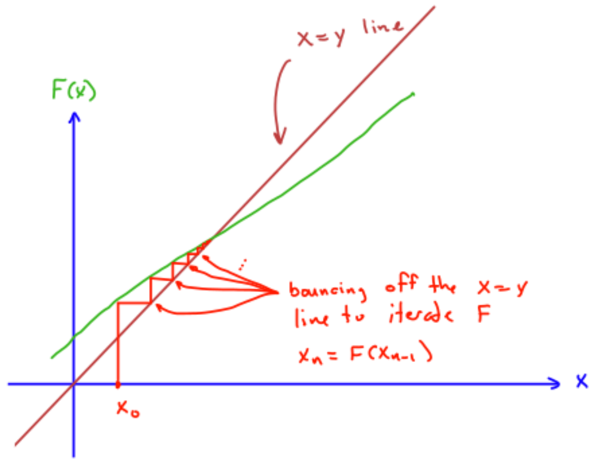


Figure 77: Figure that shows the use of the $x = y$ to quickly plot an iteration. See Exercise 13.1.1.

Exercise 13.1.2. (Stability of Fixed Points) Let $F : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable mapping with a fixed point x^*. (See Figure 78.)

1 Suppose that
$$-1 < F'(x^*)(= \frac{dF}{dx}(x^*)) < 1.$$
Show that for some $\epsilon > 0$, and any $x \in (x^* - \epsilon, x^* + \epsilon)$ we have
$$|F(x) - x^*| = |F(x) - F(x^*)| < |x - x^*|.$$
We will call such a fixed point a stable fixed point.

2 Suppose that
$$1 < |F'(x^*)|.$$
Show that for some $\epsilon > 0$, and any $x \in (x^* - \epsilon, x^* + \epsilon)$ we have
$$|F(x) - x^*| = |F(x) - F(x^*)| > |x - x^*|.$$
We will call such a fixed point an unstable fixed point.

3 What can you say about the case in which $F'(x^*) = 1$? See if you can find examples where (a) $F(x^*) = x^*$, $F'(x^*) = 1$ and x^* is a stable fixed point and (b) $F(x^*) = x^*$, $F'(x^*) = 1$ and x^* is an unstable fixed point.

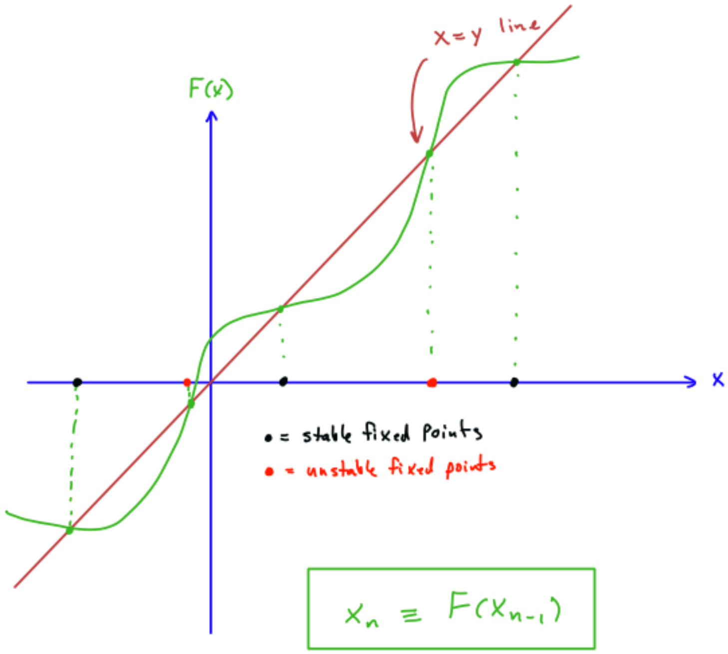


Figure 78: The stability of the fixed points is determined by the slope of the graph at the fixed point. See Exercise 13.1.2.

Exercise 13.1.3. Exercise 13.1.2 can be shown analytically and it can be shown with a geometric argument. Assuming you only gave one argument (say analytic or geometric) when you did Exercise 13.1.2, provide the other of those proofs (resp. geometric or analytic).

Exercise 13.1.4. This exercise moves you to the continuous analogs of the iterations studied above. It turns out that 1-dimensional iterations can be very complex – the map $x_n = \lambda x_{n-1}(1 - x_{n-1})$ can display truly complex behavior and, for values of λ between 3 and 4, is used as a model for chaos. In the case of flows in 1 space dimension, $\dot{x} = F(x)$ where $F : \mathbb{R} \to \mathbb{R}$ and $\dot{x} \equiv \frac{dx}{dt}$, the behavior is much simpler, though still interesting. See Figure 79. By making plots similar to that shown in Figure 79, explore the following questions: In each of these we will be looking at some aspect of a continuous flow, $\dot{x} = F(x)$.

1 Suppose that x_0 is a fixed point – $F(x_0) = 0$.
   a Use a Taylor's expansion of F about x_0,

$$F(x_0 + h) = F(x_0) + F'(x_0)h + \frac{1}{2}F''(x_0)h^2 + ...$$

   to show that for any $\eta > 0$ and a small enough ϵ, $h \in [-\epsilon, \epsilon]$ implies that

$$(F'(x_0) - \eta)h \leqslant \dot{h} \leqslant (F'(x_0) + \eta)h$$

   from which we can conclude that $h(t)$ lies between $h(0)e^{(F'(x_0)-\eta)t}$ and $h(0)e^{(F'(x_0)+\eta)t}$. Since η can be as small as you like, we say that $h(t) \approx h(0)e^{F'(x_0)t}$ (for the times t in which $h(t) \in [-\epsilon, \epsilon]$).
   b Notice that, if we discretize time with very small steps Δt then locally, around x_0, $x(t + \Delta t) = e^{F'(x_0)\Delta t}x(t)$ ... i.e. $x_n = F(x_{n-1})$ where $F(x) = e^{F'(x_0)\Delta t}x$. So that negative (resp, positive) derivatives of F in a flow corresponds to slopes that are less than (resp. greater than) 1 in the implicit iteration generated locally.

2 Suppose that $F(x) = x^2 + \alpha$. What can you say about the stability of the fixed point at 0 as α changes from a negative number to a positive number? Draw the *bifurcation diagram*. (To understand what a bifurcation diagram is, see the example in Figure 80.)

3 Suppose that $F(x) = x(x^2 + \alpha)$. What can you say about the stability of the fixed point at 0 as α changes from a negative number to a positive number? Draw the bifurcation diagram.

4 Suppose that $F(x) = x^3 - x + \alpha$. Draw the bifurcation diagram.

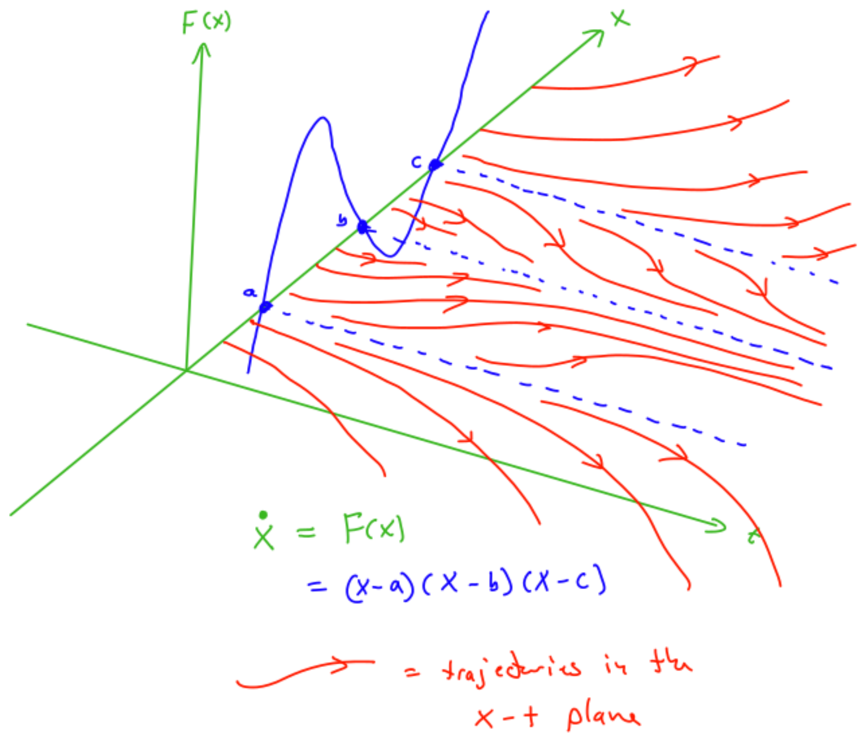


Figure 79: Example of a flow in 1 (space, i.e. x) dimension, plotted in 3 dimensions to simultaneously show the statespace and the flow in time.

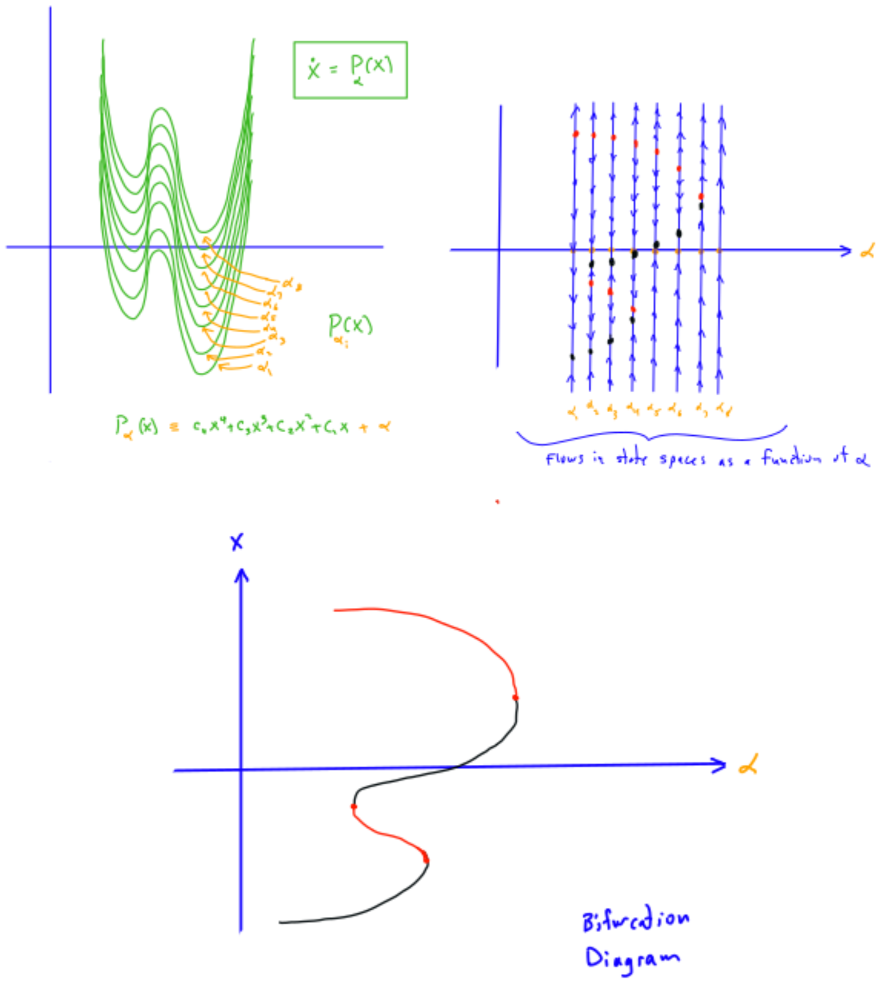


Figure 80: An example of the changes in fixed points as the parameter α changes. The resulting bifurcation diagram is an example of something that can get very complicated in high dimensions.

13.2 Transverse Intersections

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

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

Definition 13.2.1 (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* $\operatorname{span}(U_k, W_m) = \mathbb{R}^n$.

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

Definition 13.2.2 (Transverse Intersection of Submanifolds). *Two submanifolds of $\mathbb{R}^n$, M and N, intersecting at x are said to intersect transversely at x if the tangent spaces T_xM and T_xN intersect transversely as subspaces of $\mathbb{R}^n$, I.e. if* $\operatorname{span}(T_xM, T_xN) = \mathbb{R}^n$.

Example 13.2.1 (2 Curves in $\mathbb{R}^3$). *In $\mathbb{R}^3$, an intersection between 2, 1-manifolds is never transverse.*

Example 13.2.2 (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 13.2.3 (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.

Exercise 13.2.1. Suppose that M_k and N_p are k and p dimensional submanifolds of $H = \mathbb{R}^n$, M_k is a level set of a smooth map $F_M : \mathbb{R}^n \to \mathbb{R}^{n-k}$ and N_p is a level set of a smooth map $F_N : \mathbb{R}^n \to \mathbb{R}^{n-p}$. Work out the relationship of transversality of an intersection of M_k and N_p at x and the rank of the mapping $F : \mathbb{R}^n \to \mathbb{R}^{2n-p-k}$ defined by

$$x \in \mathbb{R}^n \to \begin{bmatrix} F_M(x) \\ F_N(x) \end{bmatrix} \in \mathbb{R}^{2n-k-p}.$$

Exercise 13.2.2. Suppose that $f_1(x, y) \equiv y - x^2 = \alpha$ and $f_2(x, y) \equiv y - x = \beta$.

1 Find the values of α and β such that there are intersections. Classify the intersections as transverse or not.
2 Show that an intersection, defined by $f_1(x, y) = \alpha_0$ and $f_2(x, y) = \beta_0$, that is transverse, is stable. More precisely, use the implicit function theorem to show that if $f_1(x_0, y_0) = \alpha_0$ and $f_2(x_0, y_0) = \beta_0$ defines a transverse intersection, then there is an $\epsilon > 0$ such that, defining $h_x = x - x_0$ and $h_y = y - y_0$, for $|h_x| < \epsilon$ and $|h_y| < \epsilon$, there are $g_\alpha(h_x, h_y)$ and $g_\beta(h_x, h_y)$ such that $f_1(x_0 + h_x, y_0 + h_y) = \alpha_0 + g_\alpha(h_x, h_y)$ and $f_2(x_0 + h_x, y_0 + h_y) = \beta_0 + g_\beta(h_x, h_y)$. **Hint**: consider the mapping $F : \mathbb{R}^4 \to \mathbb{R}^2$ defined by

$$F(x, y, \alpha, \beta) = \begin{bmatrix} f_1(x, y) - \alpha \\ f_2(x, y) - \beta \end{bmatrix}$$

at the point $(x_0, y_0, \alpha_0, \beta_0)$.

13.3 Convex functions and Subgradients

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $E \subset \mathbb{R}^n$.

A **function f is convex** if $f(\alpha_1 x + \alpha_2 y) \leqslant \alpha_1 f(x) + \alpha_2 f(y)$ for all $\alpha_i \geqslant 0$ 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 $\alpha_i > 0$, $i = 1, 2$, also 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 $\alpha_i \geqslant 0$ 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 13.3.1. 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 13.3.2. 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 for $y \in E$, $\langle y - y_x, x - y_x \rangle$ is always non-positive.

Exercise 13.3.3. Prove that y_x is the closet point in E to every point in the ray $x + \alpha \cdot (x - y_x)\ \forall\ \alpha \in \{\mathbb{R} \geqslant 0\}$

Definition 13.3.1 (Subgradient and Subdifferential). *Let $\mathcal{H}$ be a Hilbert space, $f : \mathcal{H} \to \mathbb{R}$ be a convex function and $H(x) = \langle v_p, x - p\rangle + f(p)$ be a supporting hyperplane at $(p, f(p))$. We refer to v_p as a **subgradient of** f **at** p and the set of subgradients of f at p, as the **subdifferential of** f **at** p.*

Exercise 13.3.4. **(Challenge)** 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 $w \in B(x, \delta) \cap L_E(c)$, $\epsilon \leqslant \langle x - y_x, w - x\rangle \leqslant 0$.

Exercise 13.3.5. **(Challenge)** Use Exercises (13.3.1-13.3.4) 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 $d(y, L_E(1)) = |x^* - y| > 1$. Here you can either

1. note that because $E \subset \mathcal{H}_{y_{x^*}, x^* - y_{x^*}}$ conclude that $d(x^*, E) > 1$ which is a contradiction, or
2. note that since $y \in \text{bdy}(E)$ there is a sequence of points in $\{x_i\}_{i=1}^{\infty} \subset E^C$ converging to y. The sequence of points $\{y_{x_i}\}_{i=1}^{\infty} \subset E$ (where y_{x_i} is the unique closest points in E to x_i) also converges to y. The points $y_{x_i} + \frac{x_i - y_{x_i}}{|x_i - y_{x_i}|}$ all lie on $L_E(1)$. Which, you then show is a contradiction.

Exercise 13.3.6. 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.

