An Invitation to Mathematical Physics *and its History*

Jont B. Allen
Help from Steve Levinson, John D'Angelo, Michael Stone

University of Illinois at Urbana-Champaign

October 24, 2016

Contents

1 Introduction **13**

1.1 Early Science and Mathematics . . . 15
1.1.1 Lecture 1: (*Week 1*) Three Streams from the Pythagorean theorem . . . 15
1.1.2 Pythagorean Triplets . . . 16
1.1.3 What is mathematics? . . . 17
1.1.4 Early Physics as Mathematics . . . 17
1.1.5 The birth of modern mathematics . . . 19
1.1.6 The Three Pythagorean Streams . . . 20
1.2 Stream 1: Number Systems (9 Lectures) . . . 21
1.2.1 Lecture 2: The Taxonomy of Numbers: $\mathbb{P}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{I}, \mathbb{R}, \mathbb{C}$ . . . 22
1.2.2 Lecture 3: The role of physics in mathematics . . . 26
1.2.3 Lecture 4: Prime numbers . . . 29
1.2.4 Lecture 5: Early number theory: The Euclidean algorithm . . . 33
1.2.5 Lecture 6: Early number theory: Continued fractions . . . 34
1.2.6 Labor day: *Week 3* . . . 36
1.2.7 Lecture 7: Pythagorean triplets (Euclid's formula) . . . 36
1.2.8 Lecture 8: (*Week 3*) Pell's Equation . . . 37
1.2.9 Lecture 9: (*Week 4*) Fibonacci sequence . . . 40
1.2.10 Lecture 10: Exam I . . . 41
1.3 Stream 2: Algebraic Equations (11 Lectures) . . . 41
1.3.1 Lecture 11: *(Week 4)* Algebra and mathematics driven by physics . . . 41
1.3.2 Lecture 12: (*Week 5*) Examples of nonlinear algebra in Physics . . . 45
1.3.3 Lecture 13: Polynomial root classification (convolution) . . . 46
1.3.4 Lecture 14: Introduction to Analytic Geometry . . . 48
1.3.5 Lecture 15: (*Week 6*) Gaussian Elimination . . . 51
1.3.6 Lecture 16: Transmission (ABCD) matrix composition method . . . 52
1.3.7 Lecture 17: Riemann Sphere . . . 53
1.3.8 Lecture 18: (*Week 7*) Complex analytic mappings (Domain coloring) . . . 55
1.3.9 Lecture 19: Signals: Fourier transforms . . . 56
1.3.10 Lecture 20: Systems: Laplace transforms . . . 59
1.3.11 Lecture 21: (*Week 8*) Network (System) Postulates . . . 61
1.3.12 Lecture 22: Exam II . . . 62
1.4 Stream 3a: Scalar Calculus (12 Lectures) . . . 63
1.4.1 Lecture 23: (*Week 8*) Integration in the complex plane . . . 64
1.4.2 Lecture 24: Cauchy-Riemann conditions . . . 66
1.4.3 Lecture 25: (*Week 9*) Complex Analytic functions and Impedance . . . 67
1.4.4 Lecture 26: Three complex integration theorems I . . . 70
1.4.5 Lecture 27 Three complex integration theorems II . . . 71
1.4.6 Lecture 28: Inverse Laplace transform (Cauchy residue theorem) . . . 72
1.4.7 Lecture 29: (*Week 11*) Inverse Laplace transform and the Cauchy Residue Thm 72
1.4.8 Lecture 30: Case for causality: closing the contour . . . 72

1.4.9 Lecture 31: Properties of the LT . . . 72
1.4.10 Lecture 33: (*Week 12*) Multi-valued functions and Branch cuts . . . 72
1.4.11 Lecture 34: The Riemann Zeta function $\zeta(s)$ . . . 72
1.4.12 Lecture 35: Exam III . . . 73
1.5 Stream 3b: Vector Calculus (7 Lectures) . . . 73
1.5.1 Lecture 36: Scalar Wave Equation . . . 73
1.5.2 Lecture 37: (*Week 13*) Partial Diff Eqs of Physics; . . . 73
1.5.3 Lecture 38: Vector dot and cross products . . . 73
1.5.4 Thanksgiving Holiday 11/19–11/27 2016 . . . 73
1.5.5 Lectures 39: (*Week 14*) Vector operators . . . 74
1.5.6 Lecture 40: Definitions of E, H, B, D; Maxwell's Equations . . . 75
1.5.7 Lecture 41: Fundamental Theorem of Vector Calculus (Helmholtz theorem) . . . 75
1.5.8 Lecture 42 (*Week 15*): Various topics . . . 77
1.5.9 Lecture 43: Final overview . . . 78

2 Number Systems: Stream 1 81
2.1 Week 2 . . . 81
2.1.1 Lec 4 Two theorems on primes . . . 81
2.1.2 Lec 5 Greatest common divisor (GCD) . . . 83
2.1.3 Lec 6 Continued Fraction Expansion (CFA) . . . 85
2.2 Week 3 . . . 86
2.2.1 Lec 7 Pythagorean triplets (PTs) and Euclid's formula , . . . 86
2.2.2 Lec 8 Pell's Equation . . . 89
2.3 Week 4 . . . 89
2.3.1 Lec 9 Fibonacci Numbers . . . 89
2.3.2 Lec 10 Exam I . . . 91

3 Algebraic Equations: Stream 2 93
3.1 Week 4 . . . 93
3.1.1 Lec 11 Algebra and geometry as physics . . . 93
3.2 Week 5 . . . 95
3.2.1 Lec 12 The physics behind complex analytic expressions: linear vs. nonlinear . . . 95
3.2.2 Lec 13 Root classification of polynomials . . . 98
3.2.3 Lec 14 Analytic Geometry . . . 100
3.3 Week 6 . . . 101
3.3.1 Lec 15 Gaussian Elimination . . . 101
3.3.2 Lec 16 Matrix composition: Bilinear and ABCD . . . 104
3.3.3 Lec 17 Introduction to the Branch cut and Riemann sheets . . . 105
3.4 Week 7 . . . 105
3.4.1 Lec 18 Complex analytic mappings (domain coloring) . . . 105
3.4.2 Lec 19 Signals and Systems: Fourier vs. Laplace Transforms . . . 107
3.4.3 Lec 20 Role of Causality and the Laplace Transform . . . 107
3.5 Week 8 . . . 107
3.5.1 Lec 21 The 9 postulates of System of algebraic Networks . . . 107
3.5.2 Lec 22 Exam II (Evening) . . . 113

4 Scalar Calculus: Stream 3a 115
4.1 Week 8 . . . 115
4.1.1 Lec 23 Newton and early calculus & the Bernoulli Family . . . 115
4.2 Week 9 . . . 117
4.2.1 Lec 24 Power series and complex analytic functions . . . 117
4.2.2 Lec 25 Integration in the complex plane . . . 117

4.2.3 Lec 26 Cauchy Riemann conditions: Complex-analytic functions . . . 117
4.3 Week 10 . . . 118
4.3.1 Lec 27 Differentiation in the complex plane . . . 118
4.3.2 Lec 28 Three complex Integral Theorems . . . 118
4.3.3 Lec 29 Inverse Laplace Transform . . . 118
4.4 Week 11 . . . 119
4.4.1 Lec 30 Inverse Laplace Transform & Cauchy residue theorem . . . 119
4.4.2 Lec 31 The case for causality . . . 119
4.4.3 Lec 32 Laplace transform properties: Modulation, time translation, etc. . . . 119
4.5 Week 12 . . . 120
4.5.1 Lec 33 Multi-valued complex functions, Branch Cuts, Extended plane . . . 120
4.5.2 Lec 34 The Riemann Zeta function $\zeta(s)$ . . . 120
4.5.3 Lec 35 Exam III . . . 121

5 Vector Calculus: Stream 3b 123
5.1 Week 13 . . . 123
5.1.1 Lec 36 Scalar Wave equation . . . 123
5.1.2 Lec 37 Partial differential equations of physics . . . 123
5.1.3 Lec 38 Gradient, divergence and curl vector operators . . . 123
5.2 Thanksgiving Holiday . . . 124
5.3 Week 14 . . . 124
5.3.1 Lec 39 Geometry of Gradient, divergence and curl vector operators . . . 124
5.3.2 Lec: 40 Introduction to Maxwell's Equation . . . 127
5.3.3 Lec: 41 The Fundamental theorem of Vector Calculus . . . 127
5.4 Week 15 . . . 127
5.4.1 Lec 42: The Quasi-static approximation and applications . . . 127
5.4.2 Lec 43: Last day of class: Review of Fund Thms of Mathematics . . . 127

A Notation 131
A.1 Number systems . . . 131
A.1.1 Greek letters . . . 131
A.1.2 Table of *double-bold* number notation . . . 131
A.2 Periodic functions . . . 132
A.3 Vectors . . . 132
A.4 Matrices . . . 132
A.5 Differential equations vs. Polynomials . . . 133
A.6 Residue expansions and the ROC . . . 134

B Linear algebra on 2x2 matrices 135
B.1 Gaussian elimination . . . 135
B.1.1 Problems . . . 137
B.2 Gaussian Elimination of a 2x2 matrix . . . 137
B.2.1 Derivation of the inverse of a 2x2 matrix . . . 138

C Eigenvector analysis 139

D Solution to Pell's Equation (N=2) 143
D.1 Pell equation for N=3 . . . 143

Abstract

An understanding of physics requires knowledge of mathematics. The contrary is not true. By definition, pure mathematics contains no physics. Yet historically, mathematics has a rich history filled with physical applications. Mathematics was developed by people with intent of making things work. In my view, as an engineer, I see these creators of early mathematics, as budding engineers. This book is an attempt to tell this story, of the development of mathematical physics, as viewed by an engineer.

The book is broken down into three topics, called streams, presented as five chapters: 1) Introduction, 2) Number systems, 3) Algebra Equations, 4) Scalar Calculus, and 5) Vector Calculus. The material is delivered as 40 "Lectures" spread out over a semester of 15 weeks, three lectures per week, with a 3 lecture time-out for administrative duties. Problems are provided for each week's assignment. These problems are written out in LaTeX, with built in solutions, that may be expressed by un-commenting one line. Once the home-works are turned in, each student is given the solution. With regard to learning the material, the students rated these Assignments as the most important part of the course. There is a built in interplay between these assignments and the lectures. On many occasions I solved the homework in class, as motivation for coming to class. Four exams were given, one at the end of each of the three sections, and a final. Some of the exams were in class and some were evening exams, that ran over two hours. The final was two hours. Each of the exams, like the assignments, is provided as a LaTeX file, with solutions encoded with a one line software switch. The Exams are largely based on the Assignments. It is my philosophy that, in principle, the students could see the exam in advance of taking it.

Author's Personal Statement

An expert is someone who has made all the mistakes in a small field. I don't know if I would be called and expert, but I certainly have made my share of mistakes. I openly state that "I love making mistakes, because I learn so much from them." One might call that the "expert's corollary."

This book has been written out of both my love for the topic of mathematical physics, and a frustration for wanting to share many key concepts, and even new ideas on these basic concepts. Over the years I have developed a certain *physical sense* of math, along with a related mathematical sense of physics. While doing my research,[1] I have come across what I feel are certain conceptual holes that need filling, and sense many deep relationships between math and physics, that remain unidentified. While what we presently teach is not wrong, it is missing these relationships. What is lacking is an intuition for how math "works." We need to start listening to the language of mathematics. We need to let mathematics guide us toward our engineering goals.

It is my strong suspicion that over the centuries many others have had similar insights, and like me, have been unable to convey this slight misdirection. I hope these views can be useful to open young minds.

Venn diagram

This marriage of math and physics will help us make progress in understanding the physical world. I turn to mathematics and physics when trying to understand the universe. I have arrived in my views following a lifelong attempt to understand human communication. This research arose from my 32 years at Bell Labs in the Acoustics Research Department. There such lifelong pursuits were not only possible, they were openly encouraged. The idea was that if you were successful at something, take it

[1] http://auditorymodels.org/index.php/Main/Publications

as far as you can. But on the other side, don't do something well that's not worth doing. People got fired for the latter. I should have left for University after a mere 20 years,[2] but the job was just too cushy.

In this text it is my goal to clarify some of the conceptual errors when telling the story about physics and mathematics. My views have been often inspired by classic works, as documented in the bibliography. This book was inspired by my careful reading of Stillwell (2002), through Chapter 21 (Fig. 2). Somewhere in Chapter 22 I stopped reading and switched to the third edition (Stillwell, 2010), where I saw there was much more to master. At that point I saw that teaching this material to sophomores would allow me to absorb the more advanced material at a reasonable pace, which led to to this book.

Back Cover Summary

This is foremost a math book, but not the typical math book. First, this book is for the engineering minded, for those who need to understand math to do engineering, to learn how things work. In that sense it is more about physics and engineering. Math skill are critical to making progress in building things, be it pyramids or computers, as clearly shown by the many great civilizations of the Chinese, Egyptians, Arabs (people of Mesopotamia), Greeks and Romans.

Second, this is a book about the math that developed to explain physics, to allow people to engineer complex things. To sail around the world one needs to know how to navigate. This requires a model of the planets and stars. You can only know where you are on earth if you understand where earth is, relative to the heavens. The answer to such a deep questions will depend on who you ask. The utility and accuracy of that answer depends critically on the depth of understanding of how the worlds and heavens work. Who is qualified to answer such question? It is best answered by those who study mathematics applied to the physical world.

Halley (1656–1742), the English astronomer, asked Newton (1643–1727) for the equation that describes the orbit of the planets. Halley was obviously interested in comets. Newton immediately answered "an ellipse." It is said that Halley was stunned by the response (Stillwell, 2010, p. 176), as this was what had been experimentally observed by Kepler (c1619), and thus he knew Newton must have some deeper insight (Stillwell, 2010, p. 176).

When Halley asked Newton to explain how he knew this correct answer, Newton said he calculated it. But when challenged to show the calculation, Newton was unable to reproduce it. This open challenge eventually led to Newton's grand treatise, *Philosophiae Naturalis Principia Mathematica* (July 5, 1687). It had a humble beginning, more as a letter to Halley, explaining how to calculate the orbits of the planets. To do this Newton needed mathematics, a tool he had mastered. It is widely accepted that Isaac Newton and Gottfried Leibniz invented calculus. But the early record shows that perhaps Bhāskara II (1114–1185 AD) had mastered this art well before Newton.[3]

Third, the main goal of this book is to teach engineering mathematics, in a way that it can be understood, remembered, and mastered, by anyone motivated to learn this topic. How can this near impossible goal be achieved? The answered is to fill in the gaps with "who did what, and when." There is an historical story that may be told and mastered, by anyone serious about the science of making things.

One cannot be an expert in a field if they do not know the history of that field. This includes who the people were, what they did, and the credibility of their story. Do you believe the Pope or Galileo, on the topic of the relative position of the sun and the earth? The observables provided by science are clearly on Galileo's side. Who were those first engineers? They are names we all know: Archimedes, Pythagoras, Leonardo da Vinci, Galileo, Newton, ... All of these individuals had mastered mathematics. This book teaches the tools taught to every engineer. Do not memorize

[2] I should have left when AT&T Labs was formed, c1997. I started around December 1970, fresh out of Graduate school, and retired in December 2002.

[3] `http://www-history.mcs.st-and.ac.uk/Projects/Pearce/Chapters/Ch8_5.html`

complex formulas, rather make the equations "obvious" by teaching the simplicity of the underlying concept.

Credits

Besides thanking my parents, I would like to credit John Stillwell for his constructive, historical summary of mathematics. My close friend and colleague Steve Levinson somehow drew me into this project, without my even knowing it. My brilliant graduate student Sarah Robinson was constantly at my side, grading home-works and exams, and tutoring the students. Without her, I would not have survived the first semester the material was taught. Her proof-reading skills are amazing. Thank you Sarah for your infinite help. Finally I would like to thank John D'Angelo for putting up with my many silly questions. When it comes to the heavy hitting, John was always there to provide a brilliant explanation that I could easily translate into Engineer'ese (Matheering?) (i.e., Engineer language).

To write this book I had to master the language of mathematics (John's language). I had already mastered the language of engineering, and a good part of physics.[4] But we are all talking about the same thing. Via the physics and engineering, I already had a decent understanding of the mathematics, but I did not know that language. Hopefully, now I can get by.

Finally I would like to thank my wife (Sheau Feng Jeng aka Patricia Allen) for her unbelievable support and love. She delivered constant piece of mind, without which this project could never have been started, much less finish.

There are many others who played important roles, but they must remain anonymous, out of my fear of offending someone I forgot to mention.

–Jont Allen, Mahomet IL, Dec. 24, 2015

[4]Each genre (i.e, group) speaks their own dialect. One of my secondary goals is to bring down this scientific Tower of Babble.

Preface

It is widely acknowledged that interdisciplinary science is the backbone of modern scientific investigation. This is embodied in the STEM (Science, Technology, Engineering, and Mathematics) programs. Contemporary research is about connecting different areas of knowledge, thus it requires an understanding of cross-disciplines. However, while STEM is being taught, interdisciplinary science is not, due to its inherent complexity and breadth. Furthermore there are few people to teach it. Mathematics, Engineering and Physics (MEP) are at the core of such studies.[5]

STEM vs. MEP

Mathematics is based on the application rigor. Mathematicians specifically attend to the definitions of increasingly general concepts. Thus mathematics advances slowly, as these complex definitions must be collectively agreed upon. Mathematics shuns controversy, and embraces rigor, the opposite of uncertainty. Physics explores the fringes of uncertainty. Physicists love controversy. Engineering addresses the advancement the technology. Engineers, much like mathematicians, are uncomfortable with uncertainty, but are trained to deal with it.

To create such an interdisciplinary STEM program, a unified MEP curriculum is needed. In my view this unification could (should) take place based on a core mathematical training, from a historical perspective, starting with Euclid or before (i.e., Chinese mathematics), up to modern information theory and logic. As a bare minimum, the *fundamental theorems of mathematics* (arithmetic, algebra, calculus, vector calculus, etc.) need to be appreciated by every MEP student. The core of this curriculum is outlined in Table 1.1 (p. 21).

If, in the sophomore semester, students are taught a common MEP methodology and vocabulary, presented in terms of the history of mathematics, they will be equipped to

1. Exercise interdisciplinary science (STEM)

2. Communicate with other MEP trained (STEM) students and professors.

The goal is a *comprehensive understanding of the fundamental concepts of mathematics*, defined as those required for engineering. We assume that students with this deep understanding will end up being in the top 0.1% of Engineering. Time will tell if this assumption is correct.

The key tool is methodology. The traditional approach is a five to six course sequence: Calc I, II, III, DiffEq IV, Linear Algebra V and Complex Variables VI, over a time frame of three years (six semesters). This was the way I learned math. Following such a formal training regime, I felt I had not fully mastered the material, so I started over. I now consider myself to be self-taught. We need a more effective teaching method. I am not suggesting we replace the standard 6 semester math curriculum, rather I am suggesting replacing Calc I, II with this mathematical physics course, based on the historical thread, for those students who have demonstrated advanced ability. One needs more than a high school education to succeed in college engineering courses.

By teaching mathematics in the context of history, the student can fully appreciate the underlying principles. Including the mathematical history provides a uniform terminology for understanding the

[5] I prefer MEP over STEM, as being better focused on the people that do the work, organized around their scientific point of view.

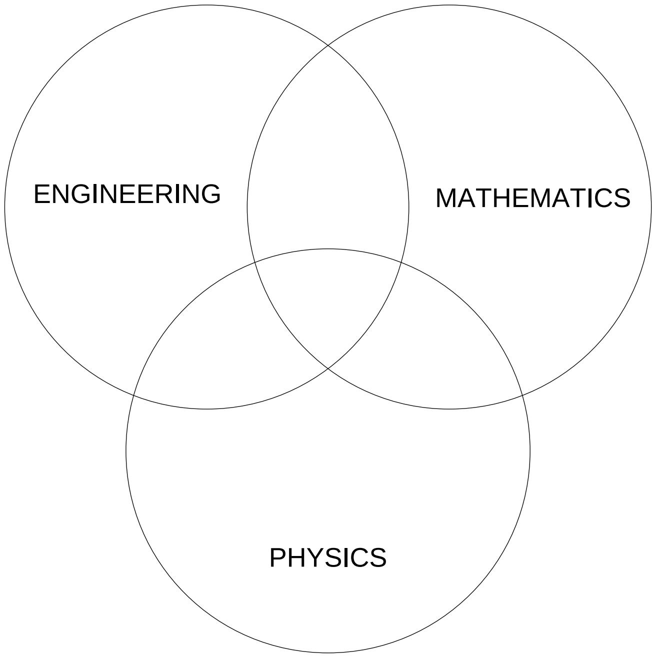


Figure 1: *There is a natural symbiotic relationship between Physics, Mathematics and Engineering, as depicted by this Venn diagram. Physics explores the boundaries. Mathematics provides the method and rigor. engineering transforms the method into technology. While these three disciplines work well together, there is poor communication due to a different vocabulary.*

fundamentals of mathematics. The present teaching method, using abstract proofs, with no (or few) figures or physical principles, by design removes intuition and the motivation that was available to the creators of these early theories. This present six semester approach does not function for many students, leaving them with a poor intuition.

Mathematics and its History (Stillwell, 2002)

Contents

1 The Theorem of Pythagoras	1
2 Greek Geometry	17
3 Greek Number Theory	37
4 Infinity in Greek Mathematics	53
5 Number Theory in Asia	68
6 Polynomial Equations	87
7 Analytic Geometry	109
8 Projective Geometry	126
9 Calculus	157
10 Infinite Series	181
11 The Number Theory Revival	203
12 Elliptic Functions	224
13 Mechanics	243
14 Complex Numbers in Algebra	275
15 Complex Numbers and Curves	295
16 Complex Numbers and Functions	313
17 Differential Geometry	335
18 NonEuclidean Geometry	359
19 Group Theory	382
20 Hypercomplex Numbers	415
21 Algebraic Number Theory	438
22 Topology	467
23 Simple Groups	495
24 Sets Logic and Computation	525
25 Combinatorics	553
Bibliography	589
Index	629

Figure 2: Table of contents of Stillwell (2002)

Chapter 1

Introduction

Much of early mathematics centered around the love of art and music, due to our sensations of light and sound. Exploring our physiological senses required a scientific understanding of vision and hearing, as first explored by Newton (1687) and Helmholtz (1863a) (Stillwell, 2010, p. 261).[1] Our sense of color and musical pitch are determined by the frequencies of light and sound. The Chinese and Pythagoreans are well known for their early contributions to music theory.

Pythagoras strongly believed that "all is integer," meaning that every number, and every mathematical and physical concept, could be explained by integral relationships. It is likely that this belief was based on Chinese mathematics from thousands of years earlier. It is also known that his ideas about the importance of integers were based on what was known about music theory in those days. For example it was known that the relationships between the musical notes (pitches) obey natural integral relationships.

image: Pythag
Newton Helmholtz

Other important modern applications of number theory are present with

- Public-private key encryption: which requires the computationally intensive factoring of large integers
- IEEE Floating point[2]

As acknowledged by Stillwell (2010, p. 16), the Pythagorean view is relevant today

> *With the digital computer, digital audio, and digital video coding everything, at least approximately into sequences of whole numbers, we are closer than ever to a world in which "all is number."*

Mersenne (1588-1647) contributed to our understanding of the relationship between the wavelength and the length of musical instruments. These results were extended by Galileo's father, and then by Galileo himself (1564-1642). Many of these musical contributions resulted in new mathematics, such as the discovery of the wave equation by Newton (c1687), followed by its one-dimensional general solution by d'Alembert (c1747).

By that time there was a basic understanding that sound and light traveled at very different speeds (thus why not the velocities of different falling weights?).

> Ole Rõmer first demonstrated in 1676 that light travels at a finite speed (as opposed to instantaneously) by studying the apparent motion of Jupiter's moon Io. In 1865, James Clerk Maxwell proposed that light was an electromagnetic wave, and therefore traveled at the speed c appearing in his theory of electromagnetism.[3]

[1] https://en.wikipedia.org/wiki/Acoustics

[2] https://en.wikipedia.org/wiki/IEEE_floating_point\#Formats

[3] https://en.wikipedia.org/wiki/Speed_of_light

Galileo famously conceptualized an experiment in 1589 where he suggested dropping two different weights from the Leaning Tower of Pisa, and showed that they must take the same time to hit the ground. Conceptually this is an important experiment, driven by a mathematical argument in which he considered the two weights to be connected by an elastic cord. This resulted in the concept of *conservation of energy*, one of the cornerstones of modern physical theory.

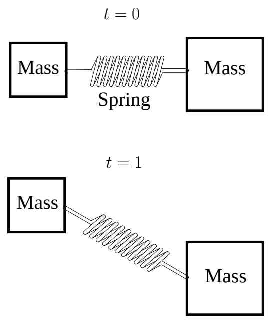


Figure 1.1: *Depiction of the argument of Galileo (unpublished book of 1638) as to why weights of different masses (weight) must fall with identical velocity. By joining them with an elastic cord they become one. Thus if the velocity were proportional to the mass, the joined masses would fall even faster. This results in a logical fallacy. This may have been the first time that the principle of* conservation of energy *was clearly stated.*

While Newton may be best known for his studies on light, he was the first to predict the speed of sound. However his theory was in error by[4] $\sqrt{c_p/c_v} = \sqrt{1.4} = 1.183$. This famous error would not be resolved for over two hundred years, awaiting the formulation of thermodynamics by Laplace, Maxwell and Boltzmann, and others. What was needed was the concept of constant-heat, or *adiabatic process*. For audio frequencies (0.02-20 [kHz]), the small temperature gradients cannot diffuse the distance of a wavelength in one cycle (Pierce, 1981; Boyer and Merzbach, 2011), “trapping” the heat energy in the wave. There were several other physical enigmas, such as the observation that sound disappears in a vacuum and that a vacuum cannot draw water up a column by more than 34 feet.

There are other outstanding examples where physiology impacted mathematics. Leonardo da Vinci is well known for his studies of the human body. Helmholtz’s theories of music and the perception of sound are excellent examples of under-appreciated fundamental mathematical contributions (Helmholtz, 1863a). Lord Kelvin (aka William Thompson),[5] was one of the first true engineer-scientists, equally acknowledged as a mathematical physicist and well known for his interdisciplinary research, knighted by Queen Victoria in 1866. Lord Kelvin coined the term *thermodynamics*, a science more fully developed by Maxwell (the same Maxwell of electrodynamics). Thus the interdisciplinary nature of science has played many key roles in the development of thermodynamics.[6] Lord Rayleigh’s book on the theory of sound (Rayleigh, 1896) is a classic text, read even today by anyone who studies acoustics.

It seems that we have detracted from this venerable interdisciplinary view of science by splitting the disciplines into into smaller parts whenever we perceived a natural educational boundary. Reforging these natural connections at some point in the curriculum is essential for the proper training of students, both scientists and engineers.[7]

WEEK 1

[4]The square root of the ratio of the specific heat capacity at constant pressure to that at constant volume

[5]Lord Kelvin was one of half a dozen interdisciplinary mathematical physicists, all working about the same time, that made a fundamental change in our scientific understanding. Others include Helmholtz, Stokes, Green, Heaviside, Rayleigh and Maxwell.

[6]Thermodynamics is another example of a course that needs reworking along historical lines.

[7]Perhaps its time to put the MEP Humpty Dumpty back together.

1.1 Early Science and Mathematics

The first 5,000 years is not well document, but the basic record is clear, as outlined in Fig. 1.2. Thanks to Euclid and later Diophantus (c250 CE), we have some limited understanding of what they studied. For example, Euclid's formula (Fig. 2.3, Eq. 2.3) provides a method for computing all Pythagorean triplets (Stillwell, 2010, pp. 4-9).

Chinese Bells and stringed musical instruments were exquisitely developed in their tonal quality, as documented by ancient physical artifacts (Fletcher and Rossing, 2008). In fact this development was so rich that one must question why the Chinese failed to initiate the industrial revolution. Specifically, why did Europe eventually dominate with its innovation when it was the Chinese who did the extensive early invention?

According to Lin (1995) this is known as the *Needham question*:

> "Why did modern science, the mathematization of hypotheses about Nature, with all its implications for advanced technology, take its meteoric rise only in the West at the time of Galileo[, but] had not developed in Chinese civilization or Indian civilization?"

Needham cites the many developments in China:[8]

> "Gunpowder, the magnetic compass, and paper and printing, which Francis Bacon considered as the three most important inventions facilitating the West's transformation from the Dark Ages to the modern world, were invented in China." (Lin, 1995)

> "Needham's works attribute significant weight to the impact of Confucianism and Taoism on the pace of Chinese scientific discovery, and emphasizes what it describes as the 'diffusionist' approach of Chinese science as opposed to a perceived independent inventiveness in the western world. Needham held that the notion that the Chinese script had inhibited scientific thought was 'grossly overrated' " (Grosswiler, 2004).

Lin was focused on military applications, missing the importance of non-military applications. A large fraction of mathematics was developed to better understand the solar system, acoustics, musical instruments and the theory of sound and light. Eventually the universe became a popular topic, and still is today.

1.1.1 Lec 1 The Pythagorean theorem

While early Asian mathematics is not fully documented, it clearly defined the course for at least several thousand years. The first recorded mathematics was that of the Chinese (5000-1200 BCE) and the Egyptians (3,300 BEC). Some of the best early record were left by the people of Mesopotamia (Iraq, 1800 BEC). Thanks to Euclid's Elements (c323 BEC) we have an historical record, tracing the progress in geometry, as defined by the Pythagorean theorem *for any right triangle*

$$c^2 = a^2 + b^2, \tag{1.1}$$

having sides of lengths (a, b, c) that are positive real numbers with $c > [a, b]$ and $a + b > c$. Solutions were likely found by trial and error rather than by an algorithm.

If a, b, c are lengths, then a^2, b^2, c^2 are areas. Equation 1.1 says that the area a^2 of a square plus the area b^2 of a square equals the area c^2 of square. Today a simple way to prove this is to compute the magnitude of the complex number $c = a + b\jmath$, which forces the right angle

$$|c|^2 = (a + b\jmath)(a - b\jmath) = a^2 + b^2.$$

[8] https://en.wikipedia.org/wiki/Joseph_Needham\#cite_note-11

However, complex arithmetic was not an option for the Greek mathematicians, since complex numbers and algebra had yet to be invented.

Almost 700 years after Euclid's *Elements*, the Library of Alexandria was destroyed (391 EC) by fire, taking with it much of the accumulated Greek knowledge. Thus one of the best technical records may be Euclid's Elements, along with some sparse mathematics due to Archimedes (c300 BEC) on geometrical series, computing the volume of a sphere, and the area of the parabola, and elementary hydrostatics. Additionally, a copy of a book by Diophantus *Arithmetic* was discovered by Bombelli (c1572) in the Vatican library (Stillwell, 2010, p. 51).

Chronological history pre 16th century **1.1.2a**

200th BCE Chinese (Primes; quadratic equation; Euclidean algorithm (GCD))
180th BCE Babylonia (Mesopotamia/Iraq) (quadratic equation)
6th BCE Pythagoras (Thales) and the Pythagorean "tribe"
4th BCE Archimedes 300BCE; Euclid (quadratic equation)

3th CE Diophantus c250CE;
4th CE Alexandria Library destroyed 391CE;
7th CE Brahmagupta (negative numbers; quadratic equation)
9th CE al-Khwārizmī (algebra) 830CE
10th CE Bhaskara (calculus) 1114-1183
15th Leonardo & Copernicus 1473-1543
16th Tartaglia (cubic eqs); Bombelli 1526-1572; Galileo Galilei 1564-1642

Time Line

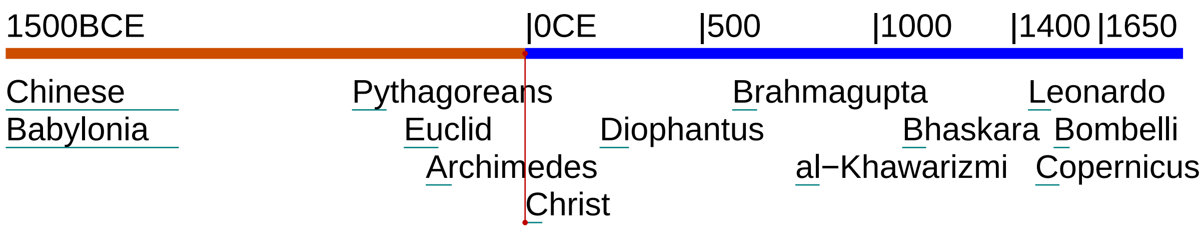


Figure 1.2: Mathematical time-line between 1500 BCE and 1650 CE.

1.1.2 Pythagorean Triplets

Well before Pythagoras, the Babylonians had tables of *Pythagorean triplets* (PTs), integers $[a, b, c]$ that obey Eq. 1.1. An example is $[3, 4, 5]$. A stone tablet (Plimpton-322) dating back to 1800 [BCE] (Fig. 1.9) was found with integers for $[a, c]$. Given such sets of two numbers, which determined a third positive integer b such that $b = \sqrt{c^2 - a^2}$, this table is more than convincing that the Babylonians were well aware of PTs, but less convincing that they had access to Euclid's formula (Eq. 1.3).

It seems likely that Euclid's Elements was largely the source of the fruitful 6th century era due to the Greek Mathematician Diophantus (Fig. 1.2), who developed the concept of *discrete mathematics*, now known as *Diophantine analysis*.

The work of Diophantus was followed by a rich mathematical era, with the discovery of 1) early calculus (Brahmagupta, 7th CE), 2) algebra (al-Khwārizmī, 9th CE), and 3) complex arithmetic (Bombelli, 15th CE). This period overlapped with the European middle (i.e., dark) ages. Presumably European intellectuals did not stop thinking during these many centuries, but what happened in Europe is presently unclear given the available records.[9]

---

[9] It might be interesting to search the archives of the monasteries, where the records were kept, to figure out what

1.1.3 What is mathematics?

Mathematics is a language, not so different from other languages. Today's mathematics is a written language with an emphasis on symbols and glyphs, biased toward Greek letters. The etymology of these symbols would be interesting to study. Each symbol is dynamically assigned a meaning, appropriate for the problem being described. These symbols are then assembled to make sentences. It is similar to Chinese in that the spoken and written version are different across dialects. In fact, like Chinese, the sentences may be read out loud in the language (dialect) of your choice, while the mathematical sentence (like Chinese) is universal.

Math is a language: It seems strange when people complain that they "can't learn math,"[10] but they claim to be good at languages. Math is a language, with the symbols taken from various languages, with a bias toward Greek, due to the popularity of Euclid's Elements. Learning a new language is fun because it opens doors to other cultures.

Math is different due to the rigor of the rules of the language, along with the way it is taught (e.g., not as a language). A third difference between math and the romance languages is that math evolved from physics, with important technical applications. This was the concept behind the Pythagorean school, a band of followers called the *Pythagoreans*. Learning languages is an advanced social skill. Thus the social outcomes are very different between learning a romance language and math. A further problem is that pre-high-school, students confuse arithmetic with math. The two topics are very different, and students need to understand this. One does not need to be good at arithmetic to be good at math (but it doesn't hurt).

There are many rules that must be mastered. These rules are defined by algebra. For example the sentence $a = b$ means that the number a has the same value as the number b. The sentence is spoken as "a equals b." The numbers are nouns and the equal sign says they are equivalent, playing the role of a verb, or action symbol. Following the rules of algebra, this sentence may be rewritten as $a - b = 0$. Here the symbols for minus and equal indicate two types of actions.

Sentences can become arbitrarily complex, such as the definition of the integral of a function, or a differential equation. But in each case, the mathematical sentence is written down, may be read out loud, has a well defined meaning, and may be manipulated into equivalent forms following the rules of algebra and calculus. This language of mathematics is powerful, with deep consequences, known as proofs.

The writer of an equation should always translate (explicitly summarize the meaning of the expression), so the reader will not miss the main point. This is simply a matter of clear writing.

Language may be thought of as mathematics (turning this idea on its head). To properly write correct English it is necessary to understand the construction of the sentence. It is important to identify the subject, verb, object, and various types of modifying phrases. If you wish to read about this, look up the distinction between the words *that* and *which*, which make a nice example of this concept. Most of us work directly with what we think "sounds right," but if you're learning English as a second language, it is very helpful to understand these mathematical rules, which are arguably easier to master than the foreign phones (i.e., speech sounds).

1.1.4 Early Physics as Mathematics

Mathematics has many functions, but basically it summarizes an algorithm (a set of rules). It was clear to Pythagoras (and many others before him), that there was an important relationship between mathematics and the physical world. Pythagoras may have been one of the first to capitalize on this relationship, using science and mathematics to design and make things.[11] This was the beginnings of technology as we know it, coming from the relationship between physics and math, impacting

happened during this strange time.

[10]"It looks like Greek to me."

[11]It is likely that the Chinese and Egyptians also did this, but this is more difficult to document.

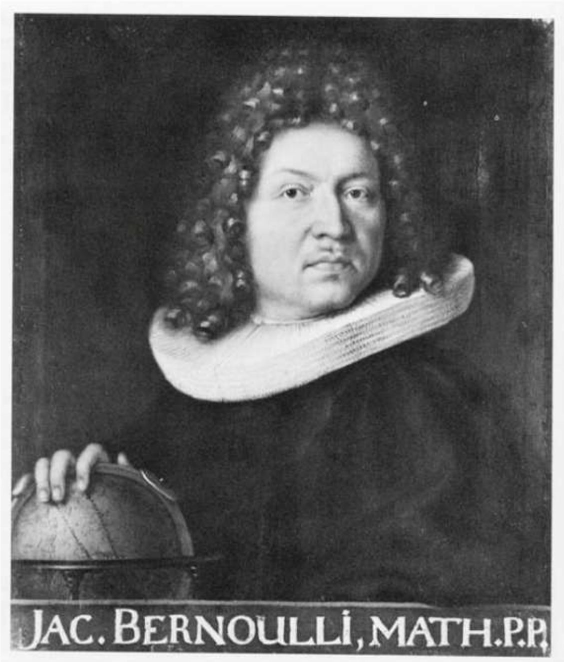


Figure 13.10: Portrait of Jakob Bernoulli by Nicholas Bernoulli

Figure 13.11: Johann Bernoulli

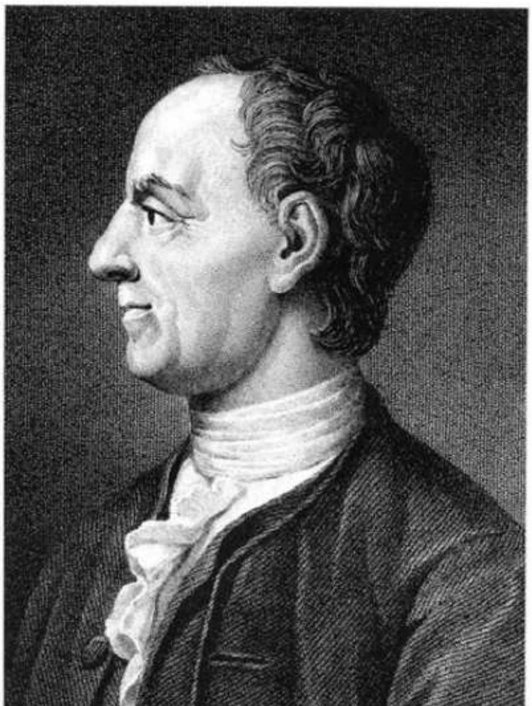


Figure 10.4: Leonhard Euler

Figure 1.3: *Above: Jakob (1655-1705) and Johann (1667-1748) Bernoulli; Below: Leonhard Euler (1707) and Jean le Rond d'Alembert (1717-1783). The figure numbers are from Stillwell (2010).*

map making, tools, implements of war (the wheel, gunpowder), art (music), sound, water transport, sanitation, secure communication, food, . . . , etc.

Why is Eq. 1.1 called a *theorem*, and what exactly needs to be proved? We do not need to prove that (a, b, c) obey this relationship, since this is a condition that is observed. We do not need to prove that a^2 is the area of a square, as this is the definition of the area of a square. What needs to be proved is that this relation only holds if the angle between the two shorter sides is $90°$.

To appreciate the significance of this development it is helpful to trace the record back to before the time of the Greeks. The Pythagorean theorem (Eq. 1.1) did not begin with Euclid or Pythagoras. Rather Euclid and Pythagoras appreciated the importance of these ideas and documented them.

In the end the Pythagoreans were destroyed by fear. This may be the danger of mixing technology and politics:

> "Whether the complete rule of number (integers) is wise remains to be seen. It is said that when the Pythagoreans tried to extend their influence into politics they met with popular resistance. Pythagoras fled, but he was murdered in nearby Mesopotamia in 497 BCE."
>
> –Stillwell (2010, p. 16)

1.1.5 The birth of modern mathematics

Modern mathematics (what we know today) was born in the 15-16th century, in the hands of Leonardo da Vinci, Bombelli, Galileo, Descartes, Fermat, and many others (Stillwell, 2010). Many of these early master were, like the Pythagoreans, secretive to the extreme about how they solved problems. They had no interest in sharing their ideas. This soon changed by Mersenne, Descartes and Newton, causing mathematics to blossom.

The amazing Bernoulli family The first individual that seems to have openly recognized the importance of mathematics, to actually teach it, was Jacob Bernoulli (Fig. 1.3). Jacob worked on what is now view as the standard package of analytic "circular" (i.e., periodic) functions: $\sin(x)$, $\cos(x)$, $\exp(x\jmath)$, $\log(x)$.[12] Eventually the full details were developed (for real variables) by Euler (Section 1.3.8 and 3.4.1).

From Fig. 1.4 we see that he was contemporary to Galileo, Mersenne, Descartes, Fermat, Huygens, Newton, and Euler. Thus it seems likely that he was strongly influenced by Newton, who in turn was influenced by Descartes,[13] Vit̀e and Wallis (Stillwell, 2010, p. 175). With the closure of Cambridge University due to the plague of 1665, Newton returned home, Woolsthorpe-by-Colsterworth (95 [mi] north of London), to worked by himself, for over a year.

Jacob Bernoulli, like all successful mathematicians of the day, was largely self taught. Yet Jacob was in a new category of mathematicians, because he was an effective teacher. Jacob taught his sibling Johann, who then taught his sibling Daniel. But most importantly, Johann taught Leonhard Euler (Figs. 1.4, 1.3), the most prolific (thus influential) of all mathematicians. This resulted in an explosion of new ideas and understanding. It is most significant that all four mathematicians published their methods and findings. Much later, Jacob studied with students of Descartes[14] (Stillwell, 2010, p. 268-9).

Euler went far beyond all the Bernoulli family, Jacob, Johann and Daniel, (Stillwell, 2010, p. 315). A special strength of Euler was the degree to which he published. First he would master a topic, and then he would publish. His papers continued to appear long after his death (Calinger, 2015).

Another individual of that time of special note, who also published extensively, was d'Alembert (Figs. 1.4, 1.3). Some of the most important tools were first proposed by d'Alembert. Unfortunately, and perhaps somewhat unfairly, his rigor was criticized by Euler, and later by Gauss (Stillwell, 2010).

[12] The log and tan functions are related by $\tan^{-1}(z) = -\frac{1\jmath}{2}\ln(\frac{1\jmath - z}{1\jmath + z})$.

[13] `https://en.wikipedia.org/wiki/Early_life_of_Isaac_Newton`

[14] It seems clear that Descartes was also a teacher.

Once the tools were being openly published, mathematics grew exponentially. It was one of the most creative times in mathematics. Figure 1.4 shows the list of the many famous names, and their relative time-line. To aid in understand the time line, note that Leonhard Euler was a contemporary of Benjamin Franklin, James Clerk Maxwell of Abraham Lincoln.[15]

Chronological history post 16^{th} century **1.1.2b**

17^{th} *Galileo 1564-1642, Kepler 1571-1630, Newton 1642-1727* Principia *1687; Mersenne; Huygen; Pascal; Fermat, Descartes (analytic geometry); Bernoullis Jakob, Johann & son Daniel*

18^{th} *Euler 1748* Student of Johann Bernoulli*; d'Alembert 1717-1783; Kirchhoff; Lagrange; Laplace; Gauss 1777-1855*

19^{th} *Möbius, Riemann 1826-1866, Galois, Hamilton, Cauchy 1789-1857, Maxwell, Heaviside, Cayley, von Helmholtz, Rayleigh*

20^{th} *Hilbert; Einstein; ...*

Time Line

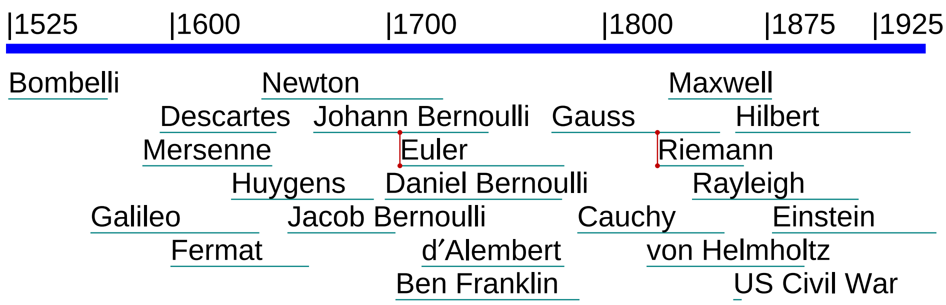


Figure 1.4: *Time-line of the four centuries from the 16^{th} and 20^{th}* CE

1.1.6 Three Streams from the Pythagorean theorem

From the outset of his presentation, Stillwell (2010, p. 1) defines "three great *streams* of mathematical thought: *Numbers, Geometry and Infinity,*" that flow from the Pythagorean theorem, as summarized in Table 1.1. Namely the Pythagorean theorem is the spring from which flow the three streams of *all* mathematics. This is a useful concept, based on reasoning not as obvious as one might think. Many factors are in play here. One of these was the strongly held opinion of Pythagoras that all mathematics should be based on integers. The rest are tied up in the long, necessarily complex history of mathematics, as best summarized by the Fundamental theorems, which are each discussed in detail in the appropriate chapter.

Stillwell's concept of three streams following from the Pythagorean theorem is the organizing principle behind the this book, organized by chapter:

1. *Introduction* (Chapter 1) A detailed overview of the fundamentals and the three streams are presented in Sections 1.2–1.5.

2. *Number Systems* (Chapter 2: Stream 1) Fundamentals of number systems, starting with prime numbers, through complex numbers, vectors and matrices.

3. *Algebraic Equations* (Chapter 3: Stream 2) Algebra and its development, as we know it today. The theory of real and complex equations and functions of real and complex variables. Complex impedance $Z(s)$ of complex frequency $s = \sigma + \omega\jmath$ is covered with some care, given its importance for engineering mathematics.

[15] Lincoln traveled through Mahomet IL (where I live) on his way to the Urbana Court house.

4. *Scalar Calculus* (Chapter 4: Stream 3a) Ordinary differential equations. Integral theorems. Acoustics.

5. *Vector Calculus*: (Chapter 5: Stream 3b) Vector Partial differential equations. Gradient, divergence and curl differential operators. Stokes, and Green's theorems. Maxwell's equations.

Table 1.1: *Three streams followed from Pythagorean theorem: Number Systems (Stream 1), Geometry (Stream 2) and Infinity (Stream 3).*

- *The Pythagorean Theorem is the mathematical spring which bore the three streams.*
- $\approx$*Several centuries per stream:*

  1) **Numbers:**
  - *6^{th}BC* $\mathbb{N}$ *counting numbers,* $\mathbb{Q}$ *(Rationals),* $\mathbb{P}$ *Primes*
  - *5^{th}BC* $\mathbb{Z}$ *Common Integers,* $\mathbb{I}$ *Irrationals*
  - *7^{th}CE* *zero* $\in \mathbb{Z}$

  2) **Geometry:** *(e.g., lines, circles, spheres, toroids, ...)*
  - *17^{th}CE* *Composition of polynomials (Descartes, Fermat)*
    Euclid's Geometry + algebra $\Rightarrow$ Analytic Geometry
  - *18^{th}CE* *Fundamental Theorem of Algebra*

  3) **Infinity:** *($\infty \rightarrow$ Sets)*
  - *17-18^{th}CE* $\mathbb{F}$ *Taylor series, Functions, Calculus (Newton)*
  - *19^{th}CE* $\mathbb{R}$ *Real,* $\mathbb{C}$ *Complex 1851*
  - *20^{th}CE* *Set theory*

1.2 Stream 1: Number Systems

This era produced a new stream of fundamental theorems. A few of the individuals who played a notable role in this development, in chronological (birth) order, include Galileo, Mersenne, Newton, d'Alembert, Fermat, Huygens, Descartes and Helmholtz. These individuals were some of the first to develop the basic ideas, in various forms, that were then later reworked into the proofs, that today we acknowledge as *The fundamental theorems of mathematics*.

Number theory (discrete, i.e., integer mathematics) was a starting point for many key ideas. For example, in Euclid's geometrical constructions the Pythagorean theorem for $\{a, b, c\} \in \mathbb{R}$ was accepted as true, but the emphasis in the early analysis was on integer constructions, such as Euclid's formula for Pythagorean triplets (Eq. 1.3, Fig. 2.3) k As we shall see, the Pythagorean theorem is a rich source of mathematical constructions, such as composition of polynomials, and solutions of Pell's equation by eigenvector and recursive analysis methods. Recursive difference equation solutions predate calculus, at least going back to the Chinese (c2000 BCE). These are early (pre-limit) forms of differential equations, best analyzed using an eigenfunction expansion (Appendix D), a powerful geometrical concept from linear algebra, of an expansion in terms of an orthogonal set of normalized (unit-length) vectors.

The first use of zero and ∞: It is hard to imagine that one would not appreciate the concept of zero and negative numbers when using an abacus. If five beads are moved up, and one is moved down, then four are left. Then if four more are move down, that leaves zero. Taking away is the opposite of addition, and taking away from four to get zero beads, is no different than taking four away from zero, to get negative four beads. Subtraction, the inverse of addition, seems like an obvious idea, on an abacus.

However, understanding the concept of zero and negative numbers is not the same as having a symbolic notation. The Roman number system had no such symbols. The first recorded use of a symbol for zero is said to be by Brahmagupta in 628 CE.[16] Thus it does not take much imagination to go from counting numbers $\mathbb{N}$ to the set of all integers $\mathbb{Z}$, including zero, but apparently it takes 600 years to develop a terminology that represents these ideas. Defining the rules of subtraction required the creation of algebra c830 CE (Fig. 1.2). The concept that caused far more difficulty was ∞. Until Riemann's thesis in 1851 it was not clear if ∞ was a number, many numbers, or even definable.

1.2.1 Lec 2: The Taxonomy of Numbers: $\mathbb{P}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{I}, \mathbb{R}, \mathbb{C}$

Once symbols for zero and negative numbers were defined (and accepted), progress was made. In a similar manner, to fully understand numbers, a transparent notation is required. First one must differentiate between the different classes (genus) of numbers, providing a notation that defines each of these classes, along with their relationships. It is logical to start with the most basic *counting numbers*, which we indicate with the double-bold symbol $\mathbb{N}$. All the double-bold symbols and their genus are summarized in Appendix A.

Counting numbers $\mathbb{N}$: These are known as the "natural numbers" $\{1, 2, 3, \cdots\} \in \mathbb{N}$, denoted by the double-bold symbol $\mathbb{N}$. For increased clarity we shall refer to the natural numbers as *counting numbers*, to clarify that *natural* means *integer*. The mathematical sentence $2 \in \mathbb{N}$ is read as *2 is a member of the set of counting numbers.* The word *set* means the *sharing of a specific property.*

Primes $\mathbb{P}$: A prime number $\mathbb{P} \subset \mathbb{N}$ (set $\mathbb{P}$ is a subset of $\mathbb{N}$) is an integer that may not be factored, other than by 1 and itself. Since $1 = 1 \cdot 1$, $1 \not\in \mathbb{P}$, as it is seen to violate this basic definition of a prime. Prime numbers $\mathbb{P}$ are a *subset* of the counting numbers ($\mathbb{P} \subset \mathbb{N}$). We shall use the convenient notation π_n for the prime numbers, indexed by $n \in \mathbb{N}$. The first 12 primes ($n = 1, \ldots, 12$) are $\pi_n = 2$, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37. Since, $4 = 2^2$ and $6 = 2 \cdot 3$ may be factored, $\{4, 6\} \not\in \mathbb{P}$ (read as: *4 and 6 are not in the set of primes*). Given this definition, multiples of a prime, i.e., $n\pi_k \equiv [2, 3, 4, 5, \ldots times\pi_k$ of any prime π_k, cannot be prime. It follows that all primes except 2 must be odd and every integer N is unique in its factorization.

Coprimes are number whose factors are distinct (they have no common factors). Thus 4 and 6 are not coprime, since they have a common factor of 2, whereas $21 = 3 \cdot 7$ and $10 = 2 \cdot 5$ are coprime. By definition all distinct primes are coprime. The notation $m \perp n$ indicates that m, n are coprime.

The *Fundamental Theorem of Arithmetic* states that all integers may be uniquely expressed as a product of primes. The *Prime Number Theorem* estimates the mean density of primes over $\mathbb{N}$.

Integers $\mathbb{Z}$: These include positive and negative counting numbers and zero. Notionally we might indicate this using *set notation* as $\mathbb{Z} : \{-\mathbb{N}, 0, \mathbb{N}\}$. Read this as *The integers are in the set composed of the negative of the natural numbers ($-\mathbb{N}$), zero, and counting numbers* $\mathbb{N}$. Note that $\mathbb{N} \subset \mathbb{Z}$.

Rational numbers $\mathbb{Q}$: These are defined as numbers formed from the ratio of two integers. Since the integers $\mathbb{Z}$ include 1, it follows that integers are a subset of rational numbers ($\mathbb{Z} \subset \mathbb{Q}$). For example the rational number $3/1 \in \mathbb{Z}$). The main utility of rational numbers is that that they can efficiently approximate any number on the real line, to any precision. For example $\pi \approx 22/7$ with a relative error of $\approx$0.04%. Of course, if the number is rational the error is zero.

Fractional number $\mathbb{F}$: The utility of rational numbers is their power to approximate irrational numbers ($\mathbb{R} \not\subset \mathbb{Z}$). It follows that a subset of the rationals, that excludes the integers, has great value. We call these numbers *Fractional numbers* and assign them the symbol $\mathbb{F}$. They are defined as the subset of rationals that are not integers. From this definition $\mathbb{F} \perp \mathbb{Z}$, $\mathbb{F} \subset \mathbb{Q} = \mathbb{Z} \cup \mathbb{F}$. Because of their

[16] The fall of the Roman Empire was Sept. 4, 476.

approximating property, the fractional set $\mathbb{F}$ represent the most important (and the largest) portion of the rational numbers, dwarfing the size of the integers, another good reason for defining the two distinct subsets.

