THE GREAT COURSES®

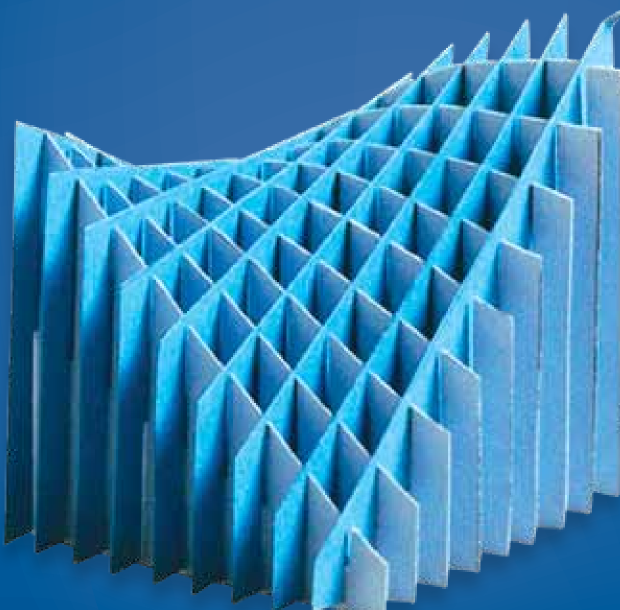
Topic
Science & Mathematics

Subtopic
Mathematics

The Art and Craft of Mathematical Problem Solving

Course Guidebook

Professor Paul Zeitz
University of San Francisco



PUBLISHED BY:

THE GREAT COURSES
Corporate Headquarters
4840 Westfields Boulevard, Suite 500
Chantilly, Virginia 20151-2299
Phone: 1-800-832-2412
Fax: 703-378-3819
www.thegreatcourses.com

Printed in the United States of America

Paul Zeitz, Ph.D.

Professor of Mathematics
University of San Francisco

Paul Zeitz is Professor of Mathematics at the University of San Francisco. He majored in History at Harvard University and received a Ph.D. in Mathematics from the University of California, Berkeley, in 1992, specializing in Ergodic Theory. Between college and graduate school, he taught high school mathematics in San Francisco and Colorado Springs.

One of Professor Zeitz's greatest interests is mathematical problem solving. He won the USA Mathematical Olympiad (USAMO) and was a member of the first American team to participate in the International Mathematical Olympiad (IMO) in 1974. Since 1985, he has composed and edited problems for several national math contests, including the USAMO. He has helped train several American IMO teams, most notably the 1994 "Dream Team," which was the first—and heretofore only—team in history to achieve a perfect score. This work, and his experiences teaching at USF, led him to write *The Art and Craft of Problem Solving* (Wiley, 1999; 2nd ed., 2007).

Professor Zeitz has also been active in events for high school students. He founded the San Francisco Bay Area Math Meet in 1994; cofounded the Bay Area Mathematical Olympiad in 1999; and currently is the director of the San Francisco Math Circle, a program that targets middle and high school students from underrepresented populations.

Professor Zeitz was honored in March 2002 with the Award for Distinguished College or University Teaching of Mathematics from the Northern California Section of the Mathematical Association of America (MAA), and in January 2003, he received the MAA's national teaching award, the Deborah and Franklin Tepper Haimo Award. ■

Table of Contents

INTRODUCTION

LECTURE GUIDES

Table of Contents

Table of Contents

The Art and Craft of Mathematical Problem Solving

Scope:

This is a course about mathematical problem solving. The phrase "problem solving" has become quite popular lately, so before we proceed, it is important that you understand how I define this term.

I contrast problems with exercises. The latter are mathematical questions that one knows how to answer immediately: for example, "What is 3 + 8?" or "What is 38^{74}?" Both of these are simple arithmetic exercises, although the second one is rather difficult, and the chance of getting the correct answer is nil. Nevertheless, there is no question about how to proceed.

In contrast, a problem is a question that one does not know, at the outset, how to approach. This is what makes mathematical problem solving so important, and not just for mathematicians. Arguably, all pure mathematical research is just problem solving, at a rather high level. But the problem-solving mind-set is important for all who take learning seriously, especially lifelong learners. Much of the current craze in brain strengthening focuses merely on exercises. These are not without merit—indeed, mental exercise is essential for everyone—but they miss out on a crucial dimension of intellectual life. Our brains are not just for doing crosswords or sudoku—they also can and should help us with intensive contemplation, open-ended experimentation, long wild goose chases, and moments of hard-earned triumph. That is what problem solving is all about.

An analogy that I frequently use compares an exerciser to a gym rat and a problem solver to a mountaineer. The latter's experience is riskier, messier, dirtier, less constrained, less certain, but much more fun. For those of you who prefer more civilized pursuits, consider 2 ways to learn Italian. One involves toiling over grammar exercises and translations of texts. The other method is to spend a few months, perhaps after a short bit of preparation, in a small town in Italy where no one else speaks English. Again, the latter approach is messier but fundamentally richer.

Becoming a good problem solver requires new skills (mathematical as well as psychological) and patient effort. My pedagogical philosophy is both experiential and analytic. In other words, you cannot learn problem solving without working hard at lots of problems. But I also want you to understand what you are doing at as high a level as possible. We will break down the process of solving a problem into investigation, strategy, tactics, and finer-grained tools, and we will often step back to discuss not just how we solved a problem but why our methods worked.

Problems, by definition, are hard to solve. Solving problems requires investigation, and successful investigations need strategies and tactics. Strategies are broad ideas, often not just mathematical, that facilitate investigation. Some strategies are psychological, others organizational, and others simply commonsense ideas that apply to problems in any field. Tactics are more narrowly focused, mostly mathematical ideas that help solve many problems that have been softened by good strategy. Additionally, there are very specialized techniques, called tricks by some, that I call tools.

This course is devoted to the systematic development of investigation methods, strategies, and tactics. Besides this "problemsolvingology," I will introduce you to mathematical folklore: classic problems as well as mathematical disciplines that play an important role in the problem-solving world. For example, no course on problem solving is complete without some discussion of graph theory, which is an important branch of math on its own but is also a very accessible laboratory for exploring problem-solving themes. Many of the lectures will include small amounts of new mathematics that we will build up and stitch together as the course progresses. The topics are largely drawn from discrete mathematics (graph theory, integer sequences, number theory, and combinatorics), because this branch of math does not require advanced skills such as calculus. That does not mean it is easy, but we will move slowly and develop new ideas carefully.

A small but important part of the course explores the culture of problem solving. I will draw on my experience as a competitor, coach, and problem writer for various regional, national, and international math contests, to make the little-known world of math Olympiads come to life. And I will discuss the recent educational reform movement (in which I am a key player) to bring Eastern European–inspired mathematical circles to the United States.

Problem solving is not a vertically organized discipline; it is not something that one learns in a linear fashion. Thus the overall organization of this course has a recursive, spiral nature. The first few lectures introduce the main ideas of strategy and tactics, which then are revisited and illuminated by different examples. We will often return to and refine previously introduced ideas. Overall, the topics get more complex toward the end of the course, but the underlying concepts do not really change. An analogy is a theme and variations musical piece, where the main theme is introduced with a slow, stately rhythm and later ends in complex avant-garde interpretations. By the end of the course, you should understand the main theme (the basic and powerful strategies and tactics of problem solving) quite well because you had to struggle with the complex interpretations (the advanced folklore problems that used the basic strategies in novel ways).

Problem solving is not just solving math problems. It is a mental discipline; successful investigations demand concentration and patient contemplation that few of us can do, at least at first. Also, problem solving is an aesthetic discipline—in other words, an art—where we create and contemplate objects of elegance and beauty. I hope that you enjoy learning about this wonderful subject as much as I have! ■

Problems versus Exercises

Lecture 1

There is always a porous boundary between problem and exercise, but a problem by its very nature requires investigation, sometimes very intense and sustained investigation. The investigation of a problem employs strategies and tactics, and that's what this course is about.

In this introductory lecture, we define the main entity that we will study in this course: problems. Problems, by definition, are difficult, and our investigation of them cannot proceed without organized strategies and tactics. Indeed, our course focuses on 3 things: investigation, strategies, and tactics. Problem solving is at the heart of mathematics. It is not just a way of thinking about math but is an intellectual lifestyle with its own mathematical folklore and culture. Most of our learning will be by example. Almost every lecture will revolve around one or more problems. You, the viewer, will need to use the Pause button and pencil and paper. This lecture will include several fun problems not requiring any special mathematical skills, but in later lectures, the problems will be more complex.

Much of the current craze in brain strengthening focuses on exercises. This is not without merit, but our brains should also work on intensive contemplation and open-ended experimentation.

Who am I, and what do I do? I have been a professor at the University of San Francisco since 1992. I received my Ph.D. in Mathematics from the University of California, Berkeley, specializing in Ergodic Theory, a sort of abstract probability theory. I first learned about problem-solving mathematics as a mathlete at Stuyvesant High School in New York City. I won national awards and participated in international competitions, some of the most formative experiences of my life. I currently teach problem-solving mathematics to high school and middle school teachers, run math clubs, write problems for math competitions, and train mathletes.

What do I mean by “problems versus exercises”? This course is devoted to the study of problem-solving investigation and the strategies and tactics that facilitate it. An exercise is a mathematical question that you know immediately how to answer. You may not answer it correctly, and it may not be easy, but there is no doubt about how to proceed. In contrast, a problem is a mathematical question that you do not know how to answer, at least initially. Problems require investigation, which employs strategies and tactics.

Why study problem solving? It helps you develop a problem solver’s mind-set, which involves both heightened mental discipline and an explorer’s attitude. Much of the current craze in brain strengthening focuses on exercises. This is not without merit, but our brains should also work on intensive contemplation and open-ended experimentation. A good problem solver is intellectually playful and fearless.

The Pill Problem

- This is a recreational problem, in that it requires little or no formal mathematics to solve.
- For 10 days, you must take exactly 1 *A* and 1 *B* pill at noon, or you will die. The pills are indistinguishable! All goes well until day 3. On this day, you shake 1 *A* and 2 *B* pills into your hand and do not know which is which. Can you survive? If so, how?
- We will discuss the solution later in this lecture. I urge you to work on this and future problems on your own, before we present the solutions. It is important to make good use of your Pause button in this course!

In this course, you will learn “problemsolvingology” in an analytic way, by systematically developing strategies and tactics. The goal is not for you to learn tricks that solve problems but to develop a mind-set that facilitates persistent, creative investigation of problems. Problem solving is also an art, with folklore and “morals.” These are classic problems whose content and solution tell important stories that teach us about problem solving and show us the beautiful interconnectedness of mathematics. From time to time, we change our focus from problem solving itself to exploring the problem-solving culture. We do this not just because it is interesting but because it is essential.

Learning to become a better problem solver requires a change in attitude. There are 2 things that you may be unaccustomed to: (1) Timescale. Most of us are not used to thinking hard for more than a minute or so. Developing expert skill requires 10,000 hours, which is somewhat more than a year without breaks. (2) Failure. One needs to cultivate an attitude that investigation is always worthwhile, even if it does not lead to solution. ■

The Subway Love-Triangle Problem

- Anna lives near a subway station located at the middle of the line. She has 2 boyfriends, Bert and Curt, who live at either end of the subway line. The men demand that she choose between them. She proposes to randomly show up at her station every day for a month and take the first train that comes. Whichever boyfriend she visits the most will be the one she chooses. The trains come every 20 minutes in each direction, every day, 24 hours per day. Is Anna's scheme fair? (You can assume that the month has 31 days, so there is no possibility of a tie.)

- Strategy: Get your hands dirty! There is no way to approach this problem without sitting down and writing a train schedule.

- Consider the following schedule. Anna's scheme seems fair.

Northbound	Southbound
12:00	12:10
12:20	12:30
12:40	12:50

- But look at this schedule!

Northbound	Southbound
12:00	12:05
12:20	12:25
12:40	12:45

- The second schedule favors the boyfriend at the northern terminus, because there is only a 5-minute window every 20 minutes during which the southbound train comes first. So the northern boyfriend has a 3-to-1 advantage. Clearly the schedule can be designed to give one boyfriend an even greater advantage. So the scheme is not fair; it is less fair than flipping a coin.

- Notice how this problem could not be investigated, let alone solved, without getting your hands dirty. Simple, confident experimentation is often the key to investigating any problem.

Solution to the Pill Problem

- Wishful thinking suggests trying to get a dosage that is at least closer to the correct dose. What you have is both unknown and unbalanced. Adding an additional *A* pill gives you a balanced dosage. Here we are using the tactic of imposing symmetry on our problem; we will see later that this is a very powerful idea.

- Now we have a balanced dose but one too big. Wishful thinking: Imagine that you are a giant (twice as large as a normal person). Then you would be done! This immediately suggests the solution: Cut the double-sized dose in half (or grind up and divide in half), producing 2 days of correct dosages to get us back on track.

- What we did was use wishful thinking and symmetry. Why we did it was to increase our ability to investigate and move our solution toward a configuration with more balance and possibly more information.

- Wishful thinking, get hands dirty, and symmetry: the core of the course. We will use these ideas over and over and add new and powerful ideas.

Suggested Reading

Lehoczky and Rusczyk, *The Art of Problem Solving*.

Vakil, *A Mathematical Mosaic.*

Zeitz, *The Art and Craft of Problem Solving*, chap. 1.

Questions to Consider

1. You are in the downstairs lobby of a house. There are 3 switches, all in the off position. Upstairs, there is a room with a light bulb that is turned off. One and only one of the 3 switches controls the bulb. You want to discover which switch controls the bulb, but you are only allowed to go upstairs once. How do you do it? (No fancy strings, telescopes, and so on are allowed. You cannot see the upstairs room from downstairs. The light bulb is a standard 60-watt bulb.)

2. You are locked in a 50- × 50- × 50-foot room that sits on 100-foot stilts. There is an open window at the corner of the room, near the floor, with a strong hook cemented into the floor by the window. So if you had a 100-foot rope, you could tie one end to the hook and climb down the rope to freedom. (The stilts are not accessible from the window.) There are two 50-foot lengths of rope, each cemented into the ceiling, about 1 foot apart, near the center of the ceiling. You are a strong, agile rope climber, good at tying knots, and you have a sharp knife. You have no other tools (not even clothes). The rope is strong enough to hold your weight, but not if it is cut lengthwise. You can survive a fall of no more than 10 feet. How do you get out alive?

Strategies and Tactics

Lecture 2

So far we've seen the difference between problems and exercises, and we've solved several problems using 2 very, very simple commonsense strategies: wishful thinking and get your hands dirty. What we'll do in this lecture is develop a careful definition of strategies and tactics, which is what we need to proceed with problem-solving investigations, and we'll look at an analytic approach to problem solving. … Along the way, of course, we'll solve some classic problems using several different approaches. We'll do some where we concentrate on strategies and others where we're concentrating more on tactics.

The main goal of this lecture is an overview of the analytic approach to problem solving, carefully defining the notions of strategy and tactics that were introduced in Lecture 1. All problems require investigation, and to facilitate investigation, we need many resources. These are collectively called strategies, and we will mention several but only focus on a few during this lecture. Tactics have a narrower and more mathematical focus and are used, generally, at a later stage of investigation, often providing the key to solution. In this lecture, we will look at 2 classic problems. An important aspect of this lecture is the stress on the need to deconstruct solutions of problems, to understand not just how, but also why, we could solve them.

Strategies are ideas, mostly nonmathematical, that facilitate investigation of almost any problem. Tactics are more narrowly focused, mostly mathematical, ideas that help one solve many problems that have been "softened" by good strategy.

Let's first look at solving standard story problems. Here is a typical story problem, the sort that freshmen business math students struggle with. Such problems are too simple for this course—they are really more like exercises for us.

Omnicorp Story Problem

- Omnicorp is sending 216 of its employees to a productivity enhancement conference, which involves an overnight stay at a hotel. The employees will stay in either 2-person or 4-person rooms, which cost \$100 and \$150 each, respectively. The housing budget for the conference is \$9600. How many rooms of each kind will be reserved?

- I employ the fantasy answer method, which applies the wishful thinking strategy.

- The method is simple: Just pretend that you have actually solved the problem, and write down an answer.

- The only catch is that the answer must make sense in terms of units.

- The student then takes the fantasy answer, which is entirely concrete, and reads the problem again, trying to figure out why the answer is not correct. This, hopefully, will lead to understanding how the problem can be set up algebraically.

- The overarching idea is the utility of wishful thinking as a means of facilitating investigation.

How does the analytic approach to problem solving work? Use strategies to begin and facilitate the investigation. Next, deploy tactics to continue the investigation and hopefully yield a solution. Use tools (a.k.a. tricks) sparingly, at the narrowest focal level. Strategies are ideas, mostly nonmathematical, that facilitate investigation of almost any problem. Tactics are more narrowly focused, mostly mathematical, ideas that help one solve many problems that have been “softened” by good strategy. Tools have very narrow applications—and very impressive results when used correctly. ■

The Census Taker Problem

- A classic example, one of my favorites, that uses the get hands dirty strategy.
- A census taker knocks on a door and asks the woman inside how many children she has and how old they are.
- "I have 3 daughters, their ages are whole numbers, and the product of the ages is 36," says the mother. "That is not enough information," responds the census taker. "I would tell you the sum of their ages, but you would still be stumped," says the mom. "I wish you would tell me something more," begs the census taker. The mom responds, "Okay, my oldest daughter, Annie, likes dogs."
- What are the ages of the 3 daughters?
- The problem consists of what appear to be too few clues. But just start with the first clue, and get your hands dirty!
- The first clue says, "The product of the ages is 36." There are only a few possible ways you can multiply 3 whole numbers to get 36; it makes sense to systematically list them.
- The second clue says that if she told him the sum, he would still be stumped. It makes sense, then, to compute the sums of the possibilities you have listed.
- Now we see what is going on. Two of the possibilities—9, 2, 2 and 6, 6, 1—have the same sum (13), and these are the only 2 with the same sum. So we know that it must be one of these, for otherwise we would not have been stumped.
- Finally, we understand the final clue, which merely indicates that there is an oldest child. So the answer is 9, 2, 2.

The Frog Problem

- The frog problem is a classic Russian math circle problem.
- Three frogs are situated at 3 of the corners of a square. Every minute, 1 frog is chosen to leap over another chosen frog, so that if you drew a line from the starting position to the ending position of the leaper, the leapee is at the exact midpoint.
- Will a frog ever occupy the vertex of the square that was originally unoccupied?
- How can we effectively investigate this problem?
- Graph paper allows us to attach numbers to the positions of the frogs. Once we have numbers, we can employ arithmetical and algebraic methods. Thus, place the frogs at $(0, 0)$, $(0, 1)$, and $(1, 1)$. The question now is, can a frog ever reach $(1, 0)$?
- Thinking about the appropriate venue for investigation is an essential starting strategy for any problem.
- Another investigative idea: Use colored pencils to keep track of individual frogs. This adds information, as it allows us to keep track of 1 frog at a time. Color the $(1, 1)$, $(0, 1)$, and $(0, 0)$ frogs red, blue, and green, respectively.
- Notice, by experimenting, that the red frog only seems to hit certain points, forming a larger (2-unit) grid.
- Some of the coordinates that the red frog hits are $(1, 1)$, $(1, 3)$, $(1, -1)$, $(-1, 1)$, $(-1, -1)$, and $(-1, -3)$. They are all odd numbers!
- Likewise, the blue frog only hits certain points on a 2-unit grid, including $(0, 1)$, $(2, 1)$, $(4, 1)$, and $(0, -1)$; these are all of the form (even, odd).

- Likewise, the green frog only hits (even, even) points.
- On the other hand, the missing southeast vertex was (1, 0), which has the form (odd, even). It seems as though it is impossible, but how can we formulate this in an airtight way?
- It is often very profitable to contemplate parity (oddness and evenness).
- The essential reason for this is that a parity focus reduces a problem from possibly infinitely many states to just 2.
- Parity involves the number 2. Where in this problem do we see this number? In doubling, because of the symmetry of the way the frogs leap. When the leaper jumps over the leapee, she adds twice the horizontal displacement to her original horizontal coordinate. The same holds for vertical coordinates.
- So when a frog jumps, its coordinates change by even numbers!
- For example, suppose the red (1, 1) frog jumps over the green frog at (0, 0). The horizontal and vertical displacements to the leapee are both −1 (since it is moving left and down), so the final change in coordinates will be −2. The horizontal coordinate will be $1 + -2 = -1$, and the vertical will also be −1.
- Suppose now that the red frog jumps over the blue frog, which is (0, 1). The horizontal displacement is +1, and the vertical displacement to the target is +2. So the new horizontal coordinate will be −1 (the starting value) $+ 2 \times 1 = +1$, and the new vertical coordinate will be −1 (the starting value) $+ 2 \times 2 = 3$. Thus the red frog jumps from (−1, −1) to (1, 3).

- In general, when a frog jumps, we will take its starting x-coordinate and add twice the horizontal displacement to its target. Likewise, we take its starting y-coordinate and add twice the vertical displacement to the target. These displacements may be positive, negative, or zero.

- In other words, you take the starting coordinates and add even numbers to them. But when you add an even number to something, its parity does not change!

- So the (odd, odd) frog—the red frog—is destined to stay at (odd, odd) coordinates, no matter what.

Suggested Reading

Polya, *How to Solve It.*

Zeitz, *The Art and Craft of Problem Solving*, chap. 2.

Questions to Consider

1. Write the numbers from 1 to 10 in a row and place either a minus or a plus sign between the numbers. Is it possible to get an answer of zero?

2. A group of jealous professors is locked up in a room. There is nothing else in the room but pencils and 1 tiny scrap of paper per person. The professors want to determine their average (mean, not median) salary so that each can gloat or grieve over his or her personal situation compared to the others. However, they are secretive people and do not want to give away salary information to anyone else. Can they determine the average salary in such a way that no professor can discover any fact about the salary of anyone but herself? For example, even facts such as "one professor earns less than $90,000" are not allowed.

The Problem Solver's Mind-Set

Lecture 3

There are 2 lovely quotes about concentration. If we want to paraphrase them, the first one says that concentration is a human virtue, and the second one says obsession begets concentration. You need to relax; you need to develop good work habits; and you need to find, most importantly, problems that are interesting to you, approachable by you, and addictive.

In this lecture, we discuss some of the mental tools needed for successful problem solving. A good problem solver needs concentration, confidence, and creativity—but how can one acquire these, and how can these attributes be enhanced? We explore some classic puzzlers and begin to develop some number theory ideas and to investigate a problem about the famous Fibonacci numbers.

The 3 *c*'s of mental attitude are concentration, creativity, and confidence. How do we enhance these qualities? Confidence is the least important, as it is derived from the other qualities, and creativity is the most elusive, so it is best to start by developing your concentration. You need to relax, develop good work habits, and find problems that are interesting, approachable, and addictive. As with any art or craft, you must set aside a quiet time and place for your work and start building your concentration. Start building up a stock of back-burner problems. Make good use of unstructured time.

Let's look at confidence and creativity by examining some truly creative mathematics. We will look at 2 tools. Recall that tools are narrowly focused mathematical ideas that solve certain types of problems. The first tool, Gaussian pairing, was made famous by Carl Gauss (1777–1855), who is universally recognized as one of the greatest mathematicians ever.

When Gauss was only 10, as legend has it, he was faced with the sum $1 + 2 + 3 + \cdots + 100$. How was he to compute it, in 1787, when there were no calculators? He simply paired the terms: $(1 + 100) + (2 + 99) + \cdots + (50 + 51)$. Thus the sum is $50(101) = 5050$.