Definition 13.3.2 (Epigraph). *The epigraph,* E_f*, of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is given by*

$$E_f \equiv \{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leqslant y\}.$$

These are just the points in the graph space on or above the graph of f*.*

Exercise 13.3.7. This exercise looks at the relationship between epigraphs and convexity.

1 Show that $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if the epigraph E_f is convex. Show this is still true if we allow functions to take on infinite values. i.e. $f : \mathbb{R}^1 \to \{\mathbb{R}^1 \cup +\infty\}$
2 Suppose that $f : D \subset \mathbb{R}^1 \to \mathbb{R}^1$ is convex and D is a closed (possibly infinite) interval in $\mathbb{R}$. Show that the epigraph E_f is closed and convex in $\mathbb{R}^2$.
3 Show by an example of a convex function $f : \mathbb{R}^1 \to \{\mathbb{R}^1 \cup +\infty\}$, that the previous conclusion can be false when f takes on the value $+\infty$ somewhere.

Exercise 13.3.8. 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 13.3.9. Suppose that E_α is convex 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 13.3.10. Prove that $E = \bigcap_{x \in E^c} H_{y_x, x - y_x}$.

Exercise 13.3.11. 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 13.3.12. The *uniformly bounded* condition in Exercise 13.3.11 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 $\{(x, y) \mid f(x) \leqslant y < \infty\} \subset \mathbb{R}^2$. 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 13.3.13. 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 $f(\alpha_1 x + \alpha_2 y) \geqslant \alpha_1 f(x) + \alpha_2 f(y)$ for all $\alpha_1, \alpha_2 \geqslant 0$ and $\alpha_1 + \alpha_2 = 1$.

Exercise 13.3.14. 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 13.3.15. 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 13.3.16. Let H_f be the collection of supporting lines of the convex function f. Show that $f(x) = \sup\limits_{h \in H_f} h(x)$. Consequently, the epigraph of f is the intersection of the upper half-planes defined by the supporting lines.

Exercise 13.3.17. Let $f \in C^2(\mathbb{R}^1, \mathbb{R}^1)$ and suppose that $f''(x) \geqslant 0$ 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 13.3.18. Suppose that $f \in C^2(\mathbb{R}^1, \mathbb{R}^1)$ and f is convex. Show that $f''(x) \geqslant 0$ for all $x \in \mathbb{R}$.

Exercise 13.3.19. 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 13.3.20. Suppose $f : \mathbb{R}^n \to \mathbb{R}^1$ is convex. Show that the sublevel sets $S_f(c) \equiv \{x \in \mathbb{R}^n \mid f(x) \leqslant c\}$ are convex. Give an example of a non-convex function g whose sublevel sets $S_g(c)$ are all convex.

Exercise 13.3.21. Suppose that $f : \mathbb{R}^2 \to \mathbb{R}$. 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 closed ball $\bar{B}(0, C) = \{x \mid |x| \leqslant C\}$ big enough that $f(x) > 2(|f(0)| + 1)$ for $x \in \bar{B}(0, C)^c$ and use the fact that $\bar{B}(0, C)$ is a compact set.

Exercise 13.3.22. Give an example of a convex function that does not have a minimal value.

Definition 13.3.3 (Directionally Coercive)**.** *Suppose that* X *is a normed linear space. We will say that* $f : X \to \mathbb{R}$ *is* ***directionally coercive*** *if, for all* $v \in \partial B(0, 1) \subset X$, $f(sv) \underset{s\to\infty}{\to} \infty$.

Exercise 13.3.23. **(Challenge)** Give an example $f : \mathbb{R}^2 \to \mathbb{R}^1$ that is continuous and directionally coercive but not coercive.

Exercise 13.3.24. Prove that when $f : \mathbb{R}^n \to \mathbb{R}$ is convex, directionally coercive implies coercive. **Hint**: suppose that f is directionally coercive and (without loss of genreality) $f(0) = 0$. For any $C > 0$ define

$$R_C(v) = \sup\{r | f(rv) \leqslant C\}$$

i.e. $f(R_C(v)v) = C$ and $f(rv) \geqslant C$ for all $r > R_C(v)$. Now, suppose f is not coercive so that, for some $0 < C < \infty$, $\sup\limits_{v \in \partial B(0,1)} R_C(v) = \infty$. 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 $i \to \infty$. Use the convexity of $f(sv_i)$ in s for each v_i to conclude $f(sv_i) \leqslant C$ for $s \leqslant R_C(v_i)$. Use this, and the continuity of f, to prove that $f(sv^*) \leqslant C$ for all $s > 0$, implying that f is not directionally coercive, which is a contradition.

Exercise 13.3.25. Show that a coercive, strictly convex function has a unique minimizer x^* such that $f(x^*) < f(x)$ for all $x \neq x^*$.

Exercise 13.3.26. **(Challenge)** Suppose that $f : x \in \mathbb{R}^n \to y \in \mathbb{R}^1$ is convex, $f \in C^1(\mathbb{R}^n, \mathbb{R}^1)$ and f is minimal at x^*. Prove 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 $\nabla f \neq 0$.

Exercise 13.3.27. **(Challenge)** Even though Exercise 13.3.26 implies that gradient descent cannot converge unless we are converging to a minimizer, we are not guaranteed we are converging very fast. Create C^1, strictly convex f's with unique minimizers 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 any $T \leqslant \infty$. **Hint**: consider $f(x) = |x|^\alpha$.

Exercise 13.3.28. Define f^*, the Legendre-Fenchel transform of f, by

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

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,

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

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

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

Exercise 13.3.30. Prove that f^{**} is always convex even if f is not.

13.4 ... And the Three Integrals in Section 11.5, Again

We began with a very simple smooth function $f : \mathbb{R} \to \mathbb{R}$ (see Figure 81) which we probed with three integrals, the generalizations

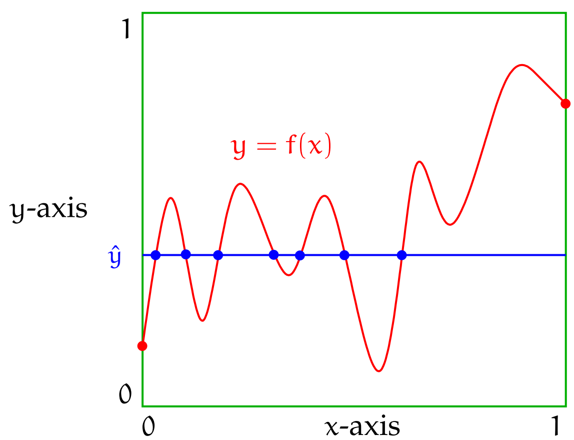


Figure 81: The level set $X_{\hat{y}}$ has 7 elements, shown as 7 blue dots in this figure. Same figure can be used to illustrate each of the three integrals.

of which turned out to be important tools for nonlinear, geometric analysis ...

Degree Theory

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

$\rightarrow$ special case of degree theory

$\rightarrow$ brought up Sard's Theorem for us.

Remark 13.4.1. *Again, the integral immediately above need only be over the set* $f([0,1])$ *but because* $f([0,1]) \subset [0,1]$, *integrating from* 0 *to* 1 *works.*

Area/Coarea

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

$\rightarrow$ special case of area and coarea formulas.

Stokes Theorem

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

$\rightarrow$ simple case of divergence theorem

$\rightarrow$ which is itself a simple case of Stokes Theorem.

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

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

$\operatorname{rank}(D_y f) = \min(n, m) \quad \forall \;\; y \in X_c$

$\rightarrow$ Sard's Theorem also comes up

$\rightarrow$ Which brings up the 5R covering theorem

$\rightarrow$ Which becomes a good place to begin looking at outer measures.

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

$$(B(y,\epsilon) \cap \{y + V_y\} \cap E) \sim (B(y,\epsilon) \cap X_c) \quad \forall \ y \in X_c$$

$\to$ Really the same idea as Derivative = linear approximation

$\to$ Introduces Manifolds.

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

$$\left(B(y,\epsilon) \cap \{y + D_y f^{-1}(0)\} \cap E\right) \sim (B(y,\epsilon) \cap X_c) \quad \forall \ y \in X_c$$

$\to$ Level sets corresponding to Regular values = manifolds.

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

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

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

... where the Jacobian $J^* f \equiv \begin{cases} \sqrt{Df^t \circ Df} & n < m \\ \sqrt{Df \circ Df^t} & n \geqslant m \end{cases}$

$\to$ a very powerful general tool for tracking and computing mapped volumes

$\to$ We encounter outer measures and Hausdorff measures in earnest here!

Stokes Theorem – Briefly

$$\begin{aligned} \int_{\partial\Omega} \omega \quad &= \quad \int_\Omega d\omega \text{ (Stokes Theorem)} \\ &\to \quad \int_{\partial\Omega} v \cdot \vec{n} \ d\sigma = \int_\Omega \nabla \cdot v \ dx \text{ (Divergence Theorem)} \\ &\to \quad \oint_{\partial\Omega} \vec{v} \cdot T_{\partial\Omega} = \int_\Omega \nabla \times \vec{v} \ dx \text{ (Little Stokes Theorem).} \end{aligned}$$

Exercise 13.4.1. Spend time thinking about the three integrals and the generalizations indicated in the comments on the integrals.

1 Find a presentation of Stokes theorem for differential forms and how it connects with the divergence theorem and the theorem about integrating the curl of a vector field being the same as integrating the vector field around the boundary (this is often also known as Stokes theorem). Both of these familiar theorems from vector calculus are actually special cases of the Stokes theorem for differential forms.
2 Read Chapter 3 of Frank Morgan's book *Geometric Measure Theory: a Beginner's Guide* [32] to get a feeling for the proofs of the area and coarea formulas, and then see Evans and Gariepy [12] for more detailed proofs if you are interested.

14 Derivatives, Again, Measure Theoretically

The first section reviews the classical definition and the second recalls derivatives as linear approximation as a launching point to a very geometric measure theoretical way of looking at derivatives. The remaining sections significantly expand our previous explorations using measure theoretic tools.

14.1 Secants and Derivatives

The derivative that is encountered for the first time in calculus is defined as the limit of a ratio of the "rise" over "run" of the graph of a function. For $y = f(x)$, this becomes

$$\frac{df}{dx}(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

This is visualized as the slope of the secant lines approaching a limit – the slope of the tangent line – as the free ends of those lines approach $(a, f(a))$. Figure 82 illustrates this.

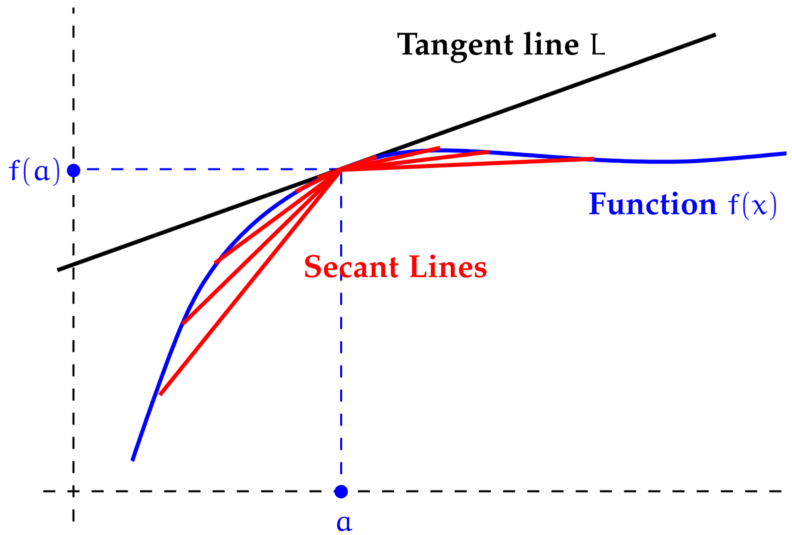


Figure 82: The traditional definition of the derivative.

14.2 The Derivative as an Approximation

The derivative as L_a, the optimal linear approximation to f at a, is another, very useful way to think about the derivative. Here, we focus on the fact that the tangent line at $(a, f(a))$ approximates the graph of $f(x)$ at $(a, f(a))$ as we zoom in on the graph. More precisely, writing $x = h + a$,

$$f(x) = f(h + a) = f(a) + L_a(h) + g(h)h,$$

where L_a is linear in h, $g(h) \to 0$ as $h \to 0$, and the tangent line L is the graph of the function $y = f(a) + L_a(x - a)$.

Exercise 14.2.1. Use the facts that (1) linear $L_a : \mathbb{R} \to \mathbb{R}$ have the form $h \to sh$, s a scalar, and (2) $g(h) \to 0$ as $h \to 0$, to rearrange this last equation for $f(x)$ into the original definition of a derivative.

Using the equation above to get

$$|f(x) - (f(a) + L_a(x - a))| \leqslant \left(\sup_{|s| \in [0,\epsilon]} |g(s)| \right) |h| \text{ for } h \in [-\epsilon, \epsilon],$$

we obtain the nice geometric interpretation illustrated in Figure 83. The

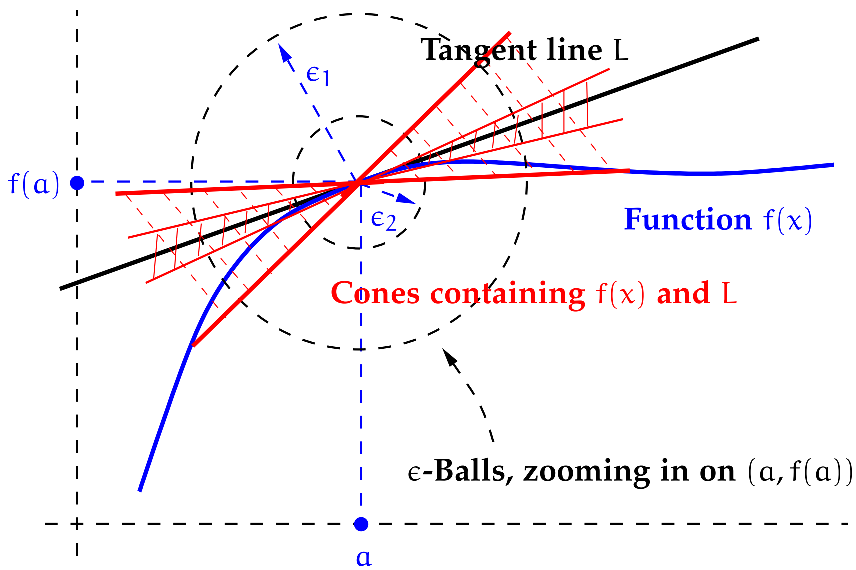


Figure 83: As we zoom into a point of differentiability, the graph is contained in cones that get thinner.

figure illustrates the fact that the graph of $f(x)$ lies in cones centered on L, whose angular widths go to zero as we restrict ourselves to smaller and smaller ϵ-balls centered on $(a, f(a))$. Inside the ϵ_1-ball, the graph stays in the wider cone, while in the smaller, ϵ_2-ball the graph stays in the narrower cone.

Let's restate this. Defining

- $p \equiv (a, f(a))$,
- $B(\epsilon)$ to be the ball of radius ϵ centered on p,
- $F \equiv \{(x, y) | y = f(x)\}$,
- $C_L(p, \epsilon)$ to be the smallest closed cone, symmetrically centered on L, with vertex at p such that $F \cap B(\epsilon) \subset C_L(p, \epsilon)$, and
- $\theta(\epsilon)$ to be the angular width of $C_L(p, \epsilon)$,

we have that f is differentiable at $a \Leftrightarrow \theta(\epsilon) \to 0$ as $\epsilon \to 0$. Figure 84 illustrates this.

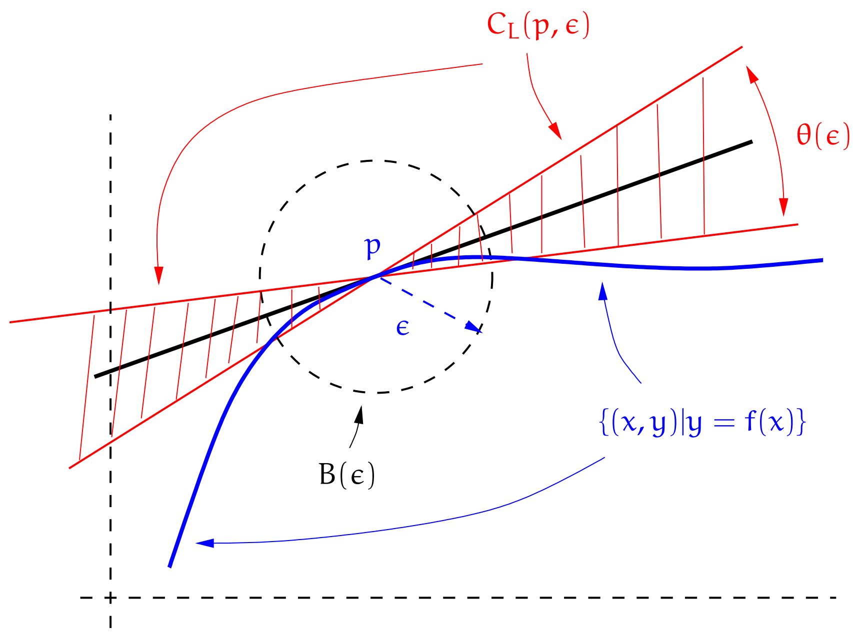


Figure 84: f is differentiable at $a \Leftrightarrow \theta(\epsilon) \to 0$ as $\epsilon \to 0$.