Once factored and common factors canceled, the subset $\mathbb{F} \subset \mathbb{Q}$ of rational numbers is always the ratio of coprimes. For example $\pi \approx 22/7 = 11 \cdot 2/7 = 3 + 1/7$ with $22 \perp 7$, and $9/6 = 3/2 = 1 + 1/2$ with $3 \perp 2$.[17]

Irrational numbers $\mathbb{I}$: Every real number that is not rational ($\not\mathbb{Q}$) is *irrational* ($\mathbb{Q} \perp \mathbb{I}$). Irrational numbers include π, e and the square roots of most integers (i.e., $\sqrt{2}$). These are decimal numbers that never repeat, thus requiring infinite precision in their representation.

Irrational numbers ($\mathbb{I}$) were famously problematic for the Pythagoreans, who incorrectly theorized that all numbers were rational. Like ∞, irrational numbers require a new and difficult concept before they may even be defined: They were not in the set of fractional numbers ($\mathbb{I} \not\subset \mathbb{F}$). It was easily shown, from a simple geometrical construction, that most, but not all of the square roots of integers are irrational. It was essential to understand the factorization of counting numbers before the concept of irrationals could be sorted out.

Real numbers $\mathbb{R}$: Reals are the union of rational and irrational numbers, namely $\mathbb{R} : \{\mathbb{I}, \mathbb{Q}\}$ ($\mathbb{R} = \mathbb{Z} \cup \mathbb{F} \cup \mathbb{I}$). Reals are the lengths in Euclidean geometry. Many people assume that *IEEE 754 floating point numbers* (c1985) are real (i.e., $\in \mathbb{R}$). In fact they are rational ($\mathbb{Q} : \{\mathbb{F} \cup \mathbb{Z}\}$) approximations to real numbers, designed to have a very large dynamic range. There can be no machine realization of irrational numbers, since such a number would require infinite precision (∞ bits). The hallmark of fractional numbers ($\mathbb{F}$) is their power in making highly accurate approximations of any real number.

Using Euclid's compass and ruler methods, one can make line length proportionally shorter or longer, or (approximately) the same. A line may be made be twice as long, an angle bisected. However, the concept of an integer length in Euclid's geometry was not defined.[18] Nor can one construct an imaginary or complex line as all lines are assumed to be real.

Real numbers were first fully accepted only after set theory was developed by Cantor (1874) (Stillwell, 2010, pp. 461, 525...). It seems amazing, given how widely accepted real numbers are today. But in some sense they were accepted by the Greeks, as lengths of real lines.

Complex numbers $\mathbb{C}$: Complex numbers are best defined as *ordered pairs of real numbers*.[19] They are quite special in engineering mathematics, since roots of polynomials having either real or complex coefficients may be complex. The best known example is the quadratic formula for the roots of a 2^d degree polynomial, with either real or complex coefficients.

The common way to write a complex number is using the common notation $z = a + b\jmath \in \mathbb{C}$, where $a, b \in \mathbb{R}$. Here $1\jmath = \sqrt{-1}$. We also define $1\imath = -1\jmath$ to account for the two possible signs of the square root. Accordingly $1\jmath^2 = 1\imath^2 = -1$.

Multiplication of complex numbers follows the rules of real algebra, similar to multiplying two polynomials. Multiplication of two first degree polynomials gives

$$(a + bx)(c + dx) = ac + (ad + bc)x + bdx^2$$

If we substitute $1\jmath$ for x, and use the definition $1\jmath^2 = -1$, we obtain the product of the two complex numbers

$$(a + b\jmath)(c + d\jmath) = ac - bd + (ad + bc)\jmath.$$

[17]HW problem: How to define $\mathbb{F}$ given two integers $(n, m) \subset Z$? Sol: Not sure how to approach this, but it seems like a fun problem. Here two simple methods that do *not* work: (1) One cannot define $\mathbb{F}$ as the ratio $x = n/m$, since given $m = 1, x \in \mathbb{Z}$. (2) One cannot define $\mathbb{F}$ as the ratio of two coprimes, since then $x = 1/m \notin \mathbb{F}$ (since $1 \not\mathbb{F}$).

[18]As best I know.

[19]A polynomial $a + bx$ and a 2-vector $[a, b]^T = \begin{bmatrix} a \\ b \end{bmatrix}$ are also examples of ordered pairs.

Thus multiplication of complex numbers obeys the accepted rules of algebra.

Polar representation: A alternative for complex multiplication is to work with polar coordinates. The polar form of complex number $z = a + b\jmath$ is written in terms of its magnitude $\rho = \sqrt{a^2 + b^2}$ and angle $\theta = \angle z = \tan^{-1}(z) = \arctan z$, as $z = \rho e^{\theta\jmath}$. From the definition of the complex natural log function

$$\ln \rho e^{\theta\jmath} = \ln \rho + \theta\jmath,$$

which is useful in engineering calculations.[20]

Matrix representation: A second alternative and useful way to represent complex numbers is in terms of 2x2 matrices. This relationship is defined by the mapping from a complex number to a 2x2 matrix (along with several examples)

$$a + b\jmath \leftrightarrow \begin{bmatrix} a & -b \\ b & a \end{bmatrix}, \quad 2 \leftrightarrow \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad -\jmath \leftrightarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad e^{\theta\jmath} \leftrightarrow \begin{bmatrix} \sin(\theta) & -\cos(\theta) \\ \cos(\theta) & \sin(\theta) \end{bmatrix}$$

You might verify that

$$\frac{a + b\jmath}{c + d\jmath} = \frac{ab + bd + (bc - ad)\jmath}{c^2 + d^2} \longleftrightarrow \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & -d \\ d & c \end{bmatrix}^{-1} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & d \\ -d & c \end{bmatrix} \frac{1}{c^2 + d^2}.$$

By taking the inverse of the 2x2 matrix one can define the ratio of one complex number by another, Until you try out this representation, it may not seem obvious, or even that it could work.

This representation proves that $1\jmath$ is not necessary to define a complex number. What $1\jmath$ can do is simplify the algebra, both conceptually and for numerical results. It is worth your time to become familiar with the matrix representation, to clarify any possible confusions you might have about multiplication and division of complex numbers. This matrix representation can save you time, heartache and messy algebra. Once you have learned how to multiply two matrices, it's a lot simpler than doing the complex algebra. In many cases we will leave the results of our analysis in matrix form, to avoid the algebra altogether.[21] More on this topic may be found in Chapter 2.

Real versus complex numbers: All numbers may be viewed as complex. Namely every real number is complex if we take the imaginary part to be zero (Boas, 1987). For example, $2 \in \mathbb{P} \subset \mathbb{C}$. Likewise every purely imaginary number (e.g., $0 + 1\jmath$) is complex with zero real part. It follows that $2\jmath \in \mathbb{P}\jmath$. Integers are a subset of reals, which are a subset of complex numbers[22] *Gaussian integers* are complex integers ($\mathbb{Z} \subset \mathbb{R} \subset \mathbb{C}$).[23] From the above discussion it should be clear that each of these different classes of number are nested in a hierarchy, in the following embeddings

$$\pi_k \in \mathbb{P} \subset \mathbb{N} \subset \mathbb{Z} \subset \mathbb{Z} \cup \mathbb{F} = \mathbb{Q} \subset \mathbb{Q} \cup \mathbb{I} = \mathbb{R} \subset \mathbb{C}.$$

The integers $\mathbb{Z}$ and fractionals $\mathbb{F}$ split the rationals ($\mathbb{Q} : \mathbb{Z} \cup \mathbb{F}$, $\mathbb{Z} \perp \mathbb{F}$), each of which is a subset of the rationals ($\mathbb{Z} \in \mathbb{Q}$, $\mathbb{F} \subset \mathbb{Q}$). The rationals $\mathbb{Q}$ and irrationals $\mathbb{I}$ split the reals ($\mathbb{R} : \mathbb{Q} \cup \mathbb{I}$, $\mathbb{Q} \perp \mathbb{I}$), each of which is a subset of the reals ($\mathbb{Q} \in \mathbb{R}$, $\mathbb{I} \in \mathbb{R}$).

The roots of polynomials x_k are complex ($x_k \in \mathbb{C}$), independent of the genus of the coefficients (e.g., real integer coefficients give rise to complex roots). Each genus plays an important role in algebra, with prime numbers at the bottom (root of the tree) and complex numbers at the top. We shall explore this further in Chaps. 2 and 3.

---

[20] Chapter 2 discusses the definition of the phase, i.e., how is it computed (i.e., $\arctan(e^{\theta\jmath})$, `arctan2(x,y)`), and the importance of the unwrapped phase, as in the example $\delta(t - \tau) \leftrightarrow e^{-\tau\jmath}$.

[21] Sometimes we let the computer do the final algebra, numerically, as 2x2 matrix multiplications.

[22] The plural *complexs* (a double /s/) seems an unacceptable word in English.

[23] It follows that integers are a subset of Gaussian integers (the imaginary or real part of the Gaussian integer may be zero).

Finally, note that complex numbers $\mathbb{C}$ do not have "order," meaning one cannot be larger, smaller or equal to another. It makes no sense to say that $\jmath > 1$ or $\jmath = 1$ (Boas, 1987). The real and imaginary parts and the magnitude and phase have order.

History of complex numbers: It is notable how long it took for complex numbers to be accepted (1851), relative to when they were first introduced by Bombelli (16^{th} century CE). In fact, complex integers (aka, *Gaussian integers*) were accepted before non-integral complex numbers. Apparently real numbers ($\mathbb{R}$) were not accepted (i.e., proved to exist, thus mathematically defined) until even later. It took the development of *set theory* in the late 19^{th} century to sort out a proper definition of the real number, due to the existence of irrational numbers.

Computer Representations of $\mathbb{I}, \mathbb{R}, \mathbb{C}$: When doing numerical work, one must consider how we may compute within the family of reals (i.e., irrationals). There can be no irrational number representation on a computers. IEEE floating point numbers, which are the international standard of computation, are actually rational approximations. The mantissa and the exponent are each integers, having sign and magnitude. The size of each integer depends on the precision of the number being represented. An IEEE floating-point number is rational because it has a mantissa integer multiplied by a base to the power of an exponent integer.

Floating point numbers contain irrational numbers, which must be approximate by rational numbers. This leads to the concept of *fractional representation*, which requires the definition of the *mantissa*, *exponent* and *base*. Numerical results must not depend on the base. For example, when using base ten[24] Make a table

$$\pi \cdot 10^5 \approx 314\,159.27\ldots = 3 \cdot 10^5 + 1 \cdot 10^4 + 4 \cdot 10^3 + \cdots + 9 \cdot 10^0 + 2 \cdot 10^{-1} \cdots .$$

According to Matlab's DEC2BIN() routine, the binary representation is ck 4 errors

$$\pi \cdot 2^{17} \approx 131072_{10} \cdot 22/7 = 110,0100,1001,0010,0101_2,$$

where 1 and 0 are multipliers of powers of 2, which are then added together as follows

$$2^{18} + 2^{17} + 2^{14} + 2^{11} + 2^8 + 2^5 + 2^2 + 2^0.$$

In base 16 (i.e, hexadecimal) $2^{17} \cdot 22/7 = 2^{18} \cdot 8_{16}/7_{16}$.

One may keep track of the decimal point using the exponent, which in this case is a factor of 2^{17} $= 131072_{10}$. The concept of a number having a decimal point is replaced by an integer, having the desired precision, and a scale factor of any base (radix). This scale factor may be thought of as moving the decimal point to the right (larger number) or left (smaller number). The mantissa "fine-tunes" the value about a scale factor (the exponent).

Here is $x = 2^{17} \times 22/7$ at IEEE 754 full double precision, as computed by an IEEE-754 floating point converter[25] $x = 411940.5625_{10} = 2^{54} \times 1198372 = 010010,001,10010,010010,010010,010010_2 =$ $0x48c92492_{16}$. The commas in the binary string of ones and zeros, are to help visualize the quasi-periodic nature of the bit-stream. The mantissa is 4793490_{10} and the exponent is 2^{18}. The numbers are stored in a 32 bit format, with 1 bit for sign, 8 bits for the exponent and 23 bits for the mantissa. Perhaps a more instructive number is $x = 4793490.0 = 01,001,010,100,100,100,100,100,100,100,100_2$ $= 0x4a,924,924_{16}$ which has a repeating binary bit pattern of $((100))_3$, only broken by the scale factor $0x4a$. Another with even higher symmetry is $x = 6.344,131,191,146,9 \times 10^{-17} = 0x24,924,924_{16} =$ $00,100,100,100,100,100,100,100,100,100,100_2$. In this example the repeating pattern is clear in the Hex representation as a repeating $((942))_3$. As before, the commas are to help with readability, and have no other meaning.

[24] Base 10 is the natural world-wide standard simply because we have 10 fingers.

[25] http://www.h-schmidt.net/FloatConverter/IEEE754.html

There are other important types of representations. As pairs of reals, complex numbers have similar approximate representations. An important representations of complex numbers is $e^z = \cosh(z) + j\sinh(z)$, which includes the famous formula of Euler $e^{j\theta} = \cos\theta + j\sin\theta$. Some of these concepts can be generalized to include vectors, matrices and polynomials.

Integers and the Pythagoreans The integer is the corner stone of the Pythagorean doctrine, so much so that it caused a fracture within the Pythagoreans when it was discovered that not all numbers are rational. The famous example is the isosceles triangle $1, 1, \sqrt{2}$, which lead to the next triangle $[1, 2, \sqrt{3}]$, etc. This is known as the Spiral of Theodorus: the short side is 1 and the hypotenuse is extended by one, using a simple compass-ruler construction.

There are right-triangles with integral lengths, the best known being $[3, 4, 5]$. Such triplets of integers $[a, b, c]$ that satisfy the Pythagorean formula (Eq. 1.1) are denoted *Pythagorean triplets*, which may be verified using Euclid's formula (Eq. 1.3).

To form triangles with perfect $90°$ angles, the lengths need to satisfy Eq. 1.1. Such triangles are also useful in constructing buildings or roads made from of bricks having a uniform size.

Public-private key Security: An important application of prime numbers is public-private key (RSA) encryption, essential for internet security applications (e.g., online banking). To send secure messages the security (i.e., utility) of the internet is dependent on key encryption.[26] Most people assume this is done by a personal login and passwords. Passwords are simply not secure, for many reasons. The proper method depends on factoring integers formed from products of primes having thousands of bits.[27] The security is based on the relative ease in multiplying large primes, but the virtual impossibility of factoring them.

When a computation is easy in one direction, but its inverse is impossible, it is called a *trap-door function*. We shall explore the reasons for this in Chapter 2. If everyone switched from passwords to public key encryption, the internet would be much more secure.

Puzzles: A third application of integers are imaginative problems that use integers. An example is the classic Chinese *Four stone problem:* "Find the weight of four stones that can be used with a scale to weigh any object (e.g., salt, gold) between 0, 1, 2, ..., 40 [gm]." As with the other problems, the answer is not as interesting as the method, since the problem may be easily recast into a related one. This type of problem can be found in airline magazines as entertain on a long flight. The solution to this problem is best cast as a linear algebra problem, with integer solutions. Again, once you know the trick, it is "easy."[28]

1.2.2 Lec 3: The role of physics in mathematics

Bells, chimes and Eigenmodes Integers naturally arose in art, music and science. An example are the relations between musical notes, the natural eigenmodes (tones) of strings and other musical instruments. These relations were so common and well studied, it appeared that to understand the physical world (aka, the Universe), one needed to understand integers. This was a seductive view, but not actually correct. As will be discussed in Sections 1.3.1 and 3.1.1, it is best to view the relationship between acoustics, music and mathematics as historical, since these topics played such an important role in the development of mathematics. Also interesting is the role that integers seem to play in quantum mechanics, for much the same reasons.

[26] One might say this is either: i) a key application of primes, or ii) it is primary application of keys. Its a joke.

[27] It would seem that public key encryption could work by having two numbers with a common prime, and then by using Euclidean algorithm, that GCD could be worked out. One of the integers could be the public key and the second could be the private key. Given the difficulty of factoring the numbers into their primes, and ease of finding the GCD using Euclidean algorithm, a practical scheme may be possible. Ck this out.

[28] When ever someone tells you something is "easy," you should immediately appreciate that it is very hard, but there is a concept, that once you learn, the difficulty evaporates.

Engineers are so accustomed to working with real (or complex) numbers, the distinction between real (i.e., irrational) and fractional numbers are rarely acknowledged. Integers on the other hand arise in many contexts. One cannot master programming computers without understanding integer, hexadecimal, octal, and binary representations, since all numbers in a computer are represented in numerical computations in terms of rationals ($\mathbb{Q} = \mathbb{Z} \cup \mathbb{F}$).

As discussed in Section 1.2.1, the primary reason integers are so important is their absolute precision. Every integer $n \in Z$ is unique,[29] and has the *indexing property*, which is essential for making lists that are ordered, so that one can quickly look things up. The alphabet also has this property (e.g., a book's index). Other than for hexadecimal numbers, which for notional reasons use the alphabet, letters are equivalent to integers.

Because of the integer's absolute precision, the digital computer overtook the analog computer, once it was practical to make logic circuits that were fast. The first digital computer was thought to be the *Eniac* at the University of Pennsylvania, but it turned out that the code-breaking effort in Bletchley Park, England, under the guidance of Alan Turing, created the first digital computer (The Colossus) to break the WWII German "Enigma" code. Due to the high secrecy of this war effort, the credit was only acknowledged in the 1970s when the project was declassified.

There is zero possibility of analog computing displacing digital computing, due to the importance of precision (and speed). But even with binary representation, there is a non-zero probability of error, for example on a hard drive, due to physical noise. To deal with this, error correcting codes have been developed, to reduce the error by several orders of magnitude. Today this is a science, and billions of dollars are invested to increase the density of bits per area, to increasingly larger factors. A few years ago the terabyte drive was unheard of; today it is the standard. In a few years petabyte drives will certainly become available. It is hard to comprehend how these will be used by individuals, but they are essential for on-line (cloud) computing.

Fundamental theorems

Modern mathematics is build on a hierarchical construct of *fundamental theorems*, as summarized in Table 1.2. The importance of such theorems cannot be overemphasized. Every engineering student needs to fully appreciate the significance of these key theorems. If necessary, memorize them. But that will not do over the long run, as each and every theorem must be fully understood. Fortunately most students already know several of these theorems, but perhaps not by name. In such cases, it is a matter of mastering the vocabulary.

The theorems are naturally organized, starting with two theorems on prime numbers (Table 1.2). These may also be thought of in terms of Stillwell's three streams. For Stream 1 there is the *Fundamental Theorem of Arithmetic* and the *Prime Number Theorem*. For Stream 2 there is the *Fundamental Theorem of Algebra* and Bézout's theorem, while for Stream 3 there are a host of theorems on calculus, ordered by their dimensionality. Some of these theorems verge on trivial (e.g., the Fundamental Theorem of Arithmetic). Others are more challenging, such as the *Fundamental Theorem of Vector Calculus* and *Green's theorem*.

Complexity should not be confused with importance. Each of these theorems is, as stated, fundamental. Taken as a whole, they are a powerful way of summarizing mathematics.

Stream 1: Prime Number theorems:

There are two fundamental theorems about primes,

1. *The Fundamental Theorem of Arithmetic*: This states that every counting number $n > 1 \in \mathbb{N}$ may be uniquely factored into prime numbers. This raises the question of the meaning of *factor* (split into a product).

[29] Check out the history of $1729 = 1^3 + 12^2 = 9^3 + 10^3$.

Table 1.2: *The Fundamental theorems of mathematics*

1. Fundamental theorems of:
   (a) **Number Systems: Stream 1**
      - Arithmetic
      - Prime Number
   (b) **Geometry: Stream 2**
      - Algebra
      - Bézout
   (c) **Calculus: Stream 3**[a]
      - Leibniz $\mathbb{R}^1$
      - Complex $\mathbb{C} \subset \mathbb{R}^2$
      - vectors $\mathbb{R}^3, \mathbb{R}^n, \mathbb{R}^\infty$
2. Other key concepts:
   - Complex analytic functions (complex roots are finally accepted!)
      - Complex Taylor Series (complex analytic functions)
      - Region of convergence (ROC) of complex analytic series
      - Laplace transform, and its inverse
      - Causal time $\implies$ complex frequency s
      - Cauchy Integral Theorem
      - Residue integration (i.e., Green's Thm in $\mathbb{R}^2$)
   - Riemann mapping theorem (Gray, 1994; Walsh, 1973)
   - Complex Impedance (Ohm's Law) Kennelly

[a] Flanders, Harley (June–July 1973). "Differentiation under the integral sign." *American Mathematical Monthly* 80 (6): 615-627. doi:10.2307/2319163. JSTOR 2319163.

2. *The Prime Number Theorem:* One would like to know how many primes there are. That is easy: $|\mathbb{P}| = \infty$. (The *cardinality*, or size of the set of primes, is infinite). The proper way of asking this questions is *What is the average density of primes, in the limit as $n \to \infty$?* This question was answered, for all practical purposes, by Gauss, who as a pastime computed the first million primes by hand. He discovered that, to a good approximation, the primes are equally likely on a log scale. This is nicely summarized by the jingle attributed to the mathematician Pafnuty Chebyshev

> Chebyshev said, and I say it again: *There is always a prime between n and 2n.*

(Stillwell, 2010, p. 585)

When the ratio (interval) of two frequencies (pitch) is 2, the relationship is called an *octave*. Thus we might say there is at least one prime per octave. This makes on wonder about the maximum number of primes per octave. In modern music the octave is further divided into 12 intervals called *semitones* (factors), equal to the $\sqrt[12]{2}$. The product of 12 semitones is an octave. Thus we must wonder how many primes there is per semitone?

Stream 2: Fundamental theorem of Algebra

This theorem states that every polynomial has at least one root. When that root is removed then the degree of the polynomial is reduced by 1. Thus when applied recursively, a polynomial of degree N has N roots.

Besides the Fundamental Theorem of Algebra, a second important theorem is Bézout's theorem, which is a generalization of the Fundamental Theorem of Algebra. It says[30] that the *composition* of two polynomials has degree equal to the product of the degrees of each polynomial. For example, if $P_3(x) = x^3$ and $P_5(x) = x^5$, then $P_3(P_5)(x) = (x^5)^3 = x^{15}$. It further states that when counting the N roots of a polynomial of degree N, one must include the imaginary roots, double roots and roots at infinity, some of which may difficult to identify.

Stream 3: Fundamental theorems of calculus

In Sections 1.5.3, 1.5.5 and 5.1.3 we will deal with each of the theorems for Stream 3, where we consider the several Fundamental theorems of integration, starting with Leibniz's formula for integration on the real line ($\mathbb{R}$), then progressing to complex integration in the complex plane ($\mathbb{C}$) (*Cauchy's theorem*), which is required for computing the inverse Laplace transform. Then we discuss Gauss' and Stokes' Laws for $\mathbb{R}^2$, with closed and open surfaces. One cannot understand Maxwell's equations, fluid flow, or acoustics without understanding these theorems. Any problem that deals with the wave equation in more than one dimension requires an understanding of these concepts. The derivation of the Kirchhoff voltage and current laws is based on these theorems.

Other key concepts

Besides the widely recognized fundamental theorems for the three streams, there are a number of equally important theorems that have not yet been labeled as "fundamental."[31]

The widely recognized *Cauchy Integral Theorem* is an excellent example, since it is a stepping stone to the *Fundamental Theorem of Complex Integral Calculus*. In Chapter 4 we clarify the contributions of each of these special theorems.

Once these Fundamental theorems of integration (Stream 3) have been mastered, the student is ready for the *complex frequency domain*, which takes us back to Stream 2 and the *complex frequency plane* ($s = \sigma + \omega \jmath \in \mathbb{C}$). While the Fourier and Laplace transforms are taught in Mathematics courses, typically few physical connections are made, accordingly the concept of *complex frequency* is rarely mentioned. The *complex frequency domain* and *causality* are fundamentally related, and critical for the analysis of signals and systems.

Without the concept of time and frequency, one cannot develop an intuition for the Fourier and Laplace transform relationships, especially within the context of engineering and mathematical physics.

WEEK 2

1.2.3 Lec 4: Prime numbers

If someone came up to you and asked for a theory of counting numbers, I suspect you would look them in the eye with a blank stare, and start counting. It sounds like either a bad joke or a stupid question. Yet integers are rich topic, so the question is not even slightly dumb. It is somewhat amazing that even birds and bees can count. While I doubt birds and bees can recognize primes, cicadas and other

[30] Statements of the theorem speak of *intersections* and constructions of curves, rather than compositions. I find this somewhat confusing. For example, how does intersection differ from elimination, or construction from composition (Stillwell, 2010, p. 119)?

[31] It is not clear what it takes to reach this more official sounding category.

insects only crawl out of the ground in multiples of prime years, (e.g., 13 or 17 year cycles). If you have ever witnessed such an event (I have), you will never forget it. Somehow they know. Finally, there is an analytic function, first introduced by Euler, based on his analysis of the Sieve, now known as the *Riemann zeta function* $\zeta(s)$, which is complex analytic, with its poles at the logs of the prime numbers. The exact relationship between the primes and the poles will be discussed in Sections 1.4.11 and 4.5.2. The properties of this function are truly amazing, even fun. It follows that primes are fundamental properties of the counting numbers, that the theory of numbers (and primes) is an important topic of study. Many of the questions, and some of the answers, go back to at least the time of the Chinese (Stillwell, 2010).

The most relevant question at this point is "Why are integers so important?" We addressed this question in Section 1.2.9. First we count with them, so we can keep track of "how much." But there is much more to numbers than counting: We use integers for any application where absolute accuracy is essential, such as banking transactions (making change), and precise computing of dates (Stillwell, 2010, p. 70) or location (I'll meet you at location L$\in \mathbb{N}$ at time T$\in \mathbb{N}$), building roads or buildings out of bricks (objects of a uniform size). If you go to 34^{th} street and Madison and they are at 33th and Madison, that's a problem. To navigate we need to know how to predict the tides, the location of the moon and sun, etc. Integers are important because they are precise: Once a month there is a full moon, easily recognizable. The next day its slightly less than full.

Sieves

A recursive sieve method for finding primes was first devised by the Greek Eratosthenes,[32] and summarized in Fig. 1.5.

1. Write $N-1$ counting number from 2 to N (List)
2. Define loop index $k=1$ and a multiplier $n \in \mathbb{N}$ denoted $n := \{2, \cdots, N\}$.
3. The next number on the list is prime $\pi_k \in \mathbb{P}$
4. Remove (Cross out) all multiples $n \cdot \pi_n$ of π_k
5. $k = k+1$: return to step 3.

Starting from the first prime $\pi_1 = 2$, one successively strikes out all the multiples of that prime. For example, starting from $\pi_1 = 2$ one strikes out $2 \cdot 2, 2 \cdot 3, 2 \cdot 4, 2 \cdot 5, \cdots, 2 \cdot (N/2)$. By definition the multiples are products of the target prime (2 in our example) and every another integer ($n \geq 2$). All the even numbers are removed in the first iteration. One then considers the next integer not struck out (3 in our example), which is identified as the next (second) prime π_2. Then all the $(N-2)/2$ non-prime multiples of π_2 are struck out. The next number which has not been struck out is 5, thus is prime π_3. All remaining multiples of 5 are struck out ($\cancel{10}$, $\cancel{15}$, $\cancel{25}$, $\cdots$). This process is repeated until all the numbers on the starting list have been processed.

As the word sieve implies, this sifting process takes a heavy toll on the integers, rapidly pruning the non-primes. In four loops of the sieve algorithm, all the primes below $N = 50$ are identified in **red**. The final set of primes is displayed at the bottom of Fig. 1.5.

Once a prime greater than $N/2$ has been identified, we may stop, since twice that prime is greater than N, the maximum number under consideration. Once you have reached $\sqrt{N}$ all the primes have been struck out (this follows from the fact that the next prime π_n is multiplied by an integer $n = 1, \ldots N$). Once this number $n\pi_n > N$ the list has been exhausted, which must be $n < \sqrt{N}$.

There are various schemes for making the sieve more efficient. For example the recursion $n\pi_k = (n-1)\pi_k + \pi_k$. could speed up the process, by replacing the multiply by an add by a known quantity. When using a computer, memory efficiency and speed are the main considerations.

[32] `https://en.wikipedia.org/wiki/Sieve_of_Eratosthenes\#Euler.27s_Sieve`

1. Write N integers from 2 to $N-1$. Let $k=1$. The first element $\pi_1 = 2$ is a prime. Cross out $n \cdot \pi_n$: (e.g., $n \cdot \pi_1 = 4, 8, 16, 32, \cdots$).

	2	3	~~4~~	5	~~6~~	7	~~8~~	9	~~10~~
11	~~12~~	13	~~14~~	15	~~16~~	17	~~18~~	19	~~20~~
21	~~22~~	23	~~24~~	25	~~26~~	27	~~28~~	29	~~30~~
31	~~32~~	33	~~34~~	35	~~36~~	37	~~38~~	39	~~40~~
41	~~42~~	43	~~44~~	45	~~46~~	47	~~48~~	49	~~50~~

2. Let $k=2$, $\pi_2 = 3$. Cross out $n\pi_k$ $(6, 9, 12, 15, \ldots)$:

	2	**3**	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	25	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	35	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	49	~~50~~

3. Let $k=3, \pi_3 = 5$. Corss out $n\pi_3$. (Cross out 25, 35).

	2	**3**	~~4~~	**5**	~~6~~	**7**	~~8~~	~~9~~	~~10~~
11	~~12~~	**13**	~~14~~	~~15~~	~~16~~	**17**	~~18~~	**19**	~~20~~
~~21~~	~~22~~	**23**	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	**29**	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	**37**	~~38~~	~~39~~	~~40~~
41	~~42~~	**43**	~~44~~	~~45~~	~~46~~	**47**	~~48~~	~~49~~	~~50~~

4. Finally let $k = 4, \pi_4 = 7$. Remove $n\pi_4$: (Cross out 49). Thus there are 15 primes less than $N = 50$: $\pi_k = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47\}$.

Figure 1.5: *Sieve of Eratosthenes for the case of $N = 50$.*

The importance of prime numbers

Likely the first insight into the counting numbers starts with the sieve, shown in Fig. 1.5. A sieve answers the question "What is the definition of a prime number?" which is likely the first question to be asked. The answer comes from looking for irregular patterns in the counting numbers, by playing the counting numbers against themselves.

A prime is that subset of positive integers $\mathbb{P} \in \mathbb{N}$ that cannot be factored. The number 1 is not a prime, for some non-obvious reasons, but there is no pattern in it since it is always a (useless) factor of every counting number.

To identify the primes we start from the first candidate on the list, which is 2 (since 1 is not a prime), and strike out all multiples by the counting numbers greater than 1 $[(n+1) \cdot 2 = 4, 6, 8, \cdots]$. While not obvious, this is our first result, that 2 is a prime, since it has no other factors but 1 and itself. This leaves only the odd numbers. We need a notation to indicate this result so we shall set $\pi_1 = 2$, as the first prime.[33]

Two Fundamental Theorems of Primes: Early theories of numbers revealed two fundamental theorems (there are many more than two), as discussed in Section 1.2.2. The first of these is the *Fundamental Theorem of Arithmetic*, which says that every integer greater than 1 may be uniquely

[33]There is a potentially conflicting notation since $\pi(N)$ is commonly defined as the number of primes less than index N. Be warned that here we define π_n as the n^{th} prime, and $\Pi(N)$ as the number of primes $\leq N$, since having a convenient notation for the n^{th} prime is more important that for the number of primes less than N.

factored into a product of primes (Eq. 2.1). Our demonstration of this is empirical, using Matlab's `factor(N)` routine, which delivers the prime numbers that compose N.[34] Typically the prime factors appear more than once, for example $4 = 2^2$. To make the notation compact we define the *multiplicity* β_k of each prime factor π_k (Eq. 2.1).

Each counting number is *uniquely* represented by a product of primes. There cannot be two integers with the same factorization. Once you multiply the factors out, the result is a unique N. Note that it's easy to multiply integers (e.g., primes), but nearly impossible to factor them. Factoring is not the same as dividing, as one needs to know what to divide by. Factoring means dividing by some integer and obtaining another integer with a zero remainder. This is what makes it so difficult (nearly impossible).

So the question remains: "What is the utility of the FTA?" which brings us to the topic of *internet security*. Unfortunately at this time I can not give you a proper summary of how it works. The full answer requires a proper course in number theory, beyond what is presented here.

The basic concept is that it is easy to construct the product of two primes, even very long primes having hundreds, or even thousands, of digits. It is very costly (but not impossible) to factor them. Why not use Matlab's `factor(N)` routine to find the factors? This is where *cost* comes in. The numbers used in RSA are too large for Matlab's routine to deliver an answer. In fact, even the largest computer in the world (such as the University of Illinois' super computer (NCSA Water) cannot do this computation. The reason has to do with the number of primes. If we were simply looking up a few numbers from a short list of primes, it would be easy, but the density of primes among the integers, is huge (see Section 2.1.1). This take us to the *Prime Number Theorem* (PNT). The security problem is the reason why these two theorems are so important: 1) Every integer has a unique representation as a product of primes, and 2) the number of primes is very dense (their are a very large number of them). Security reduces to the needle in the haystack problem, the cost of a search. A more formal way to measure the density is called the *entropy*, which is couched in terms of the probability of events, which in this case is "What is the probability of finding a prime between N and $2N$?"

Rational numbers $\mathbb{Q}$

The most important genus of numbers are the rational numbers since they maintain the utility of absolute precision, and they can approximate any irrational number (e.g., $\pi \approx 22/7$) to any desired degree of precision. However, the subset of rationals we really are interested in are the fractionals $\mathbb{F}$. Recall that $\mathbb{Q} : \mathbb{F} \cup \mathbb{Z}$ and $\mathbb{F} \perp \mathbb{Z}$. The fractionals are the numbers with the approximation utility, with arbitrary accuracy. Integers are equally important, but for a very different reason. All numerical computing today is done with $\mathbb{Q}$. Indexing uses integers $\mathbb{Z}$, while the rest of computing (flow dynamics, differential equations, etc.) is done with the fractionals $\mathbb{F}$. Computer scientists are trained on these topics, and engineers need to be at least conversant with them.

Irrational numbers: The cardinality of numbers may be ordered: $|\mathbb{I}| \ggg |\mathbb{Q}| \gg |\mathbb{N}| = |\mathbb{P}|$

The real line may be split into the irrationals and rationals. The rationals may be further split into the integers and the fractionals. Thus, all is *not* integer. If a triangle has two integer sides, then the hypotenuse must be irrational ($\sqrt{2} = \sqrt{1^2 + 1^2}$). This leads us to a fundamental question: "Are there integer solutions to Eq. 1.1?" We need not look further than the simple example $\{3, 4, 5\}$. In fact this example does generalize, and the formula for computing an infinite number of integer solutions is called *Euclid's Formula*, which we will discuss in Section 2.1.3.

However, the more important point is that the cardinality of the irrationals is much larger than any set other than the reals (i.e., complex numbers). Thus when we use computers to model physical systems, we are constantly needing to compute with irrational numbers. But this is impossible since every irrational numbers would require an infinite number of bits to represent it. Thus we must compute with rational approximations to the irrationals. This means we need to use the fractionals. In the end,

[34] If you wish to be a Mathematician, you need to learn how to prove theorems. If you're an Engineer, you are happy that someone else has already proved them, so that you can use the result.

we must work with the IEEE 754 floating point numbers,[35] which are fractionals, more fully discussed in Section 2.1.1.

1.2.4 Lec 5: Greatest common divisor (Euclidean algorithm)

The *Euclidean algorithm* is a method to find the greatest common divisor (GCD) k between two integers n, m, denoted $k = gcd(n, m)$, where $n, m, k \in \mathbb{N}$. For example $15 = \mathrm{gcd}(30, 105)$ since when factored $(30, 105) = (2 \cdot 3 \cdot 5,\ 7 \cdot 3 \cdot 5) = 3 \cdot 5 \cdot (2, 7) = 15 \cdot (2, 7)$. The Euclidean algorithm was known to the Chinese (i.e., not discovered by Euclid) (Stillwell, 2010, p. 41).

Why is the GCD important? Computing the GCD is simple, whereas a full factoring is extremely expensive. The GCD is important precisely because of the fundamental difficulty of factoring large integers into their primes. This utility surfaces when the two numbers are composed of very large primes. When two integers have no common factors they are said to be *coprime*, thus their GCD is 1. The ratio of two integers which are coprime is automatically in *reduced form* (they have no common factors).

For example $4/2 \in \mathbb{Q}$ is not reduced since 2=gcd(4,2). Canceling out the common factor 2, gives the reduced form $2/1{=}2 \in \mathbb{N}$. Thus if we wish to form the ratio of two integers, first compute the gcd and remove it from the two numbers, then form the ratio. This assures the rational number is in its reduced form. If the GCD were 10^3 digits it is obvious that the common factor must be removed before any computation should proceed.

An example: Take the two integers $[873, 582]$. In factored form these are $[\pi_{25} \cdot 3^2, \pi_{25} \cdot 3 \cdot 2]$. Given the factors, we see that the largest common factor is $\pi_{25} \cdot 3 = 291$. When we take the ratio of the two numbers this common factor cancels

$$\frac{873}{582} = \frac{\cancel{\pi_{25}} \cdot \cancel{3} \cdot 3}{\cancel{\pi_{25}} \cdot \cancel{3} \cdot 2} = \frac{3}{2}.$$

Of course if we divide 582 into 873 this we will numerically obtain the answer $1.5 \in \mathbb{R}$. If the common factor is large, a floating point number in $\mathbb{F}$ is returned, since all floating point numbers are in $\mathbb{F}$. But due to rounding errors, it may not be $3/2$. To obtain the exact answer, in $\mathbb{F}$, we need the GCD. Removing large common factors, without actually factoring the two numbers, has obvious practical utility.

Euclidean algorithm: The algorithm is best explained by a trivial example: Let the two numbers be $6, 9$. At each step the smaller number (6) is subtracted from the larger (9) and the difference (the remainder) and the smaller numbers are saved. This process continues until the two resulting numbers are equal, at which point the GCD equals that final number. If we were to take one more step, the final numbers would be the gcd and zero. For our example step 1 gives 9-6=3, leaving 6 and 3. Step 2 gives 6-3=3 and 3. Since the two numbers are the same, the GCD=3. If we take one more difference we obtain (3,0). We can easily verify this result since this example is easily factored (e.g., $3 \cdot 3, 3 \cdot 2) = 3(3, 2)$. It may be numerically verified using the Matlab GCD command `gcd(6,9)`, which returns 3.

In Chapter 2, Section 2.1.2 (p. 83), we shall describe two methods for implementing this procedure using matrix notation, and explore the deeper implications.

Coprimes

Related to the prime numbers are *co-primes*, which are integers that when factored, have no common primes. For example $20 = 5{\cdot}2{\cdot}2$ and $21 = 7{\cdot}3$ have no common factors, thus they are coprime. Coprimes $[m, n]$ may be indicated with the "perpendicular" notation $n \perp m$, spoken as "n is perpendicular (perp)

[35] IEEE 754: `http://www.h-schmidt.net/FloatConverter/IEEE754.html`.

Greatest Common Divisor: `k=gcd(m,n)`

- *Examples* $(m, n, k \in \mathbb{Z})$:
  - *gcd(13*5,11*5) = 5 (The common 5 is the gcd)*
  - *gcd(13*10,11*10) = 10 (The gcd(130,110) = 10 = 2*5, is not prime)*
  - *gcd(1234,1024) = 2 (1234=2*617, 1024=*2^{10}*)*
  - *gcd*$(\pi_k\pi_m,\ \pi_k\pi_n) = \pi_k$
  - `k=gcd(m,n)` *is the part that cancels in the fraction* $m/n \in F$
  - $m/\gcd(m,n) \in \mathbb{Z}$
- Co-primes $(m \perp n)$ *are numbers with no common factors: i.e., gcd(m,n)=1*
  - *The gcd of two primes is always 1: gcd(13,11) = 1, gcd*(π_m, π_n)*=1*
  - $m = 7 \cdot 13, n = 5 \cdot 19 \Rightarrow (7 \cdot 13) \perp (5 \cdot 19)$
  - *If* $m \perp n$ *then gcd*$(m, n) = 1$
  - *If gcd*$(m, n) = 1$ *then* $m \perp n$
- *The GCD may be extended to polynomials: e.g., gcd*$(ax^2 + bx + c,\ \alpha x^2 + \beta x + \gamma)$
  - *gcd*$((x-3)(x-4),\ (x-3)(x-5)) = (x-3)$
  - *gcd*$(x^2 - 7x + 12,\ 3(x^2 - 8x + 15)) = 3(x-3)$
  - *gcd*$(x^2 - 7x + 12,\ (3x^2 - 24x + 45) = 3(x-3)$
  - *gcd*$(\ (x-2\pi)(x-4),\ (x-2\pi)(x-5)\ ) = (x-2\pi)$ *(Needs long division)*

Figure 1.6: The Euclidean algorithm for finding the GCD of two numbers is one of the oldest algorithms in mathematics, and is highly relevant today. It is both powerful and simple. It was used by the Chinese during the Han dynasty (Stillwell, 2010, p. 70) for reducing fractions. It may be used to find pairs of integers that are coprime *(their gcd must be 1), and it may be used to identify factors of polynomials by long division. It has an important sister algorithm called the* continued fraction algorithm *(CFA), that is so similar in concept that Gauss referred to the Euclidean algorithm as the"continued fraction algorithm" (Stillwell, 2010, p. 48).*

to m." One may use the GCD to determine if two numbers are coprime. When $\gcd(m, n) = 1$, m and n are coprime. For example since `gcd(21,20)=1` (i.e., $21 \perp 20$) the are coprime.

1.2.5 Lec 6: Continued fraction algorithm (CFA)

The Continued fraction algorithm was mentioned in Section 1.2.4 at the end of the discussion on the GCD. These two algorithms (CFA vs. GCD) are closely related, enough that Gauss referred to the Euclidean algorithm as the Continued fraction algorithm (i.e., the name of the CFA algorithm) (Stillwell, 2010, P. 48). This question of similarity needs some clarification, as it seems unlikely that Gauss would be confused about such a basic algorithm.

In its simplest form the CFA starts from a real decimal number and recursively expands it as a fraction. It is useful for finding rational approximations to any real number. The GCD uses the Euclidean algorithm on a pair of integers $m > n \in \mathbb{N}$ and finds their greatest common divisor $k \in \mathbb{N}$. At first glance it is not clear why Gauss would call the CFA the Euclidean algorithm. One must assume that Gauss had some deeper insight into the relationship. If so, it would be valuable to understand.

In the following we refine the description of the CFA and give examples that go beyond the simple cases of expanding numbers. The CFA of any number, say x_0, is defined as follows:

1. Start with $n = 0$ and input target (starting value) $x_0 \in \mathbb{R}$.
2. If $|x_n| \geq 1/2$ define $a_n = \texttt{round}(x_n)$, which rounds to the nearest integer.
3. $r_n = x_n - a_n$ is the remainder. If $r_n = 0$, the recursion terminates.
4. Define $x_{n+1} \equiv 1/r_n$ and return to step 2 (with $n = n + 1$).

An example: Let $x_0 \equiv \pi \approx 3.14159\ldots$. Thus $a_0 = 3$, $r_0 = 0.14159$, $x_1 = 7.065 \approx 1/r_0$, and $a_1 = 7$. If we were to stop here we would have

$$\widehat{\pi}_1 \approx 3 + \frac{1}{7 + 0.0625\ldots} \approx 3 + \frac{1}{7} = \frac{22}{7}. \tag{1.2}$$

This approximation of $\pi \approx 22/7$ has a relative error of 0.04%

$$\frac{22/7 - \pi}{\pi} = 4 \times 10^{-4}.$$

For the second approximation we continue by reciprocating the remainder $1/0.0625 \approx 15.9966$ which rounds to 16, resulting in the second approximation

$$\widehat{\pi}_2 \approx 3 + 1/(7 + 1/16) = 3 + 16/(7 \cdot 16 + 1) = 3 + 16/113 = 355/113.$$

Note that if we had truncated 15.9966 to 15, the remainder would have been much larger, resulting in a less accurate rational approximation. The recursion may continue to any desired accuracy as convergence is guaranteed.

Rational approximation examples

$$\begin{aligned}
\frac{22}{7} &= [3; 7] && \approx \pi + O(1.3 \times 10^{-3}) \\
\frac{355}{113} &= [3; 7, 16] && \approx \pi + O(2.7 \times 10^{-7}) \\
\frac{104348}{33215} &= [3; 7, 16, -249] && \approx \pi + O(3.3 \times 10^{-10})
\end{aligned}$$

Figure 1.7: The expansion of π to various orders using the CFA, along with the order of the error of each rational approximation. For example 22/7 has an absolute error ($|22/7 - \pi|$) of about 0.13%.

Notation: Writing out all the fractions can become tedious. For example, expanding e using the Matlab command `rat(exp(1))` gives the approximation

$$3 + 1/(-4 + 1/(2 + 1/(5 + 1/(-2 + 1/(-7))))).$$

A compact notation for this these coefficients of the CFA is $[3; -4, 2, 5, -2, -7]$. Note that the leading integer may be indicated by an optional semicolon to indicate the decimal point. Unfortunately Matlab does not support the bracket notation.

If the process is carried further, the values of $a_n \in \mathbb{N}$ give increasingly more accurate rational approximations. If the floor rounding is used $\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, \ldots]$ whereas true rounding gives $\pi = [3; 7, 16, -294, 3, -4, 5, -15, \ldots]$, thus rounding introduces negative coefficients each time a number rounds up.

HW problem?

When the CFA is applied and the expansion terminates ($r_n = 0$), the target is rational. When the expansion does not terminate (which is not always easy to determine), the number is irrational. Thus the CFA has important theoretical applications regarding irrational numbers. You may try this yourself using Matlab's `rats(pi)` command. Also try the Matlab command `rat(1+sqrt(2))`.

One of the useful things about the procedure, besides its being so simple, are its generalizations, one of which will be discussed in Section 2.1.2 (p. 83).

A *continued fraction expansion* can have a high degree of symmetry. For example, the CFA of

$$\frac{1+\sqrt{5}}{2} = 1 + \frac{1}{1+\frac{1}{1+\cdots}} = 1.618033988749895\ldots,$$

Here the lead term in the fraction is *always* 1 ($a_n = [1; 1, 1, \cdots]$), thus the sequence will not terminate, proving that $\sqrt{5} \in \mathbb{I}$. A related example is `rat(1+sqrt(2))`, which gives $[2; 2, 2, 2, \ldots]$.

When expanding a target irrational number ($x_0 \in \mathbb{I}$), and the CFA is truncated, the resulting rational fraction approximates the irrational target. For the example above, if we truncate at three coefficients ($[1; 1, 1]$) we obtain

$$1 + \frac{1}{1+\frac{1}{1+0}} = 1 + 1/2 = 3/2 = 1.5 = \frac{1+\sqrt{5}}{2} + 0.118 + \ldots.$$

Truncation after six steps gives

$$[1.\ 1, 1, 1, 1, 1, 1] = 13/8 \approx 1.6250 = \frac{1+\sqrt{5}}{2} + .0070\ldots.$$

Because all the coefficients are 1, this example converges very slowly, When the coefficients are large (i.e., remainder small), the convergence will be faster. The expansion of π is an example of faster convergence.

In summary: Every rational number $m/n \in \mathbb{F}$, with $m > n > 1$, may be uniquely expanded as a continued fraction, with coefficients a_k determined using the CFA. When the target number is irrational ($x_0 \in \mathbb{Q}$), the CFA does not terminate, thus each step produces a more accurate rational approximation, converging in the limit as $n \to \infty$.

Thus the CFA expansion is an algorithm that can, in theory, determine when the target is rational, but with an important caveat: one must determine if the expansion terminates. In cases where the expansion produces a repeating coefficient sequence, it is clear that the sequence cannot terminate. The fraction $1/3 = 0.33333\ldots$ is an example of such a target where the CFA will terminate.[36]

WEEK 3

1.2.6 Labor day

1.2.7 Lec 7: Pythagorean triplets (Euclid's formula)

Euclid's formula is a method for finding three integer lengths $[a, b, c] \in \mathbb{N}$, that satisfy Eq. 1.1. It is important to ask "Which set are the lengths [a,b,c] drawn from?" There is a huge difference, both practical and theoretical, if they are from the real numbers $\mathbb{R}$, or the counting numbers $\mathbb{N}$. Given $p > q \in \mathbb{N}$, the three lengths $[a, b, c] \in \mathbb{N}$ of Eq. 1.1 are given by

$$a = p^2 - q^2, \qquad b = 2pq, \qquad c = p^2 + q^2. \tag{1.3}$$

This result may be directly verified, since

$$[p^2 + q^2]^2 = [p^2 - q^2]^2 + [2pq]^2$$

or

$$p^4 + q^4 + \cancel{2p^2q^2} = p^4 + q^4 - \cancel{2p^2q^2} + \cancel{4p^2q^2}.$$

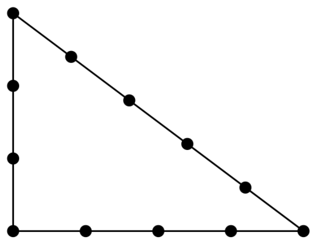


Figure 1.8: *Beads on a string form perfect right triangles when number of beads on each side satisfy Eq. 1.1.*

Thus, this result is easily proven, given the solution. Construction the solution is more difficult.

A well known example is the right triangle depicted in Fig. 1.8, defined by the integers $[3, 4, 5] \in N$, having angles $[0.54, 0.65, \pi/2]$ [rad], which satisfies Eq. 1.1. As quantified by Euclid's formula Eq. 1.3 (Section 2.2.1), there are an infinite number of *Pythagorean triplets* (PTs). Furthermore the seemingly simple triangle, having angles of $[30, 60, 90] \in N$ [deg] (i.e., $[\pi/6, \pi/3, \pi/2] \in \mathbb{I}$ [rad]), has one irrational ($\mathbb{I}$) length ($[1, \sqrt{3}, 2]$).

The technique for proving Euclid's formula for PTs $[a, b, c] \in \mathbb{Q}$, derived in Fig. 2.3 (p. 87) of Section 2.1.3, is much more interesting than the PTs themselves.

The set from which the lengths $[a, b, c]$ are drawn was not missed by the Indians, Chinese, Egyptians, Mesopotamians, Greeks, etc. Any equation whose solution is based on integers is called a *Diophantine equation*, named after the Greek mathematician Diophantus of Alexandria (c250 CE).

A stone tablet having the numbers engraved on it, as shown in Table 1.9 was discovered in Mesopotamia and from the 19^{th} century [BCE] and cataloged in 1922 by George Plimpton.[37] [38] These numbers are a and c pairs from PTs [a,b,c]. Given this discovery, it is clear that the Pythagoreans were walking in the footsteps of those well before them. Recently a second similar stone, dating between 350 and 50 [BCE] has been reported, that indicates early calculus on the orbit of Jupiter's.[39]

1.2.8 Lec 8: Pell's Equation

Pell's equation

$$x^2 - Ny^2 = 1, \tag{1.4}$$

with non-square $N \in \mathbb{N}$ specified and $a, b \in \mathbb{N}$ unknown, is related to the Euclidean algorithm (Stillwell, 2010, 48). For example, with $N = 2$, one solution is $a = 17, b = 12$ ($17^2 - 2 \cdot 12^2 = 1$). This equation has a long history (Stillwell, 2010).

A 2x2 matrix recursion algorithmic, used by the Pythagoreans to investigate the $\sqrt{2}$,

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_{n-1} \\ y_{n-1} \end{bmatrix}, \tag{1.5}$$

with $[1, 0]^T$, results in solutions of Pell's equations (Stillwell, 2010, p. 44). Their approach was likely motivated by the Euclidean algorithm (GCD, p. 33), since $y_n/x_n \to \sqrt{2}$ (Stillwell, 2010, p. 37,55). Note that this is a composition method, of 2x2 matrices, since the output of one matrix multiply is the input to the next. flow graph (p. 152

Asian solutions: The first intended solutions of Pell's was presented by Brahmagupta (c628), who independently discovered the equation (Stillwell, 2010, p. 46). Brahmagupta's novel solution introduced

[36] Taking the Fourier transform of the target number, represented as a sequence, could identify a periodic component. The number $1/7 = [[1, 4, 2, 8, 5, 7]]_6$ has a 50 [dB] notch at 0.8π [rad] due to its 6 digit periodicity, carried to 15 digits (Matlab precision), Hamming windows, and then zero padded to 1024 samples.

[37] http://www.nytimes.com/2010/11/27/arts/design/27tablets.html

[38] https://en.wikipedia.org/wiki/Plimpton_322

[39] http://www.nytimes.com/2016/01/29/science/babylonians-clay-tablets-geometry-astronomy-jupiter.html

Exercises

The integer pairs (a, c) in Plimpton 322 are

a	c
119	169
3367	4825
4601	6649
12709	18541
65	97
319	481
2291	3541
799	1249
481	769
4961	8161
45	75
1679	2929
161	289
1771	3229
56	106

Figure 1.3: Pairs in Plimpton 322

1.2.1 For each pair (a, c) in the table, compute $c^2 - a^2$, and confirm that it is a perfect square, b^2. (Computer assistance is recommended.)

You should notice that in most cases b is a "rounder" number than a or c.

1.2.2 Show that most of the numbers b are divisible by 60, and that the rest are divisible by 30 or 12.

Figure 1.9: *"Plimpton-322" is a stone tablet from 1800 BCE, displaying a and c values of the Pythagorean triplets $[a, b, c]$. Numbers ($a, c \in \mathbb{N}$), with the property $b = \sqrt{c^2 - a^2} \in \mathbb{N}$, known as* Pythagorean triplets, *were found carved on a stone tablet from the 19th century [BCE]. Several of the c values are primes, but not the a values. The stone is item 322 (item 3 from 1922) from the collection of George A. Plimpton. –Stillwell (2010, Exercise 1.2)*

a different *composition method* (Stillwell, 2010, p. 69), and like the Greek result, these solutions were incomplete.

Then in 1150CE, Bhâskara II obtained solutions using Eq. 1.5 (Stillwell, 2010, p.69). This is the solution method we shall explore here, as summarized in Fig. 1.10.

The best way to see how this recursion results in solutions to Pell's equation, is by example.