The second tool is telescoping. We will apply it to a harder sum involving fractions:

$$\frac{1}{1\times 2}+\frac{1}{2\times 3}+\frac{1}{3\times 4}+\cdots+\frac{1}{99\times 100}$$

Do not let such a problem make you panic: Panicking is bad for your confidence and harms investigation. You must look for things that foster investigation. The wishful thinking strategy works for this reason: Pretending you have solved a problem—even an easier one—keeps you happy, and that keeps you thinking about the problem!

- A corollary of wishful thinking is the make it easier strategy. This strategy is common sense: If the current problem is too hard, make it easier by reducing its size or eliminating one or more of the elements that make it hard.

- In our case, that just means we should replace the 99-term sum with, say, a 1-term or 2-term sum. When we look at these smaller problems, it is easy to conjecture that the 99-term sum equals 99/100.

- Telescoping is the way to see why this conjecture is true.

- Write the terms as differences.

$$\begin{aligned} 1 - 1/2 &= 1/(1 \times 2) \\ 1/2 - 1/3 &= (3 - 2)/(2 \times 3) = 1/(2\times 3) \\ 1/3 - 1/4 &= (4 - 3)/(3 \times 4) = 1/(3\times 4) \end{aligned}$$

- Notice that for any k, $1/k - 1/(k + 1) = 1/[k(k + 1)]$.

- Thus all terms cancel (telescope) except the first and last, yielding the answer $1 - 1/100 = 99/100$.

How do you get more creative? The 3 *c*'s are inextricably linked: Concentration leads to confidence, which frees you to explore, which facilitates investigation and creativity. You need to set up a problem-solving routine. And think about peripheral vision: Many problems cannot be solved with direct focus. Many problems need to percolate in your unconscious. You need to cultivate a good supply of back-burners and get in the habit of *not* solving problems. The more you can cultivate a state of investigation and purposeful contemplation, the more powerful your mind will get.

People are endowed unequally with confidence, creativity, and power of concentration, but all of these are trainable skills.

We end with a fun open-ended problem designed to facilitate uninhibited investigation. There are no wrong answers. First, a little introduction to modular arithmetic: We write $a \equiv b \pmod{m}$, read "*a* is congruent to *b* modulo *m*," if $a - b$ is divisible by m. The nice thing about congruence is that it preserves addition, subtraction, and multiplication. For example, if $17 \equiv 2 \pmod 5$ and $8 \equiv 3 \pmod 5$, then $17 \times 8 \equiv 2 \times 3 \pmod 5$. You can think of congruence as a sort of myopia in which we lump the infinitude of integers into just a few categories. In the (mod 5) universe, there are only 5 types of numbers, those congruent to 0, 1, 2, 3, and 4 (mod 5). If we restrict ourselves to the (mod 2) universe, that is the same as only worrying about parity. The Fibonacci numbers are defined by $f_1 = f_2 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for *n* greater than 2. The first few terms are 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, and 144.

Your assignment: Investigate divisibility patterns of the Fibonacci numbers. For example, which Fibonacci numbers are even? Odd? Multiples of 3? Of 5?

- Experiment and conjecture! Don't worry about why at this point.

- Investigate by making it easier and getting your hands dirty.

- Start with mod 2 (parity). Then the sequence is 1, 1, 0, 1, 1, 0, 1, 1, 0, … . It is evident that every third one is even.

- Trying mod 3, we get 1, 1, 2, 0, 2, 2, 1, 0, 1, 1, 2, 0, … , and we see that every fourth one is a multiple of 3.

- Do more experiments, until you have some good conjectures. We will not worry about proofs yet.

- People are endowed unequally with confidence, creativity, and power of concentration, but all of these are trainable skills. It is possible to practice them and improve them—you just need to see lots of creativity in action, and you need lots of open-ended opportunity to experiment. ■

Suggested Reading

Gardner, *Aha!*

Honsberger, *Ingenuity in Mathematics*.

Zeitz, *The Art and Craft of Problem Solving*, sec. 2.1.

Questions to Consider

1. How do you bring a 1.5-meter sword onto a train if no baggage item with dimension greater than 1 meter is allowed?

2. One day Martha said, "I have been alive during all or part of 5 decades." Rounded to the nearest year, what is the youngest she could have been?

Searching for Patterns

Lecture 4

The moral of the story that we've seen is that uninhibited experimentation is lots of fun and it often leads to many fun conjectures, but mere pattern hunting is not enough if we do not understand the why behind what we see.

In this lecture, we step back a bit and examine the power of simple strategies that allow us to simplify problems, make numerical experiments, and develop conjectures. We also look at 2 cautionary examples that show that experimentation and conjecture is not always enough. The core of this lecture is the beginning of an investigation into trapezoidal numbers and a search for patterns in Pascal's triangle. This lecture is devoted to the search for patterns by getting one's hands dirty. We will look at several examples where this strategy succeeds, as well as ones where it is clearly not enough.

Perhaps the most important mathematical playground of all is Pascal's triangle.

It is helpful to build up a stock of knowledge to aid our receptiveness to discovery. You must be at least passively aware of some of the most important subsets of the integers. It is important to develop an obsession with numbers and sequences.

- Squares: 1, 4, 9, 16, 25, 36, 49, 64, 81, 100 … .
- Cubes: 1, 8, 27, 64, 125, 216, 343, 512, 729, 1000, … .
- Primes: 2, 3, 5, 7, 11, 13, 17, 19, 23, … .
- Powers of 2: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, … .
- Fibonacci sequence: 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144… .

- Factorials:

  $1! = 1,$
  $2! = 1 \times 2 = 2,$
  $3! = 1 \times 2 \times 3 = 6,$
  $4! = 1 \times 2 \times 3 \times 4 = 24,$
  $5! = 4! \times 5 = 120,$
  $6! = 720,$
  $7! = 5040.$

- Triangular numbers (called triangular because they can be rearranged as dots that form triangles):

  $1,$
  $1 + 2 = 3,$
  $1 + 2 + 3 = 6,$
  $1 + 2 + 3 + 4 = 10.$

Let's look at a problem involving a slightly more exotic sequence, the trapezoidal numbers. Find all positive integers that can be written as a sum of at least 2 consecutive positive integers. We call such numbers trapezoidal since when we depict them with dots, the pattern is trapezoidal. Examples:

- $6 = 1 + 2 + 3.$

- $36 = 11 + 12 + 13.$

An investigation of the first dozen or so trapezoidal numbers yields the conjecture that the powers of 2 (1, 2, 4, 8, 16, …) are not trapezoidal.

Perhaps the most important mathematical playground of all is Pascal's triangle. You should add to your passive stock of integer knowledge the first 10 or so rows. Pascal's triangle is defined by the funnel property: Each term is equal to the sum of the 2 above it. For example, $10 = 4 + 6$. We label the rows starting with row 0. We call the k^{th} number in row n "n choose k" and write it $\binom{n}{k}$. We call these binomial coefficients. We use the word "choose" in binomial coefficients because *n* choose *k* also has a combinatorial meaning—the number of ways to choose *k* things from a set of *n* objects.

Our goal for now is to find at least 5 interesting patterns in Pascal's triangle. Carefully write out rows 0 to 10, at least. The sums of the elements in each row are powers of 2. The alternating sums, however, are always 0. The hockey-stick pattern: For example, 1 + 4 + 10 = 15. Triangular numbers (1, 3, 6, 10, …). Fibonacci numbers even appear; it is easier to see this if we draw Pascal's triangle this way:

1					
1	1				
1	2	1			
1	3	3	1		
1	4	6	4	1	
1	5	10	10	5	1.

The sums of each southwest-to-northeast diagonal are Fibonacci numbers (e.g., 1 + 4 + 3 = 8). The coefficients of $(1+x)^n$ are the numbers in row *n*. For example, $(1+x)^4 = x^4 + 4x^3 + 6x^2 + 4x + 1$.

What about divisibility in Pascal's triangle? At the very least, we should investigate parity. When we count the number of odd terms in each row, we see that these numbers are not only even but seem to be powers of 2. But which power of 2? And what are we observing here? We will investigate this later.

There are limits to inductive reasoning. Here are 2 examples of why seeing a pattern is not sufficient if you do not understand why the pattern is there.

- Can a polynomial output nothing but primes? Consider $P(x) = x^2 + x + 41$. $P(x)$ is prime for all positive integers up to 40. But this numerical investigation distracts us from the *why* of the problem. $P(41) = 41^2 + 41 + 41$, which is obviously a multiple of 41!

- The 5 circles problem: If n points are placed on a circle and all pairs of points are joined by line segments, into how many regions is the circle divided? Assume that the points are in general position (i.e., no 3 lines intersect in a single point). Investigation quickly yields the sequence 1, 2, 4, 8, 16. The obvious conjecture is that the number of regions is 2^{n-1}, where n is the number of points. But a careful count of the 6-point circle yields only 31 regions!

The moral of the story is clear: You must understand what you are looking at. ■

Suggested Reading

Gardiner, *Discovering Mathematics.*

Graham, Knuth, and Patashnik, *Concrete Mathematics.*

Guy, "The Strong Law of Small Numbers."

Questions to Consider

1. For each positive integer n, find distinct positive integers x and y such that $1/x + 1/y = 1/n$.

2. Draw triangles with lattice point vertices. Count the number of lattice points in the interior (I) and boundary (B). Is there a formula relating these 2 numbers to the area (A)?

Closing the Deal—Proofs and Tools

Lecture 5

Investigation, as I've said before, trumps almost everything. Coupled with strategy and tactics, it's the paramount way to think about problems, but it's not always enough. Sometimes we need to close the deal with a creative logical argument or with a clever algebraic trick, such as proof by contradiction, direct proof, or even algorithmic proof.

This lecture focuses on closing the deal: turning your investigative ideas into rigorous arguments. We first develop the ideas of deductive proof and proof by contradiction. Many problems also require finely focused ideas known as tricks, or tools. We briefly discuss some of the most important of these, while strenuously arguing against their overuse (a common beginner's error).

Infinite processes such as sums and fractions are defined by looking at the finite version and considering what happens when they converge.

It is critical to become comfortable with mathematical logic. Math is fundamentally different from most other fields of inquiry, because things are usually either right or wrong. Types of statements include conjectures, lemmas, and theorems. Proofs are arguments that demonstrate the truth of a theorem. There are direct proofs and indirect proofs; let's look at an example of an indirect proof.

Indirect Proof for the Irrationality of $\sqrt{2}$

- Assume to the contrary that $\sqrt{2}$ is rational.
- Then $\sqrt{2} = a / b$, for some integers *a* and *b*. Algebraic manipulations yields $2b^2 = a^2$.

- Now count the 2s in the prime factorizations and use the fundamental theorem of arithmetic.

- The left-hand side has an odd number of 2s, while the right-hand side has an even number—a contradiction!

Now we can prove the conjecture about trapezoidal numbers from the last lecture.

- Recall that trapezoidal numbers are sums of at least 2 consecutive positive integers. Then if a number T is trapezoidal, there exist positive integers a, n, and l, for the starting value, number of terms, and ending value, respectively. For example, if $T = 18 = 3 + 4 + 5 + 6$, then $a = 3$, $n = 4$, and $\mathrm{l} = 6$.

- Using the Gaussian pairing trick, we can add up $a + (a + 1) + \cdots + \mathrm{l}$ and get the important formula $T = n(a + \mathrm{l})/2$. This makes sense, since it says that the sum is equal to the average value of the terms, times the number of terms.

- Thus we have the formula $T = n(a + a + n - 1)/2 = n(2a + n - 1)/2$.

- We want to show 2 things: that T cannot be a power of 2 and that if T is anything else, we can find an a and n that work.

    - For the first goal, since powers of 2 are totally even, it makes sense to think about parity.

    - If n is even, then $2a + (n - 1)$ is odd. Hence $T = (\text{even})(\text{odd})/2$.

    - If n is odd, then $2a + (n - 1)$ is even, but again $T = (\text{odd})(\text{even})/2$.

    - So in every case, T must have an odd factor! T cannot be a power of 2! Notice the importance of the penultimate step strategy here.

- In the other direction (showing that all non–powers of 2 work), let T be such a number. Then we have to find a and n, both positive, with $n > 1$, such that $T = n(2a + n - 1)/2$.

- Remove the fraction, getting $2T = n(2a + n - 1)$.

    - Since a is at least 1, the second factor of this expression is at least $n + 1$.

    - Thus, for any value of T that has at least one odd factor (besides 1), compute $2T$ and factor it into a product of an odd and an even. The smaller factor will be n, and the larger will be $2a + n - 1$, and we will be able to solve for a.

    - Example: $T = 10$. $2T = 20 = 4 \times 5$. So $n = 4$, and $2a + 3 = 5$, making $a = 1$.

The infinitude of primes is a classic proof by contradiction.

- Assume, to the contrary, that there are finitely many primes, ending with the prime number L.

- Define Q to equal the product of all these primes, plus 1. In other words, $Q = (2 \times 3 \times \cdots \times L) + 1$.

- Q cannot be prime; it is much bigger than L!

- But Q cannot be divisible by 2, 3, 5, or any other prime number.

- Thus there must be a prime that is not in the list 2, 3, 5, … , L—a contradiction!

- We conclude that there cannot be a finite number of primes.

Now let's apply these methods to the Fibonacci divisors problem. First we need a little tiny lemma. Lemma: If p is prime and ab is congruent to 0 (mod p), then $a = 0$ or $b = 0$. This lemma is not true for composites [2 ×

3 is congruent to 0 (mod 6)]. We will use this to prove our conjecture about Fibonacci divisors. For example, $F_4 = 3$, so we conjecture that every fourth Fibonacci is a multiple of 3. Modulo 3, the first 4 Fibonacci numbers are 1, 1, 2, and 0. But the fifth and sixth Fibonacci numbers are 2 and 2. The rest of the sequence behaves like the first terms, only now they are multiplied by 2! So the second 4 terms are equal to the first 4 terms, but multiplied by 2. Since 3 is a prime, the only way we will get a zero is when we multiply by zero. This argument will work for prime divisors only. But we have proven that if a prime p divides F_n, then p divides F_{2n}, F_{3n}, and so on.

Now let's focus on useful tools involving algebraic sums. The mother of all tools is telescoping. Sometimes we can take a sum and perturb it so that most of the new terms cancel with the old terms. Here is an example of a geometric series.

$$S = a + ar + ar^2 + ar^3 + ar^4$$
$$rS = ar + ar^2 + ar^3 + ar^4 + ar^5$$
$$S - rS = a - ar^5$$
$$S(1-r) = a - ar^5$$

In other words, $S(1 - r)$ turned the sum into a telescoping series. We can extend this to infinite geometric series, as long as r is less than 1 in absolute value. As n grows, the limiting value will be $S = a/(1-r)$.

The massage tool is less common than telescoping but quite powerful. It says, feel free to mess around with an expression to make it simpler for your particular context; do not worry if you have altered its value a little. Let's work an example: The harmonic series is the infinite sum 1 + 1/2 + 1/3 + 1/4 + 1/5 + $\cdots$. Does it converge or diverge? We will prove divergence by showing that we can make the sum arbitrarily large if we go out far enough.

- Key idea: $1/n \geq 1/m$ if $m \geq n$.

- Thus, $1/3 + 1/4 \geq 1/4 + 1/4$ (which equals 2/4, which simplifies to 1/2).

- Likewise, 1/5 + 1/6 + 1/7 + 1/8 ≥ 1/8 + 1/8 + 1/8 + 1/8 (which equals 4/8 and simplifies to 1/2).

- Hence if we look at the terms from 1/9 to 1/16, there are 8 terms, each at least as large as 1/16, so their sum is at least 8/16 = 1/2.

- Likewise, the sum of the 16 terms from 1/17 to 1/32 is at least 1/2.

- Each time, we go out twice as far and extract a sum that is at least 1/2.

- So we can make the sum as large as we please: It diverges!

We end with an example of a hard infinite sum that succumbs to a simple algebraic substitution. Compute the infinite continued fraction

$$1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}.$$

In general, infinite processes such as sums and fractions are defined by looking at the finite version and considering what happens when they converge. Get your hands dirty: Compute successive fractions. We get 1, 2, 3/2, 5/3, 8/5, and so on. Conjecture: These are quotients of successive Fibonacci numbers. We compute the limiting value using the creative substitution tool. Let x equal the whole limiting value. Then $x = 1 + 1/x$. The quadratic formula yields $\frac{1+\sqrt{5}}{2}$, which is approximately 1.6803. This number is known as the golden ratio, and it is ubiquitous in mathematics. ■

Suggested Reading

Solow, *How to Read and Do Proofs*.

Velleman, *How to Prove It*.

Questions to Consider

1. Prove that there is no smallest positive real number.

2. Find the sum $1 \times 1! + 2 \times 2! + 3 \times 3! + \cdots + 100 \times 100!$.

Pictures, Recasting, and Points of View

Lecture 6

You'll soon see how powerful it is to stay flexible, to not commit to a particular point of view or even to a particular branch of mathematics.

In this lecture, we explore 3 strategies: draw a picture, change your point of view, and recast your problem. We begin with proofs without words; then explore the amazing utility of the simple distance-time graph; and next move on to harder problems, where the crux idea is the discovery of the natural point of view. We will look at an example where the problem is geometric, at least on the surface, but cannot be solved until we turn it into a logic puzzler.

Let's begin with an example that makes strong use, as many pictorial problems do, of symmetry: Prove that if T is a triangular number then $8T + 1$ will be a perfect square.

- For example, 55 is a triangular number (namely, $1 + 2 + \cdots + 10$), and $8 \times 55 + 1 = 441$, which is 21×21.
- But why is it true? It is not too hard to prove using algebra, but it is more fun to use pictures.
- The key idea is that 2 identical triangular numbers can be adjoined to form a rectangle.
- Four such rectangles can be arranged symmetrically to form a square with a hole.

And we get for free the algebraic identity $(2n+1)^2 = 8T_n$.

Let's switch gears to a pure word problem, one with no obvious picture. Pat works in the city and lives in the suburbs with Sal. Every afternoon, Pat gets on a train that arrives at the suburban station at exactly 5 pm. Sal leaves the house before 5 and drives at a constant speed so as to arrive at the train station at exactly 5 pm to pick up Pat. The route that Sal drives never changes. One day, this routine is interrupted, because there is a power failure at work. Pat gets to leave early and catches a train that arrives at the suburban station at 4 pm. Instead of phoning Sal to ask for an earlier pickup, Pat decides to get a little exercise and begins walking home along the route that Sal drives, knowing that eventually Sal will intercept Pat and make a U-turn, and they will head home together in the car. This is indeed what happens, and Pat ends up arriving at home 10 minutes earlier than on a normal day. Assuming that Pat's walking speed is constant, that the U-turn takes no time, and that Sal's driving speed is constant, for how many minutes did Pat walk?

The problem can be solved by drawing a distance-time graph. It is clear geometrically that this length is 10 minutes. Hence Pat walks for 55 minutes.

Every problem has a natural point of view; you need to find it. If a problem is hard, perhaps you need a new point of view. What is the first time after noon at which the hour and minute hands meet? This is an amusing and moderately hard algebra exercise. However, this problem can be solved in a few seconds in your head if you avoid messy algebra and just consider the natural point of view. Don't use a fixed frame of reference. Instead, look at it from the point of view of the hour hand. At noon, it coincides with the minute hand. Then, after somewhat more than an hour, the minute hand visits it again!

The most common form of recasting is between algebra and geometry.

How long until the next visit? The same amount of time, of course! From the point of view of the hour hand, nothing has changed from the situation at noon! Clearly there will be 11 meetings between noon and midnight, so the time between meetings is exactly 12/11 of an hour. The first meeting after noon is at 1:05:27.18.

Here is a somewhat harder problem: A person dives from a bridge into a river and swims upstream through the water for 1 hour at constant speed. She then turns around and swims downstream through the water at the same rate of speed. As the swimmer passes under the bridge, a bystander tells her that her hat fell into the river as she originally dived. The swimmer continues downstream at the same rate of speed, catching up with the hat at another bridge exactly 1 mile downstream from the first one. What is the speed of the current in miles per hour?

Look at things from the hat's point of view. The hat thinks that it is sitting still in the water. From its point of view, the swimmer abandoned it and then swam away for an hour at a certain speed (namely, the speed of the swimmer in still water). Then the swimmer turned around and headed back, going at exactly the same speed, since the current is always acting equally on both hat and swimmer. Therefore, the swimmer retrieves the hat exactly 1 hour after turning around. The whole thing took 2 hours, during which the hat traveled 1 mile downstream. So the speed of the current is 1/2 mile per hour.

Let's turn to recasting. The most common form of recasting is between algebra and geometry. Another common recasting is between number theory and combinatorics. There are other possibilities, as you will see with the private planets problem.

The Private Planets Problem

- Consider n identical spherical planets in space, with $n > 2$. Call a point on the surface of a planet private if it cannot be seen from any other planet. What is the total private area among all the planets?

  - It is easy to experiment with $n = 2$; the total private area is equal to the area of a single planet.

  - For $n = 3$, it is a similar situation, because the centers of the planets lie in a plane.

  - But for $n > 3$, it gets ugly. The planets may no longer lie on a plane, so we are forced to contemplate spherical geometry.

- But perhaps it is not really a geometry problem! Crux move: Construct a universal north. Imagine that the planets lie in a room, and thus the point on each planet that is closest to the ceiling of the room is that planet's north pole.

    - Now imagine a planet *A* whose north pole is a private point.

    - For this to be so, there can be no "eyes" on other planets that are north of the plane tangent to *A*'s north pole. In other words, every point on every other planet must lie south of this tangent plane.

    - If the north pole is private, the centers of all other planets lie south of the equator.

    - Thus, the north poles of all other planets will be visible from planet *A* or else visible from a planet to the south of planet *A*.

    - Generalizing, if a point *P* is a private point on some planet, then it must be public on all other planets!

    - Conversely, if *P* is public on one planet, it must be private on at least one other planet.

- Conclusion: If a universal location is private on one planet, it cannot be private on any other planet, and every universal location must be private on at least one planet. Putting these 2 together, every universal location is private on exactly one planet.

- So the total private area is equal to the surface area of one planet!

How did we solve the planet problem? The universal frame of reference facilitated investigation of the relationships between private and public on different planets. We then realized that private/public is not just a geometric relationship but a binary logical relationship. We recast a hard geometry problem into a relatively simple logic problem! ■

Suggested Reading

Nelson, *Proofs without Words*.

Zeitz, *The Art and Craft of Problem Solving*, sec. 2.4.

Questions to Consider

1. Find a "proof without words" that the sum of the first n positive odd integers is equal to n^2.

2. Sonia walks up an escalator that is going up. When she walks at 1 step per second, it takes her 20 steps to get to the top. If she walks at 2 steps per second, it takes her 32 steps to get to the top. She never skips over any steps. How many steps does the escalator have?

The Great Simplifier—Parity

Lecture 7

Any time you have a situation of mutuality, where you might have 2 players, in some sense, connected by some relationship, there should be a nice systematic way of analyzing it. ... There's a branch of mathematics designed for handling these situations, called graph theory.

Parity is such an important tactical idea that we devote most of this lecture to it. Parity analysis helps to solve problems because it allows us to reduce the possibly infinite complexity of a situation to just 2 states. We apply parity to a number of diverse problems involving information flow, changing configurations, and maps of countries. We also introduce a topic that will recur several times in this course: graph theory, the study of networks.

The Evil Wizard Hat Problem

An evil wizard has imprisoned 64 people. The wizard announces the following: Tomorrow I will have you stand in a line, and I will put a hat on each of your heads. The hat will be either white or black. You will be able to see the hats of everyone in front of you, but you will not be able to see your hat or the hats of the people behind you. I will begin by asking the person at the back of the line to guess his or her hat color. If the guess is correct, that person will be released. If the guess is wrong, that person will be killed in a painful way. Then I will ask the next person in line, and so on. When it is your turn to speak, you are only allowed to say the single word 'black' or 'white,' and otherwise you are not allowed to communicate with each other while you are standing in line. Although you will not be able to see the people behind you, you will know (by hearing) if they have answered correctly or not.

