Benjamin McKay

Concrete Algebra

With a View Toward Abstract Algebra

November 8, 2019
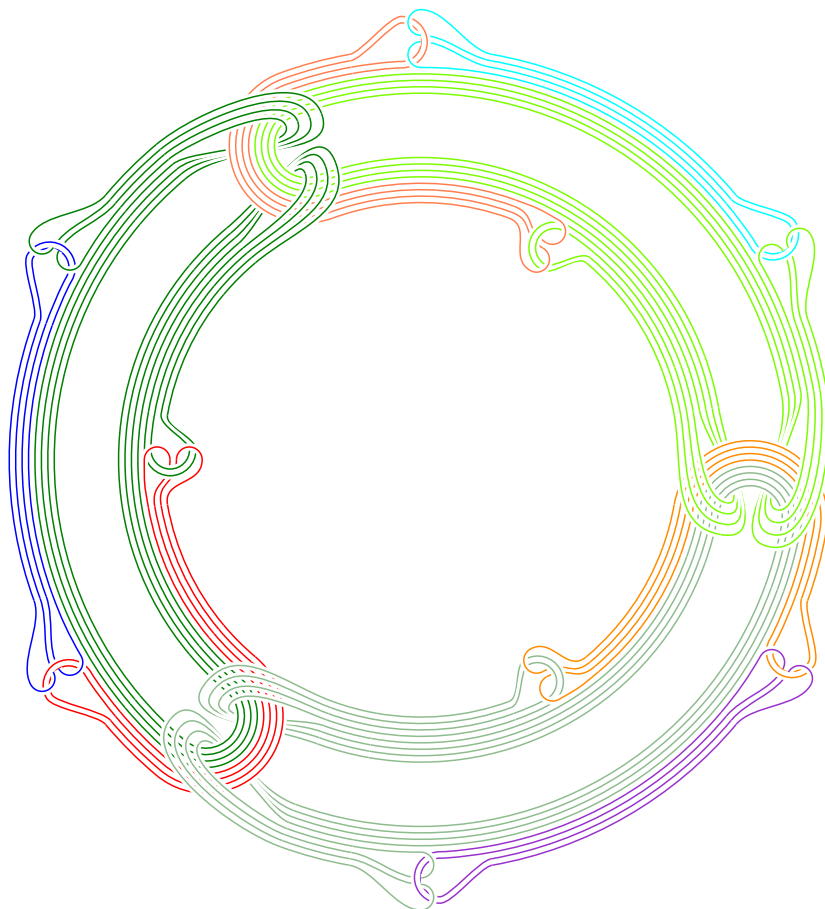
Preface

These notes are from lectures given in 2015 at University College Cork. They aim to explain the most concrete and fundamental aspects of algebra, in particular the algebra of the integers and of polynomial functions of a single variable, grounded by proofs using mathematical induction. It is impossible to learn mathematics by reading a book like you would read a novel; you have to work through exercises and calculate out examples. You should try all of the problems. More importantly, since the purpose of this class is to give you a deeper feeling for elementary mathematics, rather than rushing into advanced mathematics, you should reflect about how the simple ideas in this book reshape your vision of algebra. Consider how you can use your new perspective on elementary mathematics to help you some day guide other students, especially children, with surer footing than the teachers who guided you.

The temperature of Heaven can be rather accurately computed. Our authority is Isaiah 30:26, "Moreover, the light of the Moon shall be as the light of the Sun and the light of the Sun shall be sevenfold, as the light of seven days." Thus Heaven receives from the Moon as much radiation as we do from the Sun, and in addition $7 \times 7 = 49$ times as much as the Earth does from the Sun, or 50 times in all. The light we receive from the Moon is one $1/10\,000$ of the light we receive from the Sun, so we can ignore that. . . . The radiation falling on Heaven will heat it to the point where the heat lost by radiation is just equal to the heat received by radiation, i.e., Heaven loses 50 times as much heat as the Earth by radiation. Using the Stefan-Boltzmann law for radiation, (H/E) temperature of the earth $(\sim 300\,\mathrm{K})$, gives H as $798\,\mathrm{K}$ $(525\,°\mathrm{C})$. The exact temperature of Hell cannot be computed. . . . [However] Revelations 21:8 says "But the fearful, and unbelieving . . . shall have their part in the lake which burneth with fire and brimstone." A lake of molten brimstone means that its temperature must be at or below the boiling point, $444.6\,°\mathrm{C}$. We have, then, that Heaven, at $525\,°\mathrm{C}$ is hotter than Hell at $445\,°\mathrm{C}$.
>   — Applied Optics , vol. 11, A14, 1972

In these days the angel of topology and the devil of abstract algebra fight for the soul of every individual discipline of mathematics.
>   — Hermann Weyl
>   Invariants, Duke Mathematical Journal 5, 1939, 489–502

— and so who are you, after all?
— I am part of the power which forever wills evil and forever works good.
>   — Goethe
>   Faust

This Book is not to be doubted.
>   — Quran , 2:1/2:6-2:10 *The Cow*

Contents

<center>Chapter 1</center>

<center>## The integers</center>

<center>*God made the integers; all else is the work of man.*</center>

<center>— Leopold Kronecker</center>

Notation

We will write numbers using notation like $1\,234\,567.123\,45$, using a decimal point . at the last integer digit, and using thin spaces to separate out every 3 digits before or after the decimal point. You might prefer 1,234,567·123,45 or 1,234,567.123,45, which are also fine. We reserve the · symbol for multiplication, writing $2 \cdot 3 = 6$ rather than $2 \times 3 = 6$.

The laws of integer arithmetic

The integers are the numbers $\ldots, -2, -1, 0, 1, 2, \ldots$. Let us distill their essential properties, using only the concepts of addition and multiplication.

Addition laws:

    a. The associative law: For any integers a, b, c: $(a + b) + c = a + (b + c)$.

    b. The identity law: There is an integer 0 so that for any integer a: $a + 0 = a$.

    c. The existence of negatives: for any integer a: there is an integer b (denoted by the symbol $-a$) so that $a + b = 0$.

    d. The commutative law: For any integers a, b: $a + b = b + a$.

Multiplication laws:

    a. The associative law: For any integers a, b, c: $(ab)c = a(bc)$.

    b. The identity law: There is an integer 1 so that for any integer a: $a1 = a$.

    c. The zero divisors law: For any integers a, b: if $ab = 0$ then $a = 0$ or $b = 0$.

    d. The commutative law: For any integers a, b: $ab = ba$.

The distributive law:

    a. For any integers a, b, c: $a(b + c) = ab + ac$.

<center>1</center>

Sign laws:

Certain integers are called *positive*.

    a. The succession law: An integer b is positive just when either $b = 1$ or $b = c + 1$ for a positive integer c.

    b. Determinacy of sign: Every integer a has precisely one of the following properties: a is positive, $a = 0$, or $-a$ is positive.

We write $a < b$ to mean that there is a positive integer c so that $b = a + c$.

The law of well ordering:

    a. Any nonempty collection of positive integers has a least element; that is to say, an element a so that every element b satisfies $a < b$ or $a = b$.

All of the other arithmetic laws we are familiar with can be derived from these. For example, the associative law for addition, applied twice, shows that $(a+b)+(c+d) = a + (b + (c + d))$, and so on, so that we can add up any finite sequence of integers, in any order, and get the same result, which we write in this case as $a + b + c + d$. A similar story holds for multiplication.

Of course, we write $1 + 1$ as 2, and $1 + 1 + 1$ as 3 and so on. Write $a > b$ to mean $b < a$. Write $a \leq b$ to mean $a < b$ or $a = b$. Write $a \geq b$ to mean $b \leq a$. Write $|a|$ to mean a, if $a \geq 0$, and to mean $-a$ otherwise, and call it the *absolute value* of a. An integer a is *negative* if $-a$ is positive.

To understand mathematics, you have to solve a large number of problems.

> *I prayed for twenty years but received no answer until I prayed with my legs.*

> — Frederick Douglass, statesman and escaped slave

1.1 For each equation below, what law above justifies it?

    a. $7(3 + 1) = 7 \cdot 3 + 7 \cdot 1$

    b. $4(9 \cdot 2) = (4 \cdot 9)2$

    c. $2 \cdot 3 = 3 \cdot 2$

1.2 Use the laws above to prove that for any integers a, b, c: $(a + b)c = ac + bc$.

1.3 Use the laws above to prove that $0 + 0 = 0$.

1.4 Use the laws above to prove that $0 \cdot 0 = 0$.

1.5 Use the laws above to prove that, for any integer a: $a \cdot 0 = 0$.

1.6 Use the laws above to prove that, for any integer a: there is exactly one integer b so that $a + b = 0$; of course, we call this b by the name $-a$.

1.7 Use the laws above to prove that, for any integer a: $(-1)a = -a$.

1.8 Use the laws above to prove that $(-1)(-1) = 1$.

1.9 Use the laws above to prove that, for any integers b, c: $|bc| = |b||c|$.

1.10 Our laws ensure that there is an integer 0 so that $a + 0 = 0 + a = a$ for any integer a. Could there be two different integers, say p and q, so that $a + p = p + a = a$ for any integer a, and also so that $a + q = q + a = a$ for any integer a? (Roughly speaking, we are asking if there is more than one integer which can "play the role" of zero.)

1.11 Our laws ensure that there is an integer 1 so that $a \cdot 1 = 1 \cdot a = a$ for any integer a. Could there be two different integers, say p and q, so that $ap = pa = a$ for any integer a, and also so that $aq = qa = a$ for any integer a?

Theorem 1.1 (The equality cancellation law for addition). *Suppose that a, b and c are integers and that $a + c = b + c$. Then $a = b$.*

Proof. By the existence of negatives, there is a integer $-c$ so that $c + (-c) = 0$. Clearly $(a + c) + (-c) = (b + c) + (-c)$. Apply the associative law for addition: $a + (c + (-c)) = b + (c + (-c))$, so $a + 0 = b + 0$, so $a = b$. $\qquad\square$

1.12 Prove that $-(b + c) = (-b) + (-c)$ for any integers b, c.

1.13 A *predecessor* of an integer b is an integer c so that $b = c + 1$. Prove that every integer has a unique predecessor.

We haven't mentioned subtraction yet.

1.14 Suppose that a and b are integers. Prove that there is a unique integer c so that $a = b + c$. Of course, from now on we write this integer c as $a - b$.

1.15 Prove that, for any integers a, b, $a - b = a + (-b)$.

1.16 Prove that subtraction *distributes over multiplication*: for any integers a, b, c, $a(b - c) = ab - ac$.

1.17 Prove that any two integers b and c satisfy just precisely one of the conditions $b > c$, $b = c$, $b < c$.

Theorem 1.2. *The equality cancellation law for multiplication: for an integers a, b, c if $ab = ac$ and $a \neq 0$ then $b = c$.*

Proof. If $ab = ac$ then $ab - ac = ac - ac = 0$. But $ab - ac = a(b - c)$ by the distributive law. So $a(b - c) = 0$. By the zero divisors law, $a = 0$ or $b = c$. $\qquad\square$

1.18 We know how to add and multiply 2×2 matrices with integer entries. Of the various laws of addition and multiplication and signs for integers, which hold true also for such matrices?

1.19 Use the laws above to prove that the sum of any two positive integers is positive.

1.20 Use the laws above to prove that the product of any two positive integers is positive.

1.21 Use the laws above to prove that the product of any two integers is positive just when (1) both are positive or (2) both are negative.

1.22 Use the laws above to prove that the product of any two negative integers is positive.

1.23 Prove the inequality cancellation law for addition: For any integers a, b, c: if $a + c < b + c$ then $a < b$.

1.24 Prove the inequality cancellation law for multiplication: For any integers a, b, c: if $a < b$ and if $0 < c$ then $ac < bc$.

1.25 Suppose that S is a set of integers. A *lower bound* on S is an integer b so that $b \leq c$ for every integer c from S; if S has a lower bound, S is *bounded from below*. Prove that a nonempty set of integers bounded from below contains a least element.

Division of integers

> *Can you do Division? Divide a loaf by a knife—what's the answer to that?*
>
> — Lewis Carroll
>   THROUGH THE LOOKING GLASS

We haven't mentioned division yet. *Danger:* although 2 and 3 are integers,

$$\frac{3}{2} = 1.5$$

is *not* an integer.

1.26 Suppose that a and b are integers and that $b \neq 0$. Prove from the laws above that there is *at most* one integer c so that $a = bc$. Of course, from now on we write this integer c as $\frac{a}{b}$ or a/b.

1.27 *Danger:* why *can't* we divide by zero?

We already noted that $3/2$ is *not* an integer. At the moment, we are trying to work with integers only. An integer b *divides* an integer c if c/b is an integer; we also say that b is a *divisor* of c.

1.28 Explain why every integer divides into zero.

1.29 Prove that, for any two integers b and c, the following are equivalent:

a. b divides c,

b. $-b$ divides c,

c. b divides $-c$,

d. $-b$ divides $-c$,

e. $|b|$ divides $|c|$.

Proposition 1.3. *Take any two integers b and c. If b divides c, then $|b| < |c|$ or $b = c$ or $b = -c$.*

Proof. By the solution of the last problem, we can assume that b and c are positive. If $b = c$ the proposition holds, so suppose that $b > c$; write $b = c + k$ for some $k > 0$. Since b divides c, say $c = qb$ for some integer q. Since $b > 0$ and $c > 0$, $q > 0$ by problem 1.21 on the facing page. If $q = 1$ then $c = b$, so the proposition holds. If $q > 1$, then $q = n+1$ for some positive integer n. But then $c = qb = (n+1)(c+k) = nc+c+nk+k$. Subtract c from both sides: $0 = nc+nk+k$, a sum of positive integers, hence positive (by problem 1.20 on the preceding page) a contradiction. $\square$

Theorem 1.4 (Euclid). *Suppose that b and c are integers and $c \neq 0$. Then there are unique integers q and r (the quotient and remainder) so that $b = qc + r$ and so that $0 \le r < |c|$.*

Proof. To make our notation a little easier, we can assume (by perhaps changing the signs of c and q in our statement above) that $c > 0$.

Consider all integers of the form $b - qc$, for various integers q. If we were to take $q = -|b|$, then $b - qc = b + |b| + (c-1)|b| \ge 0$. By the law of well ordering, since there is an integer of the form $b - qc \ge 0$, there is a smallest integer of the form $b - qc \ge 0$; call it r. If $r \ge c$, we can replace $r = b - qc$ by $r - c = b - (q+1)c$, and so r was not smallest.

We have proven that we can find a quotient q and remainder r. We need to show that they are unique. Suppose that there is some other choice of integers Q and R so that $b = Qc + R$ and $0 \le R < c$. Taking the difference between $b = qc + r$ and $b = Qc + R$, we find $0 = (Q - q)c + (R - r)$. In particular, c divides into $R - r$. Switching the labels as to which ones are q, r and which are Q, R, we can assume that $R \ge r$. So then $0 \le r \le R < c$, so $R - r < c$. By proposition 1.3, since $0 \le R - r < c$ and c divides into $R - r$, we must have $R - r = 0$, so $r = R$. Plug into $0 = (Q - q)c + (R - r)$ to see that $Q = q$. $\square$

How do we actually find quotient and remainder, by hand? Long division:

$$
\begin{array}{r}
14 \\
17 \overline{)\,249} \\
17 \\
\overline{79} \\
68 \\
\overline{11}
\end{array}
$$

So if we start with $b = 249$ and $c = 17$, we carry out long division to find the quotient $q = 14$ and the remainder $r = 11$.

1.30 Find the quotient and remainder for b, c equal to:

a. $-180, 9$

b. $-169, 11$

c. $-982, -11$

> *d.* $279, -11$
>
> *e.* $247, -27$

The greatest common divisor

A *common divisor* of some integers is an integer which divides them all.

1.31 Given any collection of one or more integers, not all zero, prove that they have a *greatest common divisor*, i.e. a largest positive integer divisor.

Denote the greatest common divisor of integers $m_1, m_2, \ldots, m_n$ as

$$\gcd \{m_1, m_2, \ldots, m_n\}.$$

If the greatest common divisor of two integers is 1, they are *coprime*. We will also define $\gcd \{0, 0, \ldots, 0\} := 0$.

Lemma 1.5. *Take two integers b, c with $c \neq 0$ and compute the quotient q and remainder r, so that $b = qc + r$ and $0 \le r < |c|$. Then the greatest common divisor of b and c is the greatest common divisor of c and r.*

Proof. Any divisor of c and r divides the right hand of the equation $b = qc + r$, so it divides the left hand side, and so divides b and c. By the same reasoning, writing the same equation as $r = b - qc$, we see that any divisor of b and c divides c and r. $\square$

> This makes very fast work of finding the greatest common divisor by hand. For example, if $b = 249$ and $c = 17$ then we found that $q = 14$ and $r = 11$, so $\gcd \{249, 17\} = \gcd \{17, 11\}$. Repeat the process: taking 17 and dividing out 11, the quotient and remainder are $1, 6$, so $\gcd \{17, 11\} = \gcd \{11, 6\}$. Again repeat the process: the quotient and remainder for $11, 6$ are $1, 5$, so $\gcd \{11, 6\} = \gcd \{6, 5\}$. Again repeat the process: the quotient and remainder for $11, 6$ are $1, 5$, so $\gcd \{11, 6\} = \gcd \{6, 5\}$. In more detail, we divide the smaller integer (smaller in absolute value) into that the larger:
>
> $$
> \begin{array}{r}
> 14 \\
> 17 \overline{)\,249} \\
> 17 \\
> \hline
> 79 \\
> 68 \\
> \hline
> 11
> \end{array}
> $$
>
> Throw out the larger integer, 249, and replace it by the remainder, 11, and divide again:
>
> $$
> \begin{array}{r}
> 1 \\
> 11 \overline{)\,17} \\
> 11 \\
> \hline
> 6
> \end{array}
> $$
>
> Again we throw out the larger integer (in absolute value), 17, and replace

with the remainder, 6, and repeat:

$$6 \overline{)\begin{matrix} 1 \\ 11 \\ \underline{6} \\ 5 \end{matrix}}$$

and again:

$$5 \overline{)\begin{matrix} 1 \\ 6 \\ \underline{5} \\ 1 \end{matrix}}$$

and again:

$$1 \overline{)\begin{matrix} 5 \\ 5 \\ \underline{5} \\ 0 \end{matrix}}$$

The remainder is now zero. The greatest common divisor is therefore 1, the final nonzero remainder: $\gcd \{249, 17\} = 1$.

This method to find the greatest common divisor is called the *Euclidean algorithm*.

If the final nonzero integer is negative, just change its sign to get the greatest common divisor.

$$\gcd \{-4, -2\} = \gcd \{-2, 0\} = \gcd \{-2\} = 2.$$

1.32 Find the greatest common divisor of

   a. $4233, 884$

   b. $-191, 78$

   c. $253, 29$

   d. $84, 276$

   e. $-92, 876$

   f. $147, 637$

   g. $266\,664, 877\,769$

1.33* Take a nonempty collection of integers, at least one of which is not zero. Allow yourself to "build new integers" by adding or subtracting the ones you have from one another. Prove that the greatest common divisor of the collection is the smallest positive integer you can "build".

The least common multiple

The *least common multiple* of a finite collection of integers $m_1, m_2, \ldots, m_n$ is the smallest positive integer $\ell = \mathrm{lcm} \{m_1, m_2, \ldots, m_n\}$ so that all of the integers $m_1, m_2, \ldots, m_n$ divide ℓ.

Lemma 1.6. *The least common multiple of any two integers b, c (not both zero) is*

$$\operatorname{lcm}\{b, c\} = \frac{|bc|}{\gcd\{b, c\}}.$$

Proof. For simplicity, assume that $b > 0$ and $c > 0$; the cases of $b \le 0$ or $c \le 0$ are treated easily by flipping signs as needed and checking what happens when $b = 0$ or when $c = 0$ directly; we leave this to the reader. Let $d := \gcd\{b, c\}$, and factor $b = Bd$ and $c = Cd$. Then B and C are coprime. Write the least common multiple ℓ of b, c as either $\ell = b_1 b$ or as $\ell = c_1 c$, since it is a multiple of both b and c. So then $\ell = b_1 Bd = c_1 Cd$. Cancelling, $b_1 B = c_1 C$. So C divides $b_1 B$, but doesn't divide B, so divides b_1, say $b_1 = b_2 C$, so $\ell = b_1 Bd = b_2 CBd$. So BCd divides ℓ. So $(bc)/d = BCd$ divides ℓ, and is a multiple of b: $BCd = (Bd)C = bC$, and is a multiple of c: $BCd = B(Cd) = Bc$. But ℓ is the least such multiple. $\square$

Sage

> *Computers are useless. They can only give you answers.*
>
> — Pablo Picasso

These lecture notes include optional sections explaining the use of the sage computer algebra system. At the time these notes were written, instructions to install sage on a computer are at www.sagemath.org, but you should be able to try out sage on sagecell.sagemath.org or even create worksheets in sage on cocalc.com over the internet without installing anything on your computer. If we type

```
gcd(1200,1040)
```

and press *shift–enter*, we see the greatest common divisor: 80. Similarly type

```
lcm(1200,1040)
```

and press *shift–enter*, we see the least common multiple: 15600. Sage can carry out complicated arithmetic operations. The multiplication symbol is *:

```
12345*13579
```

gives 167632755. Sage uses the expression 2^3 to mean 2^3. You can invent variables in sage by just typing in names for them. For example, ending each line with *enter*, except the last which you end with *shift–enter* (when you want sage to compute results):

```
x=4
y=7
x*y
```

to print 28. The expression 15 % 4 means the remainder of 15 divided by 4, while 15//4 means the quotient of 15 divided by 4. We can calculate the greatest common divisor of several integers as

```
gcd([3800,7600,1900])
```

giving us 1900.

Mathematical induction

It is sometimes required to prove a theorem which shall be true whenever a certain quantity n which it involves shall be an integer or whole number and the method of proof is usually of the following kind. 1st. The theorem is proved to be true when n = 1. 2ndly. It is proved that if the theorem is true when n is a given whole number, it will be true if n is the next greater integer. Hence the theorem is true universally.

— George Boole

*So nat'ralists observe, a flea
Has smaller fleas that on him prey;
And these have smaller fleas to bite 'em.
And so proceeds Ad infinitum.*

— Jonathan Swift
ON POETRY: A RHAPSODY

It is often believed that everyone with red hair must have a red haired ancestor. But this principle is not true. We have all seen someone with red hair. According to the principle, she must have a red haired ancestor. By the same principle, he must have a red haired ancestor too, and so on. So the principle predicts an infinite chain of red haired ancestors. But there have only been a finite number of creatures (so the geologists tell us). So some red haired creature had no red haired ancestor.

We will see that
$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$
for any positive integer n. First, let's check that this is correct for a few values of n just to be careful. *Danger:* just to be *very* careful, we put $\stackrel{?}{=}$ between any two quantities when we are *checking* to see if they are equal; we are making clear that we don't yet know.

For $n = 1$, we get $1 \stackrel{?}{=} 1(1+1)/2$, which is true because the right hand side is $1(1+1)/2 = 2/2 = 1$.

For $n = 2$, we need to check $1 + 2 \stackrel{?}{=} 2(2+1)/2$. The left hand side of the equation is 3, while the right hand side is:

$$\frac{2(2+1)}{2} = \frac{2(3)}{2} = 3,$$

So we checked that they are equal.

For $n = 3$, we need to check that $1 + 2 + 3 \stackrel{?}{=} 3(3+1)/2$. The left hand side is 6, and you can check that the right hand side is $3(4)/2 = 6$ too. So we checked that they are equal.

But this process will just go on for ever and we will never finish checking all values of n. We want to check *all* values of n, all at once.

Picture a row of dominoes. If we can make the first domino topple, and we can make sure that each domino, if it topples, will make the next domino topple, then they will all topple.

Theorem 2.1 (The Principle of Mathematical Induction). *Take a collection of positive integers. Suppose that 1 belongs to this collection. Suppose that, whenever all positive integers less than a given positive integer n belong to the collection, then so does the integer n. (For example, if $1, 2, 3, 4, 5$ belong, then so does 6, and so on.) Then the collection consists precisely of all of the positive integers.*

Proof. Let S be the set of positive integers *not* belonging to that collection. If S is empty, then our collection contains all positive integers, so our theorem is correct. But what is S is *not* empty? By the law of well ordering, if S is not empty, then S has a least element, say n. So n is *not* in our collection. But being the least integer not in our collection, all integers $1, 2, \ldots, n-1$ less than n *are* in our collection. By hypothesis, n is also in our collection, a contradiction to our assumption that S is not empty. $\qquad\square$

Let's prove that $1 + 2 + 3 + \cdots + n = n(n+1)/2$ for any positive integer n. First, note that we have already checked this for $n = 1$ above. Imagine that we have checked our equation $1 + 2 + 3 + \cdots + n = n(n+1)/2$ for any positive integer n up to, but not including, some integer $n = k$. Now we want to check for $n = k$ whether it is still true: $1 + 2 + 3 + \cdots + n \stackrel{?}{=} n(n+1)/2$. Since we already know this for $n = k - 1$, we are allowed to write that out, without question marks:

$$1 + 2 + 3 + \cdots + (k-2) + (k-1) = (k-1)(k-1+1)/2.$$

Simplify:
$$1 + 2 + 3 + \cdots + k - 1 = (k-1)k/2.$$

Now add k to both sides. The left hand side becomes $1 + 2 + 3 + \cdots + k$. The

right hand side becomes $(k-1)k/2 + k$, which we simplify to

$$\frac{(k-1)k}{2} + k = \frac{(k-1)k}{2} + \frac{2k}{2},$$
$$= \frac{k^2 - k + 2k}{2},$$
$$= \frac{k^2 + k}{2},$$
$$= \frac{k(k+1)}{2}.$$

So we have found that, as we wanted to prove,

$$1 + 2 + 3 + \cdots + k = \frac{k(k+1)}{2}.$$

Note that we *used* mathematical induction to prove this: we prove the result for $n = 1$, and then suppose that we have proven it already for all values of n up to some given value, and then show that this will ensure it is still true for the next value.

The general pattern of induction arguments: we start with a statement we want to prove, which contains a variable, say n, representing a positive integer.

a. *The base case:* Prove your statement directly for $n = 1$.

b. *The induction hypothesis:* Assume that the statement is true for all positive integer values less than some value n.

c. *The induction step:* Prove that it is therefore also true for that value of n.

We can use induction in definitions, not just in proofs. We haven't made sense yet of exponents. (Exponents are often called *indices* in Irish secondary schools, but nowhere else in the world to my knowledge).

a. *The base case:* For any integer $a \neq 0$ we define $a^0 := 1$. *Watch:* We can write $:=$ instead of $=$ to mean that this is our *definition* of what a^0 *means*, not an equation we have somehow calculated out. For any integer a (including the possibility of $a = 0$) we also define $a^1 := a$.

b. *The induction hypothesis:* Suppose that we have defined already what

$$a^1, a^2, \ldots, a^b$$

means, for some positive integer b.

c. *The induction step:* We then define a^{b+1} to mean $a^{b+1} := a \cdot a^b$.

For example, by writing out

$$a^4 = a \cdot a^3,$$
$$= a \cdot a \cdot a^2,$$
$$= a \cdot a \cdot a \cdot a^1,$$
$$= \underbrace{a \cdot a \cdot a \cdot a}_{4 \text{ times}}.$$

Another example:
$$a^3 a^2 = (a \cdot a \cdot a)(a \cdot a) = a^5.$$
In an expression a^b, the quantity a is the *mantissa* or *base* and b is the *exponent*.

2.1 Use induction to prove that $a^{b+c} = a^b a^c$ for any integer a and for any positive integers b, c.

2.2 Use induction to prove that $\left(a^b\right)^c = a^{bc}$ for any integer a and for any positive integers b, c.

Sometimes we start induction at a value of n which is *not* at $n = 1$.

Let's prove, for any integer $n \geq 2$, that $2^{n+1} < 3^n$.

  a. *The base case:* First, we check that this is true for $n = 2$: $2^{2+1} < 3^2$? Simplify to see that this is $8 < 9$, which is clearly true.

  b. *The induction hypothesis:* Next, suppose that we know this is true for all values $n = 2, 3, 4, \ldots, k - 1$.

  c. *The induction step:* We need to check it for $n = k$: $2^{k+1} < 3^k$? How can we relate this to the values $n = 2, 3, 4, \ldots, k - 1$?

$$2^{k+1} = 2 \cdot 2^k,$$
$$< 3 \cdot 3^{k-1} \text{ by assumption,}$$
$$= 3^k.$$

We conclude that $2^{n+1} < 3^n$ for all integers $n \geq 2$.

2.3 All horses are the same colour; we can prove this by induction on the number of horses in a given set.

  a. *The base case:* If there's just one horse then it's the same colour as itself, so the base case is trivial.

  b. *The induction hypothesis:* Suppose that we have proven that, in any set of at most k horses, all of the horses are the same colour as one another, for any number $k = 1, 2, \ldots, n$.

  c. *The induction step:* Assume that there are n horses numbered 1 to n. By the induction hypothesis, horses 1 through $n - 1$ are the same color as one another. Similarly, horses 2 through n are the same color. But the middle horses, 2 through $n - 1$, can't change color when they're in different groups; these are horses, not chameleons. So horses 1 and n must be the same color as well.

Thus all n horses are the same color. What is wrong with this reasoning?

2.4 We could have assumed much less about the integers than the laws we gave in chapter 1. Using only the laws for addition and induction,

  a. Explain how to define multiplication of integers.

    b. Use your definition of multiplication to prove the associative law for multiplication.

    c. Use your definition of multiplication to prove the equality cancellation law for multiplication.

2.5 Suppose that x, b are positive integers. Prove that x can be written as
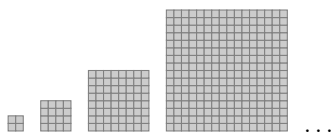
$$x = a_0 + a_1 b + a_2 b^2 + \cdots + a_k b^k$$

for unique integers $a_0, a_1, \ldots, a_k$ with $0 \leq a_i < b$. Hint: take quotient and remainder, and apply induction. (The sequence $a_k, a_{k-1}, \ldots, a_1, a_0$ is the sequence of digits of x in base b notation.)

2.6 Prove that for every positive integer n,

$$1^3 + 2^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2} \right)^2.$$

2.7 Picture a 2×2 grid, a 4×4 grid, an 8×8 grid, a 16×16 grid, and so on.

Fix a positive integer n. Show that it is possible to tile any $2^n \times 2^n$ grid, but with exactly one square removed, using 'L'-shaped tiles of three squares:

Recall that $\binom{n}{k}$ means the binomial coefficient, i.e. the coefficient of x^k (or of x^{n-k}) in the expansion of $(1+x)^n$ in powers of x:

$$(1 + x)^n = \binom{n}{0} + \binom{n}{1} x + \binom{n}{2} x^2 + \cdots + \binom{n}{n} x^n.$$

2.8 Prove that the binomial coefficients satisfy $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$. Explain how this gives Pascal's triangle:

$$1$$

$$1 \quad 1$$

$$1 \quad 2 \quad 1$$

$$1 \quad 3 \quad 3 \quad 1$$

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

$$1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1$$

$$1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1$$

$$1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1$$

$$1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1$$

$$1 \quad 9 \quad 36 \quad 84 \quad 126 \quad 126 \quad 84 \quad 36 \quad 9 \quad 1$$

$$1 \quad 10 \quad 45 \quad 120 \quad 210 \quad 252 \quad 210 \quad 120 \quad 45 \quad 10 \quad 1$$

$$1 \quad 11 \quad 55 \quad 165 \quad 330 \quad 462 \quad 462 \quad 330 \quad 165 \quad 55 \quad 11 \quad 1$$

$$1 \quad 12 \quad 66 \quad 220 \quad 495 \quad 792 \quad 924 \quad 792 \quad 495 \quad 220 \quad 66 \quad 12 \quad 1$$

$$1 \quad 13 \quad 78 \quad 286 \quad 715 \quad 1287 \quad 1716 \quad 1716 \quad 1287 \quad 715 \quad 286 \quad 78 \quad 13 \quad 1$$

$$1 \quad 14 \quad 91 \quad 364 \quad 1001 \quad 2002 \quad 3003 \quad 3432 \quad 3003 \quad 2002 \quad 1001 \quad 364 \quad 91 \quad 14 \quad 1$$

$$1 \quad 15 \quad 105 \quad 455 \quad 1365 \quad 3003 \quad 5005 \quad 6435 \quad 6435 \quad 5005 \quad 3003 \quad 1365 \quad 455 \quad 105 \quad 15 \quad 1$$

2.9 For each positive integer n, let

$$p_n(x) := 1 + x + \cdots + x^{n-1} = \sum_{k=0}^{n-1} x^k.$$

Prove that

$$p_n(1+x) = \sum_{k=0}^{n-1} \binom{n}{k+1} x^k.$$

Sage

> *A computer once beat me at chess, but it was no match for me at kick boxing.*
>
> — Emo Philips

To define a function in sage, you type `def`, to mean *define*, like:

```
def f(n):
    return 2*n+1
```

and press *shift–enter*. For any input n, it will return a value $f(n) = 2n + 1$. Careful about the * for multiplication. Use your function, for example, by typing

```
f(3)
```

and press *shift–enter* to get 7.

A function can depend on several inputs:

```
def g(x,y):
    return x*y+2
```

A *recursive function* is one whose value for some input depends on its value for other inputs. For example, the sum of the integers from 1 to n is:

```
def sum_of_integers(n):
    if n<=0:
        return 0
    else:
        return n+sum_of_integers(n-1)
```

giving us `sum_of_integers(5)` $= 15$.

2.10 The *Fibonnaci sequence* is $F(1) = 1$, $F(2) = 1$, and $F(n) = F(n-1) + F(n-2)$ for $n \geq 3$. Write out a function `F(n)` in sage to compute the Fibonacci sequence.

Lets check that $2^{n+1} < 3^n$ for some values of n, using a loop:

```
for n in [0..10]:
    print(n,3^n-2^(n+1))
```

This will try successively plugging in each value of n starting at $n = 0$ and going up to $n = 1$, $n = 2$, and so on up to $n = 10$, and print out the value of n and the value of $3^n - 2^{n+1}$. It prints out:

```
(0, -1)
(1, -1)
(2, 1)
(3, 11)
(4, 49)
(5, 179)
(6, 601)
(7, 1931)
(8, 6049)
(9, 18659)
(10, 57001)
```

giving the value of n and the value of the difference $3^n - 2^{n+1}$, which we can readily see is positive once $n \geq 2$. Tricks like this help us to check our induction proofs.

If we write `a=a+1` in sage, this means that the new value of the variable `a` is equal to one more than the old value that `a` had previously. Unlike mathematical variables, sage variables can change value over time. In sage the expression `a<>b` means $a \neq b$.

Although sage knows how to calculate the greatest common divisor of some integers, we can define own greatest common divisor function:

```
def my_gcd(a, b):
    while b<>0:
        a,b=b,a%b
    return a
```

We could also write a greatest divisor function that uses repeated subtraction, avoiding the quotient operation %:

```
def gcd_by_subtraction(a, b):
    if a<0:
        a=-a
    if b<0:
        b=-b
    if a==0:
        return b
    if b==0:
        return a
    while a<>b:
        if a>b:
            a=a-b
        else:
            b=b-a
    return a
```

Lists in sage

Sage manipulates lists. A list is a finite sequence of objects written down next to one another. The list L=[4,4,7,9] is a list called L which consists of four numbers: $4, 4, 7$ and 9, in that order. Note that the same entry can appear any number of times; in this example, the number 4 appears twice. You can think of a list as like a vector in $\mathbb{R}^n$, a list of numbers. When you "add" lists they concatenate: [6]+[2] yields $[6, 2]$. Sage uses notation L[0], L[1], and so on, to retrieve the elements of a list, instead of the usual vector notation. *Warning:* the element L[1] is *not* the first element. The length of a list L is len(L). To create a list $[0, 1, 2, 3, 4, 5]$ of successive integer, type range(0,6). Note that the number 6 here tells you to go up to, but not include, 6. So a L list of 6 elements has elements $L_0, L_1, \ldots, L_5$, denoted L[0], L[1], ..., L[5] in sage. To retrieve the list of elements of L from L_2 to L_4, type L[2:5]; again strangely this means up to but not including L_5.

For example, to reverse the elements of a list:

```
def reverse(L):
    n=len(L)
    if n<=1:
        return L
    else:
        return L[n-1:n]+reverse(L[0:n-1])
```

We create a list with 10 entries, print it:

```
L=range(0,10)
print(L)
```

yielding $[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]$ and then print its reversal:

```
print(reverse(L))
```

yielding $[9, 8, 7, 6, 5, 4, 3, 2, 1, 0]$.

We can construct a list of elements of the Fibonacci sequence:

```
def fibonacci_list(n):
    L=[1,1]
    for i in range(2,n):
        L=L+[L[i-1]+L[i-2]]
    return L
```

so that `fibonacci_list(10)` yields $[1, 1, 2, 3, 5, 8, 13, 21, 34, 55]$, a list of the first ten numbers in the Fibonacci sequence. The code works by first creating a list `L=[1,1]` with the first two entries in it, and then successively building up the list L by concatenating to it a list with one more element, the sum of the two previous list entries, for all values $i = 2, 3, \ldots, n$.

Greatest common divisors

The laws of nature are but the mathematical thoughts of God.

> — Euclid

Euclid alone has looked on Beauty bare.
Let all who prate of Beauty hold their peace,
And lay them prone upon the earth and cease
To ponder on themselves, the while they stare
At nothing, intricately drawn nowhere
In shapes of shifting lineage; let geese
Gabble and hiss, but heroes seek release
From dusty bondage into luminous air.
O blinding hour, O holy, terrible day,
When first the shaft into his vision shone
Of light anatomized! Euclid alone
Has looked on Beauty bare. Fortunate they
Who, though once only and then but far away,
Have heard her massive sandal set on stone.

> — Edna St. Vincent Millay
> Euclid Alone

The extended Euclidean algorithm

Theorem 3.1 (Bézout). *For any two integers b, c, not both zero, there are integers s, t so that $sb + tc = \gcd\{b, c\}$.*

These *Bézout coefficients* s, t play an essential role in advanced arithmetic, as we will see.

To find the Bézout coefficients of $12, 8$, write out a matrix

$$\begin{pmatrix} 1 & 0 & 12 \\ 0 & 1 & 8 \end{pmatrix}.$$

Repeatedly add some integer multiple of one row to the other, to try to make the bigger number in the last column (bigger by absolute value) get smaller

(by absolute value). In this case, we can subtract row 2 from row 1, to get rid of as much of the 12 as possible. All operations are carried out a whole row at a time:

$$\begin{pmatrix} 1 & -1 & 4 \\ 0 & 1 & 8 \end{pmatrix}$$

Now the bigger number (by absolute value) in the last column is the 8. We subtract as large an integer multiple of row 1 from row 2, in this case subtract $2\,(\text{row } 1)$ from row 2, to kill as much of the 8 as possible:

$$\begin{pmatrix} 1 & -1 & 4 \\ -2 & 3 & 0 \end{pmatrix}$$

Once we get a zero in the last column, the other entry in the last column is the greatest common divisor, and the entries in the same row as the greatest common divisor are the Bézout coefficients. In our example, the Bézout coefficients of $12, 8$ are $1, -1$ and the greatest common divisor is 4. This trick to calculate Bézout coefficients is the *extended Euclidean algorithm*.

Again, we have to be a little bit careful about minus signs. Find Bézout coefficients of $-8, -4$:

$$\begin{pmatrix} 1 & 0 & -8 \\ 0 & 1 & -4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & -4 \end{pmatrix},$$

the process stops here, and where we expect to get an equation $sb + tc = \gcd\{b, c\}$, instead we get

$$0(-8) + 1(-4) = -4,$$

which has the wrong sign, a negative greatest common divisor. So if the answer pops out a negative for the greatest common divisor, we change signs: $s = 0, t = -1, \gcd\{b, c\} = 4$.

Theorem 3.2. *The extended Euclidean algorithm calculates Bézout coefficients. In particular, Bézout coefficients s, t exist for any integers b, c with $c \neq 0$.*

Proof. At each step in the extended Euclidean algorithm, the third column proceeds exactly by the Euclidean algorithm, replacing the larger number (in absolute value) with the remainder by the division, except for perhaps a minus sign. So at the last step, the last nonzero number in the third column is the greatest common divisor, except for perhaps a minus sign.

Our extended Euclidean algorithm starts with

$$\begin{pmatrix} 1 & 0 & b \\ 0 & 1 & c \end{pmatrix},$$

a matrix which satisfies

$$\begin{pmatrix} 1 & 0 & b \\ 0 & 1 & c \end{pmatrix} \begin{pmatrix} b \\ c \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

It is easy to check that if some 2×3 matrix M satisfies

$$M \begin{pmatrix} b \\ c \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

then so does the matrix you get by adding any multiple of one row of M to the other row.

Eventually we get a zero in the final column, say

$$\begin{pmatrix} s & t & d \\ S & T & 0 \end{pmatrix},$$

or

$$\begin{pmatrix} s & t & 0 \\ S & T & D \end{pmatrix}.$$

It doesn't matter which since, at any step, we could always swap the two rows without changing the steps otherwise. So suppose we end up at

$$\begin{pmatrix} s & t & d \\ S & T & 0 \end{pmatrix}.$$

But

$$\begin{pmatrix} s & t & d \\ S & T & 0 \end{pmatrix} \begin{pmatrix} b \\ c \\ -1 \end{pmatrix} = \begin{pmatrix} sb + tc - d \\ 0 \end{pmatrix}.$$

This has to be zero, so: $sb + tc = d$. $\qquad\square$

Proposition 3.3. *Take two integers b, c, not both zero. The number $\gcd\{b, c\}$ is precisely the smallest positive integer which can be written as $sb + tc$ for some integers s, t.*

Proof. Let $d := \gcd\{b, c\}$. Since d divides into b, we divide it, say as $b = Bd$, for some integer B. Similarly, $c = Cd$ for some integer C. Imagine that we write down some positive integer as $sb + tc$ for some integers s, t. Then $sb + sc = sBd + tCd = (sB + tC)d$ is a multiple of d, so no smaller than d. $\qquad\square$

3.1 Find Bézout coefficients and greatest common divisors of

    a. $2468, 180$

    b. $79, -22$

    c. $45, 16$

    d. $-1000, 2002$

3.2 How would you find the greatest common divisor of several integers? The Bézout coefficients?

Sage

> *Computer science is no more about computers than astronomy is about telescopes.*
>
> — Edsger Dijkstra

The extended Euclidean algorithm is built in to sage: `xgcd(12,8)` returns $(4, 1, -1)$, the greatest common divisor and the Bézout coefficients, so that $4 = 1 \cdot 12 + (-1) \cdot 8$. We can write our own function to compute Bézout coefficients:

```
def bez(b,c):
    p,q,r=1,0,b
    s,t,u=0,1,c
    while u<>0:
        if abs(r)>abs(u):
            p,q,r,s,t,u=s,t,u,p,q,r
        Q=u//r
        s,t,u=s-Q*p,t-Q*q,u-Q*r
    return r,p,q
```

so that `bez(45,210)` returns a triple (g, s, t) where g is the greatest commmon divisor of $45, 210$, while s, t are the Bézout coefficients.

Chapter 4

Prime numbers

I think prime numbers are like life. They are very logical but you could never work out the rules, even if you spent all your time thinking about them.

— Mark Haddon
The Curious Incident of the Dog in the Night-Time

Prime numbers

A positive integer $p \geq 2$ is *prime* if the only positive integers which divide p are 1 and p.

4.1 Prove that 2 is prime and that 3 is prime and that 4 is not prime.

Lemma 4.1. *If b, c, d are integers and d divides bc and d is coprime to b then d divides c.*

Proof. Since d divides the product bc, we can write the product as $bc = qd$. Since d and b are coprime, their greatest common divisor is 1. So there are Bézout coefficients s, t for b, d, so $sb + td = 1$. So then $sbc + tdc = c$, i.e. $sqd + tdc = c$, or $d(sq + dc) = c$, i.e. d divides c. □

Corollary 4.2. *A prime number divides a product bc just when it divides one of the factors b or c.*

An expression of a positive integer as a product of primes, such as $12 = 2 \cdot 2 \cdot 3$, written down in increasing order, is a *prime factorization*.

Theorem 4.3. *Every positive integer n has a unique prime factorization.*

Proof. Danger: if we multiply together a collection of integers, we must insist that there can only be finitely many numbers in the collection, to make sense out of the multiplication.

More danger: an *empty* collection is always allowed in mathematics. We will simply say that if we have *no* numbers at all, then the product of those numbers (the product of no numbers at all) is defined to mean 1. This will be the right definition to use to make our theorems have the simplest expression. In particular, the integer 1 has the prime factorization consist of *no* primes.

First, let's show that there is a prime factorization, and then we will see that it is unique. It is clear that $1, 2, 3, \ldots, 12$ have prime factorizations, as in our table above.

$$2 = 2,$$
$$3 = 3,$$
$$4 = 2 \cdot 2,$$
$$5 = 5,$$
$$6 = 2 \cdot 3,$$
$$7 = 7,$$
$$8 = 2 \cdot 2 \cdot 2,$$
$$9 = 3 \cdot 3,$$
$$10 = 2 \cdot 5,$$
$$11 = 11,$$
$$12 = 2 \cdot 2 \cdot 3,$$
$$\vdots$$
$$2395800 = 2^3 \cdot 3^2 \cdot 5^2 \cdot 11^3,$$
$$\vdots$$

Suppose that all integers $1, 2, \ldots, n-1$ have prime factorizations. If n does not factor into smaller factors, then n is prime and $n = n$ is a factorization. Suppose that n factors, say into a product $n = bc$ of positive integers, neither equal to 1. Write down a prime factorization for b and then next to it one for c, giving a prime factorization for $n = bc$. So every positive integer has at least one prime factorization.