Initializing the recursion with the trivial solution $x_0 = [1, 0]^T$, gives

$$
\begin{array}{lr}
\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & 1^2 - 2 \cdot 1^2 = -1 \\
\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} & 3^2 - 2 \cdot 2^2 = 1 \\
\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} & (7)^2 - 2 \cdot (5)^2 = -1 \\
\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 17 \\ 12 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 5 \end{bmatrix} & 17^2 - 2 \cdot 12^2 = 1 \\
\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \begin{bmatrix} 41 \\ 29 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 17 \\ 12 \end{bmatrix} & (41)^2 - 2 \cdot (29)^2 = -1
\end{array}
$$

Thus the recursion results in a modified version of Pell's equation

$$x_n^2 - 2y_n^2 = (-1)^n,$$

thus only even values of n are solutions. This sign change had no effect on the Pythagoreans, who only cared about $y_n/x_n \to \sqrt{2}$.

Solution to Pell's equation: By multiplying the matrix by $1\jmath$, all the solutions to Pell's equation are determined. This solution is shown in Fig. 1.10 for the case of $N = 2$, and again in Appendix D, Eq. D.1, for $N = 3$. The math is straightforward and is easily verified using Matlab. From Fig. 1.10 we can see that every output this slightly modified matrix recursion gives solutions to Pell's equation (Eq. 1.4).

For $n = 0$ (the initial solution) $[x_0, y_0]$ is [1,0], $[x_1, y_1] = \jmath[1, 1]$, and $[x_2, y_2] = -[3, 2]$. These are easily computed by this recursion, and easily checked on a hand calculator (or using Matlab). Without the $\jmath$ factor the sign would alternate; the $1\jmath$ factor corrects the alternation in sign, so every iteration yields a solution.

- Case of $N = 2$ & $[x_0, y_0]^T = [1, 0]^T$
  Note: $x_n^2 - 2y_n^2 = 1, \qquad x_n/y_n \underset{\infty}{\longrightarrow} \sqrt{2}$

$$
\begin{array}{lr}
\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \jmath \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \jmath^2 - 2 \cdot \jmath^2 = 1 \\
\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \jmath^2 \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \jmath \begin{bmatrix} 1 \\ 1 \end{bmatrix} & 3^2 - 2 \cdot 2^2 = 1 \\
\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \jmath^3 \begin{bmatrix} 7 \\ 5 \end{bmatrix} = \jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \jmath^2 \begin{bmatrix} 3 \\ 2 \end{bmatrix} & (7\jmath)^2 - 2 \cdot (5\jmath)^2 = 1 \\
\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \begin{bmatrix} 17 \\ 12 \end{bmatrix} = \jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \jmath^3 \begin{bmatrix} 7 \\ 5 \end{bmatrix} & 17^2 - 2 \cdot 12^2 = 1 \\
\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \jmath \begin{bmatrix} 41 \\ 29 \end{bmatrix} = \jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 17 \\ 12 \end{bmatrix} & (41\jmath)^2 - 2 \cdot (29\jmath)^2 = 1
\end{array}
$$

Figure 1.10: *This summarizes the solution of Pell's equation for $N = 2$ using a slightly modified matrix recursion. Note that $x_n/y_n \to \sqrt{2}$ as $n \to \infty$, which was what the Pythagoreans were pursuing.*

At each iteration, the ratio x_n/y_n approaches $\sqrt{2}$ with increasing accuracy, coupling it to the Euclidean algorithm (GCD). The value of $41/29 \approx \sqrt{2}$, with a relative error of <0.03%. The solution for $N = 3$ is discussed at the end of Appendix D.

Relations to digital signal processing: Today we recognize Eq. 1.5 as a *difference equation,* which is a pre-limit (pre Stream 3) form of differential equation. The Greek 2x2 form is an early precursor to 17^{th} and 18^{th} century developments in linear algebra. Thus the Greek's recursive solution for the $\sqrt{2}$ and Bhâskara's (1030 CE) solution of Pell's equation, is an early precursor to discrete-time processing, as well as to calculus. Newton was fully aware of these developments as he reconstructed Diophantus chord/tangent method (Stillwell, 2010, p. 7, 49, 218).

Given the development of linear algebra c19^{th} century, as discussed in Section 2.2.2 (page 89), this may be evaluated by eigenvector diagonalization.[40]

There are similarities between Pell's Equation and the Pythagorean theorem. As we shall see in Chapter 2, Pell's equation is related to the geometry of a hyperbola, just as the Pythagorean equation is related to the geometry of a circle. One might wonder if there is a Euclidean formula for the solutions of Pell's Equations. After all, these are all conic sections with closely related geometry, in the complex plane.

Pell's Equation and irrational numbers: Since the eigenvalues of Eq. 1.5 ($\lambda_\pm = 1 \mp \sqrt{N} \not\in \mathbb{N}$), solutions to Pell's equation raised the possibility that all numbers are not rational. This discovery of irrational numbers forced the jarring realization that the Pythagorean dogma "all is integer" was wrong. The significance of irrational numbers was far from understood.

WEEK 4

1.2.9 Lec 9: Fibonacci sequence

Another classic problem, formulated by the Chinese, was the Fibonacci sequence, generated by the relation

$$f_{n+1} = f_n + f_{n-1}. \tag{1.6}$$

Here the next number f_{n+1} is the sum of the previous two. If we start from $[0, 1]$, this difference equation leads to the Fibonacci sequence $f_n = [0, 1, 1, 2, 3, 5, 8, 13, \cdots]$. The solution may be generated by the recursion of a 2x2 matrix equation, or by the z-transform method. Alternatively, if we define $y_{n+1} = x_n$, then

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \tag{1.7}$$

is equivalent to Eq. 1.6. The correspondence is easily verified. Starting with $[x_n, y_n]^T = [0, 1]^T$ we obtain for the first few steps

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \cdots$$

From the above $x_n = [0, 1, 1, 2, 3, 5, \ldots]$ is the Fibonacci sequence since the next number is the sum of the last two.

Note that this 2x2 equation is similar to Pell's equation, suggesting that an eigenfunction expansion of Eq. 1.7 may be used to analyze the sequence, as shown in Section 2.3.1 (p. 89) (Stillwell, 2010, 192).

[40] `https://en.wikipedia.org/wiki/Transformation_matrix#Rotation`

1.2.10 Lec 10: Exam I (In class)

1.3 Algebraic Equations: Stream 2

1.3.1 Lec 11 Algebra and geometry as physics

Following Stillwell's history of mathematics, Stream 2 is geometry, which led to the merging of Euclid's geometrical methods and the 9^{th} century development of algebra by al-Khwarizmi (830 CE). This integration of ideas lead Descartes and Fermat to develop of *analytic geometry*. While not entirely a unique and novel idea, it was late in coming, given what was known at that time.

The mathematics up to the time of the Greeks, documented and formalized by Euclid, served students of mathematics for more than two thousand years. Algebra and geometry were, at first, independent lines of thought. When merged, the focus returned to the Pythagorean theorem, generalized as analytic conic sections rather than as geometry in Euclid's Elements. With the introduction of Algebra, numbers, rather than lines, could be used to represent a geometrical length. Thus the appreciation for geometry grew given the addition of the rigorous analysis using numbers.

Physics inspires algebraic mathematics: The Chinese used music, art, navigation to drive mathematics. With the invention of algebra this paradigm did not shift. A desire to understand motions of objects and planets precipitated many new discoveries. Galileo investigated gravity and invented the telescope. Kepler investigated the motion of the planets. While Kepler was the first to appreciate that the planets were described by ellipses, it seems he under-appreciate the significance of this finding, and continued with his epicycle models of the planets. Using algebra and calculus, Newton formalized the equation of gravity, forces and motion (Newton's three laws) and showed that Kepler's discovery of planetary elliptical motion naturally follows from these laws. With the discovery of Uranus [in 1781] "Kepler's theory was ruined." (Stillwell, 2010, p. 23).

Once Newton proposed the basic laws of gravity, he proceed to calculate, for the first time, the speed of sound. This required some form of the *wave equation*

$$\frac{\partial^2}{\partial x^2}p(x,t) = \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(x,t), \tag{1.8}$$

a key equation in mathematical physics. The speed of sound is $c = 343$ [m/s], which is a function of the density $\rho = 1.12$ [kg/m^3] and the dynamic stiffness ηP_0 of air.[41]

If we substitute for the pressure

$$p(x,t) = e^{\jmath 2\pi(ft \pm kx)}, \tag{1.9}$$

where t is time and x is position, we find that $(\jmath k)^2 = s^2/c^2$, or $k = 2\pi/\lambda = 2\pi f/c$.

While Newton's value for c was incorrect by the thermodynamic constant $\sqrt{\eta}$, a problem that would take more than two hundred years to resolve, his success was important because it quantified the physics behind the speed of sound, and demonstrated that momentum mv, not mass m, was transported by the wave. His concept was correct, and his formulation using algebra and calculus represented a milestone in science.

Newton's *Principia* was finally published in 1687, and the general solution to Newton's wave equation [i.e., $p(x,t) = G(t \pm x/c)$], where G is *any* function, was first published 60 years later by d'Alembert (c1747), which showed that for sounds of a single frequency, the wavelength λ and frequency f were related by

$$f\lambda = c.$$

[41] $c = \sqrt{\eta P_0/\rho}$, $\eta = c_p/c_v = 1.4$ is the ratio of two thermodynamic constants, and $P_0 = 10^5$ [Pa] is the barometric pressure of air.

Today d'Alembert's analytic wave solution is frequently written as Eq. 1.9 where $k = c/\lambda$ is the *wave number*. This formulation led to the frequency domain concept of Fourier analysis, based on the *linearity* (i.e., superposition) property of the wave equation (Postulate P2: Lec. 1.3.11, p. 61).

The corresponding discovery for the formula for the speed of light was made 114 years later by Maxwell (c1861). Maxwell's formulation also required great ingenuity, as it was necessary to hypothesize an experimentally unmeasured term in his equations, to get the mathematics to correctly predict the speed of light.

The first Algebra: Prior to the invention of algebra, people worked out problems as sentences using an obtuse description of the problem (Stillwell, 2010, p. 93). Algebra solved this problem. Algebra may be thought of as a compact language, where numbers are represented as abstract symbols (e.g., x and α). The problems they wished to solve could be formulated in terms of sums of powers of smaller terms, the most common being powers of some independent variable (i.e., time or frequency). Today we call such an expression a *polynomial* of *degree n*

$$P_n(z) \equiv z^n + a_{n-1}z^{n-1} + \cdots + a_0 z^0 = \sum_{k=0}^{n} a_k z^k = \prod_{k=0}^{n}(z - z_k). \tag{1.10}$$

Here we have set $a_n = 1$. The coefficient a_n cannot be zero, or the polynomial would not have degree n. The solution is to force $a_n = 1$. This does not change the roots.

The key question was "What values of the $z = z_k$ result in $P_n(z_k) = 0$." In other words, what are the roots z_k of the polynomial? The quest for the answer to this question consumed thousands of years, with intense efforts by many aspiring mathematicians. In the earliest attempts, it was a competition to evaluate mathematical acumen. Results were held as a secret to the death bed. It would be fair to view this effort as an obsession. Today the roots of any polynomial may be found by numerical methods, to very high accuracy. There are also a number of important theorems.

Of particular interest was composing a circle with a line, for example when the line does not touch the circle, and finding the roots. There was no solution to this problem using geometry.

See assignments

Finding roots of polynomials The problem of factoring polynomials has a history more than a millennium in the making. While the quadratic (degree $N = 2$) was solved by the time of the Babylonians (i.e., the earliest recorded history of mathematics), the cubic solution was finally published by Cardano in 1545. The same year, Cardano's student solved the quartic ($N = 4$). In 1826 it was proved that the quintic ($N = 5$) could not be factored by analytic methods.

As a concrete example we begin with trivial but important case of the quadratic

$$P_2(x) = ax^2 + bx + c. \tag{1.11}$$

The *roots* are those values of x such that $P_2(x_k) = 0$. One of the first results (recorded by the Babylonians, c2000 BCE) was the factoring of this equation by *completing the square* (Stillwell, 2010, p. 93). One may rewrite Eq. 1.11 as

$$\frac{1}{a}P_2(x) = (x + b/2a)^2 - (b/2a)^2 + c/a, \tag{1.12}$$

which is easily verified by expanding the squared term and canceling $(b/2a)^2$

$$\frac{1}{a}P_2(x) = [x^2 + (b/a)x + \cancel{(b/2a)^2}] - \cancel{(b/2a)^2} + c/a.$$

Setting Eq. 1.12 to zero and solving for the two roots $x_\pm$, gives the *quadratic formula*[42]

$$x_\pm = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{1.13}$$

[42] By direct substitution demonstrate that Eq. 1.13 is the solution of Eq. 1.11.

If $b^2 + ac > 0$, then the two roots are real ($x_\pm \in \mathbb{R}$). Otherwise, they are complex. If $a, b, c \in \mathbb{R}$, which is typically the case, then this condition simplifies to $ac < 0$ (i.e., $c/a > 0$).

No insight is gained by memorizing the quadratic formula (Eq. 1.13). On the other hand, an important concept is gained by learning Eq. 1.12, which can be very helpful when doing analysis. I suggest that instead of memorizing Eq. 1.13, memorize Eq. 1.12. Arguably, the factored form is easier to remember (or learn). Perhaps more importantly, the term $b/2a$ has significance $[P_2(-b/2a) = c/a - (b/2a)^2]$, the sign of which determines if the roots are real or complex.

In third grade I learned the trick[43]

$$9 \cdot n = (n-1) \cdot 10 + (10 - n). \tag{1.14}$$

With this simple rule I did not need to depend on my memory for the 9 times tables. How one thinks about a problem can have great impact.

Analytic Series: When the degree of the polynomial is infinite (i.e., $n = \infty$), $P_\infty(x)$, $x \in \mathbb{R}$ the series is called a *power series*. For values of x where the power series converges, it is said to be *analytic*. When the coefficients are determined by derivatives of $P(x)$ evaluated at $x = 0$, then it is called a *Taylor series*. These various series play a special role in mathematics, as the coefficients of the series uniquely determine a function (e.g., via the derivatives). Two well known examples are the geometric series

$$\frac{1}{1-x} = 1 + x + x^2 + x^2 + \ldots = \sum_{n=0}^{\infty} x^n$$

and exponential

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3 \cdot 2}x^3 + \frac{1}{4 \cdot 3 \cdot 2}x^4 + \ldots = \sum_{n=0}^{\infty} \frac{1}{n!}x^n. \tag{1.15}$$

Region of convergence: The set of values for which the series is analytic is called the *region of convergence*, or simply the ROC. Knowing how to determine the ROC for a given analytic function is quite important, and may not always be obvious. For the geometric series, the ROC is $|x| < 1$. The function $1/(x^2+1)$ has the same ROC as the geometric series, since it may be written as (Section A.6, p. 134)

$$\frac{1}{x^2+1} = \frac{1}{(x+1\jmath)(x-1\jmath)} = \frac{1}{2\jmath}\left(\frac{1}{x-1\jmath} - \frac{1}{x+1\jmath}\right).$$

Each term has an ROC of $|x| < |1\jmath| = 1$. In other words, it is the sum of two geometric series, each having a pole at $\pm 1\jmath$.

The exponential series converges for every finite value of x (the ROC is the entire open plane), thus the exponential is called an *entire function*.

Analytic functions: Any function that has an analytic series representation is called an *analytic function*. Polynomials, $1/(1 - x\jmath)$ and $e^{x\jmath}$ are analytic functions. Because analytic functions are easily manipulated, term by term, they may be used to find solutions of differential equations. The derivatives are easily computed, since they may be uniquely determined, term by term. Every analytic function has a corresponding differential equation, that is determined by the coefficients of the analytic power series. An example is the exponential, which has the property that it is the eigenfunction of the derivative operation

$$\frac{d}{dx}e^{ax} = ae^{ax},$$

which may be verified using Eq. 1.15. This relationship is a common definition of the exponential function, which is a very special function.

[43]E.G.: $9 \cdot 7 = (7-1) \cdot 10 + (10-7) = 60 + 3$ and $9 \cdot 3 = 2 \cdot 10 + 7 = 27$. As a check note that the two digits of the answer must add to 9.

Analytic functions may also be easily integrated, term by term. Newton took full advantage of these properties of analytic functions. To fully understand the theory of differential equations (DE), one needs to master single valued analytic functions and their analytic power series. Newton used the analytic series (*Taylor series*) to solve many problems, especially for working out integrals, allowing him to solve DEs.

During the 16th and 17th century, it had becoming clear that DEs can characterize a law of nature at a single point in space and time. For example the law of gravity (first formulated by Galileo to explain the dropping to two objects of different masses) must obey conservation of energy. Newton (c1687) went on to show that there must be a gravitational potential between to masses (m_1, m_2) of the form

$$\phi(r) \propto \frac{m_1 m_2}{r}, \tag{1.16}$$

where $r = |x_1 - x_2|$ is the Euclidean distance between the two point masses at locations x_1 and x_2. Note that this a power series, but with exponent of -1, which is called a *pole*.

Single- vs. multi-valued functions: Polynomials are single valued functions: for each x there is a single value of $P_n(x)$. The set of x values of a function are called the *domain* and the set of $y(x)$ values are called the *codomain*. The roles of the domain and codomain may be swapped, to obtain the inverse function, which is typically quite different in its properties compared to the function. For example $y(x) = x^2 + 1$ has the inverse $x = \pm\sqrt{y-1}$, which is double valued, and complex when $y < 1$. Periodic functions such as $y(x) = \sin(x)$ are more exotic, since $x(y) = \arcsin(x) = \sin^{-1}(x)$ has an ∞ number of $x(y)$ values for each y.

Complex analytic functions: When the argument of and analytic function is complex, that is, $x \in \mathbb{R}$ is replaced by $z = x + y\jmath \in \mathbb{C}$ (recall that $\mathbb{R} \in \mathbb{C}$), the function is said to be a *complex analytic*. We shall delve further into this topic in Section 3.1.1 (p. 93).

When the argument of the exponential becomes complex, it is periodic since

$$e^{st} = e^{(\sigma+\omega\jmath)t} = e^{\sigma t}e^{\jmath\omega t} = e^{\sigma t}\left[\cos(\omega t) + \jmath \sin(\omega t)\right],$$

thus

$$\cos(\omega) = \frac{e^{\omega\jmath} + e^{-\omega\jmath}}{2}.$$

It should be clear that the exponential and circular functions are fundamentally related. This exposes the family of *entire circular functions* [i.e., $e^s, \sin(s), \cos(s), \tan(s), \cosh(s), \sinh(s)$] and their inverses [$\ln(s), \arcsin(s), \arccos(s), \arctan(s), \cosh^{-1}(s), \sinh^{-1}(s)$], first fully elucidated by Euler (c1750). Thus there is a deep relationships between single vs. multi-valued functions, periodic functions, functions of a complex variable, and their inverse functions, all of which are complex analytic. We shall explore this in Lec 1.3.8 (p. 55).

Given a complex analytic function of a complex variable, one must resort to the *extended complex plane*, *Riemann sheets* and *branch cuts*, as discussed in Section 1.3.7 (p. 53). The extended complex plane is a tool that is needed throughout complex analytic functions, and into higher mathematics, since at least the 16^{th} century, and probably before. This topic is critically important in engineering mathematics, and will be discussed in length in Sections 1.3.7-1.3.10 (pp. 53-59).

There seems to be some disagreement as to the status of multivalued functions: Are they functions, or is a function strictly single valued? If so, then we are missing out on a host of possibilities, including all the inverses of every complex analytic function.

Impact on Physics: It seems likely, if not obvious, that the success of Newton in mathematics was his ability to find mathematics in physics. He was inventing new mathematics at the same time as he was explaining new physics. The same might be said for Galileo. It seems likely that Newton was extending the successful techniques and results of Galileo. Galileo died on Jan 8, 1642, and Newton was born Jan 4, 1643, just short of one year later. Certainly Newton was well aware of Galileo's great success, and naturally would have been influenced by them.

The application of complex analytic functions to physics was dramatic, as may be seen in the six volumes on physics by Arnold Sommerfeld (1868-1951), and from the productivity of his many (36) students (e.g., Debye, Lenz, Ewald, Pauli, Guillemin, Bethe, Heisenberg[44] and Seebach, to name a few), notable coworkers (i.e., Leon Brillouin) and others (i.e., John Bardeen), upon whom he had a strong influence. Sommerfeld is known for having many students who were awarded the Nobel Prize in Physics, yet he was not (the prize is not awarded in Mathematics). Sommerfeld brought mathematical physics (the merging of physical and experimental principles with mathematics) to a new level with the use of complex integration of analytic functions to solve otherwise difficult problems, thus following the lead of Newton who used real integration of Taylor series to solve differential equations, and later Cauchy. While much of this work is outside the scope of the present discussion, it is helpful to know who did what and when, and how people and concepts are connected.

WEEK 5 12.5.0

L 12 Examples of algebraic expressions in physics
Fundamental Thm of Algebra (d'Alembert, ≈1760)
Analytic Geometry: Algebra + Geometry (Euclid to Descartes)
Newton and power series; Taylor series & ROC Composition of polynomial equations in two variables.

L 13 Root classification for polynomials of Degree * = 1–4 (p.102);
Convolution of monomials gives polynomial construction; Work out convolution for cubic
Show that a_{n-1} is sum of roots and a_0 is product of roots. Quintic (* = 5) cannot be solved

L 14 First Analytic Geometry (Fermat 1629; Descartes 1637) (p. 118) Descartes' insight: Composition of two polynomials of degrees (m,n → one of degree $m \cdot n$)
Examples: $x^4 \circ x^2 = x^8$. Discuss Composition vs. intersection of functions.

1.3.2 Lec 12 Physical equations quadratic in several variables

When lines and planes are defined, the equations are said to be *linear* in the independent variables. In keeping with this definition of *linear*, we say that the equations are *non-linear* when the equations have degree greater than 1 in the independent variables. The term *bilinear* has a special meaning, in that both the domain and codomain are linearly related by lines (or planes). As an example, an impedance is defined in frequency as the ratio of the voltage over the current

$$Z(s) = \frac{V(\omega)}{I(\omega)} = \frac{N(s)}{D(s)},$$

where $Z(s)$ is the impedance and V and I are the voltage and current at radian frequency ω. The impedance is typically specified as the ratio of two polynomials, $N(s)$ and $D(s)$, as functions of complex Laplace frequency $s = \sigma + \jmath\omega$. An example will be given in Section 1.3.6, Fig. 1.11. The bilinear function may be written as $D(s)V = N(s)I$. Since $D(s)$ and $N(s)$ are both polynomials in s, this is called bilinear. It comes from a corresponding scalar differential equation (Appendix A, Eq. A.1, p. 133).

[44] https://www.aip.org/history-programs/niels-bohr-library/oral-histories/4661-1

As an example consider the well known problem in geometry: the intersection of a plane with a cone, which leads to the conic sections: the circle, hyperbola, ellipse and parabola, along with some degenerate cases, such as the intersection of two straight lines[45]. If we stick to such 3-dimensional objects, we can write equations in the three variables $[x, y, z]$, and be sure that they each represent some physical geometry. For example $x^2 + y^2 + z^2 = r_0^2$ is a sphere of radius r_0.

The geometry and the algebra do not always seem to agree. Which is correct? In general the geometry only looks at the real part of the solution, unless you know how to tease out the complex solutions. However the roots of any polynomial are from $\mathbb{C}$, so we may not ignore the imaginary roots, as Newton did. There is an important related fundamental theorems, known as Bézout's theorem, that address this case, as described next.

1.3.3 Lec 13: Polynomial root classification by convolution

Following the exploration of algebraic relationships by Fermat and Descartes, the first theorem was being formulated by d'Alembert. The idea behind this theorem is that every polynomial of degree N (Eq. 1.10) has at least one root. This may be written as the product of the root and a second polynomial of degree of $N-1$. By the recursive application of this concept, it is clear that every polynomial of degree N has N roots. Today this result is known as the *Fundamental Theorem of Algebra:*

> Every polynomial equation $p(z) = 0$ has a solution in the complex numbers. As Descartes observed, a solution $z = a$ implies that $p(z)$ has a factor $z - a$. The quotient
>
> $$q(z) = \frac{p(z)}{z-a}$$
>
> is then a polynomial of one lower degree. ...We can go on to factorize $p(z)$ into n linear factors.[46]
>
> —Stillwell (2010, p. 285).

The ultimate expression of this theorem is given by Eq. 1.10 (p. 42), which indirectly states that an n^{th} degree polynomial has n roots.

Today this theorem is so widely accepted we may fail to appreciate it. Certainly about the time you learned the quadratic formula, you were prepared to understand the concept. The simple quadratic case may be extended a higher degree polynomial. The Matlab/Octave command `roots`($[a_3, a_2, a_1, a_0]$) will provide the roots of the cubic equation, defined by the four coefficients $a_3, \ldots, a_0$. I don't know the largest degree that can be accurately factored by Matlab/Octave, but I'm sure its well over $N = 10^3$. Today, finding the roots numerically is a solved problem.

Factorization versus convolution: The best way to gain insight into the polynomial factorization problem is through the inverse operation, multiplication of monomials. Given the roots x_k, there is a simple algorithm for computing the coefficients a_k of $P_n(x)$ for any n, no matter how large. This method is called *convolution*. Convolution is said to be a *trap-door* since it is easy, while the inverse, factoring (*deconvolution*), is hard, and analytically intractable for degree $N \geq 5$ (Stillwell, 2010, p. 102).

Convolution of monomials

As outlined by Eq. 1.10, a polynomial has two descriptions, first as a series with coefficients a_n and second in terms of its roots x_r. The question is "What is the relationship between the coefficients and the roots?" The simple answer is that they are related by *convolution*.

[45]Such problems were first studied algebraically and Descartes (Stillwell, 2010, p. 118) and Fermat (c1637).

[46]Look into expressing this in terms of complex 2x2 matrices, as on p. 26.

Let us start with the quadratic

$$(x+a)(x+b) = x^2 + (a+b)x + ab,$$

where in vector notation $[-a, -b]$ are the roots and $[1, a+b, ab]$ are the coefficients.

To see how the result generalizes, we may work out the coefficients for the cubic ($N = 3$). Multiplying the following three factors gives

$$(x-1)(x-2)(x-3) = (x^2-3x+2)(x-3) = x(x^2-3x+2) - 3(x^2-3x+2) = x^3 - 6x^2 + 11x - 6.$$

When the roots are $[1, 2, 3]$ the coefficients of the polynomial are $[1, -6, 11, -6]$. To verify, substitute the roots into the polynomial, and show that they give zero. For example $r_1 = 1$ is a root since $P_3(1) = 1 - 6 + 11 - 6 = 0$.

As the degree increases, the algebra becomes more difficult; even a cubic becomes tedious. Imagine trying to work out the coefficients for $N = 100$. What is needed is an simple way of finding the coefficients from the roots. Fortunately, *convolution* keeps track of the book-keeping, by formalizing the procedure.

Convolution of two vectors: To get the coefficients by convolution, write the roots as two vectors $[1, a]$ and $[1, b]$. To find the coefficients we must convolve the root vectors, indicated by $[1, a] \star [1, b]$, where $\star$ denotes convolution. Convolution is a recursive operation. The convolution of $[1, a] \star [1, b]$ is done as follows: reverse one of the two monomials, padding unused elements with zeros. Next slide one monomial against the other, forming the local *dot product* (element-wise multiply and add):

$$\begin{array}{cccc} a & 1 & 0 & 0 \\ 0 & 0 & 1 & b \\ = & 0 & & \end{array} \quad \begin{array}{ccc} a & 1 & 0 \\ 0 & 1 & b \\ = & x^2 & \end{array} \quad \begin{array}{ccc} a & 1 & 0 \\ 1 & b & 0 \\ = & (a+b)x & \end{array} \quad \begin{array}{ccc} 0 & a & 1 \\ 1 & b & 0 \\ = & abx^0 & \end{array} \quad \begin{array}{cccc} 0 & 0 & a & 1 \\ 1 & b & 0 & 0 \\ = & 0 & & \end{array},$$

resulting in coefficients $[\cdots, 0, 0, 1, a+b, ab, 0, 0, \cdots]$.

By reversing one of the polynomials, and then taking successive dot products, all the terms in the sum of the dot product correspond to the same power of x. This explains why convolution of the coefficients gives the same answer as the product of the polynomials.

As seen by the above example, the position of the first monomial coefficients are reversed, and then slid across the second set of coefficients, the dot-product is computed, and the result placed in the output vector. Outside the range shown, all the elements are zero. In summary,

$$[1, -1] \star [1, -2] = [1, -1-2, 2] = [1, -3, 2].$$

In general

$$[a, b] \star [c, d] = [ac, bc + ad, bd],$$

Convolving a third term $[1, -3]$ with $[1, -3, 2]$ gives

$$[1, -3] \star [1, -3, 2] = [1, -3-3, 9+2, -6] = [1, -6, 11, -6],$$

which is identical to the cubic example, found by the algebraic method.

By convolving one monomial factor at a time, the overlap is always two elements, thus it is never necessary to compute more than two multiplies and an add for each output coefficient. This greatly simplifies the operations (i.e., they are easily done in your head). Thus the final result is more likely to be correct. Comparing this to the algebraic method, convolution has the clear advantage.

The product of monomial $P_1(x)$ with a polynomial $P_N(x)$ gives $P_{N+1}(x)$: This statement is another way of stating the *Fundamental Theorem of Algebra*. Each time we convolve a monomial with a polynomial of degree N, we obtain a polynomial of degree $N+1$. The convolution of two monomials results in a quadratic (degree 2 polynomial). The convolution of three monomials give a cubic (degree 3). In general, the degree k, of the product of two polynomials, of degree n, m, is the sum of the degrees $(k = n + m)$. For example, if the degrees are each 5 $(n = m = 5)$, then the resulting degree is 10.

In summary, the product of two polynomials of degree m, n having m and n roots, give a polynomial of degree $m+n$ having $m+n$ roots. This is an analysis process, of merging polynomials, by coefficient convolution. Multiplying polynomials is a merging process, into a single polynomial.

Note that the degree of a polynomial is one less than the length of the vector of coefficients. The coefficient on the lead term should always be set to 1 since it cannot be zero, resulting in an illogical result. *Always* normalize the lead term of $P_n(x)$ to 1 (i.e., $a_N = 1$). This has no effect on the roots.

While you already know this theorem from high school algebra class, it is important to explicitly state the *Fundamental theorem of algebra*.

1.3.4 Lec 14: Introduction to Analytic Geometry

Analytic geometry is the natural consequence of Euclid's Geometry (which deals with conical geometry (e.g., points, lines, triangles, circles, spheres, ellipses, cones, etc., in two and three physical dimensions), merged with algebra (which deals with simultaneous equations, roots of polynomials, analytic functions, and ultimately, solutions of differential equations). The combination of Euclid's (323 BEC) geometry and al-Khwarizmi's (830 CE) algebra provides a new powerful tool, *analytic geometry*.

There are many important relationships between Euclidean Geometry and 16th century algebra. Important similarities include vectors, their Pythagorean lengths $[a, b, c]$

$$c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{1.17}$$

$a = x_2 - x_1$ and $b = y_2 - y_1$, and the angles. Geometry had no concept of coordinates of the vectors. This is one of the main differences, the ability of algebra to compute with numbers. A detailed comparison is in order, attempted in Table 1.3.

There are several basic concepts that merged under the development of analytic geometry

1. Composition of functions: If $y = f(x)$ and $z = g(y)$ then the composition of functions f and g is denoted $z(x) = g \circ f(x) = g(f(x))$.

2. Elimination: Given two functions $f(x, y)$ and $g(x, y)$, elimination removes either x or y. This procedure, called *Gaussian Elimination*, was known to the Chinese.

3. Vectors: Euclidean geometry includes the concept of a vector, as a line with a length and a slope. Analytic geometry defines a vector as an ordered set of points.

4. Analytic geometry extends the ideas of geometry with the introduction of the product of two vectors, the dot product $\vec{\mathbf{f}} \cdot \vec{\mathbf{g}}$ and the cross product $\vec{\mathbf{f}} \times \vec{\mathbf{g}}$ (Appendix B).

5. Intersection: While one may speak of the intersection of two lines, the term *intersection* is also a concept of set theory. To distinguish the very different uses of this important word, we shall coin a new spelling for the case of set theory: *intersetsion*. The intersetsion of two sets is indicate by $A \cap B$.

What algebra added to geometry was the ability to compute with numbers. For example, the length of a line (Eq. 1.17) was measured in Geometry with a compass: numbers played no role. Once algebra was available, the line's Euclidean length could be computed from the coordinates of the two ends, defined by the *3-vector*

$$\vec{e} = x\hat{x} + y\hat{y} + z\hat{z},$$

Table 1.3: Comparison between Euclidean Geometry and Algebra. The concepts that are included in Euclidean geometry are under the column *Geometry.* Those under the column *Algebra* are unique to Algebra (not present in Geometry). The third column are uncertain.

Geometry	Algebra	Uncertain
Proof, Vector, Length, Point, Direction, $\le$ 3 dimensional, Intersection, Conic Section, Dot product, Square Root	$>$ 1 dimensional, Numbers, Analytic series and functions, Composition, Elimination, Fund. Thm. Algebra, $\sin(\theta)$, $\cos(\theta)$, $e^{\theta \jmath}$, $\log(z)$, Derivative, Calculus, Polynomial & Roots	Cross product, Recursion, Iteration (Newton's method)

which represents a point at $(x, y, z) \in \mathbb{R}^3 \subset \mathbb{C}^3$ in three dimensions, having *direction*, from the origin $(0, 0, 0)$ to (x, y, z). A point, by itself, has no direction, thus is a scalar. An alternative matrix notation is $\vec{e} = [x, y, z]^T$, a column vector of three numbers. These two notations are very different ways of representing exactly the same thing. View them as equivalent concepts.

By defining the vector, analytic geometry allows Euclidean geometry to become quantitative, beyond the physical drawing of an object (e.g., a sphere, triangle or line). With analytic geometry we have the Euclidean concept of a vector, a line having a magnitude and direction (a slope), but in terms of physical coordinates (i.e., numbers). The difference between two vectors defines a length, a concept already present in Euclidean geometry, but not quantified other than by the spread of the compass points placed at the two ends of the line.

Dot product of two vectors: When using algebra, many concepts in geometry are made precise. There are many examples of how algebra extends Euclidean geometry, the most basic being the *dot product* between two vectors

$$\begin{aligned}\vec{x} \cdot \vec{\zeta} &= (x\hat{x} + y\hat{y} + z\hat{z}) \cdot (\alpha\hat{x} + \beta\hat{y} + \gamma\hat{z}) \\ &= \alpha x + \beta y + \gamma z.\end{aligned}$$

The dot product takes two vectors in $\mathbb{R}^3$ and returns a real positive scalar function of the coordinates of the two vectors in $\mathbb{R}$, which represents a length. In matrix notation the dot product is written as

$$\vec{x} \cdot \vec{\zeta} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}^T \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x, y, z \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \alpha x + \beta y + \gamma z.$$

Recall that in matrix notation a vector, by default, is always a column. The transpose is a row-vector.

Norm of a vector: The *vector* $\vec{e} \in \mathbb{C}^3$ encompasses the scalar point $(x, y, z) \in \mathbb{C}$ along with the scalar origin $(0, 0, 0)$, to define a the vector, having both a *magnitude*

$$||\vec{e}|| = \sqrt{\vec{e} \cdot \vec{e}} = \sqrt{x^2 + y^2 + z^2},$$

and *direction* (i.e., three angles). The magnitude of the vector (its length) is given by the Pythagorean theorem, in $\mathbb{R}^3$. Notionally, $||\vec{e}||$ is called the *norm* of the vector and is defined as the positive square root of the dot product of the vector with itself. This is the generalization of a length, denoted the *norm* of $\vec{e}$, is

$$||\vec{e}|| \equiv +\sqrt{\vec{e} \cdot \vec{e}}.$$

The sign of the square-root must be positive as the length of a vector is not a concept from complex analytic functions. Rather the length is a concept of Euclidean geometry, and it must always be positive and real. A zero-length vector is a point, which has no direction (thus it is not a vector).

More generally, the Euclidean length of a line is given as the *norm* of the difference between two vectors

$$\begin{aligned}||\vec{e_1} - \vec{e_2}||^2 &= (\vec{e}_1 - \vec{e}_2) \cdot (\vec{e}_1 - \vec{e}_2) \\ &= (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2,\end{aligned}$$

which is the Euclidean length between the two vectors. This length is not an analytic concept. Thus the minus sign on the root makes no sense in this context.

Cross product of two vectors: The second type of product between two vectors is the cross-product $\vec{e_1} \times \vec{e_2}$. While the dot product results in a scalar, the dot product results in a third vector, having a third component, not in the plane of the two being multiplied. The cross product results in a vector that points out of the plane defined by $\vec{e_1}$ and $\vec{e_2}$. For example, if the two vectors are in $\hat{x}$ and j, then the cross-product has a component in $\hat{z}$. It is strictly in $\hat{z}$ if the two vectors are perpendicular to each other (i.e., $\hat{z} = \hat{x} \times \hat{y}$).

The formula for computing the cross product is

$$\vec{a} \times \vec{b} = (a_1\hat{x} + a_2\hat{y} + a_3\hat{z}) \times (b_1\hat{x} + b_2\hat{y} + b_3\hat{z}) = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

The cross product of a vector with itself (or the difference between two vectors) is zero.

The dot product of a third vector $\vec{c}$ with the cross product $\vec{x} \times \vec{\zeta}$ is computed from

$$\vec{c} \cdot (\vec{a} \times \vec{b}) = \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

This computes the volume of a parallelepiped (sort of a squashed sugar cube).

Impact of Analytic Geometry: The most obvious development of algebra, given the creation of analytic geometry, was a detailed analysis of the conic section, using algebra, rather than using drawings made with a compass and ruler. A useful example is the composition of the line and circle, a construction what was used many times over the history of mathematics. Once algebra was invented the composition could be done using formulas.

The first two mathematicians to appreciate this mixture of Euclid's geometry and the new algebra were the French, Fermat and Descartes (Stillwell, 2010, p. 111-115); soon Newton contributed to this effort, by the addition of physics (calculations in acoustics, orbits of the planets, and the theory of gravity and light (Stillwell, 2010, p. 115-117).

Given these new methods, many new solutions to problems emerged. The complex roots of polynomials continued to appear, without any obvious physical meaning. This seem to have been viewed as more of an inconvenience that a problem. Newton's solution to this dilemma was to simply ignore the imaginary cases (Stillwell, 2010, p. 119). The resolution of this was eventually to be found in Bézout's theorem, which states the number of roots of composition of two functions is determined by the product of their degrees. This problem is described as the *construct of equations* (Stillwell, 2010, p. 118). It was finally proved much later, by Bézout (1779), who showed how one needs to count all the roots, including multiple roots, and roots at infinity (Stillwell, 2010, p. 295).

1.3.5 Lec 15 Gaussian Elimination

The method for finding the intersection of equations is based on the recursive elimination of all the variables but one. This method, known as *Gaussian elimination*, works across a broad range of cases, but may be defined in a systematic procedure when the equations are linear in the variables.[47] Rarely do we even attempt to solve problems in several variables of degree greater than 1. But Gaussian eliminations can still work in such cases (Stillwell, 2010, p. 90).

In Appendix B the inverse of a 2x2 linear system of equations is derived. Even for a 2x2 case, the general solution requires a great deal of algebra. Working out a numeric example of Gaussian elimination is more instructive. For example, suppose we wish to find the intersection of the equations

$$\begin{aligned} x - y &= 3 \\ 2x + y &= 2. \end{aligned}$$

This 2x2 system of equations is so simple that you may immediately see the solution: Adding the two equations, and the y term is eliminated, giving $3x = 5$. But doing it this way takes advantage of the specific example, and we need a method for larger systems of equations. We need a generalized (algorithmic) approach. This general approach is *Gaussian elimination*.

Start by writing the equations in a standardized *matrix* format

$$\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}. \tag{1.18}$$

Next, eliminate the lower left term ($2x$) using a scaled version of the upper left term (x). Specifically, multiply the first equation by -2, add it to the second equation, replacing the second equation with the result. This gives

$$\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 2 - 3 \cdot 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}. \tag{1.19}$$

Note that the top equation did not change. Once the matrix is "upper triangular" (zero below the diagonal) you have the solution. Starting from the bottom equation, $y = -4/3$. Then the upper equation then gives $x - (-4/3) = 3$, or $x = 3 - 4/3 = 5/3$.

In principle Gaussian elimination is easy, but if you make a calculation mistake along the way, it is very difficult to find the error. The method requires a lot of mental labor, with a high probability of making a mistake. You do not want to apply this method every time. For example suppose the elements are complex numbers, or polynomials in some other variable such as frequency. Once the coefficients become more complicated, the seeming trivial problem becomes highly error prone. There is a much better way, that is easily verified, which puts all the numerics at the end in a single step.

The above operations may be automated by finding a carefully chosen lower-diagonalization matrix L that does the same operation. For example let

$$L = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}. \tag{1.20}$$

Multiplying Eq. 1.18 by U we find

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix} \tag{1.21}$$

we obtain Eq. 1.19. With a little practice one can quickly and easily find a U that does the job of removing elements below the diagonal.

[47] `https://en.wikipedia.org/wiki/System_of_linear_equations`

In Appendix B the inverse of a general 2x2 matrix is summarized in terms of three steps: 1) swap the diagonal elements, 2) reverse the signs of the off diagonal elements and 3) divide by the determinant $\Delta = ab - cd$. Specifically

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{\Delta}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{1.22}$$

There are very few things that you must memorize, but the inverse of a 2x2 is one of them. It needs to be in your tool-bag of tricks, as you once did for the quadratic formula.

While it is difficult to compute the inverse matrix from scratch (Appendix B), it takes only a few seconds to verify it (steps 1 and 2)

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & -ab + ab \\ cd - cd & -bc + ad \end{bmatrix} = \begin{bmatrix} \Delta & 0 \\ 0 & \Delta \end{bmatrix}. \tag{1.23}$$

Finally, dividing by the determinant gives the 2x2 identity matrix. A good strategy, when you don't trust your memory, is to write down the inverse as best you can, and then verify.

Using 2x2 matrix inverse on our example, we find

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{1+2}\begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 5 \\ -6+2 \end{bmatrix} = \begin{bmatrix} 5/3 \\ -4/3 \end{bmatrix}. \tag{1.24}$$

If you use this method, you will rarely (never) make a mistake, and the solution is easily verified. Either you can check the numbers in the inverse, as was done in Eq. 1.23, or you can substitute the solution back into the original equation.

1.3.6 Lec 16: Transmission (ABCD) matrix composition method

In this section we shall derive the method of composition of linear systems, known by several names as the *ABCD Transmission matrix method*, or in the mathematical literature as the Möbius (bilinear) transformation. By the application of the method of composition, a linear system of equations, expressed in terms of 2x2 matrices, can represent a large family of differential equation networks.

By the application of Ohm's law to the circuit shown in Fig. 1.11, we can model a cascade of such cells. Since the CFA can also treat such circuits, as shown in Fig. 2.2 and Eq. 2.3, the two methods are related to each other via the 2x2 matrix expressions.

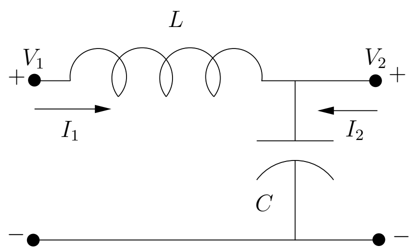


Figure 1.11: *This is a single LC segment (cell) of the transmission line show in Fig. 2.2. It may be modeled by the ABCD model as the product of two matrices, as show below. A more detailed analysis shows that the wave number is $\kappa = \sqrt{ZY} = \sqrt{sL \cdot sC} = s\sqrt{LC}$. It follows that the wave velocity is $c = 1/\sqrt{LC}$. The characteristic resistance is $r_c = \sqrt{Z/Y} = \sqrt{sL/sC} = \sqrt{L/C}$.*

Example of the use of the ABCD matrix composition: In Fig. 1.11 we see the network is composed of a series inductor (mass) having an impedance $Z_l = sL$, and a shunt capacitor (compliance) having an impedance $Z_c = 1/sC$, where s is the Laplace frequency. By Ohm's Law, each impedance is describe by a linear relation between the current and the voltage. For the inductive impedance, applying Ohm's law, we find

$$V_1 - V_2 = Z_l I_1$$

where Z_l is the impedance of the inductor. For the capacitive impedance, applying Ohm's law we find

$$V_2 = (I_1 + I_2)Z_c,$$

where Z_c is the impedance of the capacitor. Each of these linear impedance relations may be written in matrix form. The series inductor equation gives

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & Z_l \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_1 \end{bmatrix}, \tag{1.25}$$

while the shunt capacitor equation yields

$$\begin{bmatrix} V_2 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ Y_c & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix}, \tag{1.26}$$

where $Y_c = 1/Z_c$ is called the *admittance*.

When the second matrix equation for the shunt admittance (Eq. 1.26) is substituted into the series impedance equation (Eq. 1.25), we find the composed form ABCD matrix for the cell, as the product of two matrices

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} 1 & Z_l \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ Y_c & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ I_2 \end{bmatrix} = \begin{bmatrix} 1 + Z_l Y_c & Z_l \\ Y_c & 1 \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}. \tag{1.27}$$

This equation characterizes the relations between the input and output voltage and current of the cell of Fig. 1.11 (p. 52).

For example, the ratio of the output to input voltage with the output unloaded ($I_2 = 0$)

$$\left.\frac{V_2}{V_1}\right|_{I_2=0} = \frac{1}{A} = \frac{1}{1 + Z_l Y_c} = \frac{Z_c}{Z_c + Z_l}$$

is known as the *voltage divider equation*. Can you derive the formula for the current divider equation? Hint: $V_2 = 0$ For each matrix the input voltage and current are on the left (e.g., $[V_1, I_1]^T$), while the output voltage and current is on the right (e.g., $[V_2, -I_2]^T$). The sign of the output current must be negative, so that it is into the input node of the adjoining cell. In this way we may chain these matrices together. On the left the input current is into the node. The sign of current I_2 must be reversed in sign, so that it will be into the node of the adjacent cell to the right (every current is defined to be into each node).

Matrix composition: Matrix multiplication represents a composition of 2x2 matrices, because the output of the second matrix is the input of the first (this follows from the definition of composition). Thus the ABCD matrix is also known as the *chain matrix*. The general expression for a "chain matrix" is

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}, \tag{1.28}$$

which is discussed in more detail in Section 3.3.2 (p. 104).

1.3.7 Lec 17: Riemann Sphere: 3^d extension of chord and tangent method

Once algebra was formulated c830 CE, mathematics was able to expand beyond the limits placed on it by geometry on the real plane, and the verbose descriptions of each problem in prose (Stillwell, 2010, p. 93). The geometry of Euclid's Elements had paved the way, but after 2000 years, the addition of the language of algebra would change everything. The analytic function was a key development, heavily used by both Newton and Euler. Also the investigations of Cauchy made important headway with his work on complex variables. Of special note was integration and differentiation in the complex plane of complex analytic functions, which is the topic of stream 3.

It was Riemann, working with Gauss, who made the breakthrough, with the concept of the *extended complex plane*. The idea was based on the composition of a line with the sphere, similar to the derivation of Euclid's formula for Pythagorean triplets (Fig. 2.3, p. 87). But the impact was unforeseen, and it changed analytic mathematics forever, and the physics that was supported by it, by simplifying many important integrals, to the extreme. This idea is captured by the *Fundamental Theorem of Complex Integral Calculus* (Table 1.6 p. 63) and 4.3.1, p. 118).

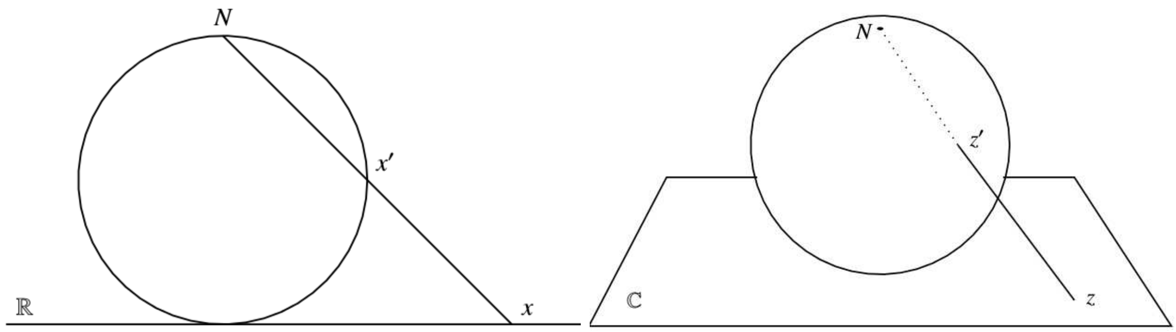


Figure 1.12: *The left panel shows how the real line may be composed with the circle. Each real x value maps to a corresponding point x' on on the unit circle. The point $x \to \infty$ then naturally maps to the north pole N. This simple idea may be extended with the composition of the complex plane with the unit sphere, thus mapping the plane onto the sphere. As with the circle, the point on the complex plane $z \to \infty$ maps onto the north pole N. This construction is important because while the plane is open (does not include $z \to \infty$), the sphere is analytic at the north pole. Thus the sphere defines the* closed extended plane. *Figure from Stillwell (2010, pp. 299-300).*

The idea is outlined in Fig. 1.12. On the left is a circle and a line. The difference between this case and the derivation of the Pythagorean Triplets is, that the line starts at the north pole, and ends on the real $x \in \mathbb{R}$ axis at point x. At point x', the line cuts through the circle. Thus the mapping from x to x' takes every point on $\mathbb{R}$ to a point on the circle. For example, the point $x = 0$ maps to the south pole (not indicated). To express x' in terms of x one must composition of the line and the circle, similar to the composition used in Fig. 2.3 (p. 87). The points on the circle, indicated here by x', require a traditional polar coordinate system, having a unit radius and an angle defined between the radius and a vertical line passing through the north pole. When $x \to \infty$ the point $x' \to N$, known as the *point at infinity*. But this idea must go much further, as shown on the right half of Fig. 1.12.

Here the real tangent line is replaced by the a tangent complex $z \in \mathbb{C}$, and the complex puncture point $z' \in \mathbb{C}$, in this case on the complex sphere, called the *extended complex plane*. This is a natural extension of the chord/tangent method on the left, but with significant consequences. The main difference between the complex plane z and the extended complex plane, other than the coordinate system, is what happens at the north pole. On the plane the point at $|z| = \infty$ is not defined, whereas on the sphere the point at the north pole is simply another point, like every other point on the sphere.

Mathematically the plane is said to be an *open set* since the limit $z \to \infty$ is not defined, whereas on the sphere the point z' is a member of a *closed set*, since the north pole *is* defined. The distinction between an open and closed set is important because the closed set allows the function to be analytic at the north pole, which it cannot be on the plane (since the point at infinity is not defined).

The z plane may be replaced with another plane, say the $w = F(z) \in \mathbb{C}$ plane, where w is some function F of $z \in \mathbb{C}$. We shall limit ourselves to *complex analytic functions* of z, namely $w = F(z) = u(x,y) + v(x,y)\jmath = \sum_{n=0}^{\infty} z^n$.

In summary, given a point $z = x + y\jmath$ on the open complex plane, we map it to $w = F(z) \in \mathbb{C}$, the complex $w = u + v\jmath$ plane, and from there to the closed extended complex plane $w'(z)$. The point of doing this is that it allows us to allow the function $w'(z)$ to be analytic at the north pole, meaning it can have a convergent Taylor series at the point at infinity $z \to \infty$.

Möbius bilinear transformation

In mathematics the *Möbius transformation* has special importance because it is linear in its action. In the engineering literature this transformation is known as the *bilinear transformation.* Since we are engineers we shall stick with the engineering terminology. But if you wish to read about this on the internet, be sure to also search for the mathematical term, which may be better supported.

When a point on the complex plane $z = x+y\jmath$ is composed with the bilinear transform $(a, b, c, d \in \mathbb{C})$, the result is $w(z) = u(x, y) + v(x, y)\jmath$

$$w = \frac{az + b}{cz + d} \tag{1.29}$$

the transformation is a cascade of four independent compositions

1. translation $(w = z + b)$
2. scaling $(w = |a|z)$
3. rotation $(w = \frac{a}{|a|}z)$ and
4. inversion $(w = \frac{1}{z})$

Each of these transformations are a special case of Eq. 1.29, with the inversion the most complicated. A wonderful video showing the effect of the bilinear (Möbius) transformation on the plane is available that I highly recommended you watch it: https://www.youtube.com/watch?v=0z1fIsUNhO4

When the extended plane (Riemann sphere) is analytic at $z = \infty$, one may take the derivatives there, and meaningfully integrate through ∞. When the bilinear transformation rotates the Riemann sphere, the point at infinity is translated to a finite point on the complex plane, revealing normal characteristics. A second way to access the point at infinity is by inversion, which takes the north pole to the south pole, swapping poles with zeros. Thus a zero at infinity is the same as a pole at the origin, and vice-versa.

This construction of the Riemann sphere and the Mb̈ious (bilinear) transformation allow us to fully understand the point at infinity, and treat it like any other point. If you felt that you never understood the meaning of the point at ∞ (likely), then this should help.

1.3.8 Lec 18: Complex analytic mappings (Domain coloring)

One of the most difficult aspects of complex functions of a complex variable is understanding what's going on. For example, $w = \sin(\sigma)$ is trivial when $\omega = 0$ (i.e., s is real), because $\sin(\sigma)$ is then real. But the general case, $w(s) = \sin(s) \in \mathbb{C}$, is not so easily visualized, because with argument $s \in \mathbb{C}$, the function $w \in \mathbb{C}$, since the mapping is from the $s = \sigma + \omega\jmath$ plane to the $w(\sigma, \omega) = u(\sigma, \omega) + v(\sigma, \omega)\jmath$ plane, which in general is complicated. Fortunately with computer software today, this problem can be solved by adding color to the graph. A Matlab script `zviz.m` has been used to make these make the charts shown here.[48] This tool is known as *Domain coloring.* Rather than plotting $u(\sigma, \omega)$ and $v(\sigma, \omega)$ separately, domain coloring allows us to display the entire function on one plot. Note that for this visualization we see the complex polar form of $w(s)$, rather than a rectangular (u, v) graph.

In Fig. 1.13 we show this color code as a 2x2 dimensional domain-coloring graph. The color (hue) represents the phase, and intensity (dark to light) the magnitude. On the left is the reference condition, the identity mapping $(w = s)$. Red is $0°$, violet is $90°$, blue is $135°$, blue-green is $180°$ and sea-green is $-90°$ (or $270°$). The hue (darkness) represents the magnitude. Since the function $w = s$ has a dark spot (a zero) at $s = 0$, becoming brighter as we move radially away from the zero. The figure on the right is $w = F(z - 1)$, which moves the zero to the right by 1. In summary, domain coloring gives the full picture of the complex analytic function mappings $w(x, y) = u(x, y) + v(x, y)\jmath$.