Exercise 14.2.2. Provide the missing details taking us from the above inequality bounding the deviation from linearity to the above statement that {f is differentiable at $a \Leftrightarrow \theta(\epsilon) \to 0$ as $\epsilon \to 0$} using the facts that (1) the above inequality defines cones that are almost symmetric about L and (2) the ϵ-ball centered at p is contained in the vertical strip $(x - a, y - f(a)) \in [-\epsilon, \epsilon] \times (-\infty, \infty)$.

With this shift to a geometric perspective, we are now in a position to take a step in the direction of geometric measure theory. Note that in our definition the cones contain all of the graph as they narrow down and we zoom in. What if instead of the cones converging to a line, we converge to two or three lines, or that we converge to a cone that does not narrow, or two cones that do not narrow? Then we are interested in the general tangent cone, a special case of which is the usual tangent line. Alternatively, what if all we know is that a larger and larger fraction of the graph is in a narrower and narrower cone as we zoom into p? That is precisely the idea that approximate tangent lines capture. We now turn to the first idea, the tangent cone.

14.3 Tangent Cones

The tangent line discussed above is also the tangent cone. The tangent cone of a set in $\mathbb{R}^n$ can have any dimension from 1 to n. For nicely behaved k-dimensional sets, the tangent cone will also be k-dimensional. In the case of the usual derivative of functions from $\mathbb{R}$ to $\mathbb{R}$, we are working in the graph space $\mathbb{R}^2$ with 1-dimensional sets. Moving to tangent cones, we can approximate one dimensional sets which are not graphs or, more generally, arbitrary subsets of $\mathbb{R}^n$.

We now build up to a definition of the tangent cone of $F \subset \mathbb{R}^n$ at p. Begin by translating F by $-p$. (This moves p to 0.) Define $F(\epsilon) \equiv (F \cap B(\epsilon)) \setminus p$. Use a projection center at 0 to project the translated $F(\epsilon)$ onto the sphere of radius 1. Take the closure of the resulting subset of the 1-sphere. Finally take the cone over this set. Call this set $T_p^\epsilon(F)$ (We will sometimes refer to this as the tangent cone at scale ϵ). Putting all this together,

$$T_p^\epsilon(F) = \{\mathbb{R} \geqslant 0\}(\text{Closure}(\cup_{x \in F(\epsilon)} \frac{x-p}{|x-p|}))$$

where the cone over E, $[\{\mathbb{R} \geqslant 0\}\ E]$, is the union of all possible non-negative scalings of all points in E, i.e. the union of rays from the origin through points in E. Now the tangent cone of F at p is the intersection of $T_p^\epsilon(F)$ at any sequence of ϵ_i's going to zero; $\epsilon_i = \frac{1}{i}$ will do. Thus the tangent cone of F at p, $T_p(F)$ is given by:

$$T_p(F) = \bigcap_i T_p^{\frac{1}{i}}(F).$$

Summarizing, we get:

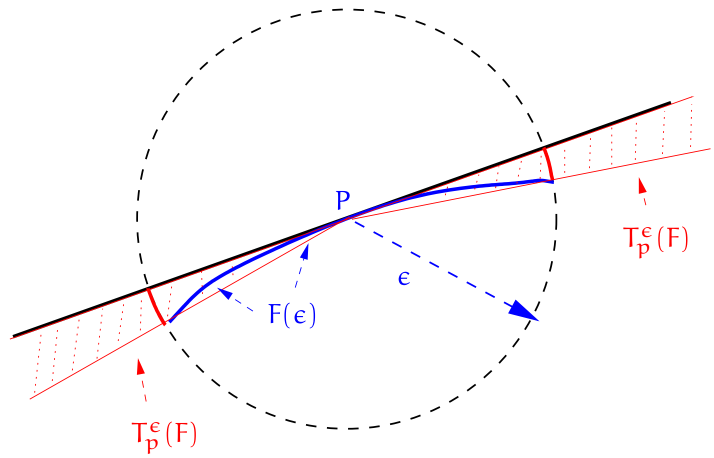


Figure 85: As we zoom in, the tangent cone at scale ϵ, $T_p^\epsilon(F)$, converges to the tangent line through p.

Definition 14.3.1 (Tangent Cone). *The tangent cone of* F *at* p *is given by*

$$T_p(F) = \bigcap_i T_p^{\frac{1}{i}}(F).$$

where $T_p^\epsilon(F)$ *is given by*

$$T_p^\epsilon(F) = \{\mathbb{R} \geqslant 0\}(\mathit{Closure}(\cup_{x \in F(\epsilon)} \frac{x-p}{|x-p|})).$$

In the case of a differentiable function $f : \mathbb{R} \to \mathbb{R}$, this tangent cone is the usual 1-dimensional tangent line. Figure 85 illustrates this.

Remark 14.3.1. *The tangent cone is centered on the origin,* 0*, but I will be plotting it as though it were centered on* p*. Similarly, the tangent lines will sometimes be thought of as linear subspaces (i.e. centered on the origin* 0*) and at other times, as the shift of that linear subspace to* p*.*

If the curve we are considering does not have a derivative at p, then we can get tangent cones that are not lines. Figure 86 illustrates an example of this. The function generating the cone in the figure keeps oscillating between the upper and lower lines as we zoom into p.

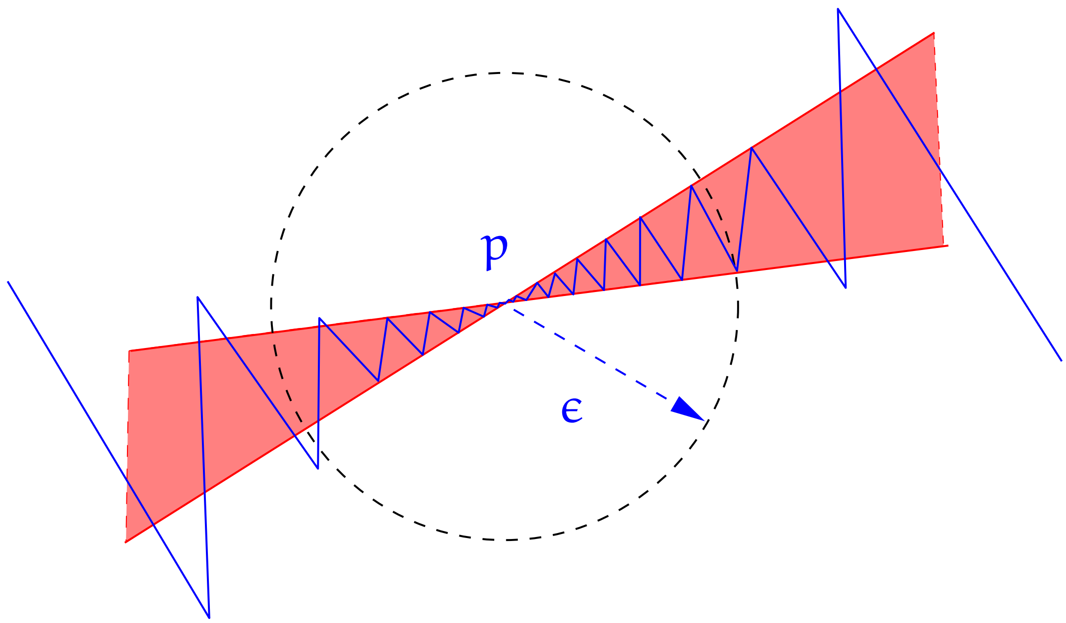


Figure 86: The tangent cone of the blue curve at the point p is shown in red and does not narrow down to a line as $\epsilon \to 0$.

Exercise 14.3.1. Construct an concrete example of a function that has a tangent cone like the tangent cone shown in Figure 86.

Exercise 14.3.2. Come up with examples of one dimensional sets $F \subset \mathbb{R}^2$ whose tangent cone at p, $T_p(F)$, are:

- Two lines passing through p,
- An infinite number of lines passing through p, yet **not** equaling all of $\mathbb{R}^2$
- one convex cone with vertex at p.
- a line through p and a convex cone with vertex at p, intersecting the line only at p.

Pick F's so that $F \neq T_p(F)$. Can you find four 0-dimensional F's yielding the same four tangent cones?

Exercise 14.3.3. Show that any $E \subset \mathbb{R}^2$ such that $\mathcal{H}^0(E) < \infty$ has empty tangent cones at every point in $\mathbb{R}^2$.

14.4 Approximate Tangent Cones

14.4.1 Densities

Now we need $\theta^k(F,p)$, the k-dimensional density of F at p. Let $\omega(k)$ be the volume of the unit ball in $\mathbb{R}^k$ when k is an integer, and something that interpolates sensibly otherwise. (See Definition 12.3.6 on page 285). Choose some measure μ. (Typically this will be k-dimensional Hausdorff measure, $\mathcal{H}^k$, restricted to some set, possibly with some weight function). Now, $\theta^k(F,p)$ is given by

$$\theta^k(F,p) = \lim_{\epsilon \to 0} \frac{\mu(F \cap B(\epsilon))}{\omega(k)\epsilon^k}$$

when this limit exists. When the limit does not exist, we work with the $\limsup$ and $\liminf$ of the right hand side which are called upper and lower densities of F at p and are denoted by $\theta^{*k}(F,p)$ and $\theta_*^k(F,p)$ respectively. (See Chapter 15 for a somewhat more detailed look at densities and their uses.)

14.4.2 Using Densities to get Approximate Tangents

We now define the **approximate tangent cone at p of F** to be the intersection of closed cones whose complements intersected with F have density zero at p,

Definition 14.4.1 (approximate tangent cone at p of F)**.** *The approximate tangent cone of* F *at* p, $\tilde{T}_p(F)$ *is given by*

$$\tilde{T}_p(F) = \bigcap \left\{ \textit{closed cones } C \textit{ with vertex } p \mid \theta^k((\mathbb{R}^n \setminus C) \cap F, p) = 0 \right\}.$$

Originally (in this section), we were aiming at having a definition of approximate tangent line that was invariant to (small) pieces of the set F outside the sequence of cones, provided those pieces got small enough, quick enough. Now we can make that more precise. We want a definition of approximate tangent line that ignores such excursions of F provided these excursions have density zero at p.

Rather anti-climatically then, here is the definition we have been waiting for (though you might have already guessed it!).

Definition 14.4.2 (**Approximate Tangent Line**)**.** *A 1-dimensional set has an approximate tangent line at* p *when the approximate tangent cone is equal to a line through* p*.*

When the 1-dimensional set is an embedded differentiable curve, the tangent line and the approximate tangent line are the same.

Remark 14.4.1. *In general, when we are dealing with* k*-dimensional sets in* $\mathbb{R}^n$*, we will get approximate tangent* k*-planes. That is because most things* *we deal with will be* rectifiable sets *having approximate tangent* k*-planes* $\mathcal{H}^k$ *almost everywhere. Rectifiable sets are introduced in Chapter 15.*

Exercise 14.4.1. Can you create examples of one dimensional sets which have a (density based) approximate tangent line at p but not the usual tangent line at p?

Exercise 14.4.2. Prove that a tangent line to a continuous curve is also the (density based) approximate tangent line at p.

14.5 Weak Tangents

There is different version of approximate tangent k-plane based on integration. We will call these tangents, *weak tangents*.

We start with the fact that we can integrate functions defined on $\mathbb{R}^n$ over k-dimensional sets using k-dimensional measures μ (typically $\mathcal{H}^k$). We zoom in on the point p, through dilation of the set F:

$$F_\rho(p) = \{x \in \mathbb{R}^n | \ x = \frac{y-p}{\rho} + \ p \text{ for some } y \in F\}.$$

Definition 14.5.1 (weak tangent k-plane)**.** *The k-dimensional subspace of* $\mathbb{R}^n$*,* L*, is the weak tangent k-plane of* F *at* p *if* $F_\rho(p)$ *converges weakly to* L*: i.e. if*

$$\int_{F_\rho} \phi \ d\mu \underset{\rho\to 0}{\to} \int_L \phi \ d\mu$$

for all continuously differentiable, compactly supported $\phi : \mathbb{R}^n \to \mathbb{R}$*.*

In Figure 87, we illustrate this for the case of 1-planes (i.e. lines) in the top illustration, L is the weak limit of the dilations of F, while in the bottom it is not.

Exercise 14.5.1. Can you create an example of a one dimensional curve in $\mathbb{R}^2$ which has the usual tangent line at p but does not have a weak

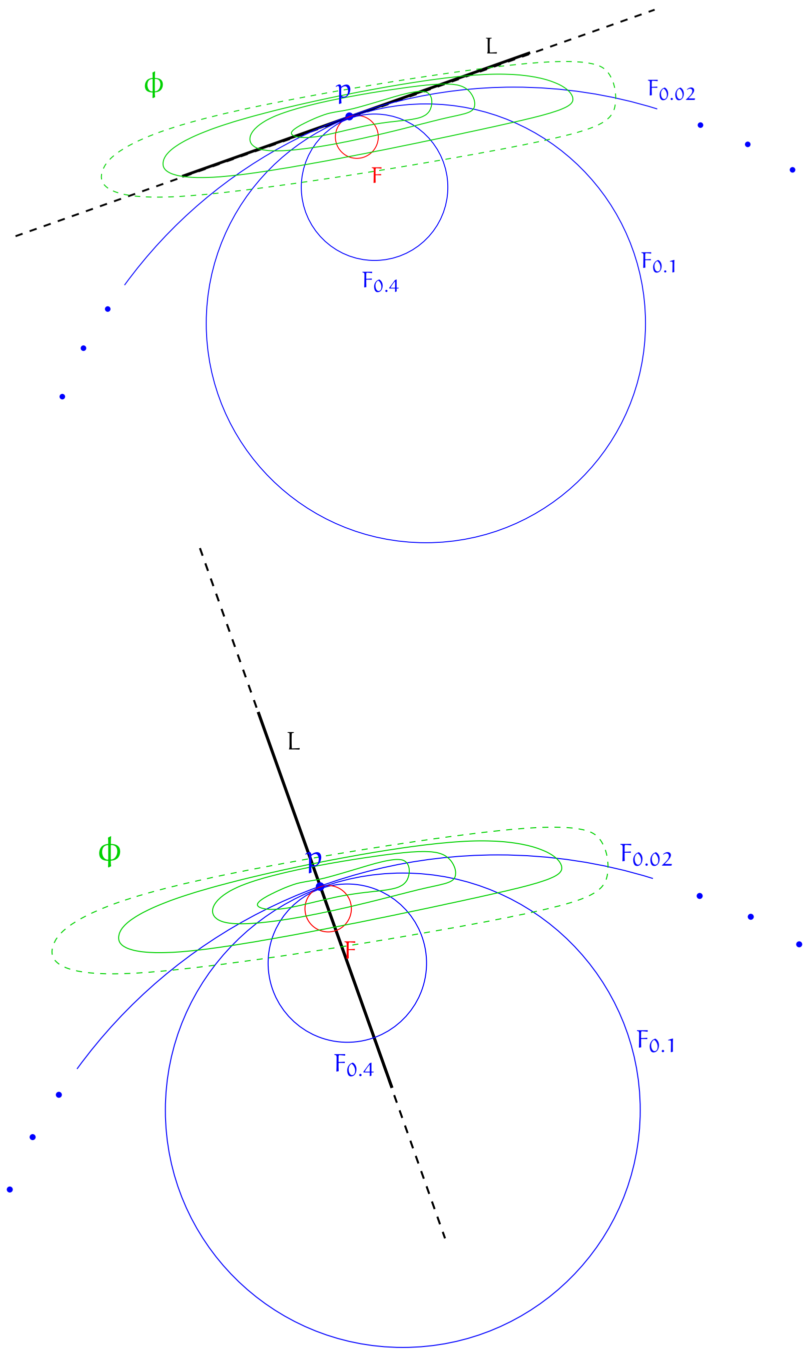


Figure 87: Illustration of a line that is (top) and a line that is not (bottom) the weak tangent. The solid green lines are the level sets of ϕ while the dashed green line indicates the boundary of the support of ϕ. Note also that the ρ's of 0.4, 0.1, and 0.02 are approximate.

tangent line at p? **Hint**: such examples exist and depend on the fact that $\mathcal{H}^1$ of a set that looks thin locally, can be as large as you like! You might start by restricting yourself to the region between $y = x^2$ and $y = -x^2$ and creating a curve in that region, focusing on the tangent structure at $(0, 0)$.