Let p be the smallest integer so that $p \geq 2$ and p divides n. Clearly p is prime. Since p divides the product of the primes in any prime factorization of n, p divides one of the primes in the factorization. But then p must equal one of these primes, and must be the smallest prime in the factorization. This determines the first prime in the factorization. So write $n = p n_1$ for a unique integer n_1, and by induction we can assume that n_1 has a unique prime factorization, and so n does as well.   □

Theorem 4.4 (Euclid). *There are infinitely many prime numbers.*

Proof. Write down a list of finitely many primes, say $p_1, p_2, \ldots, p_n$. Let $b := p_1 p_2 \ldots p_n$ and let $c := b + 1$. Then clearly b, c are coprime. So the prime decomposition of c has none of the primes $p_1, p_2, \ldots, p_n$ in it, and so must have some other primes distinct from these.   □

4.2 Write down the positive integers in a long list starting at 2:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, \ldots$$

Strike out all multiples of 2, except 2 itself:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, \ldots$$

Skip on to the next number which isn't struck out: 3. Strike out all multiples of 3, except 3 itself:

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, \ldots$$

Skip on to the next number which isn't struck out, and strike out all of its multiples except the number itself. Continue in this way forever. Prove that the remaining numbers, those which do *not* get struck out, are precisely the prime numbers. Use this to write down all prime numbers less than 120.

Greatest common divisors

We speed up calculation of the greatest common divisor by factoring out any obvious common factors: $\gcd\{46, 12\} = \gcd\{2 \cdot 23, 2 \cdot 6\} = 2 \gcd\{23, 6\} = 2$. For example, this works well if both numbers are even, or both are multiples of ten, or both are multiples of 5, or even if both are multiples of 3.

4.3 Prove that an integer is a multiple of 3 just when its digits sum up to a multiple of 3. Use this to see if 12345 is a multiple of 3. Use this to find $\gcd\{12345, 123456\}$.

4.4 An *even integer* is a multiple of 2; an *odd integer* is *not* a multiple of 2. Prove that if b, c are positive integers and

  a. b and c are both even then $\gcd\{b, c\} = 2 \gcd\{b/2, c/2\}$,

  b. b is even and c is odd, then $\gcd\{b, c\} = \gcd\{b/2, c\}$,

  c. b is odd and c is even, then $\gcd\{b, c\} = \gcd\{b, c/2\}$,

d. b and c are both odd, then gcd $\{b, c\} = $ gcd $\{b, b - c\}$.

Use this to find the greatest common divisor of $4864, 3458$ without using long division. Answer: gcd $\{4864, 3458\} = 19$.

4.5 Prove that any integer divides some integers just when it divides their greatest common divisor.

4.6 Prove that, for any positive integer m and integers b, c, not both zero, m gcd $\{b, c\} = $ gcd $\{mb, mc\}$.

Sage

> *Mathematicians have tried in vain to this day to discover some order in the sequence of prime numbers, and we have reason to believe that it is a mystery into which the human mind will never penetrate.*
>
> — Leonhard Euler

Sage can factor numbers into primes: `factor(1386)` gives $2 \cdot 3^2 \cdot 7 \cdot 11$. Sage can find the next prime number larger than a given number: `next_prime(2005)`= 2011. You can test if the number 4 is prime with `is_prime(4)`, which returns `False`. To list the primes between 14 and 100, type `prime_range(14,100)` to see

```
[17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89,
97]
```

The number of primes less than or equal to a real number x is traditionally denoted $\pi(x)$ (which has nothing to do with the π of trigonometry). In sage, $\pi(x)$ is denoted `prime_pi(x)`. For example, $\pi(10^6) = 78498$ is calculated as `prime_pi(10^6)`. The prime number theorem says that $\pi(x)$ gets more closely approximated by

$$\frac{x}{-1 + \log x}$$

as x gets large. We can plot $\pi(x)$, and compare it to that approximation.

```
p=plot(prime_pi, 0, 10000, rgbcolor='red')
q=plot(x/(-1+log(x)), 5, 10000, rgbcolor='blue')
show(p+q)
```

which displays:
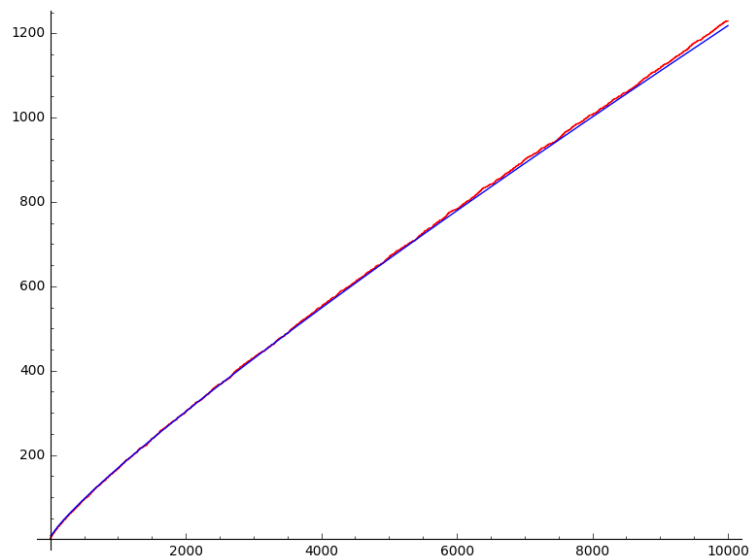
4.7 Write code in sage to find the greatest common divisor of two integers, using the ideas of problem 4.4 on page 24.

Chapter 5

Modular arithmetic

Mathematics is the queen of the sciences and number theory is the queen of mathematics.

— Carl Friedrich Gauss

Definition

If we divide integers by 7, the possible remainders are

$$0, 1, 2, 3, 4, 5, 6.$$

For example, $65 = 9 \cdot 7 + 2$, so the remainder is 2. Two integers are *congruent mod* 7 if they have the same remainder modulo 7. So 65 and 2 are congruent mod 7, because $65 = 9 \cdot 7 + 2$ and $2 = 0 \cdot 7 + 2$. We denote congruence modulo 7 as

$$65 \equiv 2 \pmod{7}.$$

Sometimes we will allow ourselves a sloppy notation, where we write the remainder of 65 modulo 7 as $\overline{65}$. This is sloppy because the notation doesn't remind us that we are working out remainders modulo 7. If we change 7 to some other number, we could get confused by this notation. We will often compute remainders modulo some chosen number, say m, instead of 7.

If we add multiples of 7 to an integer, we don't change its remainder modulo 7:

$$\overline{65 + 7} = \overline{65}.$$

Similarly,

$$\overline{65 - 7} = \overline{65}.$$

If we add, or multiply, some numbers, what happens to their remainders?

Theorem 5.1. *Take a positive integer m and integers a, A, b, B. If $a \equiv A \pmod{m}$ and $b \equiv B \pmod{m}$ then*

$$a + b \equiv A + B \pmod{m},$$
$$a - b \equiv A - B \pmod{m},$$
$$\text{and } ab \equiv AB \pmod{m}.$$

27

The bar notation is more elegant. If we agree that $\bar{a}$ means remainder of a modulo the fixed choice of integer m, we can write this as: if

$$\bar{a} = \bar{A} \text{ and } \bar{b} = \bar{B}$$

then

$$\overline{a+b} = \overline{A+B},$$
$$\overline{a-b} = \overline{A-B}, \text{ and}$$
$$\overline{ab} = \overline{AB}.$$

Note that $9 \equiv 2 \pmod 7$ and $4 \equiv 11 \pmod 7$, so our theorem tells us that $9 \cdot 4 \equiv 2 \cdot 11 \pmod 7$. Let's check this. The left hand side:

$$9 \cdot 4 = 36,$$
$$= 5(7) + 1,$$

so that $9 \cdot 4 \equiv 1 \pmod 7$. The right hand side:

$$2 \cdot 11 = 22,$$
$$= 3(7) + 1,$$

so that $2 \cdot 11 \equiv 1 \pmod 7$.

So it works in this example. Let's prove that it always works.

Proof. Since $a - A$ is a multiple of m, as is $b - B$, note that

$$(a+b) - (A+B) = (a-A) + (b-B)$$

is also a multiple of m, so addition works. In the same way

$$ab - AB = (a-A)b + A(b-B)$$

is a multiple of m. $\qquad\square$

5.1 Prove by induction that every "perfect square", i.e. integer of the form n^2, has remainder $0, 1$ or 4 modulo 8.

5.2 Take an integer like 243098 and write out the sum of its digits $2+4+3+0+9+8$. Explain why every integer is congruent to the sum of its digits, modulo 9.

5.3 Take integers a, b, m with $m \neq 0$. Let $d := \gcd\{a, b, m\}$. Prove that $a \equiv b \pmod m$ just when $a/d \equiv b/d \pmod{m/d}$.

Arithmetic of remainders

We now *define* an addition law on the numbers $0, 1, 2, 3, 4, 5, 6$ by declaring that when we add these numbers, we then take remainder modulo 7. This is *not* usual addition. To make this clear, we write the remainders with bars over them, always. For example, we are saying that in this addition law $\bar{3} + \bar{5}$ means $\overline{3+5} = \bar{8} = \bar{1}$,

since we are working modulo 7. We adopt the same rule for subtraction, and for multiplication. For example, modulo 13,

$$\left(\bar{7}+\bar{9}\right)\left(\overline{11}+\bar{6}\right)=\overline{16}\cdot\overline{17},$$
$$=\overline{13+3}\cdot\overline{13+4},$$
$$=\bar{3}\cdot\bar{4},$$
$$=\overline{12}.$$

If we are daring, we might just drop all of the bars, and state clearly that we are calculating modulo some integer. In our daring notation, modulo 17,

$$16\cdot 29 - 7\cdot 5 = 16\cdot 12 - 7\cdot 5,$$
$$= 192 - 35,$$
$$= (11\cdot 17 + 5) - (2\cdot 17 + 1),$$
$$= 5 - 1,$$
$$= 4.$$

5.4 Expand and simplify $\bar{5}\cdot\bar{2}\cdot\left(\bar{6}-\bar{9}\right)$ modulo 7.

The addition and multiplication tables for remainder modulo 5:

+	0	1	2	3	4		·	0	1	2	3	4
0	1	2	3	4	0		0	0	0	0	0	0
1	2	3	4	0	1		1	0	1	2	3	4
2	3	4	0	1	2		2	0	2	4	1	3
3	4	0	1	2	3		3	0	3	1	4	2
4	0	1	2	3	4		4	0	4	3	2	1

5.5 Describe laws of modular arithmetic, imitating the laws of integer arithmetic. If we work modulo 4, explain why the zero divisors law fails. Why are there no sign laws?

5.6 Compute the remainder when dividing 19 into 37^{200}.

5.7 Compute the last two digits of 9^{2000}.

5.8 Prove that the equation $a^2 + b^2 = 3c^2$ has no solutions in nonzero integers a, b and c. Hint: start by proving that modulo 4, $a^2 = 0$ or 1. Then consider the equation modulo 4; show that a, b and c are divisible by 2. Then each of a^2, b^2 and c^2 has a factor of 4. Divide through by 4 to show that there would be a smaller set of solutions to the original equation. Apply induction.

To carry a remainder to a huge power, say 2^{2005} modulo 13, we can build up the power out of smaller ones. For example, $2^2 = 4$ modulo 13, and therefore

modulo 13,

$$2^4 = \left(2^2\right)^2,$$
$$= 4^2,$$
$$= 16,$$
$$= 3.$$

Keeping track of these partial results as we go, modulo 13,

$$2^8 = \left(2^4\right)^2,$$
$$= 3^2,$$
$$= 9.$$

We get higher and higher powers of 2: modulo 13,

k	2^k	$2^{2^k} \bmod 13$
0	1	2
1	2	4
2	4	3
3	8	9
4	16	$9^2 = 81 = 3$
5	32	$3^2 = 9$
6	64	$9^2 = 3$
7	128	$3^2 = 9$
8	256	$9^2 = 3$
9	512	$3^2 = 9$
10	1024	$9^2 = 3$
11	2048	

The last row gets into 2^{2048}, too large to be relevant to our problem. We now want to write out exponent 2005 as a sum of powers of 2, by first dividing in 1024:

$$2005 = 1024 + 981$$

and then dividing in the next power of 2 we can fit into the remainder,

$$= 1024 + 512 + 469,$$
$$= 1024 + 512 + 256 + 128 + 64 + 16 + 4 + 1.$$

Then we can compute out modulo 13:

$$2^{2005} = 2^{1024+512+256+128+64+16+4+1},$$
$$= 2^{1024}2^{512}2^{256}2^{128}2^{64}2^{16}2^{4}2^{1},$$
$$= 3 \cdot 9 \cdot 3 \cdot 9 \cdot 3 \cdot 3 \cdot 3 \cdot 2,$$
$$= (3 \cdot 9)^3 \cdot 2,$$
$$= 27^2 \cdot 2,$$
$$= 1^2 \cdot 2,$$
$$= 2.$$

5.9 Compute 2^{100} modulo 125.

Reciprocals

Every nonzero rational number b/c has a reciprocal: c/b. Since we now have modular arithmetic defined, we want to know which remainders have "reciprocals". Working modulo some positive integer, say that a remainder x has a reciprocal $y = x^{-1}$ if $xy = 1$. (It seems just a little too weird to write it as $y = 1/x$, but you can if you like.) Reciprocals are also called *multiplicative inverses*.

Modulo 7

$$\bar{1} \cdot \bar{1} = \bar{1},$$
$$\bar{2} \cdot \bar{4} = \bar{1},$$
$$\bar{3} \cdot \bar{5} = \bar{1},$$
$$\bar{4} \cdot \bar{2} = \bar{1},$$
$$\bar{5} \cdot \bar{3} = \bar{1},$$
$$\bar{6} \cdot \bar{6} = \bar{1}.$$

So in this weird type of arithmetic, we can allow ourselves the freedom to write these equations as identifying a reciprocal.

$$\bar{1}^{-1} = \bar{1},$$
$$\bar{2}^{-1} = \bar{4},$$
$$\bar{3}^{-1} = \bar{5},$$
$$\bar{4}^{-1} = \bar{2},$$
$$\bar{5}^{-1} = \bar{3},$$
$$\bar{6}^{-1} = \bar{6}.$$

A remainder that has a reciprocal is a *unit*.

5.10 *Danger:* If we work modulo 4, then prove that $\bar{2}$ has *no* reciprocal. Hint: $\bar{2}^2 = 0$.

5.11 Prove that, modulo any integer m, $(m-1)^{-1} = m - 1$, and that modulo m^2, $(m-1)^{-1} = m^2 - m - 1$.

Theorem 5.2. *Take a positive integer m and a remainder $\bar{r}$ modulo m. Take the Bézout coefficients of r and m: $sr + tm = d$, so that d is the greatest common divisor of r and m.*

*   If $d = 1$ then $\bar{r}^{-1} = \bar{s}$ modulo m. If $d \neq 1$ then $\bar{r}^{-1}$ does not exist. In particular, in the remainders modulo m, a remainder $\bar{r}$ is a unit just when r, m are coprime integers.*

Proof. If r, m are coprime integers, so their greatest common divisor is 1, then write Bézout coefficients $sr + tm = 1$, and quotient by m:

$$\bar{s}\bar{r} = \bar{1}.$$

On the other hand, if

$$\bar{s}\bar{r} = \bar{1},$$

then sr is congruent to 1 modulo m, i.e. there is some quotient q so that $sr = qm + 1$, so $sr - qm = 1$, giving Bézout coefficients $s = s, t = -q$, so the greatest common divisor of r, m is 1. $\qquad\square$

Working modulo 163, let's compute 14^{-1}. First we carry out the long division

$$
\begin{array}{r}
11 \\
14 \overline{)163} \\
14 \\
\hline
23 \\
14 \\
\hline
9
\end{array}
$$

Now let's start looking for Bézout coefficients, by writing out matrix:

$$\begin{pmatrix} 1 & 0 & 14 \\ 0 & 1 & 163 \end{pmatrix}$$

and then add $-11 \cdot$ row 1 to row 2:

$$\begin{pmatrix} 1 & 0 & 14 \\ -11 & 1 & 9 \end{pmatrix}.$$

Add $-$row 2 to row 1:

$$\begin{pmatrix} 12 & -1 & 5 \\ -11 & 1 & 9 \end{pmatrix}.$$

Add $-$row 1 to row 2:

$$\begin{pmatrix} 12 & -1 & 5 \\ -23 & 2 & 4 \end{pmatrix}.$$

Add $-$row 2 to row 1:

$$\begin{pmatrix} 35 & -3 & 1 \\ -23 & 2 & 4 \end{pmatrix}.$$

Add $-4 \cdot$ row 1 to row 2:

$$\begin{pmatrix} 35 & -3 & 1 \\ -163 & 14 & 0 \end{pmatrix}.$$

Summing it all up: $35 \cdot 14 + (-3) \cdot 163 = 1$. Quotient out by 163: modulo 163, $35 \cdot 14 = 1$, so modulo 163, $14^{-1} = 35$.

5.12 Use this method to find reciprocals:

  a. 13^{-1} modulo 59

  b. 10^{-1} modulo 11

  c. 2^{-1} modulo 193.

  d. 6003722857^{-1} modulo 77695236973.

5.13 Suppose that b, c are remainders modulo a prime. Prove that $bc = 0$ just when either $b = 0$ or $c = 0$.

5.14 Suppose that p is a prime number and n is an integer with $n < p$. Explain why, modulo p, the numbers $0, n, 2n, 3n, \ldots, (p-1)n$ consist in precisely the remainders $0, 1, 2, \ldots, p-1$, in some order. (Hint: use the reciprocal of n.) Next, since every nonzero remainder has a reciprocal remainder, explain why the product of the nonzero remainders is 1. Use this to explain why

$$n^{p-1} \equiv 1 \pmod{p}.$$

Finally, explain why, for any integer k,

$$k^p \equiv k \pmod{p}.$$

The Chinese remainder theorem

> *An old woman goes to market and a horse steps on her basket and crushes the eggs. The rider offers to pay for the damages and asks her how many eggs she had brought. She does not remember the exact number, but when she had taken them out two at a time, there was one egg left. The same happened when she picked them out three, four, five, and six at a time, but when she took them seven at a time, there were none left. What is the smallest number of eggs she could have had?*
>
> — Brahmagupta (580CE–670CE)
> BRAHMA-SPHUTA-SIDDHANTA (BRAHMA'S CORRECT SYSTEM)

Take a look at some numbers and their remainders modulo 3 and 5:

n	$n \bmod 3$	$n \bmod 5$
0	0	0
1	1	1
2	2	2
3	0	3
4	1	4
5	2	0
6	0	1
7	1	2
8	2	3
9	0	4
10	1	0
11	2	1
12	0	2
13	1	3
14	2	4
15	0	0

Remainders modulo 3 repeat every 3, and remainders modulo 5 repeat every 5, but the pair of remainders modulo 3 and 5 together repeat every 15.

How can you find an integer x so that

$$
\begin{aligned}
x &\equiv 1 \pmod 3, \\
x &\equiv 2 \pmod 4, \\
x &\equiv 1 \pmod 7?
\end{aligned}
$$

A recipe: how to find an unknown integer x given only the knowledge of its remainders modulo various integers. Suppose we know x has remainder r_1 modulo m_1, r_2 modulo m_2, and so on. Let

$$
m := m_1 m_2 \ldots m_n.
$$

For each i, let

$$
u_i := \frac{m}{m_i} = m_1 m_2 \ldots m_{i-1} m_{i+1} m_{i+2} \ldots m_n.
$$

Each u_i has some reciprocal modulo m_i, given as the remainder of some integer v_i. Let

$$
x := r_1 u_1 v_1 + r_2 u_2 v_2 + \cdots + r_n u_n v_n.
$$

We can simplify this a little: add or subtract multiples of m until x is in the range $0 \le x \le m - 1$.

Theorem 5.3. *Take some positive integers $m_1, m_2, \ldots, m_n$, coprime to one another. Suppose that we want to find an unknown integer x, given only the knowledge of*

its remainders modulo $m_1, m_2, \ldots, m_n$; so we know its remainder modulo m_1 is r_1, modulo m_2 is r_2, and so on. There is such an integer x, given by the recipe above, and x is unique modulo

$$m_1 m_2 \ldots m_n.$$

Proof. All of $m_1, m_2, \ldots, m_n$ divide u_i, except m_i. So if $j \neq i$ then modulo m_j, $u_i = 0$. All of the other m_j are coprime to m_i, so their product is coprime to m_i, so has a reciprocal modulo m_i. Thus modulo m_i, $u_i \neq 0$. Then modulo m_i, $x = r_i$, and so on. So the recipe gives the required integer x. If there are two, then their difference vanishes modulo all of the m_i, so is a multiple of every one of the m_i, which are coprime, so is a multiple of their product m. □

Let's find an integer x so that

$$x \equiv 1 \pmod{3},$$
$$x \equiv 2 \pmod{4},$$
$$x \equiv 1 \pmod{7}.$$

So in this problem we have to work modulo $(m_1, m_2, m_3) = (3, 4, 7)$, and get remainders $(r_1, r_2, r_3) = (1, 2, 1)$. First, no matter what the remainders, we have to work out the reciprocal mod each m_i of the product of all of the other m_j. So let's reduce these products down to their remainders:

$$4 \cdot 7 = 28 = 9 \cdot 3 + 1 \equiv 1 \pmod{3},$$
$$3 \cdot 7 = 21 = 5 \cdot 4 + 1 \equiv 1 \pmod{4},$$
$$3 \cdot 4 = 12 = 1 \cdot 7 + 5 \equiv 5 \pmod{7}.$$

We need the reciprocals of these, which, to save ink, we just write down for you without writing out the calculations:

$$1^{-1} \equiv 1 \pmod{3},$$
$$1^{-1} \equiv 1 \pmod{4},$$
$$5^{-1} \equiv 3 \pmod{7}.$$

(You can easily check those.) Finally, we add up remainder times product times reciprocal:

$$x = r_1 \cdot 4 \cdot 7 \cdot 1 + r_2 \cdot 3 \cdot 7 \cdot 1 + r_3 \cdot 3 \cdot 4 \cdot 3,$$
$$= 1 \cdot 4 \cdot 7 \cdot 1 + 2 \cdot 3 \cdot 7 \cdot 1 + 1 \cdot 3 \cdot 4 \cdot 3,$$
$$= 106.$$

We can now check to be sure:

$$106 \equiv 1 \pmod{3},$$
$$106 \equiv 2 \pmod{4},$$
$$106 \equiv 1 \pmod{7}.$$

The Chinese remainder theorem tells us also that 106 is the unique solution modulo $3 \cdot 4 \cdot 7 = 84$. But then $106 - 84 = 22$ is also a solution, the smallest positive solution.

5.15 Solve the problem about eggs. Hint: ignore the information about eggs taken out six at a time.

5.16 Use the Chinese remainder theorem to determine the smallest number of soldiers possible in Han Xin's army if the following facts are true. When they parade in rows of three soldiers, two soldiers will be left. When they parade in rows of five, three will be left, and in rows of seven, two will be left.

5.17 Use the Chinese remainder theorem to determine the smallest number of soldiers possible in Han Xin's army if the following facts are true. When they parade in rows of three soldiers, one soldier will be left. When they parade in rows of seven, two will be left, and in rows of 19, three will be left.

5.18 Use the Chinese remainder theorem to find the smallest positive integer x so that

$$x = 2 \quad (\mathrm{mod}\ 22),$$
$$x = 8 \quad (\mathrm{mod}\ 39).$$

Bird's eye view of the Chinese remainder theorem

Given some integers $m_1, m_2, \ldots, m_n$, we consider sequences $(b_1, b_2, \ldots, b_n)$ consisting of remainders: b_1 a remainder modulo m_1, and so on. Add sequences of remainders in the obvious way:

$$(b_1, b_2, \ldots, b_n) + (c_1, c_2, \ldots, c_n) = (b_1 + c_1, b_2 + c_2, \ldots, b_n + c_n).$$

Similarly, we can subtract and multiply sequences of remainders:

$$(b_1, b_2, \ldots, b_n)(c_1, c_2, \ldots, c_n) = (b_1 c_1, b_2 c_2, \ldots, b_n c_n),$$

by multiplying remainders as usual, modulo the various $m_1, m_2, \ldots, m_n$.

> Modulo $(3, 5)$, we multiply
> $$(2, 4)(3, 2) = (2 \cdot 3, 4 \cdot 2),$$
> $$= (6, 8),$$
> $$= (0, 3).$$

Let $m := m_1 m_2 \ldots m_n$. To each remainder modulo m, say b, associate its remainder b_1 modulo m_1, b_2 modulo m_2, and so on. Associate the sequence $\vec{b} := (b_1, b_2, \ldots, b_n)$ of all of those remainders. In this way we make a map taking each remainder b modulo m to its sequence $\vec{b}$ of remainders modulo all of the various m_i. Moreover, $\overrightarrow{b + c} = \vec{b} + \vec{c}$ and $\overrightarrow{b - c} = \vec{b} - \vec{c}$ and $\overrightarrow{bc} = \vec{b}\vec{c}$, since each of these works when we take remainder modulo anything.

> Take m_1, m_2, m_3 to be $3, 4, 7$. Then $m = 3 \cdot 4 \cdot 7 = 84$. If $b = 8$ modulo 84,

then

$$b_1 = 8 \mod 3,$$
$$= 2,$$
$$b_2 = 8 \mod 4,$$
$$= 0,$$
$$b_3 = 8 \mod 7,$$
$$= 1,$$
$$\vec{b} = (b_1, b_2, b_3) = (2, 0, 1).$$

Corollary 5.4. *Take some positive integers $m_1, m_2, \ldots, m_n$, so that any two of them are coprime. The map taking b to $\vec{b}$, from remainders modulo m to sequences of remainders modulo $m_1, m_2, \ldots, m_n$, is one-to-one and onto, identifies sums with sums, products with products, differences with differences, units with sequences of units.*

Euler's totient function

Euler's totient function ϕ assigns to each positive integer $m = 2, 3, \ldots$ the number of all remainders modulo m which are units (in other words, coprime to m) [in other words, which have reciprocals]. It is convenient to *define* $\phi(1) := 1$ (even though there *isn't* actually 1 unit remainder modulo 1).

5.19 Explain by examining the remainders that the first few values of ϕ are

m	$\phi(m)$
1	1
2	1
3	2
4	2
5	4
6	2
7	6

5.20 Prove that a positive integer $m \geq 2$ is prime just when $\phi(m) = m - 1$.

Theorem 5.5. *Suppose that $m \geq 2$ is an integer with prime factorizaton*

$$m = p_1^{a_1} p_2^{a_2} \ldots p_n^{a_n},$$

so that $p_1, p_2, \ldots, p_n$ are prime numbers and $a_1, a_2, \ldots, a_n$ are positive integers. Then

$$\phi(m) = \left(p_1^{a_1} - p_1^{a_1-1}\right)\left(p_2^{a_2} - p_2^{a_2-1}\right)\ldots\left(p_n^{a_n} - p_n^{a_n-1}\right).$$

Proof. If m is prime, this follows from problem 5.20.

Suppose that m has just one prime factor, or in other words that $m = p^a$ for some prime number p and integer a. It is tricky to count the remainders coprime to m, but

easier to count those not coprime, i.e. those which have a factor of p. Clearly these are the multiples of p between 0 and $p^a - p$, so the numbers pj for $0 \le j \le p^{a-1} - 1$. So there are p^{a-1} such remainders. We take these out and we are left with $p^a - p^{a-1}$ remainders left, i.e. coprime.

If b, c are coprime integers ≥ 2, then corollary maps units modulo bc to pairs of a unit modulo b and a unit modulo c, and is one-to-one and onto. Therefore counting units: $\phi(bc) = \phi(b)\phi(c)$. $\qquad\qquad\square$

Theorem 5.6 (Euler)**.** *For any positive integer m and any integer b coprime to m,*

$$b^{\phi(m)} \equiv 1 \pmod{m}.$$

Proof. Working with remainders modulo m, we have to prove that for any unit remainder b, $b^{\phi(m)} = 1$ modulo m.

Let U be the set of all units modulo m, so U is a subset of the remainders $0, 1, 2, \ldots, m-1$. The product of units is a unit, since it has a reciprocal (the product of the reciprocals). Therefore the map

$$u \in U \mapsto bu \in U$$

is defined. It has an inverse:

$$u \in U \mapsto b^{-1}u \in U,$$

where b^{-1} is the reciprocal of b. Writing out the elements of U, say as

$$u_1, u_2, \ldots, u_q$$

note that $q = \phi(m)$. Then multiplying by b scrambles these units into a different order:

$$bu_1, bu_2, \ldots, bu_q.$$

But multiplying by b just scrambles the order of the roots, so if we multiply them all, and then scramble them back into order:

$$(bu_1)(bu_2)\ldots(bu_q) = u_1 u_2 \ldots u_q.$$

Divide every unit $u_1, u_2, \ldots, u_q$ out of boths sides to find $b^q = 1$. $\qquad\square$

5.21 Take an integer $m \ge 2$. Suppose that b is a unit in the remainders modulo m. Prove that the reciprocal of b is $b^{\phi(m)-1}$.

Euler's theorem is important because we can use it to calculate quickly modulo prime numbers (and sometimes even modulo numbers which are not prime).

Modulo 19, let's find $123456789^{987654321}$. First, the base of this expression is $123456789 = 6497725 \cdot 19 + 14$ So modulo 19:

$$123456789^{987654321} = 14^{987654321}.$$

That helps with the base, but the exponent is still large. According to Euler's theorem, since 19 is prime, modulo 19:

$$b^{19-1} = 1$$

for any remainder b. In particular, modulo 19,

$$14^{18} = 1.$$

So every time we get rid of 18 copies of 14 multiplied together, we don't change our result. Divide 18 into the exponent:

$$987654321 = 54869684 \cdot 18 + 9.$$

So then modulo 19:

$$14^{987654321} = 14^9.$$

We leave the reader to check that $14^9 = 18$ modulo 19, so that finally, modulo 19,

$$123456789^{987654321} = 18.$$

5.22 By hand, using Euler's totient function, compute 127^{162} modulo 120.

Lemma 5.7. *For any prime number p and integers b and k, $b^{1+k(p-1)} = b$ modulo p.*

Proof. If b *is* a multiple of p then $b = 0$ modulo p so both sides are zero. If b is not a multiple of p then $b^{p-1} = 1$ modulo p, by Euler's theorem. Take both sides to the power k and multiply by b to get the result. $\square$

Theorem 5.8. *Suppose that $m = p_1 p_2 \ldots p_n$ is a product of distinct prime numbers. Then for any integers b and k with $k \geq 0$,*

$$b^{1+k\phi(m)} \equiv b \pmod{m}.$$

Proof. By lemma 5.7, the result is true if m is prime. So if we take two prime numbers p and q, then $b^{1+k(p-1)(q-1)} = b$ modulo p, but also modulo q, and therefore modulo pq. The same trick works if we start throwing in more distinct prime factors into m. $\square$

5.23 Give an example of integers b and m with $m \geq 2$ for which $b^{1+\phi(m)} \neq b$ modulo m.

Sage

In Sage, the quotient of 71 modulo 13 is `mod(71,13)`. The tricky bit: it returns a "remainder modulo 13", so if we write

```
a=mod(71,13)
```

this will define a to be a "remainder modulo 13". The value of a is then 6, but the value of a^2 is 10, because the result is again calculated modulo 13.

Euler's totient function is

```
euler_phi(777)
```

which yields $\phi(777) = 432$. To find 14^{-1} modulo 19,

```
inverse_mod(14,19)
```

yields $14^{-1} = 15$ modulo 19.

We can write our own version of Euler's totient function, just to see how it might look:

```
def phi(n):
    return prod(p^a-p^(a-1) for (p,a) in factor(n))
```

where `prod` means product, so that `phi(666)` yields $\phi(666) = 216$. The code here uses the function `factor()`, which takes an integer n and returns a list `p=factor(n)` of its prime factors. In our case, the prime factorisation of 666 is $666 = 2 \cdot 3^2 \cdot 37$. The expression `p=factor(n)` when $n = 666$ yields a list `p=[(2,1), (3,2), (37,1)]`, a list of the prime factors together with their powers. To find these values, `p[0]` yields `(2,1)`, `p[1]` yields `(3,2)`, and `p[2]` yields `(37,1)`. The expression `len(p)` gives the length of the list `p`, which is the number of entries in that list, in this case 3. For each entry, we set $b = p_i$ and $e = a_i$ and then multiply the result r by $b^e - b^{e-1}$.

To use the Chinese remainder theorem, suppose we want to find a number x so that x has remainders 1 mod 3 and 2 mod 7,

```
crt([1,2],[3,7])
```

gives you $x = 16$; you first list the two remainders 1,2 and then list the moduli 3,7.

Another way to work with modular arithmetic in sage: we can create an object which represents the remainder of 9 modulo 17:

```
a=mod(9,17)
a^(-1)
```

yielding 2. As long as all of our remainders are modulo the same number 17, we can do arithmetic directly on them:

```
b=mod(7,17)
a*b
```

yields 12.

Chapter 6

Secret messages

And when at last you find someone to whom you feel you can pour out your soul, you stop in shock at the words you utter— they are so rusty, so ugly, so meaningless and feeble from being kept in the small cramped dark inside you so long.

> — Sylvia Plath
> The Unabridged Journals of Sylvia Plath

Man is not what he thinks he is, he is what he hides.

> — André Malraux
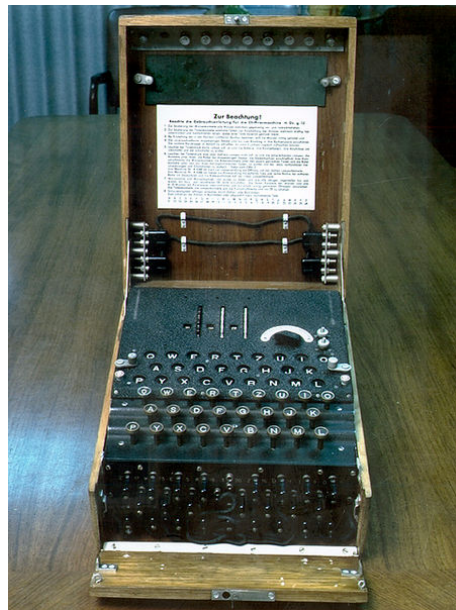
Enigma, Nazi secret code machine

RSA: the Cocks, Rivest, Shamir and Adleman algorithm

Alice wants to send a message to Bob, but she doesn't want Eve to read it.
She first writes the message down in a computer, as a collection of 0's and
1's. She can think of these 0's and 1's as binary digits of some large integer x,
or as binary digits of some *remainder* x modulo some large integer m. Alice
takes her message x and turns it into a secret coded message by sending Bob
not the original x, but instead sending him x^d modulo m, for some integer
d. This will scramble up the digits of x unrecognizably, if d is chosen "at
random". For random enough d, and suitably chosen positive integer m, Bob
can unscramble the digits of x^d to find x.

For example, take $m := 55$ and let $d := 17$ and $y := x^{17} \bmod 55$:

x	y	x	y	x	y	x	y
0	0	14	9	28	8	42	37
1	1	15	5	29	39	43	43
2	7	16	36	30	35	44	44
3	53	17	52	31	26	45	45
4	49	18	28	32	32	46	51
5	25	19	24	33	33	47	42
6	41	20	15	34	34	48	38
7	17	21	21	35	40	49	14
8	13	22	22	36	31	50	30
9	4	23	23	37	27	51	6
10	10	24	29	38	3	52	2
11	11	25	20	39	19	53	48
12	12	26	16	40	50	54	54
13	18	27	47	41	46	55	0

If Alice sends Bob the secret message $y = x^{17}$, then Bob decodes it by $x = y^{33}$,
as we will see.

Theorem 6.1. *Pick two different prime numbers p and q and let $m := pq$. Recall
Euler's totient function $\phi(m)$. Suppose that d and e are positive integers so that $de = 1$
modulo $\phi(m)$. If Alice maps each remainder x to $y = x^d$ modulo m, then Bob can
invert this map by taking each remainder y to $x = y^e$ modulo m.*

Proof. We have to prove that $\left(x^d\right)^e = x$ modulo m for all x, i.e. that $x^{de} = x$ modulo
m for all x. For x coprime to m, this follows from Euler's theorem, but for other
values of x the result is not obvious.

By theorem 5.5 on page 37, $\phi(m) = (p-1)(q-1)$. Since $de = 1$ modulo $\phi(m)$,
clearly
$$de - 1 = k(p-1)(q-1)$$

for some integer k. If $k = 0$, then $de = 1$ and the result is clear: $x^1 = x$. Since d and e are positive integers, $de - 1$ is not negative, so $k \geq 0$. So we can assume that $k > 0$. The result follows from theorem 5.8 on page 39. $\qquad\square$

6.1 Take each letter in the alphabet, ordered

$$abc \ldots zABC \ldots Z,$$

and one "blank space" letter _, so $26 + 26 + 1 = 53$ letters in all. Use the rule that each letter is then represented by a number from 100 to 152, starting with $a \mapsto 100$, $b \mapsto 101$, and so on to _ $\mapsto 152$. Then string the digits of these numbers together into a single number. For example, ab_c is $100\,101\,152\,102$.

    a. Write out the message *Hail_Ceasar* as a number.

    b. Translate the number $127\,114\,114\,111\,104$ into letters by this encoding.

    c. Apply the method above, taking

$$p = 993\,319,$$
$$q = 999\,331,$$
$$d = 13$$

and use a computer to compute the secret code on the number $x = 127\,114\,114\,111\,104$. You should get:
$$y = 202\,002\,770\,451$$

    d. Use a computer to check that the associated number e for the given numbers p, q, d in this case is
$$e = 839\,936\,711\,257.$$

How can you use a computer to find this number e, if you only know the numbers p, q, d in this case?

    e. Use the process described in theorem 6.1 on the preceding page to decode the message
$$y = 660\,968\,731\,660.$$

Warning: you might want to factor e first and then compute the power y^e modulo m by computing one factor of e at a time.

Sage

Sage shines because we can't do any of the computations of this chapter by hand.

Alice

To encode a single letter as a number, sage has a function `ord()`:

```
ord('A')
```

yields 65, giving each letter (and punctuation sign and space, etc.) a different positive integer less than 100. (If Alice wants to use lower case letters, we need to go beyond 100, so let's not do that.) To apply this to an entire string of text,

```
m = "HELLO WORLD"
m = map(ord, m)
m
```

yields $[72, 69, 76, 76, 79, 32, 87, 79, 82, 76, 68]$. Alice turns this list of numbers into a single number, by first reversing order:

```
m.reverse()
m
```

yields $[68, 76, 82, 79, 87, 32, 79, 76, 76, 69, 72]$. Then the expression

```
ZZ(m,100)
```

adds up successively these numbers, multiplying by 100 at each step, so that

```
x=ZZ(m,100)
x
```

yields x. This is the integer Alice wants to encode. Alice sets up the values of her primes and exponent:

```
p=993319
q=999331
d=13
m=p*q
f=euler_phi(m)
e=inverse_mod(d,f)
```

The secret code Alice sends is $y = x^d \pmod{m}$. Sage can compute powers in modular arithmetic quickly (using the same tricks we learned previously):

```
y=power_mod(x, d, m)
y
```

yielding 782102149108. This is Alice's encoded message. She can send it to Bob openly: over the phone, by email, or on a billboard.

Bob

Bob decodes by

```
x=power_mod(y, e, m)
x
```

yielding 53800826153. (Alice has secretly whispered e and m to Bob.) Bob turns this number x back into text by using the function chr(), which is the inverse of ord():

```
def recover_message(x):
    if x==0:
        return ""
    else:
        return recover_message(x//100)+chr(x%100)
```

Bob applies this:

    print(recover_message(x))

to yield HELLO WORLD.

Chapter 7

Rational, real and complex numbers

A fragment of the Rhind papyrus, 1650BCE, containing ancient Egyptian calculations using rational numbers

There is a remote tribe that worships the number zero. Is nothing sacred?

— Les Dawson

Rational numbers

A *rational number* is a ratio of two integers, like

$$\frac{2}{3}, \quad \frac{-7}{9}, \quad \frac{22}{-11}, \quad \frac{0}{2}.$$

We also write them $2/3, -7/9, 22/-11, 0/2$. In writing $\frac{2}{3}$, the number 2 is the *numerator*, and 3 is the *denominator*. We can think of rational numbers two different ways:

   a. Geometry: Divide up an interval of length b into c equal pieces. Each piece has length b/c.

   b. Algebra: a rational number b/c is just a formal expression we write down, with two integers b and c, in which $c \neq 0$. We agree that any rational number $\frac{b}{c}$ is equal to $\frac{ab}{ac}$ for any integer $a \neq 0$.

Of course, the two ideas give the same objects, but in these notes, the algebra approach is best. For example, we see that $22/11 = 14/7$, because I can factor out:

$$\frac{22}{11} = \frac{11 \cdot 2}{11 \cdot 1},$$

and factor out

$$\frac{14}{7} = \frac{7 \cdot 2}{7 \cdot 1}.$$

Write any rational number $\frac{b}{1}$ as simply b, and in this way see the integers sitting among the rational numbers.

   More generally, any rational number b/c can be simplified by dividing b and c by their greatest common divisor, and then changing the sign of b and of c if needed, to ensure that $c > 0$; the resulting rational number is *in lowest terms*.

7.1 Bring these rational numbers to lowest terms: $224/82$, $324/-72$, $-1000/8800$.

   Multiply rational numbers by

$$\frac{b}{c}\frac{B}{C} = \frac{bB}{cC}.$$

Similarly, divide a rational number by a *nonzero* rational number as

$$\frac{\frac{b}{c}}{\frac{B}{C}} = \frac{b}{c}\frac{C}{B}.$$

Add rationals with the *same denominator* as:

$$\frac{b}{d} + \frac{c}{d} = \frac{b+c}{d}.$$

The same for subtraction:

$$\frac{b}{d} - \frac{c}{d} = \frac{b-c}{d}.$$

But if the denominators don't match up, we rescale to make them match up:

$$\frac{b}{c} + \frac{B}{C} = \frac{bC}{cC} + \frac{cB}{cC}.$$

7.2 If we replace $\frac{b}{c}$ by $\frac{ab}{ac}$ and we replace $\frac{B}{C}$ by $\frac{AB}{AC}$, for some nonzero integers a, A, prove that the result of computing out $\frac{b}{c} + \frac{B}{C}$, $\frac{b}{c} - \frac{B}{C}$ or $\frac{b}{c}\frac{B}{C}$ only changes by scaling both numerator and denominator by the same nonzero integer. Hence the arithmetic operations don't contradict our agreement that we declare $\frac{b}{c}$ to equal $\frac{ab}{ac}$.

7.3 By hand, showing your work, simplify

$$\frac{2}{3} - \frac{1}{2}, \frac{324}{49} \cdot \frac{392}{81}, \frac{4}{5} + \frac{7}{4}.$$

Lemma 7.1. *The number $\sqrt{2}$ is irrational. (To be more precise, no rational number squares to 2.)*

Proof. If it is rational, say $\sqrt{2} = \frac{b}{c}$, then $c\sqrt{2} = b$, so squaring gives $c^2 \cdot 2 = b^2$. Every prime factor in b occurs twice in b^2, so the prime factorization of the right hand side has an even number of factors of each prime. But the left hand side has an odd number of factors of 2, since c^2 has an even number, but there is another factor of 2 on the left hand side. This contradicts uniqueness of prime factorization. $\square$

7.4 Suppose that N is a positive integer. Prove that either $\sqrt{N}$ is irrational or N is a *perfect square*, i.e. $N = n^2$ for some integer n.

7.5 Prove that there are no rational numbers x and y satisfying $\sqrt{3} = x + y\sqrt{2}$.

A rational number is *positive* if it is a ratio of positive integers. We write $x > y$ to mean that $x - y$ is positive, and similarly define $x < y$, $x \le y$, and so on.

7.6 Prove that every positive rational number has a unique expression in the form

$$\frac{p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}}{q_1^{\beta_1} q_2^{\beta_2} \cdots q_\ell^{\beta_\ell}}$$

where

$$p_1, p_2, \ldots, p_k, q_1, q_2, \ldots, q_\ell$$

are prime numbers, all different, with $p_1 < p_2 < \cdots < p_k$ and $q_1 < q_2 < \cdots < q_\ell$ and

$$\alpha_1, \alpha_2, \ldots, \alpha_k, \beta_1, \beta_2, \ldots, \beta_\ell$$

are positive integers. For example

$$\frac{234}{16} = \frac{3^2 \cdot 13}{2^3}.$$

7.7 Of the various laws of addition and multiplication and signs for integers, which hold true also for rational numbers?

The most important property that rational numbers have, that integers don't, is that we can divide any rational number by any nonzero rational number.

Real numbers

> *Order is an exotic in Ireland. It has been imported from England but will not grow. It suits neither soil nor climate.*
>
> — J.A. Froude

Think of the rational numbers geometrically, so that b/c is the length of each piece when a rod of length b is cut into c equal lengths. Draw an straight line, infinite in both directions, and mark a point 0 on it. Then mark at point 1 at 1 unit of distance from 0, and so on marking all of the integers to form a *number line*. The points of this continuous number line are the *real numbers*. This physical definition doesn't make it possible to prove anything about the real numbers.

The rational numbers are insufficient for algebra: they have no $\sqrt{2}$. Imagine taking all positive rational numbers b/c for which $b^2/c^2 > 2$. Thinking geometrically, drawing a number line:

$$\sqrt{2}$$

Our rational numbers sit along the line getting very close to the circled point of the line, which should be the point $\sqrt{2}$. But there is no rational number there. From now on, we assume that the reader is familiar with the real numbers and their basic properties; see [12] for the complete story of real numbers.