[48] URL for `zviz.m`: http://jontalle.web.engr.illinois.edu/uploads/298/zviz.m

Two examples are given in Fig. 1.14 to help you interpret the two complex mappings $w = e^s$ (left) and its inverse $s = \ln(w)$. The exponential is relatively easy to understand because $w = e^{\sigma}e^{\omega \jmath}$. The red region is where $\omega \approx 0$ in which case $w \approx e^{\sigma}$. As σ becomes large and negative, $w \to 0$ so the entire field at the left becomes dark. The field is becoming light on the right where $w = e^{\sigma} \to \infty$. If we let $\sigma = 0$ and look along the ω axis, we see that the function is changing phase, green (-90°) at the top and violet (90°) at the bottom.

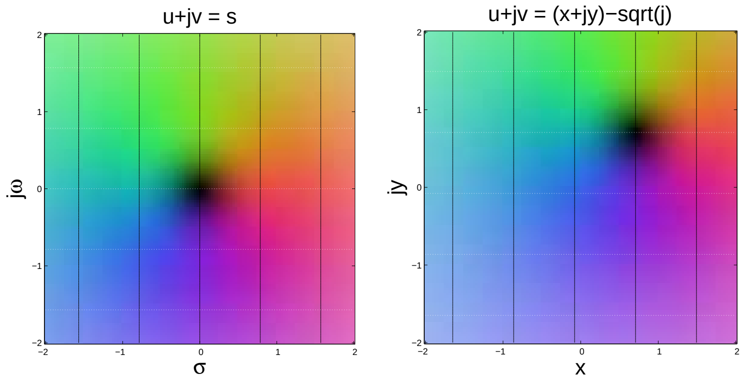


Figure 1.13: *On the left is a color map showing the definition of the complex mapping from the $s = \sigma + \omega\jmath$ plane to the $w(s) = u(\sigma,\omega) + v(\sigma,\omega)\jmath$ plane. This mapping from one complex plane $s = \sigma + \omega\jmath$ to another $w = u(\sigma,\omega) + v(\sigma,\omega)\jmath$, is visualized by the use of intensity (light/dark) to indicate magnitude, and color (hue) to indicate angle (phase). On the left $w(s) = s = \sigma + \omega\jmath$ (i.e., $u = \sigma$ and $v = v$). On the right is $w(z) = z - \sqrt{\jmath}$, a shift to the right and up by $\sqrt{2}/2 = 0.707$.*

Note the zero in $\ln(w)$ at $w = 1$. The root of the log function is $\log(s_r) = 0$ is $s_r = 1$, since $\log(1) = 0$. More generally, taking the log of $w = |w|e^{\phi\jmath}$ gives $s = log(|w|) + \phi\jmath$. Thus s can only be zero when the angle of w is zero (i.e., $\phi = 0$).

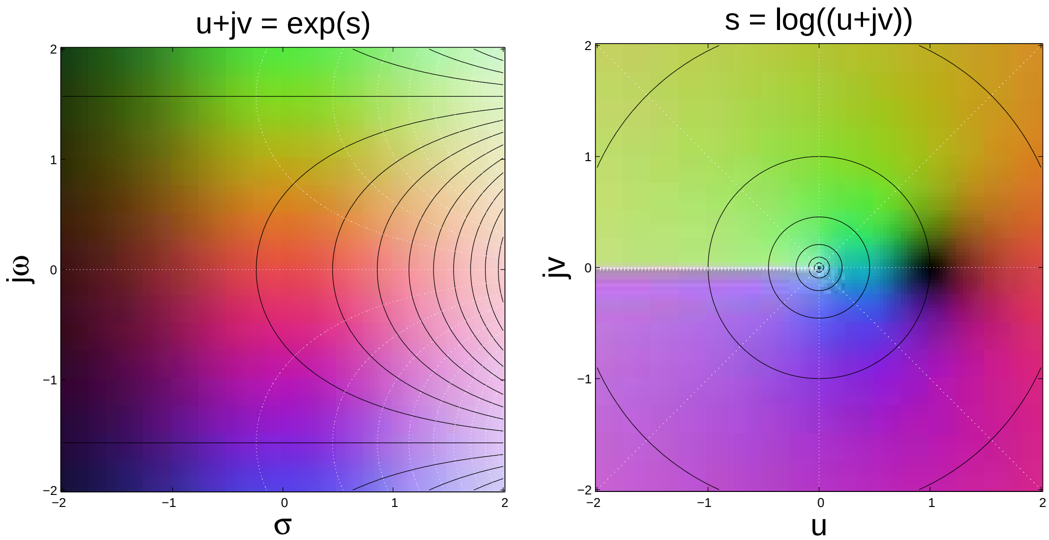


Figure 1.14: *On the left is the function $w(s) = e^s$ and on the right is $s = log(w)$.*

1.3.9 Lec 19: Signals: Fourier transforms

Two basic transformations in engineering mathematics are the Fourier and the Laplace transforms, which deal with time–frequency analysis.

The Fourier transform takes a time domain signal $f(t) \in \mathbb{R}$ and transforms it to the frequency domain $\omega \in \mathbb{R}$, where it is complex ($F(\omega) \in \mathbb{C}$). For the Fourier transform, both the time $-\infty < t < \infty$ and frequency $\infty < \omega < \infty$ are strictly real. The time vs. frequency relationship, between $f(t)$ and its transform $F(\omega)$, is indicated by the double arrow symbol $\leftrightarrow$

$$f(t) \leftrightarrow F(\omega).$$

Since the FT obeys superposition, it is possible to define a FT of a complex time function. This is useful in certain applications (i.e, Hilbert envelope, Hilbert transforms), but because the FT is not complex analytic, can be error prone if your not fully in charge of the method.

The Laplace transform takes real signals $f(t) \in \mathbb{R}$ as a function of real time $t \in \mathbb{R}$, that are strictly zero for negative time ($f(t) = 0$ for $t < 0$), and transforms them into complex functions $F(s) \in \mathbb{C}$ of complex frequency $s = \sigma + \omega\jmath$. As for the Fourier transform, there is the convenient notation $f(t) \leftrightarrow F(s)$.

When a signal is zero for negative time $f(t < 0) = 0$, it is said to be *causal*, and the resulting transform $F(s)$ is then complex analytic over significant regions of the s plane. For a function of time to be causal, time *must* be real, since if it were complex, it would lose the order property.

Restriction on a function (e.g., real, causal, periodic, positive real part, etc.) are called a *symmetry* property. There are many forms of symmetry (Section 1.3.11, p. 61). The concept of symmetry is very general and widely used in both mathematics and physics, where it is more generally known as *Group theory*. We shall restrict ourselves to a few very basic symmetries.

Periodic signals: Besides these two basic types of time–frequency transforms, there are several variants that depend on the symmetry in time and frequency. For example, when the time signal is sampled (discrete in time), the frequency response becomes periodic, leading to the *Discrete-time Fourier transform* (DTFT). When a time response is periodic, the frequency response is sampled (discrete in frequency), leading to the *Fourier Series*. These two symmetries may be simply characterized as *periodic in time* $\Rightarrow$ discrete in frequency, and *periodic in frequency* $\Rightarrow$ discrete in time. In Section 3.4.2 we shall explain these concepts in greater detail, with examples. When a function is both discrete in time and frequency, it is necessarily periodic in time and frequency, leading to the *Discrete Fourier Transform* (DFT). The DFT is typically computed with an algorithm called the *Fast Fourier Transform* (FFT), which can dramatically speed up the calculation when the data is composite (e.g., a power of 2 in length).

A very important case is that of functions that are both causal (in time) and periodic (in frequency). The best know example is the class of signals that have *z transforms*. These are causal (one-sided in time), discrete-time signals. The harmonic series is the z-transform step function, discrete and one-sided in time, and analytic within the ROC, in the complex frequency (z) domain.

Summary: The definitions of the FT and LT transforms are superficially similar. The key difference is that the time response of the Laplace transform is causal, leading to complex analytic frequency responses. The frequency response of the Fourier transform is real, thus typically not analytic. These are not superficial differences. The concept of symmetry is helpful in understanding the many different types of time-frequency transforms. Two fundamental types of symmetry are causality and periodicity.

Definition of the Fourier transform: The forward transform takes $f(t)$ to $F(\omega)$ while the inverse transform takes $F(\omega)$ to $\widetilde{f}(t)$. The tilde symbol indicates that in general recovered inverse transform signal can be slightly different from $f(t)$. We will give examples of this below.

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-\jmath\omega t} dt \qquad \widehat{f}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega) e^{\jmath\omega t} d\omega \tag{1.30}$$

$$F(\omega) \leftrightarrow f(t) \qquad \widehat{f}(t) \leftrightarrow F(\omega). \tag{1.31}$$

Properties of Fourier Transforms:

1. Both time t and frequency ω are real.
2. For the forward transform (time to frequency), the sign of the exponential is negative.
3. The limits on the integrals in both the forward and reverse FTs are $[-\infty, \infty]$.

4. When taking the inverse FT (IFT), the normalization factor of $1/2\pi$ is required to cancel the 2π in the differential of the integral $d\omega/2\pi = df$, where f is frequency in [Hz], and ω is the radian frequency [rads].

5. The Fourier step function may be defined by the use of superposition of 1 and $\text{sgn}(t) = t/|t|$ as

$$\widetilde{u}(t) \equiv \frac{1+\text{sgn}(t)}{2} = \begin{cases} 1 & \text{if } t > 0 \\ 1/2 & t = 0 \\ 0 & \text{if } t < 0 \end{cases}.$$

   Taking the FT of a delayed step function

$$\widetilde{u}(t-T_0) \leftrightarrow \frac{1}{2}\int_{-\infty}^{\infty} [1 - \text{sgn}(t-T_0)]\, e^{-j\omega t} dt = \pi\tilde{\delta}(\omega) + \frac{e^{-j\omega T_0}}{j\omega}$$

   Thus the FT of the step function has the term $\pi\delta(\omega)$ due to the 1 in the definition of the Fourier step. This term introduces a serious flaw with the FT of the step function: While it appears to be causal, it is not.

6. The convolution $\widetilde{u}(t) \star \widetilde{u}(t)$ is not defined because both $1 \star 1$ and $\widetilde{\delta}^2(\omega)$ do not exist (and cannot be defined).

7. The inverse FT has convergence problems whenever there is a discontinuity in the time response. This we indicate with a hat over the reconstructed time response. The error between the target time function and the reconstructed is zero in the root-mean sense, but not point-wise.

   Specifically, at the discontinuity point for the Fourier step function ($t = 0$), $\widetilde{u}(t) \neq u(t)$, yet $\int |\widetilde{u}(t) - u(t)|^2 dt = 0$. At the point of the discontinuity the reconstructed function displays *Gibbs ringing* (it oscillates around the step, hence does not converge at the jump).[49] More about this in Section 3.4.2.

8. The FT is not always analytic in ω, as in this example of the step function. The step function cannot be expanded in a Taylor series about $\omega = 0$, because $\widetilde{\delta}(\omega)$ is not analytic in ω.

9. The Fourier δ function is denoted $\widetilde{\delta}(t)$, to differentiate it from the Laplace delta function $\delta(t)$. They differ because the step functions differ, due to the convergence problem described above.

10. One may define

$$\widetilde{u}(t) = \int_{-\infty}^{t} \widetilde{\delta}(t) dt,$$

   and define the somewhat questionable notation

$$\widetilde{\delta}(t) =: \frac{d}{dt}\widetilde{u}(t),$$

   since the Fourier step function is not analytic.

11. The $\text{rec}(t)$ function is defined as

$$\text{rec}(t) = \frac{\widetilde{u}(t) - \widetilde{u}(t-T_0)}{T_0} = \begin{cases} 0 & \text{if } t > 0 \\ 1/T_0 & 0 < t < T_0 \\ 0 & \text{if } t < 0 \end{cases}.$$

   It follows that $\widetilde{\delta}(t) = \lim_{T_0 \to 0}$. Like $\widetilde{\delta}(t)$, the $\text{rec}(t)$ has unit area.

[49] https://en.wikipedia.org/wiki/Gibbs_phenomenon

Table 1.4: *The following table provides a short table of important Fourier Transforms. Note $a > 0 \in \mathbb{R}$ has units [rad/s]. To flag this necessary condition, we use $|a|$ to assure this condition will be met. The other constant $T_0 \in \mathbb{R}$ [s] has no restrictions, other than being real. Complex constants may not appear as the argument to a delta function, since complex numbers do not have the order property.*

$f(t) \leftrightarrow F(\omega)$	Name
$\tilde{\delta}(t) \leftrightarrow 1(\omega)$	Dirac
$1(t) \leftrightarrow 2\pi\tilde{\delta}(\omega)$	Dirac
$\text{sgn}(t) = \frac{t}{\lvert t \rvert} \leftrightarrow \frac{2}{\jmath\omega}$	
$\tilde{u}(t) = \frac{1(t) + \text{sgn}(t)}{2} \leftrightarrow \pi\tilde{\delta}(\omega) + \frac{1}{\jmath\omega}$	step
$\tilde{\delta}(t - T_0) \leftrightarrow e^{-\jmath\omega T_0}$	delay
$\tilde{\delta}(t - T_0) \star f(t) \leftrightarrow F(\omega)e^{-\jmath\omega T_0}$	delay
$\tilde{u}(t)e^{-\lvert a \rvert t} \leftrightarrow \frac{1}{\jmath\omega + \lvert a \rvert}$	exp
$\text{rec}(t) = \frac{1}{T_0}\left[\tilde{u}(t) - \tilde{u}(t - T_0)\right] \leftrightarrow \frac{1}{T_0 s}\left(1 - e^{-\jmath\omega T_0}\right)$	pulse
$\tilde{u}(t) \star \tilde{u}(t) \leftrightarrow \tilde{\delta}^2(\omega)$ Not defined	NaN
FT Properties	
$\frac{d}{dt}v(t) \leftrightarrow \jmath\omega V(\omega)$	deriv
$f(t) \star g(t) \leftrightarrow F(\omega)G(\omega)$	conv

1.3.10 Lec 20: Laplace transforms

When dealing with engineering problems it is convenient to separate the signals we use from the systems that process them. We do this by treating signals, such as a music signal, differently from a system, such as a filter. In general signals may start and end at any time. The concept of causality has no mathematical meaning in signal space. Systems, on the other hand, obey very rigid rules (to assure that they remain physical). These physical restrictions are described in terms of nine *Network Postulates*, which are briefly discussed in Lecture 1.3.11, and again in greater detail in Section 3.5.1.

Definition of the Laplace transform: The forward and inverse Laplace transforms are

$$F(s) = \int_{0^-}^{\infty} f(t)e^{-st}dt \qquad f(t) = \frac{1}{2\pi\jmath}\int_{\sigma_0-\infty\jmath}^{\sigma_0+\infty\jmath} F(s)e^{st}ds \tag{1.32}$$

$$F(s) \leftrightarrow f(t) \qquad f(t) \leftrightarrow F(s) \tag{1.33}$$

1. Time $t \in \mathbb{R}$ and *Laplace frequency* is defined as $s = \sigma + \omega\jmath \in \mathbb{C}$.

Table 1.5: *The following table provides a brief table of important Laplace Transforms. Assume that:* $f(t), \delta(t), u(t), rec(t), T_0, e, \in \mathbb{R}$ *and* $F(s), G(s), s, a \in \mathbb{C}$.

$f(t) \leftrightarrow F(s)$	Name
$\delta(t) \leftrightarrow 1$	Dirac
$\delta(t-T_0) \leftrightarrow e^{-sT_0}$	delay
$\delta(t-T_0) \star f(t) \leftrightarrow F(s)e^{-sT_0}$	–
$u(t) \leftrightarrow \dfrac{1}{s}$	step
$u(t) \star u(t) = tu(t) \leftrightarrow 1/s^2$	ramp
$u(t) \star u(t) \star u(t) = \dfrac{1}{2}t^2u(t) \leftrightarrow 1/s^3$	–
$e^{at}u(t) \leftrightarrow \dfrac{1}{s-a}$	exp
$\mathrm{rec}(t) = \dfrac{1}{T_0}\left[u(t) - u(t-T_0)\right] \leftrightarrow \dfrac{1}{T_0 s}\left(1 - e^{-sT_0}\right)$	pulse
LT Properties	
$\dfrac{d}{dt}f(t) \leftrightarrow sF(s)$	deriv
$f(t) \star g(t) \leftrightarrow F(s)G(s)$	conv
$u(t) \star f(t) = \displaystyle\int_{0^-}^{t} f(t)dt \leftrightarrow \dfrac{F(s)}{s}$	conv

2. When taking the forward transform ($t \to s$), the sign of the exponential is negative. This is necessary to assure that the integral converges when the integrand $f(t) \to \infty$ as $t \to \infty$. For example, if $f(t) = e^t u(t)$ (i.e., without the negative σ exponent), the integral does not converge.

3. The target time function $f(t < 0) = 0$ (i.e., it must be causal). The time limits are $0^- < t < \infty$. Thus the integral must start from slightly below $t = 0$ to integrate over a delta functions at $t = 0$. For example if $f(t) = \delta(t)$, the integral must include both sides of the impulse. If you wish to include non-causal functions such as $\delta(t + 1)$ it is necessary to extend the lower limit to include it. In such cases simply set the lower limit to $-\infty$, and let the integrand to determine the limits.

4. The limits on the integrals of the forward transform are $t : (0^-, \infty) \in \mathbb{R}$, and the reverse LTs are $[\sigma_0 - \infty\jmath, \sigma_0 + \infty\jmath] \in \mathbb{C}$. These limits will be further discussed in Section 1.4.9 (p. 72).

5. When taking the inverse FT (IFT), the normalization factor of $1/2\pi\jmath$ is required to cancel the $2\pi\jmath$ in the differential ds of the integral.

6. The frequency for the LT must be is complex, and in general $F(s)$ is complex analytic for $\sigma > \sigma_0$. It follows that the real and imaginary parts of $F(s)$ are related. Given $\Re F(s)$ it is possible to find $\Im F(s)$ (Boas, 1987). More on this in Section 3.4.2 (p. 107).

7. To take the inverse Laplace transform, we must learn how to integrate in the complex s plane. This will be explained in Section 1.4.5 (p. 71).

8. The Laplace step function is defined as

$$u(t) = \int_{-\infty}^{t} \delta(t)dt = \begin{cases} 1 & \text{if } t > 0 \\ \text{NaN} & t = 0 \\ 0 & \text{if } t < 0 \end{cases}.$$

   Alternatively one could define $\delta(t) = du(t)/dt$.

9. It is easily shown that $u(t) \leftrightarrow 1/s$ by direct integration

$$F(s) = \int_{0}^{\infty} u(t)\, e^{-st}dt = -\frac{e^{-st}}{s}\bigg|_0^\infty = \frac{1}{s}.$$

   With the LT there is no Gibbs effect, as the step function, with a true discontinuity, is exactly represented by the inverse LT.

10. In many physical applications, the Laplace transform takes the form of a ratio of two polynomials. In such case the roots of the numerator polynomial are call the *zeros* while the roots of the denominator polynomial are called the *poles*. For example the LT of $u(t) \leftrightarrow 1/s$ has a pole at $s = 0$, which represents integration, since

$$u(t) \star f(t) = \int_{-\infty}^{r} f(\tau)d\tau \leftrightarrow \frac{F(s)}{s}.$$

11. The LT is quite different from the FT in terms of its analytic properties. For example, the step function $u(t) \leftrightarrow 1/s$ is complex analytic everywhere, except at $s = 0$. The FT of $1 \leftrightarrow 2\pi\tilde{\delta}(\omega)$ is not analytic anywhere.

    Once complex integration in the complex plane has been defined (Section 1.4.2, p. 117), we can justify the definition of the inverse LT (Eq. 1.32).

1.3.11 Lec 21: The 9 postulates of systems

Systems of differential equations, such as the wave equation, need a mathematical statement of underlying properties, which we present here in terms of nine *network postulates*:

(P1) *causality* (non-causal/acausal): Causal systems respond when acted upon. Virtually all physical systems obey causality. An example of a causal system is an integrator, which has a response of a step function. Filters are also examples of causal systems. Signals represent acausal responses. They do not have a clear begining or end, such as the sound of the wind or traffic noise.

(P2) *linearity* (nonlinear): Linear systems obey superpositon. If two signals $x(t)$ and $y(t)$ are the inputs to a linear system, producing outputs $x'(t)$ and $y'(t)$, then if the inputs are presented together as $ax(t) + by(t)$, with weights $a, b \in \mathbb{R}$, then the output will be $ax'(t) + by'(t)$. If either a, b are zero, that signal is removed from the output.

Nonlinear system mix two inputs, thereby producing other signals not present in the input. For example, if the inputs to a nonlinear system are two sine waves, the output will contain distortion components, having frequencies not present at the input. An example of a nonlinear system is one that multiplies the two inputs. A second is a diode, which rectifies a signal, letting current flow only in one direction. Most physical systems have some degree of nonlinear response, but this is not always desired. Other systems are designed to be nonlinear, such as the diode example.

(P3) *passive* (active): An active system has a power source, such as a battery while a passive system has no power source. While you may consider a transistor amplifier to be active, it is only so when connected to a power source.

(P4) *real* (complex) time response : Typically systems are "real in, real out." They do not naturally have complex responses (real and imaginary parts). While a Fourier transform takes real inputs and produces complex outputs, this is not an example of a complex time response. P4 is a characterization of the input signal, not its Fourier transform.

(P5) *time-invariant* (time varying): For a system to be time varying system its properties mus depend on when the signal starts and stops. If the output, relative to the input, is independent of the starting time, then the system is time-invariant.

(P6) *reciprocal* (anti-reciprocal): In many ways this is the most difficult propriety to understand. It is best characterized by the ABCD matrix. If $B = C$ it is said to be reciprocal. If $B = -C$ it is said to be anti-reciprocal. The loudspeaker is anti-reciprocal, which is why it is modeled by the gyrator rather than a transformer. All non-reciprocal systems are modeled by such gyrator, which swap the force and flow variables. In some ways this property is beyond the scope of this book.

(P7) *reversibility* (non-reversible): If the system can be flipped, between input and output, the it is said to be reversible. In terms of the ABCD parameters, if $A = D$ is is reversible.

(P8) *space-invariant* (space-variant): If a system operates independently as a function of where it physically is in space, then it is space-invariant. When the parameters that characterize the system depend on position, it is space-variant.

(P9) *quasi-static* (multi-modal): Quasi-statics follows from systems that are small compared to the wavelength. This is a very general assumption, that must be false when the frequency is raised and the wavelength becomes short. Thus this is also know as the *long-wavelength* approximation. It is a very powerful tool in modeling systems such as transmission lines.

Each postulate has two (in one case three) categories. For example for (P2) a system is either linear or non-linear and for (P1) is either causal, non-causal or acausal. P6 and P9 only apply to 2-port networks (those having an input and an output). The others can apply to both a 2- or 1-port networks (e.g., an impedance is a 1-port).

Include figs examples of different types of systems.

Related forms of these postulates had been circulating in the literature for many years, widely accepted in the network theory literature (Van Valkenburg, 1964a,b; Ramo et al., 1965). The first six of these were formally introduced Carlin and Giordano (1964), and (P7-P9) were added by Kim et al. (2016).

1.3.12 Lec 22: Exam II (Evening Exam)

1.4 Stream 3: Scalar (Ordinary) Differential Equations

Stream 3 is ∞, a concept which inspires "very large." In the case of calculus, ∞ means "very small," since taking a limit requires small numbers. Taking the limit means you never reaching the target. This is a concept that the Greeks called *Zeno's paradox* (Stillwell, 2010, p. 76).

When speaking of the class of *ordinary* (versus *vector*) differential equations, the term *scalar* is preferable, since the term "ordinary" is vague. There are a special subset of Fundamental theorems for scalar calculus, all of which are about integration, as summarized in Table 1.6, starting with Leibniz's theorem. These will be discussed below, and in Sections 1.4.1 (p. 115) and 1.4.4–1.4.6 (pp. 117-118).

Table 1.6: *Summary of the fundamental theorems of integral calculus, each of which deals with integration. There are at least four theorems related to scalar calculus, and four more to vector calculus.*

Name	Mapping	p.	Description
Leibnez	$\mathbb{R}^1 \to \mathbb{R}^0$	64	Area under a real curve.
Cauchy	$\mathbb{C} \to \mathbb{C}^0$	64	Residue integration and analytic functions.
Gauss's Law	$\mathbb{R}^3 \to \mathbb{R}^2$	74	Conservation of mass and charge crossing a closed surface.
Stokes	$\mathbb{R}^3 \to \mathbb{R}^1$	74	Relates line integrals to the rate of change of the flux crossing an open surface.
Green	$\mathbb{R}^2 \to \mathbb{R}_0$	74	Special case of Stokes for a plane
Helmholtz		75	Every differentiable vector field may be decomposed into a dilatation and a rotation.

Following the integral theorems on scaler calculus are those on vector calculus, without which there could be no understanding of Maxwell's equations, the crown jewel of modern science. Of these the Fundamental theorem of complex calculus, also known as Helmholtz theorem, Gauss's Law and Stokes theorem, form the corner stone of modern vector field analysis. These theorems allow one to connect the differential (point) and macroscopic (integral) relationships. For example, Maxwell's equations may be written as either vector differential equations, as shown by Heaviside (along with Gibbs and Hertz).[50]

The beginning of modern Mathematics

The birth of mathematics as we know it today, occurred during the 16th to 18th centuries, perhaps starting with Galileo, Descartes, Fermat, Newton, the Bernoulli family, and Euler, as outlined in Fig. 1.4. Galileo was formative due to his fame, fortune, and his stance against the powerful Catholic church and its establishment. His creativity in scientific circles was certainly well known due to his many skills and accomplishments. Fermat and Descartes were at the forefront of merging algebra and geometry, but while Fermat kept meticulous notebooks, he did not publish what he knew, and tended to be secretive.

Much was yet to be done, by Newton and Leibniz, with the development of calculus, again based on the analytic function (i.e., Taylor series), by integration of functions based on their power series representation, term by term. This was powerful technique, but as stated, incomplete, because the series is only valid within the ROC, but more importantly, Newton (and others) failed to recognize the powerful generalization to complex analytic series, and their functions. The breakthrough was Newton's publication of Principia (1687).

[50] `https://en.wikipedia.org/wiki/History_of_Maxwell\%27s_equations`

Following Newton's lead, the secretive and introverted behavior of the typical mathematician, dramatically changed with the Bernoulli family. The oldest brother, Jacob, taught his much younger brother Johann, who then taught his son Daniel, and Johann's star pupil, Euler. Euler first mastered all the tools and then published, with prolific intensity previously unknown.

Euler and the circular functions: The first major task was to understand the family of analytic circular functions, $e^x, \sin(x), \cos(x)$, and $\log(x)$, a task begun by the Bernoulli family, but mastered by Euler. Euler sought relations between these many functions, some of which are not thought of as being related, such as the log and sin functions. The connection that may "easily" be made is through the Taylor series. By the manipulation of the analytic series representations, the relationship between e^x, and the sin and cos was precisely captured with Euler's formula

$$e^{x\jmath} = \cos(x) + \jmath \sin(x).$$

Euler's formula not only relates the exponential function to the circular functions, but implicates $j = \sqrt{-1}$ as well. While every high school student has this relation drilled into their collective heads, it is doubtful (i.e., rare) that they fully appreciate its meaning. While Euler found these relationships with the use of the power series, namely analytic functions, neither he, nor those who followed, seemed to appreciated the importance of $\jmath$ and its potential role. For example, Newton famously ignored imaginary numbers, and called them imaginary, in a disparaging (pejorative) way. Given Newton's prominence, that certainly must have attenuated any interest in any complex algebra, even though it had been previously quantified by Bombelli in 1525, likely based on his amazing finding in the Vatican library of Diophantus' book *Arithmetic*.

While Euler was fluent with $\jmath = \sqrt{-1}$, he did not consider functions to be complex analytic. That concept was first explored by Cauchy, almost a century later. The missing link to the concept of complex analytic is the definition of the derivative with respect to the complex argument (i.e., $dF(z)/dz$ with $z = x + y\jmath$), without which the Taylor coefficients are not defined.

1.4.1 Lec 23: Fundamental theorem of complex calculus

History of the Fundamental theorem of calculus: It some real sense, the story of calculus begins with the *Fundamental theorem of calculus* (FTC), also know generically as *Leibniz's formula*. Simply stated, the area under a real curve $x, f(x) \in \mathbb{R}$, only depends on the end points

$$f(x) = f(0) + \int_0^x F(\xi) d\xi. \tag{1.34}$$

This makes sense intuitively, by a simple drawing of $f(x)$ (or $f(t)$), where x represents a spatial dimension [meters], and t a temporal variable [seconds]. As before, when we start from a physical intuition, to create a key mathematical concept, the area under a function agrees with common sense. For example, if we travel from point a to $x = b$, along a straight line measured by ξ, the distance traveled depends on the length of the line. This is reminiscent of Eq. 1.1 (p. 15), which is in $\in R^2$, since is about lengths and areas.

Fundamental theorems of real and complex calculus: Leibniz's formula gives the area under a real analytic function $F(x), f(x), x \in \mathbb{R}$, the integral only depends only on the end points. It follows (Leibniz Rule) that

$$\frac{d}{dx} f(x) = F(x).$$

The integral is also known as the *anti-derivative*.

Given a complex analytic impedance $Z(s)$, due to the *Fundamental theorem of complex calculus* (FTCC), we may integrate $Z(s)e^{st}$ in the complex $s = \sigma + \omega\jmath$ plane

$$\zeta(s) = \zeta(s_0) + \int_{s_0}^{s} Z(s) e^{st} ds. \tag{1.35}$$

If we compare this to Eq. 1.34, they only differ in that the path of the integral is complex, making the integral a line integral over $s \in \mathbb{C}$, rather than a real integral over $x \in \mathbb{R}$. Because the integrand includes the complex exponential, the integral becomes the Laplace transform (Section 1.3.10, p. 59) if we choose the limits to be $\sigma_0 \pm \omega\jmath$. This is precisely why it is so important to study integration in the complex plane, so that we can evaluate inverse Laplace transforms.

Yet the FTCC states that the integral only dependents on the end points, if the integrand is complex analytic. If $Z(s)$ is complex analytic, then

$$\frac{d}{ds}\zeta(s) = Z(s)e^{st} = \sum_{n=0}^{\infty} c_n s^n \tag{1.36}$$

must be complex analytic, since e^{st} is *entire* (it is analytic for all $s \neq \infty$). Note that the poles and zeros of the integrand are determined by $Z(s)$, since e^{st} has no poles or zeros (it and its inverse e^{-st} are entire (complex analytic over the entire open s plane). Thus the FTCC is valid.

Definition of a derivative with respect to a complex variable: When defining a complex analytic function of $Z(s)$ the coefficients are defined by Taylor's formula

$$c_n = \frac{1}{n!}\frac{d^n}{ds^n} Z(s)\big|_{s=s_0}, \tag{1.37}$$

where s_0 is the expansion point. This requires taking derivatives with respect to s at s_0. This formula naturally follows from the definition of the derivative, applied to the infinite series. If the function is an impedance $Z(s)$, a function of $s = \sigma + \omega\jmath \in \mathbb{C}$, the coefficients $c_n \in \mathbb{C}$ are determined by taking derivatives of $Z(s)$ with respect to s at s_0. Since the derivative of s^n is ns^{n-1}, the Taylor series formula follows.

But what does it mean to take the derivative of a function with respect to s, where s defines a plane, not the a real line. We only have learned how to form the derivative on the real line. Can the derivative concept be extended to the complex plane? The problem is resolved by applying the same old rules, but this time in the complex domain.

Specifically, given an analytic function $Z(s)$, is the partial derivative with respect to σ different from the partial derivative with respect to $\omega\jmath$? For complex analytic functions, the FTCC states that the integral is independent of the path in the s plane. Based on the chain rule, the derivative must also be independent of direction at s_0. This directly follows from the FTCC. The *Cauchy-Riemann conditions* follow. From the observation that if the integral of a function of a complex variable is to be independent of the path, the derivative of a function with respect to a complex variable must be independent of the direction. This follows from Taylor's formula, Eq. 1.37 (p. 65) for the coefficients of the complex analytic formula.

It is rarely stated that the variable that we are integrating over, either x (space) or t (time), is real $x, t \in R$, since that obvious fact is implicit, due to the physical nature of the formulation of the integral. But this intuition must change once complex numbers are included with $s \in \mathbb{C}$, where $s = \sigma + \omega\jmath$, as first developed by Cauchy (1789-1857), with the aid of Gauss (1777-1855).

The fact that time and space are real variables is more than an assumption, rather it is a requirement, due to the *order property*. Real numbers have order. For example, if $t = 0$ is now (the present), then $t < 0$ is the past and $t > 0$ is the future. The order property of time and space allows one to order these along a real axis. If the axis were complex, as in frequency s, the order property is lost. It follows that if we desire order, time and space must be real ($t, x \in \mathbb{R}$). Interestingly, it has was shown by d'Alembert (1747) that time and space are related and are almost the same thing, connected, by the wave speed c. To obtain a solution to the governing wave equation, that Newton first proposed, for sound waves, that $x, t \in \mathbb{R}^3$ may be combined as

$$\zeta = ct \pm x,$$

where c [m/s] is the phase velocity of the waves. The d'Alembert solution to the wave equation, describing waves on a string under tension, is

$$u(x,t) = F(ct - x) + G(ct + x),$$

which describes the transverse velocity (or displacement) of two independent waves $F(\xi), G(\xi)$ on the string, which represent forward and backward traveling waves. For example, starting with a string at rest, if one displaces the left end, at $x = 0$, by a step function $u(t)$, then that step displacement will propagate to the right as $u(ct - x)$, arriving at location x_0, at time ct_0. Before this time, the string will not move to the right of the wave-front, at $x = ct$, and after ct_0 it will have displacement 1 at x_0. Since the wave equation obeys superposition (postulate P1, p. 61), it follows that the "plane-wave" eigen-functions of the wave equation for $\vec{x}, \vec{k} \in \mathbb{R}^3$ are given by

$$\Psi_{\pm}(\vec{x}, t) = e^{st \pm \jmath \vec{k}\cdot\vec{x}}. \tag{1.38}$$

where $s = \sigma + \omega\jmath$ and $\vec{k} = \jmath 2\pi/\vec{\lambda}$ is the *wave number*, $\lambda = |\vec{\lambda}|$ is the wavelength, and $|s/\vec{k}| = \lambda f = c$. When propagation losses are considered, we must replace $\jmath\vec{k}$ with a complex analytic wave number $\vec{\kappa}(s) = \vec{k}_r + \jmath\vec{k}$, known as the *dispersion relation.* An important example is the case of electron waves propagating in crystals (i.e., silicon). For this special case, $\vec{\kappa}(s)$ describes the crystal's dispersion relations, know as *Brillouin zones* (Brillouin, 1953).

1.4.2 Lec 24: Cauchy-Riemann conditions

Lec 24: Cauchy-Riemann conditions
Real and imaginary parts of analytic functions obey Laplace's equation.
Infinite power Series and analytic function theory; ROC

For the integral of $Z(s) = R(\sigma,\omega) + X(\sigma,\omega)\jmath$ to be independent of the path, the derivative of $Z(s)$ must be independent of the direction of the derivative. As we show next, this leads to a pair of equations known as the *Cauchy-Riemann* conditions. Taking partial derivatives of $Z(s)$ with respect to σ and $\jmath\omega$, and then equating the real and imaginary parts, gives

$$\frac{\partial R(\sigma,\omega)}{\partial\sigma} = \jmath\frac{\partial X(\sigma,\omega)}{\partial\jmath\omega} \quad \text{and} \quad \frac{\partial R(\sigma,\omega)}{\partial\jmath\omega} = \jmath\frac{\partial X(\sigma,\omega)}{\partial\sigma}. \tag{1.39}$$

These are the necessary conditions that the integral, and thus the derivative, of $Z(s)$ be independent of the path, expressed in terms of conditions on the real and imaginary parts of Z. This is a very strong condition on $Z(s)$, which follows assuming that $Z(s)$ may be approximated by a Taylor series in s

$$Z(s) = Z_0 + Z_1 s + \frac{1}{2}Z_2 s^2 + \cdots, \tag{1.40}$$

where $Z_n \in \mathbb{C}$ are complex constants given by the Taylor series formula (Eq. 1.37, p. 65).

Every function that may be expressed as a *Taylor series* about a point s is said to be *complex analytic* at that point. This series, which is single valued, is said to converge within a *radius of convergence* (ROC). This highly restrictive condition has significant physical consequences. As an important example, every impedance function $Z(s)$ obeys the CR conditions over large regions of the s plane, including the entire right half plane (RHP) ($\sigma > 0$). This condition is summarized by the Brune condition $\Re Z(\sigma > 0) \geq 0$ (Eq. 1.44, Section 1.4.3).

When this conditions is generalize to volume integrals, it is called *Green's theorem*, which is a special case of both Gauss's and Stokes's Laws, used heavily in the solution of boundary value problems in Engineering-Physics (e.g., solving EM problems that start from Maxwell's equations). The last chapter of this course shall depend heavily on these concepts.

We may merge these equations into a pair of second order derivative equations by taking a second round of partials. Specifically, eliminating the real part $R(\sigma,\omega)$ of Eq. 1.39 gives

$$\frac{\partial^2 R(\sigma,\omega)}{\partial\sigma\partial\omega} = \frac{\partial^2 X(\sigma,\omega)}{\partial^2\omega} = -\frac{\partial^2 X(\sigma,\omega)}{\partial^2\sigma}, \tag{1.41}$$

which may be compactly written as $\nabla^2 X(\sigma,\omega) = 0$. Eliminating the imaginary part gives

$$\frac{\partial^2 X(\sigma,\omega)}{\partial\omega\partial\sigma} = \frac{\partial^2 R(\sigma,\omega)}{\partial^2\sigma} = -\frac{\partial^2 R(\sigma,\omega)}{\partial^2\omega}, \tag{1.42}$$

may be written as $\nabla^2 R(\sigma,\omega) = 0$.

In summary, for a function $Z(s)$ to be analytic, the derivative dZ/ds must be independent of direction (path), which requires that the real and imaginary parts of the function obey Laplace's equation, i.e.,

$$\nabla^2 R(\sigma,\omega) = 0 \quad \text{and} \quad \nabla^2 X(\sigma,\omega) = 0. \tag{1.43}$$

As we shall see in the next few lectures, analytic functions must be very smooth due to the above condition. Such functions define *Conservative fields*, which means that energy is conserved. The work done in moving a mass from a to b in such a field is conserved. If you return the mass from b back to a, the energy is retrieved, and zero net work has been done.

1.4.3 Lec 25: Complex Analytic functions and Brune Impedance

Lec 25: Power series and integration of functions
Taylor series formula for coefficients
Fundamental Theorem of integral calculus (Leibniz formula)

Complex impedance functions

One of the most important and obvious applications of complex functions of a complex variable is the impedance function.

Every impedance must obey conservation of energy (P3): According to Postulate P4 Section 1.3.11 (p. 61), a system is passive if it does not contain a power source. Drawing power from an impedance violates conservation of energy. This propriety is also called *positive real*, which is defined as (Brune, 1931a,b)

$$\Re Z(s \ge 0) \ge 0, \tag{1.44}$$

namely that the real part of every impedance must be positive for $\sigma \ge 0$. If this is true, then one cannot draw more power that they have stored in the impedance. A second condition requires that the impedance has simple poles. If there were a pole in the region $\sigma > 0$, then the first condition would not be met. Therefore there can only be simple poles (degree of 1) in the region $\sigma \le 0$. This region is called the *left half plane* (LHP). The complementary region $\sigma > 0$ is called the *Right half plane* (RHP). The condition on the simple poles is sufficient, but not necessary, as $Z(s) = 1/\sqrt{s}$ is a physical impedance, without a first order pole. The impedance function $Z(s) = R(\sigma,\omega) + \jmath X(\sigma,\omega)$ has resistance R and reactance X as a function of complex frequency $s = \sigma + \jmath\omega$. The function $z(t) \leftrightarrow Z(s)$ are defined by a Laplace transform pair. From the causality postulate (P1) of Section 3.5.1, $z(t < 0) = 0$.

As an example, a resistor R_0 in series with an capacitor C_0 has an impedance given by

$$Z(s) = R_0 + 1/sC_0 \tag{1.45}$$

with $R_0, C_0 \in \mathbb{R} > 0$. In mechanics an impedance composed of a dash-pot (damper) and a spring have the same form. A resonant system has an inductor, resistor and a capacitor, with an impedance given by

$$Z(s) = R_0 + 1/sC_0 + sM_0, \tag{1.46}$$

is a second degree polynomial in the complex resonant frequency s. Thus it has two roots (eigenvalues). When $R_0 > 0$ these roots are in the left half s plane.

Systems (networks) containing many elements, and transmission lines, can be much more complicated, yet still have a simple frequency domain representation. This is the key to understanding how these physical systems work, as will be described below.

To solve a differential equation, or integrate a function, Newton used the Taylor series to integrate one term at a time. However he only used real functions of a real variable due to the fundamental lack of appreciation of the complex analytic series. This same method is how one finds solutions to differential equations today, but in a "plug-and-chug" approach, that makes the solution method less transparent. Rather than working directly with the Taylor series, today we use the complex exponential. The reasoning given is that the complex exponential is the eigenfunction of the derivative, namely

$$\frac{d}{dt}e^{st} = se^{st}.$$

Of course e^{st} may be expressed as a Taylor series having coefficients $c_n = 1/n!$, so in some real sense this is doing what Newton did. Thus every linear constant coefficient differential equation in time may be simply transformed into a polynomial in complex Laplace frequency s, by looking for solutions of the form $A(s)e^{st}$, transforming the differential equation into a polynomial $A(s)$ in complex frequency. For example

$$\frac{d}{dt}f(t) + af(t) \leftrightarrow (s+a)F(s).$$

The root of $A(s_r) = s_r + a = 0$ is the eigenvalue of the differential equation. The key to understanding the solutions of differential equations, both scalar and vector, is to work in the Laplace frequency domain. The Taylor series has been replaced by e^{st}, transforming Newton's real Taylor series into the complex exponential eigenfunction. In some sense, these are the same method, since

$$e^{st} = \sum_{n=0}^{\infty} \frac{(st)^n}{n!}. \tag{1.47}$$

Taking the derivative with respect to time gives

$$\frac{d}{dt}e^{st} = se^{st} = s\sum_{n=0}^{\infty} \frac{(st)^n}{n!}, \tag{1.48}$$

which is also complex analytic. Thus if the series for $F(s)$ is valid (i.e., it converges), then its derivative is also valid. This was a very powerful concept, exploited by Newton for real functions of a real variable, and later by Cauchy and Riemann for complex functions of a complex variable. The key here is "Where does the series fail to converge?" in which case, the entire representation fails. This is the main message behind the *Fundamental Theorem of Complex Calculus* (Greenberg, 1988, p. 1197)

$$F(s) = F(s_0) + \int_{s_0}^{s} f(\zeta)d\zeta. \tag{1.49}$$

This result is formally the same as the Fundamental theorem of calculus (Leibniz formula), the key difference being that the argument of the integrand $s \in \mathbb{C}$. Thus this integration is a line integral in the complex plane. One would naturally assume that the value of the integral depend on the path of integration. But behold, accord to this theorem, it does not. In fact is it is indistinguishable from is simple cousin the Fundamental theorem of calculus. And the reasoning is the same. If $f(s) = dF(s)/ds$ is complex analytic (i.e., has a power series $f(s) = \sum_k c_k s^k$, with $f(s), c_k, s \in \mathbb{C}$), then it may be integrated, and *the integral does not depend on the path.* This is sort of amazing. The key is that $F(s)$ and $f(s)$ must be complex analytic, which means they are differentiable. This all follows from the Taylor series formula Eq. 1.37 (p. 65) for the coefficients of the complex analytic series. For Eq. 1.49 to hold, the derivatives must be independent of the direction, as discussed in Section 1.4.2. The concept of a complex analytic function therefore has eminence convergences.

The generalization of the Taylor series formula to the complex plane generalizes the functions they describe, with unpredictable consequences, as nicely shown by the domain coloring diagrams presented in Section 1.3.8 (p. 55). These tools were first exploited in physics by Sommerfeld (1952), to explain the onset transients in waves, as explained in detail in Brillouin (1960, Chap. 3). Up to 1910, when Sommerfeld first published his results using complex analytic signals and saddle point integration in the complex plane, there was a poor understanding of the implications of the causal wave-front. It would be reasonable to say that his insights changed our understanding of wave propagation, for both light and sound. Sadly this insight has not been fully appreciated, even to this day. If you question this summary, please read Brillouin (1960, Chap. 1).

The full power of the analytic function was first appreciated by Bernard Riemann (1826-1866), in his University of Göttingen PhD Thesis of 1851, under the tutelage of Carl Friedrich Gauss (1777-1855), and drawing heavily on the work of Cauchy.

The key definition of a complex analytic function is that it has a Taylor series representation over a region of the complex frequency plane $s = \sigma + j\omega$, that converges in a *region of convergence* (ROC) about the expansion point, with a radius determined by the nearest pole of the function. A further surprising feature of all analytic functions is that within the ROC, the inverse of that function also has a complex analytic expansion. Thus given $w(s)$, one may also determine $s(w)$ to any desired accuracy, critically depending on the ROC.

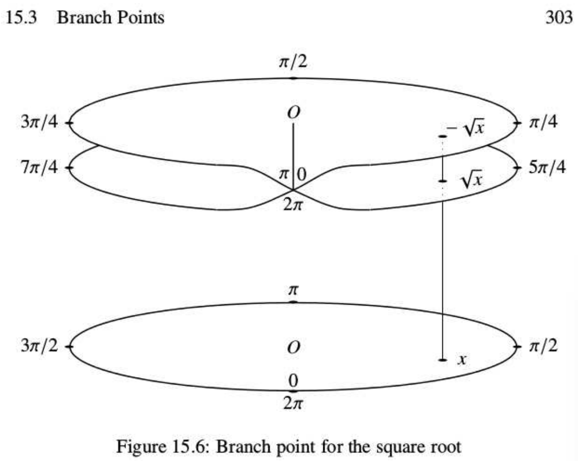


Figure 1.15: *Here we see the mapping for the square root function $w(z) = \pm\sqrt{z}$ which has two single-valued sheets, corresponding to the two signs of the square root. The lower sheet is $+\sqrt{z}$, and the upper sheet is $-\sqrt{z}$. The location of the branch cut may be moved by rotating the z coordinate system. For example, $w(z) = \pm\jmath\sqrt{z}$ and $w(z) = \pm\sqrt{-z}$ have a different branch cuts, as may be easily verified using the Matlab commands `j*zviz(z)` and `zviz(-z)`. A function is analytic on the branch cut, since the cut may be moved. If a Taylor series is formed on the branch cut, it will describe the function on the two different sheets. Thus the Taylor series is agnostic to the location of the cut, as it only describes the function uniquely (as a single valued function), valid in its local region of convergence. Figure from Stillwell (2010, p. 303).*

This concept of analytic inverses becomes rich when the inverse function is multi-valued. For example, if $F(s) = s^2$ then $s(F) = \pm\sqrt{F}$. Riemann dealt with such extensions with the concept of a *branch-cut* with multiple *sheets*, labeled by a sheet number. Each sheet describes an analytic function (Taylor series), that converges within some ROC, with a radius out to the nearest pole of that function. This Riemann's branch cut and sheets explicitly deal with the need to define a unique single valued inverses to multi-valued functions.

Branch Cuts: The concept of a *branch cut* allows one to manipulate (and visualize) multi-valued functions, by breaking each region into a single valued *sheets*. The concepts of the branch cut and the sheets, along with the *extended plane*, were first devised by Riemann, working with Gauss (1777-1855), first described in his thesis of 1851. Of course it was these three mathematical constructions that provide the deep insight to complex analytic functions, supplementing the important earlier work of Cauchy (1789-1857), on the calculus of complex analytic functions.

The branch cut is a line that separates the various single valued parts of a multi-valued function. For example, in Fig. 1.15 we see the double-valued mapping of $w(z) = \pm\sqrt{z}$. Since the square root function has two overlapping regions, corresponding to the $\pm$ of the radical, there must be two connected regions, sort of like mathematical Siamese-twins, distinct, yet the same.

The branch cuts emanate from poles that have non-integral degrees, and terminate at either another pole of the same degree, or at ∞. For example, suppose that in the neighborhood of the pole, at s_0 the function is

$$f(s) = \frac{K}{(s-s_0)^k},$$

where $f, s, K, k \in \mathbb{C}$. Here K is the residue, $s = \sigma + \omega \jmath$, $s_0 = \sigma_0 + \omega_0 \jmath$ and k is some complex or real constant that defines the degree of the pole. Up to this point we have taken $k \in \mathbb{Z}$ or fractionals ($k \in \mathbb{F}$). When $k \in \mathbb{Z}$ there is no branch cut. When $k \in \mathbb{F}$ there must be a branch cut, of degree k. For example if $k = 1/2$, the pole is of degree $1/2$, and there are two Riemann sheets, as shown in Fig. 1.15. An important example is the Bessel function

$$J_0(t)u(t) \leftrightarrow \frac{1}{\sqrt{1+s^2}},$$

which is a solution to the wave equation in cylindrical coordinates. This plays a key role in the solutions to waves in round pipes, or surface waves on the earth (seismic waves) or on the surface of a pond.

When $k \in \mathbb{I}$ is irrational the pole, and thus the branch cut, has irrational order. Accordingly there must be an infinite number of Riemann sheets, as in the case of the log function, but, I suspect, much worse. An example might be $k = \pi$, for example. When the irrational number is very close to 1, the branch cut could be very subtle, and could even go unnoticed, but it would have a huge impact on the nature of the function, and of course, on the inverse Laplace transform. I am not aware of any physical systems that behave this way, but I would take a wild guess that they exist (I have no evidence for this possibly bizarre speculation).

It is important to understand that the function is analytic on the branch cut, but no at the branch point (the pole). In fact, one is free to move the branch cut, almost at will. It does not need to be on a line, it could be cut as a spiral. The only rule is that it must start and stop at two poles of the same degree, or at ∞, which must have a pole of degree k. It is because the function is analytic on the branch cut, that it may be moved.

1.4.4 Lec 26: Cauchy Integral Theorem

There are three theorems that collectively are in the category of a fundamental theorem of complex calculus, all attributed to Cauchy, and named after him and his important discoveries. The names of the three theorems are similar, thus potentially confusing. They are the

1. Cauchy integral theorem
$$\oint_{\mathcal{C}} f(s)ds = 0 \tag{1.50}$$

2. Cauchy integral formula
$$\oint_{\mathcal{C}} \frac{f(s)}{s-s_0} ds = \begin{cases} 2\pi\jmath & \text{if } s_0 \in \mathcal{C} \\ 0 & \text{if } s_0 \not\in \mathcal{C} \end{cases}. \tag{1.51}$$

3. Cauchy residue theorem.
$$\oint_{\mathcal{C}} F(s)ds = 2\pi\jmath \sum_{k}^{K} \text{Res}_k \tag{1.52}$$

The integral sign with the circle $\oint_{\mathcal{C}}$ means that the integral's path $\mathcal{C}$ is closed. The lower case $f(s)$ denotes a complex analytic function inside and on $\mathcal{C}$. The upper case $F(s)$ denotes a function with K poles in $\mathcal{C}$. In general it makes no sense to integrate *through* a pole, thus the poles (or other singularities) must not be on $\mathcal{C}$.

The Cauchy integral theorem, follows trivially from the Fundamental theorem of complex calculus, since if the integral is independent of the path, and the path returns to the starting point, the closed integral must be zero. Thus

$$\oint_C f(s)ds = 0,$$

if, and only if, $f(s)$ is complex analytic within C.

This theorem follows trivially from the FTCC, since if the integral is independent of the path, and the path returns to the starting point, the closed integral must be zero. Thus

$$\oint_C f(s)ds = 0,$$

if, and only if, $f(s)$ is complex analytic within C.

Since the real and imaginary parts of every complex analytic function obeys Laplace's equation (Eq. 1.43, p. 67), it follows that every closed integral over a Laplace field, i.e., one defined by Laplace's equation, must be zero. This corresponds to many physical situations. If a closed box has fixed potentials on the walls, with any distribution what so ever, and a point charge (i.e, an electron) is placed in the box, then a force equal to $F = qE$ is required to move that charge, thus work is done. However if the point is returned to its starting location, the net work done is zero.

Work is done in charging a capacitor, and energy is stored. However when the capacitor is discharged, all of the energy is returned to the load.

Soap bubbles and rubber sheets on a wire frame, obey Laplace's equation.

These are all cases where the fields are Laplacian, thus closed line integrals must be zero. Laplacian fields are perhaps the most common observed, because they are so basic.

We have presented the impedance as the primary example of complex analytic functions. Physically the impedance defines the power. The impedance is defined as the force over the flow (i.e., voltage over current). The power is defined as the voltage times the current (see Section 3.2.1, p. 95, Eq. 3.5).

1.4.5 Lec 27: Cauchy Integral Formula & Residue Theorem

The Cauchy integral formula (Eq. 1.51) is an important extension of the Cauchy integral theorem (Eq. 1.50) in that a pole has been injected into the integrand at $s = s_0$. If the pole location is outside of the curve $\mathcal{C}$, the result of the integral is zero, in keeping with Eq. 1.50. However, when the pole is inside of $\mathcal{C}$, the integrand is no longer analytic, and a new result follows. By a manipulation of the contour, the pole can be isolated with a small circle around the pole. It may be shown that for a very small circle around the pole,

$$\oint_{|s|=1} \frac{f(s_0)}{s_0} ds = 2\pi\jmath.$$

For the related *Cauchy residue theorem* (Eq. 1.52) the same result holds, except it is assumed that there are K poles in the function $F(s)$. This requires the repeated application of Eq. 1.51, K times, so it represents a minor extension of Eq. 1.51.

Non-integral degree poles The key point is that this theorem applies when $n \in \mathbb{I}$, including fractionals $n \in \mathbb{F}$. The function $1/\sqrt{s}$ has no residue, which is strictly defined as the amplitude of $1/s$ following the definition of the residue c_{-1} of $f(s)$

$$c_{-1} = \lim(s - s_k)f(s).$$

When $n \in \mathbb{F}$, the residue is, by definition, zero. When $n \in \mathbb{I}$, the residue is, by definition, zero. When $n = 1$, the residue is c_{-1}.