14.6 More Exercises

Exercise 14.6.1. Let E be the set of all points in the unit square $[0, 1] \times [0, 1]$ with rational coordinates.

- Find all possible tangent cones generated by E.
- Find all possible approximate tangent cones generated by E and the measure $\mu = \mathcal{L}^2$ (2-dimensional Lebesgue measure).

Note: Though the point p at which you compute $T_p(E)$ or $\tilde{T}_p(E)$ need not be in E for these tangent cones to be non-empty, if $p \notin \mathrm{clos}(E)$, then the tangent cone at p and the approximate tangent cone at p are both empty (why?) and so you need only consider points in $\mathrm{clos}(E)$, the closure of E.

Exercise 14.6.2. Suppose E is a spiral that spirals around an infinite number of times as it spirals into the origin in $\mathbb{R}^2$.

- What is the tangent cone at $p = (0, 0)$?
- What is the approximate tangent cone at $p = (0, 0)$ using the measure $\mu = \mathcal{L}^2$?
- What is the approximate tangent cone at $p = (0, 0)$ using the measure $\mu = \mathcal{H}^1$ (1-dimensional Hausdorff Measure)?
- What can you say about the tangent cones and approximate tangent cones at all other $p \neq (0, 0)$?

From Nonlinear to Nonsmooth: An invitation to Geometric Measure Theory 15

15.1 Lipschitz Functions and Rectifiable Sets

As might be surmised from a quick look at the table of contents, derivatives form a major thread in this book. In Chapter 14 we viewed derivatives as approximating sets obtained by zooming in on the set at some point. In this chapter, we continue the exploration of the many ways derivatives, in one form or another, are doors to all sorts of ideas in analysis and geometric analysis. We begin by looking at Lipschitz functions, and the sets that can be constructed using Lipschitz functions. Some of the proofs or outlines of proofs are included but several are not and for those, I refer the reader to Evans and Gariepy [12] and Morgan [32].

The last section of the chapter is an invitation to geometric measure theory that flows rather naturally from the first part of the chapter on non-smooth functions and sets.

15.1.1 Lipschitz Functions

Recall from Section 4.4 that we define a *Lipschitz mapping* $f : \mathbb{R}^m \to \mathbb{R}^n$ to be one that does not expand space too much: that is,

$$|f(x) - f(y)| \leqslant K\,|x - y|$$

where $K < \infty$ is some constant independent of x and y. Graphs of Lipschitz functions can have corners, but not cusps. Figure 88 illustrates a Lipschitz function with a Lipschitz constant of 1.

Lipschitz functions seem like they can have a lot less regularity than C^1 functions. And, in fact, one can define a Lipschitz mapping g :

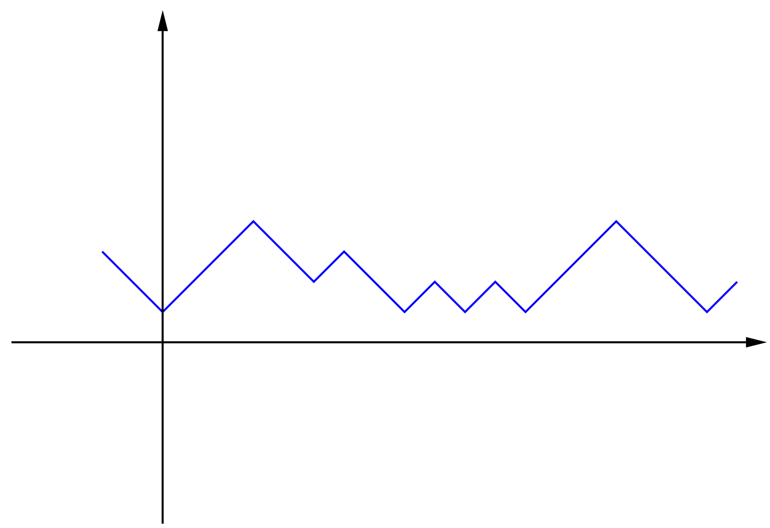


Figure 88: A function satisfying the Lipschitz condition for any $K \geqslant 1$.

$[0,1] \subset \mathbb{R} \to [0,1]$ that does not have a derivative on a set which is dense in $[0,1]$ – for example $[0,1] \cap \mathbb{Q}$. See exercise 15.1.1.

Exercise 15.1.1. Define $g_n(x) = \int_0^x h_n(t)\, dt$ where

$$h_n(t) \equiv \begin{cases} 1 & \text{for } x \in \left(\frac{2k}{2 \cdot 3^n}, \frac{2k+1}{2 \cdot 3^n}\right] \\ -1 & \text{for } x \in \left(\frac{2k+1}{2 \cdot 3^n}, \frac{2k+2}{2 \cdot 3^n}\right] \end{cases} \quad k = 0, \ldots, 3^n - 1.$$

1 Show that

$$f(x) \equiv \sum_{i=1}^{\infty} 2^{-i} g_i(x)$$

(a) is Lipschitz with Lipschitz constant 1 and (b) has no derivative at each point in the dense set $\cup_{n=1}^{\infty} \{x \mid g_n \text{ is not differentiable at } x\}$

2 **(Challenge)** Now play with functions of the form

$$f_{\hat{\alpha}}(x) \equiv \sum_{i=1}^{\infty} \alpha_i g_i(x)$$

with

$$\hat{\alpha} = (\alpha_1, \alpha_2, \ldots) \text{ and } \alpha_i \geqslant 0 \ \forall i$$

to get functions whose points of non-differentiability are dense in $[0, 1]$. **Hint**: show that $\hat{\alpha} = (\alpha_1, \alpha_2, \ldots)$ yields a Lipschitz function whose points of non-differentiability are dense in $[0, 1]$ if and only if $\alpha_i \neq 0$ for an infinite number of i's and $\sum_{i=1}^{\infty} \alpha_i < \infty$. (Keep in mind that we assume the α_i's are all non-negative.)

But, we also have the following theorem that says Lipschitz functions are differentiable everywhere except a set whose Lebesgue measure is zero. Reminding ourselves of the definition of *almost everywhere*,

Definition 15.1.1 (**Almost Everywhere**). *A property is said to hold almost everywhere or μ-almost everywhere if the property fails on a subset $E \subset \mathbb{R}^n$ having measure zero: $\mu(E) = 0$,*

we are prepared for Rademacher's theorem.

Theorem 15.1.1 (Rademacher). *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz. Then f is differentiable $\mathcal{L}^n$ almost everywhere in $\mathbb{R}^n$.*

Proof. See Chapter 3 of Evans and Gariepy [12]. □

15.1.1.1 *Exercises on Lipschitz Functions*

Definition 15.1.2 (**Approximate Continuity**). *We say that f is approximately continuous at x, with respect to the measure μ, if, for any $\epsilon > 0$,*

$$\lim_{r \to 0} \frac{\mu(\{y \mid |f(y) - f(x)| > \epsilon\})}{\mu(B(x, r))} = 0.$$

Exercise 15.1.2. **(Challenge)** Suppose that the absolute value of $f : [0, 1] \to \mathbb{R}$ is bounded by C (i.e. $|f| \leqslant C$) and Lebesgue measurable. Show that $g(x) \equiv \int_0^x f(t)dt$ is Lipschitz with Lipschitz constant C. Using the fact that measurable functions are approximately continuous, show that g is differentiable almost everywhere. (The point of this exercise is to prove Rademacher for a restricted set of functions. The more general proof is more involved of course.)**Hint**: prove it first for the case in which f is continuous.

Theorem 15.1.2 (C^1 Approximation)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz, μ is n-dimensional Lebesgue measure on $\mathbb{R}^n$ and $\epsilon > 0$. Then there is a C^1 function $g : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\mu\{x|f(x) \neq g(x) \text{ or } Df(x) \neq Dg(x)\} < \epsilon.$$

We also have that $|Dg(x)| \leqslant C(n)\mathrm{Lip}(f)$ for all $x \in \mathbb{R}^n$, where $\mathrm{Lip}(f)$ is the Lipschitz constant of f and $C(n)$ is a constant that depends only on n.

See Chapter 6 of [12] for a proof.

Exercise 15.1.3. Suppose that $F = f(E)$ where f is Lipschitz, $f : E \subset \mathbb{R}^k \to \mathbb{R}^n$, and $k \leqslant n$. Use Theorem 15.1.2 to show that F can be expressed as the union of a set E_0 with k-dim Hausdorff measure zero plus a countable union of pieces of images of C^1 maps $g_i : \mathbb{R}^k \to \mathbb{R}^n$, $E_i \subset g_i(\mathbb{R}^k)$: $F = \{\bigcup_i E_i\} \cup E_0$.

Exercise 15.1.4. Think about the following questions.

1 We already know from exercise 15.1.1 that a Lipschitz function can have an infinite number of corners. Can it have an uncountably infinite set of non-differentiable points?
2 What can the graph of a Lipschitz function look like? What can image of a Lipschitz function look like? (What is the difference?)

Exercise 15.1.5. Suppose that f is Lipschitz with Lipschitz constant K (i.e. $|f(x) - f(x)| \leqslant K|x - y|$). Prove that $\mathcal{H}^s(f(E)) \leqslant K^s\mathcal{H}^s(E)$. Use this to show that orthogonal projections P in $\mathbb{R}^n$ cannot increase the measure of a set. I.e. $\mathcal{H}^s(P(E)) \leqslant \mathcal{H}^s(E)$. Recall that by choosing coordinates to your advantage, you can always think of an orthogonal projection to be the map taking $(x_1, x_2, ..., x_k, x_{k+1}, ..., x_n)$ to $(x_1, x_2, ..., x_k, 0, ..., 0)$ for some k, that is it is equivalent to setting the last $n-k$ coordinates to zero.

15.1.2 *Rectifiable Sets*

Now we are ready to define rectifiable sets which are capable of a lot of craziness (for example, all the sets in the figure in Section 4.1 are rectifiable), and yet are also quite nice, locally, measure-theoretically. In fact, they are nice enough that you can do analysis on them as though they were smooth, though this often requires nuanced reasoning.

15.1.2.1 *Definitions*

Definition 15.1.3 (Rectifiable). *A set $E \subset \mathbb{R}^n$ is k-rectifiable if there are Lipschitz $\{f_i\}_{i=1}^{\infty}$, $f_i : \mathbb{R}^k \to \mathbb{R}^n$ such that*

$$E \subset \{\bigcup_i f_i(\mathbb{R}^k)\} \cup N_0.$$

where $N_0 \subset \mathbb{R}^n$ (called the null set) satisfies $\mathcal{H}^k(N_0) = 0$.

Here are two theorems that yield alternative, equivalent definitions.

Theorem 15.1.3 (Rectifiable sets as subsets of C^1 submanifolds)**.** E *is a* k*-rectifiable subset of* $\mathbb{R}^n$ *if and only if it can be represented as a countable union of pieces of embedded* C^1 k*-dimensional submanifolds* $M_i \subset \mathbb{R}^n$ *plus a set with k-dimensional measure* 0*. More succinctly,*

$$\mathcal{H}^k\left(E \setminus \bigcup_i M_i\right) = 0.$$

Exercise 15.1.6. Show that if, in Theorem 15.1.3 we define $E_i \equiv M_i \cap E$ we can choose $\hat{E}_i$ such that (1) $\hat{E}_i \subset E_i$ for all i, (2) $\hat{E}_i \cap \hat{E}_j = \emptyset$ whenever $i \neq j$, i.e.they are pairwise disjoint, and (3) $\mathcal{H}^k(E \setminus \cup_i \hat{E}_i)$. **Hint**: all you need is the statement of the theorem and a bit of set manipulation to prove this. Note that some of the $\hat{E}_i$ may be empty if there are superfluous M_i.

Definition 15.1.4 (**Lipschitz Graph**). *A set $E \subset \mathbb{R}^n$ is k-dimensional Lipschitz Graph in $\mathbb{R}^n$ if it is a subset of a (possibly translated and rotated) graph of some Lipschitz $f : \mathbb{R}^k \to \mathbb{R}^{n-r}$.*

Theorem 15.1.4 (Rectifiable sets are unions of Lipschitz graphs). *E is a k-rectifiable subset of $\mathbb{R}^n$ if and only if almost all of E is contained in a countable union of k-dimensional Lipschitz graphs $\{W_i\}_{i=1}^{\infty}$,*

$$\mathcal{H}^k\left(E \setminus \bigcup_i W_i\right) = 0.$$

Remark 15.1.1 (Note on Notation). *We differ a bit in our terminology with Federer – see Section 3.2.14 of Federer [13].*

Remark 15.1.2 (Different Types of Rectifiability). *We are often interested in k-rectifiable sets E which have finite $\mathcal{H}^k$ measure – i.e. $\mathcal{H}^k(E) < \infty$. Federer calls such sets $(\mathcal{H}^k, k)$-rectifiable sets.*

I encourage you to explore examples of rectifiable sets in the following exercises.

15.1.2.2 *Exercises on Rectifiability*

Exercise 15.1.7. Let E be a circle in $\mathbb{R}^2$. Show that it is $(\mathcal{H}^1, 1)$-rectifiable by explicitly constructing Lipschitz functions from $\mathbb{R}^1$ into $\mathbb{R}^2$ that cover E. Can you do this with one mapping?

Exercise 15.1.8. Let E be a union of some countably infinite collection of lines and line segments in $\mathbb{R}^2$. Show that E 1-rectifiable. What do you need for E to be $(\mathcal{H}^1, 1)$-rectifiable?

Exercise 15.1.9. Show that a countable union of k-rectifiable subsets of $\mathbb{R}^n$ is again a k-rectifiable subset of $\mathbb{R}^n$.

Exercise 15.1.10. Show that $E \equiv [0,1]^2 \cap \mathbb{Q}^2$ is $(\mathcal{H}^1, 1)$-rectifiable but not $(\mathcal{H}^0, 0)$-rectifiable.

Exercise 15.1.11. Define $E \subset \mathbb{R}^3$ to be the union of D_i, where D_i is horizontal disk with radius $\frac{1}{i}$, centered at $(i, 0, 0)$, and $i \in \mathbb{N} \cap [2, \infty)$. Show that E is $(\mathcal{H}^2, 2)$-rectifiable but ∂E is only 1-rectifiable, not $(\mathcal{H}^1, 1)$-rectifiable. Can you choose a sequence of radii for the disks so that ∂E is now $(\mathcal{H}^1, 1)$-rectifiable?

Remark 15.1.3. *Quite remarkably, one of the properties of* k*-rectifiable sets* E *such that* $\mathcal{H}^k(E) < \infty$*, is that at* $\mathcal{H}^k$ *almost all points in* E*, the approximate tangent cone is a plane. In other words, if you ignore a set with density* 0 *at* p*, then there is a tangent plane at* p*.*

In the next section, we look at a surprising theorem about measuring k-rectifiable sets. We take a very close look at the case of 1-rectifiable sets in $\mathbb{R}^2$.

15.2 Crofton's Magic Formula: Measuring Sets with Projections

Crofton's formula tells us that we can find the k-dimensional measure of sets in $\mathbb{R}^n$ by integrating over k-dimensional projections. To say what this means, we need to be more detailed. We will give the details for 1-dimensional curves in $\mathbb{R}^2$ first, including a very careful proof that does not use big machinery, but rather uses geometric insights and some work(!) to reach the goal.

First a little terminology: $l(\theta)$ will be the line through the origin in $\mathbb{R}^2$, whose angle with the positive horizontal axis is $\theta \in [0, \pi)$. We will use $r \in (-\infty, \infty)$ to refer to the coordinate along this line. The line perpendicular to $l(\theta)$, through the point $r \in l(\theta)$, will be denoted by $p_\theta(r)$. Figure 89 illuminates the terminology.

Now suppose we have a 1-dimensional set E. Note that $\mathcal{H}^0(p_\theta(r) \cap E)$ simply counts the number of intersections between E and $p_\theta(r)$. Then *Crofton's formula* is given by:

Theorem 15.2.1 (Crofton's Formula: 1-dim in $\mathbb{R}^2$).

$$\mathcal{H}^1(E) = C(1,2) \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap E) dr \, d\theta$$

where the constant $C(1,2)$ does not depend on E.