Complex numbers

> *Entre deux vérités du domaine réel, le chemin le plus facile et le plus court passe bien souvent par le domaine complexe.*
>
> The shortest and easiest path between any two facts about the real domain often passes through the complex domain.
>
> — Paul Painlevé
> Analyse des travaux scientifiques

Just as the rational numbers have a deficiency, no $\sqrt{2}$, so the real numbers have a deficiency: no $\sqrt{-1}$. We can fix this by introducing the complex numbers. Here are two definitions:

a. *Algebra:* We agree to play a game, where we write down algebraic expressions in real numbers and in an abstract variable i, but we treat the symbol i according to the following rule. Whenever we see two copies of i multiplied together (or we see i^2) appear anywhere, we are allowed to wipe them out and replace with -1, and vice versa when we see a -1 we can replace it with an i^2. We force by hand the existence of a $\sqrt{-1}$ by forcing the abstract symbol i to be $\sqrt{-1}$.

b. *Geometry:* We draw the usual x, y axes on the plane. We pretend that each point of the plane represents a single number, which we write as $z = x + yi$ to represent the point (x, y). If we have two points, $z = x + yi$ and $Z = X + Yi$, we add them by the rule:

$$z + Z = (x + X) + (y + Y)i$$

and multiply them by the rule:

$$zZ = (xX - yY) + (xY + yX)i.$$

For example,

$$(2 + 4i)(3 - i) = (2 \cdot 3 + 4 \cdot 1) + (-2 + 4 \cdot 3)i = 10 + 10i.$$

It isn't at all clear that these two games we can play arrive at the same result. Note that the rule for multiplication seems to come out of the air in the geometric theory, but in the algebraic theory it is just what you get by expanding out both sides and insisting that $i^2 = -1$. Algebra is in this respect clearly superior. But the geometric approach makes very precise what sort of objects we are really dealing with, so we will use it as our definition.

Just like with rational numbers, if a complex number $z = x + yi$ is not zero, or in other words x and y are not both zero, then it has a reciprocal

$$\frac{1}{z} = \frac{x - yi}{x^2 + y^2}.$$

7.8 Check that this is a reciprocal. Why is it defined?

Complex conjugation is the operation taking $z = x + yi$ to $\bar{z} := x - iy$.

7.9 Prove that a complex number z is a real number just when $z = \bar{z}$.

7.10 Prove that addition, subtraction, multiplication and division of complex numbers passes through conjugation:

$$\overline{z + w} = \bar{z} + \bar{w},$$
$$\overline{z - w} = \bar{z} - \bar{w},$$
$$\overline{zw} = \bar{z}\,\bar{w},$$
$$\overline{\left(\frac{z}{w}\right)} = \frac{\bar{z}}{\bar{w}},$$

for any complex numbers z and w. Explain by induction why, for any polynomial $p(z)$ with real number coefficients,

$$\overline{p(z)} = p(\bar{z}),$$

for any complex number z.

The *norm* or *modulus* of a complex number $z = x + yi$ is the real number $|z| := \sqrt{x^2 + y^2}$.

7.11 Prove that $z\bar{z} = |z|$ for any complex number z.

Sage

> *Was ist also Wahrheit? Ein bewegliches Heer von Metaphern, Metonymien, Anthropomorphismen, kurz eine Summe von menschlichen Relationen, die, poetisch und rhetorisch gesteigert, übertragen, geschmückt wurden, und die nach langem Gebrauch einem Volke fest, kanonisch und verbindlich dünken: die Wahrheiten sind Illusionen, von denen man vergessen hat, daßsie welche sind, Metaphern, die abgenutzt und sinnlich kraftlos geworden sind, Münzen, die ihr Bild verloren haben und nun als Metall, nicht mehr als Münzen, in Betracht kommen.*

> What, then, is truth? A mobile army of metaphors, metonyms, and anthropomorphisms—in short, a sum of human relations, which have been enhanced, transposed, and embellished poetically and rhetorically, and which after long use seem firm, canonical, and obligatory to a people: truths are illusions about which one has forgotten that this is what they are; metaphors which are worn out and without sensuous power; coins which have lost their pictures and now matter only as metal, no longer as coins.
>
> — Friedrich Nietzsche
> Über Wahrheit und Lüge im aussermoralischen Sinne

Sage represents numbers in two different ways: as exact symbolic expressions, or as a finite decimal approximation (often called a *numeric* or *floating point* representation). For example we compute $\frac{7}{2} + \frac{5}{3}$ symbolically as

    7/2+5/3

yielding $\frac{31}{6}$. We can also play with irrational numbers symbolically: represent $\sqrt{27}$ as

    sqrt(27)

which sage simplifies to yield $3\sqrt{3}$.

We will *never* use decimal approximations in this class (please don't), but here is how you can use them if you wish. By writing out a number with a decimal point, we are asking sage to approximate numerically, keeping only some number of decimals accuracy:

    sqrt(27.0)

yields 5.19615242270663. If we want to ask sage to give us a particular number of decimal places accuracy in a decimal approximation of a symbolic expression, say 50 decimals:

    numerical_approx(pi, digits=50)

yields
$$3.1415926535897932384626433832795028841971693993751.$$

To use complex numbers, try

    z=3+4*i
    real_part(z)

yielding 3, the real part, while `imag_part(z)` yields 4.

Chapter 8

Polynomials

I was x years old in the year x^2.

— Augustus De Morgan (when asked his age)

A *polynomial* in a variable x is an expression like

$$1, x + 1, 5 + 3x^2 - 2x.$$

Let us be more precise.

$$\boxed{\text{coefficient}} \quad \boxed{\text{variable}} \quad \boxed{\text{degree}}$$
$$75\, x^{18}$$
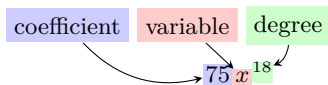
A *monomial* in a variable x is the product of a number (the *coefficient* of the term) with a nonnegative integer (the *degree*) power of x. Write x^1 as x. Write x^0 as 1, and so write any monomial of degree zero as just its coefficient, and call it a *constant.* If the coefficient of a monomial is 1, write it without a coefficient: $1\,x^{57} = x^{57}$.

A *polynomial* is a finite sum of monomials, its *terms.* We can write the sum in any order, and drop any terms with zero coefficient, without changing the polynomial. If the sum has no terms, write the polynomial as 0. A *constant* is a polynomial which has only a constant term or is 0. Add polynomials by adding coefficients of terms of the same degree:

$$2x^2 + 7x^2 + x = (2 + 7)x^2 + x = 9x^2 + x.$$

Hence we can write any polynomial so that all terms have different degree, say decreasing in degree, or increasing in degree: $1 + x^3 + 5x = 1 + 5x + x^3 = x^3 + 5x + 1$, and drop all terms with zero coefficient. Two polynomials written this way are equal just when they have the same coefficient in every degree. The *degree* is the largest degree of any term. The degree of 0 is defined to be $-\infty$. Multiply monomials like $(7x^2)(5x^3) = (7 \cdot 5)x^{2+3}$, multiplying coefficients and adding degrees. Multiply polynomials by the distributive law, term by term. The coefficients can be integers, rational numbers, real numbers, or complex numbers (in which case, we usually use z as the name of the variable, instead of x). They can even be remainders modulo some integer. As soon as we fix our choice of coefficients, the addition laws, multiplication laws, and distributive law (familiar from chapter 1) follow: write $b(x)$ as a polynomial, instead of an integer b, and so on, and change the word *integer* to *polynomial.* Ignore the sign laws and the law of well ordering.

Let's work modulo 2, in honour of George Boole. Let $p(x) := x^2 + x$. We write $x^2 + x$ to mean $\bar{1}x^2 + \bar{1}x$, i.e. the coefficients are remainders. This polynomial

is not zero, because its coefficients are not zero. *Danger:* we might want to think of this polynomial as a *function* of a variable x, and ask that x also be a remainder modulo 2. But then that function *is* zero, because if we let $x = \bar{0}$ then

$$p(\bar{0}) = \bar{1} \cdot \bar{0}^2 + \bar{1} \cdot \bar{0} = \bar{0}$$

and if we let $x = \bar{1}$ then

$$p(\bar{1}) = \bar{1} \cdot \bar{1}^2 + \bar{1} \cdot \bar{1} = \bar{0}$$

modulo 2. So a polynomial is *not* just a function: the function can be zero while the polynomial is not. A polynomial is a purely algebraic object.

8.1 Work modulo 5: for which remainders a can we find a remainder b so that $a = b^2$? In other words, which elements have square roots?

8.2 Give an example of a polynomial $b(x)$ none of whose coefficients are integers, so that, if we set x to an integer value, then $b(x)$ takes an integer value.

The extended Euclidean algorithm

You are familiar with long division of polynomials. For example, to divide $x - 4$ into $x^4 - 3x + 5$, we have to order the terms in decreasing degrees as usual, and add in zero terms to get all degrees to show up then

$$
\begin{array}{r}
x^3 + 4x^2\ +16x\ +61 \\
\hline
x-4\ \big)\ x^4 +0x^3\ +0x^2\ -3x\ +5 \\
x^4 -4x^3 \\
\hline
4x^3\ +0x^2\ -3x\ +5 \\
4x^3 -16x^2 \\
\hline
16x^2\ -3x\ +5 \\
16x^2 -64x \\
\hline
61x\ +5 \\
61x -244 \\
\hline
249
\end{array}
$$

The quotient is $x^3 + 4x^2 + 16x + 61$ and the remainder is 249. Summing up, this calculation tells us that

$$x^4 - 3x + 5 = (x - 4)\left(x^3 + 4x^2 + 16x + 61\right) + 249,$$

We stop when the remainder has small *degree* than the expression we want to divide in, in this case when 249 has degree zero, smaller than the degree of $x - 4$.

In this simple problem, we never needed to divide. But in general, to divide $b(x)$ by $c(x)$, we might need to divide coefficients of $b(x)$ by those of $c(x)$ at some stage. Therefore from now on, we will restrict the choice of coefficients to ensure that we can always divide. For example, we can carry out long division of any two polynomials with rational coefficients, and we always end up with a quotient and a remainder,

which are themselves polynomials with rational coefficients. The same is true if we work with real number coefficients, or complex number coefficients. But if we work with integer coefficients, the resulting quotient and remainder might only have rational coefficients.

8.3 Determine the greatest common divisor of $b(x) = x^3 + 4x^2 + x - 6$ and $c(x) = x^5 - 6x + 5$ and the associated Bézout coefficients $s(x), t(x)$.

It is trickier to specify the restrictions on coefficients if they are to be remainders modulo an integer. From now on, we will only work with remainders modulo a prime. The reason is that every nonzero remainder modulo a prime has a reciprocal, so we can always divide coefficients.

8.4 Prove that every nonzero remainder modulo a prime has a reciprocal modulo that prime.

With these restrictions on coefficients, we can immediately generalize the Euclidean and extended Euclidean algorithms, and compute Bézout coefficients and least common multiples, using the same proofs, so we won't repeat them.

Take the polynomials $b(x) = 3 + 2x - x^2$ and $c(x) = 12 - x - x^2$. We want to find the Bézout coefficients and the greatest common divisor among polynomials over the rational numbers. At each step, force down the degree of one of the polynomials.

$$\begin{pmatrix} 1 & 0 & 3 + 2x - x^2 \\ 0 & 1 & 12 - x - x^2 \end{pmatrix} \text{ add -row 2 to row 1,}$$

$$\begin{pmatrix} 1 & -1 & -9 + 3x \\ 0 & 1 & 12 - x - x^2 \end{pmatrix} \text{ add } (x/3)\text{row 1 to row 2,}$$

$$\begin{pmatrix} 1 & -1 & -9 + 3x \\ \frac{x}{3} & 1 - \frac{x}{3} & 12 - 4x \end{pmatrix} \text{ add } \frac{4}{3}\text{row 1 to row 2,}$$

$$\begin{pmatrix} 1 & -1 & -9 + 3x \\ \frac{4}{3} + \frac{x}{3} & -\frac{1}{3} - \frac{x}{3} & 0 \end{pmatrix}$$

The equation of the Bézout coefficients is therefore

$$1\left(3 + 2x - x^2\right) - 1\left(12 - x - x^2\right) = -9 + 3x = 3(x - 3).$$

The greatest common divisor, once we divide away any coefficient in front of the highest term, is $x - 3$. A *monic polynomial* is one whose highest degree term has coefficient 1: the greatest common divisor will always be taken to be a monic polynomial, and then there is a unique greatest common divisor.

Fix a choice of whether we are working with rational, real or complex coefficients or if our coefficients will be remainders modulo a prime. We will say this choice is a choice of *field*, and our polynomials are said to live *over* that field. A polynomial over a chosen field is *irreducible* if it does not split into a product $a(x)b(x)$ of polynomials over the same field, except for having $a(x)$ or $b(x)$ constant. The same proof for prime numbers proves that a polynomial over a field admits a factorization into a

product of irreducibles over that field, unique up to reordering the factors and perhaps multiplying any of the factors by a nonzero constant.

> Let's working over the remainders modulo 2. Notice that $2x = 0x = 0$, and similarly that $-1 = 1$ so $-x^2 = x^2$. Let's find the Bézout coefficients of $b(x) := x^3 + x$ and $c(x) := x^2$:
>
> $$\begin{pmatrix} 1 & 0 & x^3 + x \\ 0 & 1 & x^2 \end{pmatrix}, \text{ add } x(\text{row } 2) \text{ to row } 1,$$
>
> $$\begin{pmatrix} 1 & x & x \\ 0 & 1 & x^2 \end{pmatrix}, \text{ add } x(\text{row } 1) \text{ to row } 2,$$
>
> $$\begin{pmatrix} 1 & x & x \\ x & x^2 + 1 & 0 \end{pmatrix}.$$
>
> The Bézout coefficients are $1, x$, and the greatest common divisor is x:
>
> $$1\left(x^3 + x\right) + x\left(x^2\right) = x.$$

8.5 Find the Bèzout coefficients of $b(x) = x^3 + x$, $c(x) = x^4 + 1$ with coefficients in remainders modulo 3.

Factoring

> *- What is a multiple root of a polynomial?*
> *- Well, it is when we substitute a number in the polynomial and get zero. Then do it again and again get zero and so k times .... But on the $(k + 1)$-st time the zero does not appear.*
>
> — A. A. Kirillov
> WHAT ARE NUMBERS?

Proposition 8.1. *Take any constant c and any polynomial $p(x)$ over any field. Then $p(x)$ has remainder $p(c)$ when divided by $x - c$.*

Proof. Take quotient and remainder: $p(x) = (x - c)q(x) + r(x)$, using the Euclidean algorithm. The degree of the remainder is smaller than the degree of $x - c$, so the remainder is a constant, say r_0. Plug in $x = c$: $p(c) = (c - c)q(c) + r_0 = r_0$. $\square$

A *root* of a polynomial $p(x)$ is a number c in whichever field we work over so that $p(c) = 0$.

Corollary 8.2. *A polynomial $p(x)$ over any field has a root at $x = c$, i.e. $p(c) = 0$, for some constant c in the field, just when the linear polynomial $x - c$ divides into $p(x)$.*

8.6 Over the field of real numbers, or the field of rational numbers, or the field of complex numbers, or the field of remainders modulo any odd prime number: show that every constant d has either two square roots, which we denote as $\pm\sqrt{d}$, or one square root (if $2 = 0$ in our field or if $d = 0$) or has no square roots.

8.7 Work over the field of real numbers, or the field of rational numbers, or the field of complex numbers, or the field of remainders modulo any odd prime number. Show that the solutions of the quadratic equation

$$0 = ax^2 + bx + c$$

(where a, b, c are constants and $a \neq 0$) are precisely the numbers x so that

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

Prove that the solutions of the quadratic equation, over any field, are

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

just when the required square roots exist.

8.8 Prove that a quadratic or cubic polynomial (in other words of degree 2 or 3) over any field is reducible just when it has a root.

8.9 Prove that the polynomial $x^2 + x + 1$ is irreducible over the field of remainders modulo 2. By writing down all quadratic polynomials over that field, and factoring all of them but this one, show that this is the only irreducible quadratic polynomial over that field.

8.10 Prove that the polynomial $x^3 + x + 1$ is irreducible over the field of remainders modulo 2.

> The polynomial $x^3 + x + 1$ is irreducible over the field of remainders modulo 2, as we saw in the previous problem. But then, if it is reducible over the integers, say $a(x)b(x)$, then quotient out the coefficients modulo 2 to get a factorization $\bar{a}(x)\bar{b}(x)$ over the field of remainders modulo 2, a contradiction. Therefore $x^3 + x + 1$ is irreducible over the integers.

Corollary 8.3. *Every polynomial over any field has at most one factorisation into linear factors, up to reordering the factors.*

Proof. The linear factors are determined by the roots. Divide off one linear factor from each root, and apply induction on the degree of the polynomial to see that the number of times each linear factor appears is uniquely determined. □

8.11 Suppose that $p(x)$ and $q(x)$ are polynomials over some field, both of degree at most n, and that $p(c) = q(c)$ for $n + 1$ different numbers c from the field. Prove that $p(c) = q(c)$ for all numbers c from the field.

8.12 Why are we working over fields? Working modulo 6, consider the polynomial $b(x) = x^2 + 5x$. How many roots does it have among the remainders module 6? How many factorizations can you make for it, working modulo 6?

A root c of a polynomial $p(x)$ has *multiplicity* k if $(x - c)^k$ divides $p(x)$.

Corollary 8.4. *Over any field, the degree of a polynomial is never less than the sum of the multiplicities of its roots, with equality just when the polynomial is a product of linear factors.*

Proof. Degrees add when polynomials multiply, by looking at the leading term.   □

Rational roots

A powerful trick for guessing rational roots of integer coefficient polynomials:

Lemma 8.5. *If a polynomial $p(x)$ with integer coefficients has a rational root $x = n/d$, with n and d coprime integers, then the numerator n divides the coefficient of the lowest term of $p(x)$, while the denominator d divides the coefficient of the highest term of $p(x)$.*

Proof. Write out

$$p(x) = a_k x^k + a_{k-1} x^{k-1} + \cdots + a_1 x + a_0,$$

so that these integers $a_0, a_1, \ldots, a_k$ are the coefficients of $p(x)$. Plug in:

$$0 = a_k (n/d)^k + a_{k-1} (n/d)^{k-1} + \cdots + a_1 (n/d) + a_0.$$

Multiply both sides by d^k:

$$0 = a_k n^k + a_{k-1} n^{k-1} d + \cdots + a_1 n d^{k-1} + a_0 d^k.$$

All terms but the last term contain a factor of n, so n divides into the last term. But n and d are coprime, so n divides into a_0. The same trick backwards: all terms but the first term contain a factor of d, so d divides into the first term. But n and d are coprime, so d divides into a_k.   □

> The polynomial $p(x) = x^3 - 3x - 1$, if it had a rational root n/d, would need to have n divide into -1, while d would divide into 1. So the only possibilities are $n/d = 1$ or $n/d = -1$. Plug in to see that these are not roots, so $x^3 - 3x - 1$ has no rational roots.

8.13 Using lemma 8.5, find all rational roots of

  a. $x^2 - 5$,

  b. $x^2 - 3x - 5$,

  c. $x^{1000} - 3x - 5$.

  d. $x^2/3 - x - 5/3$,

Sage

> *The world of the future will be an even more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.*
>
> — Norbert Wiener
>   THE HUMAN USE OF HUMAN BEINGS: CYBERNETICS AND SOCIETY

To construct polynomials in a variable x, we first define the variable, using the strange expression:

```
x = var('x')
```

We can then solve polynomial equations:

```
solve(x^2 + 3*x + 2, x)
```

yielding [x == -2, x == -1]. To factor polynomials:

```
x=var('x')
factor(x^2-1)
```

yields (x + 1)*(x - 1). To use two variables x and y:

```
x, y = var('x, y')
solve([x+y==6, x-y==4], x, y)
```

yielding [[x == 5, y == 1]]. We can take greatest common divisor:

```
x=var('x')
b=x^3+x^2
c=x^2+x
gcd(b,c)
```

yielding $x^2 + x$. For some purposes, we will need to tell sage which field the coefficients of our polynomials come from. We can get the quotient and remainder in polynomial division by

```
R.<x> = PolynomialRing(QQ)
b=x^3+x^2
c=x^2+x
b.quo_rem(c)
```

yielding $(x, 0)$. To find Bézout coefficients,

```
xgcd(x^4-x^2,x^3-x)
```

yields $\left(x^3 - x, 0, 1\right)$.

The expression R.<x> = PolynomialRing(QQ) tells sage to use polynomials in a variable x with coefficients being rational numbers. The set of all rational numbers is called $\mathbb{Q}$, written QQ when we type into a computer. To work with coefficients that are integers modulo a prime, for example with integers modulo 5:

```
R.<x> = PolynomialRing(GF(5))
b=x^3+x^2+1
c=x^2+x+1
b.quo_rem(c)
```

yields $(x, 4x + 1)$. The set of all integers modulo a prime p is sometimes called $\mathrm{GF}(p)$.

We can define our own function to calculate Bézout coefficients, which does exactly what the `xgcd()` function does on polynomials of one variable:

```
def bezpoly(b,c):
    p,q,r=1,0,b
    s,t,u=0,1,c
    while u<>0:
        if r.degree()>u.degree():
            p,q,r,s,t,u=s,t,u,p,q,r
        Q=u//r
        s,t,u=s-Q*p,t-Q*q,u-Q*r
    return (r,p,q)
```

Then the code

```
R.<t> = PolynomialRing(QQ)
f=(2*t+1)*(t+1)*(t-1)
g=(2*t+1)*(t-1)*t^2
bezpoly(f,g)
```

returns $\left(2t^2 - t - 1, -t + 1, 1\right)$.

Interpolation

> *Il faut bien s'arrêter quelque part, et pour que la science soit possible, il faut s'arrêter quand on a trouvé la simplicité.*
>
> We have to stop somewhere, and for science to be possible, we must stop when we have found simplicity.
>
> — Henri Poincaré
>   LA SCIENCE ET L'HYPOTHÈSE

Theorem 8.6. *Over any field, if we specify distinct values $x_0, \ldots, x_n$ and arbitrary values $y_0, \ldots, y_n$, there is a unique polynomial $p(x)$ of degree n so that $p(x_i) = y_i$.*

The polynomial $p(x)$ *interpolates* the values y_i at the points x_i.

Proof. We associate to the numbers $x_0, \ldots, x_n$ the polynomials

$$p_j(x) = \frac{(x - x_0) \ldots (x - x_{j-1})(x - x_{j+1}) \ldots (x - x_n)}{(x_j - x_0) \ldots (x_j - x_{j-1})(x_j - x_{j+1}) \ldots (x_j - x_n)}.$$

for $j = 0, 1, 2, \ldots, n$. The reader can check that $p_j(x_j) = 1$ while $p_j(x_i) = 0$ for $i \neq j$. Then

$$p(x) = y_0 p_0(x) + y_1 p_1(x) + \cdots + y_n p_n(x)$$

is our required polynomial. To check that it is unique, if there are two such polynomials, say $p(x)$ and $q(x)$, both of degree at most n, both equal to y_i at $x = x_i$, their difference $p(x) - q(x)$ vanishes at $x = x_i$, so is divisible by $(x - x_0)(x - x_1) \ldots (x - x_n)$, which has degree $n + 1$. But their difference has degree at most n, so is zero.   □

8.14 What can go wrong with interpolation if the numbers are drawn from the remainders modulo 6?

Chapter 9

Real polynomials, complex polynomials

I cherish the greatest respect towards everybody's religious obligations, never mind how comical.

> — Herman Melville
> MOBY DICK

Hell is an idea first born on an undigested apple-dumpling.

> — Herman Melville
> MOBY DICK

In this chapter (and only this chapter), we assume that the reader is familiar with real numbers, with continuity, and with the intermediate value theorem: every continuous function $y = f(x)$ defined on an interval $a \le x \le b$ takes on all values between $f(a)$ and $f(b)$. We also assume that the reader knows that polynomials are continuous. See [12] for the complete story of continuity and the intermediate value theorem.

Polynomials over the real numbers

If $f(x)$ is a polynomial with positive coefficients, like

$$f(x) = -4 + 7x + 17x^2 + 674x^4,$$

then making x very large makes $f(x)$ very large positive, clearly, since x^4 eventually gets much larger than all other terms. The equation $f(x) = 0$ must have a solution $x > 0$, because $f(x)$ starts at $f(0) = -4$ (negative) and then gets much larger than anything we like eventually for large enough x, so gets positive, so must be zero somewhere in between.

9.1 Prove that the equation

$$394x^3 - 92843x^2 + 209374x - 2830$$

has a real number solution x. More generally, prove that any odd-order polynomial equation in a real variable has a real solution.

Counting positive roots

Theorem 9.1 (Descartes)**.** *The number of positive roots, counted with multiplicities, of a real polynomial $b(x)$ is either equal to the number of changes of sign of its coefficients (when we write the polynomial $b(x)$ with terms in order of degree), or less by a multiple of* 2.

> The polynomial
> $$b(x) = 12x^7 + x^5 - x^3 - x^2 + x + 1$$
> has coefficients with signs $+ + - - + +$. We ignore the zero coefficients, for example in front of x^6 and x^4. So as we travel along the signs, they change twice, once from $+$ to neighboring $-$, and once from $-$ to neighboring $+$. So this $b(x)$ has at most 2 roots $x > 0$, and the number of roots with $x > 0$ is even. So there are either 2 positive roots or none.

> The polynomial $c(x) = x^5 - 3x^4 + 3x^3 + 9x^2 - 27x + 3$ has 4 sign changes, so 4 or 2 or 0 roots x with $x > 0$. If we look at the polynomial $c(-x)$, expand out to get $c(-x) = -x^5 - 3x^4 - 3x^3 + 9x^2 + 27x + 3$ has 1 sign change, so 1 root with $x > 0$. Therefore $c(x)$ has 1 root with $x < 0$.

Proof. Take a polynomial $b(x)$. If x divides $b(x)$, dividing out factors of x from $b(x)$ doesn't change positive roots or signs of coefficients, so we can assume that there are no factors of x. In other words, assume that the constant coefficient of $b(x)$ is not zero:
$$b(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n$$
with $b_0 \neq 0$ and $b_n \neq 0$. For large positive numbers x, $b(x)$ has the same sign as b_n:
$$\frac{b(x)}{x^n} = \frac{b_0}{x^n} + \frac{b_1}{x^{n-1}} + \cdots + \frac{b_{n-1}}{x} + b_n \to b_n.$$

Suppose that $b(x)$ has no positive roots. By the intermediate value theorem, $b(x)$ stays that same sign for all $x > 0$. For $x = 0$, this sign is $b(x) = b_0$, while for x large it is the sign of b_n. So if $b(x)$ has no roots, then b_0 and b_n have the same sign, so there are an even number of sign changes, to start and end at the same sign. In particular, the theorem is true for any polynomial $b(x)$ with no roots.

Suppose that $b(x)$ has a positive root, say at $x = a$. Factor out:
$$b(x) = (x - a)c(x)$$

for some polynomial $c(x)$. We claim that $b(x)$ has one more sign change than $c(x)$. Write out the coefficients of $c(x)$, say as
$$c(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_{n-1} x^{n-1}.$$

Expand out the equation
$$b(x) = (x - a)c(x)$$
to find the relations between the coefficients:
$$b_j = c_{j-1} - ac_j.$$

Starting at the top, $b_n = c_{n-1}$, the top coefficients match in sign. As we go down the coefficients by degree, pick out the first sign change in the coefficients of $c(x)$, say $c_{j-1} < 0$ while $c_j > 0$. Then $b_j = c_{j-1} - ac_j < 0$, so the sign of the b_j coefficient is the same as that of c_{j-1}. If the $b(x)$ coefficients have not already changed sign before we got to b_j, then they certainly must change sign by the time we get to b_j. The same argument works for the next sign change and so on: at least as many sign changes. Since the signs of the highest coefficients of $b(x)$ and $c(x)$ agree, while the signs of the lowest coefficients disagree, the total number of sign changes of $b(x)$ must be greater than those of $c(x)$ by an odd number: one more sign change or three more or five more, etc.

By induction, the number of positive roots of $c(x)$ is equal to the number of sign changes of $c(x)$ or less by a multiple of two. Since $b(x)$ has one more positive root, and one more sign change or three more or five more, etc., the result holds by induction on degree. $\qquad\square$

Sage

The number of sign changes in a polynomial is easy to spot by eye, but we can also write sage code. Each polynomial p has associated list of coefficients p.list(). This list has the coefficients in order, but with zero entries in the list at zero coefficients. For example, if p=t^5-2*t+7 then p.list() yields $[7, -2, 0, 0, 0, 1]$. We will travel along the coefficients one by one, with a for loop. The code (which we explain below):

```
def descartes(p):
    sign = 0
    sign_changes = 0
    for c in p.list():
        if c <> 0:
            if sign == 0:
                if c < 0:
                    sign = -1
                else:
                    if c > 0:
                        sign = 1
            else:
                if c*sign < 0:
                    sign_changes = sign_changes + 1
                    sign = -sign
    return sign_changes
```

We store the sign of the last nonzero coefficient in a variable sign, which is set to zero initially to mean that we haven't yet encountered a nonzero coefficient. If we encounter our first nonzero coefficient, we just set sign to its value. But if we find any subsequent nonzero coefficient, we compare whether it has the same sign as sign, and if it has an opposite sign, i.e. if L[i]*sign<0, then we count a new sign change. Now calling the function:

```
t = var('t')
descartes(t^(79)-4*t^(17)+t^(10)+t^6+t-1)
```

yields 3.

Counting roots of a real polynomial in an interval

Suppose that $p(x)$ is a polynomial with real coefficients. Let $p_0(x)$ be just another name for $p(x)$. Let $p_1(x) := p'(x)$. From then on, compute quotients and remainders: $p_2(x)$ is the negative of the remainder of $p_0(x)$ divided by $p_1(x)$, and so on: $p_{j+1}(x)$ is the negative of the remainder of $p_{j-1}(x)$ divided by $p_j(x)$, until you get to have no remainder, say $p_m(x)$ divides into $p_{m-1}(x)$. The *Sturm sequence* of $p(x)$ is the sequence $p_0(x), p_1(x), \ldots, p_m(x)$.

> The Sturm sequence of $p(x) = x^6 - 12x + 10$ is:
>
> $$p_0(x) = x^6 - 12x + 10,$$
> $$p_1(x) = 6x^5 - 12,$$
> $$p_2(x) = 10x - 10,$$
> $$p_3(x) = 6.$$
>
> Note that $p_3(x)$ divides into $p_2(x)$, because $p_3(x)$ is a nonzero constant.

Take a polynomial $p(x)$, and its Sturm sequence $p_0(x), p_1(x), \ldots, p_m(x)$. For any real number x, let $s(x)$ be the number of sign changes (ignoring zeroes) in the sequence of numbers
$$p_0(x), p_1(x), \ldots, p_m(x).$$
The *expected number of distinct roots* of $p(x)$ in the interval $a < x \leq b$ is $s(a) - s(b)$.

> To find the expected number of distinct roots of $p(x) = x^6 - 12x + 10$ over $0 < x \leq 1$, first note that $p(0) = 10$ and $p(1) = -1$ are not roots, so not multiple roots. The
>
> $$p_0(0) = 10, \qquad p_0(1) = -1,$$
> $$p_1(0) = -12, \quad p_1(1) = -6,$$
> $$p_2(0) = -10, \quad p_2(1) = 0,$$
> $$p_3(0) = 6, \qquad p_3(1) = 6,$$
>
> so $s(0) = 2$ sign changes, and $s(1) = 1$ sign changes, with expected number of roots $2 - 1 = 1$, i.e. we expect that $p(x)$ has all 1 of its roots in the interval $0 < x \leq 1$.

Theorem 9.2 (Sturm). *If a polynomial $p(x)$ with real coefficients has no multiple roots at $x = a$ or at $x = b$ then the expected number of distinct roots $p(x)$ in the interval $a < x \leq b$ is the number of distinct roots of $p(x)$ in that interval.*

Proof. If two successive polynomials in the Sturm sequence share a root, say if $p_{i-1}(x)$ and $p_i(x)$ have a common root at $x = x_0$, then take quotient and remainder: $p_{i-1}(x) = q(x)p_i(x) - p_{i+1}(x)$, and plug in to see that $p_{i+1}(x_0) = 0$ too. The same idea works backwards: if $p_{i+1}(x)$ and $p_i(x)$ have a common root at $x = x_0$, then take quotient and

remainder: $p_{i-1}(x) = q(x)p_i(x) - p_{i+1}(x)$, and plug in to see that $p_{i-1}(x_0) = 0$ too. Hence the common roots of any two successive polynomials in the Sturm sequence lie at precisely the multiple roots of $p(x)$.

Suppose that $x = x_0$ is a root of $p_m(x)$. By definition, $p_m(x)$ is the greatest common divisor (up to $\pm$ sign) of $p(x)$ and $p'(x)$, so divides into all of the $p_0(x), p_1(x), \ldots, p_m(x)$. So its zeroes are already zeroes of all of those, and so $x = x_0$ is a root of all of $p_0(x), p_1(x), \ldots, p_m(x)$. Since $p_0(x) = p(x)$ and $p_1(x) = p'(x)$, $x = x_0$ is a zero of both $p(x)$ and $p'(x)$, so a double root of $p(x)$.

Take a root $x = x_0$ of one of the polynomials $p_i(x)$ in the Sturm sequence. Suppose that $p_i(x)$ is not $p_0(x)$ or $p_m(x)$, at the two ends of the sequence, but just one of the polynomials in middle of the sequence. Suppose that $x = x_0$ is not a multiple root of $p(x)$. From the above reasoning, $x = x_0$ is not a root of either $p_{i-1}(x)$ or of $p_{i+1}(x)$. Take quotient and remainder: $p_{i-1}(x) = q(x)p_i(x) - p_{i+1}(x)$ and plug in to see that $p_{i-1}(x_0) = -p_{i+1}(x_0)$. So if $x = x_0$ is not a root of all of the polynomials in the chain, but a root of one of them, p_i, then his neighbours p_{i-1}, p_{i+1} disagree in their sign at this point $x = x_0$, and so also disagree near $x = x_0$.
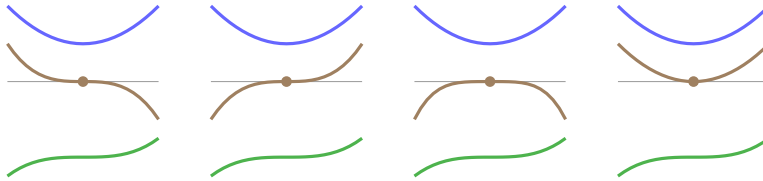
Suppose in addition that $p_i(x)$ has an odd number of roots at $x = x_0$. As x travels from a little less than $x = x_0$ to a little more than $x = x_0$, the sign of $p_i(x)$ changes across this gap, but that of $p_{i-1}(x)$ and $p_{i+1}(x)$ stay the same. So they together contribute the same total number of sign changes for $x > x_0$ as they did for $x < x_0$.
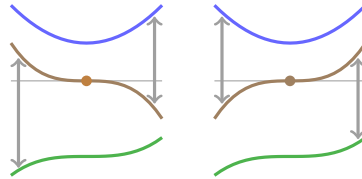
Suppose instead in addition that $p_i(x)$ has an even number of roots at $x = x_0$. As x travels from a little less than $x = x_0$ to a little more, the sign of all of $p_{i-1}(x), p_i(x), p_{i+1}(x)$ stay the same across this gap. So they together contribute the same total number of sign changes for $x > x_0$ as they did for $x < x_0$.
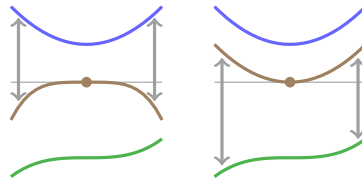
Lets go back to the start of the Sturm sequence. If $p(x)$ has an odd number of roots at some point $x = x_0$, then $p'(x)$ has an even number, and so $p(x)$ changes sign as x goes from $x < x_0$ to $x > x_0$, while $p'(x)$ doesn't change sign.
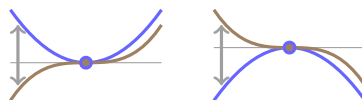
If $p(x)$ increases across this gap, then $p'(x) > 0$ on either side, so the number of sign changes contributed by $p_0(x), p_1(x)$ decreases:
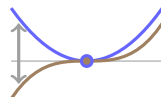
Similarly if $p(x)$ decreases:
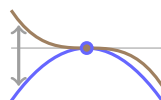
Similarly if $p(x)$ has an even number of roots:
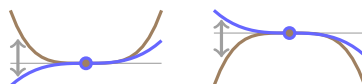
Hence as x pass through any root of $p(x)$, the terms $p_0(x), p_1(x)$ contribute different numbers of sign changes in the Sturm sequence by at least one. $\square$

9.2 Explain why, at each step in the Sturm sequence, we can allow ourselves to divide or multiply each polynomial $p_i(x)$ by any positive constant number. (That can help to simplify the computation.)

9.3 Apply the Sturm theorem to the polynomial $p(x) = 4x^4 + 2x^2 - 1$ to find the number of roots between $x = 0$ and $x = 1$.

9.4 Apply the Sturm theorem to the polynomial $p(x) = x^4 + 4x^3 - 1$ to find the number of roots between $x = 0$ and $x = 1$.

9.5 Apply the Sturm theorem to the polynomial $p(x) = x^6 + 4x^3 - 2$ to find the number of roots between $x = 0$ and $x = 1$.

9.6 Apply the Sturm theorem to the polynomial $p(x) = 2x^4 + 4x^3 + 2x^2 - 1$ to find the number of roots between $x = -2$ and $x = 1$.

9.7 What do the signs in a Sturm sequence say when you plug in a quadratic function $p(x) = ax^2 + bx + c$, say with $a \neq 0$?

To count out zeroes:

```
def count_sign_changes(L):
    n=0
    # Collect up only the nonzero elements of the list L
    M=[ x for x in L if x!= 0 ]
    for i in range(0,len(M)-1):
        p=M[i]*M[i+1]
        if p<0:
```

```
        n=n+1
    return n
def expected_number_of_zeroes(p,a,b):
    if p == 0:
        return oo
        # Returns infinity
    # Create a list called L containing the Sturm sequence.
    L = [p,diff(p(x),x)]
    n = 2
    q,r=L[n-2].quo_rem(L[n-1])
    while r!=0:
        # Every nonzero remainder r gets -r stuck in the Sturm sequence.
        L.append(-r)
        n=n+1
        q,r=L[n-2].quo_rem(L[n-1])
    # Make a list A of values of the Sturm sequence polynomials at x=a.
    A=[ t.subs(x=a) for t in L ]
    B=[ t.subs(x=b) for t in L ]
    return count_sign_changes(A)-count_sign_changes(B)
```

Then the code:

```
R.<x>=PolynomialRing(QQ)
b=(x-1)*x*(x+1)^3
expected_number_of_zeroes(b,0,1)
```

yields 1.

9.8 For a polynomial $p(x) = a_n x^n + \cdots + a_0$ with real coefficients, find a constant $c > 0$ so that every real root x of $p(x)$ lies in the interval $-c \leq x \leq c$. (By repeatedly cutting this interval in half, and applying Sturm's theorem, we can rapidly zoom in on the zeroes of $p(x)$, using a computer.)

Factoring complex polynomials

> *For every complex problem there is an answer that is clear, simple, and wrong.*
>
> — H. L. Mencken

The complex numbers are in a very strong sense free of the deficiencies of the rational and real numbers:

Theorem 9.3 (The fundamental theorem of algebra). *Every nonconstant polynomial function*

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$$

with complex number coefficients $a_0, a_1, \ldots, a_n$ has a complex root, i.e. a complex number $z = z_0$ so that $p(z_0) = 0$.

The proof of the theorem uses analysis, but we want to focus on algebra; see [12] p. 513 chapter 2 theorem 2 for a complete proof. For every complex polynomial problem, there is an answer that is perhaps unclear, complex and right.

9.9 Draw the roots of $z^5 - 1$.

Theorem 9.4. *Every complex coefficient polynomial function*

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$$

of any degree n splits into a product of linear factors

$$p(z) = a_n \left(z - r_1 \right) \left(z - r_2 \right) \ldots \left(z - r_n \right).$$

In particular, a complex coefficient polynomial is irreducible just when it is linear.

Proof. Apply the fundamental theorem of algebra to find a root r_1, and then divide $p(z)$ by $z - r_1$ and apply induction. □

Factoring real polynomials

Theorem 9.5. *Every real coefficient polynomial function*

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

of any degree n splits into a product of real linear and quadratic factors

$$p(x) = a_n \left(x - r_1 \right) \left(x - r_2 \right) \ldots \left(x - r_k \right) q_1(x) q_2(x) \ldots q_\ell(x),$$

where $q_j(x) = x^2 + b_j x + c_j$ is quadratic with no real roots. In particular, a real coefficient polynomial is irreducible just when it is linear or quadratic $ax^2 + bx + c$, with $a \neq 0$ and with $b^2 - 4ac < 0$.

Proof. If the polynomial $p(x)$ has a real root, divide off the associated linear factor, and apply induction on degree of $p(x)$. So we can assume $p(x)$ has no real roots. Take a complex root, say $z_1 = x_1 + y_1 i$. Because all coefficients of $p(x)$ are real, $p(x) = \overline{p(x)}$, for any real number x, and more generally for a complex number z:

$$\overline{p(z)} = p(\bar{z}).$$

In particular, if $z_1 = x_1 + y_1 i$ is a root, then $\bar{z}_1 = x_1 - y_1 i$ is also a root. So $p(z)$ is divisible by

$$(z - z_1)(z - \bar{z}_1) = z^2 - 2x_1 z + |z_1|^2,$$

a quadratic function with real coefficients. Divide this factor off and apply induction on degree of $p(x)$. □

Partial fractions

Lemma 9.6. *Take any rational function*

$$\frac{p(x)}{q(x)}$$

over any field. Suppose that its denominator factors, say into coprime factors $q(x) = u(x)v(x)$. Then it splits into a sum:

$$\frac{p(x)}{q(x)} = \frac{t(x)p(x)}{u(x)} + \frac{s(x)p(x)}{v(x)}$$

where $s(x), t(x)$ are the Bézout coefficients of $u(x), v(x)$.

Proof. Write out $s(x)u(x) + t(x)v(x) = 1$ and then multiply both sides by $p(x)$ and divide both sides by $q(x)$. $\qquad\square$

A *partial fraction* is a rational function $\frac{b(x)}{c(x)^n}$ where $b(x)$ has smaller degree than $c(x)$ and $c(x)$ is irreducible. A *partial fraction decomposition* of a rational function $\frac{p(x)}{q(x)}$ is an expression as a sum of a polynomial function together with finitely many partial fractions, so that the denominators of the partial fractions multiplied together divide $q(x)$.

Theorem 9.7. *Every rational function over any field has a partial fraction decomposition, unique up to reordering the partial fractions.*

Proof. Just repeat the process described in lemma 9.6. $\qquad\square$

$$\frac{4}{x^2 + 2x - 3} = \frac{-1}{x + 3} + \frac{1}{x - 1}.$$

Over the field of real numbers, the rational function

$$\frac{p(x)}{q(x)} = \frac{x^9 - 2x^6 + 2x^5 - 7x^4 + 13x^3 - 11x^2 + 12x - 4}{x^7 - 3x^6 + 5x^5 - 7x^4 + 7x^3 - 5x^2 + 3x - 1}$$

has partial fraction decomposition

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

Note that we cannot get rid of the square in the last denominator, because we have to allow powers. Also note that if we allow complex numbers here instead of real numbers, then we can factor $x^2 + 1 = (x - i)(x + i)$, so there is a different partial fraction decomposition over the complex numbers, with lower degrees in denominators.

In sage:

```
x=var('x')
f = 4/(x^2+2*x-3)
f.partial_fraction(x)
```

yields $-\frac{1}{x+3} + \frac{1}{x-1}$.

Integrals

Corollary 9.8. *Every rational function $f(x)$ with real coefficients has indefinite integral $\int f(x)\,dx$ a polynomial in functions of the form*

$$x, \frac{1}{x-c}, \log(x-c), \log\left(x^2 + bx + c\right), \arctan(x-c),$$

for finitely many constant numbers b, c.

Proof. The partial fraction decomposition of $f(x)$ is a sum of a polynomial and some partial fractions. The integral of a polynomial is a polynomial, and the integral of a sum is a sum, so it is enough to prove the result assuming that $f(x)$ is a single partial fraction $\frac{p(x)}{q(x)^n}$. Write out $p(x)$ as a sum of terms, and thereby break up the integral, so it is enough to assume that $p(x) = x^k$ for some integer $k \geq 0$. Since $q(x)$ is irreducible, it is linear or quadratic with no real roots.

Suppose that $q(x)$ is linear and $n = 1$. Since the degree of $p(x)$ is smaller than that of $q(x)$, $p(x)$ is a constant, and we can rescale to get $p(x) = 1$ and $q(x)$ monic, and translate in x by a constant to get $q(x) = x$:

$$\int f(x)\,dx = \int \frac{dx}{x} = \log x + C.$$

Suppose that $q(x)$ is quadratic and $n = 1$. Rescale to get $q(x)$ and $p(x)$ monic. Since the degree of $p(x)$ is smaller than that of $q(x)$, $p(x)$ is a constant or linear. This $q(x)$ has two complex roots, say z_1 and $\bar{z}_1$. Translate x by a real number constant to get z_1 to have vanishing real part: $z_1 = y_1 i$. Rescale variables to get $y_1 = 1$, so that $q(x) = x^2 + 1$. If $p(x)$ is linear, say $p(x) = x + c$, then

$$\frac{p(x)}{q(x)} = \frac{x}{q(x)} + c\frac{1}{q(x)},$$

so we can assume that $p(x) = x$ or $p(x) = 1$. So then

$$f(x) = \frac{x}{x^2+1}, \quad \int f(x)\,dx = \frac{\log\left(x^2+1\right)}{2} + C,$$

or

$$f(x) = \frac{1}{x^2+1}, \quad \int f(x)\,dx = \arctan(x) + C.$$

If $q(x) = (x-c)^n$ with $n \geq 2$, translate to get $q(x) = x^n$ so that we can reduce the problem to

$$\int \frac{dx}{x^n} = \frac{x^{1-n}}{1-n} + C.$$

If $q(x)$ is an irreducible quadratic, brought to a power $n \geq 2$, translation gives us $q(x) = 1 + x^2$ as above. If $p(x) = 1$, then we use

$$\int \frac{dx}{(1+x^2)^{n+1}} = \frac{x}{2n\,(1+x^2)^n} + \frac{2n-1}{2n} \int \frac{dx}{(1+x^2)^n},$$

by induction. If $p(x) = x^{2k+1}$ an even power, then use substitution $u = x^2$, to reduce the problem down by induction. In general, differentiate the expression

$$\frac{x^k}{(1+x^2)^n}$$

and integrate both sides to find

$$\frac{x^k}{(1+x^2)^n} + C = -2n \int \frac{x^{k+1}\,dx}{(1+x^2)^{n+1}} + k \int \frac{x^{k-1}dx}{(1+x^2)^n}.$$

This enables us, by induction, to lower the order of the numerator, at the cost of raising the order of the denominator, to reduce the integration to a previous case. □

Sage can integrate:

```
integral((1/x)+(x-1)/x^3,x)
```

yields $-\frac{2\,x-1}{2\,x^2} + \log(x)$.

