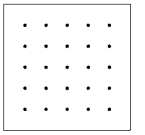
Problem Set 1

Important Stuff

Each Problem Set is divided into Important Stuff, Neat Stuff, and Tough Stuff.

Picture a piece of graph paper. Now picture a dot at each intersection. We'll call this *square dot paper*. A 5-by-5 piece of square dot paper would have five dots in each direction—also known as a "geoboard". But the dot paper can be any size, really. We'll say that the distance from a dot to its nearest neighbor is 1. Segments drawn on square dot paper must start and end at dots, but can be horizontal, vertical, or diagonal at any angle.

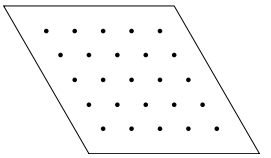


Opener

On a 6-by-6 piece of square dot paper, what *lengths* of segments are possible?

Stuff in boxes is more important than other Important Stuff!

There's another type of dot paper, made from a grid of equilateral triangles instead of squares. This is usually called *isometric dot paper*. A 5-by-5 piece of isometric dot paper looks like a "squished" version of the square dot paper, but still has 25 dots. Again, we'll say that the distance from a dot to its nearest neighbor is 1. As before, segments drawn on isometric dot paper must start and end at dots, but can be horizontal, vertical, or diagonal at any angle.



You may find having square and isometric dot paper useful throughout the course. One copy of each is included with this problem set, or your instructor may have extra copies.

Opener, Part 2

On a 6-by-6 piece of isometric dot paper, what lengths of segments are possible?

This problem might take a while! Keep with it and look for shortcuts and patterns.

Neat Stuff

Here are some more good questions to think about.

1. What happens on larger pieces of the two types of dot paper? Make a conjecture before trying.
2. How many *different* lengths are possible on a piece of n-by-n square dot paper? isometric?
3. What kinds of numbers can be distances on square dot paper? isometric? Be as specific as possible.

Tough Stuff

Here are two much more difficult problems to try.

4. Find a non-right triangle with integer side lengths and no horizontal or vertical segments that can be drawn on square dot paper, or prove that no such triangle exists.
5. Find a scalene triangle with integer side lengths that can be drawn on isometric dot paper, or prove that no such triangle exists.

Throughout, "drawn on" means that all the triangle's vertices must be at dots on the paper.

Problem Set 2

Opener

What positive integers can be written as the sum of two squares? Using the table below, determine whether each n can be written as $n = x^2 + y^2$. Look for patterns... there's more than one!

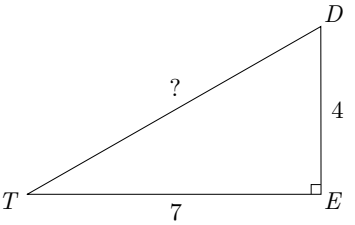
5 can, 6 can't.

1	2	3	4	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72
73	74	75	76	77	78	79	80	81	82	83	84
85	86	87	88	89	90	91	92	93	94	95	96

1. For each prime p, find integers x and y that satisfy $p = x^2 + y^2$, or determine that it's impossible.

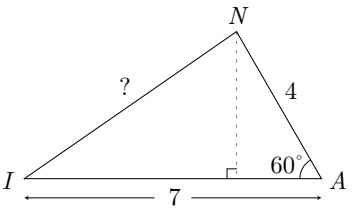
a. 101 **b.** 127 **c.** 419 **d.** 421 **e.** 10009

2. Right triangle *TED* has leg lengths of 7 and 4 as shown. Find the exact length of the third side.



3. Triangle *IAN* has two sides with lengths 7 and 4, and a 60° angle as shown. Read the side note. Find the exact length of the third side.

This question can and should be answered without the use of "cosine" and "sine" or any associated laws.



4. Find all solutions to each equation.

a. $x^2 - 12x + 32 = 0$
b. $x^2 - 12x + 33 = 0$
c. $x^2 - 12x + 34 = 0$
d. $x^2 - 12x + 35 = 0$

e. $x^2 - 12x + 36 = 0$
f. $x^2 - 12x + 37 = 0$

5. A rectangle has perimeter 24 and area 33. What are its exact side lengths?

6. Did you finish finding all of the possible segment lengths on a 6-by-6 piece of isometric dot paper? If not, please do that now.

As mentioned in Problem Set 1, this problem might take a while!

7. Find a segment on a 6-by-6 piece of square dot paper that *isn't* horizontal or vertical, but still has integer length. Find some other segments with this property (on larger pieces of square dot paper), and describe how you could find more.

Neat Stuff

8. Describe some ways to find Pythagorean triples.

Here are some *Pythagorean triples*: $(3,4,5)$, $(5,12,13)$ and $(21,20,29)$.

9. Are there any other right triangles with integer leg lengths and the same hypotenuse length as triangle *TED*?

10. Modify triangle *IAN* from problem 3 so that it has the same $m\angle A = 60°$ and length *IN*, but different (integer) lengths for *IA* and/or *AN*.

11. What positive integers n can be written as the difference of two squares?

Unless specified otherwise, all variables must be integers. Here, you want to describe integers n so that $n = x^2 - y^2$ with x and y integers.

12. What positive integers n can be written as the sum of three squares, the form $n = x^2 + y^2 + z^2$?

13. What kinds of numbers can be distances on square dot paper? isometric?

Tough Stuff

14. What positive integers n can be written as the sum of four squares, the form $n = x^2 + y^2 + z^2 + w^2$?

15. How many *different* lengths are possible on a piece of n-by-n square dot paper? isometric?

16. Find a non-right triangle with integer side lengths and no horizontal or vertical segments that can be drawn on square dot paper, or prove that no such triangle exists.

17. Find a scalene triangle with integer side lengths that can be drawn on isometric dot paper, or prove that no such triangle exists.

18. The quadratic equation $x^2 - 10x + 22 = 0$ has two roots.

- **a.** Find a quadratic equation whose roots are the *squares* of the roots of $x^2 - 10x + 22 = 0$.
- **b.** Find a quadratic equation whose roots are the nth powers of the roots of $x^2 - 10x + 22 = 0$.

19. Find all integer solutions to this system of equations.

Warning: there are many solutions!

$$\begin{aligned} a + b &= cd \\ c + d &= ab \end{aligned}$$

20.

- **a.** Find the dimensions of a rectangle whose perimeter is 20 and whose area is 21, or show that one doesn't exist.
- **b.** Find the dimensions of a rectangular box whose edge-perimeter is 44, whose surface area is 62, and whose volume is 21, or show that one doesn't exist.

CHAPTER

2 Facilitator Notes

The facilitator notes are designed to be used as needed. Each set has two components:

1. **Goals of Each Problem Set:** here we lay out what the principal ideas of each set are.
2. **Notes on Selected Problems:** we identify a few problems that are worth going over in a whole group discussion.

We will put our emphasis on the main goals of each problem set, drawn from the problems in the "Important Stuff."

Problem Set 1

Goals of the Set

Complex numbers were invented to solve problems with real numbers. We're using complex numbers to answer questions that are—at their core—not questions about complex numbers at all, but questions about integers. For example, the question, "how can we find Pythagorean triples?" can most easily be answered by looking at complex numbers.

A big question that frames the first part of this course is, "Which integers can be written as the sum of two squares?" One way to think about this question is to reach into the Gaussian Integers, $\mathbb{Z}[i] = \{a + bi | a, b \in \mathbb{Z}\}$— in other words, the set of complex numbers whose real and imaginary parts are integers.

A more complete treatment of this in Chapter 4.

In this problem set, it's not yet time to mention complex numbers, or Gaussian integers, or even the question,

"which integers can be written as a sum of two squares?" These things will come up later. For this set, we simply want to introduce a geoboard—dot paper with a square lattice. Later, participants will begin to think of this as a geometric representation (in the first quadrant) of $\mathbb{Z}[i]$.

Similarly, eventually participants will begin to think of the isometric dot paper (made up of a lattice of equilateral triangles) as a geometric representation of $\mathbb{Z}[w]$, where $w = \frac{-1+i\sqrt{3}}{2}$. But again, none of this should come up in this set.

Here, $\mathbb{Z}[w]$ is the system of complex numbers of the form $a + bw$ where a and b are integers.

This set is all about introducing the two problems, using square and isometric dot paper. If participants work on nothing but these two problems for this set, consider their work a success!

Notes on the Problems

There are two main problems from this set. Both ask participants to find all possible lengths of segments that connect lattice points on dot paper. The first dot paper they look at is square dot paper, and the second is isometric. Participants should just mess around here. There are a few things people will discover: for example, they may discover that they can find all of the lengths by starting in a corner.

It's not obvious at first that that strategy will work.

We picked the dimensions 6-by-6 on square dot paper here because it's the smallest dimension that we could have chosen where there would two significantly different ways of finding a given length (five). That's because 25 can be written as a sum of two squares in more than one way. For example,

$$25 = 3^2 + 4^2 \quad \text{and} \quad 25 = 0^2 + 5^2$$

Problem Set 2

Goals of the Set

In this set, participants investigate what numbers can be written as the sum of two squares. They will begin to build connections between that question and the question of what lengths are possible on square dot paper. Although these connections are not made explicit yet, the problems lead participants in that direction.