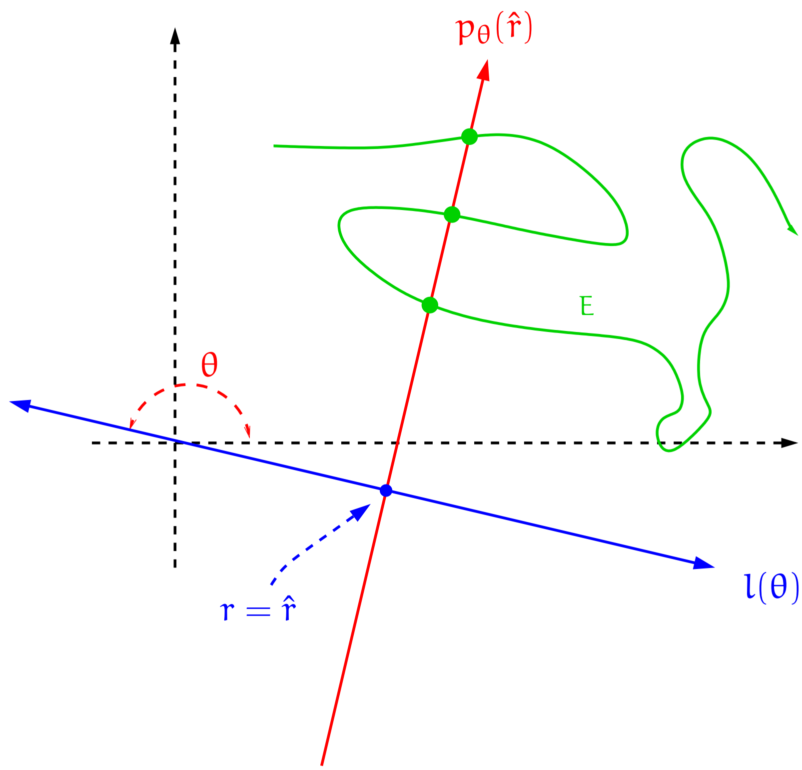


Figure 89: Illustration of E, $l(\theta)$, and $p_\theta(r)$. In this example, $\mathcal{H}^0(p_\theta(\hat{r}) \cap E) = 3$.

Question: For what 1-dimensional sets does this hold? **Answer:** rectifiable sets, which as we have seen can be pretty wild. The keys to the proof are the facts that (1) proving this for polygonal curves is easy,

(2) rectifiable sets are unions of pieces of C^1 curves and are therefore close to polygonal curves, and (3) Hausdorff measure behaves nicely under projections. The details are important here! We go through this proof in great detail in the next section.

What about sets in higher dimensions? The formula is completely analogous. Again, we begin with some terminology. Let $G(k, n)$ denote the set of all k-dimensional linear subspaces of $\mathbb{R}^n$. This is a compact manifold and we can put a measure on it, invariant to all rotations and reflections. It is called the Haar measure. $H_{G(k,n)}$ will denote the normalized Haar measure on $G(k, n)$). We will denote the orthogonal space to V by $V^\perp$, and the orthogonal space through $r \in V$ by $V_r^\perp \equiv V^\perp + r$. Figure 90 illustrates the case of $(k, n) = (1, 3)$.

Remark 15.2.1. *In the case of* $G(1, 2)$ *– the set of all lines through the origin in* $\mathbb{R}^2$ *– the measure is* $\frac{1}{\pi}\mathcal{H}^1 \llcorner C^+$*, where* C^+ *is the half unit circle parameterized by the angles* $[0, \pi)$*. Of course in this case each* V *is a line* $l(\theta)$ *for some particular* θ*,* $V^\perp$ *is the line through the origin orthogonal to* $l(\theta)$*, and* $V_r^\perp = V^\perp + r = p_\theta(r)$*. Note: in the* 1*-dimensional version of Crofton appearing in Theorem 15.2.1 we use* $\mathcal{H}^1 \llcorner C^+$*, not the normalized version* $\frac{1}{\pi}\mathcal{H}^1 \llcorner C^+$*, and as a result the* $C(1, 2)$ *we get is* $\frac{1}{2}$ *instead of* $\frac{\pi}{2}$*.*

Now we can state *Crofton's formula* for k-dimensional sets in $\mathbb{R}^n$.

Theorem 15.2.2 (Crofton's Formula: k-dim in $\mathbb{R}^n$). *Suppose that* E *is a k-dimensional rectifiable set. Then*

$$\mathcal{H}^k(E) = C(k, n) \int_{V \in G(k,n)} \int_{r \in V} \mathcal{H}^0(V_r^\perp \cap E) \, d\mathcal{H}^k(r) \, dH_{G(k,n)}(V)$$

where the constant $C(k, n)$ *does not depend on* E*.*

Exercise 15.2.1. Draw a figure like Figure 90 illustrating the case $(k, n) = (2, 3)$.

Exercise 15.2.2. Compute the value of the RHS of Crofton's formula in Theorem 15.2.1 for the case of E = the unit circle in $\mathbb{R}^2$, ignoring the

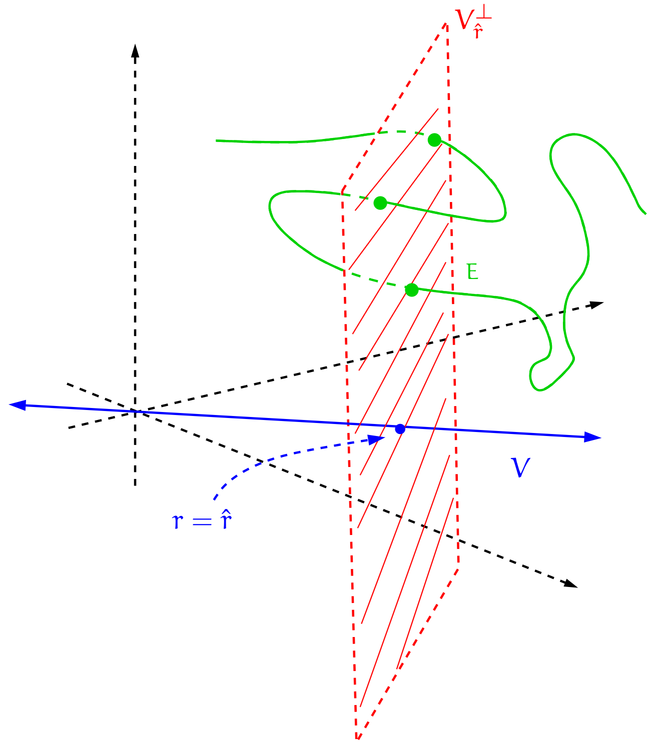


Figure 90: Illustration of E, V, and $V_{\hat{r}}^{\perp}$. In this example, $\mathcal{H}^0(V_{\hat{r}}^{\perp} \cap E) = 3$.

value of $C(1,2)$. Use the result you get to compute $C(1,2)$. Remember, this constant does not depend on E.

Exercise 15.2.3. Compute the value of Crofton's formula, using the $C(1,2)$ found in exercise 15.2.2, for the case in which E is a line segment.

Exercise 15.2.4. Use the results in the previous exercise to prove that Crofton works for polygonal curves. Why does this not actually prove the theorem for C^1 curves, even though it seems like it should? In other words, what details are there that make that leap bigger than it first appears? (This is precisely what we will be working through in the next (long) section, but before you look at the answers there, give it a try yourself.

Now we move to a careful proof of Crofton's formula for curves in $\mathbb{R}^2$.

15.2.1 Crofton's Formula in $\mathbb{R}^2$

Proof of Theorem 15.2.1: We prove this in steps. First two lemmas.

Lemma 15.2.1. *If $f : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz with Lipschitz constant L, then*

$$\mathcal{H}^k(E) \leqslant L^k \mathcal{H}^k(f(E))$$

for all $E \subset \mathbb{R}^m$.

Proof. See exercise 15.1.5

Lemma 15.2.2. *Suppose that E is an open subset of $\mathbb{R}^1$. Then there is a countable family of closed intervals F_i with mutually disjoint interiors, satisfying $\bigcup_i F_i = E$.*

Proof. We prove it in small steps:

(I) We will use dyadic intervals $D_{k,n} = [\frac{k-1}{2^n}, \frac{k}{2^n}]$ for $k \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$. Call this collection of closed intervals **Dy**. Note that **Dy** is countable.
(II) For every point $e \in E$, we may chose $D \in$ **Dy** such that $e \in D$ and $D \subset E$. Call the set of D's so obtained $\mathcal{D}$. It is a countable set, so we can enumerate the elements of $\mathcal{D}$, D_i. Note that $\bigcup_{D_i \in \mathcal{D}} D_i = E$.
(III) Note that if $D_1, D_2 \in \mathcal{D}$ then either
a: $D_1 \cap D_2 = \emptyset$ or
b: $D_1 \cap D_2 =$ a single endpoint or
c: $D_1 \cap D_2 =$ either D_1 or D_2.
(IV) Now we go through $\mathcal{D}$ and throw away all the D_i's which are contained in other D_j's. This takes a little bit of care but it can be done. (See exercise 15.2.5.) We now have a collection of closed intervals whose interiors are disjoint and whose union is E.

□

Remark 15.2.2. *It is easy to show that Lemma 15.2.2 holds for $E \in \mathbb{R}^2$ homeomorphic to a circle, but we will not need this in our proof.*

Step 1: For almost every θ, $\int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr \leqslant \mathcal{H}^1(E)$.

Proof.

1 We assume that E is an open subset of a C^1 1-dimensional embedded submanifold of $\mathbb{R}^2$ and that it is homeomorphic to an open subset of $\mathbb{R}^1$. We will call this an R^1*-submanifold* of $\mathbb{R}^2$.
2 Let $E_p(\theta) \equiv \{e \in E$ such that $T_e E$ is parallel to $p_\theta(r).\}$. Define $E_t(\theta) \equiv E \setminus E_p(\theta) = \{e \in E$ such that $T_e E$ is not parallel to $p_\theta(r).\}$. Since $\mathcal{H}^1(E) < \infty$, there are at most a countable number of angles θ_i such that $\mathcal{H}^1(E_p(\theta_i)) > 0$. Since we are integrating over θ we can ignore those θ's. We will therefore assume $\mathcal{H}^1(E_p(\theta)) = 0$. Notice that $E_p(\theta)$ is closed.

3 Now we partition E_t into a countably (possibly infinite) family of closed arcs $E_{t,i}$, meeting other arcs only at end points, such that $\mathcal{H}^0(p_\theta(r) \cap E_{t,i}) =$ either 0 or 1 for all r. We know we can make such a partition by Lemma 15.2.2, the fact that E_t is the homeomorphic image of an open set in $\mathbb{R}^1$, and the fact that E is a subset of a C^1 submanifold.

4 Then we can decompose $\mathcal{H}^0(p_\theta(r) \cap E)$ into a sum $\sum_i \mathcal{H}^0(p_\theta(r) \cap E_{t,i}) + \mathcal{H}^0(p_\theta(r) \cap E_p)$.

5 Denote the orthogonal projection of a set $F \in \mathbb{R}^2$ onto $l(\theta)$ by $P_\theta(F)$. Integrating $\mathcal{H}^0(p_\theta(r) \cap E)$ over $r \in l(\theta)$, we get

$$
\begin{aligned}
\int \mathcal{H}^0(p_\theta(r) \cap E)\, dr &= \int \sum_i \mathcal{H}^0(p_\theta(r) \cap E_{t,i}) dr + \int \mathcal{H}^0(p_\theta(r) \cap E_p) dr \\
&= \sum_i \int \mathcal{H}^0(p_\theta(r) \cap E_{t,i}) dr + \int \mathcal{H}^0(p_\theta(r) \cap E_p) dr \\
&\quad \text{(by the monotone convergence theorem)} \\
&= \sum_i \mathcal{H}^1(P_\theta(E_{t,i})) + \int \mathcal{H}^0(p_\theta(r) \cap E_p) dr \\
&\leqslant \sum_i \mathcal{H}^1(E_{t,i}) + \int \mathcal{H}^0(p_\theta(r) \cap E_p) dr \\
&\quad \text{(implied by } P_\theta \text{ being Lipschitz and Lemma 15.2.1)} \\
&= \sum_i \mathcal{H}^1(E_{t,i}) + \int_{P_\theta(E_p)} \mathcal{H}^0(p_\theta(r) \cap E_p) dr \\
&\leqslant \sum_i \mathcal{H}^1(E_{t,i}) + 0 \\
&\quad \text{(because } \mathcal{H}^1(P_\theta(E_p)) = 0 \text{ implies } \int_{P_\theta(E_p)} \mathcal{H}^0(p_\theta(r) \cap E_p) dr = 0) \\
&\leqslant \mathcal{H}^1(E).
\end{aligned}
$$

□

Remark 15.2.3. ***A note about thinking about proofs versus writing proofs down:*** *when I came up with the proof above, the visualization of it was very simple, yet looking at this string of equations above, I struggle to find that simplicity. And that is because (1) visually/geometrically the ideas are easy to represent and manipulate and (2) visually/geometrically, we can organize pieces of our thoughts into patterns that are far from sequential, yet in writing them down this is very difficult. Nevertheless, I will try to give a sense for how I was actually thinking about it here:*

- *If I could grab pieces of E that were small enough to project to 0 or 1 and miss only a measure zero piece, then I would use the Lipschitz mapping result which says that pieces project to pieces whose measure is not larger.*
- *Now sum everything up and you get that the integral of the projection function has to be dominated by the length of the thing you started with.*
- *To get that the problem points – those whose tangent is parallel to the projection direction – have measure zero, we exploit the fact that this can fail (I.e. $\mathcal{H}^1(E_p) > 0$) only at a countable number of θ's.*
- *And of course, that means we can ignore those θ's because we will be integrating over the θ's in the end.*

That is really all that the above array of equations contains!

Exercise 15.2.5. Show that the collection of intervals $\mathcal{D}$ generated in the proof of Lemma 15.2.2 can be culled as called for in the last step of the proof.

Exercise 15.2.6. Show that Lemma 15.2.2 holds for E homeomorphic to the circle.

Step 2: We can partition any R^1-submanifold into closed arcs $\bar{E}_i^\epsilon$ with disjoint interiors and the property that all the tangent directions in any one interval $\bar{E}_j^\epsilon$ lie in an arc of the unit circle of width

ϵ. That is, identifying tangent directions with $\theta \in S^1$, we can choose the $\bar{E}_i^\epsilon$ such that $|T_{e_1}E - T_{e_2}E| \leqslant \epsilon$ for any $e_1, e_2 \in \bar{E}_i^\epsilon$, for all i.

Proof. Use the fact that E is C^1 and homeomorphic to an open subset of $\mathbb{R}^1$, and then use the idea in the proof of Lemma 15.2.2 □

Prep for Step 3: Now define the associated polygonal curve P^ϵ to the partition in Crofton step 2 to be the polygonal curve formed on the endpoints of the closed arcs $\bar{E}_i^\epsilon$, with connectivity determined by the connectivity of E. Define the piece of P^ϵ sharing endpoints with $\bar{E}_i^\epsilon$ to be P_i^ϵ. (Note that this polygonal curve can have an infinite number of sides.)

Step 3:

$$\int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap \bar{E}_i^\epsilon) dr\, d\theta - \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap P_i^\epsilon) dr\, d\theta \leqslant 2\epsilon \mathcal{H}^1(\bar{E}_i^\epsilon).$$

Proof.

1 Note first that when the angle between P_θ and P_i^ϵ is greater than ϵ, $\mathcal{H}^0(p_\theta(r) \cap P_i^\epsilon) = \mathcal{H}^0(p_\theta(r) \cap \bar{E}_i^\epsilon)$ for all r. This is because both P_i^ϵ and $\bar{E}_i^\epsilon$ are Lipschitz graphs over $l(\theta)$ and they share endpoints.
2 When the angle between P_θ and P_i^ϵ is less than ϵ we have
   a $\int \mathcal{H}^0(p_\theta(r) \cap P_i^\epsilon)\ dr \geqslant 0$
   b $\int \mathcal{H}^0(p_\theta(r) \cap \bar{E}_i^\epsilon)\ dr \leqslant \mathcal{H}^1(\bar{E}_i^\epsilon)$ (from Step 1)
   c we get that $\int_0^\pi \int_{-\infty}^\infty \left|\mathcal{H}^0(p_\theta(r) \cap \bar{E}_i^\epsilon) - \mathcal{H}^0(p_\theta(r) \cap P_i^\epsilon)\right| dr\, d\theta \leqslant 2\epsilon \mathcal{H}^1(\bar{E}_i^\epsilon)$.
3 Putting 1 and 2 together, we get the desired result.

□

Step 4: For an R^1-submanifold, with partition $\{\bar{E}_i^\epsilon\}_i$ and associated polygonal curve P^ϵ we can therefore get that

$$\int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta - \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap P^\epsilon) dr\, d\theta \leqslant 2\epsilon \mathcal{H}^1(E)$$

.

Proof. This is immediate from Step 3. □

Step 5: Crofton holds for (possibly infinite, possibly disconnected) polygonal curves.