This point is equally important when defining the inverse Laplace transform. When integrating over $\omega \in \mathbb{R}$, the value passes through all possible exponents, including all rational and irrational numbers. The only value of ω that has a residue, are those at the poles of the integrand. For example, the inverse Laplace transform of $F(s) = 1/(s+1)$ has a residue of 1 at $s = -1$, thus that is the only contribution to the integral. A case that is more demanding is the Laplace transform pair

$$\frac{1}{\sqrt{t}}u(t) \leftrightarrow \sqrt{\frac{\pi}{s}}.$$

Another interesting case is proving that $u(t)$ is not defined at $t = 0$.

1.4.6 Lec 28: Inverse Laplace transform (Cauchy residue theorem)

Use of the Residue theorem to evaluate the inverse Laplace transform. Discuss causal and anti-causal cases. How does this relate to Green's theorem (in 2 dimensions).

1.4.7 Lec 29: Inverse Laplace transform and the Cauchy Residue Thm

1.4.8 Lec 30: Case for causality: closing the contour

1.4.9 Lec 31: Properties of the LT (e.g., modulation, translation, etc.)

WEEK 12 33.12.0

L 33 Multi-valued functions in the complex plane; Branch cuts
The extended complex plane (regularization at ∞) (Stillwell, 2010, p. 280)
Complex analytic functions of Genus 1 (Stillwell, 2010, p. 343)

L 34 Euler's vs. Riemann's Zeta function $\zeta(s)$: Poles at the primes
colorized plot of $\zeta(s)$
??Sterling's formula??

L 35 Exam III

1.4.10 Lec 33: Multi-valued functions Branch cuts

1.4.11 Lec 34: The Riemann zeta function

The LT of the complex Riemann zeta function $\zeta(x)$ (Fig. 4.1), as introduced by Euler for real arguments. $x \in \mathbb{R}$ as his way of proving that the number of primes is infinite (Goldstein, 1973). In the end, this formulation provided detailed information about the structure of the primes. The zeta function depends explicitly on the primes, which is why it is interesting (Section 4.5.2).

One might wonder why Euler's zeta function is known as the Riemann zeta function. It is because Riemann showed its properties when the argument is complex, namely he extended $\zeta(s)$ into the complex plane ($s \in \mathbb{C}$) (Section 4.5.2). Given that $\zeta(s)$ is a function of complex (Laplace) frequency, one might naturally ask if $\zeta(s)$ has an inverse Laplace transform. There seems to be very little written on this topic,[51] but we shall explore this interesting question further (Table 4.1). Perhaps even more important is the taxonomy of $\zeta(s)$ is in question here, namely where are its poles and zeros? About this there are volumes written.

[51] Cite book chapter on inverse LT of $\zeta(s)$.

The Riemann Zeta function is analytic with poles at log-primes

Why does the zeta function have zeros? Perhaps this is some extension of the Euler function that has zeros, rather than zeta itself. Ask Andrew Odlyzko about this problem. Go to the Math dept first and find someone qualified to discuss this with.

1.4.12 Lec 35: Exam III (Evening Exam)

WEEK 13 36.13.0

L 36 Scaler wave equations and the Webster Horn equation; WKB method
A real-world example of large delay, where the branch-cut placement is critical

L 37 Partial differential equations of Physics
Scaler wave equation and its solution in 1 and 3 Dimensions
VC-1

L 38 Vector dot and cross products $A \cdot B, A \times B$
Gradient, divergence and curl

– Thanksgiving Holiday 11/19–11/27 2016

1.5 Vector Calculus (Stream 3b)

1.5.1 Lec 36: Scalar Wave Equation (Acoustics)

Acoustic waves; The scalar wave equation: scalar differential equation in the frequency domain

The Webster Horn equation

The effect of a spatial area functions for waves in horns (the horn equation).
Derivation of the Horn equation, starting from the basic equations of acoustics.
Development of the basic equations of acoustics: Conservation of mass and momentum.
Sound in a uniform tube.
Sound propagation away from a point source (Helmholtz's Equation)

$$\nabla^2\psi + k^2\psi = \delta(r).$$

1.5.2 Lec 37: Partial Diff Eqs of Physics

1.5.3 Lec 38: Vector dot and cross products

1.5.4 Thanksgiving Holiday 11/19–11/27 2016

WEEK 14 37.14.0

L 39 Gradient, divergence and curl: Gauss's (divergence) and Stokes's (curl) theorems

L 40 J.C. Maxwell unifies Electricity and Magnetism with the formula for the speed of light
Basic definitions of E, H, B, D
O. Heaviside's (1884) vector form of Maxwell's EM equations and the *vector wave equation*
How a loud-speaker works

L 41 *The Fundamental Thm of vector calculus*
Incompressable and *Irrotational* fluids and the two defining vector identities

1.5.5 Lec 39 $\nabla, \nabla\cdot, \nabla\times$ & Vector operators

There are three key vector differential operators that we need to understand Maxwell's equations. The *gradient* transforms a potential, such as a voltage $V(x, y, z)$ into a vector, such as the electric field vector $\mathbf{E}$. The *divergence* $\nabla \cdot \mathbf{E}(x, y, z)$ transforms a vector field into a scalar field. Finally the curl $\nabla \times \mathbf{A}(x, y, z)$ transforms a vector into a vector.

To define these three operations we first need to define scalar and vector fields. These are concepts that you already understand. It is the terminology that needs to be mastered, not a new concept. Think of a voltage field in space, say between two finite sized capacitor plates. In such a case, the voltage is given by a *scalar field* $V(x, y, z)$. A scalar field is also called a *potential*. Somewhat confusing is that one may also define *vector potentials* which is three scalar potentials turned into a vector. So this term is more than one use. It is therefore important to recognize the intended use of the filed. This can be gleaned from the units. Volts is a scalar field.

The simplest example of a scalar potential is the voltage between two very large (think ∞) conducting parallel planes, or plates (large so that we can ignore the edge effects). In this case the voltage varies linearly between the two plates. For example

$$V(x, y, z) = V_0(1 - x)$$

is a scalar potential, thus it is scalar field (i.e., potential). At $x = 0$ the voltage is V_0 and at $x = 1$ the voltage is zero. Between 0 and 1 the voltage varies linearly. Thus $V(x, y, z)$ defines a *scalar field*.

If the same setup were used but the two plates were 1x1 [cm^2], with a 1 [mm] air gap, there will be a small "fringe" effect at the edges that would slightly modify the ideal fields. The hope is that this effect can be made small so that it does not ruin the capacitor composed of the two plates. If we are given a set of three scalar fields, we define a *vector field*. If the three elements of the vector are potentials, then we have a vector potential.

Gradient operator ∇

The gradient operator takes a scalar fields and outputs a vector field. This is exactly what the gradient does. Given any scalar field $V(x, y, z)$ it outputs a vector field[52]

$$\mathbf{E}(x, y, z) = [E_x(x, y, z), E_y(x, y, z), E_z(x, y, z)]^T = -\nabla V(x, y, z).$$

To understand these three operations we therefore need to define the domain and range of their operation, as specified in Table 5.1.

[52]As before vectors are columns, which take up space on the page, thus we write them as rows and take the transpose to properly format them.

1.5.6 Lec 40: Definitions of E, H, B, D and Maxwell's equations

Maxwell's Equations

Once you have mastered the three basic vector operations, the gradient, divergence and curl, you are able to understand Maxwell's equations. Like the vector operations, these equations may be written in integral or vector form. The notation is basically the same since the concept is the same. The only difference is that with Maxwell's equations we are dealing with well defined physical quantities. The scalar and vector fields take on meaning, and units. Thus to understand these important equations, one must master the units, and equally important, the names of the four fields that are manipulated by these equations.

We may now restate everything defined above in terms of two types of vector fields that decompose every vector field. Thus another name for the Fundamental Theorem of Vector Calculus is the *Helmholtz decomposition*. An *irrotational field* is define as one that is "curl free," namely the vector potential is a constant. An *incompressible field* is one that is "diverge free," namely the scalar potential is a constant. Just to confuse matters, the incompressible field is also called a *solenoidal field*. I recommend that you know this term (as it is widely used), but never use it. Rather use incompressible as a more meaningful and physical term.

1.5.7 Lec 41 Fundamental Theorem of Vector calculus (Helmholtz theorem)

The Fundamental Theorem of Vector Calculus

Here we define the basic vector operations based on the ∇ "operator," the *gradient*, *divergence* and the *curl*. These operations may be defined in terms of integral operations on a surface in 1, 2 or 3 dimensions, and then taking the limit as that surface goes to zero. These operators are required to understand Maxwell's Equations, the crown jewel of mathematical physics.

Incompressible and Irrotational vector fields

One of the most important fundamental theorems is that of vector calculus. This is also known as *Helmholtz theorem*. This theorem is very easily stated but less easily to appreciate. But a physical description will help.

A vector field may be split into two independent parts. Think linear and angular momentum. They are independent in that they represent different ways to store energy, leading to independent degrees of freedom of stored energy. An object with mass can be moving along a path and rotating at the same time. The two modes of motion define two different types of kinetic energy, transnational and rotational. In some real sense, Helmholtz theorem quantifies this independence.

The Fundamental Theorem of Vector Calculus: This theorem is also known as Helmholtz' theorem. It states that every differentiable (i.e., analytic) vector field may be written as the sum of two terms, a scalar part and a vector part, expressed in terms of a scalar potential $\phi(x, y, z)$ (think voltage) and a vector potential (think magnetic vector potential). Specifically

$$\mathbf{E} = -\nabla\phi + \nabla \times \mathbf{A}, \tag{1.53}$$

where ϕ is the scalar and $\mathbf{A}$ is the vector potential. To show how this relationship splits the vector field $\mathbf{E}$ into two parts, we include two key vector identities, that are always equal to zero for complex analytic fields: the *curl of the divergence*

$$\nabla \times \nabla\phi(x, y, z) = 0, \tag{1.54}$$

and the *divergence of the gradient*

$$\nabla \cdot \nabla \times \mathbf{A} = 0. \tag{1.55}$$

These identities are easily verified by working out a few specific examples, based on the definitions of the three operators, gradient, divergence and curl, or in terms of the operator's integral definitions, defined below. The identities have a physical meaning, as stated above: every vector field may be split into its transnational and rotational parts. If $\mathbf{E}$ is the electric field [V/m], ϕ is the voltage and $\mathbf{A}$ is the induced rotational part, induced by a current. We shall explore this in our discussion of Maxwell's equations, in Chapter 5.

By applying these two identities to Helmholtz's theorem (Eq. 1.53), we can appreciate the significance of the theorem. It is a form of proof actually, once you have satisfied yourself that the vector identities are true. In fact one can work backward using a physical argument, that rotational momentum (rotational energy) is independent of the transnational momentum (transnational energy). Again this all goes back to the definitions of rotation and transnational forces, hidden in the vector operations. Once these forces are made clear, the meaning of the vector operations all take on a very well defined meaning, and the mathematical constructions, centered around Helmholtz's theorem, begins to provide some common-sense meaning. One might conclude that the physics is related to the geometry.

Specifically, if we take the divergence of Eq. 1.53, and use the divergence vector identity

$$\nabla \cdot \mathbf{E} = \nabla \cdot \{-\nabla\phi + \nabla \times \mathbf{A}\} = -\nabla \cdot \nabla\phi = -\nabla^2\phi.$$

since the divergence vector identity removes the vector potential $\mathbf{A}(x, y, z)$.

Likewise if we take the curl of Eq. 1.53, and use the curl vector identity

$$\nabla \times \mathbf{E} = \nabla \times \{-\nabla\phi + \nabla \times \mathbf{A}\} = \nabla \times \nabla \times \mathbf{A},$$

since using the curl vector identity, removes the scalar field $\phi(x, y, z)$.

There is a third vector identity that needs to be mentioned for later use

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2\mathbf{A}.$$

The best way to think of this relationship is that it *defines* the vector Laplacian $\nabla^2\mathbf{A}$. In other words, think of this identity the definition of the left hand side of

$$\nabla^2\mathbf{A} \equiv \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}).$$

Figure 1.16: *von Helmholtz portrait taken from the English translation of his 1858 paper "On integrals of the hydrodynamic equations that correspond to Vortex motions" (in German) (von Helmholtz, 1978).*

WEEK 15

40.15.0

- L 42 Quasi-static approximation and applications:
  The Kirchoff's Laws and the *Telegraph wave equation*, starting from Maxwell's equations The telegraph wave equation starting from Maxwell's equations
  Quantum Mechanics

- L 43 Last day of class: Review of Fund Thms of Mathematics:
  Closure on Numbers, Algebra, Differential Equations and Vector Calculus,
  The Fundamental Thms of Mathematics & their applications:
  Theorems of Mathematics; Fundamental Thms of Mathematics (Ch. 9); Normal modes vs. eigenstates, delay and quasi-statics;

- – Reading Day
  VC-1 Due

1.5.8 Lec 42: Kirchhoff's Laws and the quasi-static approximation

The term *quasi-statics* (Postulate P9, p. 62) is an approximation, used to reduce a a partial differential equations to a scalar (one-dimensional) equation (Sommerfeld, 1952). Quasi-statics is a way of reducing a three dimensional problem to a one–dimensional problem. So that it is not miss-applied, it is important to understand the nature of this approximation, which is goes to the heart of transmission line theory. The quasi-static approximation states that the wavelength λ is greater than the dimensions of the size of the object Δ (e.g., $\lambda \gg \Delta$). The best known example, Kirchhoff's current and voltage laws, KCL and KVL, almost follow from Maxwell's equations given the quasi-static approximation (Ramo et al., 1965). These laws state that the sum of the currents at a node must be zero (KCL) and the some of the voltages around a loop must be zero (KCL).

These well know laws the analogue of Newton's laws of mechanics. The sum of the forces at a point is the analogue of the sum of the voltages. Voltage ϕ is the force potential, since the electric field $\mathbf{E} = -\nabla\phi$. The sum of the currents is the analogue of the vector sum of velocities at a point is zero.

The acoustic wave equation describes how the scalar field pressure $p(\vec{x}, t)$, the vector force density potential ($f(\vec{x}, t) = -\nabla p(\vec{x}, t)$ [N/m^2], propagates in three dimensions. (The net force is the integral of the pressure gradient over an area.) If the wave propagation is restricted to a pipe (e.g., organ pipe), or to a string (e.g., an guitar or lute), the transverse directions may be ignored, due to the quasi-static approximation. What needs to be modeled by the equations is the wave propagation along the pipe (string). Thus we may approximate the restricted three-dimensional wave by a one-dimensional wave.

However if we wish to be more precise about this reduction in geometry ($\mathbb{R}^2 \to \mathbb{R}$), we need to consider the quasi-static approximation, as it makes assumptions about what is happening in the other directions, and quantifies the effect ($\lambda \gg \Delta$). Taking the case of wave propagation in a tube, say the ear canal, there is the main wave direction, down the tube. But there is also wave propagation in the transverse direction, perpendicular to the direction of propagation. As shown in Table 3.1 (p. 113), the key statement of the quasi-static approximation is that the wavelength in the transverse direction is much larger that the radius of the pipe. This is equivalent to saying that the radial wave reaches the walls and is reflected back, in a time that is small compared to the distance propagated down the pipe. Clearly the speed of sound down the pipe and in the transverse direction is the same if the medium is homogeneous (i.e., air or water). Thus the sound reaches the walls and is returned to the center line in a time that the axial wave traveled about 1 diameter along the pipe. So if the distance traveled is several diameters, the radial parts of the wave have time to come to equilibrium. So the question one must ask is, what are the conditions of such an equilibrium. The most satisfying answer to this is to look at the internal forces on the air, due to the gradients in the pressure.

The pressure $p(x, y, z, t)$ is a potential, thus its gradient is a force density $\mathbf{f}(x, y, z, t) = -\nabla p(x, y, z, t)$. What this equation tells us is that as the pressure wave approaches that of a plane wave, the radial (transverse) forces go to zero. If the tube has a curvature, or a change in area, then there will be local forces that create radial flow. But after traveling a few diameters, these forces will come to equilibrium and the wave will return to a plane wave. The internal stress caused by a change is area must settle out very quickly. There is a very important caveat however: it is only at low frequencies that the plane wave can dominate. At frequencies such that the wavelength is very small compared to the diameter, the distance traveled between reflections is much greater than a few diameters. Fortunately the frequencies where this happens are so high that they play no role in frequencies that we care about. This effect is referred to as *cross-modes* which imply some sort of radial standing waves. In fact such modes exist in the ear canal, but on the eardrum where the speed of sound is much slower that that of air. Because of the slower speed, the ear drum has cross-modes, and these may be seen in the ear canal pressure. Yet they seem to have a negligible effect on our ability to hear sound with good fidelity. The point here is that the cross modes are present, but we call upon the quasi-static approximation as a justification for ignoring them, to get closer to the first-order physics.

The quasi-static approximation failure at high frequencies: At high frequencies the quasi-static approximation must break down. Thus at higher frequencies we need to consider other significant physics of the system, known as *higher order modes*. A further complication is that at higher frequencies, damping becomes an issue.

In acoustics viscosity and thermal effects are typically ignored, by assuming that wave propagation is dictated by the wave equation. In fact, it turns out that these two loss mechanisms are related. But to understand why is quite complex. Helmholtz, with help from Kirchhoff, explained this interaction, and independently published them between 1863 (Helmholtz, 1863b) and 1868 (Kirchhoff, 1868). Their collective theory is nicely summarized by Lord Rayleigh (Rayleigh, 1896), and then experimentally verified to be correct by Warren P. Mason (Mason, 1928). The nature of the correction is that the wave number is extended to be of form

$$\kappa(s) = \frac{s + \beta_0\sqrt{s}}{c_0}, \tag{1.56}$$

where the forwarded P_- and backward P_+ pressure waves propagate as

$$P_\pm(s, x) = e^{-\kappa(s)x}, e^{-\overline{\kappa}(s)x} \tag{1.57}$$

and $\overline{\kappa}(s)$ is the complex conjugate of $\kappa(s)$, and $\Re\kappa(s) > 0$. The term $\beta_0\sqrt{s}$ affects both the real and imaginary parts of $\kappa(s)$. The real part is a frequency dependent loss and the imaginary part introduces a frequency dependent velocity (Mason, 1928).

The frequency where the loss-less part equals the lossy part is an important parameter of the system. This frequency is $s_0 + \beta_0\sqrt{s_0} = 0$, or $\sqrt{s_0} = \beta_0$ or $f_0 = \beta^2/2\pi$.

Assuming air at 23.5° [C], $c_0 = \sqrt{\eta_0 P_0/\rho_0} \approx 344$ [m/s] is the speed of sound, $\eta_0 = c_p/c_v = 1.4$ is the ratio of specific heats, $\mu = 18.5 \times 10^{-6}$ [Pa-s] is the viscosity, $\rho_0 \approx 1.2$ [kgm/m^2] is the density, $P_0 = 10^5$ [Pa] (1 atm).

The constant $\beta_0 = P\eta'/2S\sqrt{\rho_0}$

$$\eta' = \sqrt{\mu}\left[1 + \sqrt{5/2}\left(\eta^{1/2} - \eta^{-1/2}\right)\right]$$

is a thermodynamic constant, P is the perimeter of the tube and S the area (Mason, 1928).

For a cylindrical tube having radius $R = 2S/P$, $\beta_0 = \eta_0'/R\sqrt{\rho}$. To get a feeling for the magnitude of β_0 consider a 7.5 [mm] tube (i.e., the average diameter of the adult ear canal). Then $\eta' = 6.6180 \times 10^{-3}$ and $\beta_0 = 1.6110$. Using these conditions the wave-number cutoff frequency is $1.611^2/2\pi = 0.4131$ [Hz]. At 1 kHz the ratio of the loss over the propagation is $\beta_0/\sqrt{|s|} = 1.6011/\sqrt{2\pi 10^3} \approx 2\%$. At 100 [Hz] this is a 6.4% effect.[53]

Mason shows that the wave speed drops from 344 [m/s] at 2.6 [kHz] to 339 [m/s] at 0.4 [kHz], which is a 1.5% reduction in the wave speed. In terms of the losses, this is much larger effect. At 1 [kHz] the loss is 1 [dB/m] for a 7.5 [mm] tube. Note that the loss and the speed of sound vary inversely with the radius. As the radius approaches the boundary layer thickness (the radial distance such that the loss is e^{-1}), the effect of loss dominates.

In Section 5.4.1 we shall look at some simple problems where we use the quasi-static effect and derive the Kirchhoff voltage and current equations, starting from Maxwell's equations.

1.5.9 Lec 43: Final Review for Final Exam

Summary

Mathematics began as a simple way of keeping track of how many things there were. But eventually Physics and Mathematics evolved together as tools to help us navigate our environment, not just physically around the globe, but how to solve daily problems such as food, water and waste management, understand the solar system and the stars, defend ourselves, use tools of war, etc.

[53] `/home/jba/Mimosa/2C-FindLengths.16/doc.2-c_calib.14/m/MasonKappa.m`

Based on the historical record of the abacus, one can infer that people precisely understood the concept of counting, addition, subtraction and multiplication (recursive addition).

There is some evidence that the abacus, a simple counting tool, formalizing the addition of very large numbers, was introduced to the Chinese by the Romans, where it was used for trade.

However this working knowledge of arithmetic did not to show up in written number systems. The Roman numerals were not useful for doing calculations done on the abacus. The final answer would then be expressed in terms of the Roman number system.

According to the known written record, the number zero (null) had no written symbol until the time of Brahmagupta (628 CE). One should not assume the concept of zero was not understood simply because there was no symbol for it in the Roman Numeral system. Negative numbers and zero would be obvious when using the abacus. Numbers between the integers would be represented as *rational numbers* $\mathbb{Q}$ since any number may be approximated with arbitrary accuracy with rations numbers.

Mathematics is the science of formalizing a repetitive method into a set of rules, and then generalizing it as much as possible. Generalizing the multiplication and division algorithm, to different types of numbers, becomes increasingly more complex as we move from integers to rational numbers, irrational numbers, real and complex numbers and ultimately, vectors and matrices. How do you multiply two vectors, or multiply and divide one matrix by another? Is it subtraction as in the case of two numbers? Multiplying and dividing polynomials (by long division) generalizes these operations even further. Linear algebra is further important generalization, fallout from the Fundamental Theorem of Algebra, and essential for solving the generalizations of the number systems.

Many of the concepts about numbers naturally evolved from music, where the length of a string (along with its tension) determined the pitch (Stillwell, 2010, pp. 11, 16, 153, 261). Cutting the string's length by half increased the frequency by a factor of 2. One forth of the length increases the frequency by a factor of 4. One octave is a factor of 2 and two octaves a factor of 4 while a half octave is $\sqrt{2}$. The musical scale was soon factored into rational parts. This scale almost worked, but did not generalize (sometimes known as *Pythagoreas' comma*[54]), resulting in today's *well tempered scale*, which is based on 12 equal geometric steps along one octave, or 1/12 octave ($\sqrt[12]{2} \approx 1.05946 \approx 18/17 = 1 + 1/17$).

But the concept of a *factor* was clear. Every number may be written as either a sum, or a product (i.e., a repetitive sum). This led the early mathematicians to the concept of a prime number, which is based on a unique factoring of every integer. At this same time (c5000 BCE), the solution of a second degree polynomial was understood, which lead to a generalization of factoring, since the polynomial, a sum of terms, may be written in factored form. If you think about this a bit, it is sort of an amazing idea, that needed to be discovered (Stillwell, 2010, p.). This concept lead to an important string of theorems on factoring polynomials, and how to numerically describe physical quantities. Newton was one of the first to master these tools with his proof that the orbits of the planets are ellipses, not circles. This lead him to expanding functions in terms of their derivatives and power series. Could these sums be factored? The solution to this problem led to calculus.

So mathematics, a product of the human mind, is a highly successful attempt to explain the physical world. All aspects of our lives were impacted by these tools. Mathematical knowledge is power. It allows one to think about complex problems in increasingly sophisticated ways. An equation is a mathematical sentence, expressing deep knowledge. Witnessed $E = mc^2$ and $\nabla^2\psi = \ddot{\psi}$. Move to end

Reading List: The above concepts come straight from mathematical physics, as developed in the 17^{th}–19^{th} centuries. Much of this was first developed in acoustics by Helmholtz, Stokes and Rayleigh, following in Green's footsteps, as described by Lord Rayleigh (1896). When it comes to fully appreciating Green's theorem and reciprocity, I have found Rayleigh (1896) to be a key reference. If you wish to repeat my reading experience, start with Brillouin (1953), followed by Sommerfeld (1952); Pipes (1958). Second tier reading contains many items Morse (1948); Sommerfeld (1949); Morse and Feshbach (1953); Ramo et al. (1965); Feynman (1970); Boas (1987). A third tier might include Helmholtz (1863a); Fry (1928); Lamb (1932); Bode (1945); Montgomery et al. (1948); Beranek (1954); Fagen (1975); Lighthill

[54] https://en.wikipedia.org/wiki/Pythagorean_comma

(1978); Hunt (1952); Olson (1947). It would be a mistake to ignore other massive physics writings by stalwart authors, J.C. Slater[55] and Landau and Lifshitz,[56] and their impressive series of Mathematical Physics books.

You must enter at a level that allows you to understand. Successful reading of these books critically depends on what you already know. A rudimentary (high school) level of math comprehension must be mastered first. Read in the order that helps you best understand the material.

Without a proper math vocabulary, mastery is hopeless. I suspect that one semester of college math can bring you up to speed. This book is my attempt to present this level of understanding.

[55]https://en.wikipedia.org/wiki/John_C._Slater

[56]https://www.amazon.com/Mechanics-Third-Course-Theoretical-Physics/dp/0750628960

Chapter 2

Number Systems: Stream 1

This chapter is devoted to *Number Systems* (Stream 1), starting with the counting numbers $\mathbb{N}$. In this chapter we delve more deeply into the details of the topics of Lectures 4-9.

2.1 Week 2

In Section 2.1.1 we explore in more detail the two fundamental theorems of prime numbers, working out a sieve example, and explore the logarithmic integral $Li(N)$ which approximates the density of primes $\rho_k(N)$ up to prime N.

The topics of Section 1.2.4 consider the practical details of computing the *greatest common divisor* (GCD) of two integers m, n (Matlab's routine `l=gcd(m,n)`), with detailed examples and comparing the algebraic and matrix methods. Homework assignments will deal with these two methods. Finally we discuss the relationship between coprimes and the GCD. In Section 1.2.5 we defined the Continued Fraction algorithm (CFA), a method for finding rational approximations to irrational numbers. The CFA and GCD are closely related, but the relation needs to be properly explained. In Section 1.2.7 we derive *Euclid's formula*, the solution for the Pythagorean triplets (PT), based on Diophantus's *chord/-tangent* method. This method is used many times throughout the course notes, first for computing Euclid's formula for the PTs, then for finding a related formula in Section 1.2.8 for the solutions to Pell's equation, and finally for finding the mapping from the complex plane to the extended complex plane (the Riemann sphere).

Finally in Section 1.2.9 the general properties of the *Fibonacci sequence* is discussed. This equation is a special case of the second order digital resonator (well known in digital signal processing), so it has both historical and practical application for engineering. The general solution of the Fibonacci is found by taking the Z-transform and finding the roots, resulting in an eigenvalue expansion (Appendix D).

2.1.1 Lec 4 Two theorems on primes

Theorem 1: *Fundamental Theorem of Arithmetic*

Factoring integers: Every integer $n \in \mathbb{N}$ has a unique factorization (Stillwell, 2010, p. 43)

$$n = \prod_{k=1}^{K} \pi_k^{\beta_k}, \tag{2.1}$$

where $k = 1, \ldots, K$ indexes the integer's K *prime factors* π_k and their *multiplicity* β_k.

Examples: $2312 = 2^3 \cdot 17^2 = \pi_1^3\, \pi_7^2$ (i.e., $\pi_1 = 2, \beta_1 = 3; \pi_7 = 17, \beta_7 = 2$)
$2313 = 3^2 \cdot 257 = \pi_3^2\, \pi_{55}$ (i.e., $\pi_2 = 3, \beta_3 = 2; \pi_{55} = 257, \beta_{55} = 1$)

Integers 2312 and 2313 are said to be *coprime*, since they have no common factors. *Coprimes* may be identified via the *greatest common divisor*:

$$gcd(a, b) = 1$$

using the *Euclidean algorithm* (Stillwell, 2010, p. 41).

Theorem 2: *Prime Number Theorem*

The primes are a random field since there is no way to predict when the next prime will occur. Thus one needs to use statistical methods to characterize their density. Based on a sample of approximately 3 million primes, Gauss showed empirically that the average total number of primes less than N is

$$\sum_{n=1}^{N} \delta_n \sim \frac{N}{\ln N}.$$

These primes were obtained by manual calculations "as a pastime" in 1792-3 (Goldstein, 1973).

Define $\delta_n = 1$ if n is a prime, and zero otherwise.[1].

It follows that the *average density of primes* is $\rho_\pi(N) \sim 1/\ln N$, thus

$$\rho_\pi(N) \equiv \frac{1}{N}\sum_{n=1}^{N} \delta_n \approx \frac{1}{N} Li(N) \equiv \frac{1}{N}\int_2^N \frac{d\xi}{\ln \xi},$$

where $Li(N)$ is the *offset logarithmic integral* (Stillwell, 2010, p. 585). The primes are distributed as $1/\ln(n)$ since the average total number of primes is proportional to the *logarithmic integral* $Li(n)$ (Goldstein, 1973; Fine, 2007).

Here is a Matlab code that tests this formula:

```
%Computes density of primes from average density

NP=1e6; % 10^6 primes
p=primes(NP); %compute primes
delta=zeros(1,NP); delta(p)=1; %put 1 at each prime
PI=cumsum(delta); %Number of primes vs N
rho=PI./(1:NP); %estimate of the density of primes= PI(N)./N
%
figure(1)
semilogy(rho); %plot of density vs number of primes
title('Density of primes vs. N'); ylabel('\rho(N)'); xlabel('N'); grid on
```

From the Prime Number Theorem it is clear that the density of primes is large (they are not scarce). As best I know there are no methods to find primes other than by the sieve method (Section ??, p. ??). If there is any good news it is that they only need to be computed once, and saved. In practical applications this doesn't help much given their large number. In theory, given primes π_n up to $n = N$, the density $\rho_\pi(N)$ could help one search for a particular prime of known size N, by estimating how many primes there are in the neighborhood of N.

Not surprisingly, playing with primes has been a popular pastime of mathematicians. Perhaps this is because those who have made inroads, providing improved understanding, have become famous.

[1]You may view δ_n for the first 100 numbers with the one-line Matlab/Octave command `stem(isprime(1:100))`

2.1.2 Lec 5 Greatest common divisor (GCD)

Multiplying two numbers together, or dividing one by the other, is very inexpensive on today's computer hardware. However, factoring a large integer (i.e., 10^3 digits) into its primes, is very expensive. When the integers are large, it is so costly that it cannot be done in a lifetime, even with the fastest computers.

The obvious question is: "Can we find the largest common factor $k = gcd(m, n)$ without factoring (m, n)?" The answer is "yes," with the *Euclidean algorithm* (EA). While the EA falls short of factoring, it is fast and easily implemented.

If the two integer are in factored form, the answer is trivial. For example $5 = gcd(5 \cdot 13, 5 \cdot 17)$, and $17 = gcd(17 \cdot 53, 17 \cdot 3 \cdot 31)$. But what about $gcd(901, 1581)$? So the problem that computing the GCD solves is when the factors are not known. Since $901 = 53*17$ and $1581 = 3*17*31$, $gcd(901, 1581) = 17$, which is not obvious.

In Matlab the GCD may be computed using `k=gcd(m,n)`, which only allows integers as arguments (and removes the sign).

Matrix method: The GCD may be written a matrix recursion, based on Eq. 2.1.2. The two starting numbers are given by the vector (m_0, n_0). The recursion is then

$$\begin{bmatrix} m_{k+1} \\ n_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m_k \\ n_k \end{bmatrix}$$

This recursion continues until $m_{k+1} < n_{k+1}$, at which point m and n must be swapped. Because the output depends on the input, this is a nonlinear recursion (Postulate P1 (Linear/nonlinear) of Section 3.5.1, p. 107).

The direct method is inefficient because in recursively subtract n many times until the resulting m is less than n, as shown in Fig. 2.1. It also must test for $m < n$ at each iteration, and then swap m and n once that condition is met. This recursion is repeated until $m_{k+1} = 0$. At that stage the GCD is n_{k+1}. Figure 2.1, along with the above matrix relation, give the best insight into the Euclidean Algorithm, but at the cost of low efficiency.

Below is a Matlab code to find the gcd based on the strict definition of the EA:

```
n = gcd(m,n)
 while m ~=0
  A=m; B=n;
  m=max(A,B); n=min(A,B); %m>n
  m=m-n;
 end
```

This program keeps looping until $m = 0$. It first finds the min and max of the inputs, sets m as the max and n as the minimum. The next line $m = m - n$ removes the smaller number from the larger one. It then loops back and repeats the cycle. Thus the EA is a two step recursive method.

Division with rounding method: This method implements Eq. 2.2. It is not necessary to test that $m_{k+1} < n_{k+1}$. After computing the new value of n, using the floor function, the old value of m is saved as the new value of n (thus $m_{k+1} > n_{k+1}$

$$\begin{bmatrix} m_{k+1} \\ n_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -\lfloor \frac{m}{n} \rfloor \end{bmatrix} \begin{bmatrix} m_k \\ n_k \end{bmatrix}. \tag{2.2}$$

where $\lfloor x \rfloor$ finds the integer part of x ($\lfloor x \rfloor$ rounds toward $-\infty$). The method terminates when $m_{k+1} = 0$. The previous values m_k, n_k are the solutions to Bézout's identity (`gcd(n,m)=1`, namely $m_k m_0 + n_k n_0 = 1$), since the terminal state and the GCD of a, b is $m - n\lfloor m/n \rfloor = 0$, for which $n = gcd(a, b)$.

Below is 1-line vectorized code that is much more efficient than the direct matrix method:

```
k = gcd(m,n) %entry point: input m,n; output k=gcd(m,n)
M=[abs(m),abs(n)]; %init M
 while M(2) ~=0 % < n*eps to ''almost work'' with irrational inputs
  M = [M(2) - M(1)*floor(M(2)/M(1)); M(1)]; %M = [M(1); M(2)] with M(1)<M(2)
 end
```

With a minor extension in the test for "end," this code can be made to work with irrational inputs: e.g.: $(n\pi, m\pi)$.

This method calculates the number of times $n < m$ must subtract from m using the floor function. Note that the new value of m ($M(1)$) is always less than n (M(2)), and must remain greater or equal to zero. This one-line vector operation is then repeated until the remainder ($M(1)$) is 0. The gcd is then n ($M(2)$). When using irrational numbers, this still works except the error is never exactly zero, due to IEEE 754 rounding. Thus the criterion must be that the error is within some small factor times the smallest number (which in Matlab is the number `eps` $= 2.220446049250313 \times 10^{-16}$.

Thus without factoring the two numbers, the Euclidean algorithm recursively finds the gcd simply by ordering the two numbers and then updating them. Perhaps this is best seen with some examples.

The GCD is an important and venerable method, useful in engineering and mathematics, but, as best I know, is not typically taught in the traditional engineering curriculum.

Graphical meaning of the GCD: The Euclidean algorithm is actually very simple when viewed graphically. In Fig. 2.1 we show what is happening as one approaches the threshold. After reaching the threshold, the two number must be swapped, which is addressed by upper row of Eq. 2.2.

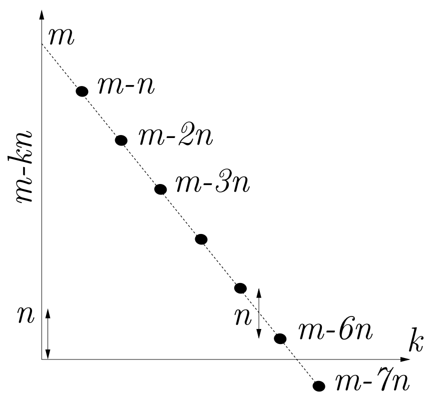


Figure 2.1: *The Euclidean Algorithm recursively subtracts n from m until the remainder $m - kn$ is either less than n or zero. For the case depicted here the value of k that renders the remainder less than n is $k = 6$. If one more step were taken ($k = 7$) the remainder becomes negative. By linear interpolation we can find that $m - an = 0$ when $a = m/n$, which for this example is close to a=6.5. In this example $6 = floor(m/n) < n$.*

Multiplication is simply recursive addition, and finding the gcd takes advantage of this fact. Lets take a trivial example, (9,6). Taking the difference of the larger from the smaller, and writing multiplication as sums, helps one see what is going on. Since 6=3*2, this difference may be written two different ways

$$9 - 6 = (3 + 3 + 3) - (3 + 3) = 0 + 0 + 3 = 3,$$

or

$$9 - 6 = (3 + 3 + 3) - (2 + 2 + 2) = 1 + 1 + 1 = 3.$$

Written out the first way, it is 3, because subtracting (3+3) from (3+3+3) leaves 3. Written out in the second way, each 3 is matched with a -2, leaving 3 ones, which add to 3. Of course the two decompositions must yield the same result because $2 \cdot 3 = 3 \cdot 2$. Thus finding the remainder of the larger number minus the smaller yields the gcd of the two numbers.

Coprimes: When the gcd of two integers is 1, the only common factor is 1. This is of key importance when trying to find common factors between the two integers. When 1=gcd(m,n) they are said to be *coprime*, which can be written as $m \perp n$. By definition, the largest common factor of coprimes is 1. But since 1 is not a prime ($\pi_1 = 2$), they have no common primes.

Generalizations of GCD: The GCD may be generalized in several significant ways. For example what is the GCD of two polynomials? To answer this question one must factor the two polynomials to identify common roots. This will be discussed in more detail in Section 3.2.2.

2.1.3 Lec 6 Continued Fraction Expansion (CFA)

Continued Fractions and circuit theory: One of the most powerful generalizations of the CFA seems to be the expansion of a function of a complex variable, such as the expansion of an impedance $Z(s)$, as a function of complex frequency s. This idea is described in Fig. 2.2 and Eq. 2.3. This is especially interesting in that it leads to a physical interpretation of the impedance in terms of a transmission line (horn), a structure well know in acoustics having a variable area $A(x)$ as function of the range variable x.

The CFA expansion is of great importance in circuit theory, where it is equivalent to an infinitely long segment of transmission line, composed of series and shunt impedance elements. Thus such a cascade network composed of 1 ohm resistors, has an input impedance of $(1+\sqrt{5})/2 \approx 1.6180$ [ohms].

The CFA may be extended to monomials in s. For example consider the input impedance of a cascade L-C transmission line as shown in Fig. 2.2. The input impedance of this transmission line is given by a continued fraction expansion of the form

$$Z_{in} = sL + \cfrac{1}{sC + \cfrac{1}{sL + \cfrac{1}{sC + \cfrac{1}{\cdots}}}} \qquad =: [sL; sC, sL, sC, \cdots]. \tag{2.3}$$

where we have again used the bracket notation to describe the CFA coefficients.

In some ways, Eq. 2.3 is reminiscent of the Taylor series expansion about $s = 0$, yet very different. In the limit, as the frequency goes to zero ($s \to 0$), the impedance of the inductors go to zero, and that of the capacitors go to ∞. In physical terms, the inductors become short circuits, while the capacitors become open circuits. The precise relation may be quantified by the use of composition, described in Fig. 1.11 of Section 2.1.3 (p. 52). Specifically

$$\begin{bmatrix} P_1 \\ U_1 \end{bmatrix} = \begin{bmatrix} 1 & sL \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \cdots \begin{bmatrix} 1 & sL \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} 1 & sL \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC & 1 \end{bmatrix} \begin{bmatrix} P_2 \\ -U_2 \end{bmatrix}. \tag{2.4}$$

It seems possible that this is the CFA generalization of the Taylor series expansion, built on composition. If we were to do the algebra we would find that $A(s), B(s), C(s), D(s)$ (i.e., Sections 1.3.6, 3.3.2) are ratios of polynomials having rational expansions as Taylor series. This seems like an important observation, that should have support beyond that of the engineering literature (Campbell, 1903; Brillouin, 1953; Ramo et al., 1965). Its interesting that (Brillouin, 1953) credits (Campbell, 1903).

In terms of the TL, it is a long piece of wire, with a delay determined by the velocity and the length, in units of cells each of length Δ. There are two basic parameters that characterize a transmission line, the characteristic resistance $r_0 = \sqrt{Z/Y}$ and the wave number

$$\kappa = 1/\sqrt{ZY} = s/\sqrt{LC} = s/c,$$

which gives $c = \sqrt{LC}$. Each of these is a constant as $\Delta \to 0$, and in that limit the waves travel as

$$f(t - x/c) = e^{-\kappa x} e^{-st},$$

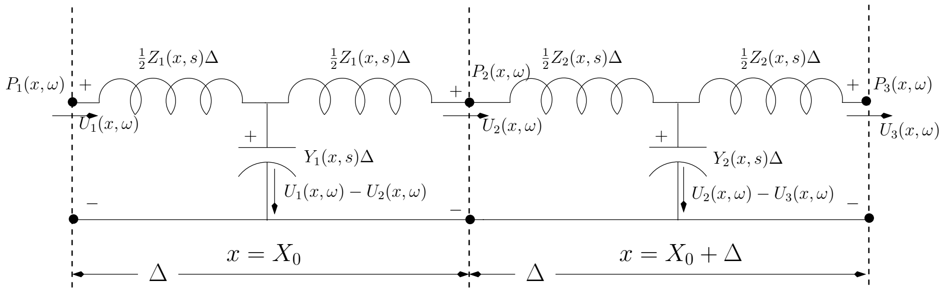


Figure 2.2: *This transmission line (TL) is known as a* low-pass *filter wave-filter (Campbell, 1922), composed of series inductors having an impedance $Z_k = sL$ and shunt admittance $Y_k = sC$. The characteristic resistance of the TL is $r_0 = \sqrt{Z_k/Y_k}$ and the wave velocity is $\kappa = \sqrt{Z_kY_k} = s/\sqrt{LC} = s/c$. Thus the wave velocity is $c = 1/\sqrt{LC}$. The length Δ [m] defines the physical size of each cell. For long wavelengths the wave-filter acts as a pure delay. But when the frequency increases above a cutoff frequency $f_c > c/\Delta$, the wavelength $\lambda = c/f$ is less than the size of a section Δ, and the system response becomes very high-order low-pass filter.*

with a wave resistance ($r_0 = \sqrt{L/C}$). The total delay $T = L/c$ where Δ is the TL cell length in meters and c is the velocity $c = 1/\sqrt{LC}$ in meters/second.

Since the CFA has a physical representation as a transmission line, as shown in Fig. 2.2, it can be of high utility for the engineer.[2] The theory behind this will be discussed in greater detail in Chapter 5. If you're ready to jump ahead, read the interesting book by Brillouin (1953) and the collected works of Campbell (1937).

2.2 Week 3

2.2.1 Lec 7 Pythagorean triplets (PTs) and Euclid's formula ,

Pythagorean triplets (PTs) have many applications in architecture and scheduling, which explains why they are important and heavily studied. For example, if one wished to construct a triangle with a perfect 90° angle, then the materials need to be squared off as shown in Fig. 1.8. The lengths of the sides need to satisfy PTs.

Derivation of Euclid's formula: The problem is to find integer solutions to the Pythagorean theorem (Eq. 1.1, p. 15). The solution method, said to be due to Diophantus, is call a *chord/tangent* method (Stillwell, 2010, p. 48). The method composes (Section 3.2.3) a line and a circle, where the line defines a chord within the circle (its not clear where the tangent line might go). The slope of the line is then taken to be rational, allowing one to determine integer solutions of the intersections points. This solution for *Pythagorean triplets* $[a, b, c]$ is known as *Euclid's formula* (Eq. 1.3, p. 1.3 (Stillwell, 2010, p. 4–9, 222).

The derivation methods of Diophantus have been lost, but Fermat and Newton figured out what Diophantus must have done (Stillwell, 2010, p. 49). Since Diophantus worked before algebra was invented, he described all the equations in prose (Stillwell, 2010, p. 93).

Derivation of Euclid's formula: The derivation is outlined in Fig. 2.3. Starting from two integers $[p > q > 0] \in \mathbb{N}$, composing a line having a rational slope $t = p/q$, with a circle (Stillwell, 2010, p. 6), reveals the formula for the Pythagorean triplets.

The construction starts with a circle and a line, which is terminated at the point $(-1, 0)$. The slope of the line is the free parameter t. By composing the circle and the line (i.e., solving for the intersection of the circle and line), the formula for the intersection point (a, b) may be determined in terms of t, which will then be taken as the rational slope $t = p/q \in \mathbb{Q}$.

[2]Continued fraction expansions of functions are know in the circuit theory literature as a *Cauer* synthesis (Van Valkenburg, 1964b).

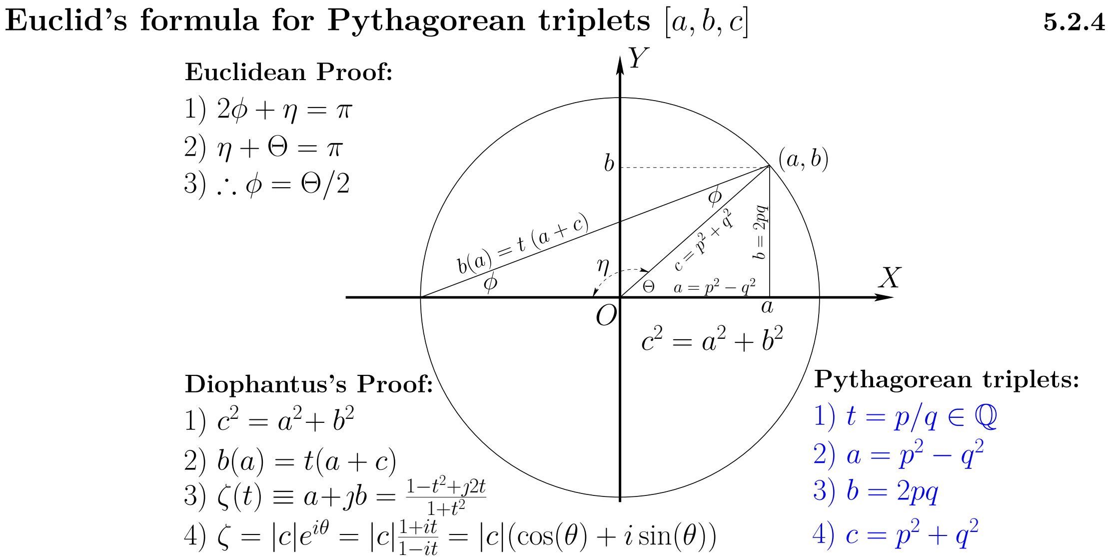


Figure 2.3: *Derivation of Euclid's formula for the Pythagorean triplets $[a, b, c]$, based on a composition of a line, having a rational slope $t = p/q \in \mathbb{Q}$, and a circle $c^2 = a^2 + b^2, [a, b, c] \in \mathbb{N}$. This analysis is attributed to Diophantus (Di·o·phan·tus) (250 CE), and today such equations are called Diophantine (Di·o·phan′·tine)* equations. *PTs have applications in architecture and scheduling, and many other practical problems.*

In Fig. 2.3 there are three panels, two labeled "Proofs." The *Euclidean Proof* shows the angle relationships of two triangles, the first an isosceles triangle formed by the chord, having slope t and two equal sides formed from the radius of the circle, and a second right triangle having its hypotenuse as the radius of the circle and its right angle vertex at $(a, 0)$. As shown, it is this smaller right triangle that must satisfy Eq. 1.1. The inner right triangle has its hypotenuse c between the origin of the circle (O) to the point (a, b). Side a forms the x axis and side b forms the y ordinate. Thus by construction Eq. 1.1 must be obeyed.

The *Diophantus Proof* is the heart of Diophantus' (250 CE) derivation, obtained by composing a line and a circle, as shown in Fig. 2.3. Diophantus's approach was to fix the line at $x = -c$ having a rational slope $t = p/q \in \mathbb{Q}$. He then solved for the intersection of the line and the circle, at (a, b).

The formula for the line is $b(a) = t(a + c)$, which goes through the points $(-c, 0)$ and (a, b). The circle is given by $a^2 + b^2 = c^2$. Composing the line with the circle gives

$$\begin{aligned}
a^2 + (t(a + c))^2 &= c^2 \\
a^2 + t^2(a^2 + 2ac + c^2) &= c^2 \\
(1 + t^2)a^2 + 2ct^2 a + c^2(t^2 - 1) &= 0
\end{aligned}$$

This last equation is a quadratic equation in a. In some sense it is not really a quadratic equation, since we know that $a = -c$ is a root. Dividing by $1 + t^2$

$$a^2 + \frac{2ct^2}{1 + t^2}a + \frac{c^2(t^2 - 1)}{1 + t^2} = 0,$$

makes it easier to complete the square, delivering the roots:

$$\left(a + \frac{ct^2}{1+t^2}\right)^2 - \left(\frac{ct^2}{1+t^2}\right)^2 + \frac{c^2(t^2-1)}{1+t^2} = 0$$
$$\left(a + \frac{ct^2}{1+t^2}\right)^2 - \frac{c^2t^4}{(1+t^2)^2} + \frac{c^2(t^2-1)(t^2+1)}{(1+t^2)^2} = 0$$
$$\left(a + \frac{ct^2}{1+t^2}\right)^2 - \frac{\cancel{c^2t^4} + c^2(\cancel{t^4}-1)}{(1+t^2)^2} = 0$$
$$\left(a + \frac{ct^2}{1+t^2}\right)^2 = \left(\frac{c}{1+t^2}\right)^2$$

The second to last equation simplifies (magic happens) because the known root $a = -c$ is embedded in the result.

Taking the square root gives the two roots

$$\begin{aligned} a_\pm + \frac{ct^2}{1+t^2} &= \pm\frac{c}{1+t^2} \\ (1+t^2)a_\pm &= -ct^2 \pm c = -c(t^2 \mp 1) \\ a_\pm &= -c\frac{t^2 \mp 1}{1+t^2}. \end{aligned}$$

The known root is $a_+ = -c$, because when the sign is $+$, the numerator and denominator terms cancel.

The root we have been looking for is a_-

$$a_- = c\,\frac{1-t^2}{1+t^2},$$

which allows us to solve for b_-

$$\begin{aligned} b_- &= \pm\sqrt{c^2 - a_-^2} \\ &= \pm c\sqrt{1 - \left(\frac{1-t^2}{1+t^2}\right)^2} \\ &= \pm c\sqrt{\frac{(1+t^2)^2 - (1-t^2)^2}{(t^2+1)^2}} \\ &= \pm\frac{2ct}{t^2+1}. \end{aligned}$$

Therefore the coordinates (a, b), the intersection point of the line and circle, are

$$(a(t), b(t)) = c\frac{[1-t^2, 2t]}{1+t^2}.$$

To obtain the Pythagorean triplets, as given in Fig. 2.3 and Eq. 1.3 of Section 1.2.7 (p. 36), set $t = p/q$, assuming $p > q \in \mathbb{Z}$, and simplify.

Complex roots: Defining the root as a complex number $\zeta(\Theta) \equiv a + b\jmath$ forces $a \perp b$ (i.e., forces the right triangle) and gives us polar coordinates, as defined by the figure as the Euclidean Proof

$$\zeta(\Theta) = |c|e^{\Theta\jmath} = |c|\left(\cos(\Theta) + \jmath\sin(\Theta)\right).$$

This naturally follows since

$$\zeta = |c|e^{\jmath\Theta(t)} = |c|\frac{1-t^2+2t\jmath}{1+t^2} = |c|\frac{\cancel{(1+\jmath t)}(1+\jmath t)}{\cancel{(1+t\jmath)}(1-t\jmath)} = (q+p\jmath)\sqrt{\frac{q+\jmath p}{q-p\jmath}}.$$

Examples of PTs include $a = 2^2 - 1^2 = 3$, $b = 2\cdot 2\cdot 1 = 4$, and $c = 2^2+1^2 = 5$, $3^2+4^3 = 5^2$.

Defining $p = q + N$ ($N \in \mathbb{N}$) gives slightly better parametric representation of the answers, as the pair (q, N) are a more systematic representation than (p, q), because the condition $p > q$ is accounted for, so the general properties of the solutions are expressed more naturally. Note that $b+c$ must always be a perfect square since $b + c = (p + q)^2 = (2q + N)^2$, as first summarized by Fermat Stillwell (2010, p. 212).

s

2.2.2 Lec 8 Pell's Equation

Eigenvalue solution to Pell's equation: To provide a full understanding of what was known to the Pythagoreans, it is helpful to provide the full solution to this recursive matrix equation, based on what we know today.

As shown in Fig. 1.10, (x_n, y_n) may be written as a power series of the 2x2 matrix A. To find the powers of a matrix, the well know modern approach is to *diagonalize* the matrix. For the 2x2 matrix case, this is relatively simple. The final result written out in detail for the general solution (x_n, y_n), as detailed in Appendix D (p. 143):

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \jmath^n \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} = E \begin{bmatrix} \lambda_+^n & 0 \\ 0 & \lambda_-^n \end{bmatrix} E^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.5}$$

The eigen-values are $\lambda_\pm = \jmath(1 \pm \sqrt{2})$ while the eigen-matrix and its inverse are

$$E = \frac{1}{\sqrt{3}}\begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.8165 & 0.8165 \\ 0.5774 & -0.5774 \end{bmatrix}, \qquad E^{-1} = \frac{\sqrt{3}}{2\sqrt{2}}\begin{bmatrix} 1 & \sqrt{2} \\ 1 & -\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.6124 & 0.866 \\ 0.6124 & -0.866 \end{bmatrix}$$

The relative "weights" on the two eigen-solutions are equal, as determined by

$$E^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\sqrt{3}}{2\sqrt{2}}\begin{bmatrix} 1 & \sqrt{2} \\ 1 & -\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\sqrt{3}}{2\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

We still need to prove that

$$\frac{x_n}{y_n} \underset{\infty}{\rightarrow} \sqrt{N},$$

which follows intuitively from Pell's equation, since as $(x_n, y_n) \to \infty$, the difference between x^2 and $2y^2$, the (± 1) becomes negligible.

WEEK 4

2.3 Week 4

2.3.1 Lec 9 Fibonacci Numbers

The Fibonacci sequence is famous in number theory. It is said that the sequence commonly appears in physical systems. Fibonacci numbers are related to the "golden ratio" $(1 + \sqrt{5})/2$, which could explain why these numbers appear in nature.

But from a mathematical point of view, the Fibonacci sequence does not seem special. It is generated by a linear recursion relationship, where the next number is the sum of the previous two (Eq. 1.6, p. 40)

$$x_{n+1} = x_n + x_{n-1}. \tag{2.6}$$

The term *linear* means that the principle of superposition holds (P1 (linear/nonlinear) of Section 3.5.1). To understand the meaning of this we need to explore the z-transform, the discrete-time version of the Laplace transform. We will return to this in Chapter 4.