Notes on the Problems

The opener asks for participants to look for patterns in a table. Participants are likely to describe patterns like this one: "numbers that are one less than multiples of four can't be written as the sum of two squares." Or: "numbers that are one more than a multiple of 12 can be written as the sum of two squares." (This turns out not to be true—the first one that fails is 133—but leave this as an open question or conjecture.) Another pattern that may emerge is that if a and b can be written as the sum of two squares, so can ab.

In Problem 1, some of these are "one less than a multiple of four," and they cannot be written as a sum of two squares. Some of the numbers are "one more than a multiple of four," and they *can* be written as a sum of two squares. Again, collect conjectures about this.

In discussing Problem 3, note the sidenote suggesting that students should not use trigonometry—instead, they should stick to 30-60-90 right triangle properties.

Participants don't need to know any trigonometry to solve this problem.

Problem Set 3

Goals of the Set

At first glance, Set 3 looks the same as Set 2, except now participants are looking at isometric dot paper. The main goals of this set are:

- Participants investigate what numbers can be written in the form $x^2 + y^2 - xy$.
- Participants begin to build connections between what numbers can be written in that form and what lengths are possible on isometric dot paper. The connection is firmly established in the first three problems.
- Participants meet complex numbers and conjugates in this set—these are treated with a very light touch at this time. Depending on their backgrounds, participants may see this as a review, a refresher, or a treatment of a subject they haven't seen in years.

Notes on the Problems

This opener is similar in flavor to that of Set 2. In this case, any number one less than a multiple of three cannot be

CHAPTER

4 Mathematical Overview

This overview contains a sample of the mathematical themes that the development team hammered out in the process of designing the course. It contains some of the mathematical background used when creating the problem sets as well as some mathematical extensions that never made it into the "soup" of problems that went out to PCMI. It is written by people who were in on the design but were not involved in the day-to-day classes at PCMI.

In addition to the teachers participating in the course, PCMI hosts research programs in mathematics and education, programs for graduate students and undergraduate faculty, and institutes for staff development professionals. See `pcmi.ias.edu` for more details.

Connecting Algebra with Geometry

Algebra and geometry, two mainstays of school mathematics, have many interconnections at every level, from elementary school through topics at the frontiers of mathematics research. At its core, this course is about connections between algebra and geometry, but the kind of "algebra" and "geometry" we consider in this course lies somewhere between high school and undergraduate mathematics.

This course was offered at PCMI in 2002 and, after extensive revision, again in 2008.

Two themes that underly much of the mathematics in the problem sets are:

- **Algebraic forms often contain attributes that give some fine-grained information about geometry.**
- **Geometric problems often lead to subtle algebraic identities and theorems.**

For example, many teachers know identities to produce Pythagorean triples, like

$$(r^2 + s^2)^2 = (r^2 - s^2)^2 + (2rs)^2 \qquad (1)$$

Sure, you can verify that it's true, and you can use it to generate Pythagorean triples, but how would someone invent it in the first place?

The course develops two paths to the identity and others like it, each of which shows how it arises naturally.

1. The algebra path will concern certain subsystems of the complex numbers, like the Gaussian integers $\mathbb{Z}[i]$—the ring of complex numbers whose real and imaginary parts are integers. The right hand side of (1) peeks out when you square the complex number $r + si$:
$$(r + si)^2 = (r^2 - s^2) + (2rs)\,i$$
This is much more than a curiosity, and the problem sets in the course develop the underlying connections between the algebra of $\mathbb{Z}[i]$ and the geometry of right triangles.

2. The geometry path will concern conic sections, especially the unit circle and the ellipse whose equation is
$$x^2 - xy + y^2 = 1$$
For example, we get another glimpse of (1) if we intersect the circle whose equation is $x^2 + y^2 = 1$ with the line ℓ through $(0, -1)$ with slope $\frac{r}{s}$. One intersection point is $(0, -1)$ itself, but the other is
$$\left(\frac{2rs}{r^2 + s^2}, \frac{r^2 - s^2}{r^2 + s^2}\right)$$
The geometry is trying to tell us something about the algebra of Pythagorean triples.

The course takes a different slant on conics than what one typically finds in school texts, and very little background is assumed. There's no consideration of, for example, foci, directrix, latus rectum, or any of the other classical accoutrements; we're concerned here with finding points that satisfy equations, and that's about all.

The mathematics involved in this course lies at the intersection of geometry, algebra, and number theory. A rigorous and precise development of all the ideas and results would require considerable time and background. But the fifteen problem sets develop all of the central ideas in transparent and special cases. Like all the PCMI courses, these problem sets give teachers a solid foundation from which they can build a more complete edifice through further reading or coursework.

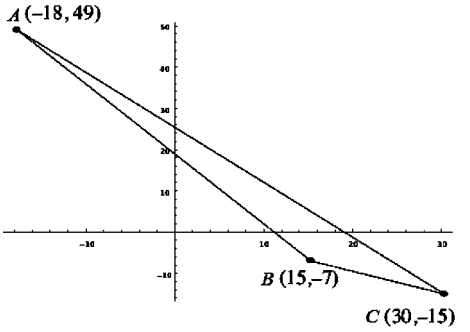
Applications to Teaching

With all the wonderful mathematics around, why did we choose algebraic systems like $\mathbb{Z}[i]$ and special conic sections as central threads? One reason is that the summer

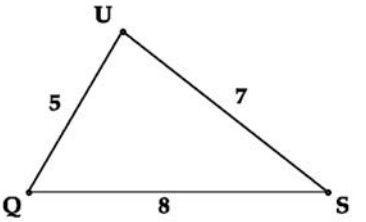
theme for PCMI 2008 was algebraic and analytic geometry. Another reason is that these threads can be used as general purpose tools for teachers, as they create problems for their students. Here are some examples for readers to try; each contains a pleasant surprise:

Mark Saul has a saying to the effect that not all the mathematics used by teachers ends up on the blackboard.

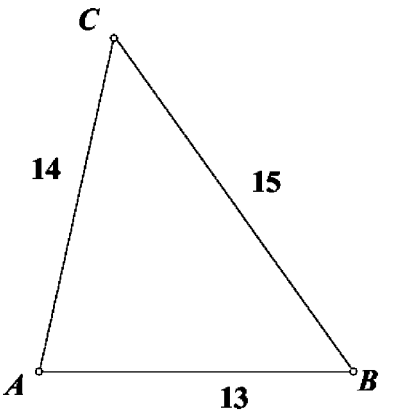
- Find the length of the sides of this triangle:



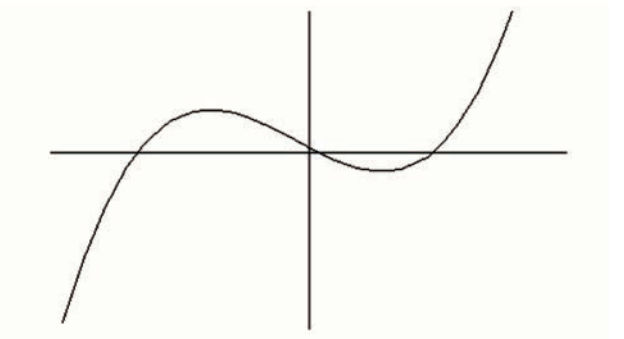
- Find the measure of $\angle Q$:



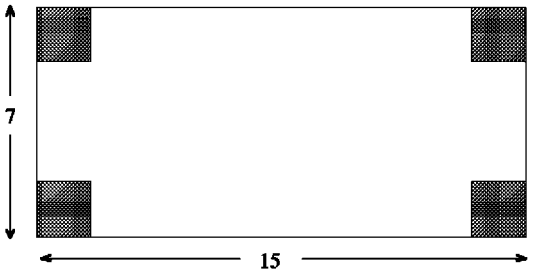
- Find the area of $\triangle ABC$:



- Find the extrema and inflection point for the graph of $y = 140 - 144x + 3x^2 + x^3$:



- Squares are cut out of the corners of a 7×15 rectangle and the sides are folded up to make a box. Find the size of the cut-out that maximizes the volume of the box:



Making up problems like these is the object of a great deal of teacher-room discussion. There are many *ad-hoc* methods for finding examples, but it turns out that both the arithmetic of systems like the Gaussian integers and the geometry of the conics can be used as general-purpose tools to solve a wide class of meta-problems.

We call these problems "meta-problems"—mathematical problems that spring from making up mathematics problems for students.

One of the oldest meta-problems has to be the generation of Pythagorean triples—triples of integers (a, b, c) so that
$a^2 + b^2 = c^2$. The form of the left-hand side makes many teachers' minds turn to either complex numbers or circles (and the distance formula):

Complex Numbers. The equation $a^2 + b^2 = c^2$ can be written as

$$(a + bi)(a - bi) = c^2$$

So, the search for Pythagorean triples is the search for Gaussian integers z so that the product of z and its complex conjugate (usually denoted by $\overline{z}$) is a perfect square. Participants who work through the problems in this course develop algebraic and geometric properties of

complex conjugation, and then define the *norm* of a complex number z to be the product of z and its conjugate:

$$N(z) = z\,\overline{z}$$

The search for Pythagorean triples, then, comes down to the search for Gaussian integers whose norms are perfect squares. But a fundamental property of the norm function, one that is developed in this course, is that the norm is *multiplicative*:

$$N(zw) = N(z)\,N(w)$$

It's unfortunate that the word "norm" is used in several different ways in mathematics. Sometimes, it's synonymous with "length" or "size." We'll use it here to mean the product of a complex number and its conjugate. There's more about norms and conjugates in section 4 on page 93.