The prisoners are allowed to chat for a few minutes before their ordeal begins. What is the largest number of prisoners that can be guaranteed to survive?

Clearly, it is possible to save half the people, where the sacrificial lambs agree to say the color of the person in front of them. But we can do much better, using the observation that as we count something, its parity alternates. This seems trivial, but sometimes it is enough! For example, suppose you can see 10 hats in front of you, and you count 7 black and 3 white. Use wishful thinking: What do you wish you knew? We need some way to communicate the parity of the number of black hats in front of us to the person directly in front of us, using the code "black" = odd and "white" = even, and the first person to speak (the one at the back of the line) will use this code for the parity of the number of black hats that he or she sees.

He may not survive, of course. The next person in line, however, is guaranteed to live, if he pays attention. He knows that the person behind him saw an even number of black hats. If he sees an odd number, then he knows his hat is black. If he sees an even number, then he knows his hat is white. And everyone else is also safe: By listening to the people behind them, they can keep track of the current parity of the number of black hats.

Here is a completely different application of parity: the classic proof (dating back to Euclid) that $\sqrt{2}$ is irrational. We use 3 simple facts.

- If a number is even, it can be written in the form $2m$ for some integer m.
- If an even integer is a perfect square, then its square root is also even.

- If a number is rational, then it can be expressed as a fraction in simplest form, in which case the numerator and denominator cannot both be even.

We will do a proof by contradiction. Assume, to the contrary, that $\sqrt{2}$ is a rational number. Then we can write it as a fraction a/b in lowest terms; so a and b are both integers and not both even. If we square this, we get 2.

- So $a^2 / b^2 = 2$, or $a^2 = 2b^2$.

- But $2b^2$ is even by fact 1 above, so a^2 is even, so a is even by fact 2. Then we can write $a = 2u$ for some integer u (by fact 1).

- Squaring this, we get $a^2 = 4u^2$, and substituting this back into $a^2 = 2b^2$, we get $4u^2 = 2b^2$, or $2u^2 = b^2$. But now we deduce that b^2 is even, so b is even.

- Conclusion: a and b are both even. That is impossible, so $\sqrt{2}$ cannot be rational.

The next problem, the locker problem, we will not solve completely in this lecture.

The Locker Problem

- Lockers in a row are numbered 1, 2, 3, … , 1000. At first, all the lockers are closed. A person walks by and opens every other locker, starting with locker 2. Thus lockers 2, 4, 6, … , 998, 1000 are open. Another person walks by and changes the state of every third locker, starting with locker 3. Then another person changes the state of every fourth locker, starting with locker 4, and so on. This process continues until no more lockers can be altered. Which lockers will be closed?

- A simple investigation quickly leads to the conjecture that squares remain closed.
- Look at the penultimate step. What determines if a locker is open or closed? The parity of the number of times the locker's state changed.
- Clearly, locker *L* will only be changed at step *S* if *S* divides *L*.
- So we have reduced our problem about lockers to the equivalent number theory question: Can we prove that a number has an odd number of divisors if and only if it is a perfect square?

The next problem comes from the Bay Area Mathematical Olympiad. A lock has 16 keys arranged in a 4×4 array, and each key is oriented either horizontally or vertically. In order to open it, all the keys must be vertically oriented. When a key is switched to another position, all the other keys in the same row and column automatically switch their positions too. Show that no matter what the starting positions are, it is always possible to open this lock. (Only one key at a time can be switched.)

Parity can be used whenever there are just 2 states in a problem, and here vertical and horizontal certainly fit the bill. We need to be able to change any key, but the problem is that when we turn a single key, it changes 7 keys, including itself. We would like to only change one key. However, if a key is changed twice, it is as if it was never changed (odd + odd is even). We need to find a move or sequence of moves that changes lots of keys an even number of times but changes only one key an odd number of times. There are not too many things to try, and a number of them work. You can check that if you pick any specific key and turn all 7 keys in its cross, this will do the trick.

Here is another problem from a regional Olympiad (Colorado): If 127 people play in a singles tennis tournament, prove that at the end of the tournament, the number of people who have played an odd number of games is even. Note that the tournament can have any structure; the only restriction is that each game requires 2 people.

Parity analysis helps to solve problems because it allows us to reduce the possibly infinite complexity of a situation to just 2 states.

Suppose each person pays a dollar to the tournament each time they play a game. Each game played makes the tournament owner $2 richer. If we add up the amount of money each person pays, it will be equal to 2 times the number of games played, and will be even. And it is the same as adding up the number of games each person plays. So, if there are 127 people, and we add up the number of games each person played, we are adding 127 numbers and getting an even sum. So the number of people who played an odd number of games is even. This problem involves mutual relationships (playing in a game). This is a ubiquitous situation and can be modeled in many other contexts. This leads to the abstract idea of a graph: an entity of vertices and edges.

Parity solves problems with ease because it reduces the complexity of a problem, which increases our information level and allows us to understand things better. ■

Suggested Reading

Fomin and Itenberg, *Mathematical Circles*, chap. 1.

Zeitz, *The Art and Craft of Problem Solving*, secs. 3.4, 4.1.

Questions to Consider

1. A Pythagorean triple *a*, *b*, *c* are integers satisfying $a^2 + b^2 = c^2$. Prove that at least one of these integers must be even.

2. Given 4 points *A*, *B*, *C*, *D* in the plane, it is possible to join *A* to *B*, *B* to *C*, *C* to *D*, and *D* back to *A* with 4 straight line segments, creating a quadrilateral. It is possible to join these line segments in such a way so that a straight line can be drawn through the interior of all 4 segments. The trick is to make the polygon intersect itself (imagine *ABDC* as the vertices of a square going around clockwise). But can this be done with 5 points?

The Great Unifier—Symmetry
Lecture 8

If an object is unchanged, in other words, invariant, with respect to some sort of transformation, a geometrical transformation, it's called symmetric, and the transformation itself is called a symmetry.

Like parity, the symmetry tactic is so important that we need to spend an entire lecture on it. Symmetry is one of the most important underlying principles in mathematics. Symmetrical structures (geometric or otherwise) are simpler than asymmetric structures and hence easier to investigate. Thus, searching for symmetry—or if need be, imposing it where it was not at first—is a powerful tactic for investigating problems. In this lecture, we explore symmetry not just visually but also algebraically.

There are hundreds of proofs of the Pythagorean theorem.

What is symmetry? An object (which may or may not be geometric) is symmetrical if a transformation leaves it invariant. This can take many forms. Geometric symmetry is the simplest to understand. Objects can have rotational, reflectional, and other symmetries. Metaphorical symmetry is more subtle. Examples of metaphorical symmetry are the pills problem (once the pills were balanced, the problem was almost solved); the private planets problem (private/public duality); and the handshake lemma (handshake is a symmetrical relation). Thus parity is, in a sense, an example of metaphorical symmetry in action.

What's so good about symmetry? A problem, by definition, is information-poor and disordered. Symmetry increases order. Thus, you should systematically search for symmetry, and if it does not appear to be present, you should attempt to impose it. The Gaussian pairing trick is an example of imposing symmetry.

- The sum $1 + 2 + 3 + 4 + \cdots + 100$ is not symmetrical.

- But the below is. It has rotational symmetry!

$$\begin{array}{l} 1 + 2 + 3 + \cdots + 100 \\ 100 + 99 + 98 + \cdots + 1 \end{array}$$

Symmetry is related to the search for natural points of view. Often this point (or line, etc.) is one of symmetry. The classic 4 bugs problem is an excellent example.

The 4 Bugs Problem

There are 4 bugs, each situated at a vertex of a unit square. Suddenly, each bug begins to chase its counterclockwise neighbor. If the bugs travel at 1 unit per minute, how long will it take for the 4 bugs to crash into one another?

- What is the critical point of view? Clearly, the center point! Hence, we need to figure out the radial speed r.

- This can be computed using vectors: Each bug moves at a constant speed, always making a 45° angle with the radial vector. So we can compute the component of speed along this vector.

- The actual value of r is $\sqrt{2}/2$ (which equals 0.707), but the precise value is not important. Since the radial speed is r units per minute, and the radial distance is also r units, the time it will take to reach the center will be r/r, which equals 1 minute!

There are hundreds of proofs of the Pythagorean theorem. Here is a simple one that uses the imposition of symmetry. First we reformulate $a^2 + b^2 = c^2$ into an equality of areas of squares. The sum of the areas of the 2 smaller squares must be equal to the area of the large square.

- Start with the large square (c^2) and add copies of the original triangle, symmetrically.

- Thus, c^2 equals the area of 4 of the original triangles plus the area of the square in the center.

- The length of this square is the equal to $(a - b)$, and the area of one triangle is $ab/2$.

- Repacking this algebraically, we get $(a-b)^2 + 4(ab/2) = c^2$, which is equivalent to $a^2 + b^2 = c^2$.

Symmetry can be applied to number theory. Recall the locker problem in the last lecture? We did not quite solve it; we need to show that an integer has an odd number of divisors if and only if it is a perfect square (a square of an integer). Remember, divisors include 1 and the number itself. For example, for $N = 12$, pair each factor with its mate. Since 12 is not a perfect square, each divisor has a distinct mate, different from itself.

1×12
2×6
3×4

Since 36 is a perfect square, the mate of 6 is itself. So 6 is literally the odd man out. What we used here is the natural correspondence (metaphorical symmetry) between the divisor d of a number N and its mate N/d.

1×36
2×18
3×12
4×9
6×6

These ideas also apply to algebra. Let's use symmetry to solve the fourth-degree equation $x^4 + x^3 + x^2 + x + 1 = 0$. This equation is nearly "symmetrical"; we can make it more so by dividing by x^2, getting $x^2 + x + 1 + 1/x + 1/x^2 = 0$. We have imposed an interesting algebraic symmetry. There is a correspondence between terms of the form x^k and terms of the form $1/x^k$.This suggests a natural symmetrical substitution: $y = x + 1/x$. Notice that $(x + 1/x)^2 = x^2 + 2 + 1/x^2$. So our equation now becomes $y^2 - 2 + y + 1 = 0$, a quadratic equation that can be solved by the quadratic formula. Then we can solve for x, again using the quadratic formula! Once again, imposing symmetry was the key. ■

Suggested Reading

Conway, Burgiel, and Goodman-Strauss, *The Symmetry of Things*.

Weyl, *Symmetry*.

Zeitz, *The Art and Craft of Problem Solving*, sec. 3.1.

Questions to Consider

1. Determine the minimum perimeter of a triangle, one of whose vertices is (4, 3), the other is on the x-axis, and the third is on the line $y = x$.

2. The set {1, 2, 3, 4, 5, 6} has 64 subsets, including the empty set and the set itself. For how many of these subsets is the sum of the elements greater than 10?

Symmetry Wins Games!

Lecture 9

In this lecture, we're going to do something a little different. Rather than focusing on problem-solving ... strategies and tactics, we're going to apply them. Specifically, we will introduce a little bit of the theory of combinatorial games, which is a nice illustration of metaphorical symmetry.

This is our first applied lecture, where we focus more on using problem-solving ideas than on developing new ones. We study combinatorial games, which apply literal and metaphorical symmetry. Our cornerstone is Wythoff's Nim, which is an absurdly easy game that is amazingly hard to play well, until you use symmetry and a few other ideas. It is also one of the many mathematical phenomena in which the Fibonacci numbers and the related golden ratio play an important role.

All combinatorial games have the same basic structure: Two players, *A* and *B*, alternate taking turns. *A* goes first. The game ends when no legal moves can be made. The last person to make a legal move wins. Our goal: Given a game, discover the winning strategy.

The Takeaway Game

Start with 17 pennies. A legal move consists of removing 1, 2, 3, or 4 pennies. What is the winning strategy?

- The get your hands dirty and make it simpler strategies compel us to change the 17 to smaller values, such as 1, 2, 3, and so on.

- Looking at these simpler games shows us that if you are left with 5, no matter what you do, your opponent will win on the next move. So if you can present your opponent with 5, you win.

- If you are presented with 6, 7, 8, or 9, then you can always present your opponent with 5 and thus win.

- Notice that the values 0 and 5 belong together, as do the values 1, 2, 3, 4 and 6, 7, 8, 9: If you move to (i.e., present to your opponent) 0 or 5, you will win; if you move to 1, 2, 3, 4 or 6, 7, 8, 9, you will lose, because your opponent will move to 0 or 5.

- The pattern clearly continues. Let us call the values 0, 5, 10, 15, … the oases and the other values the desert. If you move to an oasis value, your opponent must move to a desert value. If your opponent moves to a desert value, you can always move to an oasis. Thus, if a player moves to an oasis, he or she can control the game, always moving to oases and forcing the opponent to stay in the desert—until the end, when the lucky oasis traveler moves to the final oasis value, 0, and wins the game.

Here is a graphical way to analyze the game. We use the penultimate step strategy to work backward from the very beginning.

- Start with the value 0. Assume that you have gotten there (wishful thinking), so you have won. Color this value green.

- Analyze backward: Color red all values that can get to 0 in one move.

- The next value, 5, has not been colored yet; it is colored green.

- Notice that from this green value, one can only move to red values, and from there, one can always move to another green value.

- Continuing, we color red the next values that in one move can get to green. We continue this recursive process indefinitely.
- In the desert/oasis model, the game really switches between 2 metaphorically symmetrical states. The winner is the person who controls the switching.

The Divide and Conquer Game

Start with 100 pennies. A legal move consists of removing a proper divisor of the number of pennies left. Thus, on the first move, *A* could remove 1, 2, 4, 5, 10, 20, 25, or 50 pennies, but not 100 pennies. The game ends when there is exactly 1 penny left (since then there are no legal moves possible).

- One possible game: Start with 100. Player *A*'s moves are bolded: **98**, 97, **96**, 48, **24**, 21, **14**, 7, **6**, 4, **2**, 1. Hence *B* wins.
- What is the winning strategy? Parity is a natural first choice, since it is a good binary property; there tend to be symmetries between odd and even positions.
- Since the winning position is to present your opponent with 1, we want to see if it is possible to control oddness.
    - If we can present our opponent with an odd number, then since all divisors of an odd number are odd, the opponent must present us with an even number!
    - And if we are given an even number, we can always turn the tables and present our opponent with an odd, by subtracting 1!

- So the winning strategy is this: *A* subtracts an odd from 100 and thereafter keeps presenting odd numbers to *B*. In general, *A* can always win if the starting number is even, and *B* can always win if the starting number is odd.

The next game, which I call cat and mouse, also relies on parity and symmetry. A very polite cat chases an equally polite mouse. They take turns moving on a grid. Initially, the cat is at the point labeled *C*; the mouse is at *M*. The cat goes first and can move to any neighboring point connected to it by a single edge. Thus the cat can go to points 1, 2, or 3, but no others, on its first turn. The cat wins if it can reach the mouse in 15 or fewer moves. Can the cat win?

At first, it seems impossible; the mouse can always be one step away from the cat. If you look carefully at the grid, you will discover that it contains only rectangles glued together, with one triangle at the top left. Temporarily pretend that the triangle is not there (make it easier). Then we can color the vertices with alternating colors, like a chessboard. Let the cat's starting position be blue. Notice that the mouse's starting color is also blue. So if we ignore the triangle, the cat moves from blue to red. Then the mouse moves from blue to red. Then the cat moves from red to blue, and the mouse follows suit. The cat and mouse are always on same-colored vertices, so it is impossible for the cat to make a single move in which it catches the mouse! But if the cat moves into the triangle, it is possible for the cat to gain a tempo and change its color to the opposite of the mouse's. Then the cat has a chance at catching the mouse. And it will—try it!

All combinatorial games have the same basic structure: Two players, *A* and *B*, alternate taking turns.

The cat and mouse game illustrates the idea of bipartite graphs. A graph is bipartite if the vertices can be put into 2 nonoverlapping sets (e.g., boys and girls) in such a way that every edge joins vertices of opposites (e.g., opposite sexes). If a graph is bipartite it cannot have any odd cycles (triangles, pentagons, etc.), for if it did, you would have a chain of dance partners: *B G B G B*, and then *B* is connected to *B*.

Our final game is called Wythoff's Nim, a.k.a. "puppies and kittens." Start with a pet shelter with, say, 10 puppies and 7 kittens. *A* and *B* alternate turns adopting animals. On each turn, you must adopt at least one animal, and you must only adopt one kind of animal, unless you adopt equal numbers of both kinds. As always, the winner is the one who makes the last legal move—in this case, the one who clears out the pet shelter. We can use the oasis/desert analysis method to quickly build up a repertoire of oasis positions that will make us unbeatable.

- Crux idea: Plot positions as ordered pairs (k, p) and use graph paper!
- Here is a list of oases (leaving out symmetrical pairs): (1, 2), (3, 5), (4, 7), (6, 10), (8, 13).
- Here is another way to generate oases.
- Start with (1, 2).
- This eliminates all other ordered pairs with difference of 1 and with either *p* or *k* equal to 1 or 2.
- So we cannot use 1, 2, or any difference of 1.
- Thus the next oasis is (3, 5).
- The next one after that must be (4, 7), and so on. ■

Suggested Reading

Berlekamp, Conway, and Guy, *Winning Ways for Your Mathematical Plays*.

Gardner, *Penrose Tiles to Trapdoor Ciphers*, chap. 8.

Questions to Consider

1. Modify the takeaway game so that the last penny is "poison," and the person who takes this penny away loses.

2. Two players take turns choosing numbers from 1 to 9 inclusive. Once a number is chosen, the other player cannot choose it. The winner is the person whose numbers add up exactly to 15. Is there a winning strategy? Does this game remind you of another game?

Contemplate Extreme Values
Lecture 10

We'll look at several interesting examples, and in each case, the problem is generally rather difficult until we use the extreme principle. When we use the extreme principle, the problem becomes nearly trivial.

This lecture focuses on an incredibly productive but little known (by laypeople) tactic that simply advises one to contemplate the extremal values in a problem. For example, look at the triangle of least area, the largest variable, or the first time something occurs. This tactic has the nearly magical ability to solve hard problems in a line or 2. I have often compared expert use of the extreme principle to watching a martial artist break a board, something that looks impossible yet effortless to the uninitiated.

How do you apply the extreme principle? Put your things in order, contemplate the largest and/or smallest of these things, and be creative about just what a thing is and how to measure it.

Here is a warm-up example: There are finitely many points in the plane, colored red or blue. Between any 2 red points, there is a blue point on the line segment connecting them. Between any 2 blue points, there is a red point on the line segment connecting them. What kinds of configurations are possible?

Sometimes the hardest thing about an extreme principle problem is figuring out which entity should be contemplated.

Investigation yields the conjecture that the only configurations are linear, with points alternating colors. But how do we prove this rigorously? Assume to the contrary that there is a 2-dimensional configuration. Consider the triangle of smallest positive area. Two vertices of this triangle must be the same color, forcing a point to lie between them; this creates a smaller triangle, which contradicts the minimality of the triangle. So a 2-dimensional configuration is impossible!

Here is a problem from the Canadian Mathematical Olympiad: On a large, flat field, n people ($n > 1$) are positioned so that for each person, the distances to each of the other people are different. Each person holds a water pistol and at a given signal fires and hits the person who is closest. When n is odd, show that there is at least 1 person left dry. Is this always true when n is even?

- When n is even, it is possible to have pairs of mutual antagonists, so everyone gets wet.
- If n is odd, there must be at least some people whose victims are not shooting them.
- Consider the person among these who shoots the furthest distance.
- Claim: This person stays dry!

The Handshake Problem

In the handshake problem, we use the extreme principle to solve a problem that seems to have too little information for a solution.

- I invite 10 couples to a party at my house. I ask everyone present, including my wife, how many people they shook hands with. It turns out that everyone questioned—I did not question myself, of course—shook hands with a different number of people. If we assume that no one shook hands with his or her partner, how many people did my wife shake hands with?
- Make it easier by looking at smaller parties (with 0, 1, or 2 guests), which quickly leads to the conjecture that if there are n guest couples, the hostess must shake hands with n people. Thus we claim that the hostess shook hands with 10 people.

- But how do we prove it?
- Crux idea: Focus on the partner of the person who shakes everyone's hand (extreme principle). This person has to be the one who shakes hands with no one.
- Now what? Banish this couple from the party!
- What remains? A party with 1 fewer couple but obeying the same rules. Hence the 0-shaker must once again be partnered with the maximal person.
- Continue until only 2 people are left, me and the person who shook hands with 10 people.

Sometimes the hardest thing about an extreme principle problem is figuring out which entity should be contemplated; this is what makes the following problem rather difficult. Imagine a fixed network of homes (i.e., a graph). Each home is populated by a family that belongs to one of 2 ethnic groups. A network is called diverse if for each home, it is never the case that the majority of the neighbors come from the same ethnic group. Given any fixed network, can it be made diverse?

How do you quantify network diversity? Look at the edges. The crux idea is to maximize balanced edges! Given any fixed network, there are only finitely many ways to color the vertices. For each coloration, count the number of balanced edges. Find the coloration with the greatest number of balanced edges. This will be a diverse network!

How do we know this? Assume, to the contrary, that this network is not diverse. Then there is a "bad" vertex. Change its color! This produces a new coloration, with more balanced edges. But this is impossible! We have created our contradiction. ■

Suggested Reading

Andreescu and Savchev, *Mathematical Minatures*, chap. 25.

Zeitz, *The Art and Craft of Problem Solving*, sec. 3.2.

Questions to Consider

1. Given an infinite chessboard with positive integers in each square, arranged so that each square's value is the average of its 4 neighbors to the north, south, east, and west, prove that all the values must be equal.

2. Suppose you are given a finite set of coins in the plane, all with different diameters. Show that one of the coins is tangent to at most 5 of the others.

The Culture of Problem Solving
Lecture 11

There is a mathematical community. It's a community of people who are passionate about math. There's a parallel culture that celebrates intellectual intensity. It's one that started in Eastern Europe, in a certain sense, but it's now firmly entrenched in North America. If you've stuck with me this far, then I will declare that you are now officially inducted into this culture.

We take a brief detour from solving mathematics problems to look at problem solving as a cultural force. In Eastern Europe, mathematics has long been respected, even among children, and math contests play a role not unlike sports in the United States. We look at the history of the Mathematical Olympiad culture and assess whether this culture has a chance of taking root in the United States. I will draw upon my experiences as a member of the first U.S. team to compete in the International Mathematical Olympiad; my later career as a coach, problem writer, and editor for these and other contests; and my current efforts to make the San Francisco Bay Area become more like Bulgaria—at least with respect to mathematics.

The culture of math circles, math contests, and math nerds ("nerd" is not a pejorative) is a world where excellence in math is a social benefit and where math contests are as popular as sports contests are in the United States. This culture exists, but it is rare, both in space and time. Its roots are primarily Eastern European. My first introduction to this culture was as a student at Stuyvesant High School in New York, one of several specialized schools for math and science. Unique features of this school included social acceptance of intellectual achievement and opportunities for independent study with teachers with doctorates.

Why did this culture flourish in Eastern Europe? Mathematics and physics were ways to escape totalitarianism. More intellectuals were drawn to math and physics than in Western democracies. And the state rewarded scientists with official status.

What is a math circle? The word "circle" is a direct translation of the Russian word *kruzhok*. Math circles are like math clubs on steroids. They are highly intensive and feature interaction with college and graduate students, professors, and world-famous research mathematicians. The curriculum is based on problem solving, and there is a deliberate effort to transmit the folklore of problem solving. Most math circles are linked with math competitions.