A related linear recurrence relation is that the next output be the average of the previous two

$$x_{n+1} = \frac{x_n + x_{n-1}}{2}.$$

In some ways this relationship is more useful than the Fibonacci recursion, since it perfectly removes oscillations of the form -1^n (it is a 2-sample *moving average*, a trivial form of low-pass filter). And it is stable, unlike the Fibonacci sequence, with stable real eigenvalues (digital-poles) at $\lambda_\pm = (1, -0.5)$. Perhaps biology prefers unstable poles (to propagate growth?).

The most general 2d order recurrence relationships (i.e., digital filter) is

$$x_{n+1} = -bx_n - cx_{n-1},$$

with filter constants $b, c \in \mathbb{R}$ and poles at (completing the square), $\lambda_\pm = -b/2 \pm \sqrt{c - b/2}$.

Equation 2.6 may be written as a 2x2 matrix relationship. If we define $y_{n+1} = x_n$ then Eq. 2.6 is equivalent to (Eq. 1.7, p. 40)

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix}. \tag{2.7}$$

The first equation is $x_{n+1} = x_n + y_n$ while the second is $y_{n+1} = x_n$, which is the same as $y_n = x_{n-1}$. Note that the Pell 2x2 recursion is similar in form to the Fibonacci recursion. This removes mystique from both equations.

General properties of the Fibonacci numbers[a]

$$x_n = x_{n-1} + x_{n-2}$$

- This is a 2-sample *moving average* difference equation with an unstable pole
- $x_n = [0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \cdots]$, assuming $x_0 = 0, x_1 = 1$:
- Analytic solution (Stillwell, 2010, p. 194): $\sqrt{5}\, x_n \equiv \left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n \to \left(\frac{1+\sqrt{5}}{2}\right)^\infty$
  - $\lim_{n\to\infty} \frac{x_{n+1}}{x_n} = \frac{1+\sqrt{5}}{2}$
  - Example: $34/21 = 1.6190 \approx \frac{1+\sqrt{5}}{2} = 1.6180$ 0.10% error
- Matlab's $\mathtt{rat}(1 + \sqrt{5}) = 3 + 1/(4 + 1/(4 + 1/(4 + 1/(4)))) =: [3; 4, 4, 4, \cdots]$

[a] https://en.wikipedia.org/wiki/Fibonacci_number

Figure 2.4: Properties of the Fibonacci numbers (Stillwell, 2010, p. 28).

In the matrix diagonalization of the Pell equation we found that the eigenvalues were $\lambda_\pm = 1 \mp \sqrt{N}$, and the two solutions turned out to be powers of the eigenvalues. The solution to the Fibonacci recursion may similarly be expressed in terms of a matrix. These two cases may thus be reduced by the same 2x2 eigenvalue solution method.

The eigenvalues of the Fibonacci matrix are

$$det \begin{bmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 - \lambda - 1 = (\lambda - 1/2)^2 - (1/2)^2 - 1 = (\lambda - 1/2)^2 - 5/4 = 0$$

thus $\lambda_\pm = \frac{1\pm\sqrt{5}}{2} = [1.618, -0.618]$.

2.3.2 Lec 10 Exam I

Chapter 3

Algebraic Equations: Stream 2

3.1 Week 4

3.1.1 Lec 11 Algebra and geometry as physics

Before Newton could work out his basic theories, algebra needed to be merged with Euclid's early quantification of geometry. The key to putting geometry and algebra together is the Pythagorean theorem (Eq. 1.1), which is both geometry and algebra. To make the identification with geometry the sides of the triangle needed to be viewed as a length. This is done by recognizing that the area of a square is the square of a length. Thus a geometric proof requires one to show that the area of the square $A = a^2$ plus the area of square $B = b^2$ must equal the area of square $C = c^2$. There are many such constructions that show $A + B = C$ for the right triangle. It follows that in terms of coordinates of each vertex, the length of c is given by

$$c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \tag{3.1}$$

with $a = x_2 - x_1$ and $b = y_2 - y_1$. Thus Eq. 1.1 is both an algebraic and a geometrical statement. This is not immediately obvious.

Analytic geometry is based on coordinates of points, with the length given by Eq. 3.1. Geometry treats lines as lengths without specifying the coordinates (Eq. 1.1). Algebra gave a totally new view to the quantification of geometrical lengths by introducing a coordinate system. This resulted in an entire new way to work with conic sections, which were now explained in terms of equations having coordinate systems. When viewed through the lens of algebra, Eq. 1.1 is a circle having radius c. Complex numbers provide an equivalent representations, since if $z = x + y\jmath$, the unit circle is $z = e^{\jmath\theta}$ and $|z|^2 = x^2 + y^2$. Here we explore the relationships between points, represented as coordinates, describing geometrical objects. We shall do this with simple examples from analytic geometry.

For example, in terms of the geometry, the intersection of two circles can occur at two points, and the intersection of two spheres gives a circle. These ideas may be verified using algebra, but in a very different, since the line can traverse through the circle, like a piece of thread going through the eye of a needle. In such cases the intersections are complex intersections.

For each of these problems, the lines and circles may intersect, or not, depending on how they are drawn. Yet we now know that even when they do not intersect on the sheet of paper, they still have an intersection, but the solution is $\in \mathbb{C}$. Finding such solutions require the use of algebra rather than geometry. These ideas were in the process of being understood, first by Fermat and Descartes, then by Newton, followed by the Bernoulli family and Euler.

Complex analytic functions: A very delicate point, that seems to have been ignored for centuries, is that the roots of $P_n(x)$ are, in general, complex, namely $x_k \in \mathbb{C}$. It seems a mystery that complex numbers were not accepted once the quadratic equation was discovered, but they were not. Newton called complex roots *imaginary*, presumably in a pejorative sense. The algebra of complex numbers is first attributed to Bombelli in 1575, more than 100 years before Newton. One can only begin to imagine what Bombelli learned from Diophantus, following his discovery of Diophantus' Arithmetic, that he discovered in the Vatican library (Stillwell, 2010, p. 51).

It is interesting that Newton was using power series with fractional degree, thus requiring multi-valued solutions, much later to be known as *branch cuts* (c1851). These topics will be explored in Section 3.1.1.

When the argument is complex, analytic functions takes on an entirely new character. For example Euler's identity (1748) with $z = x + y\jmath \in \mathbb{C}$ results in $e^z \in \mathbb{C}$ (Stillwell, 2010, p. 315)

$$e^z = e^x(\cos(y) + \jmath \sin(y)).$$

It should be clear that the complex analytic functions results in a new category of algebra, with no further assumptions beyond allowing the argument to be complex.

Prior to 1851 most of the analysis *assumed* that the roots of polynomials were real ($x_k \in \mathbb{R}$), even though there was massive evidence that they were complex ($x_n \in \mathbb{C}$). This is clearly evident in Newton's work (c1687): When he found a non-real root, he ignore it (Stillwell, 2010, pp. 115-7). Euler (c1748) first derived the Zeta function as a function of real arguments $\zeta(x)$ with $\zeta, x \in \mathbb{R}$. Cauchy (c1814) broke this staid thinking with his analysis of complex analytic functions, but it was Riemann thesis (c1851), when working with Gauss (1777-1855), which had a several landmark breakthroughs. In this work Riemann introduced the *extended complex plane*, which explained the *point* at infinity. He also introduced *Riemann sheets* and *Branch cuts*, which finally allowed mathematics to better describe the physical world (Section 1.4.2).

Once the argument of an analytic function is complex, for example an impedance $Z(s)$, or the Riemann Zeta function $\zeta(s)$, The development of complex analytic functions led to many new fundamental theorems. Complex analytic functions have poles and zeros, branch cuts, Riemann sheets and can be analytic at the point at infinity. Many of these properties were first worked out by Augustin-Louis Cauchy (1789-1857), who drew heavily on the much earlier work of Euler, expanding Euler's ideas into the complex plane (Chapter 4).

Systems of equations

We don't need to restrict ourselves to polynomials in one variable. In some sense $y(x) = ax + b$ is already an equation in two variables. Moving beyond such a simple example, we can work with the equation for a circle, having radius r

$$y^2 + x^2 = r^2,$$

which is quadratic in two variables. Solving for roots $y(x_r) = 0$ ($y^2(x_r) = r^2 - x_r^2 = 0$) gives $(r - x_r)(r + x_r)$, which simply says that when the circle crosses the $y = 0$ line at $x_r = \pm r$.

This equation may also be factored as

$$(y - x\jmath)(y + x\jmath) = r^2,$$

as is easily demonstrated by multiplying out the two monomials. This does not mean that a circle has complex roots. A root is defined by either $y(x_r) = 0$ or $x(y_r) = 0$.

Writing the conic as a 2d degree polynomial gives

$$P_2(x) = ax^2 + bx + c,$$

with $y^2(x) = -P_2(x)$. Setting this equal to zero and completing the square (Eq. 1.12, p. 42), gives the equation for the roots

$$\left(x_\pm + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 + \frac{c}{a} = 0,$$

or

$$x_\pm = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}}.$$

The polynomial in factored form is

$$y^2 = -\left(x - \frac{b}{2a}\right)^2 + \left(\frac{b}{2a}\right)^2 - \frac{c}{a},$$

is a conic section, and becomes a circle with $a = 1$, $b = 0$ and $c = -r^2$.

3.2 Week 5

3.2.1 Lec 12 The physics behind complex analytic expressions: linear vs. nonlinear

The question we address here is "When do multi-variable complex analytic expressions appear in physics?" The most common example comes from the solution of the wave equation (Eq. 1.8) in three dimensions. Such cases arise in wave-guide problems, semiconductors, plasma waves, or for acoustic wave propagation in crystals (Brillouin, 1960) and the earth's mantel (e.g., seismic waves, earthquakes, etc.). The solutions to these problems are based on the *eigenfunction* for the *vector wave equation* (see Chapter 5),

$$P(s, \mathbf{x}) = e^{st} e^{-\boldsymbol{\kappa}\cdot\mathbf{x}}, \tag{3.2}$$

where $\mathbf{x} = [x\hat{x} + y\hat{y} + z\hat{z}]$ is a vector pointing in the direction of the wave, $[\hat{x}, \hat{y}, \hat{z}]$ are unit vectors in the three dimensions and $s = \sigma + \omega\jmath$ [rad] is the Laplace frequency. The function $\boldsymbol{\kappa}(s)$ is the complex vector *wave number*, which describes the propagation of a plane wave of radian frequency ω, in the $\mathbf{x}$ direction.

Just as the frequency $s = \sigma + \omega\jmath$ must be complex, it is important to allow the *wave number* function[1] to be complex,[2] because in general it will have a real part, to account for losses as the wave propagates. While it is common to assume there are no losses, in reality this assumption cannot be correct. In many cases it is an excellent approximation (e.g., even the losses of light in-vacuo are not zero) that gives realistic answers. But it is important to start with a notation that accounts for the most general situation, so that when losses must be accounted for, the notation need not change. With this in mind, we take the vector wave number to be complex

$$\boldsymbol{\kappa} = \mathbf{k}_r + \mathbf{k}\jmath,$$

where vector expression for the *lattice vector* is the imaginary part of $\boldsymbol{\kappa}$

$$\Im\boldsymbol{\kappa} = \mathbf{k} = \frac{2\pi}{\lambda_x}\hat{x} + \frac{2\pi}{\lambda_y}\hat{y} + \frac{2\pi}{\lambda_z}\hat{z}, \tag{3.3}$$

is the vector wave number for three dimensional lossless plane-wave solutions ($\mathbf{k}_r = 0$).

Add figure explain $[\lambda_x, \lambda_y, \lambda_z]$.

Equation Eq. 3.2 is linear in $\mathbf{x}$. If one takes the derivative with respect to either time or space,

$$\frac{\partial}{\partial t} e^{st} e^{-\boldsymbol{\kappa}\cdot\mathbf{x}} = s e^{st} e^{-\boldsymbol{\kappa}\cdot\mathbf{x}}, \qquad \frac{\partial}{\partial x} e^{st} e^{-\boldsymbol{\kappa}\cdot\mathbf{x}} = \frac{2\pi}{\lambda_x} e^{st} e^{-\boldsymbol{\kappa}\cdot\mathbf{x}},$$

we find the eigenvalue of that derivative.

The units of $\boldsymbol{\kappa}$ are reciprocal length [m^{-1}] since $\boldsymbol{\kappa}\cdot\mathbf{x}$ has units of radians. When there are losses $\kappa_r(s) = \Re\kappa(s)$ must be a function of frequency, due to the physics behind these losses. In many important cases, such as loss-less wave propagation in semiconductors, $\boldsymbol{\kappa}(\mathbf{x})$ is a function of direction and position (Brillouin, 1960).

[1]This function has many names in the literature, all of which are confusing. It is called the *wave number*, *propagation constant* and the Brillouin zone *dispersion function* (Brillouin, 1953, Ch. 1). However, its neither a number nor constant.

[2]In fact $\boldsymbol{\kappa}(s)$ is a complex analytic function of the Laplace frequency s.

When the eigenfunction Eq. 3.2 is applied to the wave equation, a quadratic (degree 2) algebraic expression results, known as the *dispersion relation.* The three dimensional dispersion relation

$$\left(\frac{s}{c}\right)^2 = \boldsymbol{\kappa}\cdot\boldsymbol{\kappa} = \left(\frac{2\pi}{\lambda_x}\right)^2 + \left(\frac{2\pi}{\lambda_y}\right)^2 + \left(\frac{2\pi}{\lambda_z}\right)^2 = k_x^2 + k_y^2 + k_z^2 \tag{3.4}$$

is a complex analytic algebraic relationship in four variables, frequency s and the three complex *lattice wave numbers.* This represents a three-dimensional generalization of the well know relation between wavelength and frequency, i.e., $f\lambda = c$. For lossless plane waves propagating in free space, $|\boldsymbol{\kappa}(s)| = \pm|s/c|$, where the sign accounts for the direction of the plane wave.

This scalar relation ($f\lambda = c$) was first deduced by Galileo in the 16th century and was then explored further by Mersenne a few years later.This relationship would have been important to Newton when formulating the wave equation, which he needed to estimate the speed of sound. We shall return to this in Chapters 4 and 5.

Inner product space: Another important example of algebraic expressions in mathematics is Hilbert's generalization of Eq. 1.1, known as the Schwarz inequality, shown in Fig. 3.1. What is special about this generalization is that it proves that when the vertex is 90°, the length of the leg is minimum.

Vectors may be generalize to have ∞ dimensions: $\vec{U},\vec{V} = [v_1, v_2, \cdots, v_\infty]$). The inner product (i.e., dot product) between two such vectors generalizes the finite dimentional case

$$\vec{U}\cdot\vec{V} = \sum_{k=1}^{\infty} u_k v_k.$$

As with the finite case, the *norm* $||\vec{U}|| = \sqrt{\vec{U}\cdot\vec{U}} = \sqrt{\sum u_k^2}$ is the dot product of the vector with itself, defining the *length* of the infinite component vector. Obviously there is an issue of convergence, if the norm for the vector to have a finite length.

It is a somewhat arbitrary requirement that $a, b, c \in \mathbb{R}$ for the Pythagorean theorem (Eq. 1.1). This seems natural enough since the sides are lengths. But, what if they are taken from the complex numbers, as for the lossy vector wave equation, or the lengths of vectors in $\mathbb{C}^n$? Then the equation generalizes to

$$\mathbf{c}\cdot\mathbf{c} = ||\mathbf{c}||^2 = \sum_{k=1}^{n} |c_k|^2,$$

where $||\mathbf{c}||^2 = (\mathbf{c}, \mathbf{c})$ is the inner (dot) product of a vector $\mathbf{c}$ with itself, where $|c_k|$ is the magnitude the complex c_k. As before, $||\mathbf{c}|| = \sqrt{||\mathbf{c}||^2}$ is the *norm* of vector $\mathbf{c}$, akin to a length.

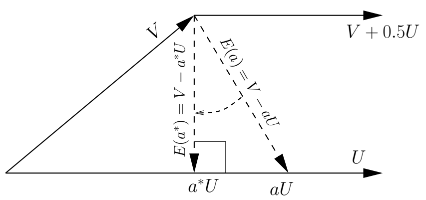


Figure 3.1: *The* Schwarz inequality *is related to the shortest distance (length of a line) between the ends of the two vectors.* $||U|| = \sqrt{(U\cdot U)}$ *as the dot product of that vector with itself.*

Schwarz inequality The *Schwarz inequality* says that the magnitude of the inner product of two vectors is less than or equal to the lengths of the product of the two vectors. This follows from Fig. 3.1 in terms of the geometry.

A proof of the Schwarz inequality is as follows: From these definitions we may define the minimum difference between the two vectors as the perpendicular from the end of one to the intersection of the second. As shown in Fig. 3.1, $U \perp V$ may be found by minimizing the length of the vector difference:

$$\begin{aligned}\min_{\alpha} ||V - \alpha U||^2 &= ||V||^2 + 2\alpha V \cdot U + \alpha^2 ||U||^2 > 0 \\ 0 &= \partial_\alpha \, (V - \alpha U) \cdot (V - \alpha U) \\ &= V \cdot U - \alpha^* ||U||^2 \\ \therefore \quad \alpha^* &= V \cdot U / ||U||^2.\end{aligned}$$

The *Schwarz inequality* follows:

$$I_{\min} = ||V - \alpha^* U||^2 = ||V||^2 - \frac{|U \cdot V|^2}{||U||^2} > 0$$

$$0 \le |U \cdot V| \le ||U|| \, ||V||$$

One may define the direction cosine between the two vectors as

$$\cos(\theta) = \frac{U \cdot V}{||U|| \, ||V||}$$

which is a measure of the projection of one of the vectors onto the other, as shown by the figure. The $\cos(\theta)$ does not represent a complex analytic function. Namely $\theta \in \mathbb{R}$, not $\in \mathbb{C}$ (as best I know).

An important example of such a vector space includes the definition of the *Fourier Transform*, where we may set

$$U(\omega) = e^{-\omega_0 \jmath t} \qquad V(\omega) = e^{\omega \jmath t} \qquad U \cdot V = \int_\omega e^{\jmath \omega t} e^{-\jmath \omega_0 t} \frac{d\omega}{2\pi} = \delta(\omega - \omega_0).$$

We shall return to this topic in Lecture 1.3.9.

Power vs. power series, linear vs. nonlinear

Another place where equations of second degree appear in physical applications is in energy and power calculations. The electrical power is given by the product of the voltage $v(t)$ and current $i(t)$ (or in mechanics as the force times the velocity). For example if we define $\mathcal{P} = v(t)i(t)$ to be the power $\mathcal{P}$ [Watts], then the *total energy* [Joules] at time t is (Van Valkenburg, 1964a, Chapter 14)

$$\mathcal{E}(t) = \int_0^t v(t)i(t)dt.$$

From this observe that the power is the rate of change of the total energy

$$\mathcal{P}(t) = \frac{d}{dt}\mathcal{E}(t).$$

Ohm's Law and impedance: The ratio of voltage over the current is call the *impedance* which has units of [Ohms]. For example given a resistor of $R = 10$ [ohms],

$$v(t) = R \, i(t).$$

Namely 1 amp flowing through the resistor would give 10 volts across it. Merging the linear relation due to Ohm's law with the definition of power, shows that the instantaneous power in a resistor is quadratic in voltage and current

$$\mathcal{P} = v(t)^2/R = i(t)^2 R. \tag{3.5}$$

Note that Ohm's law is linear in its relation between voltage and current whereas the power and energy are *nonlinear*.

Ohm's Law generalizes in a very important way, allowing the impedance (e.g., resistance) to be a linear complex analytic function of complex frequency $s = \sigma + \omega\jmath$ (Kennelly, 1893; Brune, 1931b). Impedance is a fundamental concept in many fields of engineering. For example:[3] Newton's second law $F = ma$ obeys Ohm's law, with mechanical impedance $Z(s) = sm$. Hooke's Law $F = kx$ for a spring is described by a mechanical impedance $Z(s) = k/s$. In mechanics a "resistor" is called a *dashpot* and its impedance is a positive and real constant.[4]

Force is a voltage are

Kirchhoff's Laws KCL, KVL: The laws of electricity and mechanics may be written down using Kirchoff's Laws current and voltage laws, (KCL, KVL), which lead to linear systems of equations in the currents and voltages (velocities and forces) of the system under study, with complex coefficients having positive real parts.

Points of major confusion are a number of terms that are misused, and overused, in the fields of mathematics, physics and engineering. Some of the most obviously abused terms are *linear/nonlinear*, *energy*, *power*, *power series*. These have multiple meanings, which can, and are, fundamentally in conflict.

Transfer functions (Transfer matrix): The only method that seems to work, to sort this out, is to cite the relevant physical application, in specific contexts. The most common touch point is a physical system that has an input $x(t)$ and an output $y(t)$. If the system is linear, then it may be represented by its *impulse response* $h(t)$. In such cases the system equation is

$$y(t) = h(t) \star x(t) \leftrightarrow Y(\omega) = H(s)|_{\sigma=0} X(\omega),$$

namely the convolution of the input with the impulse response gives the output. From Fourier analysis this relation may be written in the real frequency domain as a product of the Laplace transform of the impulse response, evaluated on the $\omega\jmath$ axis and the Fourier transform of the input $X(\omega) \leftrightarrow x(t)$ and output $Y(\omega) \leftrightarrow y(t)$.

Mention ABCD Transfer matrix

If the system is nonlinear, then the output is not given by a convolution, and the Fourier and Laplace transforms have no obvious meaning.

The question that must be addressed is why is the power said to be nonlinear whereas a power series of $H(s)$ said to be linear. Both have powers of the underlying variables. This is massively confusing, and must be addressed. The question will be further addressed in Section 3.5.1 in terms of the system postulates of physical systems.

Whats going on? The domain variables must be separated from the codomain variables. In our example, the voltage and current are multiplied together, resulting in a nonlinear output, the power. If the frequency is squared, this is describing the degree of a polynomial. This is not nonlinear because it does not impact the signal output, it characterizes the Laplace transform of the system response.

3.2.2 Lec 13 Root classification of polynomials

Root classification for polynomials of Degree $* = 1$–4 (p.102);
Quintic ($* = 5$) cannot be solved: Why?
Fundamental Thm of Algebra (d'Alembert, ≈1760)

Add intro & merge convolution discussions.

[3] In acoustics the pressure is a potential, like voltage. The force per unit area is given by $f = -\nabla p$ thus $F = -\int \nabla p \, dS$. Velocity is analogous to a current. In terms of the velocity potential, the velocity per unit area is $v = -\nabla\phi$.

[4] https://en.wikipedia.org/wiki/Impedance_analogy

Convolution

As we discussed in Chapter 1, given the roots, the construction of higher degree polynomials, is greatly assisted by the convolution method. This has physical meaning, and gives insight into the problem of factoring higher order polynomials. By this method we can obtain explicit relations for the coefficients of any polynomial in terms of its roots.

Extending the example of Section 1.3.3, let's find the relations for the cubic. For simplicity, assume that the polynomial has been normalized so that the lead x^3 term has coefficient 1. Then the cubic in terms of its roots $[a, b, c]$ is a convolution of three terms

$$[1, a] \star [1, b] \star [1, c] = [1, a + b, ab] \star [1, c] = [1, a + b + c, ab + c(a + b), abc].$$

Working out the coefficients for a *quartic* gives

$$[1, a+b+c, ab+c(a+b), abc] \star [1, d] = [1, a+b+c+d, d(a+b+c)+c(a+b)+ab, d(ab+ac+bc)+abc, abcd].$$

It is clear what is going on here. The coefficient on x^4 is 1 (by construction). The coefficient for x^3 is the sum over the roots. The x^2 term is the sum over all possible products of pairs of roots, The linear term x is the sum over all triple products of the four roots, and finally the last term (a constant) is the product of the four roots.

In fact this is a well known, a frequently quoted result from the mathematical literature, and trivial to show given an understand of convolution. If one wants the coefficients for the quintic, it is not even necessary to use convolution, as the pattern (rule) for all the coefficients is now clear.

You can experiment with this numerically using Matlab's convolution routine `conv(a,b)`. Once we start studying Laplace and Fourier transforms, convolution becomes critically important because multiplying an input signal in the frequency domain by a transfer function, also a function of frequency, is the same a convolution of the time domain signal with the inverse Laplace transform of the transfer function. So you didn't need to learn how to take a Laplace transform, and then learn convolution. We have learned convolution first independent of the Fourier and Laplace transforms.

When the coefficients are real, the roots must appear as conjugate pairs. This is an important symmetry.

For the case of the quadratic we have the relations between the coefficients and the roots, found by completing the square. This required isolating x to a single term, and solving for it. We then proceeded to find the coefficients for the cubic and quartic case, after a few lines of calculation. For the quartic

$$\begin{aligned}
a_4 &= 1 \\
a_3 &= a + b + c + d \\
a_2 &= d(a + b + c) + c(a + b) + ba \\
a_1 &= d(ab + ac + bd) + abc \\
a_0 &= abcd
\end{aligned}$$

These relationships are algebraically nonlinear in the roots. From the work of Galois, for $N \geq 5$, this system of equations is impossible to invert. Namely, given a_k, one may not determine the four roots $[a, b, c, d]$ analytically. One must use numeric methods.

To gain some insight, let us look at the problem for $N = 2$, which has a closed form solution:

$$\begin{aligned}
a_2 &= 1 \\
a_1 &= a + b \\
a_0 &= ab
\end{aligned}$$

We must solve for $[a, b]$ given twice the mean, $2(a+b)/2$, and the square of the geometric mean $\left(\sqrt{ab}\right)^2$. Since we already know the answer (i.e, the quadratic formula). The solution was first worked out by the

Babylonians (2000 BCE) Stillwell (2010, p. 92). It is important to recognize that for physical systems, the coefficients a_k are real. This requires that the roots come in conjugate pairs ($b = a^*$), thus $ab = |a|^2$ and $a + b = 2\Re a$, which makes the problem somewhat more difficult, due to the greater symmetry.

Once you have solved this problem, feel free to attempt the cubic case. Again, the answer is known, after thousands of years of searching. The solution to the cubic is given in (Stillwell, 2010, pp. 97-9), as discovered by Cardano in 1545. According to Stillwell "The solution of the cubic was the first clear advance in mathematics since the time of the Greeks." The ability to solve this problem required algebra, and the solutions were complex numbers. The denial of complex numbers was, in my view, the main stumbling block in the progress of these solutions. For example, how can two parallel lines have a solution? Equally mystifying, how can a circle and a line, that do not intersect, have intersections? From the algebra we know that they do. This was a basic problem that needed to be overcome. This story is still alive,[5] because the cubic solution is so difficult.[6] One can only begin to imagine how much more difficult the quartic is, solved by Cardano's student Ferrair, and published by Cardano in 1545. The impossibility of the quintic was finally resolved in 1826 by Able (Stillwell, 2010, p. 102).

Finally with these challenges behind them, Analytic Geometry, relating of algebra and geometry, via coordinate systems, was born.

3.2.3 Lec 14: Analytic Geometry

Lec 14: Early Analytic Geom (Merging Euclid and Descartes): Composition of degrees n, m gives degree $m \cdot n$

Composition, Intersection and Gaussian elimination

The first "algebra" (al-jabr) is credited to al-Khwarizmi (830 CE). Its invention advanced the theory of polynomial equations in one variable, Taylor series, and composition versus intersections of curves. The solution of the quadratic equation had been worked out thousands of year earlier, but with algebra a general solution could be defined. The Chinese had found the way to solve several equations in several unknowns, for example, finding the values of the intersection of two circles. With the invention of algebra by al-Khwarizmi, a powerful tool became available to solve the difficult problems.

Composition, Elimination and Intersection In algebra there are two contrasting operations on functions: *composition* and *Elimination*.

Composition: Composition is the merging of functions, by feeding one into the other. If the two functions are f, g then their composition is indicated by $f \circ g$, meaning the function $y = f(x)$ is substituted into the function $z = g(y)$, giving $z = g(f(x))$.

Composition is not limited to linear equations, even though that is where it is most frequently applied. To compose two functions, one must substitute one equation into the other. That requires solving for that substitution variable, which is not always possible in the case of nonlinear equations. However many tricks are available that may work around this restrictions. For example if one equation is in x^2 and the other in x^3 or $\sqrt{x}$, it may be possible to multiply the first by x or square the second. The point is that one of the variables must be isolated so that when it is substituted into the other equations, the variable is removed from the mix.

Examples: Let $y = f(x) =: x^2 - 2$ and $z = g(y) =: y + 1$. Then

$$g \circ f = g(f(x)) = (x^2 - 2) + 1 = x^2 - 1. \tag{3.6}$$

[5]M. Kac, *How I became a mathematician." American Scientist (72), 498–499.*

[6]`https://www.google.com/search?client=ubuntu&channel=fs&q=Kac+\%22how+I+became+a\%22+1984+pdf&ie=utf-8&oe=utf-8`

In general composition does not commute (i.e., $f \circ g \neq g \circ f$), as is easily demonstrated. Swapping the order of composition for our example gives

$$f \circ g = f(g(y)) = z^2 - 2 = (y+1)^2 - 2 = y^2 + 2y - 1. \tag{3.7}$$

Intersection: Complimentary to composition is *intersection* (i.e., decomposition) (Stillwell, 2010, pp. 119,149). For example, the intersection of two lines is defined as the point where they meet. This is not to be confused with finding roots. A polynomial of degree N has N roots, but the points where two polynomials intersect has nothing to do with the roots of the polynomials. The intersection is a function (equation) of lower degree, implemented with Gaussian elimination.

Intersection of two lines Unless they are parallel, two lines meet at a point. In terms of linear algebra this may be written as 2 linear equations (left) along with the intersection point $[x_1, x_2]^T$, given by the inverse of the 2x2 set of equations (right).[7]

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad\qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \tag{3.8}$$

where $\Delta = ab - cd$ is called the *determinant*. By substituting the right expression into the left, and taking the inverse we obtain the intersection point. If $\Delta = 0$ there can be no solution, in which case the two lines are parallel (they meet at infinity.)

Algebra will give the solution when geometry cannot. When the two curves fail to intersect on the real plane, the solution still exists, but is complex valued. In such cases, geometry, which only considers the real solutions, fails. For example, when the coefficients $[a, b, c, d]$ are complex, the solution exists, but the determinant can be complex. Thus algebra is much more general than geometric. Geometry fails when the solution has a complex intersection.

3.3 Week 6

3.3.1 Lec 15 Gaussian Elimination

Example problems using Gaussian Elimination: Gaussian Elimination is valid for nonlinear systems of equations. Till now we have emphasized the reduction of linear systems of equations.

Problem 1: Two lines *in a plane* either intersect or are parallel, in which case they are said to meet at ∞. Does this make sense? The two equations that describe this may be written in matrix form as $Ax = b$, which written out as

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{3.9}$$

The intersection point x_0, y_0 is given by the solution two these two equations

$$\begin{bmatrix} x_1 \\ x_1 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \tag{3.10}$$

where $\Delta = a_{11}a_{22} - a_{12}a_{21}$ is the determinant of matrix A (Matlab's `det(A)` function).

[7] It is very important to note when writing the equations in matrix format, the unknowns are x_1, x_2 whereas in the original equations they were y, x. The starting equations are $ay_1 + bx_1 = c$ and $cy_1 + dx_2 = d$ but in matrix format the names are changed. The first equation is $ax_1 + bx_2 = y_1$ and $cx_1 + dx_2 = y_2$. The first time you meet this scrambling of terminology it can be very confusing. In matrix equations the coordinates of the graph are (x_1, x_2) rather than the usual x, y.

It is useful to give an interpretation of these two equations. Each row of the 2x2 matrix defines a line in the (x, y) plane. The top row is

$$a_{11}x + a_{12}y = b_1.$$

Normally we would write this equation as $y(x) = \alpha x + \beta$, where α is the slope and β is the *intercept* (i.e., $y(0) = \beta$). In terms of the elements of matrix A, the slope of the first equation is $\alpha = -a_{11}/a_{12}$ while the slope of the second is $\alpha = -a_{21}/a_{22}$. The two slopes are equal (the lines are parallel) when $-a_{11}/a_{12} = -a_{21}/a_{22}$, or written out

$$\Delta = a_{11}a_{22} - a_{12}a_{21} = 0.$$

Thus when the determinate is zero, the two lines are parallel and there is no solution to the equations.

This 2x2 matrix equation is equivalent to a 2^d degree polynomial. If we seek an eigenvector solution $[e_1, e_2]^T$ such that

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \lambda \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{3.11}$$

the 2x2 equation becomes singular, and λ is one of the roots of the polynomial. One may proceed by merging the two terms to give

$$\begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{3.12}$$

Clearly this new matrix has no solution, since if it did, $[e_1, e_2]^T$ would be zero, which is nonsense. If it has no solution, then the determinant of the matrix must be zero. Forming this determinate gives

$$(a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0$$

thus we obtain the following quadratic equation for the roots $\lambda_\pm$ (eigenvalues)

$$\lambda_\pm^2 - (a_{11} + a_{22})\lambda_\pm + \Delta = 0.$$

When $\Delta = 0$, one eigenvalue is zero while the other is $a_{11} + a_{22}$, which is known as the *trace* of the matrix.

In summary: Given a "linear" equation for the point of intersection of two lines, we see that there must be two points of intersection, as there are always two roots of the quadratic *characteristic polynomial*. However the two lines only intersect at one point. Whats going on? What is the meaning of this second root?

Needs work.

Some simple examples will help. The eigenvalues depend on the relative slopes of the lines, which in general can become complex. The intercepts are dependent on **b**. Thus when the RHS is zero, the eigenvalues are irrelevant. This covers the very simple examples. When one eigenvalue is real and the other is imaginary, more interesting things are happening since the slope of one line is real and the slope of the other is pure imaginary. The lines can intersect in the real plane, and again in the complex plane.

Lets try an example of two lines, slopes of 1 and 2: $y_1 = x + a$ and $y_2 = 2x + b$. In matrix form Let

$$\begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} \tag{3.13}$$

The determinate is $\Delta = -1$, thus the solution is

$$\begin{bmatrix} y_0 \\ x_0 \end{bmatrix} = -1 \begin{bmatrix} -2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a - b \\ a - b \end{bmatrix}. \tag{3.14}$$

Thus the two real lines having slopes of 1 and 2 having intercepts of a and b, meet $(x_0, y_0) = (2a-b, a-b)$. We may verify by substituting $x = a - b$ into the starting equations $y_1 = (a - b) + a = 2a - b$ and $y_2 = 2(a - b) + b = 2a - b$, which each $2a - b$.

While there is a unique solution, there are two eigenvalues, given by the roots of

$$(1 - \lambda_\pm)(-2 - \lambda_\pm) + 1 = 0.$$

If we transfer the sign from one monomial to the other

$$(-1 + \lambda_\pm)(2 + \lambda_\pm) + 1 = 0$$

and reorder for simplicity

$$(\lambda_\pm - 1)(\lambda_\pm + 2) + 1 = 0$$

we obtain the quadratic for the roots

$$\lambda_\pm^2 + \lambda_\pm - 1 = 0.$$

Completing the square gives

$$(\lambda_\pm + 1/2)^2 = 3/4.$$

or

$$\lambda_\pm = -1/2 \pm \sqrt{3}/2.$$

The question is, what is the relationship between the eigenvalues and the final solution, if any? Maybe none. The solution (x_0, y_0) is reasonable, and its not clear that the eigenvalues play any useful role here, other than to predict there is a second solution. I'm confused.

Two lines in 3-space: In three dimensions

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{3.15}$$

Each row of the matrix describes a plane, which is said to be linear in the unknowns (x, y, z). Thus the system of linear equations represents three planes, which must intersect at one point. If two planes are parallel, there is no real solution. In this case the intersection by the third plane generates two parallel lines.

As in the 2x2 case, one may convert this linear equation into a cubic polynomial by setting the determinant of the matrix, with $-\lambda$ subtracted from the diagonal, equal to zero. That is, $det(A - \lambda I) = 0$. Here I is the matrix with 1 on the diagonal and zero off the diagonal.

Simple example: As a simple example, let the first plane be $z = 0$ (independent of x, y), the second parallel plane be $z = 1$ (independent of (x, y)) and the third plane be $x = 0$ (independent of y, z). This results in the system of equations

$$\begin{bmatrix} 0 & 0 & a_{13} \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{3.16}$$

Writing out the three equations we find $a_{13}z = 0$, $a_{23}z = 1$, and $a_{31}x = 0$. Note that $det(A) = 0$ (we need to learn how to compute the 3x3 determinant). This means the three planes never intersect at one point. Use Matlab to find the eigenvalues.

3.3.2 Lec 16 Matrix composition: Bilinear and ABCD transformations

The Transmission matrix

A transmission matrix is a 2x2 matrix that characterizes a 2-port circuit, one having an input and output voltage and current, as shown in Fig. 1.11. The input is the voltage and current V_1, I_1 and the output is the voltage and current $V_2, -I_2$, with the current always defined to flow into the port. For any such a linear network, the input-output relations may be written in a totally general way as

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}.$$

In Section 1.3.6 we showed that a cascade of such matrices is composition. We shall show below that the justification of this relationship is based on the composition of bilinear transformations.

Expanding Eq. 1.28 into its individual equations demonstrates the linear form of the relations

$$V_1 = A(s)V_2 - B(s)I_2 \qquad\qquad I_1 = C(s)V_2 - D(s)I_2,$$

quantifying the relationship between the input voltage and current to its output voltage and current.

Define $H(s) = V_2/V_1$ as the *transfer function*, as the ratio of the output voltage V_2 over the input voltage V_1, under the constraint that the output current $I_2 = 0$. From this definition $H(s) = 1/A(s)$.

In a similar fashion we may define the meaning of all four functions as

$$A(s) \equiv \left.\frac{V_1}{V_2}\right|_{I_2=0} \qquad\qquad B(s) \equiv -\left.\frac{V_1}{I_2}\right|_{V_2=0} \tag{3.17}$$

$$C(s) \equiv \left.\frac{I_1}{V_2}\right|_{I_2=0} \qquad\qquad D(s) \equiv -\left.\frac{I_1}{I_2}\right|_{V_2=0} \tag{3.18}$$

From Eq. 1.28 one may compute any desired quantity, specifically those quantities defined in Eq. 3.18, the open circuit voltage transfer function $(1/A(s))$, the short-circuit transfer current $(1/D(s))$ and the two transfer impedances $B(s)$ and $1/C(s)$.

In the engineering fields this matrix composition is called the *Transmission matrix*, also known as the ABCD method. It is a powerful method that is easy to learn and use, that gives important insights into transmission lines, and thus even the 1 dimensional wave equation.

Derivation of ABCD matrix for example of Fig. 1.11 (p. 52).

The derivation is straight forward by the application of Ohm's Law, as shown in Section 1.3.6.

The convenience of the ABCD matrix method is that the output of one is identically the input of the next. Cascading (composing) the results for the series inductor with the shunt compliance leads to the 2x2 matrix form that precisely corresponds to the transmission line CFA shown in Fig. 2.2,

$$\begin{bmatrix} V_n(s) \\ In(s) \end{bmatrix} = \begin{bmatrix} 1 & sL_n \\ 1 & 0 \end{bmatrix} \begin{bmatrix} V_{n+1}(s) \\ -I_{n+1}(s) \end{bmatrix}. \tag{3.19}$$

This matrix relation characterizes the series mass term sL_n. A second equation maybe be used for the shunt capacitance term $sY_n(s)$

$$\begin{bmatrix} V_n(s) \\ In(s) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ sC_n & 0 \end{bmatrix} \begin{bmatrix} V_{n+1}(s) \\ -I_{n+1}(s) \end{bmatrix}. \tag{3.20}$$

The positive constants $L_n, C_n \in \mathbb{R}$ represent the series mass (inductance) and the shunt compliance (capacitance) of the mechanical (electrical) network. The integer n indexes the series and shunt sections, that are composed one following the next.

Matrix composition and the bilinear transform: Now that we have defined the composition of two functions, we will use it to define the *Möbius* or *bilinear* transformation. Once you understand how this works, hopefully you will understand why it is the unifying element in many important engineering problems.

The bilinear transformation is given by

$$w = \frac{a + bz}{c + dz}$$

This takes one complex number $z = x + iy$ and transforms it into another complex number $w = u + iv$. This transformation is *bilinear* in the sense that its linear in both the input and output side of the equation. This may be seen when written as

$$(c + dz)w = a + bz,$$

since this relation is linear in the coefficients $[a, b, c, d]$. An important example is the transformation between impedance $Z(s)$ and reflectance $\Gamma(s)$,

$$\Gamma(s) = \frac{Z(s) - r_0}{Z(s) + r_0},$$

which is widely used in transmission line problems. In this example $w = \Gamma, z = Z(s), a = -r_0, b = 1, c = r_0, d = 1$.

If we define a second bilinear transformation (this could be the transformation from reflectance back to impedance)

$$r = \frac{\alpha + \beta w}{\gamma + \delta w},$$

and then compose the two something astray wrt arguments

$$w \circ r = \frac{a + b\, r}{c + d\, r} = \frac{a(\gamma + \delta w) + b(\alpha + \beta w)}{c(\gamma + \delta w) + d(\alpha + \beta)w} = \frac{a\gamma + b\alpha + (a\delta + b\beta)w}{c\gamma + d\alpha + (c\delta + d\beta)w},$$

something surprising happens. The composition $w \circ r$ may be written in matrix form, as the product of two matrices that represents each bilinear transform. This may be seen as true by inspecting the coefficients of the composition $w \circ r$ (shown above) and the product of the two matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} (a\gamma + b\alpha) & (a\delta + b\beta) \\ (c\gamma + d\alpha) & (c\delta + d\beta) \end{bmatrix}.$$

The the power of this composition property of the bilinear transform may be put to work solving important engineering problems, using *transmission matrices*.

3.3.3 Lec 17 Introduction to the Branch cut and Riemann sheets

Branch cuts are required to preserve the single-valued nature of complex analytic functions. When an analytic function is multi-valued, some method needs to be devised to allow the multi-valued complex analytic function to be expanded as a Taylor series, which is necessarily single-valued. It follows that each single-valued sheet must have a different expansion, valid out to the nearest pole (or singularity). We shall explain these ideas with the simplest case, the double-valued square root function $w(z) = \pm\sqrt{z}$, as shown in Fig. 1.15 (p. 69).

3.4 Week 7

3.4.1 Lec 18 Complex analytic mappings (domain coloring)

When one uses complex analytic functions it is helpful to understand their properties in the complex plane. In this sections we explore several well-known functions using domain coloring, discussed in some detail in Section 1.3.8, p. 55. For the following figures the coordinate systems are defined by $s = \sigma + \omega \jmath = \sqrt{x^2 + y^2} e^{\jmath\theta}$ and $w = u + v\jmath = \sqrt{u^2 + v^2} e^{\jmath\psi}$.

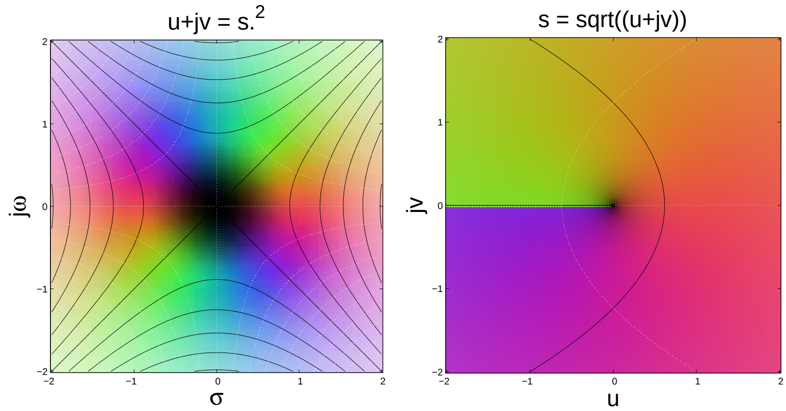


Figure 3.2: *Here the Cartesian coordinate map between $s = \sigma + \omega \jmath$ and $w = u + v\jmath$. The left shows the mapping $w(s) = s^2$. The right shows the lower branch of the inverse $s(w) = \sqrt{w}$, as shown in Fig. 1.15.*

For the first example (Fig. 3.2) $w(s) = s^2$ and its inverse $s(w) = \sqrt{w}$ are shown. On the left the red region, corresponding to 0° [degrees], appears at both 0 and 180 ($u = \pm 1$) in the w plane. This is because in polar coordinates $s^2 = |s|^2 e^{2\theta\jmath}$ where θ is the angle of $s = |s|e^{2\theta\jmath}$. The square causes the phase to rotate twice around for once around the s plane. Namely the angle is doubled, and the magnitude squared. Due to the faster changing phase in w, there are two red regions, one when $\theta = 0$ and the second at $\theta = \pi$ ($\angle w(s) = 2\theta$). The black spot is dilated due to the squaring of the radius (expanding it).

On the right the $\sqrt{w} = \sqrt{|w|}e^{\phi/2}$. Because the angle of w is divided by two, it takes twice as much phase (in w) to cover the same angle. Thus the red region (0°) is expanded. We barely see the violet 90° and yellow −90° angles. There is a *branch cut* running from $w = 0$ to $w = \infty$. As the branch cut is crossed, the function switches *Riemann sheets*, going from the top sheet (shown here) to the bottom sheet (not shown). Figure 1.15 in Section 3.3.3 depicts what is going on with these two sheets, and show the branch cut from the origin (point O) to ∞. In this depiction the first sheet $(+\sqrt{z})$ is on the bottom, while the second sheet $(\sqrt{z})$ is on top. For every value of z there are two possible outcomes, $\pm\sqrt{z}$, represented by the two sheets.

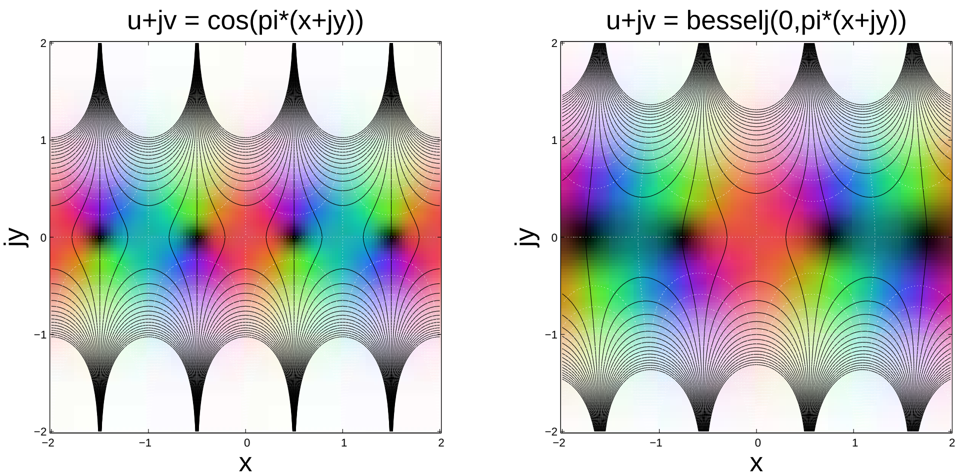


Figure 3.3: *On the left is $w(s) = \cos(\pi z)$ and on the right is the Bessel function $J_0(\pi z)$, which is similar to $\cos(\pi z)$, except the zeros are distorted away from $s = 0$ by a small amount due to the cylindrical geometry. The Bessel function is the solution to the wave equation in cylindrical coordinates while the cos is the solution in rectangular coordinates. The zeros in the function are the places where the pinned boundary condition is satisfied (where the string is restricted, by the boundary, from moving). The Bessel function $J_1(\pi z)$ is similar to $\sin(\pi z)$ in a like manner.*

Two more examples are given in Fig. 3.3 to interpret the two complex mappings $w = \cos(\pi s)$ (left) and the Bessel function $J_0(\pi z)$. Note how the white and black contour lines are always perpendicular where they cross, just as in the calibration plots for the x and y axes, shown in Fig. 1.13 in Section 1.3.8 (p. 55).

Along the σ axis the $\cos(\pi x$ is the periodic with a period of π. The dark spots are at the zeros. at $\pm\pi/2, \pm 3\pi/2, \ldots$. When we stray off the $\omega\jmath = 0$ axis, the function either goes to zero (black) or ∞ (white). This behavior carries the same π periodicity as it has along the $\omega = 0$ line. On the right is the Bessel function $J_0(\pi z)$, which is similar to $\cos(\pi z)$, except the zeros are distorted away from the orgin. These figure are worthy of careful study to develop an intuition for complex functions of complex variables. In Section 1.3.8 we shall explore more complex mappings, and in greater detail.

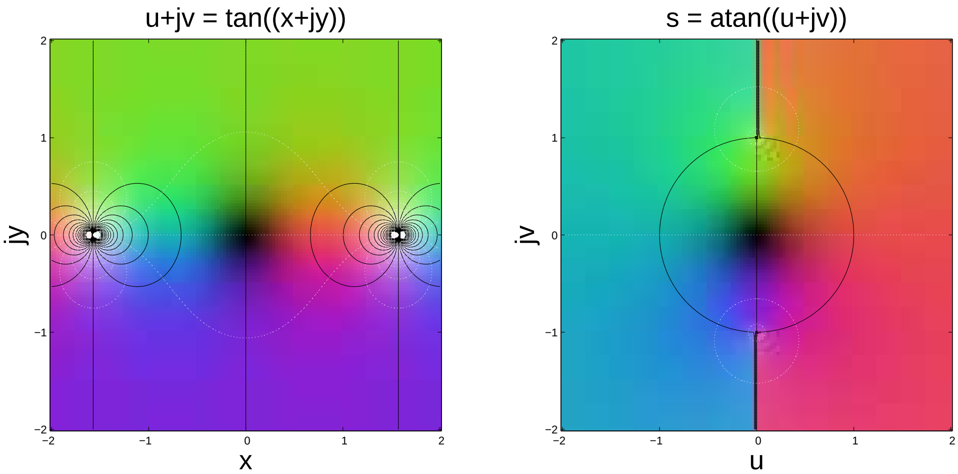


Figure 3.4: *On the left is the function $w(z) = \tan(z)$. On the right is the inverse $s = \tan^{-1}(w)$. Of special interest is* `zviz atan(i*Z) (i/2)*log((1+Z)./(1-Z))`.

In the third example (Fig. 3.4) we show $w = \tan(z)$ and its inverse $z = \tan^{-1}(w)$. The tangent function has zeros where $\sin(z)$ has zeros (e.g., at $z = 0$) and poles where $\cos(z)$ is zero (e.g., at $\pm\pi/2$. The inverse function $s = \mathrm{atan}(w)$ has a zero at $w = 0$ and branch cuts eliminating from $z = \pm\pi$.

It becomes most interesting to study polynomials of degree 5 and 4, with one zero removed, to demonstrate the Fundamental Theorem of Algebra.

3.4.2 Lec 19 Signals and Systems: Fourier vs. Laplace Transforms

Signals and Systems: Fourier vs. Laplace Transforms **AE-3**

3.4.3 Lec 20 Role of Causality and the Laplace Transform

Role of Causality and the Laplace Transform:
$u(t) \leftrightarrow 1/s$ (LT)
$2\tilde{u}(t) \equiv 1 + \mathrm{sgn}(t) \leftrightarrow 2\pi\delta(\omega) + 2/\jmath\omega$ (FT)

3.5 Week 8

3.5.1 Lec 21 The 9 postulates of System of algebraic Networks

Phyical system obey very important rules, that follow from the physics. It is helpful to summarize these physical restrictions in terms of postulates, presented in terms of a taxonomy, or catagorization method, of the fundmental properties of physical systems. Nine of these are listed below. These nine come from a recently published paper (Kim and Allen, 2013). It is possible that given time, others could be added.

A taxonomy of physical systems comes from a systematic summary of the laws of physics, which includes at least the nine basic network postulates, described in Section 1.3.11. To describe each of the network postulates it is helpful to begin with the 2-port transmission (aka ABCD, chain) matrix representation, discussed in Section 3.3.2 (p. 104).

As a specific example we show the 2-port *transmission matrix* for an acoustic transducer (loudspeaker), shown in Fig. 3.5, characterized by the equation

Add figures as sl notebook (p. 35), these postulates.

$$\begin{bmatrix} \Phi_i \\ I_i \end{bmatrix} = \begin{bmatrix} A(s) & B(s) \\ C(s) & D(s) \end{bmatrix} \begin{bmatrix} F_l \\ -U_l \end{bmatrix} = \frac{1}{T} \begin{bmatrix} z_m(s) & z_e(s)z_m(s) + T^2 \\ 1 & z_e(s) \end{bmatrix} \begin{bmatrix} F_l \\ -U_l \end{bmatrix}. \tag{3.21}$$

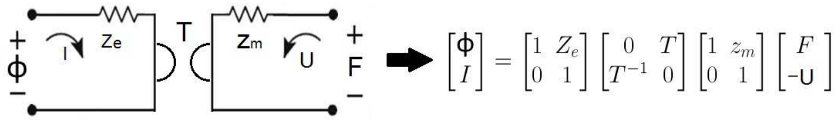


Figure 3.5: *A schematic representation of a 2-port ABCD electro-mechanic system using Hunt parameters $Z_e(s)$, $z_m(s)$, and $T(s)$: electrical impedance, mechanical impedances, and transduction coefficient (Hunt, 1952; Kim and Allen, 2013). Also $V(f)$, $I(f)$, $F(f)$, and $U(f)$ are the frequency domain voltage, current, force, and velocity respectively. Notice how the matrix method 'factors' the 2-port model into three 2×2 matrices. This allows one to separate the physical modeling from the algebra. It is a standard impedance convention that the flows $I(f), U(f)$ are always defined into the port. Thus it is necessary to apply a negative sign on the velocity $-U(f)$ so that it has an outward flow, to feed the next cell with an inward flow. Replace Φ with V.*

This equation comes from the product of the three 2x2 matrices representing each of the three elements of the figure.

This figure represents the motor of the loudspeaker (not the box that it comes in). The system consists of three elements, the electrical input impedance $Z_e(s)$, a gyrator, which is similar to a transformer, but relates current to force, and an output mechanical impedance $Z_m(s)$. This circuit describes what is needed to fully characterize it operation, from electrical input to mechanical (acoustical) output.

The input is electrical (voltage and current) $[\Phi_i, I_i]$ and the output (load) are the mechanical (force and velocity) $[F_l, U_l]$. The first matrix is the general case, expressed in terms of four unspecified functions $A(s), B(s), C(s), D(s)$, while the second matrix is for the specific example of Fig. 3.5. The four entries are the electrical driving point impedance $Z_e(s)$, the mechanical impedance $z_m(s)$ and the transduction $T = B_0 l$ where B_0 is the magnetic flux strength and l is the length of the wire crossing the flux. Since the transmission matrix is anti-reciprocal, its determinate $\Delta_T = -1$, as is easily verified.