It follows that

$$N(z^2) = (N(z))^2 \tag{2}$$

The left-hand side of this is the norm of something (it's the norm of z^2), so it has the form $a^2 + b^2$. The right hand side is the square of an integer. Hence we have a Pythagorean triple. For example, suppose $z = 3+2i$. Then $N(z) = 9 + 4 = 13$. And

$$z^2 = (3 + 2i)^2 = 5 + 12i$$

Equation (2) implies

$$5^2+12^2 = N(5+12i) = N\left((3 + 2i)^2\right) = N\left((3 + 2i)\right)^2 = 13^2$$

In other words, to find a Pythagorean triple, take a complex number $z = r + si$ and let $u = z^2$. If $u = a + bi$, then $\left(a, b, \sqrt{N(u)}\right)$ will be a Pythagorean triple.

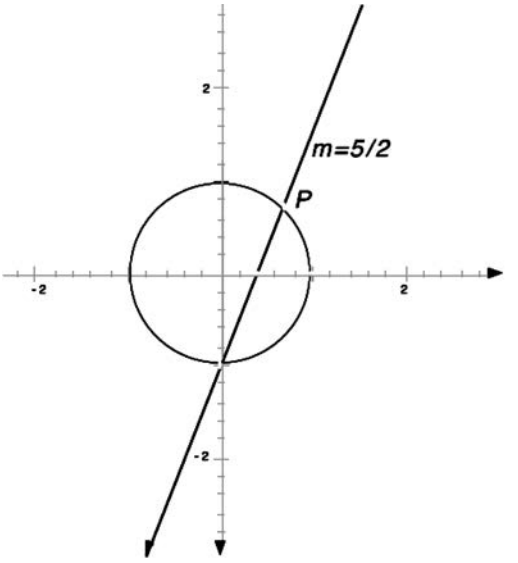
By now, you are probably itching to calculate $(r + si)^2$ and take its norm.

Circles. Another way to get Pythagorean triples is to find points with rational coordinates on the unit circle—that is, on the graph of $x^2 + y^2 = 1$. For example, $\left(\frac{3}{5}, \frac{4}{5}\right)$ is such a point, and it produces the famous $(3, 4, 5)$ triple. The reason this works is that if $c \neq 0$, $a^2 + b^2 = c^2$ if and only if

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1$$

so $\left(\frac{a}{c}, \frac{b}{c}\right)$ is a point (with rational coordinates) on the graph of $x^2 + y^2 = 1$. How does one find such points? One way is to use what some people call the "method of sweeping lines:" There are four rational (integer, even) points on the graph: $(\pm 1, 0)$, and $(0, \pm 1)$. Pick one of them, say $(0, -1)$ and draw a line through it with rational slope greater than 1.

Bill McCallum, Professor of Mathematics at the University of Arizona, is one such person who talks like this.



Without any calculations, you can reason that, if a line with rational slope intersects the circle twice, and if one intersection point is rational, then the other intersection point is rational. For example, if the slope is $\frac{5}{2}$, you'd need to solve the system

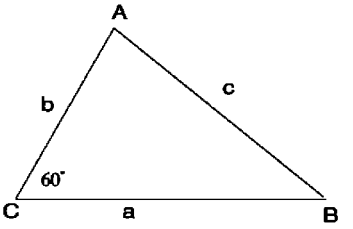
$$\begin{cases} x^2 + y^2 &= 1 \\ y &= \frac{5}{2}x - 1 \end{cases}$$

The solution in the first quadrant is $\left(\frac{20}{29}, \frac{21}{29}\right)$, producing the famous $(20, 21, 29)$ Pythagorean triple. See [3] for more examples.

$21 + 20i$ is the square of what Gaussian integer?

It turns out that these two ideas, norms and rational points on curves, are closely related and quite generalizable—these are two threads that will run through some of the neat and tough stuff in the problem sets.

For example, consider the problem of finding integer-sided triangles with a 60° angle (like the 5-8-7 triangle above). We want integers a, b, and c so that the triangle with sides (a, b, c) has a 60° angle:



Calling c the "hypotenuse" of the triangle, the law of cosines tells us that

$$\begin{aligned} c^2 &= a^2 + b^2 - 2ab\cos\angle C \\ &= a^2 + b^2 - 2ab\frac{1}{2} \\ &= a^2 - ab + b^2 \end{aligned}$$

So, we're looking for triples (a, b, c)—we call them "Eisenstein triples" after Gauss' student George Eisenstein—so that

$$a^2 - ab + b^2 = c^2 \tag{3}$$

Once again, there are two general purpose ways to generate Eisenstein triples:

Complex numbers. This time, we want to find a subring of the complex numbers where the norm has form $a^2 - ab + b^2$. Let's do some wishful thinking: if $a^2 - ab + b^2$ were the "norm" of something, and if that norm function behaved like the ordinary norm from the Gaussian integers (in particular, if the norm of a product were the product of the norms), then we'd be able to use the same method: Take a thing, square it, and then its norm would

- have the right form $(a^2 - ab + b^2)$ and
- be a perfect square.

Well, this is not the norm of $a + bi$, but suppose it were the norm of $a + b\omega$ for some complex number ω. Let's work backwards and see if we could figure out what ω would have to be. Remember, the norm is the product of the number and its conjugate, so, if a and b are integers,

$$\begin{aligned} N(a + b\omega) &= (a + b\omega)\left(\overline{a + b\omega}\right) \\ &= (a + b\omega)\left(\overline{a} + \overline{b\omega}\right) \\ &= (a + b\omega)\left(\overline{a} + \overline{b}\,\overline{\omega}\right) \\ &= (a + b\omega)\left(a + b\,\overline{\omega}\right) \\ &= a^2 + ab(\omega + \overline{\omega}) + b^2(\omega\,\overline{\omega}) \end{aligned}$$

and if we want this to be $a^2 - ab + b^2$, then we want

$$\begin{aligned} \omega + \overline{\omega} &= -1 \quad \text{and} \\ \omega\,\overline{\omega} &= 1 \end{aligned}$$

The conjugate of the sum is the sum of the conjugates, the conjugate of the product is the product of the conjugates, and the conjugate of a real number is itself.

Well, that pretty much nails ω down: we know the sum of ω and its complex conjugate (it's -1) and we know

the product $\omega\,\overline{\omega}$ (it's 1). So, ω is a root of the quadratic equation

$$x^2 + x + 1 = 0$$

This is because of the old chant from high school:

x^2-(the sum of the roots) $x+$(the product of the roots) $= 0$

Using the quadratic formula, we can take ω to be

$$\omega = \frac{-1 + i\sqrt{3}}{2}$$

and we can now generate as many triples of integers (a, b, c) so that $c^2 = a^2 - ab + b^2$ as we like. More details are in [2].

The other root is then $\overline{\omega} = \frac{-1-i\sqrt{3}}{2}$. That will work, too. We call $\mathbb{Z}[\omega]$ the "Eisenstein integers."

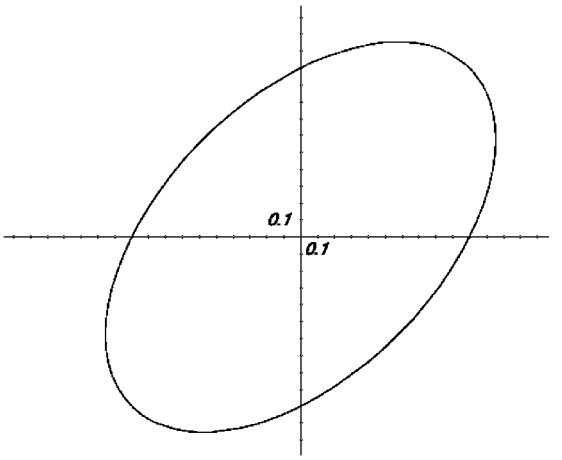
Conics. Dividing both sides by c^2, equation (3) can be written as

$$\left(\frac{a}{c}\right)^2 - \left(\frac{a}{c}\right)\left(\frac{b}{c}\right) + \left(\frac{b}{c}\right)^2 = 1$$

so that $\left(\frac{a}{c}, \frac{b}{c}\right)$ is a point with rational coordinates on the graph of

$$x^2 - xy + y^2 = 1$$

In many precalculus courses ([5], for example), students show that the graph of $x^2 - xy + y^2 = 1$ is an ellipse with major axis at a 45° angle to the coordinate axes:



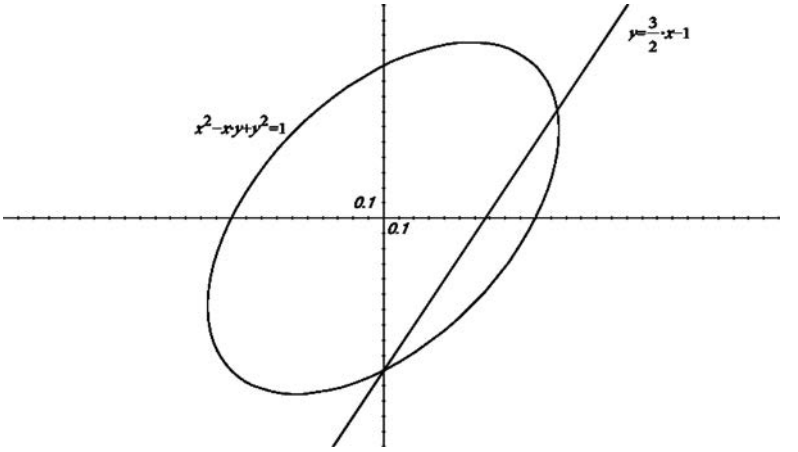
In our professional development work, we've found that teachers really like figuring out the graph.