Here is a brief and incomplete history of these contests. The first modern one was the Hungarian Problems, which began in 1906. The Moscow Olympiads began in the 1930s. The International Mathematical Olympiad began in 1959. At first, its participants were only Iron Curtain countries, but gradually it has become more inclusive. The United States first participated in 1974, and today nearly 100 nations participate. The style of these math contests is unusual. In the United States, many exams are still multiple choice, but the Olympiad style is always essay-proof.

The important thing about this culture is that it celebrates intellectual inquiry and intensity and encourages passion for investigating mathematics.

How did this culture migrate to the United States? Oversimplifying things, I will say this was due to 2 important historical events. The launch of *Sputnik* in 1957 impelled the West to emulate Eastern educational achievement. The decline and fall of communism led to a great immigration of Eastern European mathematicians.

Let's look at the "math nerd culture." There are now quite a few specialized schools. There are also several summer camps devoted to mathematical folklore transmission. There are online communities and an international nerd culture that features T-shirts, frisbees, silly word games, and fetishistic memorization of numbers such as π. Famous mathematicians visit math camps and clubs, and such luminaries become worshipped. The important thing about this culture is that it celebrates intellectual inquiry and intensity and encourages passion for investigating mathematics. ■

Suggested Reading

Bell, *Men of Mathematics*.

Olson, *Count Down*.

Recasting Integers Geometrically

Lecture 12

This and the next lecture focus on number theory, the study of the integers. We've seen a little number theory so far in earlier lectures, but now that we understand basic strategies and tactics, we can go deeper.

This lecture focuses on the chicken nuggets problem, a classic folklore puzzle that originated in England over a century ago and is now a mainstay of math clubs around the world. This rich problem can be analyzed with induction, pictures, symmetry, careful counting principles, and other approaches. We focus here on a visual approach, recasting the problem into one of counting lattice points, with symmetry playing a key role.

The fundamental objects that we will be exploring are lattice points: the points (x, y) on the coordinate plane, where both x and y are integers. By contemplating lattice points as natural objects, we can convert algebraic questions involving natural numbers into geometric questions, and vice versa.

The Chicken Nuggets Problem

Bay Area Rapid Foods sells chicken nuggets in boxes of 7 and boxes of 10. A number n is feasible if it is possible to buy n nuggets. For example, 7 is the smallest feasible number, and the next ones after that are 10, 14, 17, 20, 21, and 24. There are 2 natural questions.

- Is there a largest nonfeasible number, and if so, what is it?
- If there is a largest nonfeasible number, then how many nonfeasible numbers are there?

The values 7 and 10 are probably not critical; we should look at smaller, simpler values and try to get a general picture. In other words, given positive integers a and b, if nuggets come in boxes of size a and size b, what is the largest nonfeasible number, and how many nonfeasible numbers are there?

- Here is the algebraic formulation: A number n is feasible if there exist nonnegative integers x and y such that $ax + by = n$. If these nonnegative numbers do not exist, we say that n is nonfeasible.
- Example: Let $a = 5$ and $b = 7$. Then list the nonfeasible and feasible numbers.
- Nonfeasible: 1, 2, 3, 4, 6, 8, 9, 11, 13, 16, 18, 23.
- Feasible: 5, 7, 10, 12, 14, 15, 17, 19, 20, 21, 22, 24, 25, 26, 27, 28. It is clear that all values starting with 24 are feasible (since 5 in a row are).
- In general, we can assume that a and b are relatively prime.
    - Why? Suppose, for example, that $a = 15$ and $b = 21$, both sharing the factor of 3. Then the feasible numbers are those values of n such that there exist nonnegative integers x and y with $15x + 21y = n$.
    - But the left-hand side factors into $3(5x + 7y)$.
    - Thus, n must be a multiple of 3. So let's write $n = 3m$. We now get $3(5x + 7y) = 3m$, and we can divide both sides by 3, getting $5x + 7y = m$.
    - In other words, n is feasible for $a = 15$, $b = 21$ if and only if it is a multiple of 3 and if $n/3$ is feasible for $a = 5$ and $b = 7$.

Can we find a visual interpretation?

- Of course! Suppose that 19 is feasible. Then there exist nonnegative integers x and y such that $5x + 7y = 19$. (Indeed, $x = 3$, $y = 2$ works). But $5x + 7y = 19$ is the equation of a line in the coordinate plane!

- Visually, this is the same as saying that the line $5x + 7y = 19$ contains a lattice point in the first quadrant.

- In other words, n is feasible if and only if the line $5x + 7y = n$ has at least one lattice point in the first quadrant.

- If a and b are relatively prime, then a line with slope a/b either has no lattice points on it or hits lattice points in a systematic way, with each lattice point separated from the next by the same vector.

- What about nonfeasible numbers? For example, 11 is nonfeasible. That means that there is no lattice point on the line $5x + 7y = 11$ in the first quadrant.

- One solution is $(5, -2)$. All solutions are separated by the vector $(-7, 5)$. Remember that there are no lattice points on the dotted line between the 2 lattice points.

- In other words, we can think of the solution lines $5x + 7y = n$ as parallel lines that come with lattice points, spaced at equal intervals, and we need to find the ones that have lattice points in quadrant 1. Those n's are feasible, and the others are nonfeasible.

Breakthrough idea: Let's change our point of view and count lattice points, not feasible or nonfeasible numbers!

- The highest nonfeasible number will be the highest "bad" line stuck with a lattice point in quadrant 2. That will be the line corresponding to $(6, -1)$. If we plug $(6, -1)$ into the equation $5x + 7y$, we get 23, which we know is the largest nonfeasible number.

- In general, if a and b are relatively prime, using this exact same procedure, the lattice point corresponding to the largest nonfeasible number will be at $(b - 1, -1)$. So if we plug $x = b - 1$ and $y = -1$ into $ax + by$ when we do the algebra, we get $ab - a - b$, and that is our formula for the largest nonfeasible number.

- Is there a relationship between the feasible and nonfeasible numbers? If you look at that highest nonfeasible line and then look at the line going through zero, you see a parallelogram that is perfectly symmetrical, with rotational symmetry. Clearly the number of feasible lattice points is going to be equal to the number of nonfeasible lattice points, because the lattice points that lie on the x and y axes are in quadrant 1 proper. There are exactly the same number as the ones that lie in quadrant 2.

- Using symmetry, we can conclude that if you look at the numbers between 0 and $ab - a - b$, exactly half will be feasible, and exactly half will be nonfeasible. For example, if you go from 0 to 23, that is a total of 24 numbers: 12 of them will be feasible, and 12 of them will not be feasible. ■

Suggested Reading

Beck and Robbins, *Computing the Continuous Discretely*.

Sylvester, "Question 7382."

Questions to Consider

1. Pick's theorem (see Lecture 4, problem 2) is true for any polygon whose vertices are lattice points. Try drawing several polygons to test this empirically.

2. Suppose you accept on faith that Pick's theorem is true for triangles. How can you use this to prove that it is true for all polygons?

Recasting Integers with Counting and Series

Lecture 13

Like the last lecture, what we're presenting here is pretty difficult mathematics. We'll prove Fermat's little theorem, which is one of the most important theorems of elementary number theory, but we're going to prove it in a way that was not found in most number theory textbooks.

This lecture employs the powerful strategies of recasting and rule breaking to 2 classical theorems in number theory: Euler's proof of the infinitude of primes and Fermat's little theorem. We use the knowledge of modular arithmetic and infinite series that we developed earlier. We begin by using simple counting ideas to explore number theory, using the basic principle that if something can be counted, it is an integer.

Let's go back, for a moment, to the earlier lecture about chicken nuggets. If you were really, really observant, you may have noticed a gap in our argument. We associated each lattice point with either a feasible or nonfeasible number. But we blithely assumed for each number n that the equation $ax + by = n$ must pass through lattice points. How do we know if it does? In other words, given relatively prime numbers a and b and an integer n, can we be guaranteed that there are integers x and y such that $ax + by = n$?

We begin by using simple counting ideas to explore number theory, using the basic principle that if something can be counted, it is an integer.

We only need to consider $n = 1$. If we can find a solution for that, we can easily get one for $n = 2$ and so on. A concrete example: Show that $5x + 7y = 1$ has integer solutions. We will find a result that can easily be generalized to any values of a and b. We count lattice points. For each integer n, the equation $5x + 7y = n$ determines a line with slope $-5/7$. As n ranges from 0 to 35, we get a family of 36 parallel lines. Notice that if a line intersects one of the lattice points, it will not intersect another. So each lattice point corresponds to a different line. Thus, we just need to count the lattice points. They are arranged in a parallelogram, but we can move the bottom points up. By symmetry, we get a rectangle, and clearly there are 7×5 lattice points. Each of the 35 lines for $n = 0$ to $n = 34$ hits a lattice point. So the $n = 1$ line hits a lattice point, and we are done.

We can exploit the idea of counting in many other ways. Combinatorics has its own logic and rules, which are pretty simple, at the start. Let's look at Fermat's little theorem. We begin with an example: Find the remainder when 2^{1000} is divided by 13. We can use modular arithmetic to solve this.

- Start with $2^1 = 2 \pmod{13}$, and successively multiply by 2.
- Thus, $2^2 = 4$, $2^3 = 8$, and $2^4 = 16$, which equals $3 \pmod{13}$.
- Then $2^5 = 6$ and $2^6 = 12 \pmod{13}$. Now notice that $12 = -1 \pmod{13}$. So instead of multiplying by 2, we square both sides.
- We get $(2^6)^2 = (-1)^2 = 1 \pmod{13}$. In other words, $2^{12} = 1 \pmod{13}$.
- This is the crux move, since now we can raise this to any power we want with ease. Since $1000 = 12(83) + 4$, we have $(2^{12})^{83} = 1^{83} = 1 \pmod{13}$.
- Hence $2^{996} = 1 \pmod{13}$, and finally, $2^{1000} = 3 \pmod{13}$.

The crux was finding the exponent of 2 that equals 1 (mod 13). On the other hand, if we try a nonprime mod, like (mod 10), we discover that the powers of some numbers are never equal to 1 (mod 10), and for others we get 1, but never when we raise to the ninth power. The sensible conjecture to try is that if p is prime, then $a^{p-1} = 1 \pmod{p}$ for any number that is not a multiple of p itself. This is Fermat's little theorem. We will prove this, but first we modify the statement by multiplying both sides by a: $a^p = a \pmod{p}$.

Let's replace number theory with combinatorics. We will prove Fermat's little theorem for the concrete values $p = 7$ and $a = 4$ and demonstrate that $4^7 - 4$ is a multiple of 7 using simple counting principles. Imagine a necklace with 7 identical beads. We wish to color them using any of 4 colors. How many different necklaces are possible? Let's make it easier. If it were not a necklace but just a line of beads, then the number of different necklaces would be $4 \times 4 \times 4 \times 4 \times 4 \times 4 \times 4 = 4^7$. That is promising, since it is a number we are interested in. But since we have a necklace, we can slide beads around. In fact, there will be 7 different linear sequences of colors that are all really the same necklace.

Indeed, almost any linear color sequence is 1 of 7 "sisters" that form the same necklace. In other words, each linear sequence is in a 7-member sorority. The only exceptions are the 4 monochromatic sequences. These 4 sequences belong to exclusive sororities: Each sorority has just 1 member. Notice that we are using the fact that 7 is prime. If we had a 6-bead necklace, then the pattern black-red-black-red-black-red would only give rise to 2 sisters. We started with 4^7 linear color sequences. Of these, just 4 were monochromatic. The remaining $4^7 - 4$ sequences can be grouped into 7-member sororities where each sorority member is actually the same necklace. So the total number of different necklaces is $(4^7 - 4)/7 + 4$. So $4^7 - 4$ had to be a multiple of 7, which was what we wanted to prove!

For our final example, we will give a second proof of the infinitude of primes, due to Euler, that is notable for its surprise use of infinite series. We start with the harmonic series, which is infinite.

- Define $S_k = 1 + \frac{1}{k} + \frac{1}{k^2} + \cdots$, and consider the infinite product $S_2 S_3 S_5 S_7 \cdots$, where the subscripts run through the prime numbers.
- The infinite product begins with $(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots)(1 + \frac{1}{3} + \frac{1}{3^2} + \cdots)$ $(1 + \frac{1}{5} + \frac{1}{5^2} + \cdots) \cdots$.

- When we expand this infinite product, it will give us each term of the harmonic series; hence it is infinite.

- Each S_k is just an infinite (but convergent) geometric series.

- Now, assume to the contrary that there are only finitely many primes. Then there are only finitely many S_k terms in our product. That would make the harmonic series finite. But it is infinite! So there must be infinitely many S_k terms and hence infinitely many primes! ■

Suggested Reading

Vanden Eynden, *Elementary Number Theory.*

Zeitz, *The Art and Craft of Problem Solving*, chap. 7.

Questions to Consider

1. Use Fermat's little theorem to find the remainder when 3^{2009} is divided by 19.

2. How many different necklaces can be made using 6 beads, if each bead is a different color?

Things in Categories—The Pigeonhole Tactic

Lecture 14

What's the unifying idea behind these pigeonhole problems? The unifying idea is that we try to strive for commonality, for coincidence, for uniformity, for equality. These are often things that we want to find in a problem.

Like the extreme principle, this tactic seems almost vacuous: If you try to put $n + 1$ pigeons into n pigeonholes, at least 1 hole will contain at least 2 pigeons. Yet the pigeonhole principle allows us to solve an amazing variety of problems. Among the applications we will explore is a graph theory subject known as Ramsey theory, which provides a systematic way of finding patterns in seemingly random structures.

The simplest version of the pigeonhole principle is as follows: If you have more things (pigeons) than categories (pigeonholes), at least 2 of the things belong to the same category. For example: Suppose you color the infinite plane in 2 colors, red and blue, in any arbitrary way. Prove that there are 2 points exactly 1 meter apart that are the same color.

- Pigeonhole solution: Consider an equilateral triangle with side length 1 meter.

- Each of the 3 vertices is colored; there are only 2 possible colors.

- Let the vertices be pigeons, and the colors be holes. Since $3 > 2$, there is a hole with at least 2 pigeons.

Here is another example: People are seated around a circular table at a restaurant. The food is placed on a lazy Susan in the center of the table. Each person ordered a different dish, and it turns out that no one has the correct dish in front of him or her! Show that it is possible to rotate the platform so that at least 2 people will have the correct dish.

- People are the pigeons, but what are the holes?

- This is the challenging part of using the pigeonhole principle: carefully formulating a penultimate step that will solve the problem.

- Suppose there are n people. We can measure distance around the table where n units is a full circle. Everyone starts out a certain nonzero clockwise distance from their correct dish. The possible distances are thus $1, 2, 3, \ldots, n - 1$.

- So at least 2 of them are the same distance from their correct dish.

- Move that distance, and we are done!

A third example: Prove that among any group of people, 2 of them have the same number of friends in the group.

- Crux recasting: graph theory! Prove that in any graph, 2 of the vertices must have the same degree.

- This problem seems perfect for the pigeonhole principle, with vertices as pigeons (things) and degree values as holes (categories). For example, in a 6-vertex graph, there are 6 pigeons. So our penultimate step would be 5 possible degree values.

- But there are 6 possible values! In a 6-vertex graph, a vertex can have 0, 1, 2, 3, 4, or 5 neighbors.

- Wishful thinking and the extreme principle tell us that there must be a way to get rid of at least 1 degree value.

- Consider degree 0. If a graph has such a vertex, it is isolated from the others, in which case no vertex can have degree 5.

- Conversely, if a vertex has degree 5, it is connected to all other vertices, so no vertex can have degree 0.

- Thus the possible degree values are between 0 and 5, inclusive, but cannot include both 0 and 5. So there are only 5 possible values, and with 6 vertices, we can conclude that 2 vertices must have the same degree!

Now we look at the intermediate version of the pigeonhole principle. Suppose we put p pigeons into h holes. Then at least 1 hole contains at least $\lceil p/h \rceil$ pigeons, where the brackets mean ceiling (the ceiling of x is the least integer that is greater than or equal to x). Here is an example: Suppose you have a drawer with 23 socks, and the socks come in 4 colors. Then you must have $\lceil 23/4 \rceil = \lceil 5.75 \rceil = 6$ socks that are the same color.

The pigeonhole principle allows us to solve an amazing variety of problems.

Ramsey theory, named after Frank Ramsey, is a branch of discrete math that concerns itself with what kind of order is guaranteed, even in a random structure. A typical Ramsey theorem says something like, "Given a large enough structure, we are guaranteed to see a smaller substructure." For example, no matter how we 2-color the plane, we are guaranteed 2 points a meter apart that are the same color.

Here is a classic example: Show that among any 6 people, either 3 of them are mutual friends, or 3 are mutual strangers. Graph theory recasting: If you 2-color the edges of a complete graph with 6 vertices, then there must be a monochromatic triangle! Use green and red. We want to see lots of one color, since that would make it more likely to get a monochrome triangle. The extreme principle suggests that we search for the vertex that has, say, the maximum number of red edges emanating from it. The pigeonhole principle gives us that maximum: Each vertex has 5 edges emanating from it. At least $\lceil 5/2 \rceil = 3$ edges are the same color (say, red). Focus on the vertex we started with plus the 3 others that are joined to it with a red edge. If any of these 3 vertices are joined with a red edge, we are done. But if none of them are red, we have created a green monochromatic triangle.

Let's generalize this problem. We just showed that if the edges of a K_6 are 2-colored, then there must be a monochromatic triangle. The Ramsey number formulation of this is R(3, 3) = 6. The Ramsey number R(a, b) is defined to be the smallest number N such that if the edges of a K_N are colored blue and red, then there must be a red K_a or a blue K_b. Ramsey numbers can use more than 2 colors. For example, R(3, 3, 3) is equal to the smallest N such that if you 3-color the edges of a K_N, you must have a monochromatic triangle. Ramsey's theorem states that the numbers R(a, b, c, …) exist and are finite. Example: R(3, 3, 3), the only nontrivial Ramsey number known involving more than 2 colors, is equal to 17. ■

Suggested Reading

Soifer, *Mathematics as Problem Solving*.

Zeitz, *The Art and Craft of Problem Solving*, sec. 3.3.

Questions to Consider

1. Given a unit square, show that if 5 points are placed anywhere inside or on this square, then 2 of them must be at most $\frac{\sqrt{2}}{2}$ units apart.

2. People have at most 150,000 hairs on their head. How many people must live in a city in order to guarantee that at least 10 people have exactly the same number of hairs on their head?

The Greatest Unifier of All—Invariants

Lecture 15

The central idea of this course is the analytic approach to problem solving, looking at things from a higher level. We want, as often as possible, to have a bird's-eye view. Invariants are a very, very high-level way at looking at many, many problems. It's important to cultivate this attitude of deconstructing and analyzing problems that are solved and even unsolved to see the central underlying ideas behind them.

Invariants are central to mathematics, yet most laypeople have never heard of them. In this lecture, we show how the concept of invariants contains both symmetry and parity. We tweak it to look at monovariants and use these to study some interesting games. We also continue our study of modular arithmetic, which we now see is merely a special case of the grand unifying principle of invariants.

An invariant is any quantity or quality that stays unchanged. A geometric example is the power of a point theorem. In any circle, when 2 chords intersect inside a circle, they obey the equation $AE \times EB = DE \times EC$. Likewise, when chords intersect outside the circle, $EA \times EB = ED \times EC$.

These 2 theorems seem like related results about intersecting lines and circles, but in fact they are actually manifestations of a single fact. For any fixed point P and any fixed circle, draw any line through P that intersects the circle in points X and Y. Define the power of P to be the quantity (PX) (PY). The power of a point theorem says that for any fixed circle and any fixed point, this quantity is invariant, no matter which line we choose. This invariant formulation is true no matter where the point is located.

The Hotel Room Paradox

This is a classic brainteaser that exploits our native instinct to look for invariants. Three women check into a hotel room that advertises a rate of $117 per night. They each give $40 to the porter, and they ask him to bring back $3. The porter goes to the desk, where he learns that the room is actually only $115 per night. He gives $115 to the desk clerk and gives the guests back each $1, deciding not to tell them about the actual rate. Thus the porter has pocketed $2, while each guest has spent $39, for a total of 2 + (3 × 39) = $119. What happened to the other dollar?

- The question is not what happened to the dollar, but how do the variables relate to one another? What is invariant?

- The money the guests paid is equal to the amount that the hotel received ("hotel" means the porter and the desk). In other words, if g, p, and d are respectively equal to what the guests pay, what the porter pockets, and what the desk receives, then the quantity $g - p - d$ is an invariant, always equal to 0.

- The "paradox" is the fact that $g + p = 119$, which is close to 120. But this quantity is not invariant, and it can assume many values. For example, if the actual price of the hotel was $100, then the porter would give the desk $100, return $3 to the guests, and keep $17. Then $g + p = 119 + 17 = 126$, which seems less paradoxical.

- What confuses people is that they think $120 is invariant. And it is, as long as you think clearly about it: 120 is not the invariant amount of dollars "in circulation." Instead, 120 is the invariant "net worth" of the guests: the sum of the dollars they possess, the value of their room, and the amount that the porter stole. These numbers indeed add up to 120.

Here is another classic puzzler: Bottle A contains a quart of milk, and bottle B contains a quart of black coffee. Pour a half-pint of coffee from B into A, mix well, and then pour a half-pint of this mixture back into bottle B. What is the relationship between the fraction of coffee in A and the fraction of milk in B? It is possible to do this with algebra, and when we do so, we discover the surprising fact that the fraction of coffee in A is equal to the fraction of milk in B. But this can made obvious once we think about invariants. The coffee and milk both satisfy conservation of mass (really, volume). So if bottle A has x ounces of coffee "pollution," then bottle B is missing x ounces of coffee and thus has x ounces of milk pollution. Both bottles are equally polluted.

A monovariant is a quantity that changes, but only in one direction.

Here is an example using a parity invariant: Let $a_1, a_2, \ldots, a_n$ represent an arbitrary arrangement of the numbers 1, 2, 3, … , n. Prove that if n is odd, the product $(a_1 - 1)(a_2 - 2)\cdots(a_n - n)$ is even. First, we give a solution that uses a pigeonhole argument. An alternative solution uses invariants. Given any permutation $a_1, a_2, \ldots, a_n$, observe that the quantity $(a_1 - 1) + (a_2 - 2) + \cdots + (a_n - n)$ is invariant; namely, equal to zero! Thus the terms in the product we are interested in add up to zero, and there are an odd number of them. Clearly, they cannot all be odd, since a sum of an odd number of odd numbers is always odd, and zero is even. So one term must be even.

Here is an example that uses congruence: At first, a room is empty. Each minute, either 1 person enters or 2 people leave. After exactly an hour, could the room contain 100 people? Get your hands dirty to work out examples. Is there anything that all the possible outcomes have in common at a fixed time? Yes! Suppose the population is p at some time. A minute later, the population will be either $p + 1$ or $p - 2$. Notice that these numbers differ by 3. This pattern will continue indefinitely, so at any fixed time, there will be many different outcomes, but they will all be congruent (mod 3). In other words, for any fixed time, population is invariant modulo 3. One possible outcome is a population of 60. Thus, all possible outcomes in 60 minutes are congruent to 60 (mod 3). Since 60 is a multiple of 3, and 100 is not, we conclude that the population cannot equal 100 after an hour.

A monovariant is a quantity that changes, but only in one direction. Monovariants are useful for studying evolving systems. Here is a simple example (due to John Conway), called Belgian waffles. Two people take turns cutting up a waffle that is 6 squares × 8 squares. They are allowed to cut the waffle only along a division between the squares, and cuts can be only straight lines. The last player who can cut the waffle wins. Is there a winning strategy for the first or second player? This is actually a fake game. There is no strategy because of this simple monovariant: Each move increases the number of pieces by 1! You start with 1 piece (the whole waffle), so the game ends in 47 moves no matter what the players do!