Proof. Observe that for any line segment L,

$$\begin{aligned} \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap L) dr\, d\theta &= \int_\alpha^{\alpha+\pi} |\sin(\theta)| \mathcal{H}^1(L) d\theta \text{ (for some } \alpha) \\ &= \mathcal{H}^1(L) \int_0^\pi \sin(\theta) d\theta \\ &= 2\mathcal{H}^1(L) \end{aligned}$$

so that $C(1,2)$ for line segments is just $\frac{1}{2}$ and we can conclude that

$$\mathcal{H}^1(L) = \frac{1}{2} \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap L) dr\, d\theta.$$

Since a polygonal curve is just a bunch of line segments, we get that

$$\mathcal{H}^1(P) = \frac{1}{2} \int_0^\pi \int_{-\infty}^\infty \mathcal{H}^0(p_\theta(r) \cap P) dr\, d\theta$$

for any polygonal curve P. □

Remark 15.2.4. *This step is just Exercises 15.2.2-15.2.4*

Step 6: Crofton's formula holds for R^1-submanifolds.

Proof. As $\epsilon \to 0$, $\mathcal{H}^1(P^\epsilon) \to \mathcal{H}^1(E)$. This, combined with Steps 4 and 5, gives us our result. □

Exercise 15.2.7. Convince yourself that indeed, as Step 6 claims, $\epsilon \to 0$, $\mathcal{H}^1(P^\epsilon) \to \mathcal{H}^1(E)$.

Step 7: Crofton's Formula holds for any piece of an embedded C^1-submanifolds

1. Let $E \subset M$ where M is a 1-dimensional embedded C^1 submanifold of $\mathbb{R}^2$.
2. Since $\mathcal{H}^1 \llcorner M$ is Radon, any piece E of M can be approximated arbitrarily well with O, an open subset M. I.e. $\mathcal{H}^1(O \setminus E) \leqslant \epsilon$.
3. Now we partition each O into a countable collection of homeomorphic images of closed intervals of $\mathbb{R}$ whose images in M intersect only at the endpoints. The images of interiors, which we denote O_i, are a disjoint countable collection of $\mathbb{R}^1$-submanifolds and $O = (\cup_i O_i) \cup N$, where N has $\mathcal{H}^1$ measure 0 (actually, N is at most countably infinite).
4. The sets $F_i \equiv O_i \setminus E$ are contained in the U_i which are open subsets of O_i (and are therefore $\mathbb{R}^1$-submanifolds) with the property that $\sum_i \mathcal{H}^1(U_i) \leqslant 2\epsilon$.
5. We use **Step 1** to get

$$\begin{aligned}
\sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap F_i)dr &\leqslant \sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap U_i)dr \\
&\leqslant \sum_i \mathcal{H}^1(U_i) \\
&\leqslant 2\epsilon^i.
\end{aligned}$$

6. Define $K = N \cap E$. We note that

$$E = \left(\bigcup_i O_i \cap E \right) \cup K$$

and

$$\begin{aligned} O &= \left(\bigcup_i O_i\right) \cup N \\ &= \left(\bigcup_i F_i\right) \cup \left(\bigcup_i O_i \cap E\right) \cup N \end{aligned}$$

and because $E \subset O$ we have that

$$\int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E)dr \leqslant \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O)dr.$$

7 Putting all this together we have, except for a measure zero set of θ,

$$\begin{aligned} \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E)dr \\ &\leqslant \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O)dr \\ &= \sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap F_i)dr \\ &\quad + \sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O_i \cap E)dr \\ &\quad + \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap N)dr \\ &= \sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap F_i)dr \\ &\quad + \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E)dr \\ &\quad + 0 \text{ (See Remark 15.2.5)} \\ &= \sum_i \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap U_i)dr \\ &\quad + \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E)dr \\ &\leqslant 2\epsilon + \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E)dr. \end{aligned}$$

Remark 15.2.5. *because*

$$\int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap N) dr = \int_{P_\theta(N)} \mathcal{H}^0(p_\theta(r) \cap N) dr$$

and $\mathcal{H}^1(N) = 0 \to \mathcal{H}^1(P_\theta(N)) = 0$*, the integral is* 0*. (Reminder:* P_θ *is orthogonal projection onto* $l(\theta)$ *and* $p_\theta(r)$ *is the line orthogonal to* $l(\theta)$ *at* r*.)*

8 Note that because O is the disjoint union of a countable collection of R^1 submanifolds O_i and a null set N, we get

$$\begin{aligned} &\frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O) dr\, d\theta \\ &\qquad = \frac{1}{2}\sum_i \int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O_i) dr\, d\theta \\ &\qquad\quad + \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap N) dr\, d\theta \\ &\qquad = \sum_i \mathcal{H}^1(O_i) + 0 \\ &\qquad = \mathcal{H}^1(O). \end{aligned}$$

9 Using the previous part and integrating the results of part 7 over the semicircle,

$$\begin{aligned} \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta &\leqslant \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap O) dr\, d\theta \\ &= \mathcal{H}^1(O) \\ &\leqslant \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta + 2\epsilon\pi \end{aligned}$$

we get

$$\left| \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta - \mathcal{H}^1(O) \right| \leqslant 2\epsilon\pi.$$

Recalling $\left|\mathcal{H}^1(O) - \mathcal{H}^1(E)\right| \leqslant \epsilon$, we get that

$$\left| \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta - \mathcal{H}^1(E) \right| \leqslant 2\epsilon\pi + \epsilon$$

and because ϵ was arbitrary, leads us to

$$\mathcal{H}^1(E) = \frac{1}{2}\int_0^\pi \int_{-\infty}^{\infty} \mathcal{H}^0(p_\theta(r) \cap E) dr\, d\theta.$$

Step 8: Crofton's Formula holds for any 1-dimensional rectifiable set.

Proof. Using Theorem 15.1.3 and Exercise 15.1.6, we can write E as the union of a countable number of disjoint pieces E_i of embedded C^1 submanifolds M_i with $\mathcal{H}^1(M_i) < \infty$ and a set of $\mathcal{H}^1$ measure 0, N

$$E = \left(\bigcup_i E_i\right) \cup N$$

so that $\mathcal{H}^1(E) = \sum_i \mathcal{H}^1(E_i)$ which, combined with Step 7, proves Crofton's formula for $(\mathcal{H}^1, 1)$-dimensional rectifiable sets in R^2. □

End of Proof of Theorem 15.2.1

Remark 15.2.6. *Note that our proof of Theorem 15.2.1 works as long as* E *has locally finite* $\mathcal{H}^1$ *measure, i.e.* $\mathcal{H}^1(K \cap E) < \infty$ $\forall$ *compact* $K \subset \mathbb{R}^2$.

Remark 15.2.7. *Note that without much comment, we have used the fact that a countable union of* $\mathcal{H}^1$ *measure* 0 *sets also has* $\mathcal{H}^1$ *measure zero, a projection of a measure* 0 *set has measure* 0 *and, if you mess with a function on a set of measure* 0*, this does not change the integral of the function (when integrating with respect to* $\mathcal{H}^1$*).*

Extending the proof to higher dimensions uses the ideas introduced in the case of 1-dimensional sets in $\mathbb{R}^2$, but there are further details that make it a bit more messy. The only real added complication is that we do not have a simple way to cover the set E we are projecting using a partition like we did in the case of 1-dimensional sets in $\mathbb{R}^2$.

The basic idea is again that a rectifiable set is a union of pieces of C^1 submanifolds and that C^1 implies that the pieces are close to flat at small scales. It takes more work though because we no longer have that flat patches approximating the manifold meet up with the manifold, so we do not have that the projections match for any of the projections in the Grassmannian. But they almost do, and with extra work we can add just another small term that goes to zero and everything works out. The most efficient way to prove this theorem is to invoke the co-area formula that was introduced in Section 12.6.2. See Morgan's book [32] for this proof.

The point of our proof for the 1-dimensional case, was to show what can be done with barehanded methods. This is something I am fond of and I believe it helps you if you are aiming for an intimate, instinctive grasp of something mathematical. What can be shown using the simplest insights, using only the simplest tools (and perhaps a fair bit of work)? Approaching a theorem from multiple directions deepens the understanding and broadens your ability to use and extend it.

15.3 Calculus, Deeply Generalized

So what is **Geometric Measure Theory** anyway?

Actually, studying the first 15 chapters of this book has already exposed you to pieces of geometric measure theory and geometric analysis. In fact, Chapter 14 was titled *An Invitation to Geometric Measure Theory: Part 1* when I wrote it for my blog (see [43]). And the first two sections of this chapter are also pieces of geometric measure theory.

Here is a succinct description I like:

> *Geometric Measure Theory: calculus, generalized to wild sets, measures and functions – here we find the most general and useful versions of theorems that we often have seen in some simple form in calculus and elementary analysis.*

and

> *Geometric analysis of, and on, sets, functions and measures in* $\mathbb{R}^n$

is the shortest phrase that encompasses a large portion of what is referred to as geometric measure theory and geometric analysis.

Instead of trying to give a synopsis of the entire field, I will instead leverage what we have already learned and go a bit more deeply in the rectifiability corner of geometric measure theory. After this, I will make a few comments on references and how to take your study of geometric measure theory to the next level.

15.3.1 What Rectifiability Opens

We will follow two threads of the many we could follow into a deeper understanding of rectifiability. The first tells us how local, asymptotic properties of measures (densities) tell us precise things about the structure of locally finite measures. The second looks briefly at sets of finite perimeter in $\mathbb{R}^n$. These are essentially those sets that, if we only have integration of smooth test functions with respect to $\mathcal{L}^n$ to measure things with, look like they have boundaries with finite $\mathcal{H}^{n-1}$ measure.

15.3.1.1 *Marstrand, Preiss, and the Importance of Being Rectifiable*

In the same way that $f : \mathbb{R}^n \to \mathbb{R}^m$, satisfying the simple Lipschitz bound on expansion

$$|f(x) - f(y)| \leqslant K|x - y|,$$

lets in wildness but not so much that you don't still have linear approximations almost everywhere, we have that a subset of a countable collection of Lipschitz images

$$E \subset \bigcup_i f_i(\mathbb{R}^k),$$

where $f_i : \mathbb{R}^k \to \mathbb{R}^n$ are Lipschitz, $k < n$ and $\mathcal{H}^k(E) < \infty$, will have k-dimensional approximate tangent planes $\mathcal{H}^k$ almost everywhere.

To go further in this direction, as we will do in this section, we need a few more definitions.

Exercise 15.3.1. Suppose that

$$\mathbb{Q} \cap [0, 1] = \{q_i\}_{i=1}^{\infty}$$

and that

$$E = \cup_i \{[0, 1] \times \{q_i\}\} \subset [0, 1]^2.$$

Show that at no point of E is there a approximate tangent line.

Definition 15.3.1 (Locally Finite Borel Measure). *A Borel measure μ on $\mathbb{R}^n$ is locally finite if, for every compact set $K \subset \mathbb{R}^n$, $\mu(K) < \infty$. In $\mathbb{R}^n$, this is equivalent to requiring $\mu(B(0, r)) < \infty$ for all $r < \infty$.*

Definition 15.3.2 (k-rectifiable Measure). *A locally finite Borel measure is k-rectifiable if there is a k-rectifiable set E such that $\mu(\mathbb{R}^n \setminus E) = 0$.*

Definition 15.3.3 (Support of a Measure). *The support of a measure* μ *on* $\mathbb{R}^n$ *is the complement of the points* $x \in \mathbb{R}^n$ *such that* $\mu(B(x, r)) = 0$ *for some* $r > 0$. *Alternatively, it is the complement of* E, *the largest open set in* $\mathbb{R}^n$ *such that* $\mu(E) = 0$.

Remark 15.3.1. *We note that a* k*-rectifiable measure need not have a* k*-rectifiable support.*

Exercise 15.3.2. Find a 1-rectifiable measure in $[0, 1]^2$ whose support E is not 1-rectifiable.

Definition 15.3.4 (Restriction of a Measure to a Set). *We define* $\mu \llcorner E$ *to be the measure generated by the measure* μ *and the set* E, *defined by*

$$\mu \llcorner E(F) \equiv \mu(E \cap F).$$

Exercise 15.3.3. Suppose that

$$\mathbb{Q} \cap [0, 1] = \{q_i\}_{i=1}^{\infty},$$

$$E_i \equiv [0, 1] \times \{q_i\} \subset [0, 1]^2,$$

$$E \equiv \cup_i E_i,$$

and

$$\mu \equiv \bigcup_i \left\{ \frac{1}{2^i} \mathcal{H}^1 \llcorner E_i \right\}.$$

Show that at almost every point of E, there is a (horizontal) approximate tangent line.

To prepare us to understand the statements of Marstrand's result and Preiss' result, we look at the notion of densities (which we already encountered in Chapter 14) a bit more deeply.

Definition 15.3.5 (α**-dimensional Densities of Sets).** *We define the upper and lower α-dimensional (Hausdorff) densities of a set E at a point p, by*

$$\theta^{*\alpha}(E,p) \equiv \limsup_{r\to 0} \frac{\mathcal{H}^\alpha(B(p,r)\cap E)}{\omega(\alpha)r^\alpha}$$

$$\theta_*^{\alpha}(E,p) \equiv \liminf_{r\to 0} \frac{\mathcal{H}^\alpha(B(p,r)\cap E)}{\omega(\alpha)r^\alpha}$$

and when $\theta^{*\alpha}(E,p) = \theta_*^\alpha(E,p)$, we simply refer to this limit as the α-dimensional density of E at p

$$\theta^{\alpha}(E,p) \equiv \lim_{r\to 0} \frac{\mathcal{H}^\alpha(B(p,r)\cap E)}{\omega(\alpha)r^\alpha}.$$

See Remark 15.3.2 for a reminder of the definition of $\omega(\alpha)$.

Theorem 15.3.1 (Bounds on Upper Densities)**.** *Suppose $E \subset \mathbb{R}^n$ and $\mathcal{H}^\alpha(E) < \infty$, then*

$$\frac{1}{2^\alpha} \leqslant \theta^{*\alpha}(E,p) \equiv \limsup_{r\to 0} \frac{\mathcal{H}^\alpha(B(p,r)\cap E)}{\omega(\alpha)r^\alpha} \leqslant 1$$

Remark 15.3.2 (Reminder: $\omega(\alpha)$). *Recall that for any $\alpha \geqslant 0$, $\omega(\alpha) \equiv$ is the α-volume of the "α-dimensional unit ball" and, in the case in which $\alpha = k$, an integer, $\omega(k)$ = the usual k-volume of the k-dimensional unit ball in $\mathbb{R}^k$. We note that $\omega(0) = 1$. See Definition 12.3.6 on page 285.*

Analogously, we have:

Definition 15.3.6 (α**-dimensional Densities of Measures).** *We define the upper and lower α-dimensional densities of a measure μ at a point p, by*

$$\theta^{*\alpha}(\mu,p) \equiv \limsup_{r\to 0} \frac{\mu(B(p,r))}{\omega(\alpha)r^\alpha}$$

$$\theta_*^{\alpha}(\mu,p) \equiv \liminf_{r\to 0} \frac{\mu(B(p,r))}{\omega(\alpha)r^\alpha}$$

and when when $\theta^{*\alpha}(\mu, p) = \theta_*^{\alpha}(\mu, p)$, *we simply refer to this limit as the* α*-dimensional density of* μ *at p*

$$\theta^{\alpha}(\mu, p) \equiv \lim_{r \to 0} \frac{\mu(B(p, r))}{\omega(\alpha) r^{\alpha}}.$$

See the Figures (91-93). The first figure illustrates the geometry of a 2-dimensional density ratio, the second illustrates the limiting process leading to a 1-dimensional density, and the third illustrates a 1-dimensional set E, whose upper and lower densities are not the same at a point $p \in E$.