There are four rational points on this graph that produce "trivial" Eisenstein triples:

$$(1, 0), (0, 1), (-1, 0), (0, -1)$$

If we pick one of these points, say $(0, -1)$ and draw a line through it with rational slope greater than 1, say $\frac{3}{2}$, we

get a second intersection point, one that lies in the first quadrant:



Solving the system

$$\begin{cases} x^2 - xy + y^2 = 1 \\ y = \frac{3}{2}x - 1 \end{cases}$$

we find that the second intersection point is

$$\left(\frac{8}{7}, \frac{5}{7}\right)$$

And $(8, 5, 7)$ is an Eisenstein triple—this is the triple we used to generate the law of cosines example earlier:

$$8^2 - 8 \cdot 5 + 5^2 = 7^2$$

The ability to generate Pythagorean and Eisenstein triples turns out to be the key to solving a host of task design problems, including all of the ones we listed at the start of this section. And the method of sweeping lines gives a general purpose tool for creating problems that allow students to get used to ideas without the computational overhead of messy numbers.

Thinking Deeply about Simple Things

"To Think Deeply about Simple Things" was a famous motto of Arnold Ross, founder of the Ross programs at Notre Dame and then Ohio State, programs that have been the inspiration for many content-based professional development programs including *PROMYS for Teachers* at

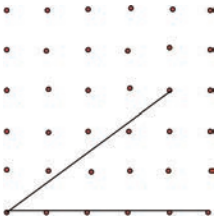
Boston University and the *Secondary School Teacher Program* at PCMI.

A perfect example of this maxim happens early in the course, when participants experiment with different lengths on various types of geoboards—square dot paper and isometric dot paper). Thinking deeply about these devices that are used by young children leads to some very sophisticated and lovely mathematics.

For example, counting all the segments of different lengths emanating from one corner of a square geoboard, it seems at first that the number of distinct lengths on an $n \times n$ geoboard is related to triangular numbers, but the pattern breaks down at $n = 6$ because 25 can be written as a sum of squares in two different ways:

$$25 = 5^2 + 0^2 = 3^2 + 4^2$$

so there are two distinct segments of length 5 that would get counted twice in the counting scheme:



This leads to the question of how many ways an integer can be written as the sum of two squares. Before asking that question, it might make sense to ask a simpler question:

> **Big Question**: *Which integers can be written as a sum of two squares?*

We've asked this simpler question of students and teachers many times in classes and workshops, and it often happens, after some numerical experiments, that people can identify which numbers are sums of two squares well before they can articulate a rule for identifying them. It's worth spending some time helping a class or group get to the point where it can formulate a precise conjecture. After that, they can worry about the "number of ways" refinement.

Numerical experiments lead to some interesting conjectures. For example, making a list of *primes* that are sums of two squares points out that they all seem to be of the form $4k + 1$. And participants can prove that a number of the form $4k+3$ can't be written as the sum of two squares, because if you square in integer, the remainder when you divide by 4 is either 0 or 1, so when you add two squares and divide by 4, the remainder is either 0, 1, or 2:

$a^2 \mod 4 \rightarrow$ $b^2 \mod 4 \downarrow$	0	1	$\swarrow\ a^2 + b^2 \mod 4$
0	0	1	
1	1	2	

It's fun to throw out some integers on the board like

$$13, 21, 45, 101, 202, 105, 260, \ldots$$

and ask for a thumbs up or thumbs down. We've often seen, after a great deal of numerical experiment like this, the following conjecture emerge:

> **The SOTS Conjecture:** An integer n is the sum of two squares if and only if in the prime factorization of n, the primes of the form $4k+3$ show up with even exponent.

Of course, the conjecture is usually not stated so precisely, and not in full generality. We've had good luck listing partial conjectures as they develop on chart paper or on the board.

If this is true, you know, with no calculations, that

13	is a SOTS
21	isn't a SOTS
45	is a SOTS
101	is a SOTS
202	is a SOTS
105	isn't a SOTS
260	is a SOTS
$2^{10} \cdot 3^8 \cdot 101$	is a SOTS
$2^{10} \cdot 3^7 \cdot 101$	isn't a SOTS

Another conjecture that often comes up is that if you multiply two numbers that are the sum of two squares together, their *product* is also a sum of two squares. For example,

$$5 = 2^2 + 1^2, 13 = 3^2 + 4^2, \text{ and } 5 \cdot 13 = 65 = 4^2 + 7^2$$

In general,

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (bc + ad)^2 \qquad (4)$$

Sure, you can show that this is true, but where does it come from? Going back to the themes on page 76, we might ask what the algebra is trying to tell us. Identity (4) has complex numbers written all over it.

Indeed, one way to think about this is to think of the geoboard as a geometric representation of $\mathbb{Z}[i]$—in the complex plane, the Gaussian integers lie at the lattice points of the Cartesian coordinate system. With this perspective, (4) is nothing other than our old friend

$$N(zw) = N(z)\,N(w)$$

One route to the proof of the SOTS Conjecture is to use this connection to the Gaussian integers and the *Law*

of Decomposition in $\mathbb{Z}[i]$, a deep theorem that tells how primes in $\mathbb{Z}$ decompose when looked at as elements of $\mathbb{Z}[i]$. The details are on page 96 of the appendix to this overview.

Once the class has pushed the Big Question as far as it can, you might want to refine it by asking *in how many ways* an integer can be written as the sum of two squares. For example, $65 = 64 + 1$ and $65 = 49 + 16$. If you "think $\mathbb{Z}[i]$," these two ways come from different factorizations of 65:

$$65 = N(8 + i) = (8 + i)(8 - i) \quad \text{and}$$
$$65 = N(7 + 4i) = (7 + 4i)(7 - 4i)$$

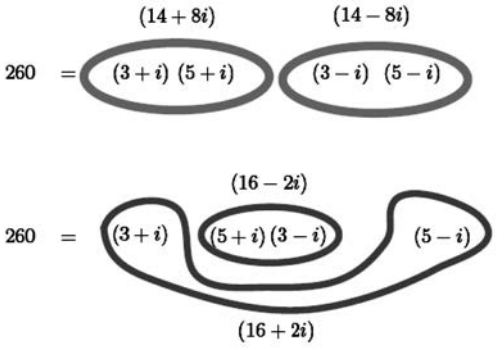
Another example: take 260. We have:

$$260 = 14^2 + 8^2 = (14 + 8i)(14 - 8i)$$
$$260 = 16^2 + 2^2 = (16 + 2i)(16 - 2i)$$

The different representations as SOTS come from a more complete factorization of 260:

$$260 = (3 + i)(5 + i)(3 - i)(5 - i)$$

Different groupings give different representations:



This is not unlike the "different" ways children factor 12, as, say, 6×2 and 4×3.

So, the question of finding the number of ways an integer n can be written as a sum of squares comes down to finding the number of "different" Gaussian integers α such that

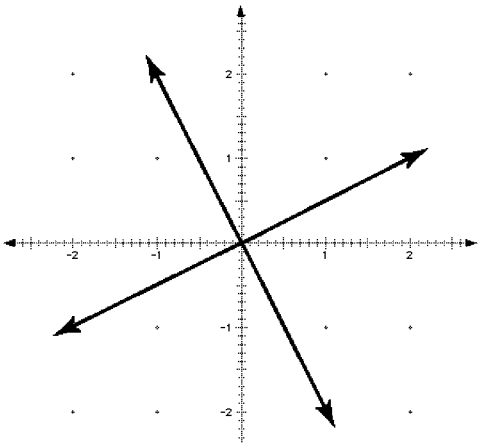
$$N(\alpha) = n$$

Let's agree about what we mean by "different." We don't want to call things different if they are the same except for a unit factor (such numbers have the same norm in a sort of trivial way). Well, if you look at a Gaussian integer α,

Here, we're using the geometric interpretations of multiplying by 1 (leave it alone), i (rotate 90° about the origin), -1 (rotate 180° about the origin), and $-i$ (rotate 270° about the origin).

you'll see that one of its "associates" (that is, either α, $-\alpha$, $i\alpha$, or $-i\alpha$) lies in the first quadrant.