The next problem illustrates the power of monovariants to cut through the complexity of evolving systems. At time $t = 0$ minutes, a virus is placed into a colony of 2009 bacteria. Every minute, each virus destroys 1 bacterium, after which all the bacteria and viruses divide in 2. For example, at $t = 1$, there will be 2008 × 2 = 4016 bacteria and 2 viruses. Will the bacteria be driven to extinction? If so, when will this happen? There are complicated algebraic formulas that will do the trick, but monovariants are a better way. Let b and v be the respective populations at a certain time, and let b' and v' be the populations 1 minute later. It is easy to see that $b' = 2(b - v) = 2b - 2v$ and that $v' = 2v$. The trick is to manipulate these quantities to recover something that is almost constant. If we divide, we get

$$\frac{b'}{v'} = \frac{2b-2v}{2v} = \frac{b}{v} - 1\,.$$

In other words, the ratio b/v is a monovariant: It decreases by 1 each minute. At $t = 0$, the ratio is 2009/1 = 2009. In 2009 minutes, it will decrease to 0, and the bacteria will be wiped out. ■

Suggested Reading

Engel, *Problem-Solving Strategies*, chap. 1.

Zeitz, *The Art and Craft of Problem Solving*, sec. 3.4.

Questions to Consider

1. A graph that can be drawn so that edges do not cross is called a planar graph. For example, a typical picture of a K_4 has the 2 diagonal edges crossing, but it is possible to draw this graph where one diagonal is "inside" and the other is "outside," so that K_4 is planar. Given any planar graph, it is easy to count the number of vertices (v), edges (e), and regions bounded by edges (r). Discover an invariant involving these variables.

2. Can you find distinct integers a, b, and c such that $a - b$ evenly divides $b - c$, $b - c$ evenly divides $c - a$, and $c - a$ evenly divides $a - b$?

Squarer Is Better—Optimizing 3s and 2s

Lecture 16

In this lecture, we will return to our old friend symmetry to explore questions of distribution and optimization. Along the way, we will develop a new proof method called algorithmic proof, where we imagine a sequence of steps, an algorithm, which [is] guaranteed to solve our problem.

Our anchor problem is an International Mathematical Olympiad problem about a maximal product: Determine, with proof, the largest number that is the product of positive integers whose sum is 1976. Intuition may tell us to try a square: 988 × 988 = 976,144. But we can do better. For example, 987 × 987 × 2 = 1,948,338. Clearly we need more investigation! First, let's consider simpler, more constrained questions.

- Warm-up: A rectangle is made of 12 inches of wire. What should the dimensions be to maximize the area?
- One solution is to appeal to symmetry: Obviously, the rectangle of largest area is the most symmetrical. So the dimensions are 3 × 3.
- How do we do this rigorously? General question: If x and y have fixed sum S, what is the maximum value of the product $P = xy$, and what will x and y be when this maximum is attained?
- Conjecture: Maximum is $(S/2)^2$, when $x = y = S/2$.
- This can be proven with calculus.

But a better way to prove this squarer-is-better principle is with a picture. Let $S = x + y$. In other words, S is the diameter of the semicircle. By similar triangles, $g/x = y/g$, so $g = \sqrt{xy}$. The maximum value of xy is attained when $x = y = S/2$. As the distance between 2 positive numbers decreases, their product increases, provided that their sum stays constant. Note that this is a dynamic principle.

A reformulation of this is the 2-dimensional arithmetic-geometric mean inequality (AM-GM): If x and y are nonnegative, then $(x + y)/2 \geq \sqrt{xy}$, with equality attained when $x = y$. The AM-GM is also true in higher dimensions. For example, in 3 dimensions, the statement is $(x + y + z)/3 \geq \sqrt[3]{xyz}$, which is very hard to prove using algebra alone. The 2-dimensional AM-GM is not too hard to prove with algebra. It is equivalent to $(x-y)^2 \geq 0$, which is certainly always true. But it is hopeless to prove the general n-dimensional AM-GM with algebraic methods.

But a better way to prove this squarer-is-better principle is with a picture.

However, we can use the 2-dimensional squarer-is-better principle to prove AM-GM in any dimension if we leave algebra behind and instead view the problem in terms of physics. The original formulation for n variables is as follows:

- Let $x_1, x_2, \ldots, x_n$ be positive real numbers.
- Then $\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 x_2 \cdots x_n}$, with equality only when the numbers are equal.

Here is a reformulation: Let the sum S of $x_1, x_2, \ldots, x_n$ be fixed, and let the product be P. Then the AM-GM asserts that $S/n \geq \sqrt[n]{P}$, with equality only when the x_i are all equal. Raising by the n^{th} power, we get $P \leq (S/n)^n$. We prove this reformulated version by using an algorithmic method with weights. We make the process of finding an optimal product a physical process. The breakthrough idea is to make the x_i unit weights placed on a number line. Then their average value corresponds to their balancing point!

Let's return to the 1976 International Mathematical Olympiad problem. We have seen how powerful symmetry is. Our instinct that equal values optimize products has been rigorously proven. But what if we are unable to make the parts equal?

- There are 2 difficulties: We are not told how many numbers are in the product. And even if we were, we could not guarantee that the values would be integers if we made them all equal.

- Begin your investigation by replacing 1976 with smaller values.

- Conjecture: Use only 3s and 2s.

- Why? Once again, try an algorithmic method. Assume that we have a maximal product of integers with a fixed sum and that we have numbers that are not 3s and 2s.

- Conclusion: The only numbers possible in an optimal product are 2s and 3s.

  - Next, notice that if you have three 2s, you can replace them with two 3s.

  - So that gives us the optimal breakdown: all 3s, unless you have to have some 2s, but never more than two 2s. For example, 12 breaks into 3 × 3 × 3 × 3, but 13 gives us 3 × 3 × 3 × 2 × 2.

- And 1976 breaks into $3^{658} \times 2$. ■

Suggested Reading

Kazarinoff, *Geometric Inequalities*.

Niven, *Maxima and Minima without Calculus*.

Questions to Consider

1. If you have a fixed length of wire and you are to make a rectangle of maximum area, you know that a square is optimal. But what if one side of your rectangle is already provided? For example, suppose you are building a rectangular fence, one of whose sides is a river? What will the optimal dimensions be? (Hint: symmetry.)

2. Use the arithmetic-geometric mean inequality and symmetry to prove the nice inequality $(a + b)(b + c)(c + a) \geq 8abc$, for positive a, b, and c.

Using Physical Intuition—and Imagination

Lecture 17

This lecture, like the earlier one about games, does not attempt to teach you new problem-solving strategies or tactics. Instead, we will look at a few problems in depth and use methods that we've already seen. Our goal is to continue to reinforce the all-important ideas of symmetry, but now we'll add its "mother," invariance, now that we've learned something about that. In particular, we will channel our intuition about the so-called "real world" to use physical invariants to solve a problem.

This lecture is inspired by a problem that I proposed for the USA Mathematical Olympiad only to find out that a variant of it had been used for years as a job interview question for a hedge fund. This challenging problem about marbles on a track combines much of what we have studied: invariants, symmetry, drawing pictures, and getting your hands dirty. Notably, we use physical intuition to get to the heart of the problem.

The Marbles on a Track Problem

Several marbles are placed on a circular track of circumference 1 meter. The width of the track and the radii of the marbles are negligible. Each marble is randomly given an orientation, clockwise or counterclockwise. At time zero, each marble begins to travel with speed 1 meter per minute, where the direction of travel depends on the orientation. Whenever 2 marbles collide, they bounce back with no change in speed, obeying the laws of inelastic collision. What can you say about the possible locations of the marbles after 1 minute with respect to their original positions? There are 3 factors to consider: the number of marbles, their initial locations, and their initial orientations.

This is a challenging problem, and we need a good venue for investigation. The geometry of the circle is irrelevant; only the fact that there is no beginning or end is important. Hence the following warm-up, Martin Gardner's classic airplane problem. Several planes are based on a small island. The tank of each plane holds just enough fuel to take it halfway around the world. Fuel can be transferred from the tank of 1 plane to the tank of another while the planes are in flight. The only source of fuel is on the island, and we assume that there is no time lost in refueling either in the air or on the ground. What is the smallest number of planes that will ensure the flight of 1 plane around the world on a great circle, assuming that the planes have the same constant speed and rate of fuel consumption and that all planes return safely to the base?

We use physical intuition to get to the heart of the problem.

- Key idea: Circular motion can be modeled on a distance-time graph; just remember that the start and end are really the same point.

- We will show that it is possible to fly around the world with just 3 planes. Call the planes A, B, and C.

  - First A, B, and C leave together, flying for 1 unit (1/8 of the way around the world).

  - Then C transfers 1 unit each to the other 2 planes. This gives C enough to return to base, while the other 2 now have full tanks.

  - A and B travel for 1 more unit, and then B transfers 1 unit to A. This leaves A with a full tank and B with enough to return to base.

- *A* now has enough fuel to get to within 2 units of the destination. As soon as *B* returns to base (when *A* has reached the halfway point, with 2 units of fuel left), *B* refuels and heads for *A*'s location (traveling backward this time).

- When *B* reaches *A*, *B* transfers 1 unit to *A*, and they both fly toward base. Meanwhile, *C* heads out again, reaching *A* and *B*, transferring 1 unit to each of them, and then all 3 head home.

- Notice how the second half of the story is symmetrical with respect to the first.

Here is another warm-up problem, one that uses reflection: A laser strikes mirror *BC* at point *C*. The beam continues its path, bouncing off mirrors *AB* and *BC* according to the rule angle of incidence equals angle of reflection. If *AB* = *BC*, determine the number of times the beam will bounce off the 2 line segments (including the first bounce, at *C*).

- Key insight: The broken-line path of the laser will be unbroken if you reflect across the mirror.

- If it works once, it can be done again. Relentless reflection solves the problem by straightening out the path.

- Instead of the actual laser path, look at the straight line. We merely need to count the number of times it intersects the boundary of one of the reflected mirrors. Answer: 6 bounces. ■

The Marbles on a Track Problem, Solved

It is possible to find starting positions that change after 1 minute.

In one example, we start with 5 balls, and they end up in different positions. However, the ending locations are a permutation of the original locations. For example, the black ball is now at the starting position of the blue ball, and the blue ball is where the green ball was. The principle of reflection and the clever use of a simple distance-time representation on graph paper help us see that the problem is not that complex. Suppose we were color-blind. A diagram of the paths would be the same as before, but we could not keep track of which marble is which. Pretend that the marbles are ghosts that can pass through one another. Then, of course, each marble ends up at exactly the same spot where it began.

This explains why the final positions of the marbles must coincide with the original positions, up to a permutation. Which permutations are possible? Remember that marbles cannot actually pass through one another, so the order of the marbles cannot change. All that can happen is a cyclic permutation. How do we predict the permutation? If we started with 6 balls, there are 6 possible permutations. However, there are 2 × 2 × 2 × 2 × 2 × 2, which equals 64, different choices of initial orientation for the balls. How do these orientations influence the final result?

An invariant saves the day: Collisions only happen between balls of opposite velocity, and when they collide, the 2 balls swap velocities (and paths). Thus the sum of the velocities is constant. This is the same as conservation of angular momentum. In general, if the net clockwise excess is c, then the net clockwise travel will be c full rotations of the circle. The only cyclic permutation that accomplishes this is the one in which each ball moves c balls clockwise. The key ideas we used were physical intuition, looking for invariants, and symmetry.

Suggested Reading

Gardner, *Martin Gardner's Sixth Book*, chap. 4.

Kendig, *Sink or Float?*

Tanton, *Solve This.*

Questions to Consider

1. A monk climbs a mountain. He starts at 8 am and reaches the summit at noon. He spends the night on the summit. The next morning, he leaves the summit at 8 am and descends by the same route that he used the day before, reaching the bottom at noon. Prove that there is a time between 8 am and noon at which the monk was at exactly the same spot on the mountain on both days. (Notice that we do not specify anything about the speed that the monk travels. For example, he could race at 1000 miles per hour for the first few minutes, then sit still for hours, then travel backward, etc. Nor does the monk have to travel at the same speeds when going up as going down.)

2. Imagine a laser beam that starts at the southwest corner of a square and moves northeast with a slope of 7/11. How many times will it bounce before it returns to its starting point?

Geometry and the Transformation Tactic

Lecture 18

There's lots of geometry to study. ... What we will look at is transformational geometry. I've chosen it because of its unexpectedness and its very much higher-order connections to problem-solving ideas that we've already seen, such as symmetry and invariance.

This is the only lecture in the course wholly devoted to geometry. We look at geometric transformation, a great example of the problem-solving strategy of reversing one's point of view. We apply this idea to a number of problems that initially appear completely intractable but become almost trivial once we are comfortable with dynamic entities like rotations, vectors, and reflections.

We devote this lecture to a tiny fraction of transformational geometry because of its connections to ideas already familiar to us, in particular, symmetry and invariants. Before we get into the details of transformational geometry, here is a problem to think about that we will solve later with the transformational methods we develop. Suppose a pentagon (not necessarily regular) is drawn on the plane. The midpoints of each side are found. Then, suppose the original pentagon is erased, leaving only the midpoints. Can the original pentagon be reconstructed?

Transformational geometry was pioneered in 1872 by the German mathematician Felix Klein and is an example of the strategy of reversing one's point of view.

Transformational geometry was pioneered in 1872 by the German mathematician Felix Klein and is an example of the strategy of reversing one's point of view. Klein suggested that the proper way to think about geometry was not to focus on the objects but instead to contemplate the transformations that act on them. Why are transformations important? Because they are not just geometric but also algebraic entities. You do not add or subtract transformations, but you can sort of multiply them by composition.

In our exploration of the algebra of transformations, we restrict our attention to the plane and look at just 3 types of transformations: reflections, rotations, and translations. Let's look at the composition of reflections: Let F_h denote a reflection across line *h*. Note that F_h leaves all the points of *h* invariant. In general, the fixed points of a reflection are a line. When composing 2 reflections, there are 3 cases: The lines are the same, parallel, or meet in a single point.

Let's look at the composition of 2 rotations. First we need a lemma about what happens to a line under rotation. Rotated line lemma: Suppose line *h* is rotated by a rotation with center *A* and angle α. Let h' be the image of *h*. Then h' makes an angle of α with *h*. The proof is a similar triangles argument. Now that you have a feel for rotating lines, here is a fantastic example. Suppose you are given 3 parallel lines: ℓ_1, ℓ_2, and ℓ_3. Is it possible to construct an equilateral triangle such that each vertex of the triangle lies on one of each of these lines? Starting with *A*, what transformation leaves parts of the triangle invariant?

- One idea: Clockwise rotation by 60° about *A*.
- Now consider this rotation, but let line ℓ_3 go along for the ride!
- Its image is ℓ_4, and where it intersects ℓ_2 is the point *B*. Now we can construct the triangle.

By using parts of the figure and the knowledge that the rotation moved one point of the unknown triangle to another unknown point, we were able to construct the location of both points! Now we are ready to understand the composition of 2 rotations. No matter where the centers are, the composition is just another rotation about this center, where the angle is just the sum of the 2 angles. But if the angles for 2 rotations add up to 360°, then the composition of the 2 rotations is a translation. This is an unexpected but helpful fact that is the tool we need for the pentagon problem. Let's call it the reflection-translation tool.

Now let's use this tool to solve the pentagon problem that we began our lecture with.

- Use wishful thinking to pretend you know the pentagon's vertices: A, B, C, D, and E. In reality, the only points that we know about are the midpoints: X, Y, Z, V, and W.

- Crux idea: Rotations of 180° about midpoints. These rotations bring (unknown) vertices to vertices.

- Compose the 5 rotations by 180° about X, Y, Z, V, and W; this brings A back to its starting point.

- So do one more rotation of 180°, about X again. That brings A to B.

- It was a composition of 6 rotations, each of 180°, and 6 × 180° = 1080° = 3 × 360°.

- So it is a translation!

- We can perform this 6-rotation composition on any point we like. It took A to B. However, there is one problem: We do not know where A is!

- Just pick a random point P. Perform the 6-rotation composition on P, getting successive points $P_1, P_2, \ldots, P_6$.

- We know that the vector from P to P_6 is the same as the vector from A to B!

- The line segment joining P_1 and P_6 has the same length as the mystery side AB of our pentagon. Draw a line parallel to this segment that goes through X, and mark off equal segments on either side of X whose length is half of AB. We have just reconstructed segment AB!

Why did it work? The higher-level reason for this is that we did not use the transformations randomly, but instead searched for transformations that leave parts of our problem invariant. ■

Suggested Reading

Liu, *Hungarian Problem Book III.*

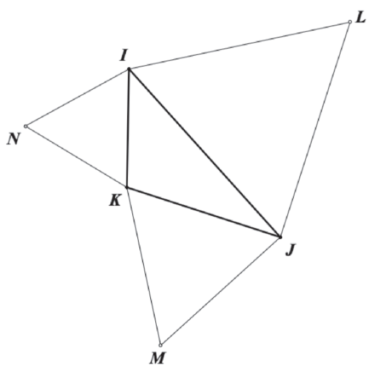
Needham, *Visual Complex Analysis*, chap. 1.

Yaglom, *Geometric Transformations I.*

Zeitz, *The Art and Craft of Problem Solving*, sec. 8.5.

Questions to Consider

1. Let *IJK* be an arbitrary triangle with equilateral triangles constructed on each edge. Thus *IJL*, *KJM*, and *IKN* are all equilateral triangles. Prove that *IM*, *KL*, and *JN* have exactly the same length. (Hint: Perform a rotation or 2.)



2. Let R_k denote rotation by 90° counterclockwise about the point $(k, 0)$ in the plane. The composition of R_0, R_1, R_2, R_3, in that order, has a total of 360° of rotation and hence is a translation (possibly the identity). What translation is it? Can you generalize to n rotations, where each is about $(k, 0)$, and the angle is $360/n$?

Building from Simple to Complex with Induction

Lecture 19

Induction has its own context. Inductive proofs usually involve problems that involve evolving structures that build upon simpler structures or problems that have recurrence, for example, the recurrence relation defining the Fibonacci numbers.

Mathematical induction is the natural way to prove assertions that are recursive, that is, where simpler cases evolve into more complex cases that depend on the earlier cases. Our cornerstone problem is a folkloric tiling of a punctured chessboard, and we also apply induction to combinatorial geometry and a probability problem from the Putnam exam.

Mathematical induction is a proof method closely related to algorithmic proof. It especially works with problems involving evolving structures that build upon simpler structures. Let's begin with a folklore problem that we will eventually solve using mathematical induction: Consider a $2^{2009} \times 2^{2009}$ chessboard with a single 1×1 square removed. Show that no matter where the small square is removed, it is possible to tile this "punctured" chessboard with L-trominos (2×2 squares with one 1×1 square removed).

We know that 2009 is a red herring, so we focus on whether it will work for any chessboard of size $2^n \times 2^n$. It certainly works when $n = 1$ or $n = 2$. Can we bootstrap from $n = 2$ to $n = 3$ to $n = 4$ and so on? Mathematical induction allows us to prove an empirical pattern and show how it extends indefinitely. In general, mathematical induction proof involves a sequence of propositions, P_n, indexed by natural numbers. We wish to prove that $P_1, P_2, \ldots$ are all true. To do this, we first show that P_1 (the base case) is true. Then we need to establish the principle that if each case is true, the next one will be as well: In other words, for all positive integers n, if P_n is true, then P_{n+1} will also be true. The assumption that P_n is true is called the inductive hypothesis. We do not know if it is true, but we use its truth to prove that P_{n+1} is true.

A simple example: The plane is divided into regions by straight lines. Show that it is always possible to color the regions with 2 colors so that adjacent regions are never the same color. Let's call colorings such as the above "nice." Define P_n to be "If the plane is divided into regions by n straight lines, then it is possible to color nicely." This is clearly true for $n = 1$ (we have proven the base case). But what about, say, $n = 10$? We need an algorithm for moving from the n^{th} case to the $(n + 1)^{\text{th}}$ case. Keep things concrete: Suppose I can nicely color any 5-line configuration. How can I use this to nicely color an arbitrary 6-line configuration? Here is the formal solution for the inductive step.

- Suppose P_n is true. Given any $(n + 1)$-line configuration, temporarily ignore one of the lines.
- Now you have an n-line configuration, which you know you can color nicely.
- Finally, invert the colors on one side of the $(n + 1)^{\text{th}}$ line.

Another problem about lines: Lines in a plane are in general position if no 2 are parallel and no 3 meet in a point. If 10 lines are drawn in general position in the plane, into how many regions do they divide the plane?

Clearly we want to discover a formula for n lines. If we let R_n denote the number of regions made by n lines in general position, we conjecture that $R_n = n + R_{n-1}$. Notice that this is the key to a rigorous induction proof, because it actually suggests the way that you go from the $(n - 1)^{\text{th}}$ case to the n^{th} case (from one case to the next case). Suppose we have $n - 1$ lines in general position, creating R_{n-1} regions.

Mathematical induction is a proof method closely related to algorithmic proof.

Imagine drawing a new line so that it is not parallel to any of the other lines and does not intersect them in any of the previous intersection points. This new line will intersect each of the $(n - 1)$ old lines, producing a new region each time an intersection is achieved. When the new line intersects the last of the old lines and exits, one final new region will be produced, for a total of n new regions.

Here is a probability example from the 2002 Putnam exam: Shanille O'Keal shoots free throws on a basketball court. She hits the first and misses the second; thereafter, the probability that she hits the next shot is equal to the proportion of shots she has hit so far. What is the probability she hits exactly 50 of her first 100 shots?

- After the second toss, the proportion of successes is 1/2, so on toss 3, we have a 1/2 chance of getting another basket.

- For toss 4, it gets more complex. We employ the draw a picture strategy loosely and create a tree diagram that shows all the scenarios.

- The outcomes have equal probability! This is somewhat surprising, but we can attempt to prove it with induction.

- Let $P(b, t)$ denote the probability that we get b baskets in t tosses. Our conjecture is that $P(b, t) = 1/(t - 1)$ for each of the valid values of b (between 1 and $t - 1$).

- We will prove this by induction on t. This time, our base case is $t = 2$, and it is trivially true.

- Suppose that $P(b, t) = 1/(t - 1)$ for some value of t greater than or equal to the base case of 2. We will use this to prove that $P(b, t + 1) = 1/t$.

- How do we get b baskets in $t + 1$ tosses? Either we get a basket on the $t + 1$ toss, or we do not.

  - Suppose we do not get a basket. Then we accumulated b baskets in the first t tosses. By the inductive hypothesis, this has probability $1/(t - 1)$. At this point, the probability we will get a new basket is equal to the current proportion, b/t. So the probability that we end up with b baskets by missing the last toss is $[1/(t - 1)](1 - b/t) = (t - b)/t(t - 1)$.

- Now suppose we do get a basket on the final toss. Then we accumulated $b - 1$ baskets in the first t tosses. By the inductive hypothesis, this also has probability $1/(t - 1)$. The probability we will get a new basket is equal to the current proportion, $(b - 1)/t$. So the probability that we end up with b baskets by missing the last toss is $[1/(t - 1)](b - 1)/t = (b - 1)/t(t - 1)$.

- Adding these 2 probabilities, we get $(b - 1)/t(t - 1) + (t - b)/t(t - 1) = (t - 1)/t(t - 1) = 1/t$.

The induction proof is formally correct but not fully illuminating. This is a feature of some induction proofs. You can verify how to get from t to $t + 1$, but you sometimes do not know why it works.

Let's return to the tromino problem. Here is a way to go from $n = 3$ to $n = 4$.

- The crux idea: symmetry!

- Given an arbitrary 16 × 16 tile board missing 1 tile, without loss of generality, the hole is in the southwest quandrant.

- Place a tromino in the center so that it takes a single bite out of each of the other quadrants.