Other common transduction examples of cross-modality transduction include current–thermal (thermoelectric effect) and force–voltage (piezoelectric effect). These systems are all reciprocal, thus the transduction has the same sign.

Impedance matrix

These nine postulates describe the properties of a system having an input and an output. For the case of an electromagnetic transducer (Loudspeaker) the system is described by the 2-port, as show in Fig. 3.5. The electrical input impedance of a loudspeaker is $Z_e(s)$, defined by

$$Z_e(s) = \left.\frac{V(\omega)}{I(\omega)}\right|_{U=0}.$$

Note that this *driving-point impedance* must be causal, thus it has a Laplace transform and therefore is a function of the complex frequency $s = \sigma + j\omega$, whereas the Fourier transforms of the voltage $V(\omega)$ and current $I(\omega)$ are functions of the real radian frequency ω, since the time-domain voltage $v(t) \leftrightarrow V(\omega)$ and the current $i(t) \leftrightarrow I(\omega)$ are signals that may start and stop at any time (they are not typically causal).

The corresponding 2-port *impedance matrix* for Fig. 3.5 is

$$\begin{bmatrix}\Phi_i\\ F_l\end{bmatrix} = \begin{bmatrix}z_{11}(s) & z_{12}(s)\\ z_{21}(s) & z_{22}(s)\end{bmatrix}\begin{bmatrix}I_i\\ U_l\end{bmatrix} = \begin{bmatrix}Z_e(s) & -T(s)\\ T(s) & z_m(s)\end{bmatrix}\begin{bmatrix}I_i\\ U_l\end{bmatrix}. \tag{3.22}$$

Such a description allows one to define *Thèvenin parameters*, a very useful concept used widely in circuit analysis and other network models from other modalities.

The impedance matrix is an alternative description of the system, but with generalized forces $[\Phi_i, F_l]$ on the left and generalized flows $[I_i, U_l]$ on the right. A rearrangement of the equations allows one to go from the ABCD to impedance set of parameters (Van Valkenburg, 1964b). The electromagnetic transducer is anti-reciprocal (P6), $z_{12} = -z_{21} = T = B_0 l$.

Additional or modified postulates

The postulates must go beyond postulates P1-P6 defined by Carlin and Giordano (Section 1.3.11, p. 61), when there are interaction of waves and a structured medium, along with other properties not covered by classic network theory. Assuming QS, the wavelength must be large relative to the medium's lattice constants. Thus the QS property must be extended to three dimensions, and possibly to the cases of an-isotropic and random media.

Causality: P1 As stated above, due to causality the negative properties (e.g., negative refractive index) must be limited in bandwidth, as a result of the Cauchy-Riemann conditions. However even causality needs to be extended to include the delay, as quantified by the d'Alembert solution to the wave equation, which means that the delay is proportional to the distance. Thus we generalize P1 to include the space dependent delay. When we wish to discuss this property we denote it *Einstein causality*, which says that the delay must be proportional to the distance x, with impulse response $\delta(t - x/c)$.

Linearity: P2 The wave properties of may be non-linear. This is not restrictive as most physical systems are naturally nonlinear. For example, a capacitor is inherently nonlinear: as the charge builds up on the plates of the capacitor, a stress is applied to the intermediate dielectric due to the electrostatic force $F = qE$. In a similar manner, an inductor is nonlinear. Two wires carrying a current are attracted or repelled, due to the force created by the flux. The net force is the product of the two fluxes due to each current.

In summary, most physical systems are naturally nonlinear, it's simply a matter of degree. An important counter example is a amplifier with negative feedback, with very large open-loop gain. There are, therefore, many types of non-linear, instantaneous and those with memory (e.g., hysteresis). Given the nature of P1, even an instantaneous non-linearity may be ruled out. The linear model is so critical for our analysis, providing fundamental understanding that we frequently take this postulate for granted.

Passive/Active: P3 This postulate is about conservation of energy and Otto Brune's *positive Real* (PR aka physically realizable) condition, that every passive impedance must obey. Following up on the earlier work of his primary PhD thesis advisor Wilhelm Cauer (1900-1945) and Ernst Guillemin, along with Norbert Weiner and Vannevar Bush at MIT, Otto Brune mathematically characterized the properties of every PR 1-port driving point impedance (Brune, 1931a).

When the input resistance of the impedance is real, the system is said to be *passive*, which means the system obeys conservation of energy. The real part of $Z(s)$ is positive if and only if the corresponding reflectance is less than 1 in magnitude. The definition of the reflectance of $Z(s)$ is defined as a bilinear transformation of the *impedance*, normalized by its *surge resistance* r_0 (Campbell, 1903)

$$\Gamma(s) = \frac{Z(s) - r0}{Z(s) + r0}.$$

The surge resistance is defined in terms of the inverse Laplace transform of $Z(s) \leftrightarrow z(t)$, which must have the form

$$z(t) = r_0\delta(t) + \zeta(t),$$

where $\zeta(t) = 0$ for $t < 0$. It naturally follows that $\gamma(t) \leftrightarrow \Gamma(s)$ is zero for negative and zero time, namely $\gamma(0) = 0, t \leq 0$. at

Given any linear PR impedance $Z(s) = R(\sigma, \omega) + jX(\sigma, \omega)$, having real part $R(\sigma, \omega)$ and imaginary part $X(\sigma, \omega)$, the impedance is defined as being PR (Brune, 1931a) if and only if

$$R(\sigma \geq 0, \omega) \geq 0. \tag{3.23}$$

That is, the real part of any PR impedance is non-negative everywhere in the right half s plane ($\sigma \ge 0$). This is a very strong condition on the complex analytic function $Z(s)$ of a complex variable s. This condition is equivalent to any of the following statements: 1) There are no poles or zeros in the right half plane ($Z(s)$ may have poles and zeros on the $\sigma = 0$ axis). 2) If $Z(s)$ is PR then its reciprocal $Y(s) = 1/Z(s)$ is PR (the poles and zeros swap). 3) If the impedance may be written as the ratio of two polynomials (a limited case, related to the quasistatics approximation, P9) having degrees N and L, then $|N - L| \le 1$. 4) The angle of the impedance $\theta \equiv \angle Z$ lies between $[-\pi \le \theta \le \pi]$. 5) The impedance *and its reciprocal* are *complex analytic* in the right half plane, thus they each obey the Cauchy Riemann conditions there.

The PR (positive real or Physically realizable) condition assures that every impedance is *positive-definite* (PD), thus guaranteeing conservation of energy is obeyed (Schwinger and Saxon, 1968, p.17). This means that the total energy absorbed by any PR impedance must remain positive for all time, namely

$$\mathcal{E}(t) = \int_{-\infty}^{t} v(t)i(t)\, dt = \int_{-\infty}^{t} i(t) \star z(t)\, i(t)\, dt > 0,$$

where $i(t)$ is *any* current, $v(t) = z(t) \star i(t)$ is the corresponding voltage and $z(t)$ is the real causal impulse response of the impedance, e.g., $z(t) \leftrightarrow Z(s)$ are a Laplace Transform pair. In summary, if $Z(s)$ is PR, $\mathcal{E}(t)$ is PD.

As discussed in detail by Van Valkenburg, any rational PR impedance can be represented as a *rational polynomial fraction expansion* (residue expansion), which can be expanded into first-order poles as

$$Z(s) = K \frac{\Pi_{i=1}^{L}(s - n_i)}{\Pi_{k=1}^{N}(s - d_k)} = \sum_n \frac{\rho_n}{s - s_n} e^{j(\theta_n - \theta_d)}, \tag{3.24}$$

where ρ_n is a complex scale factor (residue). Every pole in a PR function has only simple poles *and* zeros, requiring that $|L - N| \le 1$ (Van Valkenburg, 1964b).

Whereas the PD property clearly follows P3 (conservation of energy), the physics is not so clear. Specifically what is the physical meaning of the specific constraints on $Z(s)$? In many ways, the impedance concept is highly artificial, as expressed by P1-P7.

When the impedance is not rational, special care must be taken. An example of this is the semi-inductor $M\sqrt{s}$ and semi-capacitor $K/\sqrt{s}$ due, for example, to the *skin effect* in EM theory and viscous and thermal losses in acoustics, both of which are frequency dependent boundary-layer diffusion losses. They remain positive-real but have a branch cut, thus are double valued in frequency.

By building in the physics behind conservation of energy: Otto Brune's *positive-real* (PR) condition. Following up on the earlier work of his primary PhD thesis advisor Wilhelm Cauer (1900-1945), and working with Norbert Weiner and Vannevar Bush at MIT, Otto Brune mathematically characterized the properties of every PR 1-port driving point impedance (Brune, 1931a).

Given any linear PR impedance $Z(s) = R(\sigma, \omega) + jX(\sigma, \omega)$, having real part (resistance) $R(\sigma, \omega)$ and imaginary part (reactance) $X(\sigma, \omega)$, the impedance is defined as being PR (Brune, 1931b) if and only if

$$R(\sigma \ge 0, \omega) \ge 0. \tag{3.25}$$

That is, the real part of any PR impedance is non-negative everywhere in the right half s plane ($\sigma \ge 0$). This is a very strong condition on the complex analytic function $Z(s)$ of a complex variable s. This condition is equivalent to any of the following statements: 1) There are no poles or zeros in the right half plane ($Z(s)$ may have poles and zeros on the $\sigma = 0$ axis). 2) If $Z(s)$ is PR then its reciprocal $Y(s) = 1/Z(s)$ is PR (the poles and zeros swap). 3) If the impedance may be written as the ratio of two polynomials (a limited case) having degrees N and L, then $|N - L| \le 1$. 4) The angle of the impedance $\theta \equiv \angle Z$ lies between $[-\pi \le \theta \le \pi]$. 5) The impedance *and its reciprocal* are *complex analytic* in the right half plane, thus they each obey the Cauchy Riemann conditions.

The PR (positive real or Physically realizable) condition assures that every impedance is *positive-definite* (PD), thus guaranteeing conservation of energy is obeyed (Schwinger and Saxon, 1968, p.17).

This means that the total energy absorbed by any PR impedance must remain positive for all time, namely

$$\mathcal{E}(t) = \int_{-\infty}^{t} v(t)i(t)\, dt = \int_{-\infty}^{t} i(t)\star z(t)\; i(t)\, dt > 0,$$

where $i(t)$ is *any* current, $v(t) = z(t) \star i(t)$ is the corresponding voltage and $z(t)$ is the real causal impulse response of the impedance, e.g., $z(t) \leftrightarrow Z(s)$ are a Laplace Transform pair. In summary, if $Z(s)$ is PR, $\mathcal{E}(t)$ is PD.

As discussed in detail by Van Valkenburg, any rational PR impedance can be represented as a *rational polynomial fraction expansion* (residue expansion), which can be expanded into first-order poles as

$$Z(s) = K\frac{\Pi_{i=1}^{L}(s-n_i)}{\Pi_{k=1}^{N}(s-d_k)} = \sum_n \frac{\rho_n}{s-s_n} e^{j(\theta_n-\theta_d)}, \tag{3.26}$$

where ρ_n is a complex scale factor (residue). Every pole in a PR function has only simple poles *and* zeros, requiring that $|L-N| \le 1$ (Van Valkenburg, 1964a).

Whereas the PD property clearly follows P3 (conservation of energy), the physics is not so clear. Specifically what is the physical meaning of the specific constraints on $Z(s)$? In many ways, the impedance concept is highly artificial, as expressed by P1-P7.

When the impedance is not rational, special care must be taken. An example of this is the semi-inductor $M\sqrt{s}$ and semi-capacitor $K/\sqrt{s}$ due, for example, to the *skin effect* in EM theory and viscous and thermal losses in acoustics, both of which are frequency dependent boundary-layer diffusion losses. They remain positive-real but have a branch cut, thus are double valued in frequency.

Real time response: P4 The impulse response of every physical system is real, vs. complex. This requires that the Laplace Transform have conjugate-symmetric symmetry $H(s) = H^*(s^*)$, where the $*$ indicates conjugation (e.g., $R(\sigma,\omega) + X(\sigma,\omega) = R(\sigma,\omega) - X(\sigma,-\omega)$).

Time invariant: P5 The meaning of *time-invariant* requires that the impulse response of a system does not change over time. This requires that the system coefficients of the differential equation describing the system are constant (independent of time).

Rayleigh Reciprocity: P6 Reciprocity is defined in terms of the unloaded output voltage that results from an input current. Specifically

$$\begin{bmatrix} z_{11}(s) & z_{12}(s) \\ z_{21}(s) & z_{22}(s) \end{bmatrix} = \frac{1}{C(s)} \begin{bmatrix} A(s) & \Delta_T \\ 1 & D(s) \end{bmatrix}, \tag{3.27}$$

where $\Delta_T = A(s)D(s) - B(s)C(s) = \pm 1$ for the reciprocal and anti-reciprocal systems respectively. This is best understood in term of Eq. 3.22. The off-diagonal coefficients $z_{12}(s)$ and $z_{21}(s)$ are defined as

$$z_{12}(s) = \left.\frac{\Phi_i}{U_l}\right|_{I_i=0} \qquad\qquad z_{21}(s) = \left.\frac{F_l}{I_i}\right|_{U_l=0}$$

The these off-diagonal elements are equal $[z_{12}(s) = z_{21}(s)]$ the system is said to obey *Rayleigh reciprocity*. If they are opposite in sign $[z_{12}(s) = -z_{21}(s)]$, the system is said to be *anti-reciprocal*. If a network has neither of the reciprocal or anti-reciprocal characteristics, then we denote it as *non-reciprocal* (McMillan, 1946). The most comprehensive discussion of reciprocity, even to this day, is that of Rayleigh (1896, Vol. I). The reciprocal case may be modeled as an ideal transformer (Van Valkenburg, 1964a) while for the anti-reciprocal case the generalized force and flow are swapped across the 2-port. An electromagnetic transducer (e.g., a moving coil loudspeaker or electrical motor) is anti-reciprocal (Kim and Allen, 2013; Beranek and Mellow, 2012), requiring a gyrator rather than a transformer, as shown in Fig. 3.5.

Reversibility: P7 A second 2-port property is the *reversible/non-reversible* postulate. A reversible system is invariant to the input and output impedances being swapped. This property is defined by the input and output impedances being equal.

Referring to Eq. 3.27, when the system is *reversible* $z_{11}(s) = z_{22}(s)$ or in terms of the transmission matrix variables $\frac{A(s)}{C(s)} = \frac{D(s)}{C(s)}$ or simply $A(s) = D(s)$ assuming $C(s) \neq 0$.

An example of a non-reversible system is a transformer where the turns ratio is not one. Also an ideal operational amplifier (when the power is turned on) is non-reversible due to the large impedance difference between the input and output. Furthermore it is *active* (it has a power gain, due to the current gain at constant voltage) (Van Valkenburg, 1964b).

Generalizations of this lead to group theory, and *Noether's theorem*. These generalizations apply to systems with many modes whereas quasistatics holds when operate below a cutoff frequency (Table 3.1), meaning that like the case of the transmission line, there are no propagating transverse modes. While this assumption is never exact, it leads to highly accurate results because the non-propagating evanescent transverse modes are attenuated over a short distance, and thus, in practice, may be ignored (Montgomery et al., 1948; Schwinger and Saxon, 1968, Chap. 9-11).

We extend the Carlin and Giordano postulate set to include (P7) Reversibility, which was refined by Van Valkenburg (1964a). To satisfy the reversibility condition, the diagonal components in a system's impedance matrix must be equal. In other words, the input force and the flow are proportional to the output force and flow, respectively (i.e., $Z_e = z_m$).

Spatial invariant: P8 The characteristic impedance and wave number $\kappa(s, x)$ may be strongly frequency and/or spatially dependent, or even be negative over some limited frequency ranges. Due to *causality*, the concept of a negative group velocity must be restricted to a limited bandwidth (Brillouin, 1960). As is made clear by Einstein's theory of relativity, all materials must be strictly causal (P1), a view that must therefore apply to acoustics, but at a very different time scale. We first discuss generalized postulates, expanding on those of Carlin and Giordano.

The QS constraint: P9 When a system is described by the wave equation, delay is introduced between two points in space, which depends on the wave speed. When the wavelength is large compared to the delay, one may successfully apply the *quasistatic approximation*. This method has wide-spread application, and is frequency used without mention of the assumption. This can lead to confusion, since the limitations of the approximation may not be appreciated. An example is the use of QS in Quantum Mechanics. The QS approximation has wide spread use when the signals may be accurately approximated by a band-limited signal. Examples include KCL, KVL, wave guides, transmission lines, and most importantly, impedance. The QS property is not mentioned in the six postulates of Carlin and Giordano (1964), thus they need to be extended in some fundamental ways.

When the dimensions of a cellular structure in the material are much less than the wavelength, can the QS approximation be valid. This effect can be viewed as a *mode filter* that suppresses unwanted (or conversely enhances the desired) modes (Ramo et al., 1965). QSs may be applied to a 3 dimensional specification, as in a semiconductor lattice. But such applications fall outside the scope of this text (Schwinger and Saxon, 1968).

Although I have never seen the point discussed in the literature, the QS approximation is applied when defining Green's theorem. For example, Gauss's Law is not true when the volume of the container violates QS, since changes in the distribution of the charge have not reached the boundary, when doing the integral. Thus such integral relationships assume that the system is in quasi steady-state (i.e., that QS holds).

Formally, QS is defined as $ka < 1$ where $k = 2\pi/\lambda = \omega/c$ and a is the cellular dimension or the size of the object (k and a can be vectors, e.g., Eq. 1.38, p. 66, Section 1.4.1). Schelkunoff may have been the first to formalize this concept (Schelkunoff, 1943) (but not the first to use it, as exemplified by the Helmholtz resonator). George Ashley Campbell was the first to use the concept in the important application of a wave-filter, some 30 years before Schelkunoff (Campbell, 1903). These two men were 40

Table 3.1: *There are several ways of indicating the quasi-static (QS) approximation. For network theory there is only one lattice constant a, which must be much less than the wavelength (wavelength constraint). These three constraints are not equivalent when the object may be a larger structured medium, spanning many wavelengths, but with a cell structure size much less than the wavelength. For example, each cell could be a Helmholtz resonator, or an electromagnetic transducer (i.e., an earphone).*

Measure	Domain
$ka < 1$	Wavenumber constraint
$\lambda > 2\pi a$	Wavelength constraint
$f_c < c/2\pi a$	Bandwidth constraint

years apart, and both worked for the telephone company (after 1929, called AT&T Bell Labs) (Fagen, 1975).

There are alternative definitions of the QS approximation, depending on the geometrical cell structure. The alternatives are outlined in Table 3.1.

Summary

A transducer converts between modalities. We propose the general definition of the nine system postulates, that include all transduction modalities, such as electrical, mechanical, and acoustical. It is necessary to generalize the concept of the QS approximation (P9) to allow for guided waves.

Given the combination of the important QS approximation, along with these space-time, linearity, and reciprocity properties, a rigorous definition and characterization a system can thus be established. It is based on a taxonomy of such materials, formulated in terms of material and physical properties and in terms of extended network postulates.

3.5.2 Lec 22 Exam II (Evening)

Chapter 4

Ordinary Differential Equations: Stream 3a

WEEK 8 **23.9.0**

Week 8 Friday *Stream 3*

L 23 The amazing Bernoulli family; Fluid mechanics; airplane wings; natural logarithms
The transition from geometry → algebra → algebreic geometry → real analytic → complex analytic
From Bernoulii to Euler to Cauchy and Riemann

4.1 Week 8

4.1.1 Lec 23 Newton and early calculus & the Bernoulli Family

Newton and Calculus
Bernoulli family
Euler standard periodic (circular) function package

The period of analytic discovery:
Coming out of the dark ages, from algebra, to analytic geometry, to calculus.
Starting with real analytic functions by Euler, we move to complex analytic functions with Cauchy.
Integration in the complex plane is finally conquered.

Lect DE 25.9 Stream 3: ∞ and Sets **25.9.1**

The development of real representations proceeded at a deadly-slow pace:

- *Real numbers* $\mathbb{R}$*:* Pythagoras knew of irrational numbers ($\sqrt{2}$)
- *Complex numbers* $\mathbb{C}$*:* 1572 *"Bombelli is regarded as the inventor of complex numbers . . . "* http://www-history.mcs.st-andrews.ac.uk/Biographies/Bombelli.html http://en.wikipedia.org/wiki/Rafael_Bombelli & p. 258
- *Power Series*: Gregory-Newton interpolation formula c1670, p. 175
- Point at infinity and the Riemann sphere 1851
- Analytic functions p. 267 c1800; Impedance $Z(s)$ 1893

Stream 3 Infinity

- Infinity ∞ was not "understood" until 19^{th} CE
- ∞ is best defined in terms of a limit
- Limits are critical when defining calculus
- Set theory is the key to understanding Limits
- 2612 Open vs close sets determine when a limit exists (or not)
- Thus, to fully understand limits, one needs to understand set theory
- Related is the convergence of a series
- Every convergent series has a *Region of Convergence* (ROC)
- When the ROC is Complex:
  - Example of $\frac{1}{1-x}$ vs. $\frac{1}{i-x}$: The ROC is 1 for both cases
  - Why?
  - The case of the Heaviside step function $u(t)$ & the Fourier Transform

Irrational numbers and limits (Ch. 4)

- How are irrational numbers interleaved with the integers?
- Between n and $2n$ there is always an irrational number:

> 2613 Chebyshev said, and I say it again. There is always a prime between n and 2n. -p. 585_2

- *Prime number theorem*: The number of of primes is approximately(the density of primes is $\rho_\pi(n) \propto 1/\ln(n)$.
- The number of primes less than n is n times the density, or

$$N(n) = n/\ln(n).$$

  - The formula for *entropy* is $\mathcal{H} = -\sum_n p_n \log p_n$.
    Could there be some hidden relationship lurking here?

Stream 3: ∞ and Sets **25.9.2**

- 2614 Understanding ∞ has been a primary goals since Euclid
- The Riemann sphere solves this fundamental problem
- The point at ∞ simply "another point" on the Riemann sphere

Open vs. closed sets

Influence of open vs. closed set **7.3.6**

- Important example: LT vs. FT step function: Dirac step vs Fourier step:
- $u(t) \leftrightarrow \frac{1}{s}$ vs. $\tilde{u}(t) \leftrightarrow \pi\delta(\omega) + \frac{1}{\jmath\omega}$

WEEK 9 **23.9.0**

Week 9 Monday

L 24 Power series and integration of functions (ROC)
Fundamental Theorem of calculus (Leibniz theorm of integration)
$1/(1-x) = \sum_{k=0}^{\infty} x^k$ with $x \in \mathbb{R}$

L 25 Integration in the complex plane: Three theorems
Integration of $1/s$ on the unit circle, and on a unit circle centered about $s = 1 + i$.

L 26 Cauchy-Riemann conditions
Real and imaginary parts of analytic functions obey Laplace's equation.
Infinite power Series and analytic function theory; ROC

4.2 Week 9

4.2.1 Lec 24 Power series and complex analytic functions

L 24: Power series and complex analytic function

4.2.2 Lec 25 Integration in the complex plane

L 25: Integration in the complex plane; Infinite power Series and analytic function theory; ROC
Real and imaginary parts of analytic functions obey Laplace's equation.
Colorized plots of analytic functions. How to read the plots and what they tell us?

4.2.3 Lec 26 Cauchy Riemann conditions: Complex-analytic functions

L 26: Cauchy Riemann conditions: Complex-analytic functions

WEEK 10 **26.10.0**

L 27 $Z(s) = R(s) + \jmath X(s)$: real and imag parts obey Laplace]s Equation
Basic equations of mathematical Physics: Wave equation, Diffusion equation, Laplace's Equation
Motivation: Dispersion relation for the wave equation $\kappa \cdot \kappa = s^2/c_0^2$
Examples.

L 28 Three Fundamental theorems of complex integral calculus
$\int_0^z = F(\zeta)d\zeta = F(z) - F(0)$: $dZ(s)/ds$ independent of direction
Inverse Laplace transform
Examples.

L 29 Inverse Laplace transform: Poles and Residue expansions;
Application of the Fundamental Thm of Complex Calculus
The Inverse Laplace Transform (ILT); poles and the Residue expansion: The case for causality
ROC as a function of the sign of time in e^{st} (How does causality come into play?)
Examples.

4.3 Integration and differentiation in the complex plane

4.3.1 Lec 27 Differentiation in the complex plane

L 27: Differentiation in the complex plane: CR conditions?
Motivation: Inverse Laplace transform
ROC in the complex plane
Basic equations of mathematical Physics: Wave equation, Diffusion equation, Laplace's Equation
Motivation: Dispersion relation for the wave equation $\kappa \cdot \kappa = s^2/c_0^2$

4.3.2 Lec 28 Three complex Integral Theorems

L 28: Integration in the complex plane: Basic definitions of Three theorems
Integration of $1/s$ on the unit circle, and on a unit circle centered about $s = 1 + i$.

Moved from Lec 3 (page 31)

Set Theory: Set theory is a topic that can be inadequately addressed in the undergraduate Engineering and Physics curriculum, and is relatively young to mathematics. The set that a number is drawn from is crucially important when taking limits.

4.3.3 Lec 29 Inverse Laplace Transform

L 29: Inverse Laplace transform: Poles and Residue expansions;
Application of the Fundamental Thm of Complex Calculus
Examples.

Stream 3: Infinity and irrational numbers Ch 4 — 2.1.6

- *Limit points, open vs. closed sets are fundamental to modern mathematics*
- *These ideas first appeared with the discovery of $\sqrt{2}$, and $\sqrt{n}$* https://en.wikipedia.org/wiki/Spiral_of_Theodorus
  and related constructions (factoring the square, Pell's Eq. *p. 44**)*

Infinity and irrational $\mathbb{Q}$ numbers

WEEK 11 — 30.11.0

The fundamental theorem of calculus **2.1.7**

Let $A(x)$ be the area under $f(x)$. Then

$$\begin{aligned}\frac{d}{dx}A(x) &= \frac{d}{dx}\int^x f(\eta)d\eta \\ &= \lim_{\delta\to 0}\frac{A(x+\delta)-A(x)}{\delta}\end{aligned}$$

and/or

$$A(b)-A(a)=\int_a^b f(\eta)d\eta$$

- *Stream 3 is about limits*
- *Integration and differentiation (Calculus) depend on limits*
- *Limits are built on open vs. closed sets*

L 30 Inverse Laplace transform & Cauchy Residue Theorem

L 31 Case for causality Closing the contour as $s\to\infty$; Role of $\Re st$
DE-3

L 32 Properties of the LT:
1) Modulation, 2) Translation, 3) convolution, 4) periodic functions
Tables of common LTs

4.4 Integration in the complex plane

4.4.1 Lec 30 Inverse Laplace Transform & Cauchy residue theorem

L30: The Inverse Laplace Transform (ILT); poles and the Residue expansion: The case for causality
ROC as a function of the sign of time in e^{st} (How does causality come into play?)

4.4.2 Lec 31 The case for causality

L31: Closing the contour as $s\to\infty$; Role of $\Re st$

4.4.3 Lec 32 Laplace transform properties: Modulation, time translation, etc.

L32: Detailed examples of the Inverse LT:
1) Modulation, 2) Translation, 3) convolution, 4) periodic functions
Tables of common LTs

WEEK 12 **33.12.0**

L 33 Multi-valued functions in the complex plane; Branch cuts
The extended complex plane (regularization at ∞) (Stillwell, 2010, p. 280)
Complex analytic functions of Genus 1 (Stillwell, 2010, p. 343)

L 34 Euler's vs. Riemann's Zeta function $\zeta(s)$: Poles at the primes
colorized plot of $\zeta(s)$
??Sterling's formula??

L 35 Exam III

4.5 Complex plane concepts

4.5.1 Lec 33 Multi-valued complex functions, Branch Cuts, Extended plane

L33: Multi-valued functions in the complex plane; Branch cuts
The extended complex plane (regularization at ∞) (Stillwell, 2010, p. 280)
Complex analytic functions of Genus 1 (Stillwell, 2010, p. 343)

4.5.2 Lec 34 The Riemann Zeta function $\zeta(s)$

L34: Euler's vs. Riemann's Zeta function $\zeta(s)$: Poles at the primes
domain coloring of $\zeta(s)$
??Sterling's formula??

Table 4.1: Physical meaning of each factor of $\zeta(s)$ **4.2.7**

- *Series expansion*

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad \textit{ROC: } |x| < 1$$

- *If time T is a positive delay, then from the* Laplace transform

$$\delta(t-T) \leftrightarrow \int_0^{\infty} \delta(t-T)e^{st}dt = e^{-sT}$$

- *Each factor of $\zeta(s)$ is an ∞ sum of delays*
- *For example for $\pi_1 = 2$, ($T = \ln(2)$, thus $2^{-2} = e^{-s\ln 2}$)*

$$\sum_n \delta(t-nT) \leftrightarrow \frac{1}{1-2^{-s}} = 1 + e^{-sT} + e^{-s2T} + \cdots$$

Table 4.1: Each prime number defines a delay $T_k = \ln(\pi_k)$, which in turn defines a pole in the complex s plane. The series expansion of this pole is a train of delta functions that are one-sided periodic in the delta T. Thus each factor in the $\zeta(s)$ function defines a pole, having an incommensurate delay, since each pole is defined by a unique prime. Following this simple logic, we may interpret $\zeta(s)$ as being the Laplace transform of $Zeta(t)$, the cascade of quasi-periodic impulse responses, each with a recursive delay, determined by a prime. Note that $48100 = 10 \cdot (2 \cdot 5 \cdot 13 \cdot 37)$ is the sampling frequency [Hz] of modern CD players. This corresponds to the 20^{th} harmonic of the US line frequency (60 [Hz]).[b]

[a] since $gcd(48100, 60) = 20$ and $gcd(48100, 50) = 50$.
[b] since $gcd(48100, 60) = 20$ and $gcd(48100, 50) = 50$.

Riemann Zeta Function $\zeta(s)$

This very important analytic function is the credible argument for true deeper understanding of the power to the analytic function. Just like the Pythagorean theorem is important to all mathematics, the zeta function is important to analysis, with many streams of analysis emanating from this form. For example the analytic Gamma function $\Gamma(s)$ is a generalization of the factorial by the relationship

$$n! = \Gamma(s-1).$$

Another important relationship is

$$\sum_{k=n}^{\infty} k = nu_n = u_n \star u_n$$

where the $\star$ represents convolution. If this is treated in the frequency domain the we obtain z-transforms of a very simple second-order pole[1]

$$nu_n \leftrightarrow \frac{2}{(z-1)^2}.$$

This follows from the geometric series

$$\frac{1}{1-z} = \sum_n z^n$$

with $z = e^s$, and the definition of convolution.

The Laplace transform does not require that the series converge, rather that the series have a region of convergence that is properly specified. Thus the non-convergent series nu_n is perfectly well defined, just like

$$tu(t) = u(t) \star u(t) \leftrightarrow \frac{1!}{s^2}$$

is well defined, in the Laplace transform sense. More generally

$$t^n u(t) \leftrightarrow \frac{n!}{s^{n+1}}.$$

From this easily understood relationship we can begin to understand $\Gamma(s)$, as the analytic extension of the factorial. Its definition is simply related to the inverse Laplace transform, which is an integral. But to go there we must be able to think in the complex frequency domain. In fact we have the very simple definition for $\Gamma(p)$ with $p \in \mathbb{C}$

$$t^{p-1}u(t) \leftrightarrow \frac{\Gamma(p)}{s^p}$$

which totally explains $\Gamma(p)$. Thinking in the time domain is crucial for my understanding.

An example is a digital filter, which is linear. Such a system is shown in Fig. 4.3, where the two functions are second order digital filters. The input signal $x[n]$ enters from the left, is filtered by the first filter, producing output $y[n]$. This is then filtered again by the filter in the second box to produce signal $z[n]$. For this simple case of two linear filters the operation *commute*.

4.5.3 Lec 35 Exam III

L 35: Exam III

Thanksgiving Holiday 11/19–11/27 2016

[1]Need to verify the exact form of these relationships, not work from memory

Riemann Zeta Function $\zeta(s)$ **4.2.5**

- *Integers appear as the "roots" (aka eigenmodes) of* $\zeta(s)$
- *Basic properties* $(s = \sigma + i\omega)$

$$\zeta(s) \equiv \sum_{1}^{\infty} \frac{1}{n^s} \qquad\qquad \sigma = \Re(s) > 0$$

  - *What is the* region of convergence (ROC)?
- *The amazing* Euler-Riemann Product formula *(Stillwell, 2010, Sect. 10.7:)*

$$\zeta(s) = \prod_k \frac{1}{1-\pi_k{}^{-s}} = \prod_k \frac{1}{1-\left(\frac{1}{\pi_k}\right)^s} = \prod_k \frac{1}{1-\frac{1}{\pi_k^s}}$$
$$= \frac{1}{1-2^{-s}} \cdot \frac{1}{1-3^{-s}} \cdot \frac{1}{1-5^{-s}} \cdot \frac{1}{1-7^{-s}} \cdots \frac{1}{1-\pi_n^{-s}} \cdots$$

- *Euler c1750 assumed* $s \subset \mathbb{R}$*. Riemann c1850 extended* $s \subset \mathbb{C}$

Figure 4.1: *The zeta function arguably the most important of the special functions of analysis because it connects the primes to analytic function theory in a fundamental way.*

Plot of $\angle\zeta(s)$ **4.2.6**

Angle of Riemann Zeta function $\angle\zeta(z)$ *as a function of complex* z

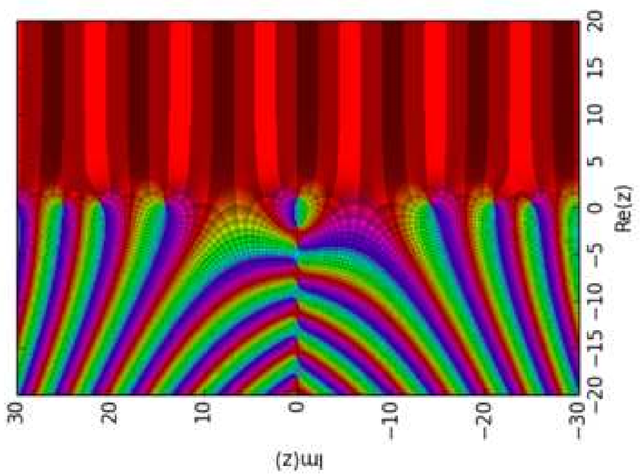


Figure 4.2: $\angle\zeta(z)$: Red $\Rightarrow \angle\zeta(z) < \pm\pi/2$

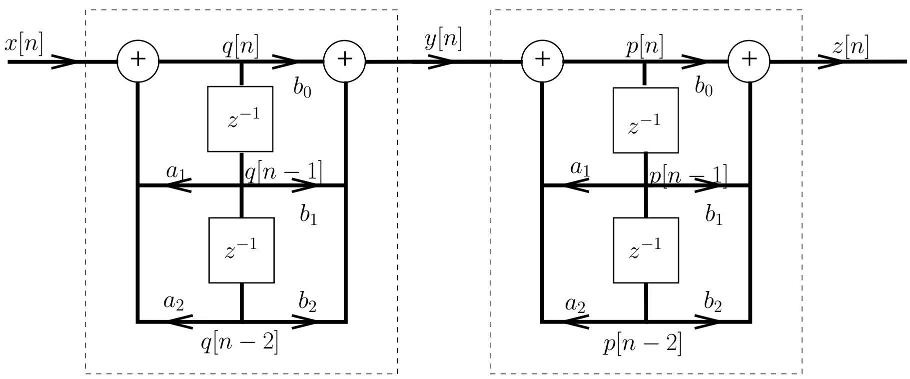


Figure 4.3: *Example of a signal flow diagram for the composition of signals* $z = g \circ f\ (x)$ *with* $y = f(x)$ *and* $z = g(y)$.

Chapter 5

Vector Calculus: Stream 3b

WEEK 13 36.13.0

L 36 Scaler wave equations and the Webster Horn equation; WKB method
A real-world example of large delay, where the branch-cut placement is critical

L 37 Partial differential equations of Physics
Scaler wave equation and its solution in 1 and 3 Dimensions
VC-1

L 38 Vector dot and cross products $A \cdot B, A \times B$
Gradient, divergence and curl

– Thanksgiving Holiday 11/19–11/27 2016

5.1 Stream 3b

5.1.1 Lec 36 Scalar Wave equation

5.1.2 Lec 37 Partial differential equations of physics

Scalar wave equations and the Webster Horn equation; WKB method
Example of a large delay, where a branch-cut placement is critical (i.e., phase unwrapping)
L 37: Partial differential equations of Physics
Scalar wave equation and its solution in 1 and 3 Dimensions

5.1.3 Lec 38 Gradient, divergence and curl vector operators

L 38: Vector dot and cross products $A \cdot B, A \times B$
Gradient, divergence and curl vector operators

WEEK 14 37.14.0

L 39 Gradient, divergence and curl: Gauss's (divergence) and Stokes's (curl) theorems

L 40 J.C. Maxwell unifies Electricity and Magnetism with the formula for the speed of light
Basic definitions of E, H, B, D
O. Heaviside's (1884) vector form of Maxwell's EM equations and the *vector wave equation*
How a loud-speaker works

L 41 *The Fundamental Thm of vector calculus*
Incompressable and *Irrotational* fluids and the two defining vector identities

5.2 Thanksgiving Holiday 11/19–11/27 2016

Thanksgiving Vacation: 1 week of rest

5.3 Vector Calculus

5.3.1 Lec 39 Geometry of Gradient, divergence and curl vector operators

Geometry of Gradient, divergence and curl vector operators

Lec 39: Review of vector field calculus **39.14.2**

- Review of last few lectures: Basic definitions
  - *Field*: i.e., Scalar & vector fields are functions of more than one variable
  - "Del:" $\nabla \equiv [\partial_x, \partial_y, \partial_z]^T$
  - Gradient: $\nabla\phi(x, y, z) \equiv [\partial_x\phi,\ \partial_y\phi,\ \partial_z\phi]^T$
- *Helmholtz Theorem:*
  Every vector field $\boldsymbol{V}(x, y, z)$ may be uniquely decomposed into *compressible* & *rotational* parts
  $$\boldsymbol{V}(x, y, z) = -\nabla\phi(x, y, z) + \nabla\times\boldsymbol{A}(x, y, z)$$
- Scalar part $\nabla\phi$ is *compressible* ($\nabla\phi = 0$ is *incompressible*)
- Vector part $\nabla\times\mathbf{A}$ is *rotational* ($\nabla\times\boldsymbol{A} = 0$ is *irrotational*)
- Key vector identities: $\nabla\times\nabla\phi = 0$; $\nabla\cdot\nabla\times\mathbf{A} = 0$
- Definitions of Divergence, Curl & Maxwell's Eqs;
- Closure: Fundamental Theorems of Integral Calculus

Name	Input	Output	Operator
Gradient	Scalar	**Vector**	$-\nabla()$
Divergence	**Vector**	Scalar	$\nabla\cdot()$
Curl	**Vector**	**Vector**	$\nabla\times()$

Table 5.1: *The three vector operators manipulate scalar and vector fields, as indicated here. The gradient converts scalar fields into vector fields. The divergence eats vector fields and outputs scalar fields. Finally the curl takes vector fields into vector fields.*

2760 **Gradient**

Gradient: $\boldsymbol{E} = \nabla\phi(x,y,z)$ **39.14.3**

- Definition: $\mathbb{R}^1 \underset{\nabla}{\mapsto} \mathbb{R}^3$

$$\mathbf{E}(x,y,z) = [\partial_x,\, \partial_y,\, \partial_z]^T\, \phi(x,y,z) = \left[\frac{\partial\phi}{\partial x},\, \frac{\partial\phi}{\partial y},\, \frac{\partial\phi}{\partial z}\right]^T_{x,y,z}$$

2761
- $\mathbf{E} \perp$ plane tangent at $\phi(x,y,z) = \phi(x_0,y_0,z_0)$
- Unit vector in direction of $\mathbf{E}$ is $\hat{\mathbf{n}} = \frac{\mathbf{E}}{||\mathbf{E}||}$, along *isocline*
- Basic definition

$$\nabla\phi(x,y,z) \equiv \lim_{|\mathcal{S}|\to 0}\left\{\frac{\iiint \phi(x,y,z)\,\hat{\mathbf{n}}\, dA}{|\mathcal{S}|}\right\}$$

$\hat{\mathbf{n}}$ is a unit vector in the direction of the gradient
$\mathcal{S}$ is the surface area centered at (x,y,z)

2762 **Divergence**

Divergence: $\nabla\cdot\boldsymbol{D} = \rho$ **39.14.4a**

- Definition: $\mathbb{R}^3 \underset{\nabla\cdot}{\mapsto} \mathbb{R}^1$

$$\nabla\cdot\mathbf{D} \equiv [\partial_x,\, \partial_y,\, \partial_z]\cdot\mathbf{D} = \left[\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}\right] = \rho(x,y,z)$$

- Examples:

2763
  - Voltage about a point charge Q [SI Units of Coulombs]

$$\phi(x,y,z) = \frac{Q}{\epsilon_0\sqrt{x^2+y^2+z^2}} = \frac{Q}{\epsilon_0 R}$$

    ϕ [Volts]; Q = [C]; Free space ϵ_0 *permittivity* (μ_0 *permeability*)

  - *Electric Displacement* (flux density) around a point charge ($\boldsymbol{D} = \epsilon_0\boldsymbol{E}$)

$$\mathbf{D} \equiv -\nabla\phi(R) = -Q\nabla\left\{\frac{1}{R}\right\} = -Q\,\delta(R)$$

Divergence: The integral definition **39.14.4b**

- Surface integral definition of *incompressible* vector field

2764
$$\nabla\cdot\mathbf{D} \equiv \lim_{|\mathcal{S}|\to 0}\left\{\frac{\iint_{\mathcal{S}} \mathbf{D}\cdot\hat{\mathbf{n}}\, dA}{|\mathcal{V}|}\right\} = \rho(x,y,z)$$

  $\mathcal{S}$ must be a closed surface
  $\hat{\mathbf{n}}$ is the unit vector in the direction of the gradient

  - $\hat{\mathbf{n}}\cdot\mathbf{D} \perp$ surface differential dA

Divergence: Gauss' Law **39.14.4c**

- General case of a *Compressible* vector field
- Volume integral over charge density $\rho(x, y, z)$ is total charge enclosed Q_{enc}
-

2765
$$\iiint_{\mathcal{V}} \nabla \cdot \boldsymbol{D} \, dV = \iint_{\mathcal{S}} \boldsymbol{D} \cdot \hat{\boldsymbol{n}} \, dA = Q_{enc}$$

- Examples
  - When the vector field is *incompressible*
    - $\rho(x, y, z) = 0$ [C/m^3] over enclosed volume
    - Surface integral is zero ($Q_{enc} = 0$)
  - Unit point charge: $D = \delta(R)$ [C/m^2]

2766 **Curl**

Curl: $\nabla \times \mathbf{H} = \mathbf{I}$ [amps/m^2] **39.14.5a**

- Definition: $\mathbb{R}^3 \underset{\nabla\times}{\mapsto} \mathbb{R}^3$

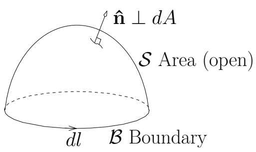
2767
$$\nabla \times \mathbf{H} \equiv \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial_x & \partial_y & \partial_z \\ H_x & H_y & H_z \end{vmatrix} = \mathbf{I}$$

- Examples:
  - Maxwell's equations: $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$, $\nabla \times \mathbf{H} = \sigma \mathbf{E} + \dot{\mathbf{D}}$,
  - $\mathbf{H} = -y\hat{x} + x\hat{y}$ then $\nabla \times \mathbf{H} = 2\hat{z}$ constant *irrotational*
  - $\mathbf{H} = 0\hat{x} + 0\hat{y} + z^2\hat{z}$ then $\nabla \times \mathbf{H} = \mathbf{0}$ is *irrotational*

2768 **Stokes' Law**

Curl: Stokes Law **39.14.5b**

- Surface integral definition of $\nabla \times \mathbf{H} = \mathbf{I}$ $\quad$ ($\boldsymbol{I} \perp$ rotation plane of $\boldsymbol{H}$)



$$\nabla \times \mathbf{H} \equiv \lim_{|\mathcal{S}| \to 0} \left\{ \frac{\iint_{\mathcal{S}} \hat{\mathbf{n}} \times \mathbf{H} \, dA}{|\mathcal{S}|} \right\} \tag{5.1}$$

2769
$$\mathcal{I}_{enc} = \iint (\nabla \times \mathbf{H}) \cdot \hat{\boldsymbol{n}} \, dA = \oint_{\mathcal{B}} \mathbf{H} \cdot d\boldsymbol{l} \tag{5.2}$$

- Eq. (1): $\mathcal{S}$ must be an *open surface*
  with closed boundary $\mathcal{B}$
  $\hat{\mathbf{n}}$ is the unit vector $\perp$ to dA
  $\boldsymbol{H} \times \hat{\boldsymbol{n}} \in$ Tangent plane of A (i.e., $\perp \hat{\boldsymbol{n}}$)
- Eq. (2): Stokes Law: Line integral of $\boldsymbol{H}$ along $\mathcal{B}$ is total current $\mathcal{I}_{enc}$

5.3.2 Lec: 40 Introduction to Maxwell's Equation

L 40: J.C. Maxwell unifies Electricity and Magnetism with the formula for the speed of light
Basic definitions of E, H, B, D
O. Heaviside's (1884) vector form of Maxwell's EM equations and the *vector wave equation*
How a loud-speaker works.

5.3.3 Lec: 41 The Fundamental theorem of Vector Calculus

L 41: *The Fundamental Thm of vector calculus*
Incompressible and *Irrotational* fluids and the two defining vector identities

WEEK 15 **40.15.0**

L 42 Quasi-static approximation and applications:
The Kirchoff's Laws and the *Telegraph wave equation*, starting from Maxwell's equations The telegraph wave equation starting from Maxwell's equations
Quantum Mechanics

L 43 Last day of class: Review of Fund Thms of Mathematics:
Closure on Numbers, Algebra, Differential Equations and Vector Calculus,
The Fundamental Thms of Mathematics & their applications:
Theorems of Mathematics; Fundamental Thms of Mathematics (Ch. 9); Normal modes vs. eigen-states, delay and quasi-statics;

– Reading Day
VC-1 Due

5.4 Kirchhoff's Laws

5.4.1 Lec 42: The Quasi-static approximation and applications

L 42: The Kirchhoff's Laws and the *Telegraph wave equation*, starting from Maxwell's equations
Quantum Mechanics

5.4.2 Lec 43: Last day of class: Review of Fund Thms of Mathematics

L 43: Closure on Numbers, Algebra, Differential Equations and Vector Calculus,
The Fundamental Thms of Mathematics & their applications:
Theorems of Mathematics; Fundamental Thms of Mathematics (Ch. 9)
Normal modes vs. eigen-states, delay and quasi-statics;

Reading Day

2801 **Properties**

Closure: Properties of fields of Maxwell's Equations **39.14.6**

The variables have the following names and defining equations:

2802

Symbol	Equation	Name	Units
$\mathbf{E}$	$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$	Electric Field strength	[Volts/m]
$\mathbf{D}$	$\nabla \cdot \mathbf{D} = \rho$	Electric Displacement (flux density)	[Col/m^2]
$\mathbf{H}$	$\nabla \times \mathbf{H} = \dot{\mathbf{D}}$	Magnetic Field strength	[Amps/m]
$\mathbf{B}$	$\nabla \cdot \mathbf{B} = 0$	Magnetic Induction (flux density)	[Weber/m^2]

In vacuo $\boldsymbol{B} = \mu_0 \boldsymbol{H}$, $\boldsymbol{D} = \epsilon_0 \boldsymbol{E}$, $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$ [m/s], $r_0 = \sqrt{\frac{\mu_0}{\epsilon_0}} = 377$ [Ω].

2803 **Vector field properties**

Closure: Summary of vector field properties **39.14.7**

- Notation: $$\boldsymbol{v}(x,y,z) = -\nabla\phi(x,y,z) + \nabla\times\boldsymbol{w}(x,y,z)$$
- Vector identities: $$\nabla\times\nabla\phi = 0; \ \nabla\cdot\nabla\times\mathbf{w} = 0$$

2804

Field type	Generator:	Test (on $\boldsymbol{v}$):
Irrotational	$\boldsymbol{v} = \nabla\phi$	$\nabla \times \boldsymbol{v} = 0$
Rotational	$\boldsymbol{v} = \nabla\times\boldsymbol{w}$	$\nabla \times \boldsymbol{v} = \boldsymbol{J}$
Incompressible	$\boldsymbol{v} = \nabla \times \boldsymbol{w}$	$\nabla \cdot \boldsymbol{v} = 0$
Compressible	$\boldsymbol{v} = \nabla\phi$	$\nabla \cdot \boldsymbol{v} = \rho$

- Source density terms: Current: $\boldsymbol{J}(x,y,z)$, Charge: $\rho(x,y,z)$
  - Examples: $\nabla\times\boldsymbol{H} = \dot{\boldsymbol{D}}(x,y,z)$, $\nabla\cdot D = \rho(x,y,z)$

2805 **Fundamental Theorem of integral Calculus**

Closure: Fundamental Theorems of integral calculus **39.14.8**

1. $f(x) \in \mathbb{R}$ (Leibniz Integral Rule): $F(x) = F(a) + \int_a^x f(x)dx$
2. $f(s) \in \mathbb{C}$ (Cauchy's formula): $F(s) = F(a) + \int_a^s f(\zeta)d\zeta$
   –When integral is independent of path, $F(s) \in \mathbb{C}$ obeys CR conditions
   –Contour integration inverts causal Laplace transforms
3. $\boldsymbol{F} \in \mathbb{R}^3$ (Helmholtz Formula): $\boldsymbol{F}(x, y, z) = -\nabla\phi(x, y, z) + \nabla\times\boldsymbol{A}(x, y, z)$
   2806 –Decompose $\boldsymbol{F}(x, y, z)$ as *compressible* and *rotational*
4. Gauss' Law (Divergence Theorem): $Q_{enc} = \iiint \nabla\cdot\boldsymbol{D}\, dV = \iint_{\mathcal{S}} \boldsymbol{D}\cdot\hat{\boldsymbol{n}}\, dA$
   –Surface integral describes enclosed compressible sources
5. Stokes' Law (Curl Theorem): $\mathcal{I}_{enc} = \iint(\nabla\times\mathbf{H})\cdot\hat{\boldsymbol{n}}\, dA = \oint_{\mathcal{B}} \mathbf{H}\cdot d\boldsymbol{l}$
   –Boundary vector line integral describes enclosed rotational sources
6. Green's Theorem ... Two-port boundary conditions
   –*Reciprocity* property (*Theory of Sound*, Rayleigh, J.W.S., 1896)

Closure: Quasi-static (QS) approximation **39.14.9**

- Definition: $ka \ll 1$ where a is the size of object, $\lambda = c/f$ wavelength
- This is equivalent to $a \ll \lambda$ or

2807

- $\omega \ll c/a$ which is a low-frequency approximation
- The QS approximation is widely used, but infrequently identified.
- All *lumped parameter models* (inductors, capacitors) are based on QS approximation as the lead term in a Taylor series approximation.

Appendix A

Notation

A.1 Number systems

The notation used in this book is defined in this appendix so that it may be quickly accessed.[1] Where the definition is sketchy, Page numbers are provided where these concepts are fully explained, along with many other important and useful definitions. For example $\mathbb{N}$ may be found on page 22.

A.1.1 Greek letters

The Greek letters used in this text include (at least) $\alpha, \beta, \gamma, \delta, \epsilon, \kappa, \rho, \xi, \omega, \sigma, \phi, \psi, \zeta$, and upper-case $\Gamma, \Xi, \Phi, \Psi, \Delta, \Omega$. Many of these imply a physical meaning. For example, ω is the radian frequency, ρ is typically a density. Φ, ϕ, Ψ, ψ are commonly used to indicate angles of a triangle. For those of you who know this, its obvious. But at some time in your life, you had to learn these conventions. They were not implanted at birth.

Likely you do not know all of these Greek letters. But they are so commonly used in mathematics, you need to recognize them. Some of them are pronounced in a way that you may not know. The symbol ξ is pronounced "c," ζ is "zeta" and χ is "kye." I must assume you know how to pronounce the others, which are more phonetic in English. One advantage of learning LaTeX is that all of these math sympols are built in, and thus easily learned, by using this powerful open-source math-oriented word-processing system.

A.1.2 Double-Bold notation

Table A.1 indicates the symbol followed by a page number and the name of the number type. For example $\mathbb{N}$ stands for the infinite set of *counting numbers* $\{1,\ 2,\ 3,\ \cdots\}$. From any counting number you may get the next one by adding 1.

Summary of various number types: Counting number ($\mathbb{N}$) are also know as the Cardinal numbers. The prime numbers ($\mathbb{P}$) cannot be further factored. The counter example of $-5 = -1 \cdot 5$ is questionable, as it could be included as a prime by a slight change in the definition. One may say that a real ($\mathbb{R}$) is a complex number ($\mathbb{C}$) with a zero imaginary part, thus real numbers are complex ($\mathbb{R} \subset \mathbb{C}$).

Note that $\mathbb{R} : \mathbb{I} \cup \mathbb{Q}$, $\mathbb{I} \perp \mathbb{Q}$, $\mathbb{Q} : \mathbb{Z} \cup \mathbb{F}$.

We say that a number is in the set with the notation $3 \in \mathbb{N} \in \mathbb{R}$, which is read as "3 is in the set of counting numbers, which in turn in the set of real numbers," or in vernacular language "3 is a real counting number."

The *cardinality* of a set is denoted by taking the absolute value (e.g., $|\mathbb{N}|$).

[1] https://en.wikipedia.org/wiki/List_of_mathematical_symbols_by_subject#Definition_symbols

Table A.1: Double-bold notation for the types of numbers. (#) is a page number.

Symbol (p. #)	Genus	Examples	Counter Examples
$\mathbb{N}$ (22)	Counting	1,2,17,3, 10^{20}	0, -10, 5j
$\mathbb{P}$ (22)	Prime	2,17,3, 10^{20}	0, 1, 4, 3^2, 12, -5
$\mathbb{Z}$ (22)	Integer	-1, 0, 17, 5j, -10^{20}	1/2,π, $\sqrt{5}$
$\mathbb{Q}$ (22)	Rational	2/1, 3/2, 1.5, 1.14	$\sqrt{2}$, $3^{-1/3}$, π
$\mathbb{F}$ (22)	Fractional	1/2, 7/22	2/1, $1/\sqrt{2}$
$\mathbb{I}$ (23)	Irrational	$\sqrt{2}$, $3^{-1/3}$, π	Vectors
$\mathbb{R}$ (23)	Reals	$\sqrt{2}$, $3^{-1/3}$, π	$2\pi j$
$\mathbb{C}$ (121)	Complex	1, $\sqrt{2}j$, $3^{-j/3}$, π^j	Vectors

A.2 Periodic functions

Any periodic function may be indicated using double-parentheses notation. This is sometimes known as modular arithmetic. For example function

$$f((t))_T = f(t) = f(t \pm kT),$$

is periodic on $t, T \in \mathbb{R}$ with a period of T and $k \in \mathbb{Z}$. This notation is useful when dealing with Fourier series of periodic functions.