The *units* in $\mathbb{Z}[i]$ are ± 1, $\pm i$. These are the only four Gaussian integers whose reciprocals are also Gaussian integers.



Let's pick that associate in the first quadrant to count when we're looking at sums of squares. By "first quadrant," we mean

For shorthand, let's denote this first quadrant by Q_1

$$\{a + bi \mid a > 0, b \geqslant 0\}$$

So, if we define

$$s(n) = \begin{array}{l}\text{the number of ways you can write } n \\ \text{as the sum of two squares}\end{array}$$

we can nail down $s(n)$ by using by counting one representation from each set of four associates whose norm is n:

Definition

$$s(n) = |\{\alpha \in Q_1 \mid N(\alpha) = n\}|$$

where $|S|$ means the number of elements of the set S.

It's great fun to let participants tabulate $s(n)$ for n between say, 1 and 100—all kinds of conjectures emerge. One of the nicest is that s seems to be multiplicative:

$s(mn) = s(m) \cdot s(n)$ provided m and n have no common factor

In other words, gather some information about the function s. Dividing up the work, complete the following table:

n	$s(n)$	n	$s(n)$	n	$s(n)$	n	$s(n)$	n	$s(n)$
1	1	11		21		31		41	
2	1	12	0	22	0	32	1	42	
3	0	13	2	23	0	33		43	
4		14		24		34		44	
5		15	0	25		35		45	
6	0	16		26		36		46	
7		17		27	0	37		47	
8		18	1	28		38		48	
9	1	19		29		39	0	49	
10		20	2	30	0	40		50	

n	$s(n)$	n	$s(n)$	n	$s(n)$	n	$s(n)$	n	$s(n)$
51		61		71		81		91	
52		62		72		82		92	
53		63		73		83		93	
54		64		74		84		94	
55		65		75		85		95	
56		66		76		86		96	
57		67		77		87		97	
58		68		78		88		98	
59		69		79		89		99	
60		70		80		90		100	

We've seen groups of teachers get very close to another very deep theorem, this one due to Fermat:

Theorem 4.1 (Fermat)

The number of representations of an integer n as the sum of two squares (that is, $s(n)$) is the excess of the number of factors of n of the form $4k+1$ over the number of factors of n of the form $4k+3$.

Ask participants to state some conjectures about the function s. When is it zero? When is it non-zero? Can you describe the n for which $s(n) > 1$? Is there a number n so that $s(n) > 5$?

Take some time to savor this theorem—write out several examples to illustrate it. The proof (at least the one we know) is hard. It's developed in the appendix to this overview.

Fermat's theorem gives a way to compute the values of s, but the tabulation shows it to be very erratic. When

there's no apparent regularity to the values of a function, sometimes it's worth looking at it's *average* value. The average value of s (or of any function) on $[1, n]$ is just what it sounds like: the sum of $s(1)$ up to $s(n)$, divided by n. In symbols, it's the function s_{av}, defined by

$$s_{av}(n) = \frac{1}{n}\sum_{k=1}^{n} s(k) = \frac{s(1) + s(2) + \cdots + s(n)}{n}$$

Gauss was quite interested in the average value of s. But Gauss was a little more liberal than Fermat: he counted all sums of squares from all *four* quadrants. So, for Gauss, 25 can be written as a sum of squares in 12 ways:

$$\begin{array}{ccc} 5^2+0^2 & 4^2+3^2 & 3^2+4^2 \\ 0^2+5^2 & (-3)^2+4^2 & (-4)^2+3^2 \\ (-5)^2+0^2 & (-4)^2+(-3)^2 & (-3)^2+(-4)^2 \\ 0^2+(-5)^2 & 3^2+(-4)^2 & 4^2+(-3)^2 \end{array}$$

Let's call Gauss' function "t."

Definition

The function t is defined by

$$t(n) = |\{\alpha \in \mathbb{Z}[i] \mid N(\alpha) = n\}|$$

Notice that $t(n) = 4s(n)$.

In the problem sets, teachers are asked to investigate the average value of t. That is, let

$$t_{av}(n) = \frac{1}{n}\sum_{k=1}^{n} t(k) = \frac{t(1) + t(2) + \cdots + t(n)}{n}$$

Some ideas:

- Tabulate t_{av}, looking for patterns.
- Look at t_{av}'s "asymptotic" behavior (what happens to $t_{av}(n)$ as n gets really big.

The surprise is that the average value of t seems to settle in on π. And it's true:

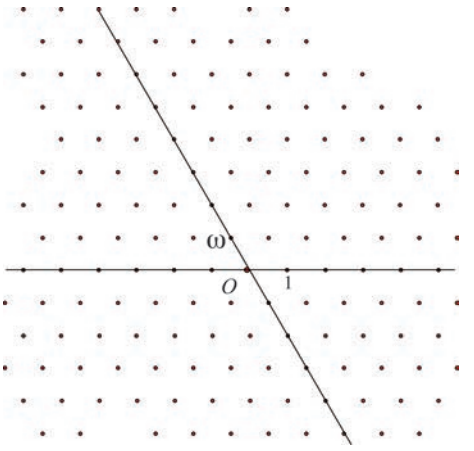
$$\lim_{n\to\infty} t_{av}(n) = \pi$$

This is known as *Gauss' circle theorem.* The participants can probably figure out why this must be true if they think about it for awhile. Essentially, $t_{av}(n)$ gives the

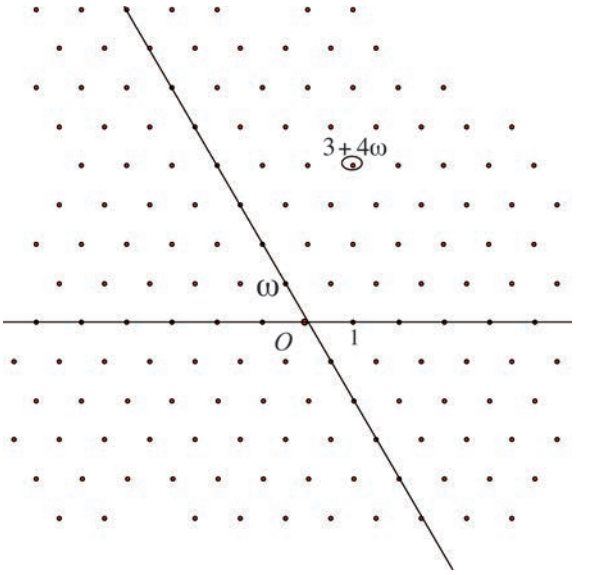
number of lattice points inside a circle of radius $\sqrt{n}$, and so this should be close to the area of that circle: πn, and this estimate gets better for large n. So, $\frac{t_{av}(n)}{n}$ should be close to $\frac{\pi n}{n} = \pi$. A nice discussion that makes this precise is in [1]. And Glenn Stevens chimes in

> Just to be weird ...
>
> According to the analytic class number formula, the average should be:
>
> $$2^{r_1}(2\pi)^{r_2}hR/w\sqrt{|D|}$$
>
> where in the case of $\mathbb{Q}(i)$ we have
>
> - $r_1 = 0$ (the number of real embeddings);
> - $r_2 = 1$ (the number of complex embeddings);
> - $h = 1$ (the class number)
> - $R = 1$ (the regulator)
> - $w = 4$ (the number of units); and
> - $D = -4$ (the discriminant).
>
> The analytic class number says the average for $s(n)$ is $\frac{2\pi}{4\sqrt{4}}$, which is $\frac{\pi}{4}$.
>
> Of course, this argument is circular, since the picture of lattice points and circles is how the analytic class number is proved in the first place.

The same kinds of things happen on the isometric geoboard. Just as the geoboard is a geometric representation of $\mathbb{Z}[i]$, this can be thought of as a representation of $\mathbb{Z}[\omega]$. Instead of perpendicular axes (real and imaginary), we have axes that intersect at a 120° angle (real and multiples of ω).



So, $3 + 4\omega$ means that you go 3 along the real axis and then 4 along the omega axis:



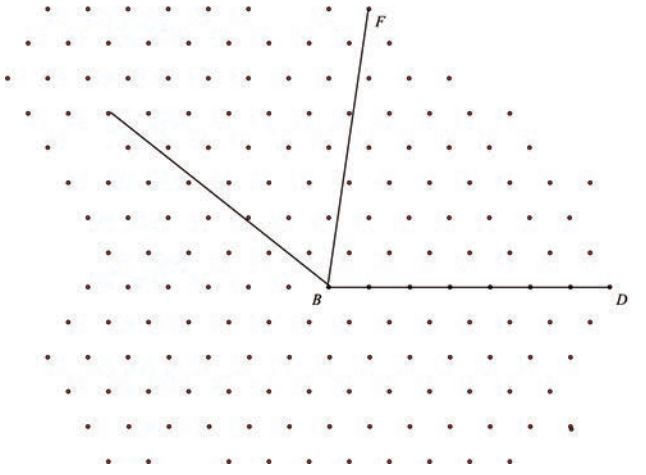
It's a very good investigation to carry over the experiments for $\mathbb{Z}[i]$ to $\mathbb{Z}[\omega]$. Both rings have unique factorization, both have a law of decomposition, and you can ask for how many ways an integer can be represented as

$$a^2 - ab + b^2 = N(a + b\omega)$$

Just as in the geoboard problem, apparent patterns break down. In fact,

$$49 = N(7 + 0\omega) = N(5 + 8\omega)$$

There are other Eisenstein integers with the same norm.