- Now all 4 quadrants have a single hole; by the inductive hypothesis (for the 8 × 8 minus 1 board), we can tile each of them!

- Clearly this can be generalized; this is our inductive algorithm. ■

Suggested Reading

Fomin and Itenberg, *Mathematical Circles*, chap. 9.

Goodaire and Parmenter, *Discrete Mathematics with Graph Theory*, sec. 5.1.

Maurer and Ralston, *Discrete Algorithmic Mathematics*, chap. 2.

Questions to Consider

1. Conjecture a formula for the sum of the first n Fibonacci numbers. Then prove your formula by induction.

2. Prove that $F_n < 2^n$, where F_n denotes the n^{th} Fibonacci number.

Induction on a Grand Scale

Lecture 20

Imagine that you had a lot of time and you wrote out many, many rows of Pascal's triangle, like a couple quadrillion rows. Then you took all of those numbers, and you put them on little pieces of paper, and you put it in a hat. Then you shook the hat really, really well, and you picked out a number at random, and you check to see if it's odd or even. What's the probability that that number will be even? That's the question we're going to ask, but since Pascal's triangle is infinite, we have to think about an infinite process.

What is the probability that a randomly chosen number in Pascal's triangle is even? This problem is surprisingly easy to investigate but requires sophistication to resolve. By this stage, you have a good grasp of investigative methods, summation, mathematical induction, and modular arithmetic, so you are ready for this investigation, the first of the advanced lectures as we approach the end of the course.

So what is the probability that a randomly chosen member of Pascal's triangle is even? This is a meaningless question as posed; Pascal's triangle is infinite! A reformulation: Let P_n be the probability that a randomly chosen element from the first n rows of Pascal's triangle is even. Does P_n converge to something as n approaches infinity? We are asking a question about density. Define the density of a subset S of the natural numbers by computing the probability that a randomly chosen integer from the first n integers is in S. Then we see what happens to P_n as n gets arbitrarily large. If it converges, then that is S's density. The density of the even integers is 1/2. The density of the perfect squares is 0, since among the first n^2 integers, exactly n are squares, so the relative frequency is $1/n$, which gets arbitrarily small.

Recall that we called the elements of Pascal's triangle binomial coefficients and asserted that the elements of row n Pascal's triangle were the coefficients of the binomial $(1+x)^n$. For example, $(1+x)^4 = x^4 + 4x^3 + 6x^2 + 4x + 1$, and indeed, the coefficients are the numbers in row 4. This is actually quite easy to prove, now that we know about induction. Let P_n be the statement that the coefficients of $(1+x)^n$ are the elements of row n. The base case is obvious, since row 0 is just 1, and row 1 is 1, 1. We want to prove the inductive step now, but let's keep it concrete and informal. We will just show that P_5 implies P_6; our argument will generalize easily.

- Start with the inductive hypothesis: $(1 + x)^5 = 1x^5 + 5x^4 + 10x^3 + 10x^2 + 5x + 1$.

- Thus $(1+x)^6$ will be this polynomial, multiplied by $(1 + x)$.

- This is exactly how we get row 6 of Pascal's triangle: $(1+x)^6 = 1x^6 + 6x^5 + 15x^4 + 20x^3 + 15x^2 + 6x + 1$.

Now let's look at the parity of the numbers of Pascal's triangle. We will work (mod 2). Look at the first 9 rows. At first, there are not many evens at all. But row 4 and row 8 are all even, except for the ubiquitous 1s that start and end every row. And notice that rows 3 and 7 are all 1s. We conjecture that row $2^n - 1$ will be all 1s and that row 2^n will be all 0s, except for the first and last terms. Clearly, the second statement follows from the first, but how do we prove the first statement?

Now look at rows 0–32. We see a fractal structure, with inverted triangles of 0s. What causes them? The seed of the 0 triangles is the row of all 1s, since this forces the next row to be all 0s (except the first and last terms). The natural way to look at the parity of Pascal's triangle is by successive doublings of it. Let T_n be the n^{th}-order triangle that ends with a row of 1s. T_1 is the triangle consisting of rows 0 and 1 (i.e., it contains three 1s), and T_2 is the triangle that is built out of 3 copies of T_1, with a 0 in the middle. In general, T_n is the triangle that ends with row $2^n - 1$, which is all 1s. This starts 2 seeds at opposite ends, with 0s in between, which then grow 2 more copies of T_n, producing a new structure, T_{n+1}.

Now that we have inductively proven the fractal structure of Pascal's triangle, we can try to count the even terms. This turns out to be complicated, but using the flip your point of view strategy, we instead look at odd terms. This is nearly trivial, since T_1 has exactly three 1s, and T_{n+1} is composed of 3 copies of T_n with 0s in the center. If we define U_n to be the number of 1s in T_n, we get the simple formula $U_n = 3^n$. Our final step is to compute the relative fraction of 1s in T_n and then let n get large.

How many elements are in T_n? It is a triangle starting with one element and ending with 2^n elements. So the number of elements in T_n is the 2^n triangular number! This is equal to $1 + 2 + \cdots + 2^n = 2^n(2^n + 1)/2$. Thus, the probability that an element in the first 2^n rows is odd is equal to $2(3^n)/2^n(2^n + 1)$. Despite the factor of 2 in the numerator, the 4^n in the denominator will eventually overpower it, so the limit is 0. A more rigorous way to see this is by dividing numerator and denominator by 4^n. As n grows larger, the entire fraction approaches 0. So the probability that an element is odd approaches 0. That means that the probability an element is even approaches 100%, which is truly surprising. In other words, essentially all binomial coefficients are even!

By this stage, you have a good grasp of investigative methods, summation, mathematical induction, and modular arithmetic.

We just proved an absolutely amazing fact about long-term convergence of parity, an asymptotic property of Pascal's triangle. But it would be nice to analyze the parity in a more exact way. In Lecture 4, we counted the number of evens and odds in each row, and the number of odd terms was a power of 2. Which power of 2? What is the appropriate point of view for investigating powers of 2? The binary (base-2) system, where we write numbers as sums of powers of 2. When we make a table of the number of odd terms in each row and look at the row numbers in binary, the conjecture is clear: The number of 1s in row n is equal to 2 raised to the number of 1s in n when n is written in base 2!

But why? Remember that the elements of Pascal's triangle are the coefficients of $(1+x)^n$. When $n = 2$, we have $(1+x)^2 = 1+2x+x^2$, but (mod 2), the middle term disappears. So $(1+x)^2 = 1+x^2$ (mod 2). If we square this again, we get $(1+x)^4 = (1+x^2)^2 = 1+x^4$ (mod 2). In general, we see that for any n, $(1+x)^{2^n} = 1+x^{2^n}$ (mod 2). This immediately explains why row 2^n has just two 1s!

But what about an arbitrary row, say, row 11?

- Write 11 in binary: 1011.
- Then look at row 11 of Pascal's triangle by expanding $(1+x)^{11}$.
- We see that $(1+x)^{11} = (1+x)^8(1+x)^2(1+x)^1$, using the binary representation. There were three 1s in the binary representation for 11, and hence there are 3 terms in the product.
- But reducing this modulo 2, we get $(1+x)^{11} = (1+x^8)(1+x^2)(1+x)$, a product of 3 binomials. When they are multiplied out, we will have 8 different nonzero terms.

Here is a slicker way to see it: When $(1+x)^{11}$ is multiplied out and simplified modulo 2, it will be a sum of powers of x^n, where the coefficients will either be 0 or 1. ■

Suggested Reading

Edwards, *Pascal's Arithmetical Triangle*.

Tabachnikov, *Kvant Selecta*, chap. 1.

Questions to Consider

1. There is a fun pattern in Pascal's triangle: Row 0 is 1, row 1 is 11, row 2 is 121, row 3 is 1331, and row 4 is 14641. Notice that for each *k*, row *k* is 11 raised to the k^{th} power! Explain why this pattern is true and why it fails for *k* greater than 4.

2. Investigate the same question that we did in the lecture, but modulo 3. In other words, look at the patterns of when elements of Pascal's triangle are multiples of 3. It is a little more subtle than before, because now there are 3 possible values (mod 3): 0, 1, and 2.

Recasting Numbers as Polynomials—Weird Dice

Lecture 21

Our overall strategy in using generating functions is we'll take a problem, we'll turn it into a sequence, we'll turn that into polynomials, we'll manipulate these polynomials in a useful way using algebra, and that will somehow inform us about the original sequence of numbers and help us to solve our problem.

This is an advanced lecture that uses algebra more than most, including infinite geometric series. Can we renumber 2 dice with positive whole numbers that are not the standard 1, 2, 3, 4, 5, and 6 in such a way that the various sums still range from 2 to 12 inclusive, with the same probabilities as standard dice? Amazingly, the answer is yes. We use generating functions, which glue most of mathematics to polynomial algebra.

Generating functions are a method of using polynomial algebra to recast many types of problems. Any sequence of numbers $a_0, a_1, a_2, \ldots$ gives rise to a generating function, the (possibly infinite) polynomial $a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$. The crux idea behind generating functions is the simple observation that $x^a x^b = x^{a+b}$. Here is the basic generating function strategy.

- A problem gives rise to sequences of numbers.
- The sequences are converted into polynomials.
- The polynomials are manipulated in a useful way with algebra, which may tell us something about our original sequence.

The Dice Problem

Ordinary dice are numbered 1, 2, 3, 4, 5, and 6. When you roll 2 dice, the probability that you get a sum of 10 will be 3/36, since there are exactly 3 ways to get a 10 and there are 36 ways of rolling 2 dice. Our question is whether it is possible to renumber 2 dice with positive integers so that neither is an ordinary die, yet all possible sums occur with the same probability as they do with a pair of ordinary dice.

- Suppose such dice exist. We will call them weird dice. If you roll 2 weird dice (and they may not be 2 identical dice), the probability of getting a sum of 10 will still be 3/36.

- The denominator of 36 is not important; what matters is the number of ways to get each sum.

- Let's label one die $a_1, a_2, \ldots, a_6$ and the other $b_1, b_2, \ldots, b_6$.

- We want the 36 possible sums of $a_i + b_j$ to behave like ordinary dice.

- Note that a weird die can have multiple faces with the same label—for example, three 1s, two 2s, and one 5.

- Our problem is simple. All we need to do is look at all the possible ways to label dice and all the possible sums. A computer could do that in microseconds, but we are not computers.

- How do we organize such masses of data? With generating functions.

Here are some examples of sequences transforming into generating functions or vice versa.

- 1, 2, 3 $\leftrightarrow 1+2x+3x^2$.
- 1, 1, 1, 1, … $\leftrightarrow 1+x+x^2+x^3+x^4+\cdots=1/(1-x)$.
- 1, 7, 21, 35, 35, 21, 7, 1 $\leftrightarrow (1+x)^7$.

Many operations are possible with generating functions, but we will stick to multiplication. Let's look at some examples. Compute $(2 + x)(1 + 3x) = 2+6x+x+3x^2$, using the FOIL method. There are 4 raw terms. What is the coefficient of x^6 in $(x^3+ 2x^2+x)(3x^4+2x^3+x)$? We want to look at the ways we can multiply terms and get the exponent of 6. The x^3 and $2x^3$ and the $2x^2$ and $3x^4$ combine to give an answer of $1 \times 2 + 2 \times 3$, which equals 8.

Generating functions can shed light on combinatorics.

Generating functions can shed light on combinatorics. Consider the simplest type of die (i.e., a coin). Put 0 on one side and 1 on the other. Then the generating function will be $1 + x$. Suppose 7 people are each flipping a coin to decide if they will get a prize (0 = no, 1 = yes). The number of prizes possible ranges from 0 to 7. There are 2^7 different outcomes. Each is encoded by the expansion $(1 + x)^7 = (1 + x)(1 + x)(1 + x)(1 + x)(1 + x)(1 + x)(1 + x) = 1 + 7x + 21x^2 + 35x^3 + 35x^4 + 21x^5 + 7x^6 + x^7$. How many outcomes had 3 prizes? 35. In other words, there are 35 ways to choose 3 prize winners out of 7 contestants. Thus $35=\binom{7}{3}$.

Now we are ready to recast the original dice problem into polynomial form. The generating function for a single die is $D(x) = x+x^2+x^3+x^4+x^5+x^6$. The generating function for the sums of 2 ordinary dice is just $D(x)D(x) = (x+x^2+x^3+x^4+x^5+x^6)^2$. This expands to $x^2+2x^3+3x^4+4x^5+\cdots+2x^{11}+x^{12}$. If weird dice exist, we must have $A(x)B(x)= (x+x^2+x^3+ x^4+x^5+x^6)^2$. Now we have converted a tricky question into a relatively simple one: Can we factor $(x+x^2+x^3+x^4+x^5+x^6)^2$ in a different way?

We can do this with algebraic tools. Use the geometric series tool to simplify $x + x^2 + \cdots + x^6 = x(x^6 - 1)/(x - 1)$. We have to factor $(x + x^2 + \ldots + x^6)^2 = x^2(x^6 - 1)^2/(x - 1)^2$. We need to get rid of the denominator. The full factorization thus is $x^2(1 + x)^2(1 - x + x^2)^2(1 + x + x^2)^2$, and now it is a matter of rearranging them in a nonsymmetrical way. In other words, if we write each distinct prime as $P(x) = x$, $Q(x) = (1 + x)$, $R(x) = 1 - x + x^2$, and $S(x) = 1 + x + x^2$, then $D(x) = PQRS$.

We want to break up the product $DD = PPQQRRSS$ into 2 factors that are not the same. $DD = (PQRS)(PQRS)$ does not work; it gets us back to where we started. There are so many choices; how do we narrow them down? The math team lemma saves the day. For each die, the sum of the coefficients must be exactly 6. So let's look at the sum of the coefficients of our prime factors.

- $P(1) = 1$.
- $Q(1) = 2$.
- $R(1) = 1$.
- $S(1) = 3$.
- Notice that the product of these numbers is 6, as it must be!

So for each die, we need exactly 1 Q and 1 S, since that is the only way to get the proper coefficient sum. We can do whatever we want with the other factors, as long as each die also has 1 P. We conclude that our weird dice are 1, 2, 2, 3, 3, 4 and 1, 3, 4, 5, 6, 8.

Even though we did not draw a single picture, there is a strong correspondence between what we did with generating functions and with transformational geometry. ■

Suggested Reading

Gardner, *Penrose Tiles to Trapdoor Ciphers*, chap. 19.

Graham, Knuth, and Patashnik, *Concrete Mathematics*, chap. 7.

Wilf, *generatingfunctionology.*

Zeitz, *The Art and Craft of Problem Solving*, sec. 4.3.

Questions to Consider

1. The coefficient of x^{13} when you expand $(1+x)^{17}$ is, of course, $\binom{17}{13}$.

But you can also write $(1+x)^{17}$ as $(1+x)^{10}(1+x)^{7}$ and expand each of those factors. What binomial coefficient identity do you get?

2. Consider the infinite series $1+D(x)+\left[D(x)\right]^2+\left[D(x)\right]^3+\cdots$, where $D(x)$ is the die generating function. When this infinite series is simplified, it is the generating function for what easily stated sequence?

A Relentless Tactic Solves a Very Hard Problem

Lecture 22

There is no limit to what we can achieve if we harness the infinitely compressible universe of mathematics and couple it with the irrepressible imagination of a good problem solver.

This advanced lecture is a continuation of the ideas begun in Lecture 14. We use the pigeonhole principle relentlessly to study Gallai's theorem, a Ramsey-style assertion. Our investigation takes us into the realm of the nearly infinite, where we contemplate numbers far larger than the number of atoms in the universe. The strategic principle we highlight is more earthbound: Don't give up.

Our focus will be on finding structure within seemingly randomly colored lattices. Our basic tool is just the pigeonhole principle and the notion of coloring. If you 2-color the points on a number line, you only need to look at 3 points, and you are guaranteed that 2 are the same color. If you have a 3×3 grid, there are 2^9 ways to color it. If you were using 10 colors, it would be 10^9.

Let's do a warm-up problem: Color the lattice points of the plane in 2 colors. Prove that there must be a rectangle (with sides parallel to the axes) each of whose vertices are the same color. The pigeonhole principle applied to 3 consecutive lattice points in a horizontal line forces there to be at least 2 points of the same color (points are pigeons, colors are holes). We would be done if we had 2 identical patterns, one on top of the other. But how do we guarantee this can happen?

Here's the crux idea: Look at 9 rows of 3 points, and we are guaranteed that 2 of the rows will be identical! After all, there are only 8 different 2-colorings of 3 points! And since each row contains 2 points of the same color (at least), we will have a monochrome rectangle. Thus every 3×9 grid of points contains a monochrome rectangle. This is a worst-case scenario, of course; we could have gotten lucky with just a 2×2 grid, but 3×9 guarantees it. This can be easily generalized. If we 5-colored the lattice points of the plane, there would still be a monochromatic rectangle, but we would need to look at rows of 6 points (to guarantee that each row contains at least 2 points of the same color) and then consider $5^6 + 1$ (or 15,626) rows.

An innocent, obvious generalization is can we get a monochrome square? This is called Gallai's theorem: If you 2-color the lattice points, you are guaranteed to find a monochrome square. We will proceed as we did with the rectangle and build up our square in stages. First we need a monochrome line, then a monochrome isosceles right triangle, and finally a monochrome square. The line is easy: Just pick any 3 points, and at least 2 of them must be the same color. But how do we get the rest of our right angle? How can we control the distance between the points of the same color?

Our investigation takes us into the realm of the nearly infinite, where we contemplate numbers far larger than the number of atoms in the universe.

We need at most 3 points to get our 2 monochrome points, so our 2 monochrome points lie in at most a 3 × 1 grid. If we are to build a right angle with this as a starting place, we will need a 3 × 3 grid. With 2 colors and 9 points in the grid, there are 2^9, which equals 512, possible 3 × 3 grids. Thus, if we looked at a row of 513 3 × 3 grids, at least 2 of them would be colored in exactly the same way. And we are guaranteed that each of these grids will have 2 points of the same color at the top. We could have a monochrome right angle in a 3 × 3 grid, but we may not. The worst-case scenario is that among our 513 grids, 2 are the same, but neither have right angles. Wishful thinking says to build a right angle! We do not have control over the colors in the third 3 × 3 grid, but we can focus on the point on the lower left. This was a worst-case scenario. We are guaranteed that in any 1539 × 1539 grid, we must have a monochromatic right angle. Let $R(c)$ equal the size of the grid that guarantees a monochrome right angle if we use c colors. So $R(2) = 1539$. Define $S(c)$ to be the size of a c-colored grid that guarantees a monochrome square. We wish to prove that $S(2)$ is finite.

How do we create a monochrome square? Certainly the monochrome right angle is a start. What if we could make a monochrome right angle out of monochrome right angles? The worst-case scenario is that the lower right corner is not red. In this case, we can guarantee a structure where 3 of the 4 vertices are red but the fourth vertex is blue. Then we can again make a right angle with these structures. Then we are done; no matter what color the lower right-hand point is, we have created a monochrome square. However, in order to do this, we needed a right-angle construction with identical right angles. We can only get monochrome right-angles so far using 2 colors. But what if we could get right angles using any number of colors? That is what we need to complete our proof.

- We know that any 2-colored 1539 × 1539 grid is guaranteed to have a monochrome right angle.

- There are only $B = 2^{(1539\times1539)}$ different ways to 2-color this 1539 × 1539 grid. This is a number with 712,996 digits. It is unimaginably larger than any "real" number.

- There are about 10^{80} particles in the universe. If every one of those particles could count at a speed of a billion billion billion billion billion numbers per second, it would take the universe 10^{712870} seconds, or 10^{712863} years, just to count this number.

- Suppose it were possible to get a monochrome right angle if we color the lattice in B colors. In other words, there is some size $G = R(B)$ such that any $G \times G$ grid that is B-colored is guaranteed to have a monochrome right angle.

- Assign a different color to each of the B possible ways to 2-color a 1539 × 1539 grid.

Now view the entire 2-colored lattice, but in 1539 × 1539 chunks. You can think of this as a B-colored lattice. If $R(B)$ is the finite value G, then there is a grid of $G \times G$ chunks that is guaranteed to contain a monochrome right angle of chunks!

Thus, if we can show that $R(B)$ is finite, we are done. It was already pretty hard to compute $R(2)$, but we will confidently compute $R(3)$, and do it in a way that can clearly be generalized to higher numbers of colors. Our guiding principles are to stick to worst-case scenarios and build structures with focal points. We start by considering a row of 4 points. This guarantees 2 identical colors somewhere in this row. There are 3^{16} different ways to 3-color this 4 × 4 grid, so just string $3^{16} + 1$ of them together in a row, and we are guaranteed to see 2 identically colored grids. Let $M = 3^{16} + 1$.

Next, focus on a grid that would contain the third vertex of our right angle. Since it is a worst-case scenario, it is possible that this focal point is colored green. Thus, we have shown that in any $4M \times 4M$ grid of 3-colored points, we may not have a monochrome right angle, but we at least are guaranteed a structure like the one we built: an almost, hopeful monochrome right angle. How many different $4M \times 4M$ grids can there be? Just 3^{16M^2}. This is a very large number, which dwarfs the superbig number B. Let's add 1 to this number and call the sum G, for gigantic. If we place G of these $4M \times 4M$ grids in a row, we are guaranteed to get 2 that have the same color. But remember that every $4M \times 4M$ grid is guaranteed to have an almost monochrome right angle structure. Once again, we focus on the hopeful lower left corner. Whether it is green, blue, or red, we have a monochrome right angle. Clearly, it is possible to keep doing this, creating mind-boggling structures until we manage to guarantee a monochrome right angle for B colors. ■

Suggested Reading

Graham, Rothschild, and Spencer, *Ramsey Theory.*

Soifer, *The Mathematical Coloring Book.*

Questions to Consider

Recall that in Lecture 14 we used the pigeonhole principle to prove that if we color the plane in 2 colors, no matter how we color the plane we are guaranteed to have 2 points that are the same color and are exactly 1 meter apart. Here are 2 problems that prove stronger things with this same hypothesis. Use the worst-case scenario methods from this lecture.

1. Color the plane in 2 colors. Prove that 1 of these colors contains pairs of points at every mutual distance. In other words, 1 of the 2 colors, say, red, has the property that for each distance x, there are 2 red points exactly x units apart. (Hint: Use proof by contradiction.) What is the negation of the assertion?

2. Color the plane in 2 colors. Prove that there will always exist an equilateral triangle with all its vertices of the same color.

Genius and Conway's Infinite Checkers Problem

Lecture 23

No course on problem solving is complete without a look at Conway's checker problem. It's a fantastic example of creative, fearless analysis of a game, and it's a mainstay of mathematical circles and competitive problem-solving teams.

In our penultimate lecture, we sketch John Conway's brilliant solution to a classic puzzle. Our focus is not just on the mathematics, which is a wonderful mix of the ubiquitous golden ratio and monovariants, but we also engage in a discussion of mathematical culture, particularly the cult of genius that surrounds Conway and other mathematical "rock stars," including Paul Erdös and Évariste Galois.

The Checkers Problem

Place checkers at every lattice point of the half plane of nonpositive y coordinates. The only legal moves are horizontal and vertical jumps. By this, we mean that a checker can leap over a neighbor, ending 2 units up, down, right, or left of its original position, provided the destination point is unoccupied. After the jump is complete, the checker that was jumped over is removed from the board. Is it possible to make a finite number of legal moves and get a checker to reach the line $y = 5$?

- It is easy to get to $y = 2$, and with work, we can get to $y = 3$. It is reasonable to conjecture that we cannot get to $y = 5$.
- What methods do we have for proving impossibility?
    - Come up with a quantity that can be calculated for each configuration.