Chapter 10

Factoring polynomials

Silva: *You've got many refinements. I don't think you need to worry about your failure at long division. I mean, after all, you got through short division, and short division is all that a lady ought to cope with.*

— Tennessee Williams
BABY DOLL AND TIGER TAIL, Act I, Scene 2

Factoring with integer coefficients

10.1 Suppose that $p(x)$ is a nonconstant polynomial with integer coefficients. Prove that there are infinitely many positive integers x at which $p(x)$ is not prime.

Proposition 10.1. *Working with coefficients remainders modulo a prime, if $0 = b(x)c(x)$ for two polynomials $b(x), c(x)$, then either $b(x) = 0$ or $c(x) = 0$.*

Proof. The highest term of the product is the product of the highest terms, so has coefficient the product of the coefficients of the highest terms. In problem 5.13 on page 33, we saw that if a product of remainders modulo a prime vanishes (modulo that prime), then one of the factors is zero. So the coefficient of the highest term of either $b(x)$ or of $c(x)$ is zero. But then by definition that isn't the highest term. $\square$

Lemma 10.2 (Gauss's lemma). *Any polynomial $p(x)$ with integer coefficients factors over rational numbers just when it factors over integers. In other words, suppose that we can factor it as $p(x) = b(x)c(x)$ into polynomials $b(x)$, $c(x)$ with rational coefficients. Then we can rescale each of $b(x)$ and $c(x)$, by multiplying by nonzero rational numbers, to arrange that $p(x) = b(x)c(x)$ still but now $b(x)$ and $c(x)$ have integer coefficients.*

Proof. The idea is to just clear denominators. First, take a least common denominator d for all of the coefficients of $b(x)$ and $c(x)$, and scale both sides with it, so that now we have an equation $d\,p(x) = b(x)c(x)$ with new polynomials $b(x), c(x)$, but with integer coefficients. Expand d into its prime factorization, say

$$d = p_1 p_2 \dots p_n,$$

where some of these primes might appear several times in the factorization. Quotient out the coefficients of both sides of $d\,p(x) = b(x)c(x)$ by the prime p_1 to see $0 = \bar{b}(x)\bar{c}(x)$. So then one of $\bar{b}(x)$ or $\bar{c}(x)$ vanishes (in other words, all of its coefficients vanish). Say, for example, $\bar{b}(x) = 0$. But then p_1 divides into all coefficients of $b(x)$, so we can divide both sides of $d\,p(x) = b(x)c(x)$ by p_1. Repeat until you have divided away all of the prime factors in d. $\square$

75

A polynomial with integer coefficients is *irreducible* if it does not split into a product $b(x)c(x)$ except for having $b(x) = \pm 1$. (Note that this notion of irreducibility is different from being irreducible over a field.)

Corollary 10.3. *Suppose that $p(x)$ is a polynomial with coprime integer coefficients. Then it is irreducible as an integer coefficient polynomial just when it is irreducible as an rational coefficient polynomial.*

Proof. By Gauss' Lemma above, if $p(x)$ factors into rational polynomials, then it factors into integer polynomials. Conversely, since $p(x)$ has coprime coefficients, if $p(x)$ factors into integer polynomials $p(x) = b(x)c(x)$ then neither $b(x)$ nor $c(x)$ can be constant polynomials. $\qquad\square$

> We saw on page 58 that $x^3 - 3x - 1$ has no rational roots. If it is reducible over the field of rational numbers, then it has a rational root; see problem 8.8 on page 57. So therefore it is irreducible over the field of rational numbers. Its coefficients are coprime, so it is irreducible over the integers.

Theorem 10.4. *Every nonzero integer coefficient polynomial has a factorization into irreducible integer coefficient polynomials, unique up to the order in which we write down the factors and up to perhaps multiplying factors by -1.*

Proof. Let d be the greatest common divisor of the coefficients of $p(x)$, so that $p(x) = d\,P(x)$, where the coefficients of $P(x)$ are coprime. Since d factors uniquely into primes, it suffices to prove that $P(x)$ can be factored uniquely into irreducibles. Thus we may assume that coefficients of $p(x)$ are coprime.

Factor $p(x)$ into irreducibles over the field of rational numbers. By Gauss' Lemma, such a factorization yields a factorization of $p(x)$ into integer coefficient factors, each a constant rational number multiple of the rational coefficient factors. We want to check that the factors remain irreducible. Since the coefficients of $p(x)$ are coprime, the coefficients in each of these factors are also coprime: a common divisor would pull out of the whole factorization. By corollary 10.3, each factor is irreducible over the integers.

Suppose that we have two factorizations of $p(x)$ into irreducible integer coefficient polynomials. Recall that the factorization over rationals is unique up to reordering factors and scaling factors by nonzero constant rationals. Therefore our two factorizations are each obtained from the other by such tricks. So if we take out one of the factors $P(x)$ from one factorization, and the corresponding factor $Q(x)$ from the other, then $Q(x) = \frac{a}{b}P(x)$ where a, b are integers, so $bQ(x) = aP(x)$. The coefficients of $P(x)$ are coprime integers, and so are those of $Q(x)$. Taking greatest common divisor, we find $b = \pm a$. So $P(x) = \pm Q(x)$. $\qquad\square$

Sage can test to see if a polynomial is irreducible:

```
R.<x> = PolynomialRing(QQ)
b=x^3+x+2
b.is_irreducible()
```

yields False. Indeed `factor(b)` yields $(x + 1) \cdot (x^2 - x + 2)$. To work over the finite field with 2 elements:

```
R.<x> = PolynomialRing(GF(2))
b=x^3+x+2
b.is_irreducible()
```

yields False, while `factor(b)` yields $x \cdot (x+1)^2$.

Eisenstein's criterion: checking that there are no more factors

Proposition 10.5 (Eisenstein's criterion). *Take a polynomial*

$$q(x) = a_n x^n + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0$$

with integer coefficients $a_0, a_1, \ldots, a_n$. Suppose that there is a prime p so that

    a. *p does not divide the highest degree coefficient a_n and*

    b. *p divides all of the other coefficients and*

    c. *p^2 does not divide the lowest coefficient a_0.*

Then $q(x)$ is irreducible over the field of rational numbers.

Proof. By corollary 10.3 on the facing page, if $q(x)$ factors over the field of rational numbers, then it factors over the integers, say

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = (b_r x^r + \cdots + b_0)(c_s x^s + \cdots + c_0)$$

with integer coefficients

$$b_0, b_1, \ldots, b_r, c_0, c_1, \ldots, c_s.$$

Since p, but not p^2, divides the lowest coefficient $a_0 = b_0 c_0$, p divides exactly one of b_0, c_0, say b_0. Now from the equation

$$a_1 = b_0 c_1 + b_1 c_0,$$

and the fact that p divides all but the highest degree coefficient a_n, we see that p divides b_1. From the equation

$$a_2 = b_0 c_2 + b_1 c_1 + b_2 c_0,$$

we see that p divides b_2. By induction, we find that p divides all coefficients

$$b_0, b_1, \ldots, b_r.$$

This contradicts the condition that p does not divide a_n. $\square$

> The polynomial $x^9 + 14x + 7$ is irreducible by Eisenstein's criterion.

> For any prime p and positive integer d, the polynomial $x^d - p$ satisfies Eisenstein's criterion. Therefore there are no rational number square roots, cube roots, and so on, of any prime number p.

> The polynomial $x^2 + 4x + x = (x+2)^2$ is not irreducible, so we *cannot* always apply Eisenstein's criterion with nonprime integers p, for instance with $p = 4$.

10.2 Give some examples of polynomials to which Eisenstein's criterion applies.

10.3 Suppose that $b(x)$ is a polynomial with integer coefficients of degree d. Prove that $c(x) := x^d b(1/x)$ is also a polynomial with integer coefficients of degree d. Prove that $b(x)$ is irreducible just when $c(x)$ is.

10.4 Explain why the lowest terms in y of a product $b(x,y)c(x,y)$ are obtained by expanding the product of the lowest y terms in each of $b(x,y)$ and $c(x,y)$. Use this to prove that
$$p(x,y) = 5 + 25x + 8x^2 + 7xy + 19y^3$$
is irreducible.

10.5 Explain why the highest terms in y of a product $b(x,y)c(x,y)$ are obtained by expanding the product of the highest y terms in each of $b(x,y)$ and $c(x,y)$. Use this to prove that

$$q(x,y) = 2 + 9xy + 40x^3y^2 + 5y^3 + 25xy^3 + 8x^2y^3$$

is irreducible.

10.6 Is
$$a(x,y) = y^9 x^9 + (y^2 + 1)x^8 + (y^2 + 1)x^7 + y + 1$$
irreducible as a polynomial in two variables x, y with coefficients remainders modulo 2?

10.7 Suppose that p is a prime number. Prove that the polynomial $c(x) := x^{p-1} + x^{p-2} + \cdots + x + 1$ is irreducible. Hint: problem .

Eisenstein's criterion in sage

To check Eisenstein's criterion, tell sage to work with integer coefficient polynomials:

```
R.<x> = PolynomialRing(ZZ)
```

Write a function to find a list of prime factors of any integer:

```
def prime_factors(x):
    x=abs(x)
    list=[]
    p=2
    while p<=x:
        if x%p==0:
            list=list+[p]
            while x%p==0:
                x=x//p
        p=p+1
    return list
```

Note that [] represents an empty list, and we add lists by concatenating them. So prime_factors(-6) yields [2,3], the prime factors in order. Finally, we make a function Eisenstein(b) to apply to integer coefficient polynomials, which returns the lowest prime p for which Eisenstein's criterion applies to our polynomial $b(x)$, or returns 0 if no such prime exists.

```
def Eisenstein(b):
    c=b.coefficients()
    highest=c.pop()
    possible_primes=prime_factors(gcd(c))
    for p in possible_primes:
        if (highest%p<>0) and (c[0]%(p^2)<>0):
            return p
    return 0
```

For example

```
Eisenstein(2*x^8+27*x^4+3*x^2+6)
```

yields 3. The expression c=b.coefficients() yields a list of coefficients of the polynomial, in order from lowest to highest degree. The expression c.pop() returns the last element in the list, and at the same time deletes that element from the list. If Eisenstein(p) is not [], then p is irreducible by Eisenstein's criterion.

Factorization over rational functions and over polynomials

10.8 Over the

   a. complex numbers,

   b. real numbers,

   c. rational numbers or

   d. integers,

   e. integer remainders modulo 2,

   f. integer remainders modulo 3,

is

$$b(x, y) = x^2 + y^2$$

reducible as a polynomial in two variables x, y?

Proposition 10.6 (Gauss' lemma)**.** *Suppose that $p(x, y)$ is a polynomial in two variables over a field, and that $p(x, y)$ factors as $p(x, y) = b(x, y)c(x, y)$, where $b(x, y)$ and $c(x, y)$ are polynomial in x with coefficients rational functions of y. Then $p(x, y)$ also factors in polynomials in x, y. To be precise, after perhaps multiplying $b(x, y)$ by a rational function of y, and $c(x, y)$ by its reciprocal, we can arrange that $b(x, y)$ and $c(x, y)$ are polynomials in both x and y.*

Proof. The coefficients in x on the right hand side of the equation $p(x, y) = b(x, y)c(x, y)$ are rational in y, hence are quotients of polynomials in y. Multiplying through by a common denominator we obtain an equation

$$d(y)p(x, y) = b(x, y)c(x, y)$$

with new $b(x, y), c(x, y)$ which are polynomials in x, y, and $d(y)$ is a nonzero polynomial. If $d(y)$ is constant, divide it into $c(x, y)$, and the result is proven.

So we can assume that $d(y)$ is nonconstant and write $d(y)$ as a product of irreducible polynomials

$$d(y) = d_1(y)d_2(y)\dots d_n(y).$$

Expand out $b(x, y)$ and $c(x, y)$ into powers of x:

$$b(x, y) = \sum b_j(y)x^j,$$

$$c(x, y) = \sum c_j(y)x^j.$$

Then $d_1(y)$ divides into all of the terms in

$$b(x, y)c(x, y) = \sum_n \sum_{j,k} b_j(y)c_k(y)x^{j+k}.$$

In particular, $d_1(y)$ divides into the lowest term $b_0(y)c_0(y)$, so into one of the factors, say into $b_0(y)$. Suppose that $d_1(y)$ divides into all of the terms $b_0(y), b_1(y), \dots, b_{j-1}(y)$, and also all of the terms $c_0(y), c_1(y), \dots, c_{k-1}(y)$. The x^{j+k} term has coefficient

$$\underline{b_0(y)}c_{j+k}(y) + \underline{b_1(y)}c_{j+k-1}(y) + \cdots + \underline{b_{j-1}(y)}c_{k+1}(y)$$
$$+ b_j(y)c_k(y) + b_{j+1}(y)\underline{c_{k-1}(y)} + \cdots + b_{j+k}(y)\underline{c_0(y)}.$$

So $d_1(y)$ divides into the underlined expressions, and must divide into one of $b_j(y)$ or $c_k(y)$. So we can increase the value of j, or the value of k, or both. By induction on both j and k, $d_1(y)$ divides into all terms in $b(x, y)$ (and hence divides into $b(x, y)$), or divides into all terms in $c(x, y)$ (and hence divides into $c(x, y)$). We cancel out a copy of $d_1(y)$ from both sides of the equation

$$d(y)p(x, y) = b(x, y)c(x, y)$$

and proceed by induction on the degree of $d(y)$.   □

Proposition 10.7 (Eisenstein's criterion)**.** *Take a polynomial $q(x, y)$ over a field. Expand in powers of x:*

$$q(x, y) = a_n(y)x^n + a_{n-1}(y)x^{n-1} + \cdots + a_1(y)x + a_0(y),$$

so that each $a_0(y), a_1(y), \dots, a_n(y)$ is a polynomial. Suppose that there is an irreducible polynomial $p(y)$ so that

a. *$p(y)$ does not divide the highest degree term $a_n(y)$ and*

b. *$p(y)$ divides all of the other terms $a_0(y), \dots, a_{n-1}(y)$ and*

c. *$p(y)^2$ does not divide the lowest coefficient $a_0(y)$.*

Then $q(x, y)$ is irreducible.

The proof is identical to the previous proof of Eisenstein's criterion.

$x^{13} + xy + y$ is irreducible, as it has a factor of y in each term in x, except the highest term.

10.9 Is
$$q(x) = 10000x^{10000} + 999x^{999} + 333x^{333} + 111$$
irreducible over the rational numbers?

10.10 Is
$$b(x,y) = (y+1)x^{10000} + (y^2+y+1)^{999}x^{999} + (y^2+y+1)^{333}x^{333} + y^2 + y + 1$$
irreducible as a polynomial in two variables x, y with coefficients

  a. remainders modulo 2?

  b. integers?

10.11 State and prove a Chinese remainder theorem for polynomials of one variable over any field. Apply it to find the smallest degree polynomial $p(x)$ so that, over the integer remainders modulo 2,

$$p(x) = x + 1 \quad (\mathrm{mod}\ x^2 + x + 1),$$
$$p(x) = x \qquad (\mathrm{mod}\ x^3 + x + 1).$$

Homogeneous polynomials

The *degree* of a term in a multivariable polynomial is the sum of the degrees in each of the variables; the *degree* of the polynomial is the highest degree of any nonzero term. A polynomial is *homogeneous* if all of its terms have the same degree.

10.12 Prove that a polynomial $b(x)$ of degree m in some variables $x_1, x_2, \ldots, x_n$ over an infinite field is homogeneous just when $b(tx) = t^m b(x)$ for every t. Give a counterexample over a finite field.

Lemma 10.8. *Over any field, every factor of a homogeneous polynomial is homogeneous.*

Proof. Write the polynomial factorized, as $b(x)c(x)$. The highest degree term in $b(x)$ multiplies by the highest degree term in $c(x)$ to give the highest degree term in $b(x)c(x)$, and the same for the lowest degree terms. Since all terms in $b(x)c(x)$ have the same degree, so do all terms in $b(x)$ and in $c(x)$. □

Factoring in many variables

Theorem 10.9. *Every polynomial $p(x_1, \ldots, x_n)$ with coefficients in a field admits a unique factorization*

$$p(x_1, \ldots, x_n) = p_1(x_1, \ldots, x_n) \ldots p_N(x_1, \ldots, x_n)$$

into irreducible factors.

Proof. We prove the result for two variables x, y, but the same proof works for any number of variables. Take two factorisations:

$$p(x,y) = p_1(x,y) \ldots p_M(x,y) = q_1(x,y) \ldots q_N(x,y).$$

If the highest order term in one of these $p_j(x, y)$ or $q_j(x, y)$ is in the x variable, then replace x, y by $x + \lambda y, y$, for some constant λ. But for a poor choice of λ, this might knock out the highest order term in y in another $p_j(x, y)$ or $q_j(x, y)$. To avoid this, we replace our field by the field of rational functions in an abstract variable λ, so no cancellations can take place. We can reverse the process by replacing x, y by $x - \lambda y, y$, so factorisations are preserved and reflected. If we can prove unique factorisation in this larger field, then any factorisation in the original field will still be unique in the original field. In particular, we can suppose that none of the factors is constant in y, and that the highest order term in y is constant in x. Thinking of p as a polynomial in y, with coefficients rational in x, by induction the factorisation in such functions is unique. So after perhaps reordering, each $p_j(x, y)$ is equal to some $q_j(x, y)$ up to some rational function $b_j(x)/c_j(x)$, with no common factor in numerator and denominator. But then $c_j(x)p_j(x, y) = b_j(x)q_j(x, y)$, and we expand out in powers of y to get $c_j(x)$ dividing into every coefficient in $q_j(x, y)$, so into the highest coefficient, which has no x in it.   □

10.13 Prove that every rational function over a field has an expression

$$\frac{b(x_1, \ldots, x_n)}{c(x_1, \ldots, x_n)}$$

as a ratio of polynomials with no common nonconstant factor, unique up to rescaling both numerator and denominator by the same nonzero constant.

<div align="center">

Chapter 11

Fields

</div>

> *A mathematician is a person who can find analogies between theorems; a better mathematician is one who can see analogies between proofs and the best mathematician can notice analogies between theories. One can imagine that the ultimate mathematician is one who can see analogies between analogies.*
>
> — Stefan Banach

Names for our favourite sets

The following are the standard names for various collections of numbers (or remainders):

$\mathbb{Z}$	the set of all integers,
$\mathbb{Q}$	the set of all rational numbers,
$\mathbb{R}$	the set of all real numbers,
$\mathbb{C}$	the set of all complex numbers,
$m\mathbb{Z}$	the set of all integer multiples of some number m,
$\mathbb{Z}/m\mathbb{Z}$	the set of all remainders modulo a positive integer m.

We can draw $\mathbb{R}$ as the number line,

and draw $\mathbb{Z}$ as equispaced dots on the number line,

and $\mathbb{C}$ as the plane:

<div align="center">

83

</div>
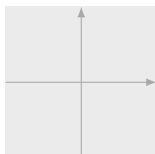
We could also draw the elements of $\mathbb{Z}/m\mathbb{Z}$ much like we would draw a clock:
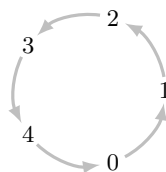
11.1 Explain how to add real numbers, how to add complex numbers, and how to add remainders modulo an integer, using these pictures. Explain how to multiply complex numbers using these pictures.

Rings

All of the above sets are examples of rings. A *ring* is a set of objects S together with two operations called *addition* and *multiplication*, and denoted by $b + c$ and bc, so that if b and c belong to S, then $b + c$ and bc also belong to S, and so that:

Addition laws:

    a. The associative law: For any elements a, b, c in S: $(a+b)+c = a+(b+c)$.

    b. The identity law: There is an element 0 in S so that for any element a in S, $a + 0 = a$.

    c. The existence of negatives: for any element a in S, there is an element b in S (denote by the symbol $-a$) so that $a + b = 0$.

    d. The commutative law: For any elements b, c in S, $b + c = c + b$.

Multiplication laws:

    a. The associative law: For any elements a, b, c in S: $(ab)c = a(bc)$.

The distributive law:

    a. For any elements a, b, c in S: $a(b + c) = ab + ac$ (left distributive) and $(b + c)a = ba + ca$ (right distributive).

> Clearly $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ and $\mathbb{Z}/m\mathbb{Z}$ (for any positive integer m) are rings.

> If X is a set, and S is a ring, then we can add and multiply any two functions mapping X to S by adding and multiplying the values of the functions as usual. The set T of all functions mapping X to S is a ring.

> If S is a ring, and x is an abstract variable (really just a pure symbol that we write down) then a *polynomial* in x with values in S is a formal expression we

write down of the form

$$a_0 + a_1 x + \cdots + a_n x^n$$

with coefficients $a_0, a_1, \ldots, a_n$ drawn from S. A polynomial is equal to 0 just when all of its coefficients are 0. Denote the set of all polynomials in x with coefficients from S as $S[x]$. Similarly, if we have two abstract variables x and y, we treat them as commuting: $xy = yx$, and then we define $S[x, y]$ to be the set of all polynomials in x and y with coefficients drawn from S. For example,

$$\frac{1}{3} - 2x + \frac{4}{7}y + 8x^2 y + y^3$$

belongs to $\mathbb{Q}[x, y]$.

If S is a ring, a *rational function* with coefficients drawn from S is a formal ratio
$$\frac{b(x)}{c(x)}$$
of polynomials $b(x), c(x)$ in $S[x]$ with $c(x)$ not the zero polynomial. We declare such a formal expression to be equal to

$$\frac{a(x)b(x)}{a(x)c(x)}$$

for any nonzero polynomial $a(x)$. Let $S(x)$ be the collection of all rational functions with coefficients in S in a variable x. *Danger:* even though we call $\frac{b(x)}{c(x)}$ a *rational function*, it is not, strictly speaking, actually a function, but only a formal expression. For example, if $S = \mathbb{Z}/2\mathbb{Z}$, then the rational function

$$\frac{1}{x(x + 1)}$$

is not defined for any value of x in $\mathbb{Z}/2\mathbb{Z}$. So we have to think of x not as a variable that varies over some numbers, but as an abstract symbol. Similarly, a rational function is really not a function but an abstract combination of symbols.

If x, y are two abstract variables, we similarly define $S(x, y)$ to be the ring of all rational functions in x, y, i.e. formal ratios of polynomials in x, y with coefficients drawn from S.

If R is a ring, $R\left[x, x^{-1}\right]$ means the ring whose elements are $b(x) + c\left(x^{-1}\right)$, polynomials in an abstract variable x and an abstract variable called x^{-1}, but so that when we multiply those two variables together, we get 1.

If R and S are rings, let $R \oplus S$ be the set of all pairs (r, s) where r is in R and s is in S. Define addition and multiplication by

$$(r_1, s_1) + (r_2, s_2) := (r_1 + r_2, s_1 + s_2),$$
$$(r_1, s_1)(r_2, s_2) := (r_1 r_2, s_1 s_2).$$

Similarly we define a finite (or even an infinite) sum of rings, as the set of finite (or infinite) sequences of elements, one from each ring, with obvious addition and multiplication, element at a time.

11.2 Let S be the set of all polynomials with real coefficients in one variable. To add elements of S, use usual polynomial addition. Take the "multiplication"operation to be composition of polynomials, *not* usual polynomial multiplication. Prove that S satisfies all of the laws above, except distributivity, and satisfies left but not right distributivity.

If a ring R is a subset of a ring S, and the addition and multiplication operations of R are just the ones of S applied to elements of R, then R is a *subring* of S. For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$, which is subring of $\mathbb{R}$, which is a subring of $\mathbb{C}$.

Special types of rings

A ring is *a ring with identity* if it satisfies the *identity law*: There is an element 1 in S so that for any element a in S, $a1 = a$.

A ring is *a division ring* if it satisfies the zero divisors law: for any elements a, b of S, if $ab = 0$ then $a = 0$ or $b = 0$.

A ring is *commutative* if it satisfies the *commutative law*: for any elements a, b of S, $ab = ba$. Note that we require *every* ring to have commutative addition; a *commutative* ring is one with commutative multiplication.

The associative law for addition, applied twice, shows that $(a + b) + (c + d) = a + (b + (c + d))$, and so on, so that we can add up any finite sequence, in any order, and get the same result, which we write in this case as $a + b + c + d$. A similar story holds for multiplication.

11.3 Let S be the set of all 2×2 matrices with integer coefficients, using the usual laws of matrix addition and multiplication as the addition and multiplication in S. Prove that S is a ring with identity, but *not* commutative.

11.4 Let $2\mathbb{Z}$ be the set of all even integers using the usual laws of integer addition and multiplication as the addition and multiplication in $2\mathbb{Z}$. Prove that $2\mathbb{Z}$ is a commutative ring but without identity.

11.5 Let $2S$ be the set of all 2×2 matrices with even integer coefficients, using the usual laws of matrix addition and multiplication as the addition and multiplication in $2S$. Prove that $2S$ is a noncommutative ring without identity.

11.6 Suppose that S is a ring with identity in which $0 = 1$. Prove that $S = \{0\}$.

From here on, we will only be interested in commutative rings with identity.

Fields

A *unit* in a ring is an element that has a multiplicative inverse, i.e. a reciprocal.

> In $\mathbb{Z}/7\mathbb{Z}$ every nonzero element is a unit, because 7 is a prime number.

> In $\mathbb{Z}$ the units are ± 1.

> If S is a ring, we denote by $S^\times$ the set of units of S.

> $(\mathbb{Q}[x])^\times = \mathbb{Q}^\times$: no polynomial of positive degree has a reciprocal polynomial.

> Every nonzero rational function is a unit in $\mathbb{Q}(x)$: $(\mathbb{Q}(x))^\times = \mathbb{Q}(x) - \{\, 0 \,\}$.

11.7 Prove that, for any ring S with identity, if b and c are units then bc is a unit.

11.8 Suppose that R is a ring and that $b \in R$ is both a zero divisor and a unit. Prove that R has exactly one element, the element zero.

A *field* is a commutative ring with identity, in which every nonzero element is a unit, i.e. has a multiplicative inverse, and in which $1 \neq 0$.

Lemma 11.1. *For any positive integer m, $\mathbb{Z}/m\mathbb{Z}$ is a field if and only if m is prime.*

Proof. If m factors, say as $m = bc$, with $b, c \geq 2$, then $b \neq 0$ modulo m, but $bc = 0$ modulo m, so both b and c are zero divisors, and therefore are not units. If m does not factor, then m is prime, and we know that every nonzero element of $\mathbb{Z}/m\mathbb{Z}$ is a unit; we even have a recipe to find its multiplicative inverse. $\square$

> The set $\mathbb{Q}$ of rational numbers is a field, as is the set $\mathbb{R}$ of real numbers and the set $\mathbb{C}$ of complex numbers.

> Any field k lies inside the field $k(x)$ of rational functions with coefficients in k. In particular, every field is contained inside an infinite field.

11.9 Note that in the field $k = \mathbb{Z}/13\mathbb{Z}$, we have the peculiar equation $13 = 0$. The *characteristic* of a ring with identity R is the smallest positive integer $c > 0$ so that c is zero in R (or in other words, if $e \in R$ is the identity element, then the element $c \cdot e = 0 \in R$). If there is no such positive number c, we say that R has characteristic zero. Prove that the characteristic of a division ring R is prime, say p, and that there is a unique map $\mathbb{Z}/p\mathbb{Z} \to R$ which takes 0 to 0, 1 to 1, and sums to sums and products to products.

11.10 A *zero divisor* r in a ring R is an element so that $rs = 0$ for some element s of R. Suppose that k is a field. Find the (i) zero divisors and (ii) the units in

a. k

b. $k[x]$

c. $k\left[x, x^{-1}\right]$

d. $k[x] \oplus k[y]$.

Let k be the set of all real numbers of the form $a + b2^{1/3} + c2^{2/3}$ for any a, b, c rational. Clearly all rational numbers (in particular 0 and 1) have this form. It is easy to check that if add any two such numbers, or subtract, we get another such. If we multiply two such numbers:

$$(a + b2^{1/3} + c2^{2/3})(p + q2^{1/3} + r2^{2/3}) = \alpha + \beta 2^{1/3} + \gamma 2^{2/3}$$

where

$$\alpha = ap + 2br + 2cq,$$
$$\beta = aq + bp + 2cr,$$
$$\gamma = ar + cp + bq.$$

We need to check division; this turns out to be difficult. Associate to each such number $a + b2^{1/3} + c2^{2/3}$ the vector (a, b, c) in $\mathbb{Q}^3$. The product $(a + b2^{1/3} + c2^{2/3})(p + q2^{1/3} + r2^{2/3})$ has associated vector

$$\begin{pmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{pmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix}.$$

The determinant of the matrix

$$A = \begin{pmatrix} a & 2c & 2b \\ b & a & 2c \\ c & b & a \end{pmatrix}$$

is $a^3 + 2b^3 + 4c^3 - 6abc$. We need to prove that this determinant is nonzero, i.e. that there is an inverse matrix. Indeed the reader can check that

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} \bar{a} & 2\bar{c} & 2\bar{b} \\ \bar{b} & \bar{a} & 2\bar{c} \\ \bar{c} & \bar{b} & \bar{a} \end{pmatrix}$$

where $\bar{a} = a^2 - 2bc$, $\bar{b} = 2c^2 - ab$, $\bar{c} = b^2 - ac$. Note that this tells us precisely that

$$\frac{1}{a + b2^{1/3} + c2^{2/3}} = \frac{\bar{a} + \bar{b}2^{1/3} + \bar{c}2^{2/3}}{\det A}.$$

So in order to see that k is a field, and check the above expression for division in k, we only need to check $\det A \neq 0$. In other words, let's check that if a, b, c are rational numbers, not all zero, then $a^3 + 2b^3 + 4c^3 \neq 6abc$. Note that we can rescale all of a, b, c by the same constant, and both sides of the inequality $a^3 + 2b^3 + 4c^3 \neq 6abc$ scale by the same factor. So we can scale a, b, c so that they become coprime integers. If $a^3 + 2b^3 + 4c^3 = 6abc$, clearly a^3 is even, so a is even, say $a = 2\alpha$. Then $a^3 + 2b^3 + 4c^3 = 6abc$ becomes

$8\alpha^3 + 2b^3 + 4c^3 = 12\alpha bc$. Divide out 2 to get $4\alpha^3 + b^3 + 2c^3 = 6\alpha bc$. Clearly b is even, say $b = 2\beta$; plug in: $4\alpha^3 + 8\beta^3 + 2c^3 = 12\alpha\beta c$. Divide out 2 to get $2\alpha^3 + 4\beta^3 + c^3 = 6\alpha\beta c$. Finally we see that c is even, contradicting the fact that a, b, c are coprime integers. Hence we conclude that k is a field.

Linear algebra

Most of linear algebra works identically over any field: matrices, Gaussian elimination, invertibility, solvability of equations, determinants, and the theory of eigenvalues and eigenvectors.

11.11 Let k be the Boolean numbers $k := \mathbb{Z}/2\mathbb{Z}$, and A the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

thought of as having entries from k. Is A invertible? If so, find A^{-1}.

11.12 If A is a matrix whose entries are rational functions of a variable t over a field k, prove that the rank of A is constant in t, except for finitely many values of t.

11.13 If A is a matrix whose entries are integers, let A_p be the same matrix, but with coefficients taken modulo a prime p. Prove that A is invertible over the rational numbers (i.e. has an inverse matrix with rational number entries) just when, for all but finitely many primes p, the matrix A_p is invertible over $\mathbb{Z}/p\mathbb{Z}$. Prove that A is not invertible over the rational numbers just when, for every prime number p, A_p is not invertible over $\mathbb{Z}/p\mathbb{Z}$. Prove that A is invertible over the integers just when, for every prime number p, the matrix A_p is invertible over $\mathbb{Z}/p\mathbb{Z}$.

Chapter 12

Field extensions

> *Mathematics, rightly viewed, possesses not only truth, but supreme*
> *beauty—a beauty cold and austere, like that of sculpture, without ap-*
> *peal to any part of our weaker nature, without the gorgeous trappings of*
> *painting or music, yet sublimely pure, and capable of a stern perfection*
> *such as only the greatest art can show.*
>
> — Bertrand Russell

Quotienting a polynomial

We return to polynomials in a single variable x. We write that $b(x) = c(x)$ modulo a polynomial $p(x)$ to mean that $b(x) - c(x)$ is a multiple of $p(x)$. Just as for integers, we add modulo $p(x)$, subtract modulo $p(x)$, and multiply modulo $p(x)$, all of which makes sense since multiples of $p(x)$ add, subtract, and multiply to give multiples of $p(x)$. When we work modulo $p(x)$, write $b(x)^{-1}$ to mean a remainder so that $b(x)b(x)^{-1} = 1$ modulo $p(x)$. To find $b(x)^{-1}$, we use Bézout coefficients.

Let $p(x) := x^3 + 2x + 1$ over the field of rational numbers. Modulo $p(x)$, clearly $x^3 = -2x - 1$. It is common to use a Greek letter, like α, for the remainder of x modulo $p(x)$, instead of calling it x. So remainders modulo $p(x)$ are just expressions like $\alpha, 7 - 4\alpha, \alpha^2/3$, but when we compute, we reduce modulo $\alpha^3 + 2\alpha + 1$, i.e. we change any α^3 to $\alpha^3 = -2\alpha - 1$. So, for example,

$$
\begin{aligned}
(\alpha^2 + 1)(\alpha^2 + \alpha + 1) &= \alpha^4 + \alpha^3 + 2\alpha^2 + \alpha + 1, \\
&= \alpha^3\alpha + \alpha^3 + 2\alpha^2 + \alpha + 1, \\
&= (-2\alpha - 1)\alpha + (-2\alpha - 1) + 2\alpha^2 + \alpha + 1, \\
&= -2\alpha^2 - \alpha - 2\alpha - 1 + 2\alpha^2 + \alpha + 1, \\
&= -2\alpha.
\end{aligned}
$$

Let $p(x) := x^2 + x + 1$ with coefficients over the field of remainders modulo 2. Write the remainder of x modulo $p(x)$ as α. To find α^{-1}, compute Bézout

coefficients:

$$\begin{pmatrix} 1 & 0 & x \\ 0 & 1 & x^2 + x + 1 \end{pmatrix}, \text{ add } (x+1)(\text{row 1}) \text{ to row 2,}$$

$$\begin{pmatrix} 1 & 0 & x \\ x+1 & 1 & 1 \end{pmatrix}, \text{ add } x(\text{row 2}) \text{ to row 1,}$$

$$\begin{pmatrix} x^2 + x + 1 & x & 0 \\ x+1 & 1 & 1 \end{pmatrix}.$$

So the Bézout coefficients are

$$(x+1)(x) + (1)(x^2 + x + 1) = 1.$$

Modulo $x^2 + x + 1$, we find

$$(\alpha + 1)\alpha = 1$$

or in other words $\alpha^{-1} = \alpha + 1$.

12.1 Work out the complete multiplication table for the remainders of polynomials in x when we quotient out by $x^2 + x$ over the field of remainders modulo 3. Which elements have reciprocals?

12.2 With coefficients being integers modulo 3, find x^{-1} modulo $x^9 + 2x^2 + 1$.

12.3 Guess: what do you think the Chinese remainder theorem might be for remainders modulo polynomials, instead of working with remainders modulo integers?

If $p(x) = x^2$, and again letting α be the remainder of x modulo $p(x)$, then quotienting out $p(x)$ yields remainders of the form $b + c\alpha$, since $\alpha^2 = 0$. The remainder α is very much like the "very small quantities" that physicists talk about, so small that the square is negligibly small and can be dropped from calculations.

If $p(x) = x(x-1)$, and again letting α be the remainder of x modulo $p(x)$, then quotienting out $p(x)$ yields remainders of the form $b + c\alpha$, but with $\alpha(\alpha - 1) = 0$, so $\alpha^2 = \alpha$. Think of α as a number which can't decide whether it wants to be zero or one, and is somehow behaving like both zero and one at the same time.

Field extensions

If k and K are fields and k is contained in K, with the same identity element, zero element, addition and multiplication operations, then k is a *subfield* of K and K is an *extension* of k.

> The complex numbers are a field extension of the real numbers.

> The real numbers are a field extension of the rational numbers.

Consider how we could write down field extensions. If $k \subset K$ is a subfield, a *basis* of K over k is a collection of elements $\alpha_1, \alpha_2, \ldots, \alpha_n$ of K so that every element of K has the form

$$a_1 \alpha_1 + a_2 \alpha_2 + \cdots + a_n \alpha_n$$

for some coefficients

$$a_1, a_2, \ldots, a_n$$

from k. The *degree* of K over k is the smallest number of elements in a basis.

> We can write every element of $\mathbb{C}$ as $x \cdot 1 + y \cdot i$, for some real coefficients x, y. So $\mathbb{C}$ is a degree 2 extension of $\mathbb{R}$ with $1, i$ as a basis.

> Clearly $1, \sqrt{2}$ is a basis for $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$.

> The real numbers $\mathbb{R}$ as an extension of $\mathbb{Q}$ has no basis; it is an infinite degree extension. This is a difficult fact to prove.

Adding a root

> Let's return to the example on page 91. Look for roots of $p(x) := x^2 + x + 1$ over the finite field $k := \mathbb{Z}/2\mathbb{Z}$. Try $x = 0$:
>
> $$p(0) = 0^2 + 0 + 1 = 1,$$
>
> no good. Try $x = 1$:
>
> $$p(1) = 1^2 + 1 + 1 = 1 + 1 + 1 = 1,$$
>
> since $1 + 1 = 0$. No good. So $p(x)$ has no roots in k.
> Let K be the ring of remainders modulo $p(x)$. As before, we write the remainder of x modulo $p(x)$ as α. Modulo $p(x) = x^2 + x + 1$, of course $x^2 = x + 1$, i.e. $\alpha^2 = \alpha + 1$. So any polynomial of degree 2 or more in α reduces to one of degree 1, by replacing α^2 by $\alpha + 1$. So every element of K is a polynomial of degree 0 or 1 in α, with coefficients 0 or 1:
>
> $$\{0, 1, \alpha, \alpha + 1\} = K.$$

Using only the fact that $1 + 1 = 0$, compute out the addition table

$+$	0	1	α	$\alpha + 1$
0	0	1	α	$\alpha + 1$
1	1	0	$\alpha + 1$	α
α	α	$\alpha + 1$	0	1
$\alpha + 1$	$\alpha + 1$	α	1	0

Using only the fact that $\alpha^2 = \alpha + 1$, compute out the multiplication table:

$\cdot$	0	1	α	$\alpha + 1$
0	0	0	0	0
1	0	1	α	$\alpha + 1$
α	0	α	$\alpha + 1$	1
$\alpha + 1$	0	$\alpha + 1$	1	α

Looking for reciprocals, we find that

$$\frac{1}{0} \text{ does not exist,}$$

$$\frac{1}{1} = 1,$$

$$\frac{1}{\alpha} = \alpha + 1,$$

$$\frac{1}{\alpha + 1} = \alpha.$$

Every nonzero element has a reciprocal: K is a field. So $K = \{0, 1, \alpha, \alpha + 1\}$ is an extension of k. The polynomial $p(x)$ splits over this field: $p(x) = (x + \alpha)(x + 1 + \alpha)$.

Theorem 12.1. *Suppose that $p(x)$ is a polynomial over a field k, Let α be the remainder of x modulo $p(x)$, and write $k[\alpha]$ for the ring of remainders modulo $p(x)$. Thinking of $p(x)$ as a polynomial, $p(x)$ has a root in $k[\alpha]$: $p(\alpha) = 0$. The ring $k[\alpha]$ is a field just when $p(x)$ is irreducible over k. If this happens, then $k(\alpha) = k[\alpha]$ and $k(\alpha)$ is an extension field of k and $p(x)$ has a linear factor over the field $k(\alpha)$.*

Proof. By definition, $p(\alpha)$ means $p(x)$ but with x replaced by α, i.e. by the remainder of x modulo $p(x)$, i.e. by x, but quotienting out any copies of $p(x)$ from the resulting expression $p(x)$, i.e. $p(\alpha)$ is just 0. Every element of $k[\alpha]$ is a polynomial $b(\alpha)$ in α, but of degree less than the degree of $p(x)$, since we can quotient out any higher order terms by $p(x)$. So $k[\alpha]$ is a field just when any nonzero $b(\alpha)$ in $k[\alpha]$ has a reciprocal in $k[\alpha]$.

Suppose that $p(x)$ is irreducible. Take Bézout coefficients for $b(x), p(x)$: $s(x)b(x) + t(x)p(x) = d(x)$, so $d(x)$ is the greatest common factor of $b(x), p(x)$. Then $b(x), p(x)$ have greatest common divisor 1, i.e. $d(x) = 1$ (after perhaps a constant rescaling), so mod out $p(x)$ to get $s(\alpha)b(\alpha) = 1$, i.e. $b(\alpha)$ has a reciprocal. So every nonzero polynomial in α has a reciprocal, i.e. $k(\alpha) = k[\alpha]$.

Suppose that $p(x)$ is reducible, say $p(x) = b(x)c(x)$, into lower degree factors, so $0 = b(\alpha)c(\alpha)$. We want to prove that $b(\alpha)$ has no reciprocal. If $b(\alpha)$ has a reciprocal, multiply both sides by it to get $0 = c(\alpha)$, i.e. $c(x)$ is a multiple of $p(x)$, but has degree less than that of $p(x)$, so $c(x) = 0$, so $p(x) = b(x)c(x) = 0$, a contradiction. $\qquad \square$

We say that the field $k[\alpha]$ arises from k by "adding a root" α to the polynomial $p(x)$. Careful: the trick only works if $p(x)$ is irreducible. If $p(x)$ is reducible, split it into irreducible factors, and you can add roots as you like for each factor.

> This theorem explains why, in our last example, adding a root to $p(x) = x^2 + x + 1$ over $k = \mathbb{Z}/2\mathbb{Z}$ gave a field $K = \{0, 1, \alpha, \alpha + 1\}$, not just a ring.

Splitting fields

A polynomial $p(x)$ with coefficients in a field k *splits* if it is a product of linear factors. If $k \subset K$ is a subfield, a polynomial $p(x)$ in $k[x]$ *splits over* K if it splits into a product of linear factors when we allow the factors to have coefficients from K.

> The real-coefficient polynomial $x^2 + 1$ splits over $\mathbb{C}$:
> $$x^2 + 1 = (x - i)(x + i).$$

12.4 Denote by $\mathbb{Q}(\sqrt{2})$ the field extension of $\mathbb{Q}$ generated by $\sqrt{2}$, i.e. the smallest subfield of $\mathbb{C}$ containing $\sqrt{2}$. Prove that $\mathbb{Q}(\sqrt{2})$ consists precisely of the numbers $b + c\sqrt{2}$ for b, c rational. Prove that $\mathbb{Q}(\sqrt{2})$ is a splitting field for $x^2 - 2$ over $\mathbb{Q}$.

An extension K of a field k is a *splitting field* of a polynomial $p(x)$ over k if $p(x)$ splits over K and every element of K is expressible as a rational function (with coefficients from k) of the roots of $p(x)$.

Theorem 12.2. *Every polynomial $p(x)$ over any field k has a splitting field K. If $p(x)$ has degree n then K has degree at most $n!$ over k.*

Proof. Split $p(x)$ into irreducible factors, and add a root of one of them. When you add a root, the elements of the extension field are polynomials of degree less than n, so the degree of extension is at most n, equal to n just when $p(x)$ is itself irreducible. In the extension, $p(x)$ splits off a linear factor $p(x) = (x - \alpha)q(x)$, so the next time you add a root, you are adding it to some $q(x)$ of lower degree. By induction, the total degree is at worst $n!$. $\qquad \square$

The splitting field is more-or-less unique, but this is difficult to make precise; see theorem .

> Over $k = \mathbb{R}$ the polynomial $p(x) = x^2 + 1$ has splitting field $\mathbb{C}$:
> $$x^2 + 1 = (x - i)(x + i).$$

Every polynomial over $\mathbb{C}$ splits into linear factors, so for any subfield $k \subset \mathbb{C}$ the splitting field K of any polynomial over k lies inside $\mathbb{C}$, a field of complex numbers.

12.5 Consider the polynomial

$$p(x) = x^3 + x^2 + 1$$

over the field $k = \mathbb{Z}/2\mathbb{Z}$. Suppose that that splitting field K of $p(x)$ contains a root α of $p(x)$. Prove that α^2 and $1 + \alpha + \alpha^2$ are the two other roots. Compute the addition table and the multiplication table of the 8 elements

$$0, 1, \alpha, 1 + \alpha, \alpha^2, 1 + \alpha^2, \alpha + \alpha^2, 1 + \alpha + \alpha^2.$$

Use this to prove that

$$K = \left\{ 0, 1, \alpha, 1 + \alpha, \alpha^2, 1 + \alpha^2, \alpha + \alpha^2, 1 + \alpha + \alpha^2 \right\}$$

so K is a finite field with 8 elements.

Algebraic closure

A field k is *algebraically closed* if every polynomial over k splits into linear factors in k. An extension K of a field k is a *splitting field* of a collection of polynomials over k if every one of those polynomials splits over K and K is generated by the roots of all of these polynomials put together. If that collection consists of *all* of the polynomials defined over k, then we say that K is an *algebraic closure* of k. We won't prove:

Theorem 12.3. *Every field k has an algebraic closure, denoted $\bar{k}$, unique up to an isomorphism which is the identity map on k. The field $\bar{k}$ is algebraically closed.*

Clearly $\bar{\mathbb{R}} = \mathbb{C}$.

Danger: $\bar{\mathbb{Q}} \neq \mathbb{C}$. The number $\pi = 3.14\ldots$ does not satisfy any polynomial with rational coefficients, so π belongs to $\mathbb{C}$ but doesn't belong to $\bar{\mathbb{Q}}$. However, $\bar{\mathbb{Q}}$ is a subfield of $\mathbb{C}$. Sit $\mathbb{Q}$ inside $\mathbb{C}$ as usual. Take the subfield of $\mathbb{C}$ generated by all roots of all polynomials with rational coefficients, all roots of polynomials with all of those as coefficients, and so on, and take the union K of all such fields. By the theorem, K is isomorphic to $\bar{\mathbb{Q}}$, by an isomorphism fixing every rational number.