When a discrete valued (e.g., time) sequence is periodic we use square brackets

$$f[[n]]_N = f[n] = f[n \pm kN],$$

with $n, k, N \in \mathbb{Z}$ and period N. This notation will be used with discrete-time signals that are periodic, such as the case of the DFT.

A.3 Vectors

Vectors are ordered sets of scalars. When we write then out, we use row notation, with the *transpose* symbol

$$[a, b, c]^T = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Vectors are always columns. Row vectors are weights not vectors. A vector dot product is normally defined between weights and vectors, resulting in a real scalar. This is said to be a 3 *dimensional* vector. for example

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 2 + 3 = 6.$$

When the elements are complex, the transpose also takes the complex conjugate.

A.4 Matrices

Unfortunately when working with matrices, the role of the weights and vectors can change, depending on the context. A useful way to view a matrix is as a set of column vectors, weighted by the elements of the column-vector of weights multiplied from the right. For example

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1M} \\ a_{21} & a_{22} & a_{32} & \cdots & a_{3M} \\ & & \ddots & & \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NM} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_M \end{bmatrix} = w_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{21} \\ \cdots \\ a_{N1} \end{bmatrix} + w_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \cdots \\ a_{N2} \end{bmatrix} \ldots w_M \begin{bmatrix} a_{1M} \\ a_{2M} \\ a_{3M} \\ \cdots \\ a_{NM} \end{bmatrix},$$

where the weights are $\Big[w_1, w_2, \ldots, w_M.\Big]^T$

Another way to view the matrix is a set of row vectors of weights, each of which re applied to the vector $[w_1, w_2, \cdots, W_M]^T$.

The determinant of a matrix is denoted either as det $\boldsymbol{A}$ or $|\boldsymbol{A}|$, as in the absolute value. The inverse of a square matrix is $\boldsymbol{A}^{-1}$ or inv$\boldsymbol{A}$. If $|A| = 0$, the inverse does not exist. $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A}$.

Matlab's notional convention for a row-vector is $[a, b, c]$ and a column-vector is $[a; b; c]$. A prime on a vector takes the complex conjugate transpose. To suppress the conjugation, place a period before the prime. The : argument converts the array into a column vector, without conjugation. A tacit notation in Matlab is that *vectors* are columns and the index to a vector is a row vector. Matlab defines the notation $1\!:\!4$ as the "row-vector" $[1, 2, 3, 4]$, which is unfortunate as it leads users to assume that the default vector is a row. This can lead to serious confusion later, as Matlab's default vector is a column. I have not found the above convention explicitly stated, and it took me years to figure this out for myself.

Units are SI; Angles in degrees [deg] unless otherwise noted. The units for π are always in radians [rad]. Ex: $\sin(\pi), e^{\jmath 90^\circ} e^{\jmath\pi/2}$.

when writing a complex number we shall try to use $1\jmath$ to indicate $\sqrt{-1}$. Matlab prefers this as well, as its explicit.

A.5 Differential equations vs. Polynomials

A polynomial has *degree N* defined by the largest power. A quadratic equation is degree 2, and a cubic has degree 3. We shall indicate a polynomial by the notation

$$P_N(z) = z^N + a_{N-1}z^{N-1} \cdots a_0.$$

Always normalize the polynomial so that $a_N = 1$. This will not change the roots, defined by Eq. 1.10 (p. 42). The coefficient on z^{N-1} is always the sum of the roots z_n ($a_{N-1} = \sum_n^N z_n$), and a_0 is always their product ($a_0 = \prod_n^N z_n$).

Differential equations have *order* (polynomials have degree). If a second order differential equation is Laplace transformed (Lec. 1.3.10, p. 59), one is left with a degree 2 polynomial. For example:

$$\frac{d^2}{dt^2}y(t) + b\frac{d}{dt}y(t) + cy(t) = \alpha\left(\frac{d}{dt}x(t) + \beta x(t)\right) \leftrightarrow \tag{A.1}$$

$$(s^2 + bs + c)Y(s) = \alpha(s + \beta)X(s). \tag{A.2}$$

$$\frac{Y(s)}{X(s)} = \alpha\frac{s+\beta}{s^2+bs+c} \equiv H(s) \leftrightarrow h(t). \tag{A.3}$$

Using the same argument as for polynomials, the lead coefficent must always be 1. The coefficent $\alpha \in \mathbb{R}$ is called the *gain*. The complex variable s is the *Laplace frequency.*

The ratio of the output $Y(s)$ over the input $X(s)$ is called the system *transfer function* $H(s)$. When $H(s)$ is the ratio of two polynomials in s, the transfer function is said to be *bilinear*, since it is linear in both the input and output. The roots of the numerator are called the *zeros* and those of the denominator, the *poles.* The inverse Laplace transform of the transfer function is called the system *impulse response*, which describes the system's output signal $y(t)$ for any given input signal $x(t)$, via convolution (i.e., $y(t) = h(t) \star x(t)$).

A.6 Residue expansions and the ROC

With the new tool of analytic functions came the concept of the *region of convergence* (ROC) that defines the regions in the complex plane where the infinite series is valid. In other words, the function $Z(s)$ and its analytic power series $\sum_0^\infty c_n s^n$, are equivalent over a region of s that lies within the ROC. When the series fails to converge, it no longer represents $Z(s)$. A helpful example is the series

$$\frac{1}{1+x^2} = \frac{1}{(1-x\jmath)(1+x\jmath)} = \frac{A}{1-x\jmath} + \frac{B}{1+x\jmath} = \frac{1}{2}\sum_{n=0}^{\infty}(+x\jmath)^n + \frac{1}{2}\sum_{n=0}^{\infty}(-x\jmath)^n,$$

which is valid for $|x| < 1$. At face value this function seems fine at $x = 1$, where it is equal to $1/2$. In fact the series fails to converge at precisely this value (the ROC is 1 for this example). Until one views x as complex, this behavior is not obvious.

A trivial analysis shows that $A = 1/2$ and $B = A$ since

$$1 = A(1 + x\jmath) + B(1 - x\jmath) = (A + Ax\jmath) + (B - Bx\jmath) = A + B + (A - B)x\jmath.$$

Appendix B

Linear algebra on 2x2 matrices

B.1 Gaussian elimination

Definitions:

1. *Scalar*: A number, e.g. $\{a, b, c, \alpha, \beta, \cdots\} \in \{\mathbb{Z}, \mathbb{Q}, \mathbb{I}, \mathbb{R}, \mathbb{C}\}$

2. *Vector*: A quantity having direction as well as magnitude, often denoted by a bold-face letter with an arrow, $\vec{\mathbf{x}}$. In matrix notation, this is typically represented as a single row $[x_1, x_2, x_3, \ldots]$ or single column $[x_1, x_2, x_3 \ldots]^T$ (where T indicates the transpose). In this class we will typically use column vectors. The vector may also be written out using unit vector notation to indicate direction. For example: $\vec{\mathbf{x}}_{3\times 1} = x_1\hat{x} + x_2\hat{y} + x_3\hat{z} = [x_1, x_2, x_3]^T$, where $\hat{x}, \hat{y}, \hat{z}$ are unit vectors in the x, y, z cartesian directions (here the vector's subscript 3×1 indicates its dimensions). The type of notation used may depend on the engineering problem you are solving.

3. *Matrix:* $A = \left[\vec{\mathbf{a}}_1, \vec{\mathbf{a}}_2, \vec{\mathbf{a}}_3, \cdots, \vec{\mathbf{a}}_M\right]_{N\times M} = \{a_{n,m}\}_{N\times M}$, can be a non-square matrix if the number of elements in each of the vectors (N) is not equal to the number of vectors (M). When $M = N$, the matrix is square. It may be inverted if its determinant $|A| = \prod \lambda_k \neq 0$ (where λ_k are the eigenvalues).

   We shall only work with 2×2 and 3×3 square matrices throughout this course.

4. *Linear system of equations*: $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ where $\vec{\mathbf{x}}$ and $\vec{\mathbf{b}}$ are vectors and matrix A is a square.

   (a) *Inverse:* The solution of this system of equations may be found by finding the inverse $\vec{\mathbf{x}} = A^{-1}\vec{\mathbf{b}}$

   (b) *Equivalence:* If two systems of equations $A_0\vec{\mathbf{x}} = \vec{\mathbf{b}}_0$ and $A_1\vec{\mathbf{x}} = \vec{\mathbf{b}}_1$ have the same solution (i.e., $\vec{\mathbf{x}} = A_0^{-1}\vec{\mathbf{b}}_0 = A_1^{-1}\vec{\mathbf{b}}_1$), they are said to be equivalent.

   (c) *Augmented matrix:* The first type of augmented matrix is defined by combining the matrix with the right-hand-side. For example, given the linear system of equations $A\vec{\mathbf{x}} = \vec{\mathbf{y}}$

   $$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix},$$

   then the augmented matrix is

   $$A|y = \left[\begin{array}{cc|c} a & b & y_1 \\ c & d & y_2 \end{array}\right]$$

   A second type of augmented matrix may be used for finding the inverse of a matrix (rather than solving a specific instance of linear equations $Ax = b$). In this case the augmented matrix is

   $$A|I = \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right]$$

Performing Gaussian elimination on this matrix, until the left side becomes the identity matrix, yields A^{-1}. This is because multiplying both sides by A^{-1} gives $A^{-1}A|A^{-1}I = I|A^{-1}$.

5. *Permutation matrix* (P): A matrix that is equivalent to the identity matrix, but with scrambled rows (or columns). Such a matrix has the properties $det(P) = \pm 1$ and $P^2 = I$. For the 2x2 case, there is only one permutation matrix:

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad P^2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

A permutation matrix P swaps rows or columns of the matrix it operates on. For example, in the 2x2 case, pre-multiplication swaps the rows

$$PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ a & b \end{bmatrix},$$

whereas post-multiplication swaps the *columns*

$$AP = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} b & a \\ \beta & \alpha \end{bmatrix}$$

For the 3x3 case there are $3 \cdot 2 = 6$ such matrices, including the original 3x3 identity matrix (swap a row with the other 2, then swap the remaining two rows).

6. *Gaussian elimination (GE) matrices* G_k: There are 3 types of elementary row operations, which may be performed without fundamentally altering a system of equations (e.g. the resulting system of equations is *equivalent*). These operations are (1) swap rows (e.g. using a permutation matrix), (2) scale rows, or (3) perform addition/subraction of two scaled rows. All such operations can be performed using matrices.

For lack of a better term, we'll describe these as 'gaussian elimination' or 'GE' matrices.[1] We will categorize any matrix that performs only elementary row operations (but any number of them) as a 'GE' matrix. Therefore, cascade of GE matrices is also a GE matrix.

Consider the GE matrix

$$G = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}$$

(a) Pre-multiplication scales and adds the *rows*

$$GA = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix} = \begin{bmatrix} a & b \\ a-\alpha & b-\beta \end{bmatrix}$$

The result is a Gaussian elimination operation.

(b) Post-multiplication adds and scales *columns*.

$$AG = \begin{bmatrix} a & b \\ \alpha & \beta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} a-b & b \\ \alpha-\beta & \beta \end{bmatrix}$$

Here the second column is subtracted from the first, and placed in the first. The second column is untouched. **This operation is *not* a Gaussian elimination**. Therefore, to

[1] The term 'elementary matrix' may also be used to refer to a matrix that performs an elementary row operation. Typically, each elementary matrix differs from the identity matrix by one single row operation. A cascade of elementary matrices could be used to perform Gaussian elimination.

put Gaussian elimination operations in matrix form, we form a cascade of **pre-multiply** matrices.

Here $\det(G) = 1$, $G^2 = I$, which won't always be true if we scale by a number greater than 1. For instance, if $G = \begin{bmatrix} 1 & 0 \\ m & 1 \end{bmatrix}$ (scale and add), then we have $\det(G) = 1$, $G^n = \begin{bmatrix} 1 & 0 \\ n \cdot m & 1 \end{bmatrix}$.

B.1.1 Problems

Find the solution to the following 3x3 matrix equation $Ax = b$ by Gaussian elimination. Show your intermediate steps. You can check your work at each step using Matlab.

$$\begin{bmatrix} 1 & 1 & -1 \\ 3 & 1 & 1 \\ 1 & -1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 8 \end{bmatrix}.$$

1. Show (i.e., verify) that the first GE matrix G_1, which zeros out all entries in the first column, is given by

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

   Identify the elementary row operations that this matrix performs. See Lab manual for solutions.

2. Find a second GE matrix, G_2, to put $G_1 A$ in upper triangular form. Identify the elementary row operations that this matrix performs. See Lab manual for solutions.

3. Find a third GE matrix, G_3, which scales each row so that its leading term is 1. Identify the elementary row operations that this matrix performs. See Lab manual for solutions.

4. Finally, find the last GE matrix, G_4, that subtracts a scaled version of row 3 from row 2, and scaled versions of rows 2 and 3 from row 1, such that you are left with the identity matrix ($G_4 G_3 G_2 G_1 A = I$). See Lab manual for solutions.

5. Solve for $[x_1, x_2, x_3]^T$ using the augmented matrix format $G_4 G_3 G_2 G_1 [A|b]$ (where $[A|b]$ is the augmented matrix). Note that if you've performed the preceding steps correctly, $x = G_4 G_3 G_2 G_1 b$. See Lab manual for solutions.

B.2 Gaussian Elimination of a 2x2 matrix

We shall now apply Gaussian elimination to find the solution $[x_1, x_2]$ for the 2x2 matrix equation $Ax = y$ (Eq. 3.8, left). We assume to know $[a, b, c, d]$ and $[y_1, y_2]$. We wish to show that the intersection (solution) is given by the equation on the right.

Here we wish to prove that the left equation has an inverse given by the right equation:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

How to take inverse:
1) Swap the diagonal, 2) change the signs of the off-diagonal, and 3) divide by Δ.

B.2.1 Derivation of the inverse of a 2x2 matrix

1. Step 1: normalize the first column to 1.
2. Step 2: subtract the top equation from the lower.
3. Step 3: express result in terms of the determinate $\Delta = ad - bc$.

$$\begin{bmatrix} 1 & \frac{b}{a} \\ 1 & \frac{d}{c} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a} y_1 \\ \frac{1}{c} y_2 \end{bmatrix} \qquad \begin{bmatrix} 1 & \frac{b}{a} \\ 0 & \frac{d}{c} - \frac{b}{a} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a} & 0 \\ -\frac{1}{a} & \frac{1}{c} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \begin{bmatrix} 1 & \frac{b}{a} \\ 0 & \Delta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a} & 0 \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

4. Step 4: These steps give the solution for x_2: $x_2 = -\frac{c}{\Delta} y_1 + \frac{a}{\Delta} y_2$.
5. Step 5: Finally the top equation is solved for x_1: $x_1 = \frac{1}{a} y_1 - \frac{b}{a} x_2 = x_1 = \frac{1}{a} y_1 - \frac{b}{a}[-\frac{c}{\Delta} y_1 + \frac{a}{\Delta} y_2]$.

In matrix format, in terms of the determinate $\Delta = ab - cd$ becomes:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{a} - \frac{bc}{a\Delta} & \frac{b}{\Delta} \\ -\frac{c}{\Delta} & \frac{a}{\Delta} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \frac{\Delta - bc}{a} & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

Summary: This is a lot of algebra, that is why it is essential you memorize the formula for the inverse.

Appendix C

Eigenvector analysis

Here we show how to compute the eigenvalues and eigenvectors for the 2x2 Pell matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & N \\ 1 & 1 \end{bmatrix}.$$

The analysis applies to any matrix, but since we are concentrated on Pell's equation, we shall use the Pell matrix, for $N = 2$. By using a specific matrix we can check all the equations below with Matlab, which I advise you to do.

The Matlab command `[E,D]=eig(A)` returns the eigenvector matrix $\boldsymbol{E}$

$$\boldsymbol{E} = [\boldsymbol{e}_+, \boldsymbol{e}_-] = \frac{1}{\sqrt{3}} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0.8165 & -0.8165 \\ 0.5774 & 0.5774. \end{bmatrix}$$

and the eigenvalue matrix Λ (Matlab's D)

$$\Lambda \equiv \begin{bmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{bmatrix} = \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix} = \begin{bmatrix} 2.4142 & 0 \\ 0 & -0.4142 \end{bmatrix}.$$

The factor $\sqrt{3}$ on $\boldsymbol{E}$ normalizes each eigenvector to 1 (i.e., The Matlab command `norm([`$\sqrt{2}, 1$`])` gives $\sqrt{3}$).

In the following discussion we show how to determine $\boldsymbol{E}$ and D (i.e, Λ), given $\boldsymbol{A}$.

Calculating the eigenvalue matrix (Λ): The matrix equation for $\boldsymbol{E}$ is

$$\boldsymbol{A}\boldsymbol{E} = \boldsymbol{E}\Lambda. \tag{C.1}$$

Pre-multiplying by $\boldsymbol{E}^{-1}$ diagonalizes $\boldsymbol{A}$, given the *eigenvalue matrix* (D in Matlab)

$$\Lambda = \boldsymbol{E}^{-1}\boldsymbol{A}\boldsymbol{E}. \tag{C.2}$$

Post-multiplying by $\boldsymbol{E}^{-1}$ recovers $\boldsymbol{A}$

$$\boldsymbol{A} = \boldsymbol{E}\Lambda\boldsymbol{E}^{-1}. \tag{C.3}$$

Matrix power formula: This last relation is the entire point of the eigenvector analysis, since it shows that any power of $\boldsymbol{A}$ may be computed from powers of the eigen values. Specifically

$$\boldsymbol{A}^n = \boldsymbol{E}\Lambda^n\boldsymbol{E}^{-1}. \tag{C.4}$$

For example, $\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A} = \boldsymbol{E}\Lambda\,(\boldsymbol{E}^{-1}\boldsymbol{E})\,\Lambda\boldsymbol{E}^{-1} = \boldsymbol{E}\Lambda^2\boldsymbol{E}^{-1}$.

Equations C.1, C.2 and C.3 are the key to eigenvector analysis, and you need to memorize them. You will use them repeatedly throughout this course, and for a long time after it is over.

Showing that $\boldsymbol{A} - \lambda_\pm \boldsymbol{I}_2$ is singular: If we restrict Eq. C.1 to a single eigenvector (one of $\boldsymbol{e}_\pm$), along with the corresponding eigenvalue $\lambda_\pm$, we obtain a matrix equations

$$\boldsymbol{A}\boldsymbol{e}_\pm = \boldsymbol{e}_\pm \lambda_\pm = \lambda_\pm \boldsymbol{e}_\pm$$

Note the important swap in the order of $\boldsymbol{e}_\pm$ and $\lambda_\pm$. Since $\lambda_\pm$ is a scalar, this is legal (and critically important), since this allows us to remove (factored out) $\boldsymbol{e}_\pm$

$$(\boldsymbol{A} - \lambda_\pm \boldsymbol{I}_2)\boldsymbol{e}_\pm = 0. \tag{C.5}$$

This means that the matrix $\boldsymbol{A} - \lambda_\pm \boldsymbol{I}_2$ must be singular, since when it operates on $\boldsymbol{e}_\pm$, which is not zero, it gives zero. It immediately follows that its determinant is zero (i.e., $|(\boldsymbol{A} - \lambda_\pm \boldsymbol{I}_2)| = 0$). This equation is used to uniquely determine the eigenvalues $\lambda_\pm$. Note the important difference between $\lambda_\pm \boldsymbol{I}_2$ and Λ (i.e., $|(A - \Lambda)| \neq 0$).

Calculating the eigenvalues $\lambda_\pm$: The eigenvalues $\lambda_\pm$ of $\boldsymbol{A}$ may be determined from $|(\boldsymbol{A} - \lambda_\pm \boldsymbol{I}_2)| = 0$

$$\begin{vmatrix} 1-\lambda_\pm & N \\ 1 & 1-\lambda_\pm \end{vmatrix} = (1-\lambda_\pm)^2 - N^2 = 0.$$

For our case of $N = 2$, $\lambda_\pm = (1 \pm \sqrt{2})$.[1]

Calculating the eigenvectors $\boldsymbol{e}_\pm$: Once the eigenvalues have been determined, they are substitute them into Eq. C.5, which determines the eigenvectors $\boldsymbol{E} = \left[\mathbf{e}_+, \mathbf{e}_-\right]$, by solving

$$(\boldsymbol{A} - \lambda_\pm)\mathbf{e}_\pm = \begin{bmatrix} 1-\lambda_\pm & 2 \\ 1 & 1-\lambda_\pm \end{bmatrix} \boldsymbol{e}_\pm = 0$$

where $1 - \lambda_\pm = 1 - (1 \pm \sqrt{2}) = \mp\sqrt{2}$.

Recall that Eq. C.5 is singular, because we are using an eigenvalue, and each eigenvector is pointing in a unique direction (This is why it is singular). You might respectively suggest that this equation has no solution. In some sense you would be correct. When we solve for $\boldsymbol{e}_\pm$, the two equations defined by Eq. C.5 *co-linear* (the two equations describe parallel lines). This follows from the fact that there is only one eigenvector for each eigenvalue.

Expecting trouble, yet proceeding to solve for $\boldsymbol{e}_+ = [e_1^+, e_2^+]^T$,

$$\begin{bmatrix} -\sqrt{2} & 2 \\ 1 & -\sqrt{2} \end{bmatrix} \begin{bmatrix} e_1^+ \\ e_2^+ \end{bmatrix} = 0$$

This gives two identical equations $-\sqrt{2}e_1^+ + 2e_2^+ = 0$ and $e_1^+ - \sqrt{2}e_2^+ = 0$. This is the price of an over-specified equation (the singular matrix is degenerate). The most we can determine is $\mathbf{e}_+ = c\,[-\sqrt{2}, 1]^T$, where c is a constant. We can determine eigenvector direction, but not its magnitude.

Following *exactly* the same procedure for λ_-, the equation for $\boldsymbol{e}_-$ is

$$\begin{bmatrix} \sqrt{2} & 2 \\ 1 & \sqrt{2} \end{bmatrix} \begin{bmatrix} e_1^- \\ e_2^- \end{bmatrix} = 0$$

In this case the relation becomes $e_1^- + \sqrt{2}e_2^- = 0$, thus $\boldsymbol{e}_- = c\,[\sqrt{2}, 1]^T$ where c is a constant.

Normalization of the eigenvectors: The two constants may be determined by normalizing the eigenvectors to have unit length. Since we cannot determine the length, we set it to 1. In some sense the degeneracy is resolved by this normalization. Thus $c = 1/\sqrt{3}$, since

$$c^2\left(\left(\pm\sqrt{2}\right)^2 + 1^2\right) = 3c^2 = 1.$$

[1] It is a convention to order the eigenvalues from largest to smallest.

Summary: Thus far we have shown

$$\boldsymbol{E} = [\boldsymbol{e}_+, \boldsymbol{e}_-] = \frac{1}{\sqrt{3}} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{bmatrix} = \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix}.$$

Verify that $\Lambda = \boldsymbol{E}^{-1}\boldsymbol{A}\boldsymbol{E}$: To find the inverse of $\boldsymbol{E}$, 1) swap the diagonal values, 2) change the sign of the off diagonals, and 3) divide by the determinant $\Delta = 2\sqrt{2}/\sqrt{3}$ (see Appendix B)

$$\boldsymbol{E}^{-1} = \frac{\sqrt{3}}{2\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} \\ -1 & \sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.6124 & 0.866 \\ -0.6124 & 0.866 \end{bmatrix}.$$

By definition for any matrix $\boldsymbol{E}^{-1}\boldsymbol{E} = \boldsymbol{E}\boldsymbol{E}^{-1} = \boldsymbol{I}_2$. Taking the product gives

$$\boldsymbol{E}^{-1}\boldsymbol{E} = \frac{\cancel{\sqrt{3}}}{2\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} \\ -1 & \sqrt{2} \end{bmatrix} \cdot \frac{1}{\cancel{\sqrt{3}}} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}_2.$$

We wish to show that $\Lambda = \boldsymbol{E}^{-1}\boldsymbol{A}\boldsymbol{E}$

$$\begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix} . = \frac{\cancel{\sqrt{3}}}{2\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} \\ -1 & \sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \cdot \frac{1}{\cancel{\sqrt{3}}} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix},$$

which is best verified with Matlab.

Verify that $\boldsymbol{A} = \boldsymbol{E}\Lambda\boldsymbol{E}^{-1}$: We wish to show that

$$\begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} = \frac{1}{\cancel{\sqrt{3}}} \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix} \cdot \frac{\cancel{\sqrt{3}}}{2\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} \\ -1 & \sqrt{2} \end{bmatrix},$$

which is best verified with Matlab (or Octave).

I suggest that you verify $\boldsymbol{E}\Lambda \neq \Lambda\boldsymbol{E}$ and $\boldsymbol{A}\boldsymbol{E} = \boldsymbol{E}\Lambda$ with Matlab. Here is the Matlab program which does this:

```
A = [1 2; 1 1]; %define the matrix
[E,D] = eig(A); %compute the eigenvector and eigenvalue matrices
A*E-E*D %This should be $\approx 0$, within numerical error.
E*D-D*E %This is not zero
```

All the equations have been verified both with Matlab and Octave.

Appendix D

Solution to Pell's Equation (N=2)

Section 2.2.2 (p. 89) showed that the solution $[x_n, y_n]^T$ to Pell's equation, for $N = 2$, is given by powers of Eq. 1.4. To find an explicit formula for $[x_n, y_n]^T$, one must compute powers of

$$\boldsymbol{A} = 1\jmath \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}. \tag{D.1}$$

We wish to find the solution to Pell's equation (Eq. 1.4), based on the recursive solution, Eq. 1.5 (p. 37). Thus we need is powers of A, that is A^n, which gives the a closed form expression for $[x_n, y_n]^T$. By the diagonalization of A, its powers are simply the powers of its eigenvalues. This diagonalization is called an *eigenvalue analysis*, a very general method rooted in linear algebra. This type of analysis allows us to find the solution to most of the linear the equations we encounter.

From Matlab with $N = 2$ the eigenvalues of Eq. D.1 are $\lambda_\pm \approx [2.4142\jmath, -0.4142\jmath]$ (i.e., $\lambda_\pm = 1\jmath(1 \pm \sqrt{2})$). The final solution to Eq. D.1 is given in Eq. 2.5 (p. 89). The solution for $N = 3$ is provided in Appendix D.1 (p. 143).

Once the matrix has been diagonalized, one may compute powers of that matrix as powers of the eigenvalues. This results in the general solution given by

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = 1\jmath^n \boldsymbol{A}^n \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1\jmath^n \boldsymbol{E} \Lambda^n \boldsymbol{E}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

The eigenvalue matrix D is diagonal with the eigenvalues sorted, largest first. The Matlab command `[E,D]=eig(A)` is helpful to find D and E given any A. As we saw above,

$$\Lambda = 1\jmath \begin{bmatrix} 1+\sqrt{2} & 0 \\ 0 & 1-\sqrt{2} \end{bmatrix} \approx \begin{bmatrix} 2.414\jmath & 0 \\ 0 & -0.414\jmath \end{bmatrix}.$$

D.1 Pell equation for N=3

This summarizes the solution of Pell's equation due to the Pythagoreans using matrix recursion, for the case of N=3. The integer solutions are shown in on the right. Note that $x_n/y_n \to \sqrt{3}$, in agreement with the Euclidean algorithm.[1] It seem likely that β_0 could be absorbed in the starting solution, and then be removed from the generating function, other than as the known factor β_0^n

Case of $N = 3$: $[x_0, y_0]^T = [1, 0]^T$, $\beta_0 = \jmath/\sqrt{2}$; Pell-3: $x_n^2 - 3y_n^2 = 1$; $x_n/y_n \underset{\infty}{\longrightarrow} \sqrt{3}$

Try other trivial solutions such as $[-1, 0]^T$ and $[\pm\jmath, 0]^T$. Perhaps this can provide a clue to the proper value of β_0.

[1]The matlab program for generating this solution is `PellSol3.m`.

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \beta_0 \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad (1\beta_0)^2 - 3(1\beta_0)^2 = 1$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \beta_0^2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \beta_0^2 \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \left(4\beta_0^2\right)^2 - 3\left(2\beta_0^2\right)^2 = 1$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \beta_0^3 \begin{bmatrix} 10 \\ 6 \end{bmatrix} = \beta_0^3 \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} \qquad \left(10\beta_0^3\right)^2 - 3\left(6\beta_0^3\right)^2 = 1$$

$$\begin{bmatrix} x_4 \\ y_4 \end{bmatrix} = \beta_0^4 \begin{bmatrix} 28 \\ 16 \end{bmatrix} = \beta_0^4 \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \end{bmatrix} \qquad \left(28\beta_0^4\right)^2 - 3\left(16\beta_0^4\right)^2 = 1$$

$$\begin{bmatrix} x_5 \\ y_5 \end{bmatrix} = \beta_0^5 \begin{bmatrix} 76 \\ 44 \end{bmatrix} = \beta_0^5 \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 28 \\ 16 \end{bmatrix} \qquad \left(76\beta_0^5\right)^2 - 3\left(44\beta_0^5\right)^2 = 1$$

List of Figures

1 There is a natural symbiotic relationship between Physics, Mathematics and Engineering, as depicted by this Venn diagram. Physics explores the boundaries. Mathematics provides the method and rigor. engineering transforms the method into technology. While these three disciplines work well together, there is poor communication due to a different vocabulary. 10

2 Table of contents of Stillwell (2002) 11

1.1 Depiction of the argument of Galileo (unpublished book of 1638) as to why weights of different masses (weight) must fall with identical velocity. By joining them with an elastic cord they become one. Thus if the velocity were proportional to the mass, the joined masses would fall even faster. This results in a logical fallacy. This may have been the first time that the principle of conservation of energy was clearly stated. 14

1.2 Mathematical time-line between 1500 BCE and 1650 CE. 16

1.3 Above: Jakob (1655-1705) and Johann (1667-1748) Bernoulli; Below: Leonhard Euler (1707) and Jean le Rond d'Alembert (1717-1783). The figure numbers are from Stillwell (2010). 18

1.4 Time-line of the four centuries from the 16^{th} and 20^{th} CE 20

1.5 Sieve of Eratosthenes for the case of $N = 50$. 31

1.6 The Euclidean algorithm for finding the GCD of two numbers is one of the oldest algorithms in mathematics, and is highly relevant today. It is both powerful and simple. It was used by the Chinese during the Han dynasty (Stillwell, 2010, p. 70) for reducing fractions. It may be used to find pairs of integers that are coprime (their gcd must be 1), and it may be used to identify factors of polynomials by long division. It has an important sister algorithm called the continued fraction algorithm (CFA), that is so similar in concept that Gauss referred to the Euclidean algorithm as the"continued fraction algorithm" (Stillwell, 2010, p. 48). 34

1.7 The expansion of π to various orders using the CFA, along with the order of the error of each rational approximation. For example 22/7 has an absolute error ($|22/7 - \pi|$) of about 0.13%. 35

1.8 Beads on a string form perfect right triangles when number of beads on each side satisfy Eq. 1.1. 37

1.9 "Plimpton-322" is a stone tablet from 1800 BCE, displaying a and c values of the Pythagorean triplets $[a, b, c]$. Numbers $(a, c \in \mathbb{N})$, with the property $b = \sqrt{c^2 - a^2} \in \mathbb{N}$, known as Pythagorean triplets, were found carved on a stone tablet from the 19^{th} century [BCE]. Several of the c values are primes, but not the a values. The stone is item 322 (item 3 from 1922) from the collection of George A. Plimpton. –Stillwell (2010, Exercise 1.2) 38

1.10 This summarizes the solution of Pell's equation for $N = 2$ using a slightly modified matrix recursion. Note that $x_n/y_n \to \sqrt{2}$ as $n \to \infty$, which was what the Pythagoreans were pursuing. 39

1.11 This is a single LC segment (cell) of the transmission line show in Fig. 2.2. It may be modeled by the ABCD model as the product of two matrices, as show below. A more detailed analysis shows that the wave number is $\kappa = \sqrt{ZY} = \sqrt{sL \cdot sC} = s\sqrt{LC}$. It follows that the wave velocity is $c = 1/\sqrt{LC}$. The characteristic resistance is $r_c = \sqrt{Z/Y} = \sqrt{sL/sC} = \sqrt{L/C}$. 52

1.12 *The left panel shows how the real line may be composed with the circle. Each real x value maps to a corresponding point x' on on the unit circle. The point $x \to \infty$ then naturally maps to the north pole N. This simple idea may be extended with the composition of the complex plane with the unit sphere, thus mapping the plane onto the sphere. As with the circle, the point on the complex plane $z \to \infty$ maps onto the north pole N. This construction is important because while the plane is open (does not include $z \to \infty$), the sphere is analytic at the north pole. Thus the sphere defines the* closed extended plane. *Figure from Stillwell (2010, pp. 299-300).* 54

1.13 *On the left is a color map showing the definition of the complex mapping from the $s = \sigma + \omega\jmath$ plane to the $w(s) = u(\sigma,\omega) + v(\sigma,\omega)\jmath$ plane. This mapping from one complex plane $s = \sigma + \omega\jmath$ to another $w = u(\sigma,\omega) + v(\sigma,\omega)\jmath$, is visualized by the use of intensity (light/dark) to indicate magnitude, and color (hue) to indicate angle (phase). On the left $w(s) = s = \sigma + \omega\jmath$ (i.e., $u = \sigma$ and $v = v$). On the right is $w(z) = z - \sqrt{\jmath}$, a shift to the right and up by $\sqrt{2}/2 = 0.707$.* 56

1.14 *On the left is the function $w(s) = e^s$ and on the right is $s = log(w)$.* 56

1.15 *Here we see the mapping for the square root function $w(z) = \pm\sqrt{z}$ which has two single-valued sheets, corresponding to the two signs of the square root. The lower sheet is $+\sqrt{z}$, and the upper sheet is $-\sqrt{z}$. The location of the branch cut may be moved by rotating the z coordinate system. For example, $w(z) = \pm\jmath\sqrt{z}$ and $w(z) = \pm\sqrt{-z}$ have a different branch cuts, as may be easily verified using the Matlab commands* `j*zviz(z)` *and* `zviz(-z)`. *A function is analytic on the branch cut, since the cut may be moved. If a Taylor series is formed on the branch cut, it will describe the function on the two different sheets. Thus the Taylor series is agnostic to the location of the cut, as it only describes the function uniquely (as a single valued function), valid in its local region of convergence. Figure from Stillwell (2010, p. 303).* 69

1.16 *von Helmholtz portrait taken from the English translation of his 1858 paper "On integrals of the hydrodynamic equations that correspond to Vortex motions" (in German) (von Helmholtz, 1978).* . . . 76

2.1 *The Euclidean Algorithm recursively subtracts n from m until the remainder $m - kn$ is either less than n or zero. For the case depicted here the value of k that renders the remainder less than n is $k = 6$. If one more step were taken ($k = 7$) the remainder becomes negative. By linear interpolation we can find that $m - an = 0$ when $a = m/n$, which for this example is close to a=6.5. In this example $6 = floor(m/n) < n$.* 84

2.2 *This transmission line (TL) is known as a* low-pass filter *wave-filter (Campbell, 1922), composed of series inductors having an impedance $Z_k = sL$ and shunt admittance $Y_k = sC$. The characteristic resistance of the TL is $r_0 = \sqrt{Z_k/Y_k}$ and the wave velocity is $\kappa = \sqrt{Z_k Y_k} = s/\sqrt{LC} = s/c$. Thus the wave velocity is $c = 1/\sqrt{LC}$. The length Δ [m] defines the physical size of each cell. For long wavelengths the wave-filter acts as a pure delay. But when the frequency increases above a cutoff frequency $f_c > c/\Delta$, the wavelength $\lambda = c/f$ is less than the size of a section Δ, and the system response becomes very high-order low-pass filter.* 86

2.3 *Derivation of Euclid's formula for the Pythagorean triplets $[a, b, c]$, based on a composition of a line, having a rational slope $t = p/q \in \mathbb{Q}$, and a circle $c^2 = a^2 + b^2, [a, b, c] \in \mathbb{N}$. This analysis is attributed to Diophantus (Di·o·phan·tus) (250 CE), and today such equations are called Diophantine (Di·o·phan′·tine)* equations. *PTs have applications in architecture and scheduling, and many other practical problems.* 87

2.4 Properties of the Fibonacci numbers (Stillwell, 2010, p. 28). 90

3.1 *The* Schwarz inequality *is related to the shortest distance (length of a line) between the ends of the two vectors. $||U|| = \sqrt{(U \cdot U)}$ as the dot product of that vector with itself.* 96

3.2 *Here the Cartesian coordinate map between $s = \sigma + \omega\jmath$ and $w = u + v\jmath$. The left shows the mapping $w(s) = s^2$. The right shows the lower branch of the inverse $s(w) = \sqrt{w}$, as shown in Fig. 1.15.* 106

3.3 *On the left is $w(s) = \cos(\pi z)$ and on the right is the Bessel function $J_0(\pi z)$, which is similar to $\cos(\pi z)$, except the zeros are distorted away from $s = 0$ by a small amount due to the cylindrical geometry. The Bessel function is the solution to the wave equation in cylindrical coordinates while the cos is the solution in rectangular coordinates. The zeros in the function are the places where the pinned boundary condition is satisfied (where the string is restricted, by the boundary, from moving). The Bessel function $J_1(\pi z)$ is similar to $\sin(\pi z)$ in a like manner.* 106
3.4 *On the left is the function $w(z) = \tan(z)$. On the right is the inverse $s = \tan^{-1}(w)$. Of special interest is `zviz atan(i*Z) (i/2)*log((1+Z)./(1-Z))`.* 107
3.5 *A schematic representation of a 2-port ABCD electro-mechanic system using Hunt parameters $Z_e(s)$, $z_m(s)$, and $T(s)$: electrical impedance, mechanical impedances, and transduction coefficient (Hunt, 1952; Kim and Allen, 2013). Also $V(f)$, $I(f)$, $F(f)$, and $U(f)$ are the frequency domain voltage, current, force, and velocity respectively. Notice how the matrix method 'factors' the 2-port model into three 2×2 matrices. This allows one to separate the physical modeling from the algebra. It is a standard impedance convention that the flows $I(f), U(f)$ are always defined into the port. Thus it is necessary to apply a negative sign on the velocity $-U(f)$ so that it has an outward flow, to feed the next cell with an inward flow. Replace Φ with V.* . . . 108

4.1 The zeta function arguably the most important of the special functions of analysis because it connects the primes to analytic function theory in a fundamental way. 122
4.2 $\angle\zeta(z)$: Red $\Rightarrow \angle\zeta(z) < \pm\pi/2$ 122
4.3 *Example of a signal flow diagram for the composition of signals $z = g \circ f\ (x)$ with $y = f(x)$ and $z = g(y)$.* 122

List of Tables

1.1 *Three streams followed from Pythagorean theorem: Number Systems (Stream 1), Geometry (Stream 2) and Infinity (Stream 3).* 21
1.2 *The Fundamental theorems of mathematics* 28
1.3 Comparison between Euclidean Geometry and Algebra. The concepts that are included in Euclidean geometry are under the column *Geometry*. Those under the column *Algebra* are unique to Algebra (not present in Geometry). The third column are uncertain. 49
1.4 *The following table provides a short table of important Fourier Transforms. Note $a > 0 \in \mathbb{R}$ has units [rad/s]. To flag this necessary condition, we use $|a|$ to assure this condition will be met. The other constant $T_0 \in \mathbb{R}$ [s] has no restrictions, other than being real. Complex constants may not appear as the argument to a delta function, since complex numbers do not have the order property.* 59
1.5 *The following table provides a brief table of important Laplace Transforms. Assume that: $f(t), \delta(t), u(t), rec(t), T_0, e, \in \mathbb{R}$ and $F(s), G(s), s, a \in \mathbb{C}$.* 60
1.6 *Summary of the fundamental theorems of integral calculus, each of which deals with integration. There are at least four theorems related to scalar calculus, and four more to vector calculus.* . . . 63

3.1 *There are several ways of indicating the quasi-static (QS) approximation. For network theory there is only one lattice constant a, which must be much less than the wavelength (wavelength constraint). These three constraints are not equivalent when the object may be a larger structured medium, spanning many wavelengths, but with a cell structure size much less than the wavelength. For example, each cell could be a Helmholtz resonator, or an electromagnetic transducer (i.e., an earphone).* 113

4.1 Physical interpretation of the Riemann zeta function 120

5.1 *The three vector operators manipulate scalar and vector fields, as indicated here. The* gradient *converts scalar fields into vector fields. The divergence eats vector fields and outputs scalar fields. Finally the curl takes vector fields into vector fields.* 124

A.1 Double-bold notation for the types of numbers. (#) is a page number. 132

Bibliography

Beranek, L. L. (1954). *Acoustics*. McGraw–Hill Book Company, Inc., 451 pages, New York.

Beranek, L. L. and Mellow, T. J. (2012). *Acoustics: Sound Fields and Transducers*. Academic Press - Elsevier Inc., Waltham, MA.

Boas, R. (1987). *An invitation to Complex Analysis*. Random House, Birkhäuser Mathematics Series.

Bode, H. (1945). *Network Analysis and Feedback Amplifier Design*. Van Nostrand, New York.

Boyer, C. and Merzbach, U. (2011). *A History of Mathematics*. Wiley.

Brillouin, L. (1953). *Wave propagation in periodic structures*. Dover, London. Updated 1946 edition with corrections and added appendix.

Brillouin, L. (1960). *Wave propagation and group velocity*. Academic Press. Pure and Applied Physics, v. 8, 154 pages.

Brune, O. (1931a). *Synthesis of a finite two-terminal network whose driving point impedance is a prescribed function of frequency*. PhD thesis, MIT. `http://dspace.mit.edu/bitstream/handle/1721.1/10661/36311006.pdf`.

Brune, O. (1931b). Synthesis of a finite two-terminal network whose driving point impedance is a prescribed function of frequency. *J. Math and Phys.*, 10:191–236.

Calinger, R. S. (2015). *Leonhard Euler: Mathematical Genius in the Enlightenment*. Princeton University Press, Princeton, NJ, USA.

Campbell, G. (1903). On loaded lines in telephonic transmission. *Phil. Mag.*, 5:313–331. See Campbell22a (footnote 2): In that discussion, ... it is tacitly assumed that the line is either an actual line with resistance, or of the limit such with R=0.

Campbell, G. (1922). Physical theory of the electric wave filter. *Bell System Tech. Jol.*, 1(2):1–32.

Campbell, G. (1937). *The collected papers of George Ashley Campbell*. AT&T, Archives, 5 Reiman Rd, Warren, NJ. Forward by Vannevar Bush; Introduction by E.H. Colpitts.

Carlin, H. J. and Giordano, A. B. (1964). *Network Theory, An Introduction to Reciprocal and Nonreciprocal Circuits*. Prentice-Hall, Englewood Cliffs NJ.

Fagen, M., editor (1975). *A history of engineering and science in the Bell System — The early years (1875-1925)*. AT&T Bell Laboratories.

Feynman, R. (1970). *Feynman Lectures On Physics*. Addison-Wesley Pub. Co.

Fine (2007). *Number Theory: An Introduction via the Distribution of Primes*, chapter 4, pages 133–196. Birkhäuser Boston, Boston, MA.

Fletcher, N. and Rossing, T. (2008). *The Physics of Musical Instruments*. Springer New York.

Fry, T. (1928). *Probability and its engineering uses*. D. Van Nostrand Co. INC., Princeton NJ.

Goldstein, L. (1973). A history of the prime number theorem. *The American Mathematical Monthly*, 80(6):599–615.

Gray, J. (1994). On the history of the Riemann mapping theorem. *Rendiconti Del Circolo Matematico di Palemrmo*, Serie II(34):47–94. Supplemento (34): 47–94, MR 1295591," Citation from `https://en.wikipedia.org/wiki/Riemann_mapping_theorem`.

Greenberg, M. D. (1988). *Advanced Engineering Mathematics*. Preentice Hall, Upper Saddle River, NJ, 07458.

Grosswiler, P. (2004). Dispelling the alphabet effect. *Canadian Journal of Communication*, 29(2).

Helmholtz, H. L. F. (1863a). *On the sensations of tone*. Dover (1954), 300 pages, New York.

Helmholtz, H. v. (1863b). Ueber den einfluss der reibung in der luft auf die schallbewegung. *Verhandlungen des naturhistoisch-medicinischen Vereins zu Heildelberg*, III(17):16–20. Sitzung vom 27.

Hunt, F. V. (1952). *Electroacoustics, The analysis of transduction, and its historical background*. The Acoustical Society of America, 260 pages, Woodbury, NY 11797.

Kennelly, A. E. (1893). Impedance. *Transactions of the American Institute of Electrical Engineers*, 10:172–232.

Kim, N. and Allen, J. B. (2013). Two-port network analysis and modeling of a balanced armature receiver. *Hearing Research*, 301:156–167.

Kim, N., Yoon, Y.-J., and Allen, J. B. (2016). Generalized metamaterials: Definitions and taxonomy. *J. Acoust. Soc. Am.*, 139:3412–3418.

Kirchhoff, G. (1868). On the influence of heat conduction in a gas on sound propagation. *Ann. Phys. Chem.*, 134:177–193.

Lamb, H. (1932). *Hydrodynamics*. Dover Publications, New York.

Lighthill, S. M. J. (1978). *Waves in fluids*. Cambridge University Press, England.

Lin, J. (1995). The Needham Puzzle: Why the Industrial Revolution did not originate in China. In Christiansen, B., editor, *Economic Behavior, Game Theory, and Technology in Emerging Markets*, volume 43(2), chapter Economic Development and Cultural Change, pages 269–292. PNAS and IGI Global, 701 E Chocolate Ave., Hershey, PA. DOI:10.1073/pnas.0900943106 PMID:19131517, DOI:10.1086/452150.

Mason, W. (1928). The propagation characteristics of sound tubes and acoustic filters. *Phy. Rev.*, 31:283–295.

McMillan, E. M. (1946). Violation of the reciprocity theorem in linear passive electromechanical systems. *Journal of the Acoustical Society of America*, 18:344–347.

Montgomery, C., Dicke, R., and Purcell, E. (1948). *Principles of Microwave Circuits*. McGraw-Hill, Inc., New York.

Morse, P. and Feshbach, H. (1953). *Methods of theoretical physics. Vol. I. II*. McGraw-Hill, Inc., New York.

Morse, P. M. (1948). *Vibration and sound*. McGraw Hill, 468 pages, NYC, NY.

Newton, I. (1687). *Philosophi_ Naturalis Principia Mathematica*. Reg. Soc. Press.

Olson, H. F. (1947). *Elements of Acoustical Engineering (2^n Edition).* D. Van Nostrand Company, 539 pages.

Pierce, A. D. (1981). *Acoustics: An Introduction to its Physical Principles and Applications.* McGraw-Hill, 678 pages, New York.

Pipes, L. A. (1958). *Applied Mathematics for Engineers and Physicists.* McGraw Hill, NYC, NY.

Ramo, S., Whinnery, J. R., and Van Duzer, T. (1965). *Fields and waves in communication electronics.* John Wiley & Sons, Inc., New York.

Rayleigh, J. W. (1896). *Theory of Sound, Vol. I 480 pages, Vol. II 504 pages.* Dover, New York.

Schelkunoff, S. (1943). *Electromagnetic waves.* Van Nostrand Co., Inc., Toronto, New York and London. 6th edition.

Schwinger, J. S. and Saxon, D. S. (1968). *Discontinuities in waveguides : notes on lectures by Julian Schwinger.* Gordon and Breach, New York, United States.

Sommerfeld, A. (1949). *Partial differential equations in Physics, Lectures on Theoretical Physics, Vol. I.* Academic Press INC., New York.

Sommerfeld, A. (1952). *Electrodynamics, Lectures on Theoretical Physics, Vol. III.* Academic Press INC., New York.

Stillwell, J. (2002). *Mathematics and its history; Undergraduate texts in Mathematics; 2d edition.* Springer, New York.

Stillwell, J. (2010). *Mathematics and its history; Undergraduate texts in Mathematics; 3d edition.* Springer, New York.

Van Valkenburg, M. (1960). *Introduction to Modern Network Synthesis.* John Wiley and Sons, Inc., New York.

Van Valkenburg, M. (1964a). *Network Analysis second edition.* Prentice-Hall, Englewood Cliffs, N.J.

Van Valkenburg, M. E. (1964b). *Modern Network Synthesis.* John Weily & Sons, Inc., New York, NY.

von Helmholtz, H. (1978). On integrals of the hydrodynamic equations that correspond to vortex motions. *International Journal of Fusion Energy*, 1(3-4):41–68. English translation 1978 of Helmholtz 1858 paper, translated by Uwe Parpart. The image is on page 41.

Walsh, J. (1973). History of the Riemann mapping theorem. *The American Mathematical Monthly*, 80(3):270–276,. DOI: 10.2307/2318448.

Index

3-vectors, 50

analytic, 43
analytic function, 43, 134
analytic geometry, 48, 50
analytic series, 43

Bézout's theorem, 46
Bardeen, 45
Beranek, 80
Bethe, 45
Bhâskara II, 37
bilinear transformation, 45, 52, 55, 104, 105, 109, 133
Boas, 80
Bode, 80
Brahmagupta, 16, 22, 37, 79
Bramagupta, 37
branch cut, 69, 105, 106
Brillouin, 45, 80
Brillouin zones, 66, 95

Cauchy, 29, 53, 69
causality, 29, 57, 62, 107, 109
Chebyshev, 28
chord–tangent methods, 40, 53, 54, 81, 86
codomain, 44
complex analytic, 44, 68, 69, 105
complex analytic function, 44, 53–55
complex analytic functions, 69
composition, 21, 37, 48, 50, 52–55
conservation of energy, 14, 44, 109–111
conservative field, 67
coprime, 22, 23, 34, 82
cross-product, 50

d'Alembert, 21, 42, 66, 73
Descartes, 19, 21, 93
DFT, 57
Dicke, 80
Diophantus, 15, 16, 37, 40
dispersion relation, 66
domain, 44
domain coloring, 55
dot product, 47
driving-point impedance, 108
DTFT, 57

eigenfunction expansion, 40
elimination, 48
encryption, 13
entire function, 43
entire functions, 65
entropy, 32, 116
Euclid, 23, 33
Euclid's formula, 15, 16, 21, 26, 36
Euclidean algorithm, 34, 82
Euler, 18–20, 26, 53, 72, 73, 93
Ewald, 45
extended complex plane, 54

factoring, 42
Fermat, 19, 21, 37, 93
Feshbach, 80
Feynmann, 80
FFT, 57
Fibonacci sequence, 40
Fourier Series, 57
Fourier transform, 56
frequency domain, 29, 42, 56, 57, 68
Fry, 80
function, circular, 44
function, inverse, 44
function, periodic, 44
functions, causal, 57
functions, periodic, 57
Fundamental Theorem of Algebra, 46
Fundamental Theorem of Complex Calculus, 68
Fundamental theorems, 27–29, 46, 63, 68

Galileo, 14, 19–21, 41
Gauss, 29, 69, 82
Gaussian elimination, 51
geometric series, 43
Gibbs phenomenon, 58, 61
Green, 14, 29, 80

Heaviside, 14, 127
Heisenberg, 45
Helmholtz, 13, 14, 21, 29, 75, 78

Helmholtz portrait, 76
Helmoltz, 80
Humpty Dumpty, 14
Hunt, 80
Huygens, 19, 21

impedance, 20, 52, 66, 67
impulse response, 133
inner-product, 50
intersection of curves, 29, 45, 48, 51, 86, 88, 93, 96, 100–103
intersection of sets, 48
intersetsion, 48

Kepler, 20, 41
Kirchhoff, 29, 78

Lagrange, 37
Lamb, 80
Laplace frequency, 57, 59
Laplace transform, 28, 29, 56, 67, 72
lattice vector, 95
laws of gravity, 41
Leibniz, 29
Lighthill, 80
linearity, 62, 107, 109
Lord Kelvin, 14

Möbius transformation, 104
M obius transformation, 52
Mason, Warren P., 78
Maxwell, 14, 29
Mersenne, 13, 19, 21
Montgomery, 80
Morse, 80

network postulates, 42, 59, 61, 62, 67, 107
Newton, 13, 14, 19–21, 37, 41, 46, 53, 73, 79, 93, 98
Newton's laws of motion, 41
norm, 49

Ohm's law, 52
Olson, 80

passive, 62, 107, 109
Pauli, 45
Pell's equation, 37, 40
Pipes, 80
point at infinity, 54, 55
poles, 43, 44, 61, 133
power, 97
power series, 43
prime, 22, 26, 27, 33, 72, 79, 82
Principia, 20, 41, 63
propagation constant, 95
Prucell, 80
Pythagorean theorem, 21
Pythagorean triplets, 16
Pythagoreans, 13, 17, 19, 23, 26, 37, 40

quasi-static, 62, 107
quasi-statics, 77, 78
quasistatics, 112

Ramo, 80
Rayleigh, 14, 78, 80
Rayleigh reciprocity, 62, 107
region of convergence, 43, 69
residue, 43
residue expansion, 43
reversability, 107
reversibility, 62, 112
RHP, 66
Riemann, 69
Riemann sheet, 69, 105, 106
Riemann Zeta Function, 72, 73, 94
right half plane, 66
ROC, 43, 69
roots, 42

Schwarz, 96
signals, 56
Sommerfeld, 45, 80
spacial-invariant, 62, 107
spatial invariant, 112
speed of light, 42
speed of sound, 41
Stokes, 14, 29, 80
stream 1, 20, 21, 27
stream 2, 20, 27, 29, 41
stream 3, 21, 27, 29, 40, 63, 73
symmetry, 57, 62
systems, 56

Taylor Series, 68
Taylor series, 43, 44, 69
Taylor series formula, 69
time domain, 56, 57
time invariant, 111
time-invariant, 62, 107
transmission line, 112

unit-vectors, 50

vector wave equation, 95

vectors, 48
visco-thermal losses, 78

wave equation, 41, 95
wave number, 42, 66, 95
wave number with losses, 78
wavelength, 41, 66

z-transform, 40, 57
Zeno's paradox, 63
zeros, 61, 133
zviz.m, 55