You might ask teachers to find them all: all the lattice points on the isometric geoboard that lie on a circle of radius 7.

And what about an analogous result for the Gauss circle theorem? Darryl Yong has a lovely Mathematica program that calculates the number of ways an integer can be represented as a norm from $\mathbb{Z}[\omega]$, and it suggests that the limit of the average value (again, counting all around the circle) is $\frac{2\pi}{\sqrt{3}}$. This agrees with Glenn's invocation of the class number formula that gives one sixth of this number.

"One sixth" because there are six units in $\mathbb{Z}[\omega]$: $\pm 1, \pm\omega, \pm\omega^2$.

Conjugates and Norms

The word "conjugate" is used inconsistently and incorrectly throughout school mathematics. Students are told that the conjugate of $3 + i$ is $3 - i$, but the conjugate of $2 + \sqrt{3}$ is $2 - \sqrt{3}$.

Students often ask, "If the conjugate of $a + bi$ is $a - bi$, the conjugate of $2 + \sqrt{3}$ should be itself, no?"

There are all kinds of abstract formulations of conjugation, and the one that ties them all together in algebra comes from Galois theory: If K/L is a finite Galois extension of fields, with Galois group G and α is in K, the *conjugates* of α over L form the set

$$\{\sigma(\alpha) \mid \sigma \in G\}$$

This is not the formulation we want to use in school or at PCMI, but it points out the important fact that conjugate is defined *over a field*—numbers in K are conjugate over L. If you change the base field, you likely change the set of conjugates.

Again, in other settings, the word "norm" is used differently. The "norm of a vector" usually means its length. In fact, in analysis, norm is often used as a proxy for a function that behaves like length. But, in the sense we use it in this course, *the norm of α is the product of α and its conjugates.*

Luckily, for our purposes, there's another equivalent definition, easier to state. Suppose α is a complex (possibly real) number, L is a number system like $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Q}[i]$, $\mathbb{Q}[\sqrt{2}]$, $\mathbb{Q}[\omega]\ldots$—the kinds of number systems that show up in school algebra. And suppose that α satisfies an irreducible polynomial equation with coefficients in L. Then you can show that there's only one such equation (up to scaling) and the **conjugates of α over** L are the roots of this equation. For example:

- The conjugates of i over $\mathbb{R}$ are i and $-i$ because they satisfy $x^2 + 1 = 0$.
- And i and $-i$ are also conjugate over $\mathbb{Q}$ because the coefficients of $x^2 + 1 = 0$ are in $\mathbb{Q}$.

- The conjugates of ω over $\mathbb{R}$ are ω and $\omega^2 = \frac{1}{\omega}$ because they satisfy $x^2 + x + 1 = 0$. And ω and ω^2 are also conjugate over $\mathbb{Q}$.
- The only conjugates of ω over $\mathbb{C}$ is ω, because it satisfies the linear equation $x - \omega = 0$.

... and the only conjugate of i over $\mathbb{C}$ is i.

- Suppose $\zeta = \cos\frac{2\pi}{5} + i\sin\frac{2\pi}{5}$. Then $\zeta^5 = 1$, but ζ satisfies a lower degree polynomial equation over $\mathbb{R}$, namely

$$x^4 + x^3 + x^2 + x + 1 = 0$$

  This left hand side is irreducible over $\mathbb{R}$, and its roots are $\{\zeta, \zeta^2, \zeta^3, \zeta^4\}$, so these are the conjugates of ζ over $\mathbb{R}$ (and over $\mathbb{Q}$ as well).
- Over $\mathbb{Q}$, $\sqrt{2}$ and $-\sqrt{2}$ are conjugate because they are roots of $x^2 - 2 = 0$ (the Galois extension here is $\mathbb{Q}[\sqrt{2}]/\mathbb{Q}$). But over $\mathbb{R}$ or $\mathbb{C}$, $\sqrt{2}$ is conjugate only to itself.
- Suppose $\zeta = \cos\frac{2\pi}{7} + i\sin\frac{2\pi}{7}$. Then the following three real numbers are conjugate over $\mathbb{Q}$:

$$\begin{aligned}\zeta + \zeta^6 &= 2\cos\tfrac{2\pi}{7},\\ \zeta^2 + \zeta^5 &= 2\cos\tfrac{4\pi}{7}, \text{and}\\ \zeta^3 + \zeta^4 &= 2\cos\tfrac{6\pi}{7}\end{aligned}$$

  because these three numbers satisfy the irreducible cubic

$$x^3 + x^2 - 2x - 1 = 0$$

  Note that each of the three roots is real, so each has only itself as a conjugate over $\mathbb{R}$.

See [5] for a proof. Technically, if L/K is a field extension of the kind described above, the norm is a mapping from K to L.

So, the norm of a number α over L is the *product of its conjugates* over L. Note that we need the referent base system L. Using the equation definition, the norm is (plus or minus) the constant term in the irreducible equation satisfied by α. Hence, the norm lies in L.

If the irreducible equation satisfied by α over L is quadratic, it has only one conjugate besides itself. So the norm is a product of two numbers, and it lies in L. That's why, for quadratic irrationalities, you can use the conjugates to rationalize denominators in fractions. So, over $\mathbb{Q}$, $7 + 2\sqrt{3}$ and $7 - 2\sqrt{3}$ satisfy

$$x^2 - (\text{the sum})\cdot x + (\text{the product}) = 0 \quad \text{or}$$
$$x^2 - 14x + 45 = 0$$

The only conjugate of a number in K over K is itself. So, the only conjugate of $7 + 2\sqrt{2}$ over $\mathbb{R}$ is $7 + 2\sqrt{2}$.

The norm is 45. Note that this definition (product of the conjugates) makes it clear that norm is multiplicative.

Another function of interest is the *sum* of the conjugates. This is called the *trace*. For example, the trace of $7 - 2\sqrt{3}$ is 14. Just as norm is multiplicative, trace is *additive*—the trace of a sum of numbers is the sum of their traces. The trace of a number is the negative of the coefficient of the penultimate term. So, if α is algebraic over L, and the irreducible equation os quadratic, you can write the equation like this:

In all these irreducible equations, we're assuming you clear common factors and make the leading coefficient 1.

$$x^2 - \mathrm{Tr}(\alpha) \cdot x + \mathrm{N}(\alpha) = 0$$

If the equation isn't quadratic, it still looks like

$$x^n - \mathrm{Tr}(\alpha) \cdot x^{n-1} + \cdots + \mathrm{N}(\alpha) = 0$$

and the other coefficeints are the elementary symmetric functions of the conjugates of α—the sum of all products taken 2 at a time, taken 3 at a time, and so on. This follows from the factor theorem.

One more thing about the quadratic case. The term *complex conjugate* has come to mean the conjugate of a complex number over $\mathbb{R}$. It has the property that it is additive and multiplicative and fixes precisely the real numbers. So, there's no choice about the complex conjugate of $7 + 3\omega$:

$$\begin{aligned}\overline{7 + 3\omega} &= \overline{7} + \overline{3} \cdot \overline{\omega} \\ &= 7 + 3 \cdot \overline{\omega} \\ &= 7 + 3\omega^2\end{aligned}$$

Appendix: Some Proofs

The following discussion offers an interesting addition for those with an interest or background in Number Theory. It is not necessary to understand this material in order to utilize the book but can provide a nice mathematical excursion.

The Law of Decomposition for $\mathbb{Z}[i]$

Just to get a flavor for how the proofs go, let's prove a few results that look at which primes are the sums of squares. This leads up to the famous *law of decomposition* that we'll state without proof (this is getting too long already). You'll have all you need to fill in the details if the mood strikes you.

Lemma 4.2

An odd prime p *is the sum of two squares* $\Leftrightarrow -1$ *is a square in* $\mathbb{Z}/p\mathbb{Z}$

$\mathbb{Z}/p\mathbb{Z}$ is the integers mod p. It has many of the same properties as the rational numbers. You can add, subtract, multiply, *and* divide.