If k is any finite field, say with elements

$$k = \left\{ \alpha_1, \alpha_2, \ldots, \alpha_n \right\},$$

then let

$$p(x) := 1 + (x - \alpha_1)(x - \alpha_2)\ldots(x - \alpha_n).$$

Clearly $p(\alpha) = 1$ for any $\alpha \in k$. Therefore k is not algebraically closed: every algebraically closed field is infinite.

12.6 Prove that, for any finite field, and any integer, there is a finite extension of that field which has more elements than that integer.

<center>Chapter 13</center>

Resultants and discriminants

The resultant

For any polynomial

$$b(x) = b_0 + b_1 x + \cdots + b_m x^m,$$

denote its vector of coefficients as

$$\vec{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Clearly the vector of coefficients of $xb(x)$ is the same, with an extra row added to the top, with a zero in that row. When we multiply two polynomials $b(x)$ and

$$c(x) = c_0 + c_1 x + \cdots + c_n x^n$$

the vector of coefficients of $b(x)c(x)$ is

$$\begin{pmatrix} b_0 & & & \\ b_1 & b_0 & & \\ \vdots & \ddots & \ddots & \\ b_m & \ddots & \ddots & b_0 \\ & b_m & \ddots & b_1 \\ & & \ddots & \vdots \\ & & & b_m \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix},$$

(with zeroes represented as blank spaces) and this $(m + n + 1) \times (n + 1)$ matrix we denote by

$$[b] = \begin{pmatrix} b_0 & & & \\ b_1 & b_0 & & \\ \vdots & \ddots & \ddots & \\ b_m & \ddots & \ddots & b_0 \\ & b_m & \ddots & b_1 \\ & & \ddots & \vdots \\ & & & b_m \end{pmatrix}.$$

So the vectors of coefficient are related by $\overrightarrow{bc} = [b]\vec{c}$.

<center>99</center>

The *resultant* of two polynomials

$$b(x) = b_0 + b_1 x + \cdots + b_m x^m \text{ and}$$
$$c(x) = c_0 + c_1 x + \cdots + c_n x^n,$$

denoted $\text{res}_{b,c}$, is the determinant of the matrix $\left([b] \quad [c]\right)$, given by stacking the matrices $[b]$ and $[c]$ of suitable sizes beside one another, to get a square matrix of size $(m+n) \times (m+n)$. To remember the sizes: the number of columns in $[b]$ is the degree of c, and vice versa.

If

$$b(x) = 4 + 3x + 7x^2 + x^3,$$
$$c(x) = 5 + 2x,$$

then $\text{res}_{b,c}$ is the determinant of the 4×4 matrix

$$\begin{pmatrix} 4 & 5 & 0 & 0 \\ 3 & 2 & 5 & 0 \\ 7 & 0 & 2 & 5 \\ 1 & 0 & 0 & 2 \end{pmatrix},$$

which is $\text{res}_{b,c} = 197$.

If we swap b and c, we swap columns, so clearly $\text{res}_{c,b} = (-1)^{mn} \text{res}_{b,c}$.

Take $b(x) = x^2 + 4x + 7$ and $c(x) = x^3$ and compute

$$\text{res}_{b,c} = \det \begin{pmatrix} 7 & 0 & 0 & 0 & 0 \\ 4 & 7 & 0 & 0 & 0 \\ 1 & 4 & 7 & 0 & 0 \\ 0 & 1 & 4 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} = 7^3.$$

The easiest examples of arbitrary degrees arise when we take any $b(x)$ but try $c(x) = x^n$:

$$\text{res}_{b,c} = \det \begin{pmatrix} b_0 & & & & & \\ \vdots & b_0 & & & & \\ \vdots & \ddots & \ddots & & & \\ b_m & \ddots & \ddots & b_0 & & \\ & b_m & \ddots & \vdots & 1 & \\ & & \ddots & \vdots & & \ddots \\ & & & b_m & & 1 \end{pmatrix} = b_0^n.$$

Sage

Computing resultants in sage requires us to tell sage what sort of numbers arise as the coefficients of our polynomials.

```
P.<x> = PolynomialRing(QQ)
a=x^3+x+7
b=x^2+2
a.resultant(b)
```

yields 51. The expression P.<x> = PolynomialRing(QQ) tells sage that we are working with polynomials with rational coefficients (QQ means rational) in a variable x.

We could write our own resultant function, just to see how it might work:

```
def resultant(b,c):
    B=b.list()
    C=c.list()
    m=len(B)-1
    n=len(C)-1
    A=matrix(m+n,m+n)
    for j in range(0,n):
        for i in range(0,m+n):
            if (0<=i-j) and (i-j<=m):
                A[i,j]=B[i-j]
            else:
                A[i,j]=0
    for j in range(n,m+n):
        for i in range(0,m+n):
            if (0<=i-j+n) and (i-j+n<=n):
                A[i,j]=C[i-j+n]
            else:
                A[i,j]=0
    return det(A)
```

Try it out:

```
t=var('t')
p = t^2+2*t+1
q = t+1
resultant(p,q)
```

yields 0.

Common factors and resultants

Lemma 13.1. *Two polynomials $b(x)$ and $c(x)$ of degrees m and n, over any field, have a nonconstant common factor just when*

$$0 = u(x)b(x) + v(x)c(x)$$

for two polynomials $u(x)$ and $v(x)$ of degrees $n - 1$ and $m - 1$ at most, not both zero.

Proof. Two polynomials $b(x)$ and $c(x)$ of degrees m and n have a nonconstant common factor just when we can write the polynomials factored, say as $b(x) = v(x)d(x)$ and $c(x) = -u(x)d(x)$. But then

$$0 = u(x)b(x) + v(x)c(x).$$

On the other hand, suppose that

$$0 = u(x)b(x) + v(x)c(x)$$

for two polynomials $u(x)$ and $v(x)$ of degrees $n-1$ and $m-1$ at most. Suppose that $b(x)$ and $c(x)$ have no common factor, so their greatest common divisor is 1. Write out Bézout coefficients

$$1 = s(x)b(x) + t(x)c(x)$$

and compute

$$\begin{aligned} v &= vsb + vtc, \\ &= vsb + tvc, \\ &= vsb - tub, \\ &= b\,(vs - tu)\,. \end{aligned}$$

But the degree of v is smaller than that of b, so $v = 0$. Swapping roles of b, c and of u, v and of s, t, we get $u = 0$. $\qquad\square$

Proposition 13.2. *Two polynomials, over any field, have a nonconstant common factor just when their resultant vanishes.*

Proof. Take two polynomials

$$\begin{aligned} u(x) &= u_0 + u_1 x + \cdots + u_{n-1} x^{n-1}, \\ v(x) &= v_0 + v_1 x + \cdots + v_{m-1} x^{m-1}. \end{aligned}$$

The vector of coefficients of $u(x)b(x) + v(x)c(x)$ is

$$\overrightarrow{bu + cv} = \begin{pmatrix} [b] & [c] \end{pmatrix} \begin{pmatrix} \vec{u} \\ \vec{v} \end{pmatrix}$$

By linear algebra, the determinant vanishes just when the matrix has a nonzero null vector, i.e. a nonzero vector in its kernel. So the resultant vanishes just when there is a choice of coefficients

$$u_0, u_1, \ldots, v_0, v_1, \ldots,$$

not all zero, so that $u(x)b(x) + v(x)c(x) = 0$. In other words, there are polynomials $u(x)$ and $v(x)$ of required degrees so that $u(x)b(x) + v(x)c(x) = 0$. The theorem now follows from lemma 13.1 . $\qquad\square$

13.1 Find the resultants, and use them to decide if there are nonconstant common factors.

    a. $x^2 - 5x + 6, x^3 - 3x^2 + x - 3$

    b. $x^2 + 1, x - 1$

The resultant is brutal to compute, even for polynomials of fairly small degree. Its importance, like that of the determinant of a matrix, arises from its theoretical power. At this point, the reader should be annoyed: we already know how to find common divisors, even the greatest common divisor, by a fast calculation, so we surely don't need a slow resultant calculation to see if there is a common divisor of positive degree, since we can already actually find that divisor, and more quickly. The reader is right, but surprisingly we will find many uses for the resultant.

13.2 Over the field of remainders modulo 2, calculate the resultant of $b(x) = x^3 + x^2 + x + 1, c(x) = x^3 + x$. *Warning*: the 6×6 determinant you get is actually easier than it looks at first, so look for tricks to find it without any hard calculation.

13.3 Calculate a resultant to prove that, for any prime integer $p \geq 2$, when we work over the field of remainders modulo p, the polynomials $x^2 + 1$ and $x^2 + 3x$ have a common factor just exactly when $p = 2$ or $p = 5$.

13.4 Suppose that $b(x)$ and $c(x)$ are polynomials in one variable x, with coefficients in the field of integer remainders modulo 2. Suppose that $b(x)$ is a sum of 1126 nonzero terms, while $c(x)$ is a sum of 8080 nonzero terms. Note: we don't say anything about the degree of either one. Prove that they have a common factor of positive degree.

Lemma 13.3. *Take two polynomials $b(x), c(x)$ in a variable x of degrees m, n. The resultant $r = \mathrm{res}_{b,c}$ is expressible as $r = u(x)b(x) + v(x)c(x)$ where $u(x)$ and $v(x)$ are polynomials of degrees $n - 1, m - 1$. One can make a particular choice of $u(x), v(x)$ so that the coefficients of $u(x)$ and of $v(x)$ are expressible as polynomial expressions in the coefficients of $b(x)$ and $c(x)$, and those polynomial expressions have only 1 or -1 as coefficient in each term.*

Proof. Look at our matrix:

$$
r = \det \begin{pmatrix}
b_0 & & & & c_0 & & & \\
b_1 & b_0 & & & c_1 & c_0 & & \\
\vdots & b_1 & \ddots & & \vdots & c_1 & \ddots & \\
\vdots & \ddots & \ddots & b_0 & \vdots & \ddots & \ddots & c_0 \\
b_m & \ddots & \ddots & b_1 & c_n & \ddots & \ddots & c_1 \\
& b_m & \ddots & \vdots & & c_n & \ddots & \vdots \\
& & \ddots & \vdots & & & \ddots & \vdots \\
& & & b_m & & & & c_n
\end{pmatrix}
$$

Add to the first row the second multiplied by x and then the third multiplied by x^2

and so on, which doesn't change the determinant:

$$r = \det \begin{pmatrix} b(x) & xb(x) & \dots & x^{n-1}b(x) & c(x) & xc(x) & \dots & x^{m-1}c(x) \\ b_1 & b_0 & & & c_1 & c_0 & & \\ \vdots & b_1 & \ddots & & \vdots & c_1 & \ddots & \\ \vdots & \ddots & \ddots & b_0 & \vdots & \ddots & \ddots & c_0 \\ b_m & \ddots & \ddots & b_1 & c_n & \ddots & \ddots & c_1 \\ & b_m & \ddots & \vdots & & c_n & \ddots & \vdots \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & b_m & & & & c_n \end{pmatrix}.$$

Expand the determinant across the top row to see that

$$r = u(x)b(x) + v(x)c(x)$$

where

$$u(x) = \det \begin{pmatrix} 1 & x & \dots & x^{n-1} & & & & \\ b_1 & b_0 & & & c_1 & c_0 & & \\ \vdots & b_1 & \ddots & & \vdots & c_1 & \ddots & \\ \vdots & \ddots & \ddots & b_0 & \vdots & \ddots & \ddots & c_0 \\ b_m & \ddots & \ddots & b_1 & c_n & \ddots & \ddots & c_1 \\ & b_m & \ddots & \vdots & & c_n & \ddots & \vdots \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & b_m & & & & c_n \end{pmatrix}$$

and

$$v(x) = \det \begin{pmatrix} & & & & 1 & x & \dots & x^{m-1} \\ b_1 & b_0 & & & c_1 & c_0 & & \\ \vdots & b_1 & \ddots & & \vdots & c_1 & \ddots & \\ \vdots & \ddots & \ddots & b_0 & \vdots & \ddots & \ddots & c_0 \\ b_m & \ddots & \ddots & b_1 & c_n & \ddots & \ddots & c_1 \\ & b_m & \ddots & \vdots & & c_n & \ddots & \vdots \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & b_m & & & & c_n \end{pmatrix}.$$

Note that we can expand out any determinant using only multiplication and addition of entries and some $\pm$ signs. $\square$

If we scale up $b(x)$ by a factor λ, the matrix gets a λ in each column in the left n columns, where n is the degree of $c(x)$. The resultant of $b(x), c(x)$ scales by λ^n. So practically we only need to compute resultants of monic polynomials.

Proposition 13.4. *Given two monic polynomials over any field, each factored into linear factors,*

$$b(x) = (x - \beta_1)(x - \beta_2) \ldots (x - \beta_m),$$
$$c(x) = (x - \gamma_1)(x - \gamma_2) \ldots (x - \gamma_n),$$

then the resultant is

$$\operatorname{res}_{b,c} = (\gamma_1 - \beta_1)(\gamma_1 - \beta_2) \ldots (\gamma_n - \beta_m),$$
$$= b(\gamma_1) b(\gamma_2) \ldots b(\gamma_n).$$
$$= (-1)^{mn} c(\beta_1) c(\beta_2) \ldots c(\beta_m),$$

In particular, the resultant vanishes just when $b(x)$ and $c(x)$ have a common root.

Proof. If we expand out the expression for $b(x)$ into a sum of monomials, with coefficients being expressed in terms of these β_i, and similarly for $c(x)$, then the resultant is a huge polynomial expression in terms of the various β_i and γ_j. This polynomial vanishes whenever $\beta_i = \gamma_j$. Thinking of β_i and γ_j as abstract variables, the expression $\gamma_j - \beta_i$ is a linear function. Assume for the moment that our field is infinite. Then the resultant is divisible by $\gamma_j - \beta_i$, for any i and j, by lemma 18.2 on page 152. Therefore the resultant is divisible by the product

$$(\gamma_1 - \beta_1)(\gamma_1 - \beta_2) \ldots (\gamma_n - \beta_m).$$

Imagine expanding out the resultant as a determinant, with $b(x), c(x)$ monic. Each term in the determinant contains exactly n of the various b_j, and each b_j has $m - j$ β's. So the highest term in the β's is the one from the diagonal. The diagonal has $b_0, b_0, \ldots, b_0, 1, 1, \ldots, 1$, giving a term b_0^n in the determinant. Expand in β's as $b_0 = (-1)^m \beta_1 \ldots \beta_m$ to give a term $(-1)^{mn} \beta_1^n \ldots \beta_m^n$ in the resultant. But the product gives exactly the same highest term in β's. By swapping columns, we see that our product and resultant also give exactly the same highest term in γ's, and so the resultant equals the product. For a finite field, embed into an infinite field, as in chapter 11. $\qquad \square$

> If $b(x) := x^7 + 2x + 1$ and $c(x) := x - 1$, then $c(x)$ has root $x = 1$, and both are monic, so
> $$\operatorname{res}_{b,c} = b(1) = 1^7 + 2 + 1 = 4.$$

13.5 Suppose that $b(x)$ is monic. Prove that

$$\operatorname{res}_{b,c} = (-1)^{mn} c(\beta_1) c(\beta_2) \ldots c(\beta_m).$$

Lemma 13.5. *For any polynomials $b(x), c(x), d(x)$,*

$$\operatorname{res}_{bd,c} = \operatorname{res}_{b,c} \operatorname{res}_{d,c}.$$

Proof. Split all of the polynomials into linear factors (over a splitting field):

$$\begin{aligned}
\operatorname{res}_{bd,c} &= b(\gamma_1)\, d(\gamma_1)\dots b(\gamma_n)\, d(\gamma_n)\,, \\
&= b(\gamma_1)\, b(\gamma_2)\dots b(\gamma_n)\, d(\gamma_1)\, d(\gamma_2)\dots d(\gamma_n)\,, \\
&= \operatorname{res}_{b,c}\operatorname{res}_{d,c}\,.
\end{aligned}$$

$\square$

13.6 Suppose that $c(x)$ is a monic polynomial. Prove that

$$\operatorname{res}_{b+dc,c} = \operatorname{res}_{b,c}$$

for any polynomials $c(x)$ and $d(x)$. A surprise: note that the degree of $b + dc$ could be either that of b or that of dc, or something smaller than either, as any number of terms in $b + dc$ might cancel one another. But surprisingly, we don't need to know the degree of $b + dc$.

13.7 A fast trick to find resultants: use the result of the previous problem to compute $\operatorname{res}_{b,c}$ where

$$\begin{aligned}
b(x) &= x^8 + 4x^2 + 2x + 1, \\
c(x) &= x^6 + 4.
\end{aligned}$$

We can make sage code for our fast trick:

```
R.<x> = PolynomialRing(QQ)
def fastresultant(b,c):
    m=b.degree()
    n=c.degree()
    if m<n:
        return (-1)^(m*n)*fastresultant(c,b)
    if n==0:
        if m==0:
            return 1
        return c(0)^m
    lb=b.leading_coefficient()
    lc=c.leading_coefficient()
    B=b/lb
    C=c/lc
    r=B % C
    if (r==0):
        return 0
    return (-1)^(m*n)*lb^n*lc^m*fastresultant(C,r)
```

so that `fastresultant(x^3+x+7,x^2+2)` yields 51.

The discriminant

For any polynomial $p(x)$, with any sort of coefficients, say

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n,$$

we define the *derivative*

$$p'(x) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^n,$$

just imitating calculus. There is no interpretation of this as a "rate of change", since the coefficients could be remainders modulo some integer, or something much stranger.

The *discriminant* Δ_p of a polynomial $p(x)$ is the resultant with the derivative: $\Delta_p := \mathrm{res}_{p,p'}$. By definition, the discriminant vanishes just when the polynomial has a common factor with its derivative. For example, over the complex numbers, the discriminant vanishes just when the polynomial has a "multiple root", i.e. a repeated linear factor:

$$p(z) = (z - z_0)^2 \ldots.$$

13.8 Find the discriminants of (a) $ax^2 + b + c$, (b) $x^3 + x^2$, (c) $x^3 + x^2 + 1$, (d) $x^3 + 2x - 1$ and explain what they tell you about these polynomials.

The polynomial $f(x) = 1 + 2x^2 + x^4$ over the real numbers has no real roots. Its derivative is $f'(x) = 4x + 4x^3$, so its discriminant is

$$\Delta_f = \det \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 4 & 0 & 0 \\ 0 & 2 & 0 & 4 & 0 & 4 & 0 \\ 1 & 0 & 2 & 0 & 4 & 0 & 4 \\ 0 & 1 & 0 & 0 & 0 & 4 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 4 \end{pmatrix} = 0$$

(after a long calculation). So $f(x)$ has a common factor with $f'(x)$. We can see the common factor more clearly if we factor:

$$f(x) = 1 + 2x^2 + x^4 = \left(1 + x^2\right)^2,$$
$$f'(x) = 4x + 4x^3 = 4x\left(1 + x^2\right).$$

So $f(x), f'(x)$ share a common factor of $1 + x^2$. They don't share a root over the real numbers, but over the complex numbers

$$f(x) = (x - i)^2 (x + i)^2,$$
$$f'(x) = 4x(x - i)(x + i),$$

so they share roots at $x = \pm i$, the double roots of $f(x)$.

The polynomial $p(x) = x^3 + 2x^2 + 2x + 2$ is irreducible over the rational numbers by Eisenstein's criterion, so has no rational roots. It has odd degree so at least one real root. There is no positive real root since all terms are positive. The derivative is $p'(x) = 3x^2 + 4x + 2$. By the quadratic formula, the zeroes of $p'(x)$ are at

$$x = -\frac{2}{3} \pm \frac{2\sqrt{2}}{3}i,$$

not real. Clearly $p'(x) > 0$ for all values of x. Therefore $p(x)$ is increasing, so has precisely one real root, irrational and negative. The two remaining complex roots are nonreal conjugates. The discriminant of $p(x)$ is

$$\Delta = \det \begin{pmatrix} 2 & 0 & 2 & 0 & 0 \\ 2 & 2 & 4 & 2 & 0 \\ 2 & 2 & 3 & 4 & 2 \\ 1 & 2 & 0 & 3 & 4 \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix} = -44.$$

Since $\Delta \neq 0$, we see that there are no double roots over the complex numbers, which is already clear.

13.9 Prove that the sign of the discriminant Δ of a monic cubic polynomial tells us that the number of roots is:

$$\begin{aligned} &\Delta > 0 \quad \text{3 distinct roots} \\ &\Delta = 0 \quad \text{a triple root or 1 real double root and 1 real single root} \\ &\Delta < 0 \quad \text{1 real and 2 complex roots.} \end{aligned}$$

13.10 Prove that, for any monic polynomial $p(x)$ with roots $x_1, x_2, \ldots, x_n$,

$$\Delta_{p(x)} = (-1)^{n(n-1)/2} \prod_{i<j} (x_i - x_j)^2.$$

Sage

Sage computes discriminants:

```
R.<t> = PolynomialRing(QQ)
(t^2-t+7).discriminant()
```

yields -27.

Parameterisation

Take the curve in the plane parameterised by

$$(x(t), y(t)) = \left(t^2 - 1, t\left(t^2 - 1\right)\right):$$

We want to find an equation for this curve, as a polynomial in x and y.

Proposition 13.6. *Draw a curve*

$$x(t) = \frac{p(t)}{q(t)}, y(t) = \frac{u(t)}{v(t)},$$

where $p(t), q(t), u(t), v(t)$ are polynomials with coefficients in a field k. The resultant in the variable t of the polynomials

$$p(t) - xq(t), u(t) - yv(t)$$

vanishes along the points (x, y) in the plane which lie along this curve; we can allow the values of x and y to be in the algebraic closure $\bar{k}$.

Proof. The resultant vanishes just at those points (x, y) where the two polynomials have a common factor. If there is a common factor of two polynomials, that factor has a root over $\bar{k}$, so there is a common root t of both polynomials, in other words a value of t where $x = p(t)/q(t)$ and $y = u(t)/v(t)$. On the other hand, a common root gives a common linear factor. So the resultant vanishes just along the image of the curve in the "plane". Careful: $k = \mathbb{R}$ then this "plane" has points given as pairs (x, y) of complex numbers, so it is not the complex plane, but two copies of the complex plane. $\square$

For the curve in the plane parameterised by

$$(x(t), y(t)) = \left(t^2 - 1, t\left(t^2 - 1\right)\right),$$

we have to take the resultant of

$$t^2 - 1 - x, t\left(t^2 - 1\right) - y,$$

as a polynomial in t; treat x and y as constants:

$$\det \begin{pmatrix} -(1+x) & & -y & & \\ 0 & -(1+x) & 0 & -1 & -y \\ 1 & 0 & -(1+x) & 0 & -1 \\ & -1 & 0 & 1 & 0 \\ & & -1 & & 1 \end{pmatrix} = y^2 - x^2 - x^3$$

So the curve is precisely the set of points (x, y) so that $y^2 = x^2 + x^3$.

For real variables, the curve $x = t^2, y = t^4$ has both $x \geq 0$ and $y \geq 0$ since $t^2 \geq 0$ and $t^4 \geq 0$. So the curve is the right half of the parabola $y = x^2$ on which $x \geq 0$. The resultant comes out as

$$\det \begin{pmatrix} -x & 0 & 0 & 0 & -y & 0 \\ 0 & -x & 0 & 0 & 0 & -y \\ 1 & 0 & -x & 0 & 0 & 0 \\ 0 & 1 & 0 & -x & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} = \left(x^2 - y\right)^2.$$

> So the resultant vanishes along $y = x^2$, the whole parabola. So we don't quite get what we wanted: we get the algebraic equation $y = x^2$ that that curve satisfies, but we miss out on the inequality $x \geq 0$.

13.11 Find a nonconstant polynomial equation $f(x, y) = 0$ satisfied by the points of the curve
$$x(t) = t^2, y(t) = t^3 - t.$$

13.12 Find a nonconstant polynomial equation $f(x, y) = 0$ satisfied by the points of the curve
$$x(t) = t^2 + t, y(t) = \frac{1}{t}.$$

13.13 Find a nonconstant polynomial equation $f(x, y) = 0$ satisfied by the points of the curve
$$x(t) = t^2 + 1, y(t) = t + \frac{1}{t}.$$

13.14 With coefficients over the field of remainders modulo 2, find a polynomial equation $0 = p(x, y)$, with $p(x, y)$ not constant, satisfied by the parameterized plane curve $(x(t), y(t)) = (t^3 + t, 1/t)$.

Sage

Sage handles 2 variable polynomials:

```
P.<x,y> = PolynomialRing(QQ)
a = x + y
b = x^3 - y^3
a.resultant(b)
```

yields $-2y^3$. If we want to use the resultant in the other variable, we specify which variable to get rid of: `a.resultant(b, y)` yields $2x^3$. Sage will take a parameterisation of a curve and give us the equation of the curve. Our example above becomes:

```
R.<x,y,t> = PolynomialRing(QQ)
(t^2-1-x).resultant(t*(t^2-1)-y,t)
```

yields $-x^3 - x^2 + y^2$.
Sage computes resultants of polynomials over any field.

> Let K be the splitting field of $x^2 + x + 1$ over $k = \mathbb{Z}/2\mathbb{Z}$, with generator a. We compute the resultant and greatest common divisor of
> $$p(t) = (at + 1)^2 (at - 1)^2, q(t) = (at + 1)(at + a^2)^2.$$
> via
> ```
> R.<x> = PolynomialRing(GF(2))
> K.<a> = (x^2 + x + 1).splitting_field()
> S.<t> = PolynomialRing(K)
> ```

```
p = S((a*t+1)^2*(a*t-1)^2)
q = S((a*t+1)*(a*t+a^2)^2)
print(p.gcd(q))
print(p.resultant(q))
```

yielding $t + a + 1$ and 0.

Permuting roots

This letter, if judged by the novelty and profundity of ideas it contains, is perhaps the most substantial piece of writing in the whole literature of mankind.

— Hermann Weyl
    On Évariste Galois's letter, written the night before Galois died in a pistol duel. Galois's letter is about permutations of roots of polynomials.

Vieta's formulas

For now, we work over any field. Which polynomials have which roots? How are the coefficients related to the roots?

> To get a quadratic polynomial $p(x)$ to have roots at 3 and 7, we need it to have $x - 3$ and $x - 7$ as factors. So it has to be
>
> $$
> \begin{aligned}
> p(x) &= a(x-3)(x-7), \\
> &= a\left(x^2 - 7x - 3x + (-3)(-7)\right), \\
> &= a\left(x^2 - (3+7)x + 3 \cdot 7\right).
> \end{aligned}
> $$

A polynomial is *monic* if its leading coefficient is 1. We can make any polynomial monic by dividing off the leading term:

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

The only monic quadratic polynomial $p(x)$ with roots $x = 3$ and $x = 7$ is

$$
p(x) = x^2 - (3+7)x + 3 \cdot 7.
$$

The only quadratic monic polynomial $p(x)$ with roots $x = r$ and $x = s$, by the same steps, must be

$$
p(x) = x^2 - (r+s)x + rs.
$$

By the same steps, the only cubic monic polynomial $q(x)$ with roots at $x = r$, $x = s$ and $x = t$ is

$$
q(x) = (x-r)(x-s)(x-t).
$$

If we multiply it all out, we get

$$q(x) = x^3 - (r + s + t)x^2 + (rs + rt + st)x - rst.$$

Ignoring the minus signs, the coefficients are

$$1$$
$$r + s + t$$
$$rs + rt + st$$
$$rst$$

There is some pattern:

a. There is a minus sign, turning on and off like a light switch, each time we write down a term.

b. Each coefficient is a sum of products of roots, taken in all possible ways with a fixed number of roots.

Proposition 14.1 (Vieta's formula (Vieté))**.** *If a monic polynomial $p(x)$ splits into a product of linear factors, say*

$$p(x) = (x - r_1)(x - r_2)\ldots(x - r_n)$$

then the numbers $r_1, r_2, \ldots, r_n$ are the roots of $p(x)$, and the coefficients of $p(x)$, say

$$p(x) = x^n - a_1 x^{n-1} + \cdots \pm a_{n-1} x \pm a_n,$$

(with signs switching each term in front of the a_j coefficients) are computed from the roots by

$$a_1 = r_1 + r_2 + \cdots + r_n,$$
$$a_2 = \sum_{i<j} r_i r_j,$$
$$\vdots$$
$$a_i = \sum_{j_1 < j_2 < \cdots < j_i} r_{j_1} r_{j_2} \ldots r_{j_i},$$
$$\vdots$$
$$a_{n-1} = r_1 r_2 \ldots r_{n-1} + r_1 r_2 \ldots r_{n-2} r_n + \cdots + r_2 r_3 \ldots r_n,$$
$$a_n = r_1 r_2 \ldots r_n.$$

The *elementary symmetric polynomials* are the polynomials $e_1, \ldots, e_n$ of variables $t_1, t_2, \ldots, t_n$, given by

$$e_1(t_1, t_2, \ldots, t_n) = t_1 + t_2 + \cdots + t_n,$$
$$e_2(t_1, t_2, \ldots, t_n) = t_1 t_2 + t_1 t_3 + \cdots + t_{n-1} t_n,$$
$$= \sum_{i<j} t_i t_j,$$
$$\vdots$$
$$e_i(t_1, t_2, \ldots, t_n) = \sum_{j_1 < j_2 < \cdots < j_i} t_{j_1} t_{j_2} \ldots t_{j_i},$$
$$\vdots$$
$$e_{n-1}(t_1, t_2, \ldots, t_n) = t_1 t_2 \ldots t_{n-1} + t_1 t_2 \ldots t_{n-2} t_n + \cdots + t_2 t_3 \ldots t_n,$$
$$e_n(t_1, t_2, \ldots, t_n) = t_1 t_2 \ldots t_n.$$

So we can restate our proposition as:

Proposition 14.2 (Vieta's formula). *If a monic polynomial $p(x)$ splits into a product of linear factors, say*

$$p(x) = (x - r_1)(x - r_2) \ldots (x - r_n)$$

then the numbers $r_1, r_2, \ldots, r_n$ are the roots of $p(x)$, and the coefficients of $p(x)$:

$$p(x) = x^n - e_1 x^{n-1} + \ldots (-1)^{n-1} e_{n-1} x + (-1)^n e_n,$$

are the elementary symmetric polynomials

$$e_i = e_i(r_1, r_2, \ldots, r_n)$$

of the roots $r_1, r_2, \ldots, r_n$.

Proof. We can see this immediately for linear polynomials: $p(x) = x - r$, and we checked it above for quadratic and cubic ones. It is convenient to rewrite the elementary symmetric polynomials as

$$e_i(r_1, r_2, \ldots, r_n) = \sum r_1^{b_1} r_2^{b_2} \ldots r_n^{b_n}$$

where the sum is over all choices of numbers $b_1, b_2, \ldots, b_n$ so that

a. each of these b_j is either 0 or 1 and

b. so that altogether
$$b_1 + b_2 + b_3 + \cdots + b_n = i,$$

or in other words, "turn on" i of the roots, and turn off the rest, and multiply out the ones that are turned on, and then sum over all choices of which to turn on.

If we expand out the product

$$p(x) = (x - r_1)(x - r_2) \ldots (x - r_n)$$

we do so by pick whether to multiply with x or with $-r_1$ from the first factor, and then whether to multiply with x or with $-r_2$ from the second, and so on, and we add

over all of these choices. Each choice write as $b_1 = 0$ if we pick to multiply in the x, but $b_1 = 1$ if we pick to multiply in the $-r_1$, and so on:

$$p(x) = \sum_{b_1,\ldots,b_n} (-r_1)^{b_1} (-r_2)^{b_2} \ldots (-r_n)^{b_n} x^{(1-b_1)+(1-b_2)+\cdots+(1-b_n)},$$

$$= \sum_j \sum_{b_1+\cdots+b_n=j} (-r_1)^{b_1} (-r_2)^{b_2} \ldots (-r_n)^{b_n} x^{n-j}.$$

$\square$

Symmetric polynomials

A polynomial $f(t_1, t_2, \ldots, t_n)$ is *symmetric* if it is unchanged by permuting the variables $t_1, t_2, \ldots, t_n$.

> $t_1 + t_2 + \cdots + t_n$ is symmetric.

For any numbers or variables $t = (t_1, t_2, \ldots, t_n)$ let

$$P_t(x) := (x - t_1)(x - t_2) \ldots (x - t_n).$$

Clearly the roots of $P_t(x)$ are precisely the entries of the vector t. Let $e(t) := (e_1(t), e_2(t), \ldots, e_n(t))$, so that e maps vectors to vectors (with entries in our field).

Recall that a field is *algebraically closed* if every polynomial with coefficients in that field splits into a product of linear factors with coefficients in that field. Recall the fundamental theorem of algebra (theorem 9.3 on page 69): the field of complex numbers $\mathbb{C}$ is algebraically closed.

Lemma 14.3. *For each vector c with entries in a field, there is a vector t, with entries in a finite degree extension field, so that $e(t) = c$. In particular, there is no relation between the elementary symmetric polynomials. Over an infinite field, there is no nontrivial polynomial equation satisfied by the elementary symmetric polynomials.*

Proof. Let $t_1, t_2, \ldots, t_n$ be the roots of the polynomial

$$P(x) = x^n - c_1 x^{n-1} + c_2 x^{n-2} + \cdots + (-1)^n c_n.$$

By proposition 14.2 on the previous page, $P_t(x)$ has coefficients precisely the same as those of $P(x)$, i.e. $P_t(x) = P(x)$. Any relation $R(e(t)) = 0$ between the elementary symmetric polynomials must be satisfied at all points c, so is trivial for any function R. For a polynomial R, which we don't think of as a function, over an infinite field, the expression $R(e(t))$ vanishing forces $R(c)$ to vanish for each c. Varying any one component of c, $R(c)$ is a polynomial function of that component with infinitely many roots, so is zero in that component for all other components fixed, so the zero polynomial by induction. $\square$

Lemma 14.4. *Over any field, the entries of two vectors s and t are permutations of one another just when $e(s) = e(t)$, i.e. just when the elementary symmetric polynomials take the same values at s and at t.*

Proof. The roots of $P_s(x)$ and $P_t(x)$ are the same numbers. $\square$

We want to see that every symmetric polynomial $f(t_1, t_2, \ldots, t_n)$ has the form $f(t) = h(e(t))$, for a unique polynomial h, and conversely if h is any function at all, then $f(t) = h(e(t))$ determines a symmetric polynomial. We want a recipe to write down each symmetric polynomial in terms of the elemenary symmetric polynomials, to find this mysterious h.

Take the function

$$f = 3x^2 yz + 3xy^2 z + 3xyz^2 + 5xy + 5xz + 5yz.$$

Clearly the terms with the 5's look like an elementary symmetric function:

$$f = 3x^2 yz + 3xy^2 z + 3xyz^2 + 5e_2.$$

(We write e_2 to mean the function $e_2(x, y, z) = xy + xz + yz$, the 2nd elementary symmetric function.) But what about the terms with the 3's? Factor them all together as much as we can:

$$f = 3xyz(x + y + z) + 5e_2.$$

Then it is clear:
$$f = 3e_3 e_1 + 5e_2.$$

If $a = (a_1, a_2, \ldots, a_n)$, write t^a to mean $t_1^{a_1} t_2^{a_2} \ldots t_n^{a_n}$. Order terms by "alphabetical" order, also called *weight*: order monomials by the order in t_1; if two monomials have the same order in t_1, break the tie by looking at the order in t_2, and so on.

The term
$$t_1^5 t_2^3$$
has higher weight than any of

$$t_1^3 t_2^5, t_1^5 t_2^2, 1, \text{ and } t_2^{1000}.$$

Showing the highest order term, and writing dots to indicate lower order terms,

$$e_1(t) = t_1 + \ldots,$$
$$e_2(t) = t_1 t_2 + \ldots,$$

$$\vdots$$

$$e_j(t) = t_1 t_2 \ldots t_j + \ldots$$

$$\vdots$$

$$e_n(t) = t_1 t_2 \ldots t_n.$$

If a symmetric polynomial contains a term, then it also contains every term obtained by permuting the variables:

$$t_1^3 t_2^5 + t_1^5 t_2^3.$$

It is somehow easier to read x, y, z than t_1, t_2, t_3, so take a symmetric polynomial
$$f(x, y, z) = 6x^{16}y^9z^7 + \ldots$$
where we only write out the highest weight term. In such notation,
$$e_1(x, y, z) = x + y + z = x + \ldots,$$
$$e_2(x, y, z) = xy + xz + yz = xz + \ldots,$$
$$e_3(x, y, z) = xyz.$$

Imagine a term in f: if the highest power in that term is in y or z, permute the variables to get that highest power into x, giving a higher weight term in f. So the highest weight term contains the variables x, y, z, with highest power in x and successively lower powers in each of y, z. By the same reasoning, inside the highest weight term, each of x, y, z appears to at least to the power of the z variable: in this case, at least to a power of 7. Factor out 7 powers of each variable; we underline that factored out part so you can see it:
$$f = 6x^9y^2\underline{(xyz)}^7 + \ldots$$

In the remaining factors, there are now no z variables, and each variable appears at least to a power of 2, so we factor out 2 of each:
$$f = 6x^7\underline{(xy)}^2 (xyz)^7 + \ldots$$

So finally it is clear that f has the same highest weight as $6e_1^7e_2^2e_3^7$. Hence
$$f = 6e_1^7e_2^2e_3^7 + \ldots$$
up to terms of lower weight.

Theorem 14.5. *Every symmetric polynomial f has exactly one expression as a polynomial in the elementary symmetric polynomials. If f has integer coefficients, then f is an integer coefficient polynomial of the elementary symmetric polynomials. Similarly if f has coefficients in any field, then f is a polynomial of the elementary symmetric polynomials, with coefficients in that same field.*

Proof. Write the highest weight term as some coefficient times $t_1^{a_1}t_2^{a_2}\ldots t_n^{a_n}$. For simplicity, write this as t^a. If a_1 is smaller than a_2, permute them to get a higher weight term, a contradiction, so $a_1 \geq a_2$. In the same way, $a_1 \geq a_2 \geq a_3 \geq \cdots \geq a_n$. It might be that various of these a_i are zero; instead we can write the term as $t_1^{a_1}\ldots t_k^{a_k}$, with all positive powers $a_1 \geq a_2 \geq \cdots \geq a_k > 0$. Let
$$b_1 := a_1 - a_k, \ldots b_{k-1} := a_{k-1} - a_k.$$

Factor out:
$$t^a = t^b(t_1 \ldots t_k)^{a_k}.$$

This is the highest weight term in
$$t^b e_k^{a_k}.$$

For simplicity, write expressions like $e_1^{d_1} \ldots e_n^{d_n}$ as e^d. By induction, we can then find some expression e^d with highest weight term t^b, so

$$t^a = t^b e_k^{a_k} + \cdots = e^d e_k^{a_k} + \ldots$$

where ... means lower weight terms. By induction, we can somehow write the lower weight terms as polynomials in the elementary symmetric functions.   □

> The sum of squares of two variables is symmetric:
>
> $$x^2 + y^2 = (x+y)^2 - 2xy.$$
>
> To compute out these expressions: $f(x,y) = x^2 + y^2$ has highest term x^2. The polynomials $e_1(x,y) = x + y$ and $e_2(x,y) = xy$ have highest terms x and xy. So we subtract off $e_1(x,y)^2$ from $f(x,y)$, and find $f(x,y) - e_1(x,y)^2 = -2xy = -2e_2(x,y)$.

14.1 Express each of the following polynomials as polynomials in the elementary symmetric polynomials:

a. $x^3 y^3 z + x^3 yz^3 + xy^3 z^3$

b. $4xyz^3 + 4xzy^3 + 4yzx^3$

c. $x^4 y^4 z^4$

d. $xyz + x^2 y^2 z^2$

Sage

Sage can compute with elementary symmetric polynomials. It writes them in a strange notation. The polynomials we have denoted by e_3 sage denotes by e[3]. More strangely, sage denotes $e_1 e_2^3 e_5^2$ as e[5,5,2,2,2,1] which is the same in sage as e[2]^3*e[1]*e[5]^2. You can't write e[1,2,2,2,5,5]; you have to write it as e[5,5,2,2,2,1]: the indices have to decrease. Set up the required rings:

```
P.<w,x,y,z>=PolynomialRing(QQ)
S=SymmetricFunctions(QQ)
e=S.e()
```

This creates a ring $P = \mathbb{Q}[w,x,y,z]$, and then lets S be the ring of symmetric polynomials. The last (mysterious) line sets up the object e to be the elementary symmetric polynomials of the ring S. We can then define a polynomial in our variables:

```
f = w^2+x^2+y^2+z^2+w*x+w*y+w*z+x*y+x*z+y*z
e.from_polynomial(f)
```

which prints out $e_{1,1} - e_2$, the expression of f in terms of elementary symmetric polynomials. To expand out a symmetric polynomial into w, x, y, z variables:

```
q = e[2,1]+e[3]
q.expand(4,alphabet=['w','x','y','z'])
```

prints out

$w^2 x + wx^2 + w^2 y + 4wxy + x^2 y + wy^2 + xy^2 + w^2 z + 4wxz + x^2 z + 4wyz + 4xyz + y^2 z + wz^2 + xz^2 + yz^2$

Sums of powers

Define $p_j(t) = t_1^j + t_2^j + \cdots + t_n^j$, the sums of powers.

Lemma 14.6 (Isaac Newton). *The sums of powers are related to the elementary symmetric polynomials by*

$$0 = e_1 - p_1,$$
$$0 = 2\, e_2 - p_1\, e_1 + p_2,$$
$$\vdots$$
$$0 = k\, e_k - p_1\, e_{k-1} + p_2\, e_{k-2} - \cdots + (-1)^{k-1} p_{k-1}\, e_1 + (-1)^k p_k,$$

Using these equations, we can write the elementary symmetric polynomials inductively in terms of the sums of powers, or vice versa.

Proof. Let's write $t^{(\ell)}$ for t with the ℓth entry removed, so if t is a vector with n entries, then $t^{(\ell)}$ is a vector with $n-1$ entries.

$$p_j e_{k-j} = \sum_\ell t_\ell^j \sum_{i_1 < i_2 < \cdots < i_{k-j}} t_{i_1} t_{i_2} \ldots t_{i_{k-j}}$$

Either we can't pull a t_ℓ factor out of the second sum, or we can:

$$= \sum_\ell t_\ell^j \sum_{\substack{i_1 < i_2 < \cdots < i_{k-j} \\ i_1, i_2, \cdots \neq \ell}}^{i_1, i_2, \cdots \neq \ell} t_{i_1} t_{i_2} \ldots t_{i_{k-j}} + \sum_\ell t_\ell^{j+1} \sum_{\substack{i_1 < i_2 < \cdots < i_{k-j-1}}}^{i_1, i_2, \cdots \neq \ell} t_{i_1} t_{i_2} \ldots t_{i_{k-j-1}}$$

$$= \sum_\ell t_\ell^j e_{k-j}\left(t^{(\ell)}\right) + \sum_\ell t_\ell^{j+1} e_{k-j-1}\left(t^{(\ell)}\right).$$

Putting in successive terms of our sum,

$$p_j e_{k-j} - p_{j+1} e_{k-j-1} = \sum_\ell t_\ell^j e_{k-j}\left(t^{(\ell)}\right) + \sum_\ell t_\ell^{j+1} e_{k-j-1}\left(t^{(\ell)}\right)$$

$$- \sum_\ell t_\ell^{j+1} e_{k-j-1}\left(t^{(\ell)}\right) - \sum_\ell t_\ell^{j+2} e_{k-j-2}\left(t^{(\ell)}\right)$$

$$= \sum_\ell t_\ell^j e_{k-j}\left(t^{(\ell)}\right) - \sum_\ell t_\ell^{j+2} e_{k-j-2}\left(t^{(\ell)}\right).$$

Hence the sum collapses to

$$p_1 e_k - p_2 e_{k-1} + \cdots + (-1)^{k-1} p_{k-1} e_1 = \sum_\ell t_\ell e_{k-1}\left(t^{(\ell)}\right) + (-1)^{k-1} \sum_\ell t_\ell^k \cdot e_0\left(t^{(\ell)}\right)$$

$$= k\, e_k + (-1)^{k-1} p_k.$$

$\square$

Proposition 14.7. *Every symmetric polynomial is a polynomial in the sums of powers. If the coefficients of the symmetric polynomial lie in a field, then it is a polynomial function of the sums of powers over that same field.*

Proof. We can solve recursively for the sums of powers in terms of the elementary symmetric polynomials and conversely. $\square$

The invariants of a square matrix

Over any field, a polynomial $f(A)$ in the entries of a square matrix A, with coefficients in the field, is *invariant* if $f\left(FAF^{-1}\right) = f(A)$ for any invertible matrix F with coefficients in the field. An invariant is independent of change of basis: if $T \colon V \to V$ is a linear map on an n-dimensional vector space, we can define the value $f(T)$ of any invariant f of $n \times n$ matrices, by letting $f(T) := f(A)$ where A is the matrix associated to T in some basis of V.

> For any $n \times n$ matrix A, write
>
> $$\det(A - \lambda I) = \chi_A(\lambda) = e_n(A) - e_{n-1}(A)\lambda + e_{n-2}(A)\lambda^2 + \cdots + (-1)^n \lambda^n.$$
>
> The expressions $e_1(A), e_2(A), \ldots, e_n(A)$ are invariants, while $\chi_A(\lambda)$ is the *characteristic polynomial* of A.

> If we write the trace of a matrix A as $\operatorname{tr} A$, then the functions
>
> $$p_k(A) = \operatorname{tr}\left(A^k\right)$$
>
> are invariants.

14.2 If A is diagonal, say

$$A = \begin{pmatrix} t_1 & & & \\ & t_2 & & \\ & & \ddots & \\ & & & t_n \end{pmatrix},$$

then prove that $e_j(A) = e_j(t_1, t_2, \ldots, t_n)$, the elementary symmetric polynomials of the eigenvalues.