Remark 15.3.3 (Details on Figure (93))**.** *We add* a *in length by wiggling the curve inside* $B(p, r_1)$ *but outside* $B(p, r_2)$, *then we repeat this, scaled for each pair of balls – we add* $\frac{a}{10^n}$ *in length to the curve by wiggling it inside* $B(p, r_{2n-1})$ *but outside* $B(p, r_{2n})$. *Reiterating from the figure the facts that* $r_{2n-1} = 10^{-(n-1)}$ *and* $r_{2n} = (0.9)10^{-(n-1)}$, *we are led to the densities*

$$\begin{aligned} \frac{\mathcal{H}^1(E \cap B(p, r_1))}{2r_1} &= \frac{2 + a + \frac{a}{10} + \frac{a}{100} + \frac{a}{1000} + \ldots}{2} \\ &= \frac{2 + a\frac{10}{9}}{2} \end{aligned}$$

$$\begin{aligned} \frac{\mathcal{H}^1(E \cap B(p, r_2))}{2r_2} &= \frac{2(0.9) + \frac{a}{10} + \frac{a}{100} + \frac{a}{1000} + \ldots}{2(0.9)} \\ &= \frac{2(0.9) + a\frac{1}{9}}{2(0.9)} \end{aligned}$$

where we have used the fact that $1.1111.. = \frac{10}{9}$. *Note now, because of the scale invariant way we have added the wiggles and chosen the* r_i*'s, we can immediately write down the following ratios:*

$$\begin{aligned} \frac{\mathcal{H}^1(E \cap B(p, r_{2n-1}))}{2r_{2n-1}} &= \frac{2 + \frac{10}{9}a}{2} \\ &= 1 + \frac{5}{9}a \end{aligned}$$

$$\frac{\mathcal{H}^1(E \cap B(p, r_{2n})}{2r_{2n}} = \frac{2(0.9) + \frac{1}{9}a}{2(0.9)}$$
$$= 1 + \frac{5}{81}a.$$

This allows us to conclude that

$$\theta^1_*(E, p) \leqslant 1 + \frac{5}{81}a < 1 + \frac{5}{9}a \leqslant \theta^{*1}(E, p)$$

implying that the density $\theta^1(E, p)$ *does not exist.*

Exercise 15.3.4. Verify that indeed, the ratios discussed in Remark 15.3.3 do not change.

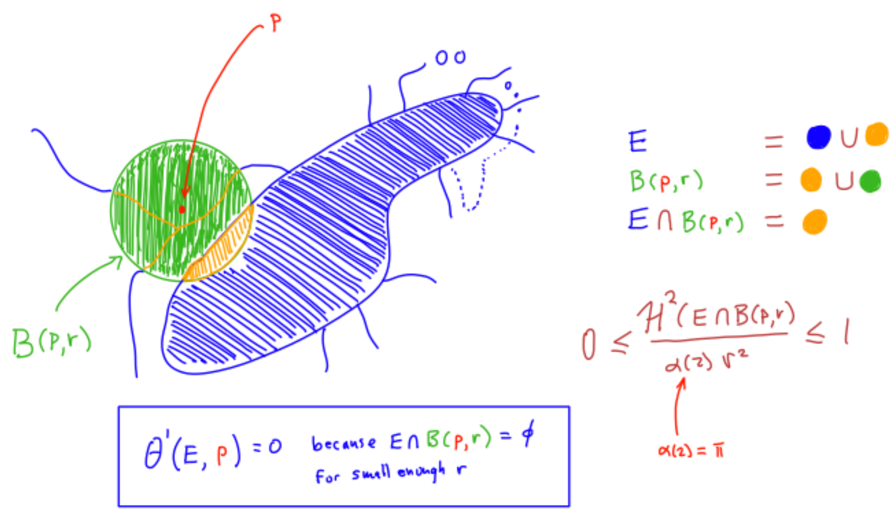


Figure 91: Illustration of the density of a 2-dimensional set.

We are ready to state the amazing results of Marstrand and Preiss.

Theorem 15.3.2 (Marstrand's Theorem). *Suppose that* μ *measures* $\mathbb{R}^n$. *If* $\alpha \geqslant 0$, E *is a Borel set with* $\mu(E) > 0$, *and*

$$0 < \theta^\alpha(\mu, p) < \infty$$

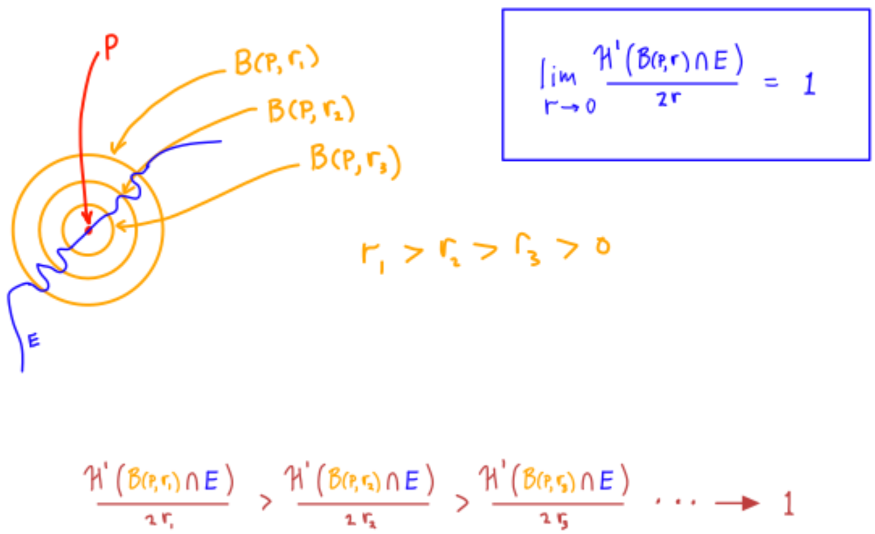


Figure 92: Illustration of the density of a 1-dimensional set.

for μ almost all $p \in E$, then α is an integer.

Remark 15.3.4. *Notice that the statement of the theorem implicitly tells you that on that set of full μ measure where the bounds hold, $\theta^{*\alpha}(\mu, p) = \theta^{\alpha}_{*}(\mu, p)$.*

Theorem 15.3.3 (Preiss' Theorem)**.** *Assume that μ is a locally finite measure on $\mathbb{R}^n$, $\alpha \geqslant 0$, and*

$$0 < \theta^{\alpha}(\mu, p) < \infty$$

for μ-almost all p. Then either

1. *$\mu = 0$ or*
2. *$\alpha = k \leqslant n$ where k is a natural number and the conditions above are equivalent to the existence of a Borel measurable function f and a collection $\{\Gamma_i\}_{i=1}^{N}$ (with $N \leqslant \infty$) of Lipschitz images of pieces of $\mathbb{R}^k$ in $\mathbb{R}^n$ such that*

$$\mu(E) = \sum_i \int_{\Gamma_i \cap E} f(p) d\mathcal{H}^k(p).$$

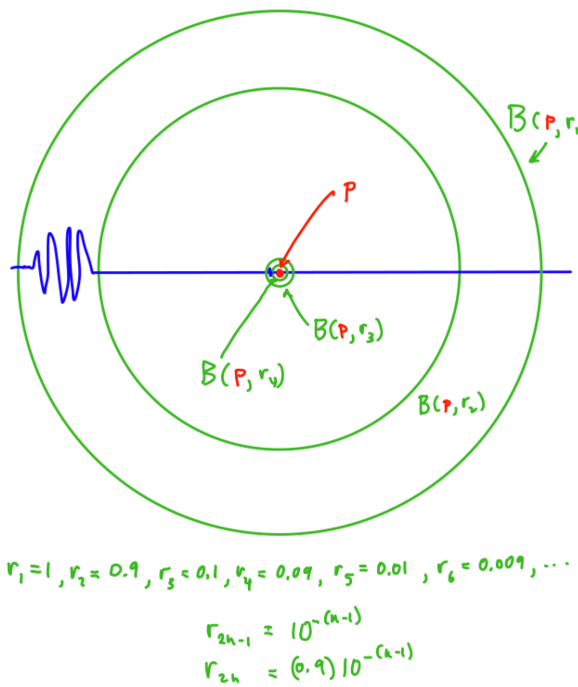


Figure 93: At each scale, we add extra length by adding a wiggle in the curve inside the $2n-1$-th ball but outside the $2n$-th ball. See the Remark (15.3.3) for the calculations.

Preiss' theorem includes Marstrand's theorem and simply put, these two results imply that, for locally finite measures,

> For locally finite measures, the measures with well behaved densities are precisely those measures that are rectifiable (i.e. well behaved!).

For more on Marstrand's theorem [27] and Preiss theorem [34], I recommend De Lellis' beautifully short exposition [11]. Federer thought that Preiss' paper was one of the most important mathematical results of the 20th century [18].

15.3.1.2 *Sets of Finite Perimeter*

When modeling many physical phenomena, boundaries are very often important to identify, to measure, to track. Whether you are solving partial differential equations on a subset E of $\mathbb{R}^n$, or thinking about the physics where you have two different phases that are interacting, modeling heat or substances diffusing from an object into the surrounding medium, or tracking chemicals diffusing across boundaries, the sizes of the boundaries are often of utmost importance. In the precise mathematical models we often want to allow complicated, even wild solutions, so we choose a space of candidate solutions that includes very wild behavior. It often becomes necessary to use a variety of definitions of boundary.

To explore this, we use densities, introduced in the last section, as well as the behavior of the density ratios as the radius goes to zero.

Suppose that $E \subset \mathbb{R}^n$. One might think of $\frac{\mathcal{H}^k(B(p,r)\cap E)}{\omega(k)r^k}$, for some $k \leqslant n$ as a measure of p's inclusion in E and E^c. This allows us to define several versions of the boundary of a set. We will only be exploring the cases of $k = n$ and $k = 0$.

r-Thick Measure Theoretic Boundaries, $\partial_{*r}E$ For any positive r and $E \subset \mathbb{R}^n$, we define

Definition 15.3.7 (r-**Thick Measure Theoretic Boundary**). *The* r*-thick measure-theoretic boundary,* $\partial_{*r}E$ *is defined by*

$$\partial_{*r}E \equiv \{p \mid \frac{\mathcal{H}^n(B(p,r)\cap E)}{\omega(n)r^n} > 0 \textit{ and } \frac{\mathcal{H}^n(B(p,r)\cap E^c)}{\omega(n)r^n} > 0\}.$$

While I am not aware of an official definition, but this is very natural and I would not be surprised if it has been used before. It probably was in the collection of things that David Caraballo and I discussed about 20 years ago, but we never pursued it further and I have forgotten.

r-Thick Topological Boundaries, $\partial_r E$ For any positive r and $E \subset \mathbb{R}^n$, we analogously define

Definition 15.3.8 (r-**Thick Topological Boundary**). *The* r*-thick topological boundary* $\partial_r E$ *is defined by*

$$\partial_r E \equiv \{p \mid \frac{\mathcal{H}^0(B(p,r)\cap E)}{\omega(0)r^0} > 0 \textit{ and } \frac{\mathcal{H}^0(B(p,r)\cap E^c)}{\omega(0)r^0} > 0\}.$$

This is a sensible definition of a topological boundary at resolution r, or r-thick topological boundary. Again, while I am not aware of an official definition of this version, it is almost certainly already defined somewhere since, if we define the r-neighborhood of a E to be $N_r(E)$,

$$N_r \equiv \{x \mid d(x,E) < r\},$$

then we have that

$$\partial_r E = N_r(E) \setminus N_r(E^c).$$

Topological boundary, ∂E This is the boundary you encounter in metric spaces – points whose neighborhoods always contain points in E and points not in E.

Definition 15.3.9 (**Topological Boundary from Densities**). *Instead of the usual topological definition, an equivalent definition of the topological boundary* ∂E *using* 0*-dimensional densities is*

$$\partial E \equiv \{p \mid \forall r > 0,\ \ \frac{\mathcal{H}^0(B(p,r)\cap E)}{\omega(0)r^0} > 0 \textit{ and } \frac{\mathcal{H}^0(B(p,r)\cap E^c)}{\omega(0)r^0} > 0\}.$$

Note that a set of measure zero in $\mathbb{R}^n$ (for example all the points with rational coordinates) can be dense in $\mathbb{R}^n$. The topological boundary of such a set is all of $\mathbb{R}^n$!

MEASURE THEORETIC BOUNDARY, $\partial_* E$ The measure theoretic boundary of $E \subset \mathbb{R}^n$, $\partial_* E$, are all the points $p \in \mathbb{R}^n$ that "measure theoretically see" both E and its complement, E^c. This is expressed with upper densities as follows,

Definition 15.3.10 (Measure Theoretic Boundary). *The point p is in the measure theoretic boundary of E – more succinctly, $p \in \partial_* E$ – if and only if*

$$\theta^{*n}(E, p) = \limsup_{r \to 0} \frac{\mathcal{H}^n(E \cap B(p, r))}{\omega(n) r^n} > 0$$

and

$$\theta^{*n}(E^c, p) = \limsup_{r \to 0} \frac{\mathcal{H}^n(E^c \cap B(p, r))}{\omega(n) r^n} > 0.$$

REDUCED BOUNDARY, $\partial^* E$ The reduced boundary is the most restrictive of all the boundaries we will be covering. The essence is that, measure theoretically speaking, a point of the measure theoretic boundary is a point where when you zoom in enough, the set looks like a half plane.

Definition 15.3.11 (The Reduced Boundary). *We will say that $p \in \mathbb{R}^n$ is in $\partial^* E$, the **Reduced Boundary** of E, if there exists a unit vector ν_p such that $H^+ \equiv \{x \in \mathbb{R}^n \mid \langle x - p, \nu_p \rangle \geqslant 0\}$ (and $H^- \equiv (H^+)^c$) satisfies*

$$\theta^n(E \cap H^+, p) = \lim_{r \to 0} \frac{\mathcal{H}^n(E \cap H^+)}{\omega(n) r^n} = 0$$

and

$$\theta^n(E^c \cap H^-, p) = \lim_{r \to 0} \frac{\mathcal{H}^n(E^c \cap H^-)}{\omega(n) r^n} = 0.$$

Remark 15.3.5. *ν_p is the measure theoretic outward unit normal to E at p.*

Remark 15.3.6. *The way the reduced boundary usually appears is in the context of sets of finite perimeter. That approach, which is explained in great detail in chapter 5 of Evans and Gariepy [12], arises from the study of functions of bounded variation. We first figure out how to compute the integral of the norm of the gradient of $\chi_E(x)$, the characteristic function* *of a set* $E \subset \mathbb{R}^n$*. Recall that* $\chi_E(x)$ *is the function that equals one on* E *and* 0 *everywhere else. When* $\int_{\mathbb{R}^n} |\nabla \chi_E|$ *(which takes work to make sense of) is bounded, we say that* E *is a* ***set of finite perimeter*** *and* $\nabla \chi_E$ *can be represented by a pair – radon measure* μ_E *and a unit vector field* ν*.* ν *points into* E *(which makes sense – the gradient point "uphill") and the measure tells about the jump up –* $\mu_E(F)$ *turns out to be* $\mathcal{H}^{n-1}(\partial^* E \cap F)$*. In this case we say that* E *is a set of finite perimeter.*

Figure 94 illustrates r-thick boundaries. In this case the set is smooth, so the r-thick measure theoretic and r-thick topological boundaries are the same.

Figure 95 illustrates the fact that for points on a smooth boundary, the density of the set is always $\frac{1}{2}$ and that these points are in every version of the boundary we define in this section.

Figures 96 and 97 illustrate the difference between the topological, measure theoretic and reduced boundaries.

Figure 98 illustrates the ideas in my definition of the reduced boundary.

Figure 99 illustrates a point p that is not in the reduced or measure theoretic boundaries, and yet for all $r > 0$, both $\frac{\mathcal{H}^2(E \cap B(p,r))}{\pi r^2} > 0$ and $\frac{\mathcal{H}^2(E^c \cap B(p,r))}{\pi r^2} > 0$.

It turns out that for sets of finite perimeter $E \subset \mathbb{R}^n$, the reduced boundary $\partial^* E$ (and thus the measure theoretic boundary $\partial_* E$) are $(\mathcal{H}^{n-1}, n-1)$-rectifiable sets. The reason the $\partial_* E$ must be a rectifiable set when $\partial^* E$ is, is that $\partial^* E \subset \partial_* E$ and $\mathcal{H}^{n-1}(\partial_* E \setminus \partial^* E) = 0$. See Evans and Gariepy [12] Chapter 5 for details.

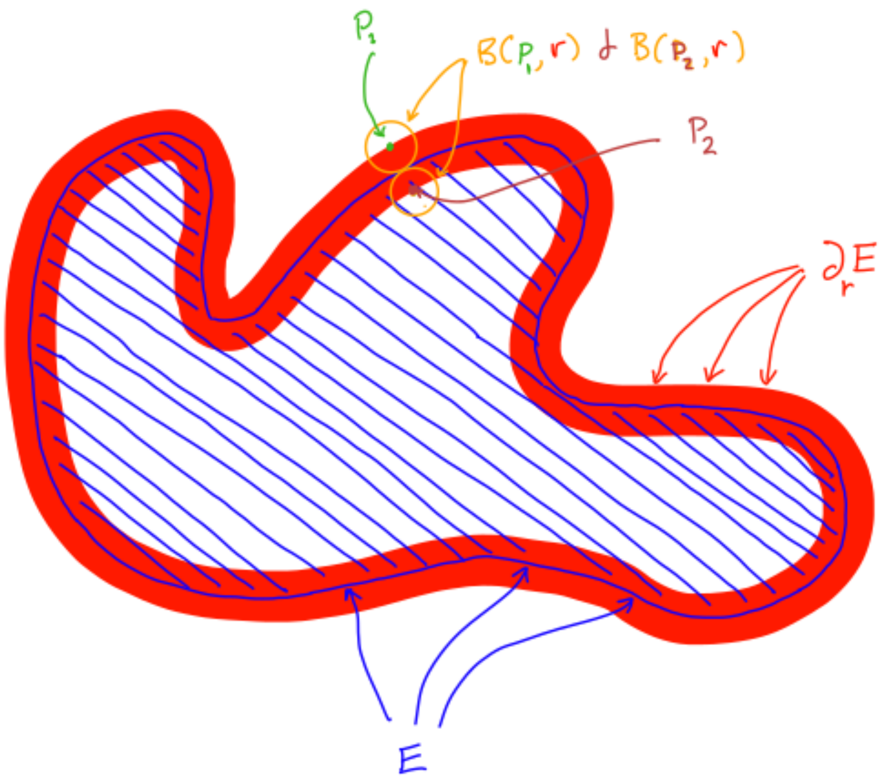


Figure 94: Here is an illustration of the r-thick boundary $\partial_r E$ of a nice (smooth set).