- This quantity should be a monovariant.
- If the quantity, say, always decreases but needs to increase in order to get to $y = 5$, we would be done.

- Conway's monovariant: Using the coordinate system with all checkers at $y = 0$ and below, let the target point be $C = (0, 5)$. We wish to prove that we can never reach this point.
- Define the number $z = (-1+\sqrt{5})/2$.
- For each point in the plane, compute its "taxicab distance" d to the target point C. For example, the point (2, 1) has distance $2 + 4 = 6$.
    - Then compute the value z^d.
    - For each configuration of checkers on the plane, add up the values z^d for each point that has a checker. This will be an infinite series. This is the Conway sum for that configuration.
- For example, the entire first row ($y = 0$) has the Conway sum $z^5 + 2(z^6 + z^7 + \cdots)$.
    - This simplifies to $z^5 + \dfrac{2z^6}{1-z}$.
    - Since $z^2 + z = 1$, we simplify this further to $z^5 + \dfrac{2z^6}{z^2} =$

$$z^5 + 2z^4 = z^3(z^2 + z) + z^4 = z^3 + z^4 = z^2(z + z^2) = z^2.$$

- Likewise, the Conway sum for the row $y = -1$ will be z^3, and so on, so the starting Conway sum of our problem is $z^2 + z^3 + z^4 + \cdots = \dfrac{z^2}{1-z} = \dfrac{z^2}{z^2} = 1$.

- Why is this a monovariant? Consider any configuration of checkers, and look at what happens to the Conway sum when a jump occurs.

- For example, suppose there are checkers at (4, 1) and (4, 2), and (4, 3) is unoccupied, so the first checker can jump over the second.

  - Before the jump, there is a checker at a distance of 8 and one at a distance of 7.

  - Afterward, the checker with a distance of 8 is now at distance 6, and the 7 checker is gone.

- So the Conway sum changes by the net amount of $z^6 - z^7 - z^8 = z^6(1 - z - z^2) = 0$. In other words, if a checker jumps toward the target point C, the Conway sum does not change!

- Consider a jump away from the target. Suppose a checker is at distance 10 and jumps over a checker at distance 11 to end up at distance 12. Then the net change is $z^{12} - z^{11} - z^{10} = z^{10}(z^2 - z - 1) = z^{10}(1 - z - z - 1) = z^{10}(-2z)$, which is negative.

- Thus if you jump away from C, the Conway sum decreases.

- There is one other case to check: that where your jump does not change the distance to C. For example, if you jump from $(-1, 2)$ to $(1, 2)$.
    - In this case, we start with 2 checkers, one at distance d (the jumper) and one at distance $d - 1$ (the jumpee). After the jump, the jumpee is gone, and the jumper is still at distance d.
    - So the net change is $z^d - (z^d + z^{(d-1)}) = -z^{(d-1)}$, which is again negative.
- Thus, the Conway sum is a true monovariant, never increasing from its initial value of 1.
- All that remains is to note that if we ever were to get a checker to C, the Conway sum would be larger than 1, since $z^0 = 1$ would be supplied by C, and there would still be infinitely many other checkers to add up.
- But our starting value is 1, and the Conway sum is a monovariant. So we can never reach C!

It takes a certain type of intellect to solve problems at this level. The key ingredient is a passion to investigate without any worry about consequence. Conway is one of a triumvirate of such heroes that also includes Paul Erdös and Évariste Galois. All 3 are iconoclastic rebels, belying the myth of the boring nerd, who supply romantic inspiration for the next generation. John Conway has led an unconventional life and made incredible contributions to math. He is like an eternal child in his ability to play, break rules, work on whatever pleases him, and continually ask questions, with a willingness to get his hands dirty. Paul Erdös

The key ingredient is a passion to investigate without any worry about consequence.

led a life of deliberate homelessness and celibacy. He wrote more papers and collaborated with more people than any mathematician in history. Évariste Galois died in a duel at age 20. His 60 pages of mathematics are considered by some to be the most important 60 pages ever written in mathematics. His greatest achievement, now called Galois theory, is a point of view flip.

All 3 of these people had passion, commitment, and a willingness to investigate relentlessly. This is something that we can all strive for, even if we cannot all possess genius. ■

Suggested Reading

Berlekamp, Conway, and Guy, "The Solitaire Army," in *Winning Ways for Your Mathematical Plays*.

Hardy, *A Mathematician's Apology.*

Hoffman, *The Man Who Loved Only Numbers*.

Honsberger, *Mathematical Gems II*, chap. 3.

Questions to Consider

1. The solution to the checkers problem was pretty subtle. Test your understanding: Why not just assign a large value—say, 100—to the point *C*? Then if a checker occupied *C*, the Conway sum would be at least equal to 100. But since the Conway sum starts at the value of 1 and never increases, it can never reach a value this large and hence never occupy *C*. What is wrong with this argument?

2. Here is a puzzle about Erdös numbers. (Assume, for simplicity, that when mathematicians write joint papers there are only 2 collaborators.) There are 5 mathematicians in a room. Each of them has written a paper with at least 1 of the others in the room. Exactly 1 of them has written papers with 3 of the others in the room, and exactly 1 has written papers with 2 others. One of the 5 mathematicians is Erdös himself. What are the possible Erdös numbers of these 5 people?

How versus Why—The Final Frontier

Lecture 24

Why complex numbers? Think of them as yet a new playground for you, like graph theory was, but this is an algebraic, geometric, physics playground, which has an incredible potential for connecting many branches of math when properly studied.

In this final lecture, we look back on what we have learned, talk about what we should study next, and reflect on what we do not know. We begin to ponder the ultimate purpose of an investigation: the quest for *why* something is true, not just how. I will share some of my favorite examples of this elusive intellectual quest.

First, some reminders about how to approach problems tactically, with the assumption that by now you have internalized key strategies.

- Proof by contradiction should be used when the thing you are trying to prove is easier to contemplate when negated.
- The extreme principle is useful when your problem has entities that become simpler at the boundary.
- The pigeonhole principle works well when the penultimate step can be formulated with 2 things belonging to the same category.
- Use induction when your problem involves recursion.
- The most important thing is to ask what type of problem you are trying to solve. This is half the battle.

What should you study next? Complex numbers! They are another playground with incredible potential for connecting many branches of math. Complex numbers are 4 things simultaneously: numbers, locations in the plane, vectors, and transformations! Complex numbers allow you to recast to and from physics, with dynamic interpretations of hitherto static objects. They also have connections to every branch of math, including number theory.

The bird's-eye view of problem solving means the following.

- We favor tactics over tools, and strategies over tactics.
- Yet we also favor investigation over rigor. In other words, we want to understand why things are true, rather than how.
- "How" arguments are rigorous and have clear details, but the global picture is murky.
- In contrast, a "why" argument is:
    - Not always rigorous.
    - Sometimes not even correct!
    - Globally clear, even if missing some details.
    - Magical yet inevitable.
    - Often a surprising yet natural point of view.

Here are some examples from our course and elsewhere.

- The proof that 8 times a triangular number equals a perfect square (Lecture 6) is a typical "why" argument.
- The bug problem (Lecture 8) can be solved analytically, with differential equations. This is a "how" argument, in contrast to our "why" argument.
- "Why" is at the heart of most mathematicians' Platonic beliefs.

- Problems fall all along the spectrum from completely opaque to completely understood in the "why" sense.

- The Shanille O'Keal problem from Lecture 19 was a good example of a "how."

- An example of something without even a "how," because all we had was a false conjecture, was the 5 circles problem of Lecture 4.

What do the "why" arguments have in common? What can we learn from them?

- Pictures ($8T + 1$ problem).

- Natural point of view and symmetry (bug problem).

- Physical intuition (arithmetic-geometric mean inequality).

- Using a physical object (Fermat's little theorem, Shanille O'Keal).

- Dynamic visualization of lines, evolving structures, and using important combinatorial facts (5 circles).

So what's next? Keep learning facts, but do not forget the need to build up flexibility, visualization, recasting, physical intuition, and the ability to see a natural point of view. Learn about complex numbers, which incorporate all of these ideas. Remember that problem solving is not just a textbook subject—it is a lifestyle, with a culture. The true underpinnings of this culture are passion and persistence. ■

Suggested Reading

Aigner and Ziegler, *Proofs from THE BOOK*.

Lansing, *Endurance.*

Needham, *Visual Complex Analysis*.

Zeitz, *The Art and Craft of Problem Solving*, sec. 4.2.

Solutions

Lecture 1

1. The problem is that there are 3 switches, but a light bulb only has 2 states: on and off. Use wishful thinking: What if you had a light bulb with 3 states? You do, since a bulb can be on, off and cold, or off and warm. So turn one switch on, leave one switch off, and turn the third one on briefly and then off. Go upstairs and check to see which state the bulb is in!

2. Once again, use wishful thinking: If you had a hook in the ceiling, you would be done, since you could tie the ends together at the bottom, climb up, hang onto the hook, cut both ropes at the top, slide one end through, climb down the doubled rope, and then pull it loose from the hook. Indeed, you can create a hook: Just cut off a small chunk at the top of one rope and make a loop right near the ceiling!

Lecture 2

1. It is not possible. The numbers 1, 3, 5, 7, and 9 are odd; the others are even. Adding or subtracting an even number does not change parity, but adding or subtracting an odd number does. We start with zero (nothing written yet) and then proceed to add or subtract 5 odd numbers and 5 even numbers. Parity changes 5 times, so the final result will always be odd. But zero is even!

2. The first professor writes a random number and shows it to the next professor, who then writes the sum of her salary and the first number on her piece of paper. This process of writing the running sum is continued until the last professor shows her sum to the first, who then subtracts her random number and adds her real salary. Now the sum (and hence the average) is known.

Lecture 3

1. Just use 3 dimensions. A $1 \times 1 \times 1$ box has a long diagonal of $\sqrt{3}$, which is greater than 1.5.

2. She could be just a few moments older than 30, if she was born at 11:59 pm on December 31, 1959, and she uttered her words early on January 1, 1990.

Lecture 4

1. Guided by $1/3 + 1/6 = 1/2$, we get $x = n(n + 1)$ and $y = n + 1$.

2. Indeed, $A = B/2 + I - 1$. This is called Pick's theorem.

Lecture 5

1. If there were a smallest number, say m, then $m/2$ would be even smaller. Contradiction!

2. Write $n(n!)$ as $(n + 1)! - n!$; then the sum telescopes to $101! - 1$.

Lecture 6

1. Draw the odds as successive Ls of dots. For example, 5 is an L that is 3 dots high and 3 dots wide (with a common dot at the corner).

2. The second trip takes 4 fewer seconds, but 12 more steps are stepped on. So the escalator is traveling 3 steps per second. Therefore, on the first trip, her net speed is 4 steps per second, and in 20 seconds she goes 80 steps, the "true" distance.

Lecture 7

1. Assume, to the contrary, that all 3 are odd. Then the sum of the squares of 2 of them would be the sum of 2 odds, which is even. Yet the square of the third one is odd. Contradiction!

2. No. Assume, to the contrary, that there is such a line. Move your diagram so that this line is horizontal. Then, without loss of generality, A lies above the line. Thus B is below, C is above, D is below, E is above, and we have a contradiction, since the segment EA should be intersected by this horizontal line!

Lecture 8

1. Let $A = (4, 3)$, and let B and C be the points on the x-axis and the line $y = x$, respectively. If we reflect the picture across the x-axis, we see that line segment AB will have the same length as the line segment from $(4, -3)$ to B. If we reflect now across the line $y = x$, we see that AC has the same length as the line segment from C to $(3, 4)$. So the perimeter of the triangle is equal to the sum of the lengths of the line segments from $(4, -3)$ to B, from B to C, and from C to $(3, 4)$. This is clearly minimized by the straight segment joining $(4, -3)$ to $(3, 4)$. By the Pythagorean theorem, this minimal length is $\sqrt{1^2 + 7^2}$, which is $\sqrt{50}$.

2. It is certainly possible to do this by hand, but if we observe that $1 + 2 + 3 + 4 + 5 + 6 = 21$ and notice that 10 is about half of that, we can use symmetry. Suppose A is subset whose elements are greater than 10. Then the complement of A (i.e., the leftover elements) has to have a sum less than or equal to 10 (or else the sum of all elements would be greater than 21). So there is a metaphorical symmetry between the subsets with sum greater than 10 and those with sum less than or equal to 10. So exactly half of the subsets, or 32, have each property.

Lecture 9

1. It is exactly equivalent to ordinary takeaway, but with one fewer starting number of pennies.

2. Consider the magic square below.

2	7	6
9	5	1
4	3	8

Each row, column, and diagonal adds up to 15. Hence our game is equivalent to tic-tac-toe, which we all know will be a draw if players play optimally. But you can win this game if you have the magic square picture in mind and your opponent does not.

Lecture 10

1. Suppose that they are not all equal. Then there is a value m that is the least value on the board, and at least somewhere on the board, since the values are not all equal, this value m must be adjacent to a value M that is greater than m. But then m is the average of 4 values, including M, which forces one of the values to be strictly less than m, a contradiction.

2. Just consider the smallest of the coins. It is easy to check (draw a picture) that no more than 5 coins can be tangent to it (remember, they all are larger).

Lecture 11

Not applicable.

Lecture 12

1. Not applicable.

2. Let's show that it is true for a quadrilateral *ABCD*. Divide it into triangles *ABC* and *ACD*. The area of the quadrilateral will be the sum of the areas of the triangles. For each triangle, the area is equal to half the boundary lattice points, plus the interior lattice points, minus 1. When we apply this to both triangles and add, the boundary count includes the lattice points on segment *AC* twice. Suppose *AC* contains exactly x lattice points (including *A* and *C*). Then half the boundary count for the 2 triangles will be equal to half the boundary count for the quadrilateral, plus $x - 1$. (The minus 1 is because *A* and *C* are counted twice.) However, the interior count of the 2 triangles does not include any of the $x - 2$ lattice points that are on *AC* (not including *A* and *C*). So the interior counts of the 2 triangles added will be equal to the interior count of the quadrilateral, minus $x - 2$. The final bit of accounting is to note that the triangle count includes two −1s, and for the rectangle, just one −1, so it balances out. This can be extended to larger polygons.

Lecture 13

1. By Fermat's little theorem, $3^{18} = 1$ (mod 19). Since $18 \times 111 = 1998$, by raising the previous equation to the 111th power, we get $3^{1998} = 1$ (mod 19). Notice that $3^{2009} = 3^{1998} \times 3^{11}$. Hence our answer will be whatever 3^{11} equals modulo 19. We compute $3^2 = 9$, $3^3 = 27 = 8$, $3^4 = 24 = 5$, and $3^5 = 15 = -4$. Squaring this last one, we get $3^{10} = 16 = -3$, and then finally $3^{11} = -3 \times 3 = -9 = 10$ (all mod 19).

2. If it was a line instead of a necklace, there would be 6 choices for the first bead, 5 for the second, and so on, for $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ choices. But since it is a necklace, we can take any arrangement such as *bcadfe* and cyclically permute it, say, to *cadfeb*. There are 6 members in each sorority, so we divide by 6, getting 120 as our answer.

Lecture 14

1. Partition the unit square into four $1/2 \times 1/2$ squares. By pigeonhole, one of these smaller squares must contain at least 2 points. Since the diagonal of each small square is $\sqrt{2}/2$, that is the maximum distance between the 2 points.

2. There are 150,001 categories (include bald people!). We need to find the minimum population p such that the ceiling of p divided by 150,001 is equal to 10. Thus $p = 9 \times 150{,}001 + 1 = 1{,}350{,}010$.

Lecture 15

1. Getting your hands dirty should yield the formula $v - e + r = 1$. This is called Euler's formula.

2. You cannot. If you could, then the quotients $(b - c)/(a - b)$, $(c - a)/(b - c)$, and $(a - b)/(c - a)$ would all be integers. However, the product of these 3 quotients is invariant; it is 1. The only way that 3 integers can multiply to 1 is if all of them are equal to 1. But that would mean that $2b = a + c$, $2c = a + b$, and $2a = b + c$. In other words, a, b, and c are each the average of the other 2. Using the extreme principle, let, say, a be the smallest of the 3 numbers (remember, the numbers are distinct). Then we have a contradiction. How can a be the average of 2 numbers that are both larger than a? The other possibility is if one of the quotients is 1 and the other 2 are -1. But if a quotient equaled -1, for example, the first quotient, we would get $b - c = b - a$, which makes $c = a$, another contradiction.

Lecture 16

1. Reflect across the stream. Then when you build your optimal rectangle with perimeter S, the mirror rectangle also has perimeter S, and collectively, you are building a rectangle with 4 sides that is to have maximal area, with a fixed perimeter of $2S$. You know that this optimal shape is a square. So the answer to the original question is a half square: a rectangle whose sides are in the ratio 1:2, with the longer side on the stream.

2. Just multiply the 3 inequalities $(a+b) \geq 2\sqrt{ab}$, $(b+c) \geq 2\sqrt{bc}$, and $(c+a) \geq 2\sqrt{ca}$, and you are done.

Lecture 17

1. Imagine, on the second day, a clone of the first monk who starts at the bottom at 8 am and exactly imitates what the first monk did the day before. Clearly the 2 monks will meet on the trail, and this time and place are the solution.

2. Draw a lattice of squares. Start at (0, 0), and begin drawing a line with slope 7/11. As it goes northeast, it will first hit another lattice point at (11, 7). To count bounces, just count the number of times this line hits a horizontal or vertical lattice line. It will hit $x = 1, 2, 3, \ldots, 10$ and $y = 1, 2, 3, \ldots, 6$, for a total of 16 bounces.

Lecture 18

1. Just rotate the diagram 60° clockwise about the center K. Then J moves to M, and N moves to I. Hence $JN = IM$. Likewise, a rotation about I will show that $LK = JN$.

2. Start with the point (0, 0). The first rotation leaves it fixed; the second one, about (1, 0), brings it to (1, −1); the third brings it to (3, −1); and the final one brings it to (4, 0). So the translation (of any starting point) will be "move 4 units to the east." You might conjecture that if you have n rotations, about (0, 0), (1, 0), … , $(n - 1, 0)$, each by $\dfrac{360}{n}^{\circ}$ counterclockwise, the net result would be a translation by n units to the right, and this is correct. The easiest way to see it is to imagine a regular n-gon with side length 1 unit whose "top" side is the line segment joining (0, −1) and (0, 0). Then the rotations are equivalent to rolling this n-gon along the x-axis, ending up n units to the right. Try this for $n = 3$, $n = 4$, or $n = 5$ to see for sure.

Lecture 19

1. It is easy to conjecture that the sum is $F_{n+2} - 1$. The base case is clearly true ($1 = F_3 - 1$). For the inductive step, assume that it is true that the sum of the first k Fibonacci numbers is equal to the $(k + 2)^{\text{th}}$ Fibonacci minus 1. Then we want to find the sum of the first $(k + 1)$ Fibonaccis. This sum is $F_1 + F_2 + \cdots + F_k + F_{k+1}$, which by the inductive hypothesis can be written as $F_{k+2} - 1 + F_{k+1}$. By the definition of Fibonacci numbers, this is equal to $F_{k+3} - 1$, which is what we wanted.

2. Once again, the base case is obvious, since the first Fibonacci number is less than 2. Thereafter, it is simple as well: Since each successive Fibonacci is equal to the sum of the 2 preceding it, and since the sequence is an increasing one, each Fibonacci number is strictly less than twice the one before it.

Lecture 20

1. Just plug in $x = 10$ into the expression $(1 + x)^k$. For example, if $k = 4$, we get $11^4 = 10^4 + 4 \cdot 10^3 + 6 \cdot 10^2 + 4 \cdot 10^1 + 1$. Since we use the base-10 system, that is the number 14,641. The reason it fails for greater values of k is because some of the coefficients of the binomial are greater than 9. For example, if $k = 5$ we get $11^5 = 10^5 + 5 \cdot 10^4 + 10 \cdot 10^3 + 10 \cdot 10^2 + 5 \cdot 10^1 + 1$. We cannot just read this off as a base-10 number as we did the last time, because this number has a thousands place digit of 10 and a hundreds place digit of 10. In base 10, digits cannot be greater than 9.

2. The analysis is similar to what we did in the lecture, but this time, recursive structures are successive powers of 3, and each new triangle consists of 6 triangles of the previous kind, with an inverted 0 in the middle. You can search the Web for "Pascal's triangle modulo 3" to find good illustrations and interactive applets; one example is at http://faculty.salisbury.edu/~kmshannon/pascal/article/twist.htm. The reason for the number 6 is that it is equal to 1 + 2 + 3. The limit of the number of nonzeros (mod 3) goes to 0 as the number of rows increases, so just as before, almost all elements of Pascal's triangle are multiples of 3!

Lecture 21

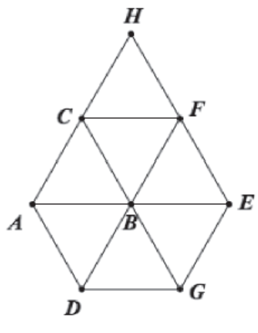
1. Using the multiplication rule for generating functions, we get $\binom{17}{13} = \binom{10}{6}\binom{7}{7} + \binom{10}{7}\binom{7}{6} + \binom{10}{8}\binom{7}{5} + \binom{10}{9}\binom{7}{4} + \binom{10}{10}\binom{7}{3}$. This has a good combinatorial interpretation: The left-hand side is the number of ways to pick 13 children from a pool of 17. The right-hand side counts the same thing but supposes that the children consist of 10 girls and 7 boys, and it breaks the outcomes into the 5 cases of 6 girls and 7 boys, 7 girls and 6 boys, 8 girls and 5 boys, 9 girls and 4 boys, and 10 girls and 3 boys.

2. The series is $1 + x + 2x^2 + 4x^3 + 8x^4 + 16x^5 + 32x^6 + 63x^7 + 125x^8 + \cdots$, and the coefficient of x^n is equal to the number of ways that any number of dice can add up to n. For example, there are 2 ways to get 2: two 1s or one 2. And there are 16 ways to get 5: 5; 1, 4; 4, 1; 1, 1, 3; 1, 3, 1; 3, 1, 1; 3, 2; 2, 3; 1, 2, 2; 2, 1, 2; 2, 2, 1; 1, 1, 1, 2; 1, 1, 2, 1; 1, 2, 1, 1; 2, 1, 1, 1; or 1, 1, 1, 1, 1. The reason this works is because for each k, $[D(x)]^k$ is the generating function for dice sums when k dice are rolled. You may enjoy looking up this sequence in The Online Encyclopedia of Integer Sequences; they are called the "hexanacci" numbers. Can you see why? Because they satisfy the recurrence formula that each number in the sequence is equal to the sum of the 6 previous numbers!

Lecture 22

1. Use proof by contradiction. Assume, to the contrary, that for neither of the 2 colors is it true that one can find 2 points of the same color at any mutual distance. In other words, there are distances x and y such that there are no 2 red points x units apart and no 2 blue points y units apart. Without loss of generality, suppose that x is greater than or equal to y (i.e., use the extreme principle on 2 numbers). Now, consider a red point. There must be at least 1 red point, for otherwise the color blue would not have a forbidden distance y. Draw a circle with radius x and center at this red point. Every point on this circle must be blue, since no 2 red points are x units apart. But now we have achieved a contradiction: Since y is less than or equal to x, it is certainly possible to find 2 (blue) points on this circle that are y units apart.

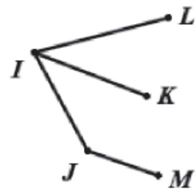
2. Start by finding 2 points that are the same color, say, red, at locations A and B. Now consider the configuration below made of equilateral triangles. Besides the small equilateral triangles, note that there are larger equilateral triangles, such as CDE and AEH. The existence of these alternatives is key for a miniature Gallai-style argument: Either C or D are red; then we would be done. Otherwise, both C and D are blue. Then if E is blue, we are done again (triangle CDE). But if E is red, then we will be done if either F or G is red. If not, they are both blue. But now we have a focal point at H. If it is red, we have a large red equilateral triangle (AEH). If it is blue, we have the small equilateral triangle HCF. No matter what, a monochrome equilateral triangle is guaranteed.