14.3 Generalize the previous exercise to A diagonalizable.

14.4 Prove that the entries of A^{-1} are rational functions of the entries of the square matrix A, over any field.

> Let Δ_A be the discriminant of the characteristic polynomial of A: $\Delta_A := \Delta_{\chi_A}$. The map $A \mapsto \Delta_A$ is also a polynomial invariant of A, as the coefficients of the characteristic polynomial are.

14.5 Take a square matrix A. Suppose that the characteristic polynomial of A splits into linear factors. Prove that $\Delta_A = 0$ just when A has an eigenvalue of multiplicity two or more.

Theorem 14.8. *Every invariant polynomial of square matrices over an infinite field has exactly one expression as a polynomial in the elementary symmetric polynomials of the eigenvalues.*

We can replace the elementary symmetric polynomials of the eigenvalues by the sums of powers of the eigenvalues.

Proof. Take an invariant polynomial $f(A)$. Every invariant polynomial $f(A)$ determines an invariant polynomial $f(t)$ by setting

$$A = \begin{pmatrix} t_1 & & & \\ & t_2 & & \\ & & \ddots & \\ & & & t_n \end{pmatrix}.$$

Taking F any permutation matrix, invariance tells us that $f\left(FAF^{-1}\right) = f(A)$. But $f\left(FAF^{-1}\right)$ is given by applying the associated permutation to the entries of t. Therefore $f(t)$ is a symmetric function. Therefore $f(t) = h(e(t))$, for some polynomial h; so $f(A) = h(e(A))$ for diagonal matrices. Replace f by $f(A) - h(e(A))$ to arrange that $f(A) = 0$ on all diagonal matrices A. By invariance, $f(A) = 0$ on all diagonalizable matrices.

The equation $f(FAF^{-1}) = f(A)$ holds for all F over our field. Imagine that F has abstract variables as entries. The difference $f(FAF^{-1}) - f(A)$ is a rational function in the entries of F. Its numerator vanishes for any choice of values of those abstract variables, and so vanishes. Hence $f(FAF^{-1}) = f(A)$ for F with abstract variable entries. In particular, if our field lies in a larger field, then f remains invariant over the larger field, because we can plug into the entries of F the values in the larger field.

Recall that a matrix has an eigenvector for each eigenvalue, so if there are n distinct eigenvalues, then A is diagonalizable. This occurs just when the characteristic polynomial $\chi_A(\lambda)$ splits into distinct linear factors, and so $\Delta_A \neq 0$. Conversely, if $\Delta_A \neq 0$ and $\chi_A(\lambda)$ splits into linear factors, then they are distinct, and so $f(A) = 0$. By theorem 12.2 on page 95, we can sit our field into a larger field in which the characteristic polynomial $\chi_A(\lambda)$ splits into linear factors, and therefore if $\Delta_A \neq 0$ then $f(A) = 0$.

Pick any matrix A_0 whose eigenvalues are all distinct. In particular, $\Delta_{A_0} \neq 0$. Take any matrix A_1. For an abstract variable t, let $A_t := (1-t)A_0 + tA_1$. Since Δ_{A_t} is a polynomial, not vanishing at $t = 0$, it is a nonzero polynomial in t. So $\Delta_{A_t} \neq 0$ except for finitely many t. Hence $f(A_t) = 0$ except for finitely many t. But $f(A)$ is a polynomial, so vanishes for all t. $\qquad\square$

The function $f(A) = e_j\left(|\lambda_1|, |\lambda_2|, \ldots, |\lambda_n|\right)$, where A has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, is a continuous invariant function of a real matrix A, and is *not* a polynomial in $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Resultants and permutations

If $b(x, y) = x^2 + y^2 - 1$ and $c(x, y) = x^2 + (y-1)^2 - 1$, how can we find the points (x, y) at which both $b(x, y) = 0$ and $c(x, y) = 0$?
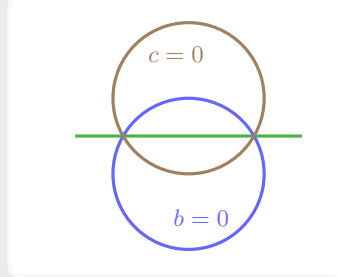
Think of $b(x,y)$ and $c(x,y)$ as polynomials in x, with coefficients rational functions of y. Then the resultant $r(y)$ of $b(x,y), c(x,y)$ is a rational function of y. Compute it: $r(y) = (2y-1)^2$ vanishes at $y = 1/2$. So there is a common factor there:

$$b(x, 1/2) = c(x, 1/2) = \left(x - \frac{\sqrt{3}}{2}\right)\left(x + \frac{\sqrt{3}}{2}\right).$$

The common factor came from two common roots.

In chapter 13, we saw that for any two polynomials split into linear factors

$$b(x) = (x - \beta_1)(x - \beta_2)\ldots(x - \beta_m),$$
$$c(x) = (x - \gamma_1)(x - \gamma_2)\ldots(x - \gamma_n),$$

the resultant is

$$\operatorname{res}_{b,c} = (\gamma_1 - \beta_1)(\gamma_1 - \beta_2)\ldots(\gamma_n - \beta_m).$$

So the resultant is homogeneous of degree mn in the variables β_i, γ_j. The resultant is invariant under any permutation of the roots of $b(x)$, and also under any permutation of the roots of $c(x)$. Indeed the resultant is by definition expressed in terms of the coefficients of $b(x)$ and $c(x)$, not the roots. The coefficients are homogeneous polynomials in the roots, elementary symmetric polynomials. Expanding out the coefficients

$$\begin{aligned}
b(x) &= x^m + b_{m-1}x^{m-1} + \cdots + b_0, \\
&= x^m - e_1(\beta)x^{m-1} + e_2(\beta)x^{m-2} + \cdots \pm e_m(\beta), \\
c(x) &= x^n + c_{n-1}x^{n-1} + \cdots + c_0, \\
&= x^n - e_1(\gamma)x^{n-1} + e_2(\gamma)x^{n-2} + \cdots \pm e_n(\gamma),
\end{aligned}$$

we see that b_j has degree $m - j$ in the roots.

Suppose now that we just take any polynomial $b(x)$ of degree m with coefficients being abstract variables b_j. We will now invent a different concept of *weight*. Assign each coefficient b_j the *weight* $m - j$ and then define the *weight* of a monomial in the b_j to be the sum of weights of its factors.

Lemma 14.9. *With this notion of weight, every term in the resultant has weight mn.*

Proof. By theorem 12.2 on page 95, we can replace our field by some larger field, to arrange that $b(x)$ and $c(x)$ split into linear factors, and then the weight of each term in the resultant is clear by expanding the product in terms of roots.   □

Proposition 14.10. *Given any two polynomials $b(x,y)$ of total degree m and $c(x,y)$ of total degree n over a field, in two variables x, y, either there are at most mn values of y for which there is some point (x,y) which satisfies both $b(x,y) = 0$ and $c(x,y) = 0$ (and the resultant in x vanishes at those values), or $b(x,y)$ and $c(x,y)$ have a common factor as polynomials in x, y. If the polynomials are homogeneous then either they have a homogeneous common factor, or their resultant is homogeneous nonzero of degree mn.*

Proof. Think of the polynomials as polynomials in x with coefficients rational in y:

$$b(x,y) = \sum_{j+k \le m} b_{jk} x^k y^j = \sum_j b_j(y) x^j,$$

we see that $b_j(y)$ has degree at most $m - j$, exactly the weight of the coefficient $b_j(y)$ as it enters into the resultant. Therefore the resultant is a polynomial of degree at most mn in y. The resultant vanishes at those values $y = a$ for which there is a common factor between $b(x,a)$ and $c(x,a)$, and in particular if there is a common root it vanishes. So there are at most mn such points or the resultant is everywhere zero. If the resultant vanishes everywhere, then $b(x,y)$ and $c(x,y)$ have a common factor with coefficients rational functions of y. By Gauss's lemma (proposition 10.6 on page 79) they also have a common factor in polynomials in x, y.

For homogeneous polynomials, each term $b_j(y)$ has degree exactly $m - j$, so the resultant either vanishes everywhere or has degree exactly mn.   □

14.6 Prove that, over any field, any symmetric rational function (i.e. permutation invariant) has an expression $b(e_1, \ldots, e_n)/c(e_1, \ldots, e_n)$ as a ratio of polynomials with no common nonconstant factor, expressed in the elementary symmetric polynomials, unique up to rescaling numerator and denominator by the same nonzero constant.

Chapter 15

Rings

We dance round in a ring and suppose, but the secret sits in the middle and knows.

— Robert Frost
THE SECRET SITS, A WITNESS TREE, 1942

Morphisms

A *morphism* of rings $f\colon R \to S$ is a map f taking elements of a ring R to elements of a ring S, so that $f(a + b) = f(a) + f(b)$ and $f(ab) = f(a)f(b)$ for all elements b, c of R. (Morphisms are also often called *homomorphisms*.)

> The map $f\colon \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$, $f(b) = \bar{b}$, remainder modulo m, is a morphism.

15.1 Prove that the map $f\colon \mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$, $f(b) = 4b$ is a morphism. Draw pictures to explain it.

15.2 Find all morphisms $f\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/q\mathbb{Z}$, where p and q are prime numbers and $p \neq q$.

Sage

We define $R = \mathbb{Q}[x, y]$ and $T = \mathbb{Q}[z]$ and define a morphism $\phi\colon R \to S$ by $\phi(x) = z^3$ and $\phi(y) = 0$:

```
R.<x,y> = PolynomialRing(QQ)
T.<z> = PolynomialRing(QQ)
phi = R.hom([z^3, 0],T)
phi(x^2+y^2)
```

yielding $\phi(x^2 + y^2) = z^6$.

Kernel and image of a morphism

If $f\colon R \to S$ is a morphism of rings, the *kernel* of f is the set

$$\ker f := \{\, r \in R \mid f(r) = 0 \,\}.$$

The *image* of f is the set

$$f(R) := \{\, s \in S \mid s = f(r) \text{ for some } r \in R \,\}.$$

> The morphism $x \in \mathbb{R} \mapsto x + 0i \in \mathbb{C}$ has kernel $\{\, 0 \,\}$ and image the set of all numbers $x + 0i$. (We often write 0 to mean $\{\, 0 \,\}$ in algebra.)

> The morphism $a \in \mathbb{Z} \mapsto \bar{a} \in \mathbb{Z}/m\mathbb{Z}$ has kernel $m\mathbb{Z}$.

15.3 Prove that the kernel of a morphism $f\colon R \to S$ of rings is a subring of R, while the image is a subring of S.

15.4 Prove that the morphism $p(x) \in \mathbb{R}[x] \mapsto p(i) \in \mathbb{C}$ has kernel consisting of all polynomials of the form

$$\left(1 + x^2\right) p(x),$$

for any polynomial $p(x)$.

15.5 Suppose that R is a ring and that $f, g\colon \mathbb{Q} \to R$ are morphisms and that $f(1) = g(1)$. Prove that $f = g$.

A morphism $f\colon R \to S$ with kernel 0 is *injective*, also called *one-to-one*, while a morphism $f\colon R \to S$ with image $f(R) = S$ is *surjective* or *onto*.

A surjective and injective morphism is an *isomorphism*. If there is an isomorphism between rings R and S, they are *isomorphic*, denoted $R \cong S$. If two rings are isomorphic, then they are identical for all practical purposes. An isomorphism $f\colon R \to R$ from a ring to itself is an *automorphism*.

> Take some positive integers $m_1, m_2, \ldots, m_n$. Let
>
> $$m := m_1 m_2 \ldots m_n.$$
>
> For each $b \in \mathbb{Z}/m\mathbb{Z}$, write its remainder modulo m_1 as $\bar{b}_1$, and so on. Write $\bar{b}$ to mean
>
> $$\bar{b} := \left(\bar{b}_1, \bar{b}_2, \ldots, \bar{b}_n\right).$$
>
> Let
>
> $$S := (\mathbb{Z}/m_1\mathbb{Z}) \oplus (\mathbb{Z}/m_2\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/m_n\mathbb{Z}).$$
>
> Map $f\colon \mathbb{Z}/m\mathbb{Z} \to S$ by $f(b) = \bar{b}$. It is clear that f is a ring morphism. The Chinese remainder theorem states that f is a ring *isomorphism* precisely when any two of the integers
>
> $$m_1, m_2, \ldots, m_n$$
>
> are coprime. For example, if $m = 3030 = 2 \cdot 3 \cdot 5 \cdot 101$, then
>
> $$\mathbb{Z}/3030\mathbb{Z} \cong (\mathbb{Z}/2\mathbb{Z}) \oplus (\mathbb{Z}/3\mathbb{Z}) \oplus (\mathbb{Z}/5\mathbb{Z}) \oplus (\mathbb{Z}/101\mathbb{Z}).$$

15.6 Prove that $\mathbb{Z}/4\mathbb{Z}$ is *not* isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \oplus (\mathbb{Z}/2\mathbb{Z})$.

15.7 Prove that any two rings that contains exactly one element are isomorphic, by a unique isomorphism.

Kernels

The main trick to understand complicated rings is to find morphisms to them from more elementary rings, and from them to more elementary rings.

15.8 Prove that every subring of any given ring is the image of some morphism of rings.

It is natural to ask which subrings of a ring are kernels of morphisms. An *ideal* of a commutative ring R is a nonempty subset I of R so that, if i, j are in I then $i - j$ is in I and if i is in I and r is in R then ri is in I: an ideal is *sticky* under multiplication.

> The even integers form an ideal $2\mathbb{Z}$ inside the integers.

> The integers do *not* form an ideal inside the rational numbers. They are not sticky, because $i = 2$ is an integer and $r = \frac{1}{3}$ is rational, but $ri = \frac{2}{3}$ is *not* an integer.

> The real polynomials in a variable x vanishing to order 3 or more at the origin form an ideal $x^3\mathbb{R}[x]$ inside $\mathbb{R}[x]$. Note why they are sticky: if $i = x^3 p(x)$ and $r = q(x)$ then $ri = x^3 p(x)q(x)$ has a factor of x^3 still, so also vanishes to order 3 or more. This is a good way to think about ideals: they remind us of the ideal of functions vanishing to some order somewhere, inside some ring of functions.

15.9 Prove that the only ideals in $\mathbb{Q}$ are 0 and $\mathbb{Q}$.

Lemma 15.1. *The kernel of any morphism of rings is an ideal.*

Proof. Take a ring morphism $f \colon R \to S$ and let $I := \ker f$. Then if i, j are in I, this means precisely that f kills i and j, so $f(i) = 0$ and $f(j) = 0$. But then $f(i - j) = f(i) - f(j) = 0 - 0 = 0$. So f kills $i - j$, or in other words $i - j$ lies in I.

Pick any i in I and any r in R. We need to prove that ir is in I, i.e. that f kills ir, i.e. that $0 = f(ir)$. But $f(ir) = f(i)f(r) = 0f(r) = 0$. $\qquad\square$

We will soon see that kernels of morphisms are precisely the same as ideals.

15.10 Prove that the ideals of $\mathbb{Z}$ are precisely $0, \mathbb{Z}$ and the subsets $m\mathbb{Z}$ for any integer $m \geq 2$. (Of course, we could just say that the ideals are the sets $m\mathbb{Z}$ for any integer $m \geq 0$.)

Lemma 15.2. *Given a ring R and any set A of elements of R, there is a unique smallest ideal containing A, denoted (A). To be precise, (A) is the set of all finite sums*

$$r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$$

for any elements

$$r_1, r_2, \ldots, r_n$$

of R and any elements

$$a_1, a_2, \ldots, a_n$$

of A.

Proof. Let (A) be that set of finite sums. Clearly A lies in (A), and (A) is an ideal. On the other hand, if A lies in some ideal, that ideal must contain all elements of A, and be sticky, so contain all products ra for r in R and a in A, and so contains all finite sums of such, and so contains I. $\qquad\qquad\square$

The ideal $I := (A)$ generated by $A := \{\, 12, 18 \,\}$ inside $R := \mathbb{Z}$ is precisely $6\mathbb{Z}$, because we use Bézout coefficients r_1, r_2 to write 6 as $6 = r_1 \cdot 12 + r_2 \cdot 18$, and so 6 lies in $I = (12, 18)$. But then $I = (12, 18) = (6)$.

More generally, the ideal $I := (m_1, m_2, \ldots, m_n)$ generated by some integers $m_1, m_2, \ldots, m_n$ is precisely $I = (m) = m\mathbb{Z}$ generated by their greatest common divisor. So we can think of ideals in rings as a replacement for the concept of greatest common divisor.

The same idea works for polynomials, instead of integers. Working, for example, over the rational numbers, the ideal $I := (p_1(x), p_2(x), \ldots, p_n(x))$ inside the ring $R = \mathbb{Q}[x]$ is just $I = p(x)\mathbb{Q}[x]$ where

$$p(x) = \gcd \{p_1(x), p_2(x), \ldots, p_n(x)\}.$$

In $R := \mathbb{Q}[x, y]$, the story is more complicated. The ideal $I := \left(x, y^2\right)$ is not expressible as $p(x, y)R$, because I cannot be expressed using only one generator $p(x, y)$. If there were such a generator $p(x, y)$, then x and y^2 would have to be divisible by $p(x, y)$, forcing $p(x, y)$ to be constant, so not generating I.

Sage

In sage we construct ideals, for example inside the ring $R = \mathbb{Q}[x, y]$, as:

```
R.<x,y>=PolynomialRing(QQ)
R.ideal( [x^3,y^3+x^3] )
```

We can then test whether two ideals are equal:

```
R.ideal( [x^3,y^3+x^3] )==R.ideal( [x^3,y^3] )
```

yields True.

Fractions

Theorem 15.3. *A commutative ring R lies inside a field k as a subring just when the product of any two nonzero elements of R is nonzero. After perhaps replacing k by the subfield generated by R, the field k is then uniquely determined up to isomorphism, and called the* field of fractions *of R. Every ring morphism $R \to K$ to a field K extends uniquely to an injective field morphism $k \to K$.*

Proof. If a commutative ring R lies in a field k, then clearly any two nonzero elements have nonzero product, as they both have reciprocals. Take two pairs (a, b) and (c, d) with a, b, c, d in R, and $0 \neq b$ and $d \neq 0$. Declare them equivalent if $ad = bc$. Write the equivalence class of (a, b) as $\frac{a}{b}$. Define, as you might expect,

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$
$$\frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd},$$
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd},$$

Tedious checking shows that these are well defined, and yield a field k. The crucial step is that the denominators can't vanish. The reciprocal in k is easy to find:

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

if $a \neq 0$ and $b \neq 0$. Map R to k by $a \mapsto \frac{a}{1}$. Check that this map is injective. Given a ring morphism $f \colon R \to K$ to a field, extend by

$$f\left(\frac{a}{b}\right) = \frac{f(a)}{f(b)}.$$

Again tedious checking ensures that this is defined. Every field morphism is injective. $\square$

The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$.

The field of fractions of the ring

$$\mathbb{Z}\left[\frac{1}{2}\right]$$

of half integers is also $\mathbb{Q}$.

The field of fractions of a field k is k.

The field of fractions of $\mathbb{Z}[x]$ is $\mathbb{Q}(x)$.

> The field of fractions of $\mathbb{Q}[x]$ is also $\mathbb{Q}(x)$.

15.11 Prove that these examples above are correct.

Sage

Sage constructs fields of fractions out of any ring with no zero divisors. For example, let $F := \mathbb{Z}/7\mathbb{Z}$ and $R := F[x, y]$ and let k be the fraction field of R, and then $S = k[t]$ and then compute the resultant of two polynomials $p(t), q(t) \in S$:

```
R.<x,y> = PolynomialRing(GF(7))
k=R.fraction_field()
S.<t> = PolynomialRing(k)
p=S(x*t+t^2/t)
q=S(y*t+1)
p.resultant(q)
```

Algebras

Almost every ring we encounter allows us to multiply by elements from some field.

> Matrices over a field k can be multiplied by elements of k.

> Polynomials over a field k can be multiplied by elements of k.

> Rational functions over a field k can be multiplied by elements of k.

An *algebra* A over a field k is a ring with an operation $s \in k, a \in A \mapsto sa \in A$, called *scalar multiplication* so that

a. $s(a + b) = (sa) + (sb)$,

b. $(r + s)a = (ra) + (sa)$,

c. $(rs)a = r(sa)$,

d. $s(ab) = (sa)b = a(sb)$,

e. $1a = a$,

for $r, s \in k, \ a, b \in A$.

> The field k is an algebra over itself.

> If $k \subset K$ is a subfield, then K is an algebra over k.

If A is an algebra over k, then the $n \times n$ matrices with entries in A are an algebra over k.

If A is an algebra over k, then the polynomials $A[x]$ with coefficients in A are an algebra over k.

Galois theory

> We have made every effort to understand Galois's proof. His reason-
> ing is not sufficiently clear, sufficiently developed, for us to judge its
> correctness and we can give no idea of it in this report.
>
> — Siméon-Denis Poisson

Automorphisms

An *automorphism* of a ring R is a morphism $f\colon R \to R$ which has an inverse $f^{-1}\colon R \to R$, which is also a morphism.

16.1 Prove that the set of all automorphisms of any ring R is a group, the *automorphism group* of R.

> The set of rational numbers of the form $b + c\sqrt{2}$, for $b, c \in \mathbb{Q}$, forms a field,
> usually denoted $\mathbb{Q}(\sqrt{2})$. Indeed, when we add such numbers
>
> $$(7 + 3\sqrt{2}) + (4 - \sqrt{2}) = (7 + 4) + (3 - 1)\sqrt{2}.$$
>
> Similarly if we subtract. If we multiply,
>
> $$(7 + 3\sqrt{2})(4 - \sqrt{2}) = (7 \cdot 4 + 3 \cdot 1 \cdot 2) + (7 \cdot (-1) + 3 \cdot 4)\sqrt{2}.$$
>
> Finally, to compute reciprocals,
>
> $$\frac{1}{7 + 6\sqrt{2}} = \frac{1}{7 + 6\sqrt{2}}\frac{7 - 6\sqrt{2}}{7 - 6\sqrt{2}} = \frac{7 - 6\sqrt{2}}{7^2 - 2 \cdot 6^2}$$
>
> which the reader can simplify. The map
>
> $$f(b + c\sqrt{2}) = b - c\sqrt{2}$$
>
> is an automorphism, i.e. a morphism of the field to itself. Note that $f^{-1} = f$.

16.2 Prove that every morphism $f\colon R \to S$ of rings preserves subtraction.

Suppose that R is a ring. Suppose that, for any elements b, c of R with $c \neq 0$, there is a unique element d of R so that $dc = b$. Write this element as $d = b/c$ and say that R has *right inverses*.

Lemma 16.1. *Every morphism $f\colon R \to S$ of rings with right inverses preserves right inverses.*

Proof. If $d = b/c$ then $dc = b$ so $f(dc) = f(b) = f(d)f(c)$ so $f(d) = f(b)/f(c)$. $\qquad\square$

Lemma 16.2. *Every automorphism of a ring preserves 0. Every automorphism of a ring with identity preserves the identity element.*

Proof. Every element b of R satisfies $b = 0 + b = b + 0$. Any automorphism $h\colon R \to R$ satisfies $h(b) = h(0 + b) = h(0) + h(b) = h(b + 0) = h(b) + h(0)$ for all $b \in R$. Since h has an inverse, h is one-to-one and onto, so every element c of R has the form $c = h(b)$ for some $b \in R$. So $c = h(0) + c = c + h(0)$ for all c in R. In particular, taking $c = 0$, $0 = h(0) + 0 = h(0)$.

Every element b of R satisfies $b = 1b = b1$. Any automorphism $h\colon R \to R$ satisfies $h(b) = h(1b) = h(1)h(b) = h(b1) = h(b)h(1)$ for all $b \in R$. Since h has an inverse, h is one-to-one and onto, so every element c of R has the form $c = h(b)$ for some $b \in R$. So $c = h(1)c = ch(1)$ for all c in R. In particular, taking $c = 1$, $1 = h(1)1 = h(1)$. $\quad\square$

Lemma 16.3. *The only automorphism of any of the fields $\mathbb{Q}, \mathbb{R}$ or $\mathbb{Z}/p\mathbb{Z}$ is the identity map $b \mapsto b$.*

Proof. Suppose that $k = \mathbb{Q}$ or $k = \mathbb{R}$ or $k = \mathbb{Z}/p\mathbb{Z}$. If $f\colon k \to k$ is an automorphism, then $f(1) = 1$. Therefore $f(1+1) = f(1) + f(1) = 1 + 1$. Similarly $f(1+1+1) = 1+1+1$, and so on. Therefore if $k = \mathbb{Z}/p\mathbb{Z}$ we have exhausted all elements of k and $f(b) = b$ for every element b of k. In the same way, if $k = \mathbb{Q}$ or $k = \mathbb{R}$, then $k(n) = n$ for all positive integers n. Similarly, $f(0) = 0$. But then $f(-n) = f(0 - n) = f(0) - f(n) = 0 - f(n) = -f(n)$. So f is the identity map on the integers. Apply f to a ratio b/c of integers, and since f preserves right inverses, $f(b/c) = f(b)/f(c) = b/c$. Hence if $k = \mathbb{Q}$, then f is the identity map.

We can assume that $k = \mathbb{R}$. Say that a number x in $\mathbb{R}$ is *positive* if $x \neq 0$ and $x = y^2$ for some number y. Clearly positivity is preserved under any automorphism. Write $x < y$ to mean that $y - x$ is positive, and $x \leq y$ to mean that $y - x$ is positive or zero. Automorphisms preserve the ordering of real numbers. Take any real number x and approximate x from below by rational numbers $r_1, r_2, \ldots, \to x$ and from above by rational numbers $s_1, s_2, \cdots \to x$. Then $r_i < x < s_i$ so applying $f\colon r_i < f(x) < s_i$. It follows that $f(x) = x$, as real numbers as completely determined by their digits, i.e. by approximation by rationals. $\qquad\square$

Lemma 16.4. *There are two automorphisms of $\mathbb{Q}(\sqrt{2})$: the identity map $\iota(b + c\sqrt{2}) = b + c\sqrt{2}$ and the map*

$$f(b + c\sqrt{2}) = b - c\sqrt{2}.$$

Proof. The same argument as above shows that any automorphism h of $\mathbb{Q}(\sqrt{2})$ fixes all rational numbers $\mathbb{Q} \subset \mathbb{Q}(\sqrt{2})$. So $h(b + c\sqrt{2}) = h(b) + h(c)h(\sqrt{2}) = b + ch(\sqrt{2})$: it suffices to determine $h(\sqrt{2})$. Note that $\sqrt{2}^2 = 2$, so apply h to get $h(\sqrt{2})^2 = 2$, i.e. $h(\sqrt{2}) = \pm\sqrt{2}$. $\qquad\square$

Galois groups

The *Galois group* of a field extension $k \subset K$, denoted $\operatorname{Aut} K/k$ is the group of automorphisms of K which are the identity on k.

We found above that $\operatorname{Aut}\mathbb{Q}(\sqrt{2})/\mathbb{Q} = \{\pm 1\}$ where we let -1 denote the transformation $b + c\sqrt{2} = b - c\sqrt{2}$.

If $b(x)$ is a polynomial with coefficients in k, and g is a automorphism of K fixing all elements of k, then $g(b(x)) = b(x)$, i.e. g fixes all coefficients. Therefore if α is a root of $b(x)$ lying in K, then so is $g\alpha$: $0 = b(\alpha) = \sum b_j\alpha^j$ so

$$0 = g(b(\alpha)) = \sum b_j g(\alpha)^j.$$

Hence the Galois group $\operatorname{Aut}K/k$ of a field extension K/k acts on the roots in K of all polynomials over k. If a polynomial splits into factors over k, then clearly the Galois group $\operatorname{Aut}K/k$ permutes the K-roots of each factor individually.

The Galois group of $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ permutes $\sqrt{2}, -\sqrt{2}$, the roots of $x^2 - 2$. Note that the elements of $\mathbb{Q}(\sqrt{2})$ have the form $a + b\sqrt{2}$, for a, b rational. Hence the elements of $\mathbb{Q}(\sqrt{2}$ fixed by the Galois group are precisely the elements of $\mathbb{Q}$.

A *Galois extension* is an extension K/k so that the elements of K fixed by $\operatorname{Aut}K/k$ are precisely the elements of k.

In chapter 11 we checked that the field $K := \mathbb{Q}(\sqrt[3]{2})$ consists of the numbers $a + b2^{1/3} + c2^{2/3}$ for a, b, c rational. Note that K is a subfield of the field of real numbers. There is only one real number α that satisfies $\alpha^3 = 2$: the real number $\alpha = 2^{1/3}$. Therefore there is only one element α of K that satisfies $\alpha^3 = 2$: the element $\alpha = 2^{1/3}$. So every automorphism of K sends $2^{1/3}$ to itself. The Galois group of $K/\mathbb{Q}$ is $\operatorname{Aut}K/\mathbb{Q} = \{1\}$. The extension $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$ is *not* a Galois extension; the Galois group is too small, fixing everything.

Let K be the splitting field of $p(x) = x^3 - 2$ over $k = \mathbb{Q}$. There are 3 complex cube roots of 2: the real number $2^{1/3}$, and its two rotations by a third of a revolution around the origin: call them α and $\bar{\alpha}$. As we will see, by theorem 19.8 on page 163, because $p(x)$ is irreducible over k, any two of these roots are swapped by some element of the Galois group of K/k. We can draw one of these permutations: complex conjugation preserves $2^{1/3}$, and preserves the polynomial, and swaps the angles of the other roots, preserving their lengths, so swaps the other two roots. Any permutation of the three roots arises as a unique element of the Galois group: once we swap roots to get any root we like to sit in the spot $2^{1/3}$, we can then permute the other two by complex conjugation if they are not already where we want them. So the Galois group of K/k is the symmetric group on 3 letters. In particular, K/k is a Galois extension.

Adding roots

Start with a field k, and take two elements $b, c \in k$. The roots of the equation $0 = x^2 + bx + c$ lie in some extension K. To be specific, the quadratic formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2}$$

shows that the roots of our quadratic equation lie in the field $K = k(\sqrt{b^2 - 4ac})$, as long as k has characteristic not equal to 2. In fields of characteristic 2, the quadratic formula doesn't hold, and the splitting field is more complicated.

Theorem 16.5. *Suppose that K is a splitting field of an irreducible polynomial $p(x)$ over a field k. If $\alpha, \beta \in K$ are roots of $p(x)$ then the Galois group of K over k has an element which takes α to β. In particular, K is a Galois extension of k.*

We will prove this result in chapter 19.

Return to our previous example: in chapter 11 we checked that the field $K := \mathbb{Q}(\sqrt[3]{2})$ consists of the numbers $a + b2^{1/3} + c2^{2/3}$ for a, b, c rational. Above, we saw that K has trivial Galois group over $k = \mathbb{Q}$. The polynomial $p(x) = x^3 - 2$ has a root in K, but not all of its roots. Our theorem says that the splitting field L of $p(x)$ over k is Galois. Let $\alpha = \sqrt[3]{2}$, so $K = k(\alpha)$. Factor $p(x) = (x - \alpha)(x^2 + \alpha x + \alpha^2)$. But $x^2 + \alpha x + \alpha^2$ has no roots in K, since its roots are not real (as we will see), while K lies inside $\mathbb{R}$. We need to extend to L to find roots for $x^2 + \alpha x + \alpha^2$, i.e. the two remaining roots of $p(x)$. Since $x^2 + \alpha x + \alpha^2$ is quadratic, L is a quadratic extension of K, i.e. $L = K(\beta)$ is given by adding

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-\alpha \pm \sqrt{\alpha^2 - 4\alpha^2}}{2} = -\frac{\alpha}{2}\left(1 \pm \sqrt{-3}\right),$$

i.e. $L = K(\sqrt{-3}) = k(\sqrt[3]{2}, \sqrt{-3})$. Automorphisms of L fixing k are determined by their action on $\sqrt[3]{2}$ and $\sqrt{-3}$. Complex conjugation is one such automorphism. Let ω be the cube root of 1 lying in the upper half plane. The three roots of $p(x)$ are $\sqrt[3]{2}, \omega\sqrt[3]{2}$ and $\bar{\omega}\sqrt[3]{2}$. Complex conjugation swaps the last two roots. Some other automorphism in the Galois group swaps the first two roots, by the theorem. That map then must leave the other root alone, as it must remain a root. So we can identify two automorphisms in the Galois group. Composing these two, we see that every permutation of the roots $\sqrt[3]{2}, \omega\sqrt[3]{2}$ and $\bar{\omega}\sqrt[3]{2}$ is obtained by a unique automorphism in the Galois group of L over k, i.e. $\operatorname{Aut} L/k$ is the symmetric group on three letters.

Pick a prime p. Over $k = \mathbb{Q}$, factor $x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$. The second factor is irreducible by problem 10.7 on page 78. The Galois group then takes any p-th root of 1 to any other (except for 1 itself, which is fixed). Once we take the root $\omega = e^{2\pi i/5}$ to some other root, ω^n, this determines the

automorphism, as all other roots are powers of ω. The Galois group is thus given by automorphisms $\phi_n(\omega) = \omega^n$, which clearly commute, forming the cyclic group of order 5.

For abstract variables $x_1, \ldots, x_n$ and any field k, the field $L = k(x_1, \ldots, x_n)$ contains the subfield $K = k(e_1, \ldots, e_n)$ generated by the elementary symmetric polynomials in $x_1, \ldots, x_n$. By problem 14.6 on page 124, this subfield is precisely the elements invariant under permutations of the variables. Hence its Galois group $\operatorname{Aut} L/K$ is the symmetric group on n letters.

16.3 Suppose that $p(x)$ is a polynomial with coefficients in a field k, and with roots in some extension K of k, say $\alpha_1, \alpha_2, \ldots, \alpha_n$. Prove that, for any symmetric polynomial q with coefficients in k, $q(\alpha_1, \ldots, \alpha_n)$ belongs to k.

Theorem 16.6. *Suppose that K is the splitting field of an irreducible monic polynomial $p(x)$ over a field k. Let d be the discriminant of $p(x)$, i.e. the resultant of $p(x), p'(x)$. Then $(-1)^{n(n-1)/2}d$ is a square in K, say α^2, so has a square root α in K. The field extension $k(\alpha) \subset K$ has Galois group $\operatorname{Aut} K/k(\alpha) \subset \operatorname{Aut} K/k$ consisting precisely of the elements of $\operatorname{Aut} K/k$ which act as even permutations of the roots.*

Proof. Problem 13.10 on page 108 shows that

$$d = (-1)^{n(n-1)/2} \prod_{i<j}(x_i - x_j)^2,$$

so we can take

$$\alpha = \prod_{i<j}(x_i - x_j),$$

or any multiple by any square of -1, or the same expression after any permutation of roots. This α changes sign when we permute roots, according to the sign of the permutation.

Every automorphism of K fixing all elements of $k(\alpha)$ fixes all elements of $k \subset k(\alpha)$, so $\operatorname{Aut} K/k(\alpha) \subset \operatorname{Aut} K/k$. Given an automorphism g in $\operatorname{Aut} K/k$, it permutes the roots of $p(x)$, so acts on α as the sign of that permutation, so fixes α just when it is even as a permutation. But it fixes k in any case, so fixes $k(\alpha)$ just when it fixes α. $\qquad\square$

Return to the splitting field K of $x^3 - 2$ over $k = \mathbb{Q}$. We saw that $\operatorname{Aut} K/k$ is the symmetric group on three letters. We compute the discriminant $\Delta = \Delta_{x^3-2} = 108$. So $d = (-1)^{3(2)/2}\Delta = -108$, which has no square root in $k = \mathbb{Q}$. We adjoin a square root to -108, or equivalently to -3 since $-108 = 2^2 3^2(-3)$, so $\alpha = \sqrt{-3}$, and $\operatorname{Aut} K/k(\alpha)$ is the alternating group on three letters.

Radical extensions

A *simple extension* of a field k is an extension K, denoted $k(\alpha)$, with an element α of K, so that every element of K is a rational function $b(\alpha)/c(\alpha)$ with coefficients in k. A *simple radical extension* of a field k is an extension $k(\alpha)$ for which α satisfies an equation $\alpha^n = c$ for some element c of k; we usually write such an extension as $k(\sqrt[n]{c})$. A *radical extension* is the result of repeated simple radical extensions $k(\alpha_1)(\alpha_2)\dots(\alpha_n)$.

> If we allow repeated radical extensions, we might as well extend by a prime root each time, since $\sqrt[6]{\alpha}$ can be added by first adding $\beta = \sqrt[2]{\alpha}$ and then adding $\sqrt[3]{\beta}$.

> The extension $\mathbb{Q}(x, y)/\mathbb{Q}$ (adding two abstract variables) admits an automorphism $x \mapsto y$, $y \mapsto x$. But the further extension $\mathbb{Q}(\sqrt{x}, y)$ doesn't admit this automorphism, since it has no $\sqrt{y}$. This is easy to fix: just add a $\sqrt{y}$.

Lemma 16.7. *Suppose that K is an extension of a field k, and that $\operatorname{Aut} K/k$ is a finite group. For every radical extension E of K, there is a radical extension F of E so that the action of $\operatorname{Aut} K/k$ on K extends to an action on F.*

Proof. If we have some element α of K and some radical $\sqrt[n]{\alpha}$, we just need to add some element $\sqrt[n]{g\alpha}$ for every g in $\operatorname{Aut} K/k$, and repeat as needed. $\square$

Chapter 17

Algebraic curves in the plane

Everyone knows what a curve is, until he has studied enough mathematics to become confused through the countless number of possible exceptions.

— Felix Klein

Example

Let's find the solutions of the algebraic equation $y^2 = x^2 + x^3$. The solutions form a curve in the plane, intuitively because the equation is one constraint on two variables. There is one obvious solution: $(x, y) = (0, 0)$. A minor miracle: almost every line through $(0, 0)$ passes through another point of the curve, which we can compute. To see this, write any line through $(0, 0)$, say with slope t, as $y = tx$. Plug this in to the equation $y^2 = x^2 + x^3$ to see if we can find any more solutions:

$$(tx)^2 = x^2 + x^3,$$

which we simplify to get either $x = 0$ or, dividing by x,

$$t^2 = 1 + x.$$

Solving for x, we find $x = t^2 - 1$. Plug in to the equation of the line $y = tx$ to get $y = t\left(t^2 - 1\right)$. Every solution, except $(x, y) = (0, 0)$, has to lie on some line through $(0, 0)$, and that line will be either $y = tx$ or a vertical line $x = 0$. But on the line $x = 0$, the only solution is $y = 0$, the origin. Moreover, if we let $t = \pm 1$, we get the solution $(x, y) = (0, 0)$ too. So we get all of the solutions.
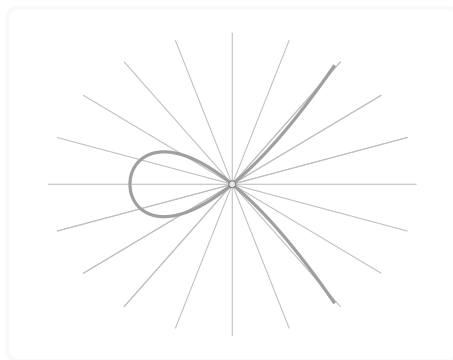
Each line strikes the curve at two points (one of which is the origin); except for the vertical line.

Definition

An *algebraic curve* is the set of zeroes of a nonconstant polynomial in two variables; the *degree* of the curve is the degree of the polynomial.

degree	appellation
1	line
2	conic
3	cubic
4	quartic
5	quintic
6	sextic
7	septic
8	octic

17.1 Use the same trick for the circle $x^2 + y^2 = 1$ as follows. Show first (using the formula for the slope of a line) that for any point (x, y) of the circle, the line through $(0, 1)$ and (x, y) passes through the horizontal axis at the point $(t, 0)$ where

$$t = \frac{x}{1 - y}.$$

It helps to draw a picture. Solve for x: $x = t(1 - y)$, and plug into the equation of the circle to find that

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

Explain now why every rational point of the circle corresponds to a rational value of t and vice versa. Explain why the same conclusion holds over any field k.

Let's try a different cubic curve. Take the cubic curve

$$y^2 = (x + 1)x(x - 1).$$
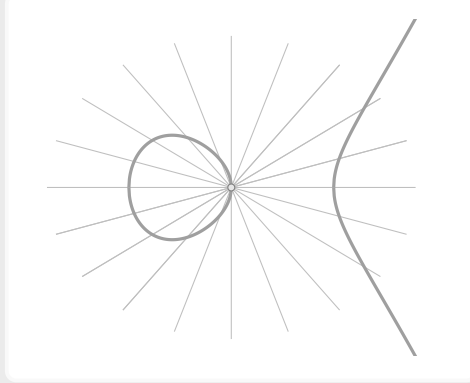
We see right away that some lines strike the curve in more than two points, one at the origin, and two others. Calculate as above that those two others are

$$x = \frac{t^2 \pm \sqrt{t^4 + 4}}{2}, \; y = tx.$$

Fermat's Last Theorem says that there are no integer solutions to $a^n + b^n = c^n$ with nonzero integers a, b, c and $n \geq 3$. Dividing both sides by c, if there were a solution, we would solve

$$x^n + y^n = 1$$

with nonzero rational numbers x, y. Conversely, any nonzero rational numbers x, y solving this equation, after clearing denominators, give a solution to $a^n + b^n = c^n$. So Fermat's Last Theorem is equivalent to the statement that the algebraic curve $x^n + y^n = 1$ has no rational number solutions except $(x, y) = (0, 1), (0, -1), (1, 0)$ and $(-1, 0)$. In particular, we are interested in algebraic curves over the rational numbers, not just the real numbers.

We make our first, naive, attempt to define algebraic curves. An *algebraic curve* is the set of points (x, y) of the plane satisfying an equation $0 = f(x, y)$, where $f(x, y)$ is a nonconstant polynomial.

Consider the polynomial equation $x^2 + y^2 = -7$. There are no real values of x, y that satisfy this, so the algebraic curve is empty.

To remedy this problem, we allow the variables x and y to take on complex values. Sticking in a value for one of the variables, or perhaps for the other, with a suitable choice of value, we will find that we have a nonconstant polynomial in one variable, so there is a solution (x, y) by the fundamental theorem of algebra. Hence algebraic curves have complex points, but perhaps no real points.

To study the existence of rational solutions to algebraic equations we can clear denominators as we did above for Fermat's Last Theorem, and look for integer solutions to different algebraic equations. We can then look at those equations modulo a prime,

and get equations over a finite field. So it is natural to consider algebraic curves over any field.

If k is a field, an algebraic curve X over k given by a nonconstant polynomial equation $f(x, y) = 0$ with coefficients in k is the set X of its $\bar{k}$-points, i.e. points (x, y) with x, y in the algebraic closure $\bar{k}$ satisfying $f(x, y) = 0$. There are infinitely many $\bar{k}$-points, because we can force either x or y to any constant value, and for infinitely many such constants $f(c, y)$ or $f(x, c)$ is not constant so has a root in $\bar{k}$.

The equation of an algebraic curve is not uniquely determined. The curve $x^2 - y = 0$ is the same curve as $\left(x^2 - y\right)^9 = 0$. The curve $xy = 0$ splits into $x = 0$ or $y = 0$, two curves. If the equation of a curve has nonconstant factors, we say that the curve is *reducible* and the equations formed from the factors, or the curves they cut out, are *components* of the curve. We can assume that the polynomial $f(x, y)$ in the equation $f(x, y) = 0$ of a curve is not a square, cube, etc. of any polynomial, nor is any constant multiple of $f(x, y)$.

Lemma 17.1. *An irreducible algebraic curve determines its equation uniquely up to scaling by a nonzero constant.*

Proof. Work over a field k with algebraic closure $\bar{k}$. Take two irreducible equations $0 = b(x, y)$ and $0 = c(x, y)$ with the same curve, i.e. the same points over $\bar{k}$. For any constant $x = a$ over $\bar{k}$, the equations $0 = b(a, y)$ and $0 = c(a, y)$ have the same solutions. Unless the x variable does not appear in either equation, these are infinitely many different points in common on the curve. But if x does not appear, swap x with y and repeat. By proposition 14.10 on page 124, $b(x, y)$ and $c(x, y)$ have a common factor. $\qquad\square$

A *regular function* on an algebraic curve is the restriction of a polynomial function. For example, on the algebraic curve $y^2 = x^2 + x^3$, the functions 0 and $y^2 - x^2 - x^3$ take on the same values at every point, and therefore are equal as regular functions. If C is an algebraic curve over field k, let $k[C]$ be the set of regular functions on C with coefficients in k.

A *regular morphism* of algebraic curves $f \colon C \to D$ is a map which can be expressed somehow as $f(x, y) = (s(x, y), t(x, y))$ for two polynomials $s(x, y), t(x, y)$, so that f, applied to any point of C, yields a point of D. For example, the map $f(x, y) = (s, t) = \left(1 - x^2, x\right)$ maps $(y = 0)$ to $t^2 = s^2 + s^3$, as we saw previously. A regular morphism with a regular inverse is *biregular*, and the associated curves are *biregular*. Clearly biregular algebraic curves have isomorphic algebras of regular functions.

Plane conics

We will see that, over any field of characteristic not 2, every conic is biregular to precisely one of

Equation	Geometry	Regular function algebra
$y = 0$	line	$k[x]$
$xy = 0$	pair of intersecting lines	$k[x] \oplus k[y]$
$xy = 1$	hyperbola	$k\left[x, x^{-1}\right]$
$y = x^2$	parabola	$k[x]$
$y = \pm 1$	pair of disjoint lines	$k[x] \oplus k[x]$
$y = \pm\alpha$	pair of disjoint lines over $k(\alpha), \alpha^2 \in k, \alpha \notin k$	$k(\alpha)[x]$

Take a conic $0 = f(x, y)$ and expand out

$$f(x, y) = ax^2 + bxy + cy^2 + rx + sy + t.$$

If $0 = a = b = c$ then this is a line, not a conic, so we can assume that at least one of a, b, c is not zero. First, suppose that $0 = a = c$, so

$$f(x, y) = bxy + rx + sy + t.$$

Rescale to arrange that $b = 1$. Factor as

$$f(x, y) = (x + s)(y + r) + t - rs.$$

Change variables by translation, replacing $x + s$ by x and $y + r$ by y and let $u := rs - t$ to get

$$f(x, y) = xy - u.$$

If $u = 0$, then our conic is a pair of lines intersecting at a single point. If $u \neq 0$, rescale x by $1/u$ to arrange that our conic is $xy = 1$.