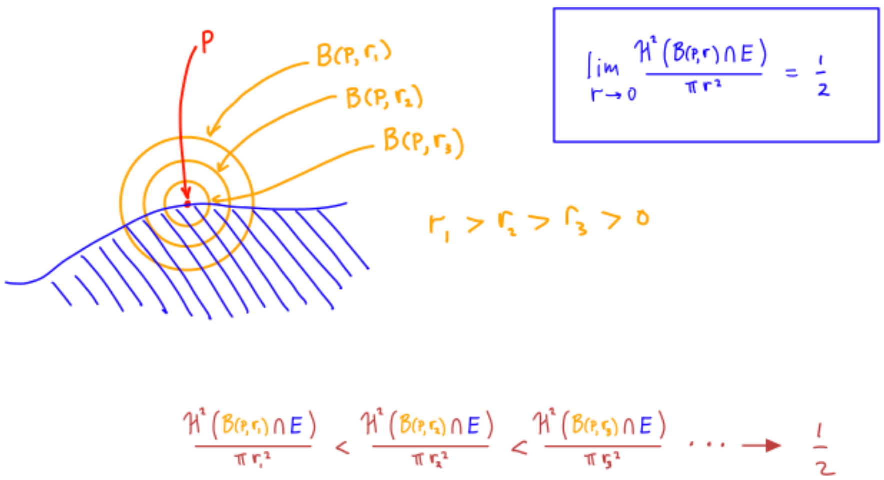


Figure 95: The density of the set at a point p where the function is smooth is always $\frac{1}{2}$. In this case, p will be in all 5 versions of boundary we define in this section.

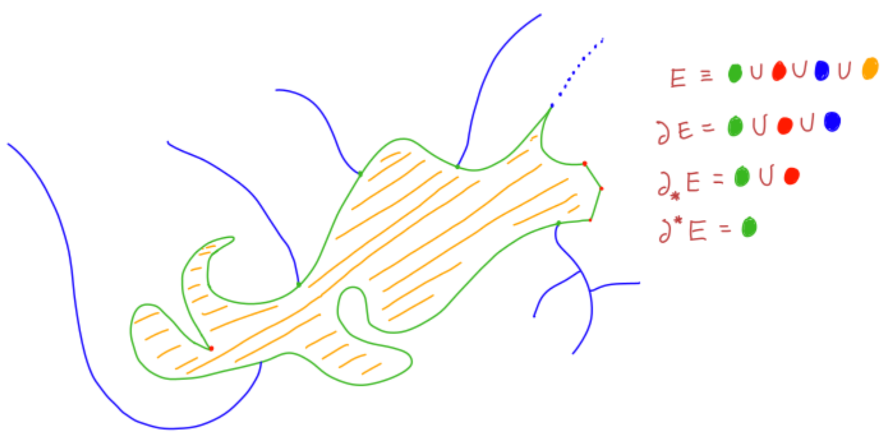


Figure 96: Example of a set and the points that are in the Topological boundary, the measure theoretic boundary and the reduced boundary. This is also an illustration of the fact that $\partial^* E \subset \partial_* E \subset \partial E$.

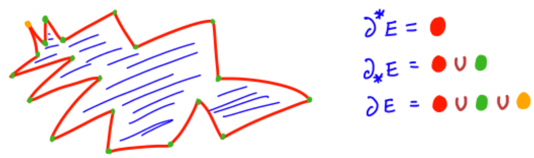


Figure 97: Another example of a set and the points that are in the Topological boundary, the measure theoretic boundary and the reduced boundary.

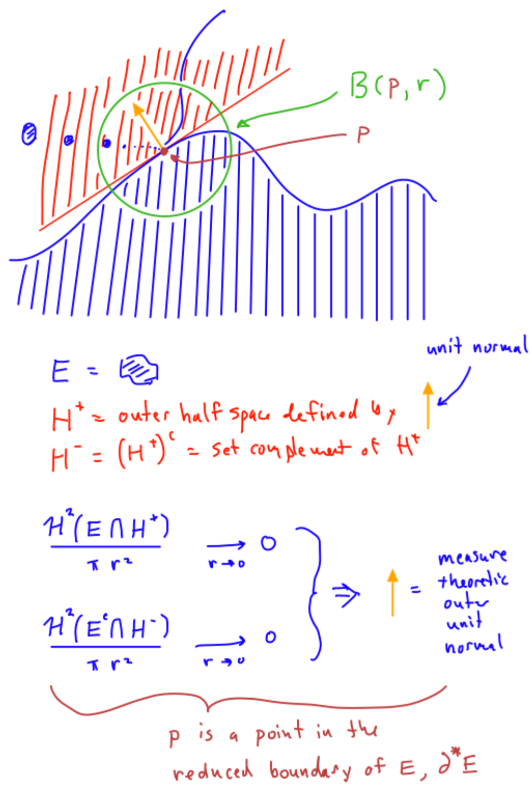


Figure 98: Illustration of how densities are used in the definition of the reduced boundary.

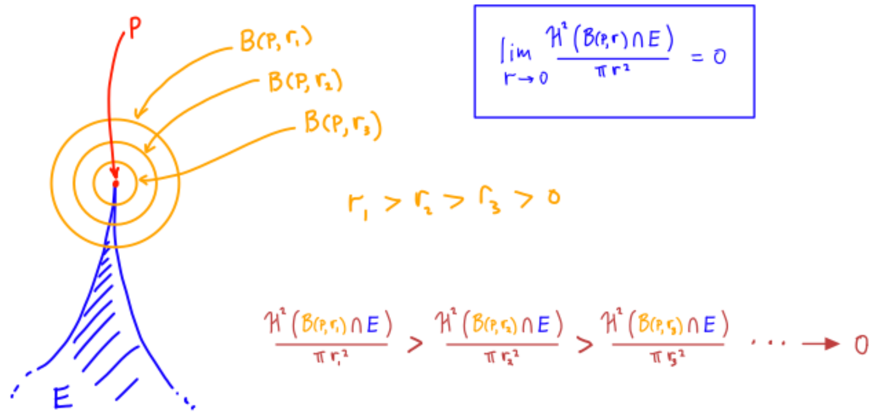


Figure 99: Illustration of the fact that cusp points are in the topological boundary, but not the reduced or the measure theoretic boundary.

Key takeaways for this section are that (1) there are lots of different versions of boundaries of a set, (2) densities are very useful in defining and understanding nuanced versions of boundary, (3) boundaries that are useful (though to be fair, I have not made an argument for useful here), are either $(\mathcal{H}^{n-1}, n-1)$-rectifiable sets or are thick boundaries of some sort.

15.3.2 *Next Steps and References*

When I met David Caraballo in 1999 and was swept away by the beauty and natural power of geometric measure theory, the first book I studied (at David's suggestion) was Evans and Gariepy's *Measure Theory and Fine Properties of Functions* [12].

Soon after, I met Bill Allard and began learning bits and pieces from him, and reading bits and pieces of Federer's tome [13]*, and (eventually) a whole host of other books I used to learn from, teach from, and reference now and then [17, 32, 37, 22, 11, 29, 26, 3, 24].†

Through the connection with Bill and David, Jean Taylor, Frank Morgan, Bob Hardt and Craig Evans began visiting, first at Los Alamos and then at WSU in Pullman. There were others as well (Harold Parks, John Garnett, and Triet Le, for example). The point of mentioning the visitors and friends is that these in person contacts, these social connections, were a very important part of the education that I and my students gained. It was an important part of how we moved into the culture and modes of thinking of geometric measure theory and geometric analysis. While it is not the only way to do that, I would encourage personal interactions if it all possible.

What is my recommendation for an efficient path into geometric measure theory?

Find collaborators (and experts) Find one or two others to talk with (at a blackboard!) while you are working towards mastery. Having an expert to ask questions of in person would be best, but having other dedicated scholars as peers is also very important.

Take time to think and time to connect Remember that you need lots of time alone and lots of time working with others.

Seek places rich in creativity Take every chance you can to go to places where there is a generous atmosphere, rich in mathematical thoughts and explorations. For example, MSRI in Berkeley, IPAM at

*When Bill was a graduate student at Brown (where Federer was), Federer was writing his famous book on geometric measure theory. Bill read through every detail of Federer's manuscript and made extensive suggestions to Federer. Anyone who knows Federer's book will be very impressed with this.

†For a more complete overview of the references you can read the article I wrote in 2012 [44]. Since then, the only other book I have added to the list of books used is Maggi's book [26].

UCLA, IMA at Minnesota, ICERM at Brown, etc. but also universities with friendly groups focused on these ideas.

Evans and Gariepy, 1-3 Master the first three chapters of Evans and Gariepy.

Morgan, 1-5 and then 6-12 Work through the first 5 chapters of Morgan's book carefully. Then understand all the ideas in chapters 6-12. Start dabbling in Federer's book.

Evans and Gariepy, 4-5 Master chapters 4 and 5 of Evans and Gariepy along with starting Mattila's book.

Mattila Finish Mattila.

Other References I also recommend Krantz and Parks [22] and Leon Simon's book [36] also. I have taught from both and like both of them.

Interest Guided Studies What you do next depends on your tastes and needs.

Exposition Always be on the lookout for expository/historical articles. For example, Almgren's famous *Questions and Answers about area–minimizing surfaces and geometric measure theory* [1] and Fleming's *Geometric Measure Theory at Brown in the 1960s* [15] and *De Giorgi and Geometric Measure Theory* [16]. (I recommend all three of these highly!)

As you launch into your next mathematical explorations, remember what was written in the Preface and Preamble to this book ... don't let becoming an expert in an area or two prevent or delay you from becoming a jack of all trades. Stay playful. Stay connected and grounded.

Blessings ... may you truly thrive!

Kevin

Bibliography

[1] Fred Almgren. Questions and Answers About Area–Minimizing Surfaces and Geometric Measure Theory. In *Proc. Symp. Pure Math*, volume 54, pages 29–53, 1993.

[2] Luigi Ambrosio. Fine Properties of Sets of Finite Perimeter in Doubling Metric Measure Spaces. *Set-Valued Analysis*, 10(2-3):111–128, 2002.

[3] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, 2000.

[4] Luigi Ambrosio and Bernd Kirchheim. Currents in Metric Spaces. *Acta Mathematica*, 185(1):1–80, 2000.

[5] Luigi Ambrosio and Bernd Kirchheim. Rectifiable Sets in Metric and Banach Spaces. *Mathematische Annalen*, 318(3):527–555, 2000.

[6] Luigi Ambrosio and Paolo Tilli. *Topics on Analysis in Metric Spaces*, volume 25. Oxford University Press, 2004.

[7] Edwin F Beckenbach and Richard Bellman. *Inequalities*, volume 30. Springer Science & Business Media, 2012.

[8] Tara Brach. Healing Trauma: Light Shines through the Broken Places. YouTube video, 2017. https://www.youtube.com/watch?v=eR-DN7JYSLo.

[9] Yurii D Burago and Viktor A Zalgaller. *Geometric Inequalities*, volume 285. Springer Science & Business Media, 2013.

[10] Robert Pemberton Burn. *Numbers and Functions: Steps into Analysis*. Cambridge University Press, 2015.

[11] Camillo De Lellis. *Rectifiable Sets, Densities and Tangent Measures*. European Mathematical Society, 2008.

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

[13] Herbert Federer. *Geometric Measure Theory – Reprint of the 1969 Edition*. Springer, 1996.

[14] Wendell H. Fleming. *Functions of Several Variables*. Springer, 1977.

[15] Wendell H Fleming. Geometric Measure Theory at Brown in the 1960s, 2015. http://www.dam.brown.edu/people/documents/Geometic_000.pdf.

[16] Wendell H Fleming. De Giorgi and Geometric Measure Theory, 2016. https://www.dam.brown.edu/people/documents/DeGiorgiandGeometricMeasureTheory_000.pdf.

[17] E Giusti. *Minimal Surfaces and Functions of Bounded Variation*, volume 80. Birkhauser, 1984.

[18] Bob Hardt. From a conversation with Bob Hardt, a well known mathematician and student of Herbert Federer.

[19] GH Hardy, JE Littlewood, and G Polya. Inequalities. *Press, Cambridge*, 1934.

[20] J. Heinonen. *Lectures on Analysis on Metric Spaces*. Springer, 2001.

[21] Juha Heinonen, Pekka Koskela, Nageswari Shanmugalingam, and Jeremy T Tyson. *Sobolev Spaces on Metric Measure Spaces: an Approach Based on Upper Gradients*. Number 27. Cambridge University Press, 2015.

[22] Steven G. Krantz and Harold R. Parks. *Geometric Integration Theory*. Birkhauser, 2008. ISBN-13: 978-0-8176-4676-9.

[23] David Levy. No Time to Think: a Google Tech Talk. YouTube, 2008. https://www.youtube.com/watch?v=KHGcvj3JiGA.

[24] Fang-Hua Lin and Xiaoping Yang. *Geometric Measure Theory: An Introduction*. Science Press, 2002.

[25] Tom L Lindstrøm. *Spaces: An Introduction to Real Analysis*, volume 29. American Mathematical Soc., 2017.

[26] Francesco Maggi. *Sets of Finite Perimeter and Geometric Variational Problems: an Introduction to Geometric Measure Theory*. Cambridge University Press, 2012.

[27] John M Marstrand. The (φ, s) regular Subsets of n-space. *Transactions of the American Mathematical Society*, 113(3):369–392, 1964.

[28] Gabor Mate. When the Body Says No: Mind/Body Unity and the Stress - Disease Connection. YouTube video, 2014. https://www.youtube.com/watch?v=qlHIWXWDuF0.

[29] Pertti Mattila. *Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability*, volume 44. Cambridge University Press, 1999.

[30] James D Meiss. *Differential Dynamical Systems*. SIAM, 2007.

[31] John W. Milnor. *Topology from the Differentiable Viewpoint*. University Press of Virginia, Charlottesville, 1965. Eighth printing,

1990.

[32] Frank Morgan. *Geometric Measure Theory: a Beginner's Guide*. Academic Press, 5th edition, 2016.

[33] Christian Pommerenke. *Boundary Behaviour of Conformal Maps*, volume 299. Springer Science & Business Media, 2013.

[34] David Preiss. Geometry of Measures in $\mathbb{R}^n$: Distribution, Rectifiability, and Densities. *Annals of Mathematics*, pages 537–643, 1987.

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

[36] Leon Simon. *Lectures on Geometric Measure Theory*. Proceedings Of The Centre For Mathematical Analysis, Australian National University, 1984.

[37] Leon Simon. *Lectures on Geometric Measure Theory*. Revision of 1984 lectures, 2014. https://web.stanford.edu/class/math285/ts-gmt.pdf.

[38] Kennan T. Smith. *Primer of Modern Analysis*. Springer-Verlag, 1983.

[39] J Michael Steele. *The Cauchy-Schwarz Master Class: an Introduction to the Art of Mathematical Inequalities*. Cambridge University Press, 2004.

[40] Steven H Strogatz. *Nonlinear Dynamics and Chaos with Student Solutions Manual: with Applications to Physics, Biology, Chemistry, and Engineering*. CRC press, 2018.

[41] William P. Thurston. On Proof and Progress in Mathematics. *Bulletin of the American Mathematical Society*, 30(2), 1994.

[42] Bessel A Van der Kolk. *The body Keeps the Score: Brain, Mind, and Body in the Healing of Trauma*. Penguin Books, 2015.

[43] Kevin R. Vixie. An Invitation to Geometric Measure Theory: Part 1, 2012. https://notesfromkevinrvixie.org/2012/12/14/226/.

[44] Kevin R. Vixie. Geometric Measure Theory by the Book, 2012. https://notesfromkevinrvixie.org/2012/11/02/geometric-measure-theory-by-the-book-2/.

Index of Symbols and Notation

Index