Proof. Suppose p is a sum of squares:

$$p = a^2 + b^2$$

Then p isn't a factor of b (why?) so there's a reciprocal b^{-1} for b in $\mathbb{Z}/p\mathbb{Z}$. Then, in $\mathbb{Z}/p\mathbb{Z}$,

$$\left(ab^{-1}\right)^2 + 1 = 0$$

and -1 is a square in $\mathbb{Z}/p\mathbb{Z}$. To go the other way, if -1 is a square in $\mathbb{Z}/p\mathbb{Z}$, there is some integer a so that $p \mid a^2 + 1$. But then, in $\mathbb{Z}[i]$

$$p \mid (a + i)(a - i)$$

But p isn't a factor of either $a + i$ or $a - i$ (because $p(x + yi) = px + pyi$ and there is no integer y so that $py = \pm 1$), so (Euclid's lemma), p is not a prime in $\mathbb{Z}[i]$. Suppose p factors

Euclid's lemma: p is a prime in $\mathbb{Z}[i] \Leftrightarrow$ whenever $p \mid \alpha\beta$, either $p \mid \alpha$ or $p \mid \beta$. Make sense?

$$p = \mathfrak{p}\mathfrak{q}$$

Then

$$p^2 = N(p) = N(\mathfrak{p})N(\mathfrak{q})$$

This is an equation in $\mathbb{Z}$. By the fundamental theorem in $\mathbb{Z}$ (and some fussiness that you can sort out),

Actually, the two factors of p are conjugates, but we don't need that.

$$N(\mathfrak{p}) = p$$

But the norm of a Gaussian integer is a sum of two squares. ∎

Lemma 4.3

-1 is a square in $\mathbb{Z}/p\mathbb{Z} \Leftrightarrow p \equiv 1 \pmod 4$

Proof. By Fermat's little theorem, $a^{p-1} - 1 = 0$ for all non-zero a in $\mathbb{Z}/p\mathbb{Z}$. So, the polynomial $x^{p-1} - 1$ factors like this in $\mathbb{Z}/p\mathbb{Z}[x]$:

This follows from the factor theorem in algebra 2.

$$x^{p-1} - 1 = (x-1)(x-2)(x-3)\dots(x-(p-1))$$

put $x = 0$. You get (again, in $\mathbb{Z}/p\mathbb{Z}$):

This implies that $(p-1)! \equiv -1 \pmod p$. This is known as Wilson's theorem.

$$-1 = (-1)(-2)(-3)\dots(-(p-1))$$

But look down from the end of this product:

$$\begin{aligned} p - 1 &= -1 \\ p - 2 &= -2 \\ p - 3 &= -3 \\ &\vdots \quad \vdots \end{aligned}$$

So:

$$-1 = (-1)(-2)(-3)\dots\left(\frac{p-1}{2}\right)\left(-\frac{p-1}{2}\right)\dots(-(-3))\,(-(-2))\,(-(-1))$$

or

$$-1 = \left(\left(\frac{p-1}{2}\right)!\right)^2 (-1)^{\frac{p-1}{2}}$$

So, suppose $p \equiv 1 \pmod 4$. Then $(-1)^{\frac{p-1}{2}} = 1$, so

$$-1 = \left(\left(\frac{p-1}{2}\right)!\right)^2$$

and -1 is a square in $\mathbb{Z}/p\mathbb{Z}$. Conversely, suppose $-1 = \alpha^2$ in $\mathbb{Z}/p\mathbb{Z}$. Then

Check that p isn't a factor of α.

$$\begin{aligned} (-1)^{\frac{p-1}{2}} &= (\alpha^2)^{\frac{p-1}{2}} \\ &= (\alpha)^{p-1} \\ &= 1 \quad \text{by little Fermat} \end{aligned}$$

so, $p \equiv 1 \pmod 4$ ■

If $p \equiv 3 \pmod 4$, $(-1)^{\frac{p-1}{2}} = -1$.

Corollary 1

An odd prime p is a sum of two squares in $\mathbb{Z} \Leftrightarrow p \equiv 1 \pmod 4$.

Another approach

There's another proof of lemma 4.3, due to Dirchlet, that capitalizes on high school algebra. Here's a sketch of one direction.

Suppose that p is a prime that's 1 mod 4. Then $p-1$ is divisible by 4, say $p-1 = 4k$. By little Fermat, over $\mathbb{Z}/p\mathbb{Z}$, we have the factorization

$$x^{p-1} - 1 = (x-1)(x-2)\dots(x-(p-1))$$

But

$$\begin{aligned} x^{p-1} - 1 &= x^{4k} - 1 \\ &= \left(x^4\right)^k - 1 \end{aligned}$$

But $\left(x^4\right)^k - 1$ is divisible by $x^4 - 1$ (why?) and hence by $x^2 + 1$, because

$$x^4 - 1 = (x^2+1)(x^2-1)$$

This means that two of the factors in the product

$$(x-1)(x-2)\dots(x-(p-1))$$

multiply to $x^2 + 1$, and hence they are roots of the equation $x^2 = -1$.

This is just the beginning. Now we know which primes are the sum of two squares. This is the same as knowing which primes factor in $\mathbb{Z}[i]$.

One of the most beautiful aspects of the structure of $\mathbb{Z}[i]$ is the classification of its primes. We've just seen that some ordinary primes like 5 are no longer primes in the Gaussian integers ($5 = (2+i)(2-i)$). Other primes like 7 stay prime even when you look at them in this larger setting. How can you tell how an ordinary prime will behave when you move up to $\mathbb{Z}[i]$. Remarkably, there's a simple test that can be carried out inside $\mathbb{Z}$ that will tell you exactly what happens. Here are the details:

Theorem 4.4

Every prime p *in* $\mathbb{Z}$ *does one of three things when you move up to* $\mathbb{Z}[i]$*:*

- *It can* split *into two (conjugate) factors:*

$$p = \mathfrak{p}\overline{\mathfrak{p}}$$

- *It can remain* inert*, staying prime in* $\mathbb{Z}[i]$*.*
- *It can* ramify *into the square of a prime in* $\mathbb{Z}[i]$*:*

$$p = \mathfrak{p}^2$$

Furthermore, we have the following decomposition law *to tell which is which:*

- p *splits* $\Leftrightarrow p \equiv 1 \pmod 4$
- p *is inert* $\Leftrightarrow p \equiv 3 \pmod 4$
- p *ramifies* $\Leftrightarrow p = 2$

This theorem was generalized independently in the 1920s by Artin and Takagi. They created a branch of number theory called "class field theory" that includes, for a very broad collection of number fields, equally simple laws of decomposition. The way a prime behaves as you move up a tower depends on its congruence class modulo a fixed (ideal) number called the *conductor*. So, the conductor of $\mathbb{Z}[i]$ is 4. There are more complicated kinds of decomposition possible if the degree of the extension is bigger than 2, but the $\mathbb{Z}$ to $\mathbb{Z}[i]$ story contains all the essential ingredients of the general case. For example, primes ramify if and only if they are factors of the conductor.

The proofs amount to tidying up the arguments so far. For example, 2 ramifies because (remember, we discount unit factors):

$$2 = -i(1+i)^2$$

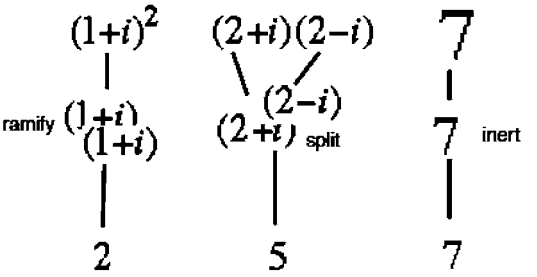
And there's more. As an exercise, prove that if the norm of a Gaussian integer $\mathfrak{p}$ is a prime in $\mathbb{Z}$, then $\mathfrak{p}$ is a prime Gaussian integer. More generally:

Theorem 4.5

Every prime $\mathfrak{p}$ *in* $\mathbb{Z}[i]$ *is one of three types:*

1. $N(\mathfrak{p}) = p$ *for some prime* $p \equiv 1 \pmod 4$
2. $\mathfrak{p} = q$ *(so* $N(\mathfrak{p}) = q^2$*) for some prime* $q \equiv 3 \pmod 4$
3. $\mathfrak{p}$ *is associate to* $1 + i$*.*

This classification of Gaussian primes is essential to the next part of our story.



Formal Algebra

Formal algebra is a wonderful bookkeeping mechanism and it often can be used to reduce complex-sounding combinatorial theorems to polynomial or power-series identities. Mathematicians in the century *before* last (especially Riemann and Dirichlet) invented a formalism that is perfect for counting things like the number of representations as the sum of two squares. Here's what they did:

A formal Dirichlet series is an expression of the form

$$\sum_{n=1}^{\infty} \frac{a(n)}{n^s}$$

where the $a(n)$ are complex numbers. Here, we are just thinking of s as an "indeterminate," a mark on the paper. So the series is a formal object that looks like this

$$\frac{a(1)}{1^s} + \frac{a(2)}{2^s} + \frac{a(3)}{3^s} + \dots$$

And we act on it via the usual rules for algebra.

The simplest of these is

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

Dirichlet series are added and multiplied formally. Algebra shows that if

$$\zeta(s) \sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \sum_{n=1}^{\infty} \frac{b(n)}{n^s}$$

then

$$b(n) = \sum_{d|n} a(d)$$

Let's state this as a theorem for later reference:

Theorem 4.6

Suppose

$$\zeta(s) \sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \left(\sum_{n=1}^{\infty} \frac{1}{n^s} \right) \left(\sum_{n=1}^{\infty} \frac{a(n)}{n^s} \right) = \sum_{n=1}^{\infty} \frac{b(n)}{n^s}$$

Then

$$b(n) = \sum_{d|n} a(d)$$

The proof is something best seen by working it out.