Suppose that one or more of a or b is not zero, and swap variables if needed to arrange that $a \neq 0$ and rescale the equation $0 = f(x, y)$ to get $a = 1$.

$$f(x, y) = x^2 + bxy + cy^2 + rx + sy + t.$$

Suppose we work over a field not of characteristic 2, i.e. where $2 \neq 0$, so we can divide by 2:

$$f(x, y) = \left(x + \frac{by}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)y^2 + rx + sy + t.$$

Replace $x + by/2$ by a new variable called x and replace $c - b^2/4$ by a new constant called c:

$$f(x, y) = x^2 + cy^2 + rx + sy + t.$$

Complete the square in x

$$f(x, y) = \left(x + \frac{r}{2}\right)^2 + cy^2 + rx + sy + t - \frac{r^2}{4}.$$

Rename $x + r/2$ to x and $t - r^2/4$ to t:

$$f(x, y) = x^2 + cy^2 + sy + t.$$

If $c = 0$ then

$$f(x, y) = x^2 + sy + t.$$

If $s \neq 0$ then replace $sy + t$ by a new variable, which we call $-y$, to get

$$f(x, y) = x^2 - y.$$

Note that then our curve is a parabola $y = x^2$. On the parabola, every polynomial $g(x, y)$ is just $g(x, x^2)$, a polynomial in x. If $0 = s = t$ then our conic is actually a line $x = 0$; linearly transform variable to get the conic to be $y = 0$. If $0 = s$ and $t \neq 0$ then our conic is $x^2 = -t$. Again linearly transform to get the conic $y^2 = -t$. If $-t$ has a square root in k, say a, then our equation is $y = \pm a$, and rescale the y variable to get $a = 1$, so a pair of disjoint lines $y = \pm 1$. If $-t$ has no square root in k, then an empty set of points with coordinates in k, but a pair of disjoint lines with coordinates in $k(\alpha)$, $\alpha = \sqrt{-t}$: $y = \pm \alpha$. This α is uniquely determined up to scaling by an element of $k^\times$. For example, over the rational numbers, the equations $x^2 = -t$ for t any prime or $x^2 = t$ for t any prime give conics with no rational points. Different primes give different conics, and the reader can work out why they are not biregular over the rational numbers. One the other hand, over the real numbers, we only have a single conic $x^2 = -1$ with no real points.

Rational functions

A *rational function* on an irreducible algebraic curve X over a field k is a rational function of x, y, whose denominator is not divisible by any nonconstant factor of the equation of the curve.

> On the curve $X = (y = x^2)$, the rational function $f = x/(x - y)$ might seem to be undefined near the origin, but we can also write it as
>
> $$f = \frac{x}{x - y},$$
> $$= \frac{x}{x - x^2},$$
> $$= \frac{1}{1 - x}.$$

> The circle $X = (x^2 + y^2 = 0)$ is irreducible over $k = \mathbb{R}$, so has well defined field of rational functions $\mathbb{R}(X)$. Note that X is reducible over $\mathbb{C}$: $x^2 + y^2 = (x + iy)(x - iy)$, so we can't make a field $\mathbb{C}(X)$, of complex coefficient rational functions, since $x + iy, x - iy$ are two rational functions which multiply to zero as complex coefficient rational functions on X. Elements of $\mathbb{R}(X)$ are ratios
>
> $$\frac{b(x) + c(x)y}{d(x) + e(x)y},$$

for any $b(x), c(x), d(x), e(x)$ real coefficient polynomials, but with $y^2 = -x^2$, i.e. we can formally write y as ix, allowing complex coefficients into any terms of positive degree in x.

17.2 Find the field of rational functions on every conic, over any field not of characteristic 2.

A rational function f is *regular* near a point of X if f can be written as $f = b/c$ for some polynomials b, c with c nonzero at that point.

A *rational morphism* of algebraic curves $f \colon C \to D$ is a map, which can be expressed somehow as $f(x, y) = (s(x, y), t(x, y))$ for two rational functions $s(x, y), t(x, y)$, each of which is regular at some point of C, so that f, applied to any point of C at which both $s(x, y)$ and $t(x, y)$ are regular, yields a point of D. A rational morphism with a rational inverse is *birational*, and the associated curves are *birational*. A rational morphism is *regular* near a point of C if $s(x, y)$ and $t(x, y)$ are regular near that point, i.e. can be expressed near that point as ratios of polynomials not vanishing near that point.

17.3 Prove that any algebraic curve is irreducible just when its algebra of rational functions forms a field.

Lemma 17.2. *Irreducible algebraic curves in the plane, over any field k, are birational just when their fields of rational functions are isomorphic extensions of k.*

Proof. Suppose that the curves are birational. Clearly the rational functions compose with the birational map, and with its inverse, identifying the fields. Conversely, suppose that the fields are isomorphic, say by an isomorphism

$$\phi \colon k(C) \to k(D).$$

Write the coordinates of the plane in which C lives as x, y, and those of the plane in which D lives as u, v. Let $s(x, y) := \phi^{-1} u$ and $t(x, y) := \phi^{-1} v$. $\qquad \square$

A curve birational to a line in the plane (with equation $y = 0$, for example) is a *rational curve*. An algebraic curve C is rational just when $k(C) \cong k(x)$.

17.4 Imitate our picture of drawing lines through a point of $y^2 = x^3 + x^2$, but replacing that cubic curve by an irreducible conic. Prove that conics are rational.

17.5 Prove that every algebraic curve over any algebraically closed field has infinitely many points. Hints: Pick a point not on the curve. (You need to prove that such a point exists, using the fact that every algebraically closed field is infinite.) Change variables by translation to arrange that this point is the origin. Any line through the origin strikes the curve at some point, except perhaps for finitely many exceptional lines. (Why?) Two lines through the origin intersect only at the origin. So any two lines through the origin strike the curve at different points. There are infinitely many lines through the origin. (Why?)

Integrals

Suppose that $g(x, y) = 0$ is the equation of an irreducible plane algebraic curve over the field of real numbers, for example $y^2 = x^2 + x^3$. Suppose that we can locally write the curve as the graph of a function $y = y(x)$, for example $y = \sqrt{x^2 + x^3}$. Pick a rational function $f(x, y)$ on the curve, and consider the integral $\int f \, dx$, by which we mean

$$\int f(x, y) \, dx = \int f(x, y(x)) \, dx.$$

If the curve is rational, then we can parameterize it by $x = x(t), y = y(t)$, as rational functions of a new variable t. So then the integral becomes

$$\int f \, dx = \int f(x(t), y(t)) \frac{dx}{dt} \, dt$$

and this is a composition of rational functions, so a rational function of t. Using partial fractions, we can integrate this explicitly. For example, on the curve $y^2 = x^2 + x^3$, if we take the rational function

$$f := \frac{x^3}{y},$$

we can write our integral as

$$\int f \, dx = \int \frac{x^3 \, dx}{\sqrt{x^2 + x^3}}.$$

We then solve this integral by parameterising as on page 139,

$$x = x(t) = t^2 - 1, y = y(t) = t \left(t^2 - 1 \right)$$

to get

$$\int \frac{x^3 \, dx}{\sqrt{x^2 + x^3}} = \int \frac{x^3 \, dx}{y},$$

$$= \int \frac{\left(t^2 - 1 \right)^3 2t \, dt}{t \left(t^2 - 1 \right)},$$

$$= 2 \int \left(t^2 - 1 \right)^2 \, dt,$$

17.6 Simplify this integral completely into a function of x.

17.7 Recall that the circle is rational, as it can be parameterized by

$$x(t) = \frac{1 - t^2}{1 + t^2}, y(t) = \frac{2t}{1 + t^2}.$$

But the circle can also be parameterized using trigonometry:

$$x(\theta) = \cos \theta, y(t) = \sin \theta.$$

Use these two facts to explain how to solve all integrals of the form

$$\int f(\cos \theta, \sin \theta) \, d\theta,$$

where f is any rational function.

Ideals

Given an algebraic curve C and a point $p \in C$ (say, to be precise, that $p \in C(k)$ is a point defined over k), we are naturally interested in the regular functions on C vanishing at p. If p has coordinates $p = (x_0, y_0)$, then the polynomial functions $x - x_0, y - y_0$ vanish simultaneously only at p. For simplicity, translate the plane so that p lies at the origin. A regular function on C vanishes at p just when it is a sum $xg(x, y) + yh(x, y)$, since every polynomial function vanishing at the origin is expressible in a finite Taylor series expansion. In other words, the regular functions on C vanishing at the origin are precisely the ideal $I = (x, y)$. Similarly, the regular functions on C vanishing at $(x, y) = (x_0, y_0)$ are the polynomials of the form $(x - x_0) g(x, y) + (y - y_0) h(x, y)$, so constitute the ideal $I = (x - x_0, y - y_0)$. This is the motivation for the name "ideal": an ideal is like a idealized notion of a point.

The regular functions on the real number line are the polynomials $p(x)$, and those which vanish at a point $x = x_0$ are those of the form $(x - x_0) p(x)$. Similarly, if we take two points $x_0 \neq x_1$, then the regular functions on the line vanishing at both points are the polynomials of the form $(x - x_0) (x - x_1) p(x)$, the polynomials divisible $(x - x_0) (x - x_1)$. If we imagine running the two points into one another, so that they collide at $x = 0$, then these polynomials approach polynomials divisible by x^2. We could think of the equation $x^2 = 0$, or of the ideal $I = (x^2)$, as representing a "double point".

Working over $k := \mathbb{Q}$ on the curve $C := (x^2 = y)$, the ideal $I := (x - 2)$ does not correspond to any pointed defined over k. Again, we can think of I as an "ideal point".

Suppose that C is an algebraic curve and $p = (x_0, y_0)$ is a point of C. Let I be the ideal of all regular functions on C vanishing at p. Then I is the kernel of the morphism $g(x, y) \mapsto g(x_0, y_0)$, which is a morphism $k[C] \to k$. So we think of all ideals of $k[C]$ as if they were "ideal points".

Sage

Sage can usually plot the real points of an algebraic plane curve. For any equation like $y^3 + x^3 - 6x^2y = 0$ the code

```
x,y=var('x,y')
contour_plot(y^3+x^3-6*x^2*y==0, (x,-10,10), (y,-10,10))
```

yields a picture of the level sets of the function $y^3 + x^3 - 6x^2y$:
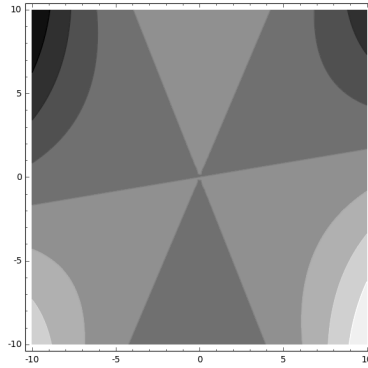
from which we can see that our curve is a union of three lines. Similarly

```
x,y=var('x,y')
contour_plot(y^2-x*(x-1)*(x-2)==0, (x,-3,4), (y,-5,5))
```

yields level sets like
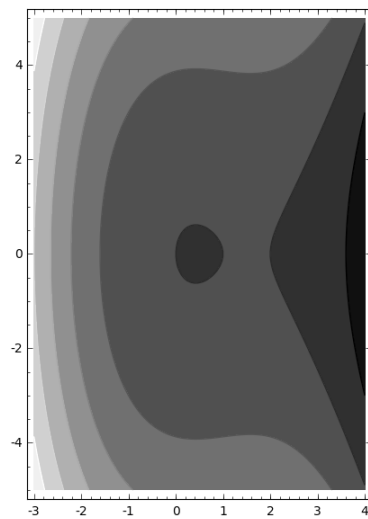
and we can also plot the algebraic curve, without the other level sets, using

```
f(x,y) = y^2-x*(x-1)*(x-2)
implicit_plot(f, (-3, 4), (-5, 5))
```

yielding
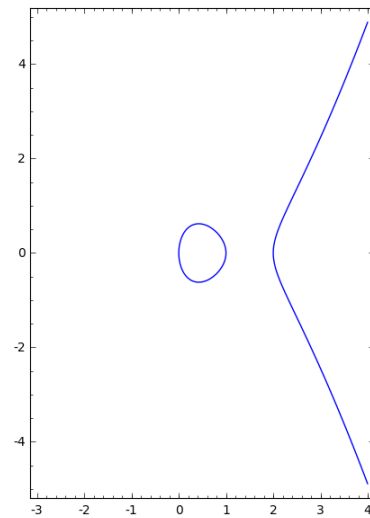
Similarly, we can draw algebraic surfaces:

    f(x,y,z) = y^2-x*(x-1)*(x-2)+z^4-1
    implicit_plot3d(f, (-3, 4), (-5, 5), (-1,1))

yielding
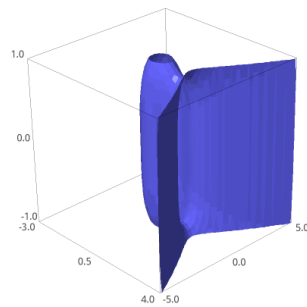
Where plane curves intersect

*Among their codified and systematic sciences is geometry, which is to-
tally useless. The sum of the three angles in a triangles is two right
angles—what benefit does it have? These theorems that are close to
their hearts—what purpose do they serve?*

— Ahmad al-Fārūqī al-Sirhindī
Maktubat

Vanishing polynomials

Over the finite field with two elements, $t(t+1)$ vanishes for any value of t in
that field.

Over any finite field, the polynomial

$$q(t) = \prod_c (t - c)$$

(where the product is over all constants c in the field) vanishes for any value
of the variable t in that field.

Lemma 18.1. *Take a nonzero polynomial in several variables, over a field k, that
vanishes for all values of those variables. Then the field k is finite and the polynomial
is expressible as*

$$p(x) = \sum_{i=1}^n p_i(x) q(x_i)$$

where $x = (x_1, x_2, \ldots, x_n)$ and each $p_i(x)$ is a polynomial and

$$q(t) = \prod_c (t - c)$$

*is our polynomial that vanishes for all values of t in k. In particular, $p(x)$ has degree
at least equal to the number of elements in the field in at least one of the variables.*

Proof. In one variable, we can factor each root as in corollary 8.2 on page 56. Suppose
we have two variables x, y and a polynomial $p(x, y)$. Set y to zero, and find that

by induction, the resulting polynomial $p(x, 0)$ is divisible as required by $q(x)$, say $p(x, 0) = q(x)p_1(x)$. So $p(x, y) = q(x)p_1(x) + yh(x, y)$, say. It is good enough to prove the result for $yh(x, y)$ and add to $q(x)p_1(x)$. So we can assume that $p(x, y) = yh(x, y)$. Moreover, since $p(x, y)$ vanishes for all x, y values in our field, $h(x, y)$ vanishes for all x, y values in our field as long as $y \neq 0$.

Define a polynomial $\delta(t)$ by

$$\delta(t) = \prod_{c \neq 0} \left(1 - \frac{t}{c} \right),$$

where the product is over all nonzero constant elements c in our field. Check that $\delta(0) = 1$ while $\delta(b) = 0$ for any nonzero element b of our field. Therefore

$$h(x, y) - \delta(y)h(x, 0)$$

vanishes for all values of x, y in our field. By induction on degree, we can write

$$h(x, y) - \delta(y)h(x, 0) = h_1(x, y)q(x) + h_2(x, y)q(y).$$

Plugging back into $p(x, y)$ gives the result. $\square$

Linear factors

A *linear function* is a polynomial of the form

$$f(x_1, x_2, \ldots, x_n) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n.$$

If not all of the coefficients $a_1, a_2, \ldots, a_n$ are zero, the set of points $x = (x_1, x_2, \ldots, x_n)$ at which $f(x) = 0$ is called a *linear hyperplane*.

Lemma 18.2. *Suppose that in variables* $x = (x_1, x_2, \ldots, x_n)$

    a. $p(x)$ *is a polynomial and*

    b. $f(x)$ *is a nonzero linear function and*

    c. $p(x) = 0$ *at every point* x *where* $f(x) = 0$ *and*

    d. *the field we are working over contains more elements than the degree of* $p(x)$ *in any variable.*

Then $f(x)$ *divides* $p(x)$, *i.e. there is a polynomial* $q(x)$ *so that* $p(x) = q(x)f(x)$.

Proof. By a linear change of variables, we can arrange that $f(x) = x_1$. Expand $p(x)$ in powers of x_1. If there is a "constant term", i.e. a monomial in $x_2, x_3, \ldots, x_n$, then setting $x_1 = 0$ we will still have some nonzero polynomial in the variables $x_2, x_3, \ldots, x_n$. But this polynomial vanishes for all values of these variables, so is zero by lemma 18.1 on the previous page. $\square$

Resultants in many variables

Lemma 18.3. *Take two polynomials $b(x), c(x)$ in a variable x, with coefficients in a commutative ring S, of degrees m, n. Then the resultant $r = \mathrm{res}_{b,c}$ is expressible as $r = u(x)b(x) + v(x)c(x)$ where $u(x)$ and $v(x)$ are polynomials, with coefficients in the same commutative ring S, of degrees $n - 1, m - 1$.*

The proof is identical to the proof of lemma 13.3 on page 103.

Take two polynomials $b(x, y)$ and $c(x, y)$ and let $r(x)$ be the resultant of $b(x, y), c(x, y)$ in the y variable, so thinking of b and c as polynomials in y, with coefficients being polynomials in x. So $r(x)$ is a polynomial in x.

If $b(x, y) = y + x$, $c(x, y) = y$, then $r(x) = x$ vanishes just at the value $x = 0$ where there is a common factor: y.
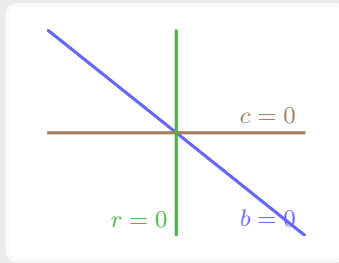
Let $b(x, y) := xy^2 + y$, $c(x, y) := xy^2 + y + 1$.
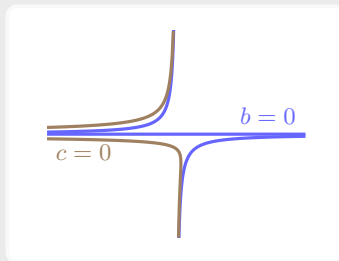
Look at the picture at $x = 0$ and near $x = 0$: because the polynomials drop degrees, the number of roots of $b(0, y)$ on the vertical line $x = 0$ is smaller than the number of roots of $b(a, y)$ on the vertical line $x = a$ for constants $x = a$ near 0. We can see roots "flying away" to infinity on those lines.
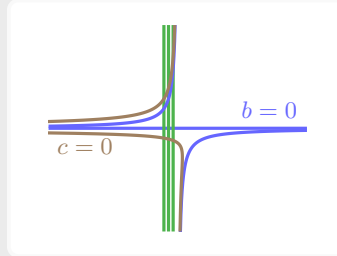
Finding the determinant of the associated 4×4 matrix tells us that $r(x) = x^2$, which vanishes just at the value $x = 0$ where $b(0, y) = y$ and $c(0, y) = y + 1$ drop in degrees, *not* due to a common factor.
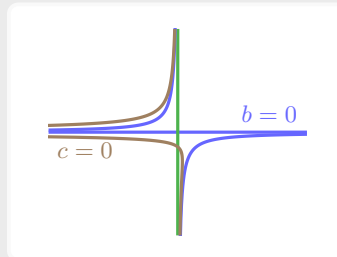
The resultant of y and $y + 1$ is *not* $r(0)$ since the degrees drop, so we would compute resultant of y and $y + 1$ using a 2×2 matrix, and find resultant -1.

If $b(x, y) = xy^2 + y$ and $c(x, y) = 2xy^2 + y + 1$ then $r(x) = x(x + 1)$ vanishes at $x = 0$ and $x = -1$. At $x = 0$, $b(0, y) = y, c(0, y) = y + 1$ have no common factor, but drop degrees. At $x = -1$, $b(-1, y) = -y(y - 1)$, $c(-1, y) = -2(y - 1)(y - 1/2)$ have a common factor, but they don't drop degrees.
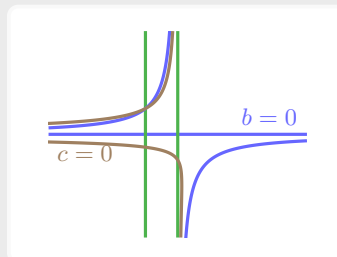
If $b(x, y) = x^2 + y^2 - 1$ and $c(x, y) = (x - 1)^2 + y^2 - 1$ then $r(x) = (2x - 1)^2$ vanishes at $x = 1/2$, where there are *two* different intersection points, a double

common factor:

$$b(1/2, y) = c(1/2, y) = \left(y - \frac{\sqrt{3}}{2}\right)\left(y + \frac{\sqrt{3}}{2}\right).$$
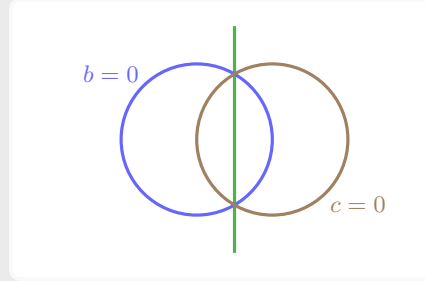
Lemma 18.4. *Suppose that k is a field and*

$$x = (x_1, x_2, \ldots, x_n)$$

are variables and y is a variable. Take $b(x, y)$ and $c(x, y)$ two nonconstant polynomials in $k[x, y]$. Then in some finite degree extension of k there are constants

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

so that (writing λy for $\sum \lambda_i y_i$) in the expressions $b(x + \lambda y, y)$, $c(x + \lambda y, y)$, the coefficients of highest order in y are both nonzero constants.

Return to our earlier example of $b(x, y) = xy^2 + y$ and $c(x, y) = 2xy^2 + y + 1$ and let

$$B(x, y) := b(x + \lambda y, y) = \lambda y^3 + xy^2 + y,$$
$$C(x, y) := c(x + \lambda y, y) = 2\lambda y^3 + 2xy^2 + y + 1.$$

For any nonzero constant $\lambda \neq 0$, the resultant of $B(x, y)$, $C(x, y)$ is

$$r(x) = \det \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ x & 1 & 0 & 2x & 1 & 1 \\ l & x & 1 & 2\lambda & 2x & 1 \\ 0 & l & x & 0 & 2\lambda & 2x \\ 0 & 0 & l & 0 & 0 & 2\lambda \end{pmatrix} = -\lambda^2\left(\lambda + 1 + x\right).$$

So the resultant now vanishes just when $x = -(\lambda + 1)$, which is precisely when $B(x, y)$, $C(x, y)$ have a common factor of $y - 1$. With a small value of $\lambda \neq 0$, the picture changes very slightly: we change x, y to $x + \lambda y, y$, a linear transformation which leaves the x-axis alone, but tilts the y-axis to the left. Crucially, we tilt the asymptotic line at which the two curves approached one another (where $b(x, y)$ and $c(x, y)$ dropped degrees), but with essentially no effect on the intersection point:
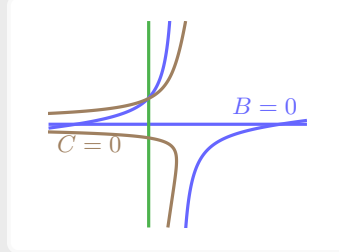
If $b(x,y) = x^2 + y^2 - 1$ and $c(x,y) = (x-1)^2 + y^2 - 1$, then $r(x) = (2x-1)^2$ vanishes at $x = 1/2$, where there are *two* different intersection points.
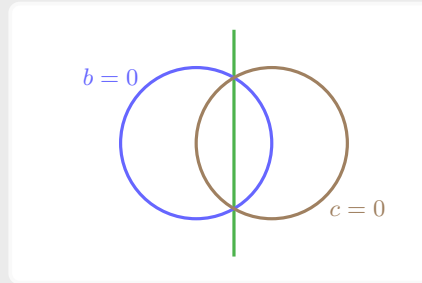
If instead we pick any nonzero constant λ and let $B(x,y) := (x + \lambda y)^2 + y^2 - 1$ and $C(x,y) := (x + \lambda y - 1)^2 + y^2 - 1$ then the resultant

$$r(x) = 4\left(\lambda^2 + 1\right)\left(x - \frac{1 + \lambda\sqrt{3}}{2}\right)\left(x - \frac{1 - \lambda\sqrt{3}}{2}\right)$$

vanishes at two distinct values of x corresponding to the two distinct roots. In the picture, the two roots now lie on different vertical lines (different values of x).
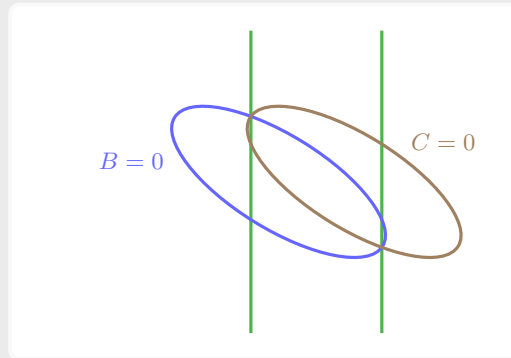
Proof. Write b as a sum of homogeneous polynomials of degrees $0, 1, 2, \ldots, d$, say

$$b = b_0 + b_1 + \cdots + b_d.$$

It is enough to prove the result for b_d, so assume that b is homogeneous of degree d. The expression

$$b(x, 1)$$

is a nonzero polynomial over an infinite field, so doesn't vanish everywhere by lemma 18.1 on page 151. Pick λ to be a value of x for which $b(\lambda, 1) \neq 0$. Then

$$b(x + \lambda y, y) = yb(x, 1) + \cdots + y^d b(\lambda, 1).$$

Similarly for two polynomials $b(x, y), c(x, y)$, or for any finite set of polynomials. $\square$

Corollary 18.5. *Take one variable y and several variables $x = (x_1, x_2, \ldots, x_n)$ and two polynomials $b(x, y)$ and $c(x, y)$ over a field. Let $r(x)$ be the resultant of $b(x, y), c(x, y)$ in the y variable. For any constant a in our field, if $b(a, y)$ and $c(a, y)$ have a common root in some algebraic extension of our field then $r(a) = 0$.*

If the coefficient of highest order in y of both $b(x, y)$ and $c(x, y)$ is constant in x (for example, perhaps after the normalization described in lemma 18.4 on page 155) then $r(a) = 0$ at some value $x = a$ in our field just exactly when $b(a, y)$ and $c(a, y)$ have a common factor. Moreover $r(x)$ is then the zero polynomial just when $b(x, y)$ and $c(x, y)$ have a common factor which is a polynomial in x, y of positive degree in y.

Proof. If, at some value $x = a$, both the degrees of $b(x, y)$ and $c(x, y)$ don't drop, then the resultant in y is expressed by the same expression whether we set $x = a$ to a constant value or leave x as an abstract variable, compute resultant, and then set $x = a$.

Work over the ring of polynomials in y, with coefficients rational in x. The resultant in y being zero as a function of x forces a common factor in that ring, i.e.

$$b(x, y) = d(x, y)B(x, y),$$
$$c(x, y) = d(x, y)C(x, y),$$

where $d(x, y), B(x, y)$ and $C(x, y)$ are rational in x and polynomial in y and $d(x, y)$ has positive degree in y. In particular, $c(x, y)$ factorises over that ring. By the Gauss lemma (proposition 10.6 on page 79), $c(x, y)$ factorises over the polynomials in x, y. But $c(x, y)$ is irreducible, so one factor is constant, and it isn't $d(x, y)$, so it must be $C(x, y)$, so we rescale by a nonzero constant to get $d(x, y) = c(x, y)$, i.e. $c(x, y)$ divides $b(x, y)$. $\square$

Corollary 18.6. *Given a finite collection of polynomial functions over a field k, for $x = (x_1, x_2, \ldots, x_n)$, either*

a. *in some finite extension of k, there is at least one point on which all polynomials in the collection vanish or*

b. *in some finite extension of k, every polynomial lies in the ideal generated by this collection.*

Proof. Suppose we just have two polynomials $p_1(x)$ and $p_2(x)$ in our collection. As above, after perhaps a finite extension, and a linear change of variables, we compute a resultant to find the values of one fewer of the variables on which there are simultaneous zeroes. The result follows by induction. Suppose instead that there are finitely many polynomials in our collection. We repeatedly replace pairs by such resultants, to eventually reduce the number of variables, and apply induction on the variables. In the end, when we have only one variable: all resultants are constants, and if they are not all zero, then there are no simultaneous solutions. Suppose that there is a nonzero resultant in among these. Rescale to get its value to be 1. That nonzero resultant is expressed as a linear combination as in lemma 18.3 on page 153. By induction we construct polynomials $q_j(x)$ so that

$$1 = q_1(x)p_1(x) + q_2p_2(x) + \cdots + q_s(x)p_s(x).$$

So 1 lies in the ideal generated by these $p_1(x), p_2(x), \ldots, p_s(x)$. But then any polynomial $f(x)$ has the form $f(x) = f(x) \cdot 1$, so also lies in that ideal. $\qquad\square$

18.1 In corollary 18.6, prove that either one case or the other occurs, but not both.

Quotient rings

Recall once again that all of our rings are assumed to be commutative rings with identity. Suppose that R is a ring and that I is an ideal. For any element $r \in R$, the *translate* of I by r, denoted $r + I$, is the set of all elements $r + i$ for any $i \in I$.

> If $R = \mathbb{Z}$ and $I = 12\mathbb{Z}$ is the multiples of 12, then $7 + I$ is the set of integers which are 7 larger than a multiple of 12, i.e. the set of all numbers
>
> $$\ldots, 7 - 12, 7, 7 + 12, 7 + 2 \cdot 12, 7 + 3 \cdot 12, \ldots$$
>
> which simplifies to
> $$\ldots, -5, 7, 19, 31, 53, \ldots$$
> But this is the same translate as $-5 + I$, since $-15 + 12 = 7$ so $-5 + I$ is the set of all numbers
>
> $$\ldots, -5 - 12, -5, -5 + 12, -5 + 2 \cdot 12, -5 + 3 \cdot 12, \ldots$$
>
> which is just
> $$\ldots, -15, -5, 7, 19, 31, 53, \ldots$$
> the same sequence.

> Working again inside $\mathbb{Z}$, $2 + 2\mathbb{Z} = 2\mathbb{Z}$ is the set of even integers.

> If $R = \mathbb{Q}[x]$ and $I = (x)$, then $\frac{1}{2} + I$ is the set of all polynomials of the form
>
> $$\frac{1}{2} + xp(x)$$
>
> for any polynomial $p(x)$.

> If i is in I then $i + I = I$.

We add two translates by

$$(r_1 + I) + (r_2 + I) := (r_1 + r_2) + I,$$

and multiply by

$$(r1 + I)(r_2 + I) = (r_1 r_2) + I.$$

Lemma 19.1. *For any ideal I in any ring R, two translates $a + I$ and $b + I$ are equal just when a and b differ by an element of I.*

Proof. We suppose that $a + I = b + I$. Clearly a belongs to $a + I$, because 0 belongs to I. Therefore a belongs to $a + I = b + I$, i.e. $a = b + i$ for some i from I.   □

Lemma 19.2. *The addition and multiplication operations on translates are well defined, i.e. if we find two different ways to write a translate as $r + I$, the results of adding and multiplying translates don't depend on which choice we make of how to write them.*

Proof. You write your translates as $a_1 + I$ and $a_2 + I$, and I write mine as $b_1 + I$ and $sb_2 + I$, but we suppose that they are the same subsets of the same ring: $a_1 + I = b_1 + I$ and $a_2 + I = b_2 + I$. By lemma 19.1 on the previous page, $a_1 = b_1 + i_i$ and $a_2 = b_2 + i_2$ for some elements i_1, i_2 of I. So then

$$
\begin{aligned}
a_1 + a_2 + I &= b_1 + i_1 + b_2 + i_2 + I, \\
&= b_1 + b_2 + (i_1 + i_2) + I, \\
&= b_1 + b_2 + I.
\end{aligned}
$$

The same for multiplication.   □

If I is an ideal in a commutative ring R, the *quotient ring* R/I is the set of all translates of I, with addition and multiplication of translates as above. The reader can easily prove:

Lemma 19.3. *For any ideal I in any commutative ring R, R/I is a commutative ring. If R has an identity element then R/I has an identity element.*

> If $R := \mathbb{Z}$ and $I := m\mathbb{Z}$ for some integer m then $R/I = \mathbb{Z}/m\mathbb{Z}$ is the usual ring of remainders modulo m.

> If $p(x, y)$ is an irreducible nonconstant polynomial over a field k and $I = (p(x, y)) = p(x, y)k[x]$ is the associated ideal, then $k[x, y]/I = k[X]$ is the ring of regular functions on the algebraic plane curve $X = (p(x, y) = 0)$.

Lemma 19.4. *A ideal I in a ring R with identity is all of R just when 1 is in I.*

Proof. If 1 lies in I, then any r in R is $r = r1$ so lies in I, so $I = R$.   □

An ideal I in a ring R is *prime* if I is not all of R and, for any two elements a, b of R, if ab is in I then either a or b lies in I.

> The ideal $p\mathbb{Z}$ in $\mathbb{Z}$ generated by any prime number is prime.

> The ring R is *not* prime in R, because the definition explicitly excludes it.

Inside $R := \mathbb{Z}[x]$, the ideal (x) is prime, because in order to have a factor of x sitting in a product, it must lie in one of the factors.

Inside $R := \mathbb{Z}[x]$, the ideal $(2x)$ is *not* prime, because $a = 2$ and $b = x$ multiply to a multiple of $2x$, but neither factor is a multiple of $2x$.

If $p(x)$ is an irreducible polynomial, then $(p(x))$ is a prime ideal.

The regular functions on an algebraic curve $p(x, y) = 0$ over a field k constitute the ring R/I where $R = k[x, y]$ and $I = (p(x, y))$. The curve is irreducible just when the ideal is prime.

More generally, intuitively, if we think of any ring R as being something like the polynomial functions in some variables, then an ideal is like the equations of some geometric object (so, roughly, ideals are like geometric objects), prime ideals are like irreducible geometric objects, and quotient rings are like the polynomial functions on those geometric objects.

Take a field k. In the quotient ring $R = k[x, y, z]/(z^2 - xy)$, the ideal (z) contains $z^2 = xy$, but does not contain x or y, so is not prime.

Lemma 19.5. *An ideal I in a commutative ring with identity R is prime just when R/I has no zero divisors, i.e. no nonzero elements α, β can have $\alpha\beta = 0$.*

Proof. Write α and β as translates $\alpha = a + I$ and $\beta = b + I$. Then $\alpha\beta = 0$ just when $ab + I = I$, i.e. just when ab lies in I. On the other hand, $\alpha = 0$ just when $a + I = I$, just when a is in I. $\qquad\square$

A ideal I in a ring R is *maximal* if I is not all of R but any ideal J which contains I is either equal to I or equal to R (nothing fits in between).

The ideal of regular functions vanishing at a point of an algebraic curve is maximal. Intuitively, we picture any ring as the ring of "regular functions" on some "geometric object". The maximal ideals we picture as the "points" of that object, while the prime ideals are the "geometric subobjects", like curves in a surface, and so on.

The maximal ideals in the integers are precisely the ideal generated by prime numbers, so the same as the prime ideals. In this sense, the integers provide a poor example, as we don't see the difference between prime and maximal.

Lemma 19.6. *If I is an ideal in a commutative ring R with identity, then I is a maximal ideal just when R/I is a field.*

Proof. Suppose that I is a maximal ideal. Take any element $\alpha \neq 0$ of R/I and write it as a translate $\alpha = a + I$. Since $\alpha \neq 0$, a is not in I. We know that $(a) + I = R$, since I is maximal. But then 1 lies in $(a) + I$, so $1 = ab + i$ for some b in R and i in I. Expand out to see that

$$\frac{1}{\alpha} = b + I.$$

So nonzero elements of R/I have reciprocals.

Suppose instead that R/I is a field. Take an element a in R not in I. Then $a + I$ has a reciprocal, say $b + I$. But then $ab + I = 1 + I$, so that $ab = 1 + i$ for some i in I. So $(a) + I$ contains $(ab) + I = (1) = R$.   □

> If p is a prime number, then $I := p\mathbb{Z}$ is maximal in $R := \mathbb{Z}$, because any other ideal J containing I and containing some other integer, say n, not a multiple of p, must contain the greatest common divisor of n and p, which is 1 since p is prime. We recover the fact that $\mathbb{Z}/p\mathbb{Z}$ is a field.

> The ideals in $R := \mathbb{Z}/4\mathbb{Z}$ are $(0), (1) = \mathbb{Z}/4\mathbb{Z}$ and $(2) \cong \mathbb{Z}/2\mathbb{Z}$, and the last of these is maximal.

> Take a field k and let $R := k[x]$. Every ideal I in R has the form $I = (p(x))$, because each ideal is generated by the greatest common divisor of its elements. We have seen that R/I is prime just when $p(x)$ is irreducible. To further have R/I a field, we need $I = (p(x))$ to be maximal. Take some polynomial $q(x)$ not belonging to I, so not divisible by $p(x)$. Since $p(x)$ is irreducible, $p(x)$ has no factors, so $\gcd\{p(x), q(x)\} = 1$ in $k[x]$, so that the ideal generated by $p(x), q(x)$ is all of R. Therefore $I = (p(x))$ is maximal, and the quotient $R/I = k[x]/(p(x))$ is a field.

> If k is an algebraically closed field and $R := k[x]$ then every irreducible polynomial is linear, so every maximal ideal in R is $I = (x - c)$ for some constant c.

19.1 For any positive integer $m \geq 2$, what are the prime ideals in $\mathbb{Z}/m\mathbb{Z}$, and what are the maximal ideals?

Sage

We take the ring $R = \mathbb{Q}[x, y]$ and then quotient out by the ideal $I = (x^2 y, xy^3)$ to produce the ring $S = R/I$. The generators x, y in R give rise to elements $x + I, y + I$ in S, which are renamed to a, b for notational convenience.

```
R.<x,y> = PolynomialRing(QQ)
I = R.ideal([x^2*y,x*y^3])
S.<a,b> = R.quotient_ring(I)
(a+b)^4
```

which yields `a^4 + b^4`.

We can define $R = \mathbb{Q}[x, y]$, $I = (y^2)$, $S = R/I$ where we label $x + I, y + I$ as a, b, $T = \mathbb{Q}[z]$ and define a morphism $\phi\colon S \to T$ by $\phi(a) = z^3, \phi(b) = 0$:

```
R.<x,y> = PolynomialRing(QQ)
I = R.ideal([y^2])
S.<a,b> = R.quotient_ring(I)
T.<z> = PolynomialRing(QQ)
phi = S.hom([z^3, 0],T)
phi(a^2+b^2)
```

yielding z^6.

Automorphisms of splitting fields

Theorem 19.7. *Suppose that k is a field and $p(x)$ is a nonconstant polynomial over k, say of degree n. Any two splitting fields for $p(x)$ over k are isomorphic by an isomorphism which is the identity on k.*

Proof. We already know that splitting fields exist, by adding roots to irreducible factors of $p(x)$. Each time we add a root, the dimension of the extension field over k is the degree of the irreducible factor. We then split off at least one linear factor in the extension field, and repeat. So at most $n!$ degree in total.

Given any splitting field K for an irreducible polynomial $p(x)$, pick any root $\alpha \in K$ of $p(x)$ and map

$$f(x) \in k[x] \mapsto f(\alpha) \in K,$$

to see that $K \cong k[x]/I$. Hence K is uniquely determined up to isomorphism.

Suppose instead that $p(x)$ is reducible, and that K, L are splitting fields of $p(x)$. Pick an irreducible factor $q(x)$ of $p(x)$. Then the subfields of K, L generated by adding a root of $q(x)$ to k are isomorphic, since $q(x)$ is irreducible, so we can assume that these subfields are the same field. We need only split $p(x)$ over that field, or quotient out all copies of $q(x)$ from $p(x)$ and split the result. Apply induction on the degree of $p(x)$. $\square$

> Adding a root to $x^2 + 1$ over $\mathbb{R}$ yields the splitting field $\mathbb{C}$.

Theorem 19.8. *Suppose that K is a splitting field of an irreducible polynomial $p(x)$ over a field k. If $\alpha, \beta \in K$ are roots of $p(x)$ then the Galois group of K over k has an element which takes α to β. In particular, K is a Galois extension of k.*

Proof. As in the proof of theorem 19.7, there is a field isomorphism $k[x]/I \to K$ taking $x \mapsto \alpha$, and another field isomorphism $k[x]/I \to K$ taking $x \mapsto \beta$. $\square$

Transcendental numbers

A number $\alpha \in \mathbb{C}$ is *algebraic* if it is the solution of a polynomial equation $p(\alpha) = 0$ where $p(x)$ is a nonzero polynomial with rational coefficients. A number which is not algebraic is called *transcendental*. More generally, given a field extension K of a field k, an element $\alpha \in K$ is *algebraic* over k if it is a root of a nonzero polynomial $p(x)$ with coefficients in k.

Theorem 19.9. *For any field extension K of a field k, an element $\alpha \in K$ is transcendental if and only if the field*

$$k(\alpha) = \left\{ \left. \frac{p(\alpha)}{q(\alpha)} \; \right| \; \frac{p(x)}{q(x)} \in k(x) \text{ and } q(\alpha) \neq 0 \right\}$$

is isomorphic to $k(x)$.

Proof. If α is transcendental, then the map

$$f \colon \frac{p(x)}{q(x)} \in k(x) \mapsto \frac{p(\alpha)}{q(\alpha)} \in k(\alpha)$$

is clearly well defined, onto, and preserves all arithmetic operations. To show that f is 1-1 is the same as showing that f has trivial kernel. Suppose that $p(x)/q(x)$ lies in the kernel of f. Then $p(\alpha) = 0$, so $p(x) = 0$, so $p(x)/q(x) = 0$. So the kernel is trivial, and so f is a bijection preserving all arithmetic operations, so f is an isomorphism of fields.

On the other hand, take a element α in K and suppose that there is some isomorphism of fields

$$g \colon k(x) \to k(\alpha).$$

Let $\beta := g(x)$. Because g is a field isomorphism, all arithmetic operations carried out on x must then be matched up with arithmetic operations carried out on β, so

$$g\left(\frac{p(x)}{q(x)} \right) = \frac{p(\beta)}{q(\beta)}.$$

Because g is an isomorphism, some element must map to α, say

$$g\left(\frac{p_0(x)}{q_0(x)} \right) = \alpha.$$

So

$$\frac{p_0(\beta)}{q_0(\beta)} = \alpha.$$

So $k(\beta) = k(\alpha)$. Any algebraic relation on α clearly gives one on β and vice versa. Therefore α is algebraic if and only if β is. Suppose that β is algebraic. Then $q(\beta) = 0$ for some polynomial $q(x)$, and then g is not defined on $1/q(x)$, a contradiction. $\qquad\square$

Finite fields

We are going to make a complete list of all finite fields. Pick a prime number p and work over the field $k = \mathbb{Z}/p\mathbb{Z}$. For any integer $n > 0$, the polynomial $x^{p^n} - x$ has two obvious linear factors: x and $x - 1$, and then factors as

$$x^{p^n} - x = x(x-1)(x^{-2+p^n} + x^{-3+p^n} + \cdots + x + 1).$$

Lemma 19.10. *For any prime p, over any extension field K of the field $k = \mathbb{Z}/p\mathbb{Z}$, the polynomial*

$$x^{p^n} - x$$

has no multiple roots.

Proof. Let $c(x) := x^{p^n} - x$. As above, write $x(x-1)b(x) = c(x)$. Note that $b(0) = 1 \neq 0$ and $b(1) = p^n - 1 = -1 \neq 0$. If, in some field K, we can factor out a multiple root from $b(x)$, that multiple root is not at $x = 0$ or $x = 1$, so we can factor out the same multiple root from $c(x) = x(x-1)b(x)$. Take the derivative of $c(x)$:

$$c'(x) = \frac{d}{dx}\left(x^{p^n} - x\right) = p^n x^{-1+p^n} - 1 = -1,$$

since $p = 0$ in k. But then $c'(x)$ has no roots in any field extension K of k, so $c(x)$ has no multiple roots over K, so $b(x)$ has no multiple roots over K. $\qquad\square$

Lemma 19.11. *For any finite field k of characteristic p, the* Frobenius morphism *$f(x) = x^p$ is an automorphism of fields $f\colon k \to k$, i.e. is a bijection preserving addition, subtraction, multiplication and division, taking $0 \mapsto 0$ and $1 \mapsto 1$.*

Proof. Clearly $0^p = 0$ and $1^p = 1$. For any $b, c \in k$, clearly $(bc)^p = b^p c^p$. The binomial theorem gives

$$(b+c)^p = b^p + p b^{p-1} c + \frac{p(p-1)}{2} b^{p-2} c^2 + \cdots + \frac{p!}{j!(p-j)!} b^{p-j} c^j + \cdots + c.$$

Every term except b and c has a factor of p in it, and $p = 0$ in k, so

$$(b+c)^p = b^p + c^p.$$

The same for $(b-c)^p = b^p - c^p$. If $f(b) = 0$, then $b^p = 0$ so $b \cdot b^{p-1} = 0$ so $b = 0$ or $b^{p-1} = 0$, and by induction, we find that $b = 0$. If $f(b) = f(c)$ then $f(b-c) = 0$ so $b - c = 0$ so $b = c$, i.e. f is 1-1. But then, since k is finite, f is a bijection. $\qquad\square$

Theorem 19.12. *For any prime p, the splitting field K of the polynomial*

$$c(x) = x^{p^n} - x$$

over the field $k = \mathbb{Z}/p\mathbb{Z}$ has p^n elements, every one of which is a root of $c(x)$. Every finite field K is obtained uniquely as the splitting field of $c(x)$ for some prime number p and integer $n \geq 1$, and so is uniquely determined up to isomorphism by its number of elements p^n.