Lecture 23

1. If we assign a value greater than 1 to the point *C*, the Conway sum is no longer a monovariant, if you look at moves involving *C* itself. For example, if a checker 2 units to the left of *C* jumped over a checker 1 unit to the left of *C* and then occupied *C*, then the Conway sum would actually increase!

2. Model it with a graph: Use vertices for mathematicians and edges to indicate joint papers. There is only one graph that satisfies the conditions.



The Erdös numbers depend on which one is Erdös. If Erdös is person *I*, then *L*, *K*, and *J* have Erdös number 1, and *M* has Erdös number 2. But if we put Erdös anywhere else, it is possible for someone to have an Erdös number of 3.

Lecture 24

Not applicable.

Timeline

1787 .... Ten-year-old Carl Gauss uses Gaussian pairing.

1796 .... Teenaged Carl Gauss solves the problem of how to construct the regular heptadecagon (17-gon), something that had eluded mathematicians since Euclid's time.

1832 .... French algebraist Évariste Galois dies in Paris from a duel at the age of 20.

1894 .... The Eotvos contest, the first Olympiad-style math competition, begins in Hungary.

1907 .... Willem Abraham Wythoff writes about the game later called Wythoff's Nim, actually an older game that probably originated in China.

1930 .... Frank Ramsey's paper "On a Problem in Formal Logic" is published; Tibor Gallai's theorem is proven around the same time.

1938 .... Annual Putnam competitions for college undergraduates begin.

1939 .... Physicist Richard Feynman is named a Putnam Fellow.

1952 .... High school mathematics contests sponsored by the Mathematical Association of America begin across the U.S.

1959.. The first International Mathematical Olympiad is held in Romania, with teams from 7 countries.

1961.. John Conway analyzes the checker problem.

1970.. John Conway invents the cellular automaton Game of Life, popularized by Martin Gardner in *Scientific American*.

1972.. The first USA Mathematical Olympiad is held.

1974.. The U.S. participates for the first time in the International Mathematical Olympiad (in East Germany), placing second after the USSR.

1977.. Wythoff's Nim is popularized by Martin Gardner in *Scientific American*.

1981.. The International Mathematical Olympiad is held in the U.S. for the first time, with 27 countries participating.

1985.. The Colorado Mathematical Olympiad begins, founded by Soviet émigré Alexander Soifer. This was the first regional Olympiad-style contest in the U.S.

1994.. The U.S. team comes in first place at the 35th International Mathematical Olympiad (held in Hong Kong), with all 6 team members receiving perfect scores. This was the first and only time a team has received a perfect score at that competition.

1996.. Hungarian-born Paul Erdös, the most prolific mathematician and problem solver of modern times, dies.

2001.. The U.S. hosts the International Mathematical Olympiad for the second time, with 83 countries attending.

Glossary

algorithmic proof: Proof where we imagine a sequence of steps that is guaranteed to solve our problem.

binomial coefficients: The numbers $\binom{n}{k}$, which are equal to (1) the coefficient of x^k in $(1+x)^n$, (2) the number of ways of choosing k things from a set of n things, and (3) the number $n(n-1)\cdots(n-k+1)/k!$.

bipartite graph: A graph whose vertices can be colored red and blue in such a way that no edge connects vertices of the same color.

congruence: Two integers are said to be congruent (modulo m) if their difference is multiple of m.

crux move: The crucial step in a problem-solving investigation that solves the problem. This step can be technical or can be a strategic breakthrough.

degree: A graph theory term; the degree of a vertex is the number of edges emanating from it.

extreme principle: The problem-solving tactic that says, "contemplate the extremal values of your problem."

Fibonacci numbers: The sequence 1, 1, 2, 3, 5, … in which each term is equal to the sum of the 2 previous. Named after Leonardo Fibonacci (c. 1170–c.1250), it is one of the most accessible playgrounds of recreational mathematics.

fundamental theorem of arithmetic: All integers can be factored into primes, and this factorization is unique (up to order).

generating function: Given a sequence $a_0, a_1, a_2, \ldots$, its generating function is the polynomial $a_0 + a_1x + a_2x^2 + \cdots$. The generating function encodes information about the entire sequence; algebraic manipulations of generating functions can thus shed light on questions about the sequence.

Goldbach conjecture: A famous unsolved problem asserting that all even numbers greater than 2 can be written as a sum of 2 primes.

golden ratio: The number $(1+\sqrt{5})/2$, which crops up in many places in mathematics, including the Fibonacci numbers.

graph theory: The branch of math that studies abstract networks, also known as graphs, which are entities of vertices joined by edges. It is easy to learn and therefore a very accessible laboratory for exploring a number of problem-solving themes.

handshake lemma: An important graph theory result stating that the sum of the degree of each of the vertices of a graph is equal to twice the number of edges in the graph.

harmonic series: The sum of the reciprocals of the positive integers, which is a divergent series (meaning the sum is infinite).

induction: Technically called mathematical induction, this is a powerful method of proving recursive statements. An inductive proof always has 2 parts: A base case proving the first stage is followed by the inductive step, where it is shown that each intermediate stage logically implies the next.

integers: The positive and negative whole numbers, including zero.

International Mathematical Olympiad (IMO): An elite problem-solving contest, initiated by Eastern Bloc countries in 1959, that now includes nearly 100 nations every year.

invariant: A very high-level tactic for looking at many problems where a quantity or quality stays unchanged. A monovariant is a quantity that changes in only one direction; monovariants are very useful for studying evolving systems and proving that they terminate or that certain states are impossible.

modulus: A mathematical entity that in a congruence divides the difference of 2 congruent members without leaving a remainder. *See also* **congruence**.

monovariant: *See* **invariant**.

number theory: The branch of math dealing with integers. Because integers are familiar to us beginning in grade school, number theory, like graph theory, is an excellent venue for mathematical investigation for beginners.

Olympiad: A style of math contest that features relatively few questions (usually fewer than 10) of the essay-proof type.

parity: The property of oddness or evenness for an integer. Parity is a powerful tactic that reduces a problem from a large, or infinite, number of states to only 2 states.

Pascal's triangle: Perhaps the greatest of all the elementary mathematical laboratories for investigation. This is a triangle of numbers where row n consists of the binomial coefficients $\binom{n}{k}, k = 0, 1, 2, \ldots, n$.

pigeonhole principle: A fundamental problem-solving tactic stating that if you have more things (pigeons) than categories (pigeonholes), at least 2 of the things belong to the same category.

prime number: A positive integer that has no positive integer divisors other than 1 and itself. The first few primes are 2, 3, 5, 7, 11, and 13.

problem: A mathematical question that one does not know, at least initially, how to approach and that therefore requires investigation, often using organized strategies and tactics. We contrast this with an exercise: a question that may be difficult but is immediately approachable with little or no investigation.

proof by contradiction: A method of proof that starts by assuming that the conclusion is false and then proceeds to a logical contradiction, concluding that the conclusion's falsehood was untenable.

Putnam exam: An Olympiad-style contest for American and Canadian undergraduates, notorious for having a median score, most years, of 0 or 1 out of 120 possible points.

Pythagorean theorem: The sum of the squares of the legs of a right triangle is equal to the square of its hypotenuse. There are hundreds of known proofs of this theorem, the earliest thousands of years old.

Ramsey theory: A branch of mathematics named in honor of Frank Ramsey (1903–1930), whose seminal 1930 paper "On a Problem of Formal Logic" began the subject with a statement and proof of what is now called Ramsey's theorem. The theorem essentially asks the question, how large must a structure be in order that it is guaranteed to contain a specified substructure? The pigeonhole principle is the trivial case of Ramsey's theorem, and Gallai's theorem about squares is an example of a Ramsey-like theorem.

recasting: The problem-solving strategy of radically changing the venue of a problem, for example, from number theory into geometry, or vice versa. Certain mathematical ideas, such as generating functions, are useful precisely because of their recasting potential.

recursive definition: A sequence or evolving structure where the later terms (or more complex structures) depend on the previous, simpler ones. The Fibonacci numbers are a simple example; the chessboard tromino problem of Lecture 19 is another.

strategies: Mostly commonsense organizational ideas that help overcome creative blocks to begin and facilitate a problem-solving investigation. Strategies in this course include wishful thinking, make it easier, get hands dirty, chainsaw the giraffe, draw a picture, change your point of view, and recast your problem.

symmetry: An object (not necessarily geometric) is symmetrical if a transformation leaves it invariant. A natural point of view is often a point of symmetry. Symmetry increases order in a problem, so you should seek, and even impose, symmetry where possible.

tactics: Narrower than strategies, tactics are broad ideas within mathematics generally used at a later stage of investigation, often providing the key to a solution. Examples in this course include symmetry, parity, the extreme principle (contemplate extreme values), the pigeonhole principle, and squarer is better.

tools: Mathematical ideas of very narrow application that are nonetheless capable of very impressive results when used correctly. Examples in this course include Gaussian pairing, telescoping, and massage. Some useful tools (e.g., creative algebraic substitution) are better understood as a narrow instance of a broader strategy (e.g., wishful thinking).

tromino: A shape made out of 3 contiguous square units. There are only 2 types of tromino, the L and the I (a straight line of 3 squares). Trominos and more complex shapes (e.g., the 12 different pentominos, made of 5 squares) are popular objects of study in recreational mathematics.

United States of America Mathematical Olympiad (USAMO): The first national Olympiad of the United States, which began in 1972 with funding from numerous mathematical societies and the Department of Defense. It is the culminating exam that begins with multiple-choice tests taken by several hundred thousand high school and middle school students. Since 1974, the top scorers on this exam have competed for places on the 6-person team to the International Mathematical Olympiad.

Wythoff's Nim: One of the many names of a simple combinatorial game whose solution involves Fibonacci numbers and the golden ratio. Nim is an ancient game in which 2 players typically take turns removing objects from piles until none are left.

Biographical Notes

Conway, John Horton (b. 1937): A British-born mathematician and professor at Princeton University who is famous for fundamental contributions to many branches of mathematics, including recreational math.

Erdös, Paul (1913–1996): The most prolific mathematician of modern times (perhaps ever), Erdös was also known for collaborating with more mathematicians than any other. He was famous for his deliberately homeless and celibate life, devoted entirely to mathematics.

Euler, Leonhard (1707–1783): A Swiss mathematician, about as prolific as Erdös for his time, who was the father of graph theory. Euler was known for unconventional, "rule-breaking" approaches to hard mathematical problems.

Fermat, Pierre de (1601–1665): A French mathematician who was one of the first investigators of what is now modern number theory.

Gallai, Tibor (1912–1992): A Hungarian mathematician who was an important collaborator and lifelong friend of Erdös.

Galois, Évariste (1811–1832): This French mathematician, famous for his short but productive life, died in a duel. He made seminal and highly original contributions to the algebra of polynomials, among other things.

Gardner, Martin (b. 1914): Editor of the Mathematical Games column of *Scientific American* from 1956 to 1981, Gardner is unquestionably the greatest modern expositor of mathematics writing in English.

Gauss, Carl Friedrich (1777–1855): A German mathematician who is universally recognized as one of the greatest 2 or 3 mathematicians in history. He made fundamental advances in all branches of mathematics, usually generations ahead of his time.

Klein, Felix (1849–1925): A German mathematician and influential expositor. He proposed the important point of view change that geometry is best understood by looking at transformations rather than objects.

Bibliography

The list of books and resources below is pretty large, even though it just scratches the surface of the literature out there. If you are just getting started and really want to become a better problem solver, you must get practice by working on problems, and it is best for the problems to be fairly gentle and nontechnical.

The best place to start is by perusing *any* book by Martin Gardner. The CD collection of his books is an economical way to go (and its searchable). If you feel overwhelmed by the sheer quantity of Gardner's work, there are 4 great single-volume choices. George Polya's classic *How to Solve It* is short and very useful. My book, *The Art and Craft of Problem Solving*, has a larger collection of problems and a more explicit treatment of strategy and tactic. The *Mathematical Circles (Russian Experience)* book lead-authored by Dmitri Fomen has a wealth of "easy" problems (i.e., intended for Russian middle school–aged kids) along with good pedagogical ideas. And perhaps the most enjoyable single-volume book to look at is Ravi Vakil's *Mathematical Mosaic*. It is not comprehensive, but the mathematical topics are chosen with great taste. Its style may seem a little juvenile—it was written for a young audience—but the mathematics is actually quite deep.

Aigner, M., and G. Ziegler. *Proofs from THE BOOK*. Berlin: Springer, 2000. A collection of proofs that satisfy Paul Erdös's criteria of elegance, simplicity, and beauty.

Andreescu, Titu, and Svetoslav Savchev. *Mathematical Miniatures.* Washington, DC: Mathematical Association of America, 2003. The dichotomy between exercises and problems was first made clear to me by Titu Andreescu, who was the head coach of the USA team at the International Mathematical Olympiad for many years. Titu has written and coauthored numerous books about problem solving at the highest levels; this is my personal favorite.

Beck, M., and S. Robbins. *Computing the Continuous Discretely*. New York: Springer, 2007. This is an elegant (albeit advanced) book that explores the relationship between combinatorics and geometry, mostly via counting lattice points.

Bell, E. T. *Men of Mathematics*. New York: Touchstone, 1986. First published in the 1930s, this is a classic history of mathematicians. It may not be the most accurate, but it is certainly responsible for many mathematicians' worship of figures such as Carl Gauss and Évariste Galois.

Berlekamp, E. R., J. H. Conway, and R. K. Guy. *Winning Ways for Your Mathematical Plays*. Vols. 1–2. London: Academic Press, 1982. A classic and groundbreaking exposition of the theory of mathematical games.

Conway, John H., Heidi Burgiel, and Chaim Goodman-Strauss. *The Symmetry of Things*. Wellesley, MA: A. K. Peters, 2008. The latest of Conway's many books, this one is bound to be a classic.

Edwards, A. W. F. *Pascal's Arithmetical Triangle: The Story of a Mathematical Idea*. Baltimore, MD: Johns Hopkins, 2002. A carefully researched history of Pascal's triangle, perhaps the most accessible mathematical playground for problem solvers.

Engel, Arthur. *Problem-Solving Strategies*. New York: Springer, 1998. An indispensable guide to strategic problem solving at the advanced level.

Fomin, Dmitri Sergey Genkin, and Ilia Itenberg. *Mathematical Circles (Russian Experience)*. Translated by Mark Saul. Providence, RI: American Mathematical Society, 1996. This is an inspiring and eye-opening guide to what Russian 12-year-olds learn in a math circle.

Gardiner, Anthony. *Discovering Mathematics: The Art of Investigation*. New York: Oxford University Press, 1986. A beautifully written elementary guide for beginners.

Gardner, Martin. *Aha! A Two Volume Collection.* Washington, DC: Mathematical Association of America, 2006. Originally published separately and now reissued as a single volume, this is a collection of some of the short puzzles of Martin Gardner. The theme of the aha puzzles is the unexpected, creative solution.

———. *Martin Gardner's Mathematical Games.* Washington, DC: Mathematical Association of America, 2005. CD-ROM. Gardner, who edited the Mathematical Games column of *Scientific American* magazine from 1956 to 1981, is in my opinion the greatest English-language mathematical expositor of modern times. He has published dozens of books based on his columns (all now collected on this single CD); any of these books contains a wealth of recreational problems along with many fun and deep essays, always written for the intelligent layperson.

———. *Martin Gardner's Sixth Book of Mathematical Diversions from Scientific American.* Chicago: University of Chicago Press, 1984.

———. *Penrose Tiles to Trapdoor Ciphers.* Rev. ed. Washington, DC: Mathematical Association of America, 1997. One of my favorites of Gardner's many collections, this book contains, among other things, a very nice discussion of the puppies and kittens game (Wythoff's Nim).

Goodaire, Edgar, and Michael Parmenter. *Discrete Mathematics with Graph Theory*. Upper Saddle River, NJ: Prentice-Hall, 2005. A very readable discrete math text; an excellent book for beginners to learn about combinatorics and graph theory.

Graham, Ronald, Donald Knuth, and Oren Patashnik. *Concrete Mathematics*. 2nd ed. Reading, MA: Addison-Wesley, 1994. This is an encyclopedic guide to many discrete math topics, from a real problem-solver's perspective. Full solutions to a wide variety of exercises and problems are at the back of the book. It is one of the essential books in my library.

Graham, Ronald, Bruce Rothschild, and Joel Spencer. *Ramsey Theory*. 2nd ed. Hoboken, NJ: Wiley, 1990. A fascinating (albeit quite advanced) discussion of many Ramsey theory topics, by one of Erdös's favorite collaborators, Ronald Graham.

Guy, Richard K. "The Strong Law of Small Numbers." *American Mathematical Monthly* 95 (1988): 697–712. This classic essay, published in *American Mathematical Monthly* but accessible to laypeople, is a play on the law of large numbers, a statement about how long-term empirical frequencies approach theoretical probabilities. In this essay, the focus is on surprising sequences that appear to do one thing but in fact do something entirely unexpected.

Hardy, G. H. *A Mathematician's Apology.* Cambridge: Cambridge University Press, 1992. First published in 1940, this is a poetic memoir about the beauty of mathematical thinking.

Hoffman, Paul. *The Man Who Loved Only Numbers*. New York: Hyperion, 1999. A beautifully written book about an unbelievable character, Paul Erdös, the most prolific researcher in the history of mathematics.

Honsberger, Ross. *Ingenuity in Mathematics*. Washington, DC: Mathematical Association of America, 1970. This and the next book are just 2 of the many wonderful works by Honsberger, whose specialty is clear essays that explain amazingly creative mathematics.

———. *Mathematical Gems II.* Washington, DC: Mathematical Association of America, 1976.

Kazarinoff, Nicholas. *Geometric Inequalities*. Washington, DC: Mathematical Association of America, 1975. A short and brilliant book that gives beginners a real intuition about inequalities by using a geometric, visual approach that minimizes algebra in favor of true insight.

Kendig, Keith. *Sink or Float? Thought Problems in Math and Physics*. Washington, DC: Mathematical Association of America, 2008. An elementary book to help develop your physical intuition in a mathematical context.

Lansing, Alfred. *Endurance.* New York: Carroll and Graf, 1999. First published in 1959, this is a riveting account of an epic tale of survival: the ill-fated Antarctic voyage of Ernest Shackleton. It is relevant to us because of the important story of mental toughness.

Lehoczky, Sandor, and Richard Rusczyk. *The Art of Problem Solving*. 7th ed. Vols 1–2. Alpine, CA: Art of Problem Solving, 2008. An excellent guide for beginners (as young as middle school), with complete solutions to problems from a very wide variety of topics.

Liu, Andy, ed. and trans. *Hungarian Problem Book III.* Washington, DC: Mathematical Association of America, 2001. There are several volumes in English of the famous Hungarian Problems, the earliest modern Olympiad-style contest. All 3 volumes of this series have excellent commentary about problem solving in general, but this is the best of them.

Maurer, Stephen B., and Anthony Ralston. *Discrete Algorithmic Mathematics*. 3rd ed. Wellesley, MA: A. K. Peters, 2004. A superb text, especially notable for its careful treatment of mathematical induction.

Needham, Tristan. *Visual Complex Analysis.* New York: Oxford University Press, 1999. An essential book for anyone who truly wants to understand *why* things are true. Needham's uncompromisingly visual approach is unique and powerful.

Nelson, Roger. *Proofs without Words*. Washington, DC: Mathematical Association of America, 1997. One of the first books to stress the importance of avoiding algebra whenever possible. Highly recommended.

Niven, Ivan. *Maxima and Minima without Calculus.* Washington, DC: Mathematical Association of America, 2005. Like the 2 books above, this is a highly recommended antidote to using too much higher-powered mathematics when simpler methods are better and ultimately more illuminating.

Olson, Steve. *Count Down: Six Kids Vie for Glory at the World's Toughest Math Competition.* Boston: Houghton Mifflin Harcourt, 2004. A fascinating account of the 2001 International Mathematical Olympiad, which was held in the United States.

Polya, G. *How to Solve It*. 2nd ed. Princeton, NJ: Princeton University Press, 1957. This is the classic book about problem solving. All later books, mine included, owe a tremendous debt to Polya's pioneering work and writing.

Soifer, Alexander. *The Mathematical Coloring Book.* New York: Spinger, 2009. This book is a mathematical history of Ramsey-style problems involving coloring, including Gallai's theorem. It also includes much fascinating history.

———. *Mathematics as Problem Solving.* New York: Springer, 2009. A very short but intense guide to problem solving for beginners, with an excellent choice of problems.

Solow, Daniel. *How to Read and Do Proofs.* 5th ed. Hoboken, NJ: Wiley, 2009. One of the standard college texts on this difficult topic.

Sylvester, J. J. "Question 7382." *Mathematical Questions from the Educational Times* 37 (1884): 26. This is the first recorded mention of what we refer to as the chicken nuggets problem, written by an eminent British mathematician.

Tabachnikov, Sergei, ed. *Kvant Selecta: Algebra and Analysis*, *I*. Providence, RI: American Mathematical Society, 1999. This book is part of a series of translations from the great Russian journal *Kvant* (Quantum), which popularized deep mathematics for a young audience.

Tanton, James. *Solve This: Math Activities for Students and Clubs.* Washington, DC: Mathematical Association of America, 2001. This is a one-of-a-kind collection of deep puzzles that have a hands-on quality. Perfect for truly doing mathematics and developing physical intuition.

Vakil, Ravi. *A Mathematical Mosaic.* Ontario: Brendan Kelly, 1996. One of my favorite books for beginners, this is an absolutely joyous excursion through many topics. Vakil is one of the best contest problem solvers in history, but only modesty and generosity come across in this remarkable little book.

Vanden Eynden, Charles. *Elementary Number Theory.* Long Grove, IL: Waveland Press, 2001. One of the clearest and simplest guides to this essential topic.

Velleman, Daniel. *How to Prove It.* New York: Cambridge University Press, 2006. This excellent book is a standard college-level text.

Weyl, H. *Symmetry.* Princeton, NJ: Princeton University Press, 1983. A classic, first published in 1952, that attempts, quite successfully, to bridge the gap between mathematics and aesthetics.

Wilf, Herbert. *generatingfunctionology.* 2nd ed. San Diego, CA: Academic Press, 1994. A beautiful (advanced) book that is as much about mathematical thinking as it is about generating functions.

Yaglom, I. M. *Geometric Transformations I.* New York: Random House, 1962. An elegant introduction to the transformational way of thinking. Excellent for beginners.

Zeitz, Paul. *The Art and Craft of Problem Solving*. 2nd ed. Hoboken, NJ: Wiley, 2007. My book parallels many of the topics in this course and has a wide variety of problems of many levels of difficulty.

Films and Online Resources:

Art of Problem Solving. http://www.artofproblemsolving.com. This is perhaps the world's preeminent online community of math enthusiasts, most of them young and interested in math contests. It has numerous resources for learning mathematics.

Csicery, George Paul. *Hard Problems*. 2008. DVD. A thought-provoking documentary about the formation of the 2006 U.S. International Mathematical Olympiad team and its adventures at the competition, which was held in Slovenia. This film initially aired on PBS stations around the United States in 2009.

———. *N is a Number*. 2007. DVD. A fascinating 1-hour documentary about the life of Paul Erdös.

The Online Encyclopedia of Integer Sequences. http://www.research.att.com/~njas/sequences/Seis.html. This encyclopedia contains more than 100,000 sequences; it is a math nerd's paradise.