Sometimes, the coefficients $a(n)$ have interesting properties. For example, they might be "multiplicative:"

$$a(mn) = a(m)a(n)$$

When this happens, the Dirichlet series has an alternate form that shows its connection with arithmetic.

Theorem 4.7

Suppose we have a Dirichlet series

$$f(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s}$$

Suppose further that $a(mn) = a(m)a(n)$ *for all integers* m *and* n*. Then*

$$f(s) = \prod_{p} \left(\frac{1}{1 - \frac{a(p)}{p^s}} \right)$$

Where the product is over all prime numbers p*.*

Proof.

Each factor on the right side is a geometric series:

To be rigorous, we should put some restrictions on the values of $a(k)$ to ensure that the series converge.

$$\begin{aligned} \frac{1}{1 - \frac{a(p)}{p^s}} &= 1 + \left(\frac{a(p)}{p^s} \right) + \left(\frac{a(p)}{p^s} \right)^2 + \left(\frac{a(p)}{p^s} \right)^3 + \dots \\ &= 1 + \left(\frac{a(p)}{p^s} \right) + \left(\frac{a(p^2)}{p^{2s}} \right) + \left(\frac{a(p^3)}{p^{3s}} \right) + \dots \end{aligned}$$

■

Multiply all these together (one for every prime) and you get the sum of every possible expression of the form:

$$\frac{a(p_1^{e_1})a(p_2^{e_2}) \dots a(p_r^{e_r})}{p_1^{e_1 s} p_2^{e_2 s} \dots p_r^{e_r s}} = \frac{a(p_1^{e_1} p_2^{e_2} \dots p_r^{e_r})}{\left(p_1^{e_1} p_2^{e_2} \dots p_r^{e_r}\right)^s}$$

Since every $n \in \mathbb{Z}$ can be written in one and only one way as a product of powers of primes (the fundamental theorem again), this is the same as the sum

$$\sum_{n=1}^{\infty} \frac{a(n)}{n^s}$$

Example

Here's a multiplicative function that's connected to our work with Gaussian integers:

$$\chi(n) = \begin{cases} 1 & \text{if } n \equiv 1 \pmod 4 \\ -1 & \text{if } n \equiv 3 \pmod 4 \\ 0 & \text{if } n \text{ is even} \end{cases}$$

χ is called a "quadratic character."

You can check that χ is multiplicative. So,

$$\sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p \left(\frac{1}{1 - \frac{\chi(p)}{p^s}} \right)$$

Notice that, by theorem 4.6, if

$$\zeta(s) \prod_p \left(\frac{1}{1 - \frac{\chi(p)}{p^s}} \right) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s}$$

then

$$a(n) = \sum_{d|n} \chi(d)$$

So, $a(n)$ is the excess of the number of divisors of n of the form $4k+1$ over the number of divisors of n of the form $4k+3$. It seems strange that an integer always has at least as many divisors of the form $4k+1$ as it has divisors of the form $4k+3$?

The Result

We're ready for the main calculation. Remember s?

$$\begin{aligned} s(n) &= \text{the number of ways } n \text{ can be a sum of squares} \\ &= |\{\alpha \in Q_1 \mid N(\alpha) = n\}| \end{aligned}$$

Consider the Dirichlet series

$$s(s) = \sum_{n=1}^{\infty} \frac{s(n)}{n^s}$$

$$= \frac{1}{1^s} + \frac{1}{2^s} + \frac{0}{3^s} + \frac{1}{4^s} + \frac{2}{5^s} + \cdots + \frac{3}{25^s} + \ldots$$

Think of $s(n)$ as the number of Gaussian integers in Q_1 whose norm is n, so the $\frac{3}{25^s}$ comes from

$$\frac{1}{N(3+4i)} + \frac{1}{N(4+3i)} + \frac{1}{N(5+0i)}$$

It follows that

The right side is called the "Dedekind zeta function" for $\mathbb{Z}[i]$.

$$\sum_{n=1}^{\infty} \frac{s(n)}{n^s} = \sum_{\alpha \in Q_1} \frac{1}{(N(\alpha))^s}$$

$$= \prod_{\mathfrak{p} \in Q_1} \sum_{k=0}^{\infty} \frac{1}{(N(\mathfrak{p}))^s} \quad \text{(use the fundamental theorem in } \mathbb{Z}[i])$$

$$= \prod_{\mathfrak{p} \in Q_1} \frac{1}{1 - \frac{1}{N(\mathfrak{p})^s}} \quad \text{(sum a geometric series)}$$

Where the product is over all Gaussian primes in the first quadrant. This is another example of a calculation that is best understood by working it out.

Now we use theorem 4.5 on page 99. Every prime in Q_1 lies over one of these:

- 2. There's just one: $1 + i$, and $N(1 + i) = 2$
- a prime p that is 1 mod 4. There are two for each such p—if
$$p = \mathfrak{p}\,\overline{\mathfrak{p}}$$
then both $\mathfrak{p}$ and $\overline{\mathfrak{p}}$ have an associate in Q_1 (and they are different), and each has norm p.
- a prime p that is 3 mod 4. There's only one such prime, because such a p is inert, and $N(p) = p^2$.

So…

$$\prod_{\mathfrak{p}\in Q_1}\frac{1}{1-\frac{1}{N(\mathfrak{p})^s}} = \frac{1}{1-\frac{1}{2^s}}\left(\prod_{p\equiv 1 \pmod 4}\frac{1}{1-\frac{1}{p^s}}\right)^2\left(\prod_{p\equiv 3 \pmod 4}\frac{1}{1-\frac{1}{p^{2s}}}\right)$$

$$= \frac{1}{1-\frac{1}{2^s}}\left(\prod_{p\equiv 1 \pmod 4}\frac{1}{1-\frac{1}{p^s}}\right)^2\left(\prod_{p\equiv 3 \pmod 4}\frac{1}{1-\frac{1}{p^s}}\right)\left(\prod_{p\equiv 3 \pmod 4}\frac{1}{1+\frac{1}{p^s}}\right)$$

$$= \frac{1}{1-\frac{1}{2^s}}\left(\prod_{p \text{ odd}}\frac{1}{1-\frac{1}{p^s}}\right)\left(\prod_{p\equiv 1 \pmod 4}\frac{1}{1-\frac{1}{p^s}}\right)\left(\prod_{p\equiv 3 \pmod 4}\frac{1}{1+\frac{1}{p^s}}\right)$$

$$= \zeta(s)\left(\prod_{p\equiv 1 \pmod 4}\frac{1}{1-\frac{\chi(p)}{p^s}}\right)\left(\prod_{p\equiv 3 \pmod 4}\frac{1}{1-\frac{\chi(p)}{p^s}}\right)$$

$$= \zeta(s)\left(\prod_{p \text{ odd}}\frac{1}{1-\frac{\chi(p)}{p^s}}\right)$$

$$= \sum_{n=1}^{\infty}\frac{a(n)}{n^s}$$

where

$$a(n) = \sum_{d|n}\chi(d)$$

But (see page 103)

$$\sum_{n=1}^{\infty}\frac{s(n)}{n^s} = \prod_{\mathfrak{p}\in Q_1}\frac{1}{1-\frac{1}{N(\mathfrak{p})^s}}$$

so

$$\sum_{n=1}^{\infty}\frac{s(n)}{n^s} = \sum_{n=1}^{\infty}\frac{a(n)}{n^s}$$

and $s(n) = a(n)$. That's the punchline:

Theorem 4.8 (Fermat)
The number of representations of an integer n as the sum of two squares is the excess of the number of factors of n of the form $4k+1$ over the number of factors of n of the form $4k+3$.

Representing $s(n)$ as $\sum_{d|n}\chi(d)$ allows you to prove that a number n can be written as sum of squares $\Leftrightarrow$ every prime divisor of n of the form $4k+3$ shows up with even exponent. To prove that, you show that $n \mapsto \sum_{d|n}\chi(d)$ is

multiplicative, so you only have to look at it on prime powers. At a power of a prime that is 1 mod 4, this function is 1. At a power of a prime that is 3 mod 4, it is 0 or 1 depending on whether the power is odd or even.

Bibliography

[1] Andrews, G. *Number Theory*. Dover, New York, 1994.

[2] Cuoco, A. "Meta-Problems in Mathematics." College Mathematics Journal, 31, 2000.

[3] Cuoco, A. "Introducing Extensible Tools in Middle- and High-School Algebra." In C. Greenes (Ed.), *Algebra and Algebraic Thinking in School Mathematics, 70th Yearbook* (pp. 51-62). National Council of Teachers of Mathematics (NCTM), Reston, VA, 2008.

[4] EDC, "Pythagorans and Cousins," in *Ways to Think About Mathematics*, Corwin Press, 2004.

[5] EDC. *CME Project*. Pearson, Boston MA, 2009.

[6] Kerins, B. et. al. "Delving Deeper: Gauss, Pythagoras, and Heron." Mathematics Teacher, May 2003.

[7] Koblitz, N. *Introduction to Elliptic Curves and Modular Forms*, Springer Verlag, New York, 1993

[8] Ireland, K. and Rosen, M. *A Classical Introduction to Modern Number Theory* Springer-Verlag, 1991.