Proof. There are p^n roots of $c(x) = x^{p^n} - x$ over its splitting field K, since $b(x)$ splits into linear factors, and, by lemma 19.10 on the previous page each linear factor gives a distinct root. Given any roots α, β of $c(x)$,

$$\alpha^{p^n} = \alpha, \ \beta^{p^n} = \beta.$$

Apply the Frobenius morphism:

$$(\alpha + \beta)^{p^n} = \alpha^{p^n} + \beta^{p^n} = \alpha + \beta,$$

so $c(\alpha + \beta) = 0$. Similarly

$$(\alpha\beta)^{p^n} = \alpha^{p^n}\beta^{p^n} = \alpha\beta,$$

so $c(\alpha\beta) = 0$. Hence the roots of $c(x)$ form a field, over which $c(x)$ splits, and so $b(x)$ splits. So every element of K is a root of $c(x)$, and so K is a splitting field of $c(x)$ with p^n elements. By the existence and uniqueness of splitting fields in theorem 19.7 on page 163, K is the unique splitting field of $c(x)$.

Take any finite field L, say of characteristic p. We map $0, 1, 2, \ldots, p-1$ in $k = \mathbb{Z}/p\mathbb{Z}$ to $0, 1, 2, \ldots, p-1$ in L, a morphism of fields. Hence L is a vector space over k, of finite dimension, say n, and so has p^n elements. The Frobenius morphism is an invertible k-linear map of L, generating a subgroup of the group of invertible k-linear maps $L \to L$. This subgroup is finite, since there are only finitely many elements of L, so finitely many permutations of those elements. This subgroup is generated by a single element, so is cyclic. By the classification of cyclic groups, this subgroup is isomorphic to $\mathbb{Z}/\ell\mathbb{Z}$, for some integer $\ell > 0$, so every element of L satisfies some equation $x^{p^\ell} = x$. $\qquad\square$

For any finite field k, say with p^n elements, we can associate to any $\alpha, \beta \in k$ the quadratic equation $(x - \alpha)(x - \beta) = 0$ which has roots α, β. On the other hand, given a quadratic equation $x^2 + bx + c = 0$, we can ask whether it has roots. Those with roots are given as $x^2 + bx + c = (x - \alpha)(x - \beta)$ for some α, β, unique up to swapping the order in which we write them down. As there are p^n elements, there are

$$\frac{p^n(p^n + 1)}{2}$$

pairs α, β of possible roots, up to swapping order. But there are p^{2n} quadratic equations $x^2 + bx + c$: choose any b, c from k. So

$$p^{2n} - \frac{p^n(p^n + 1)}{2} = \frac{p^n(p^n - 1)}{2}$$

quadratic equations over k have no solution in k.

In particular, let $k_1 \subset k_2 \subset k_3 \ldots$ be finite fields or orders $2, 4, 8$ and so on. Each k_{n+1} is the splitting field of the polynomial $z^2 + z + \alpha$ over k_n where α is any element of k_n not belonging to k_{n-1}. In particular, we can see which quadratic equations have solutions in which fields of characteristic 2. Note that we can't use the quadratic formula in fields of characteristic 2, because the quadratic formula divides by 2.

19.2 Prove by induction that if k is a finite field of characteristic 2, then every element α of k is a square, i.e. $\alpha = \beta^2$ for some β.

Sage

Lets get sage to construct the splitting field K of $p(x) := x^2 + x + 1$ over the field $k = \mathbb{Z}/2\mathbb{Z}$. Sage constructs splitting fields by first building a ring, which we will call R, of polynomials over a field. In our case, we build $R = k[x]$. Then we let K be the field generated by an element a, which corresponds to the variable x in the polynomial ring R.

```
R.<x> = PolynomialRing(GF(2))
K.<a> = (x^2 + x + 1).splitting_field()
```

We can then make a list of some elements of K, call it S, and compute out addition and multiplication tables:

```
S=[0,1,a,a+1]
print("Addition table:")
for i in range(0,4):
    for j in range(0,4):
        print("({})+({})={}".format(S[i],S[j],S[i]+S[j]))
print("Multiplication table:")
for i in range(0,4):
    for j in range(0,4):
        print("({})*({})={}".format(S[i],S[j],S[i]*S[j]))
```

Chapter 20

Field extensions and algebraic curves

We can sometimes draw pictures of field extensions. Take an irreducible algebraic curve $C = (p(x, y) = 0)$ in the plane, and look at its field of rational functions $k(C)$, our favourite example of a field extension of a field k.
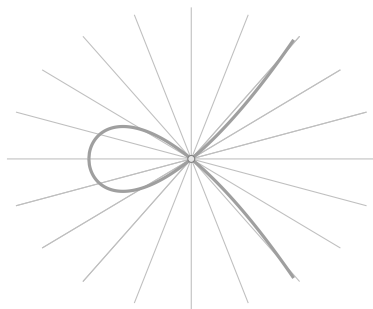
Suppose that K is a field extension of $k(C)$, say $K = k(C)(\alpha)$ given by extending by an element α. We want to draw a curve D so that $K = k(D)$.

If α is transcendental, then $k(C)(\alpha) \cong k(C)(z)$ for some abstract variable z. But $k(C)$ is already a field of rational functions of two variables x, y, quotiented out to enforce the relation $p(x, y) = 0$. So K is just the rational functions on the "cylinder" $C \times k$ inside 3-dimensional space, with variables x, y, z; our cylinder has the equation $p(x, y) = 0$, independent of z.
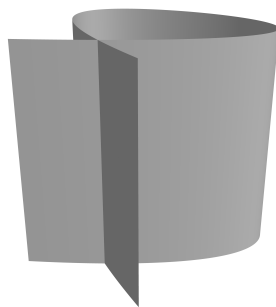
Suppose instead that α is algebraic, say satisfying a polynomial equation $f(z) = 0$ with $f(z)$ a polynomial with coefficients from $k(C)$. So each coefficient of $f(z)$ is itself a rational function on C. Each such function is expressible as a rational function of x, y. Clearing denominators, we get a polynomial equation $f(x, y, z) = 0$ with coefficients in the underlying field k. The solutions form a surface in 3-dimensional space. But the resulting surface depends on how we pick out rational functions of x, y to represent our coefficients of $f(z)$ from $k(C)$. So in fact we have to restrict

our x, y variables to lie in C, i.e. we consider the *curve D* cut out by the equations $p(x, y) = f(x, y, z) = 0$.
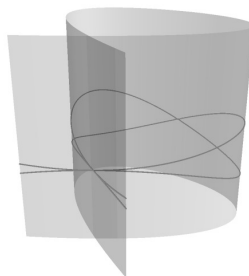
The map $(x, y, z) \in D \mapsto (x, y) \in C$ is regular.

We wonder now whether there is a $k(C)$-root to $f(z) = 0$, say β. This β is in $k(C)$, so a rational function on C, so that $f(\beta) = 0$. So this is a rational map $z = \beta(x, y)$ from C to D, lifting the curve C up from the plane into space, tracing out D on the cylinder. In particular, there is no such map if D is irreducible and $D \to C$ is many to one:
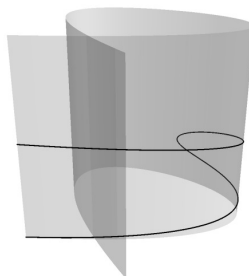
So in such a case, $k(C) \subset k(D)$ is a nontrivial field extension. We get a picture giving some intuition for nontrivial algebraic field extensions: typically we expect to see an irreducible curve D mapping to a curve C by a many-to-one map.

The story continues similarly in higher dimensions: a transcendental extension increases dimension, while an algebraic extension, say of degree n, preserves dimension, giving a new geometric object which maps regularly to the old one, by an n-to-1 map.

Rational curves

Recall that the *degree* of a field extension $k \subset K$, denoted $[K : k]$, is the smallest integer n so that there are n elements $\alpha_1, \alpha_2, \ldots, \alpha_n$ of K so that every element of K is a linear combination $\sum a_j \alpha_j$ with coefficients $a_1, a_2, \ldots, a_n$ in k. The degree of the field extension is infinite if there is no such integer n.

20.1 Suppose that $k \subset K \subset L$ are field extensions. Prove that $[L : k] = [L : K][K : k]$. Note: this works even if one or more degree is infinite: let $n\infty = \infty n = \infty\infty = \infty$ for any positive integer n. You have to prove these infinite degree cases as well.

20.2 Prove that a field extension is an isomorphism if and only if it has degree one.

20.3 Suppose that $p(x)$ is an irreducible polynomial over a field k. Let $K :=$ $k[x]/(p(x))$. Prove that $[K : k]$ is the degree of $p(x)$.

20.4 Prove that if $u(t)$ is a nonconstant rational function over a field k, then there are infinitely many values of $u(t)$ as t varies over $\bar{k}$. In particular, $k \subset k(u(t))$ is a transcendental extension.

Recall that a curve $C = (0 = f(x, y))$ is *irreducible* if $f(x, y)$ is irreducible.

Theorem 20.1. *An irreducible plane algebraic curve C is rational just when there is a nonconstant rational map $L \to C$. In other words, C is rational just when there are rational functions $x(t), y(t)$ for which the point $(x(t), y(t))$ lies in C for all t in the algebraic closure of the field.*

Proof. If our curve C is rational then by definition C is birational to a line, so there is such a map. Suppose on the other hand that there is such a map $L \to C$, say written as $t \in L \mapsto (x(t), y(t)) \in C$ with

$$f(x(t), y(t)) = 0$$

for some rational functions $x(t), y(t)$, not both constant. As we vary t through $\bar{k}$, the moving point $(x(t), y(t))$ moves through infinitely many different points of the plane. Take a rational function on C, say $g(x, y) = b(x, y)/c(x, y)$. Then $c(x, y) \neq 0$ at all but finitely points of C. (Note that this requires that our curve C is irreducible, as otherwise we might have $g(x, y)$ vanishing on a component of C.) So at infinitely many values of t, $g(x(t), y(t))$ is defined. We map $g(x, y) \in k(X) \mapsto g(x(t), y(t)) \in k(t)$, a morphism of fields. The image is some subfield $K = k(x(t), y(t)) \subset k(t)$. The result follows from the following theorem applied to K. $\square$

Theorem 20.2 (Lüroth's theorem)**.** *Suppose that $k \subset K \subset k(t)$ is a finitely generated extension of a field k, where t is an abstract variable. Then $K = k(f(t))$ is generated by a single rational function $f(t)$ in $k(t)$. In particular, either $K = k$ (if $f(t)$ is a constant function) or K is isomorphic to $k(t)$ by $t \mapsto f(t)$.*

The proof of this theorem requires a lemma:

Lemma 20.3. *Take a nonconstant rational function $f(x) = b(x)/c(x)$, with $b(x), c(x)$ coprime, over a field k. Then, as a polynomial in a variable y with coefficients in $k(f(x))$, the expression*

$$f(x)c(y) - b(y)$$

in $k(f(x))[y]$ is irreducible.

Proof. This expression is not zero, because it is nonconstant in x. It vanishes at $y = x$. Hence x satisfies this polynomial (in y) expression with coefficients in $k(f(x))$. So $k(x)$ is an algebraic extension of $k(f(x))$. Since $k(x)$ is transcendental over k, $k(f(x))$ is transcendental over k. So $f(x)$ solves no polynomial equation over k. Take an abstract variable t. Map $k(t, y) \to k(f(x), y)$ by $t \mapsto f(x)$. This map is an isomorphism of fields. Similarly $k[t, y] \to k[f(x), y]$ is an isomorphism of rings. In order to prove that $tc(y) - b(y)$ is irreducible in $k(t)[y]$, it suffices to prove that it is irreducible in $k[t, y]$, i.e. clear t from denominators, by Gauss's lemma (proposition 10.6 on page 79). If we can factor $tc(y) - b(y) = g(t, y)h(t, y)$ in $k[t, y]$ then one of $g(t, y)$ or $h(t, y)$ has degree 1 in t, since the product does, and so one of $g(t, y)$ or $h(t, y)$ has degree zero

in t, depending only on y, say $g(t, y) = g(y)$. Then $g(t)h(t, y) = tc(y) - b(y)$, so $g(y)$ divides $tc(y) - b(y)$, and since t is an abstract variable, $g(y)$ divides $b(y)$ and $c(y)$, which are coprime, so $g(y)$ is a nonzero constant. $\qquad\qquad\qquad\qquad\qquad\square$

We now prove Lüroth's theorem:

Proof. For each element $f(t) = b(t)/c(t)$ in K, let n be the larger of the degrees of $b(t), c(t)$. Pick an element $f(t) = b(t)/c(t)$ in K, not in k, for which the value of n is as small as possible. By the previous lemma, the polynomial $B(y) = f(x)c(y) - b(y)$ is irreducible in $k(f(x))[y]$, so the map

$$x, y \in k(f(x))[y]/(B(y)) \mapsto x, x \in k(x)$$

is an isomorphism of fields and $[k(x) : k(f(x))] = n$. The expression $B(y)$ is the polynomial in $K[y]$ of smallest positive y-degree which is satisfied by $y = x$, since a smaller degree one would give a smaller value for n. Indeed the y-degree of $B(y)$ is n. In particular, $B(y)$ is also irreducible in $K[y]$ and the map

$$x, y \in K[y]/(B(y)) \mapsto x, x \in k(x)$$

is an isomorphism. In particular, $[k(x) : K] = n$. But

$$\begin{aligned} n &= [k(x) : k(f(x))], \\ &= [k(x) : K][K : k(f(x))], \\ &= n[K : k(f(x))] \end{aligned}$$

so $[K : k(f(x))] = 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Chapter 21

The projective plane

*The more systematic course in the present introductory memoir ... would
have been to ignore altogether the notions of distance and metrical ge-
ometry .... Metrical geometry is a part of descriptive geometry, and
descriptive geometry is all geometry.*

— Arthur Cayley

The old fashioned term *descriptive geometry* means the geometry of straight lines.
Straight lines in the plane remain straight when you rescale the plane, or translate
the plane to the left or right, up or down, or when you rotate the plane. They even
remain straight when you carry out any linear change of variables, as you know from
linear algebra.

To see a World in a Grain of Sand,
And Heaven in a Wild Flower.
Hold Infinity in the palm of your hand,
And Eternity in an hour.

— William Blake
AUGURIES OF INNOCENCE

Railway tracks built on a flat plane appear to meet "at infinity".
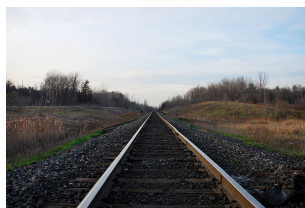
Imagine that we were to add a point to the plane "at infinity" in the direction where
these tracks (straight lines) appear to meet. *Danger:* in order to add only one point,
imagine that the two rails meet at the *same* point in one direction as they do in the
other, making them both into circles. The projective plane is the set of all of the
usual points of the plane (which we can think of as the "finite points" of the projective
plane) together with some sort of "points at infinity" to represent the directions in
the plane.

But you must note this: if God exists and if He really did create the world, then, as we all know, He created it according to the geometry of Euclid and the human mind with the conception of only three dimensions in space. Yet there have been and still are geometricians and philosophers, and even some of the most distinguished, who doubt whether the whole universe, or to speak more widely the whole of being, was only created in Euclid's geometry; they even dare to dream that two parallel lines, which according to Euclid can never meet on earth, may meet somewhere in infinity.

— Fyodor Dostoyevsky
The Brothers Karamazov

We would like a more precise mathematical definition of points "at infinity". Place yourself at a vantage point above the plane.
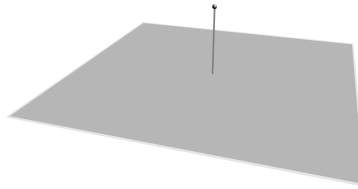
Every "finite" point of the plane lies on a line through your vantage point: the line of sight.
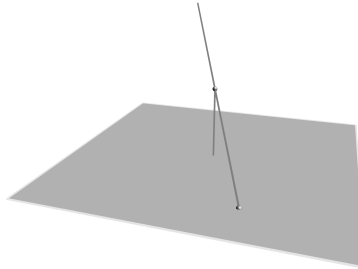
You can also look out to the points at infinity, along lines:
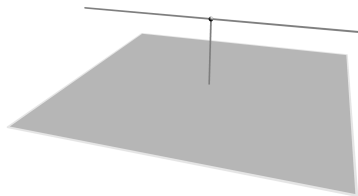
For example, there is such a line parallel to the lines of our train track:
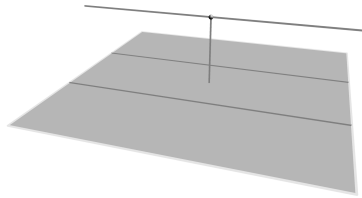
On the other hand, if we take any line through the vantage point, either it hits a "finite" point of the plane:
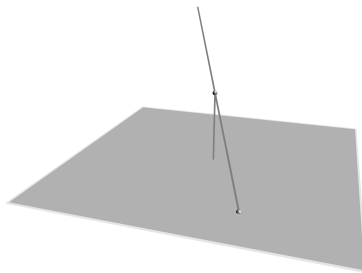
or it is parallel to the plane
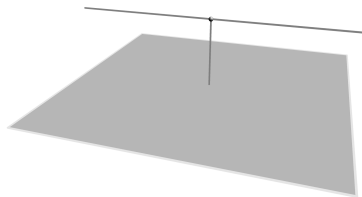
so that it is pointed along the direction of some train tracks (lines) in the plane, to some "point at infinity".
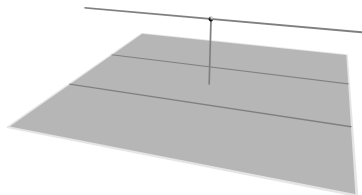
The points of the projective plane, finite points together with infinite points at infinity, are in 1-1 correspondence with lines through the vantage point.

This picture gives us a rigorous definition of points at infinity: the *projective plane* is the set of all lines through a chosen point of 3-dimensional space, called the vantage point. The *points* of the projective plane (using this definition) are the lines through

the vantage point. The "finite points" are the lines not parallel to the horizontal plane, while the "infinite points" are the lines which are parallel the horizontal plane. The set of finite points is the *affine plane*, which we think of as just the usual xy-plane. The *lines*, also called *projective lines*, of the projective plane are the planes through the vantage point.
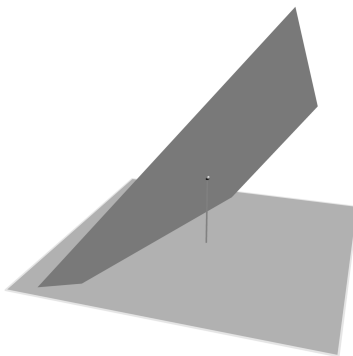
Homogeneous coordinates

We can make this more explicit by writing each point of 3-dimensional space $\mathbb{R}^3$ as a triple (x, y, z). Draw the plane not as the usual xy-plane, but as the plane $z = z_0$, for some nonzero constant z_0. Take the vantage point to be the origin. Any linear change of the 3 variables x, y, z takes lines (and planes) through the origin to one another: there are more symmetries of the projective plane than we have ever encountered in the finite plane.

Every point of 3-dimensional space not at the origin lies on a unique line through the origin. So every point (x, y, z) with not all of x, y, z zero lines on a unique line through the origin. The points of this line are precisely the rescalings (tx, ty, tz) of that point. So each point of the projective plane can be written as a triple (x, y, z), not all zero, and two such triples represent the same point just when each is a rescaling of the other. Denote such a point as $[x, y, z]$. For example, the point $[3, 1, 2]$ of the projective plane means the line through the origin consisting of the points of the form $(3t, t, 2t)$ for all values of a variable t, including $(3, 1, 2)$, $(3 \cdot 2, 1 \cdot 2, 2 \cdot 2)$, $(-3, -1, -2)$, and so on. We write this as:

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

The coordinates x, y, z are the *homogeneous coordinates* of a point $[x, y, z]$ of the projective plane.

In our pictures, the plane $z = z_0$ was below the vantage point, but it is more traditional to take the plane to be $z = 1$, above the vantage point. The *affine plane* is just the usual xy plane, but identified either with the plane $z = 1$, or with the subset of the projective plane given by points $[x, y, 1]$.

The roles of x, y, z are all the same now, and we can see that any linear change of variables of x, y, z takes the points of the projective plane to one another, and takes the lines of the projective plane to one another. Each linear change of variables is

represented by an invertible matrix

$$g = \begin{pmatrix} g_{00} & g_{01} & g_{02} \\ g_{10} & g_{11} & g_{12} \\ g_{20} & g_{21} & g_{22} \end{pmatrix}.$$

Conventionally we label our matrix columns using $0, 1, 2$ rather than $1, 2, 3$. Such a matrix takes a point $p = [x, y, z]$ of the plane to the point $q = [X, Y, Z]$ where

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = g \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

We denote this relation as $q = [g]p$ and call $[g]$ a *projective automorphism*. Clearly if g and h are 2 invertible 3×3 matrices, then $[gh] = [g][h]$, i.e. we carry out the transformation of lines through the origin by carrying out the multiplications by the matrices.

Lemma 21.1. *If a square matrix g rescales every nonzero vector by a scalar, i.e. $gx = \lambda(x)x$ for some number $\lambda(x)$, for every vector $x \neq 0$, then $\lambda(x)$ is a constant scalar independent of x.*

Proof. Suppose that g is $n \times n$. If $n = 1$ then every square matrix is a constant scalar, so suppose that $n \geq 2$. Write our vectors as $x = \sum_j x_j e_j$ in the standard basis of $\mathbb{R}^n$ and calculate

$$\begin{aligned} gx &= \lambda(x)\, x, \\ &= \sum_j \lambda(x)\, x_j e_j, \\ &= \sum_j x_j g e_j, \\ &= \sum_j x_j \lambda(e_j)\, e_j. \end{aligned}$$

So for any x with $x_j \neq 0$, we have $\lambda(x) = \lambda(e_j)$. But then if $x_i \neq 0$, and if $i \neq j$, then

$$\begin{aligned} \lambda(x) &= \lambda(e_i), \\ &= \lambda(e_i + e_j), \\ &= \lambda(e_j). \end{aligned}$$

So λ is a constant. $\square$

Lemma 21.2. *Two projective automorphisms $[g], [h]$ have precisely the same effect on all points $[x, y, z]$ of the projective plane just when the matrices agree up to a constant $h = \lambda g$, some number $\lambda \neq 0$.*

Proof. Let $\ell := h^{-1}g$, so that $[\ell] = [h]^{-1}[g]$ and $[\ell]$ fixes every point of the projective plane just when ℓ rescales every vector in $\mathbb{R}^3$ by a scalar. $\square$

In particular, the points of the projective plane are permuted by the projective automorphisms, as are the projective lines.

The *projective plane* over any field k, denoted $\mathbb{P}^2(k)$, is the set of lines through the origin in k^3. In k^3, the line through two points

$$p = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, q = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

is just the set of all points $tp + (1 - t)q$ for t in k.

Example: the Fano plane

Let $k := \mathbb{Z}/2\mathbb{Z}$; the *Fano plane* is the projective plane $\mathbb{P}^2(k)$. Each point is $[x, y, z]$ with x, y, z in k defined up to rescaling. But the only nonzero scalar you can use to rescale with is 1, since $k = \{0, 1\}$. Therefore we can just write $[x, y, z]$ as (x, y, z). Taking all possibilities for x, y, z from k, not all zero, we get 7 points

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

corresponding to the numbers $1, 2, \ldots, 7$ written in base 2.

We can also draw this diagram as a cube; for each point, you draw the corresponding point in $\mathbb{R}^3$ with the same coordinates:
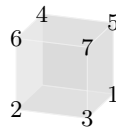
with one vertex not marked; the unmarked vertex is at the origin, so doesn't represent any point of the Fano plane. Every line in the Fano plane has three vertices. Take two numbers from 1 to 7, say $1, 5$, and write out their binary digits under one another:

$$\begin{matrix} 001 \\ 101 \end{matrix}.$$

Look at these as vectors in k^3, so add them without carrying digits:

$$\begin{matrix} 001 \\ 101 \\ \hline 100 \end{matrix}$$

The sum vector lies on the same line in the Fano plane, giving the lines:
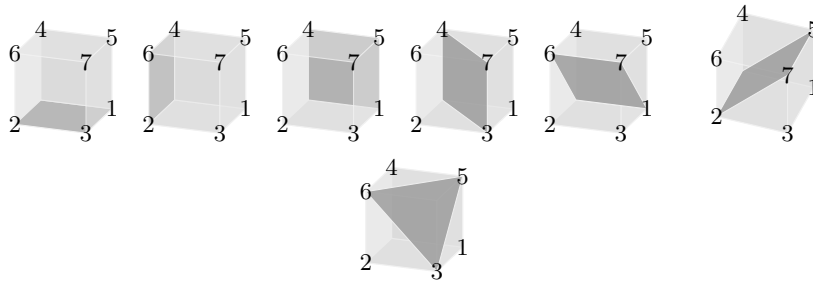
Projective space

Similarly, *projective space* $\mathbb{P}^n(k)$ is the set of lines through the origin of k^{n+1}, and a *projective automorphism* of projective space is the result of applying an invertible square matrix to those lines. Again, two invertible square matrices yield the same projective automorphism just when they agree up to a scalar multiple, by the same proof.

Algebraic curves in the projective plane

Homogenisation of plane curves

Take an algebraic curve $y^2 = x^2 + x^3$ in the affine plane over a field k.

The *cone* on the curve is the surface in k^3 given by $y^2 z = x^2 z + x^3$.
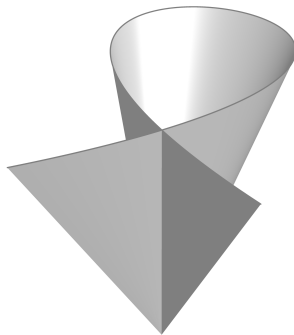
The recipe: given an algebraic equation, like $y^2 = x^2 + x^3$:

a. Find the degree n of the highest term; in this example $n = 3$.

b. Invent a new variable z.

c. Multiply each term by a power of z so that the total degree of each term reaches n.

A polynomial is *homogeneous* if all of its terms have the same total degree, i.e. sum of degrees in all variables.

22.1 Homogenise $x^2 y - x^2 + y = 1 + 2x - y^2 x^3$.

In sage:

```
R.<x,y,z> = PolynomialRing(QQ)
p = x^3+x*y+1
p.homogenize(z)
```

yields $x^3 + xyz + z^3$.

The homogenised polynomial equation $y^2 z = xz^2 + x^3$ has all terms of degree 3, so if we rescale x, y, z to tx, ty, tz, we rescale both sides of the equation by t^3, so solutions (x, y, z) rescaled remain solutions (tx, ty, tz). Thefore every solution lies on a line through the origin consisting of other solutions. The set of solutions is a surface, which is a union of lines through the origin. A *projective plane curve* is the set of lines through the origin satisfying a nonzero homogeneous polynomial.

Conics

Take a hyperbola $xy = 1$ in the affine plane,
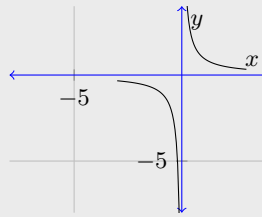
and homogenise to $xy = z^2$, as we see from three different perspectives:
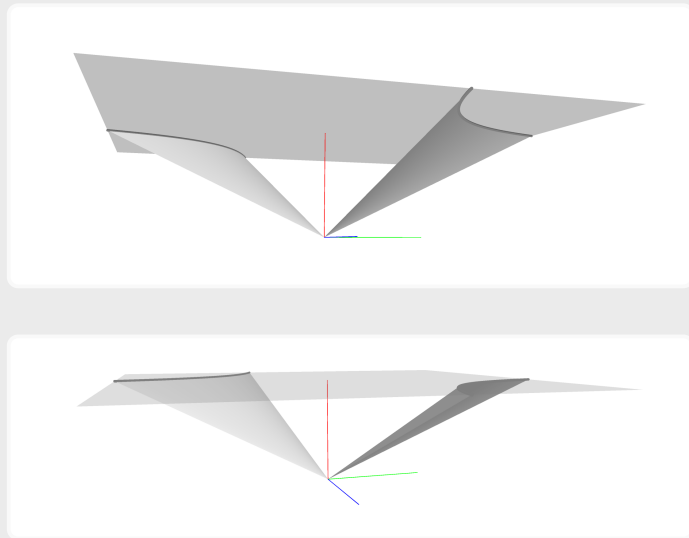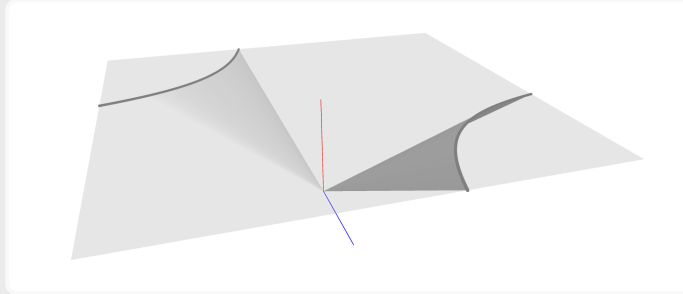
If we slice the surface along a plane like $z = 2$, parallel to the plane we drew but higher up, that plane $z = 2$ intersects our surface in a dilated hyperbola. But if we slice the surface along the plane $x = 1$, we get $xy = z^2$ to become $y = z^2$, a parabola:
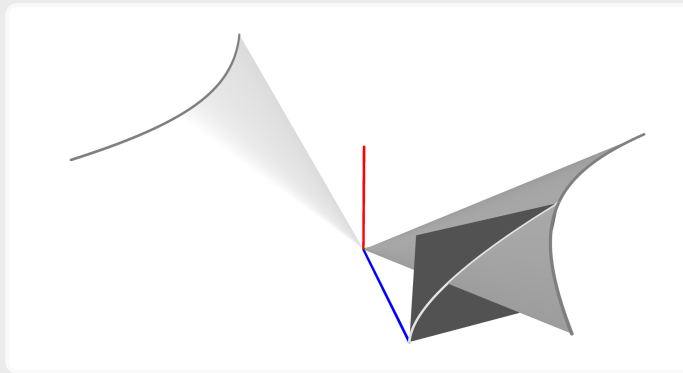
Suitable linear changes of variable interchange the x and z axes, so suitable projective automorphisms identify the two affine curves: the hyperbola and the parabola. In particular, the two curves have isomorphic fields of rational functions. They *don't* have isomorphic rings of regular functions.

If instead we slice our surface with the plane $x + y = 1$,
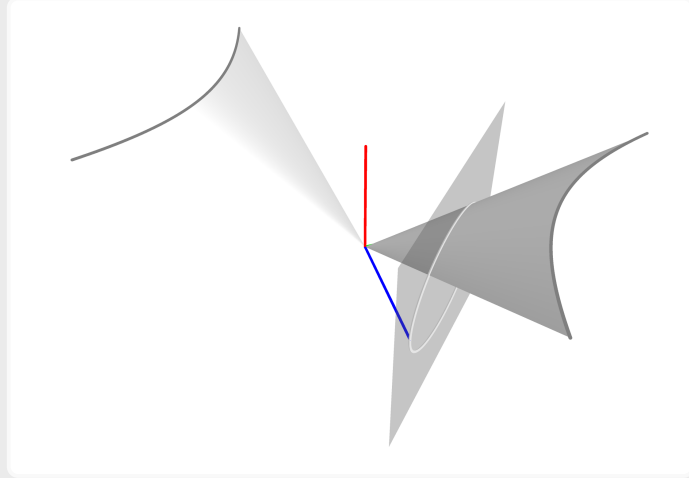
we get $y = 1 - x$ plugged in to $xy = z^2$, which simplifies to

$$\left(x - \frac{1}{2}\right)^2 + z^2 = \frac{1}{4},$$

a circle of radius $\frac{1}{2}$. So again the circle is birational to the parabola and to the hyperbola, and they all have the same rational function fields. In these pictures we can see why the conics are called *conics*: they are all slices of the same cone.

The affine curve $x^2 + y^2 = -1$ has no real points on it, but has complex points. It homogenises to $x^2 + y^2 + z^2 = 0$, a "sphere of zero radius", with the origin as its only real point, but with again many complex points. The complex linear change of variables $X = x - iy, Y = x + iy, Z = iz$ gives us $XY = Z^2$, the same projective curve as before. So these curves are all birational over the complex numbers, and indeed are all identified by automorphisms of the complex projective plane.

22.2 By the same trick of homogenising and dehomogenising in various variables, draw the same sort of pictures for the circle $x^2 + y^2 = 1$ to see what cone it homogenises to, and what curves it makes when you intersect that cone with the planes (a) $x = 1$, (b) $y = 1$ and (c) $z = y + 1$.

Counting intersections of curves

Who are you going to believe, me or your own eyes?

— Chico Marx

A line through a cubic curve
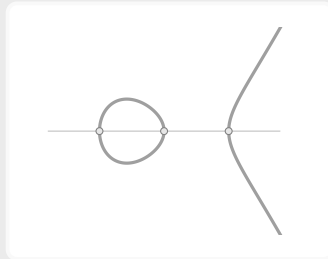
can hit as many as 3 points.

Two conics
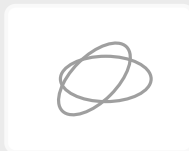
can intersect at as many as 4 points, but sometimes only at 3 points:

Some conics don't appear to intersect at all

but keep in mind that our points can have coordinates in the algebraic closure, so they actually intersect.

The conics $x^2 + y^2 = 1$ and $(x-4)^2 + y^2 = 1$:

intersect at $(x,y) = (2, \pm i\sqrt{15})$.

Some conics appear not to have even complex intersections, such as $x^2 + y^2 = 1$ and $x^2 + y^2 = 2$.
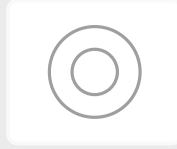
But in the projective plane, we can homogenise to $x^2 + y^2 = z^2$ and $x^2 + y^2 = 2z^2$, and then intersect with the plane $x = 1$, to get $1 + y^2 = z^2$ and $1 + y^2 = 2z^2$, which intersect at $(y,z) = (\pm i, 0)$.

22.3 Find all points of the curve $x^3 + x^2 y + xy^2 + y^3 + x^2 = 1$ which lie "at infinity", i.e. on the line $z = 0$ in homogeneous coordinates, over the field $k = \mathbb{C}$.

22.4 Find all points of the curve $y^2 + y = x^3 + x + 1$ on the projective plane over the field $k = \mathbb{Z}/5\mathbb{Z}$.

Lemma 22.1. *Suppose that k is an infinite field. Then for any finite set of lines in $\mathbb{P}^2(k)$, there are infinitely many points of $\mathbb{P}^2(k)$ not on any of those lines.*

Proof. If we work in the affine plane, the points (x,y) lying on those lines must satisfy one of the associated linear equations, say $y = m_i x + b_i$ or $x = c_i$. Pick a number x_0 not among the constants c_i. Now we have finitely many values $y = m_i x_0 + b_i$ to avoid; pick y to not be one of those. $\qquad\square$

Lemma 22.2. *Every nonzero homogeneous polynomial of degree n in two variables over a field k has n roots, counted with multiplicities, in $\mathbb{P}^1(\bar{k})$, and so splits into a product of homogeneous linear factors over $\bar{k}$.*

Proof. As part of the proof, we define multiplicities of zeroes of a homogeneous polynomial $p(x,y)$ to be multiplicities of linear factors. So we are saying precisely that $p(x,y)$ factors into n linear factors. If $p(x,y)$ has a linear factor, divide it out and apply induction. Since y is not a linear factor of $p(x,y)$, $p(x,1)$ has degree n, say $p(x,1) = a(x - x_1)\ldots(x - x_n)$. The homogeneous polynomial

$$r(x,y) := p(x,y) - a(x - x_1 y)(x - x_2 y)\ldots(x - x_n y)$$

vanishes at $y = 1$ for any x, so by homogeneity vanishes for any $y \neq 0$. Fixing any value of x, there are infinitely many values of y at which $0 = r(x, y)$, and so $r(x, y)$ vanishes everywhere. ☐

The polynomial
$$p(x, y) = x^3 + x^2 y - 2xy^2$$
becomes
$$p(x, 1) = x^3 + x^2 - 2x = x(x - 1)(x + 2),$$
so that
$$p(x, y) = x(x - y)(x + 2y).$$

The family of curves with $y^2 = ax$ contains the curve $y^2 = 0$ when $a = 0$, a line. So a family of curves of degree 2 contains a curve of degree 1. It is natural to think of the curve being a "double line" when $a = 0$.

Similarly, we can say that a *multiple component* of a curve means a curve given by an equation which is a positive integer power of an irreducible homogeneous polynomial, say $0 = f(x, y, z)^k$ giving a component with multiplicity k. From now on, we allow projective plane curves to have components with positive integer multiplicities.

Breaking curves into components

Theorem 22.3 (Study)**.** *Every algebraic curve in the projective plane has a unique decomposition into irreducible algebraic curves, its* components*. To be more precise, every nonzero homogeneous polynomial in 3 variables $f(x, y, z)$ has a decomposition*

$$f = a_0 g_1^{k_1} g_2^{k_2} \ldots g_n^{k_n},$$

into a product of a nonzero constant a_0 and several coprime irreducible homogeneous polynomials $g_j(x, y, z)$ to various positive integer powers, unique up to rescaling the constant and the polynomials by nonzero constants, and writing down the polynomials in perhaps a different order.

Proof. Suppose that g is an irreducible polynomial and that $f = 0$ at every point in $\bar{k}$ where $g = 0$. We want to prove that g divides f. Work in the affine plane $z = 1$. Think of each equation as a polynomial in one variable x, but with coefficients rational functions of y. By corollary 18.5 on page 157, after perhaps a linear change of variables, the resultant in y vanishes just at those x values where there are simultaneous roots. In particular, the number of different values of x where such roots occur is never more than the degree of the resultant, as a polynomial in x. For each value $x = c$ (over $\bar{k}$) for which $g(c, y)$ is not constant, the values of y at which $g(c, y) = 0$ also satisfy $f(c, y) = 0$, so the resultant vanishes. Therefore this resultant is the zero polynomial. If the curve where $f = 0$ is the union of distinct irreducible curves given by equations $g_i = 0$, then each of these g_i divides f, so their product divides f. Apply induction on the degree of f. ☐

Theorem 22.4. *Every rational morphism of projective plane algebraic curves* $f \colon B \to C$ *is constant or onto a finite union of components of* C, *at most one for each component of* B.

Proof. We can assume that B is irreducible, and we have already assumed that our field is algebraically closed by definition of the points of the curve. Our rational morphism is

$$f(x, y, z) = (p(x, y, z), q(x, y, z), r(x, y, z)).$$

If $B = (b(x, y, z) = 0)$ then the system of equations

$$b(a, b, c) = 0, a = a(x, y, z), b = b(x, y, z), c = c(x, y, z)$$

has (after perhaps a linear change of x, y, z variables) resultant eliminating x, y, z given by some algebraic equation in a, b, c which specifies which points lie in the image. By corollary 18.5 on page 157 this equation is not trivial. Split C into irreducibles $C_1, C_2, \ldots, C_N$, say $C_j = (c_j = 0)$. Then $c_1 c_2 \cdots c_N \circ f = 0$ on B, since f takes B to C. So if $B = (b(x, y, z) = 0)$ then b is divisible by $c_1 c_2 \cdots c_N \circ f$ in any affine chart. If $c_1 \circ f$ is not zero on B then divide by it to see that already $c_2 \cdots c_N \circ f = 0$ on B. So f takes B into $C_2 \cup \cdots \cup C_N$. We can assume that $c_j \circ f = 0$ on B for every j, i.e. that $B \subset f^{-1} C_j$ for every j, i.e. that

$$f(B) \subset \bigcap_j C_j :$$

there is only one such C_j, i.e. C is irreducible. Hence the image is precisely C. $\square$

Counting intersections

A triangle *pqr* is *generic* for two algebraic projective plane curves B, C if

    a. neither curve passes through q and

    b. no line connecting two intersection points of B and C lies through q.
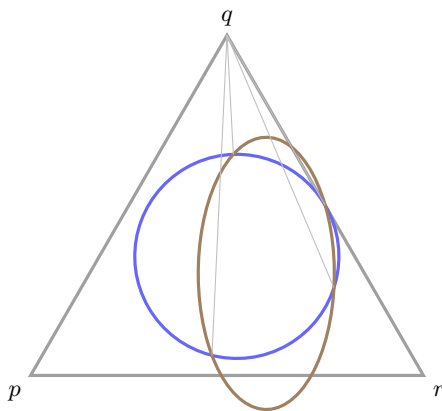
A triangle *pqr* is *very generic* for curves B and C if it is generic and there is no intersection point of B and C along the line qr. Keep in mind that we allow intersection points, and points of the triangle, to have coordinates in the algebraic closure of our field.
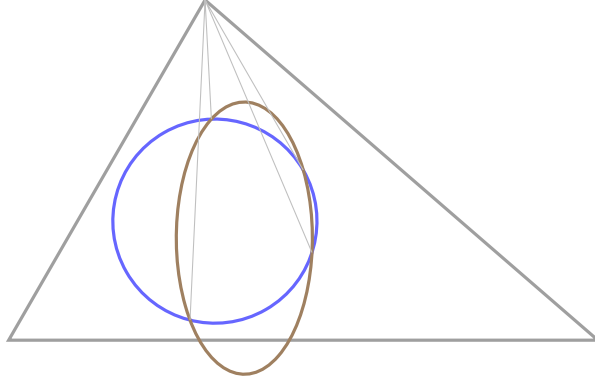
By projective automorphism, we can make any triangle become the one whose vertices are

$$[0, 0, 1], [1, 0, 0], [0, 1, 0].$$

This projective automorphism is unique up to rescaling the variables. In the affine plane, these vertices are the points $(0, 0), (\infty, 0), (0, \infty)$. Suppose that our two curves have equations $B = (b(x, y) = 0)$ and $C = (c(x, y) = 0)$. The triangle is generic just when

a. $(0, \infty) = [0 : 1 : 0]$ is not on B or C, and

b. no vertical line intersects two intersection points of B and C.

In a more algebraic description:

a. $b(x, y)$ and $c(x, y)$ both have nonzero constants in front of their highest terms in y and

b. for any constant $x = x_0$, $b(x_0, y)$ and $c(x_0, y)$ have at most one common root in y.

If the triangle is generic, it is also very generic just when there is no zero of the homogeneous polynomials $b(x, y, z)$, $c(x, y, z)$ on the line at infinity, i.e. the line $z = 0$, or, in other words, the highest order terms of $b(x, y)$ and $c(x, y)$ (which are homogeneous of the same degree as $b(x, y)$ and $c(x, y)$) have no common root on $\mathbb{P}^1$. The resultant $\mathrm{res}_{b,c}(x)$ in y depends, up to constant factor, only on the choice of curves B and C and the choice of triangle. The *intersection number* of B and C at a point $p = (x, y)$, denoted BC_p, is the multiplicity of x as a zero of $\mathrm{res}_{b,c}(x)$. If p is not a point of both B and C then define $BC_p := 0$. If p lies on a common component of B and C then let $BC_p := \infty$. Danger: so far, the intersection number depends on the choice of triangle. Let $\deg BC := \sum_p BC_p$ summed over all points p.

Let $b(x, y) := y + x$ and $c(x, y) := y$. The intersection points of the lines $(b(x, y) = 0)$ and $(c(x, y = 0)$ in the affine plane are at $0 = y = y + x$, i.e. at the origin. In the projective plane, there is no further intersection point,

as the equations are already homogeneous. So the standard triangle is very generic for these two lines. The resultant of $b(x,y), c(x,y)$ in y is $r(x) = x$. So the two lines $B = (y + x = 0)$ and $C = (y = 0)$ have $BC_{(0,0)} = 1$. Similarly any two lines have intersection number 1 at their intersection point.

Let $B := (xy^2 + y = 0)$ and $C := (xy^2 + y + 1 = 0)$. The curves intersect in the affine plane at

$$(x, y) = \left(-2, -\frac{1}{2}\right)$$

(assuming that $2 \neq 0$ in our field). Homogenize to find also the intersection point $(0, \infty) = [0 : 1 : 0]$. So the coordinate triangle is not generic. As previously, if we change x to $x + \lambda$ for any constant $\lambda \neq 0$, our curves change to

$$B = (\lambda y^3 + xy^2 + y = 0), C = (2\lambda y^3 + 2xy^2 + y + 1).$$

They now intersect at

$$(x, y) = (-1 - \lambda, 1) = [-1 - \lambda, 1, 1].$$

(We previously saw this from looking at the resultant $r(x) = -\lambda^2(\lambda + 1 + x)$.) After homogenizing and then setting $y = 1$, we also find another intersection at $[x, y, z] = [-\lambda, 1, 0]$. The standard triangle is generic for these curves, since the two intersection points lie on different lines through $[0, 1, 0]$. But it is not very generic, because our second intersection point lies outside the affine plane, on the line at infinity. We can only calculate the intersection number at the first point, where the resultant vanishes to degree 1, so an intersection number

$$BC_{(-\lambda-1,1)} = 1.$$

The other point is not in the affine plane, so intersection multiplicity is not defined at that point, at least by our definition.

Theorem 22.5 (Bézout). *Over any field, take two projective algebraic plane curves B and C not sharing a common component. Over some finite degree extension of the field there is a very generic triangle for the curves. For any generic triangle, $\deg BC \leq \deg B \deg C$. A generic triangle is very generic just when equality holds.*

Proof. Split B and C into irreducibles and add up intersections by multiplying resultants: without loss of generality we can assume that B and C are irreducible.

By proposition 14.10 on page 124, there are at most $\deg B \deg C$ values of x that occur on intersection points, and by the same reasoning at most that number of y values, so finitely many intersection points.

Since there are finitely many intersection points, there are finitely many lines through them. By lemma 22.1 on page 186 we can pick one vertex of our triangle to avoid lying on any of them (after perhaps replacing by an extension field), while arbitrarily picking the other two vertices to be any two distinct points. So we have a generic triangle, and we can assume that its vertices are $(0, 0)$, $(\infty, 0)$ and $(0, \infty)$. We can then move the line at infinity to avoid the intersection points (again in some extension field), so there is a very generic triangle.

Pick any generic triangle. No two intersection points lie on any line through $(0, \infty)$, i.e. on any vertical line: all of the intersection points have distinct x values. There is no intersection point at $(0, \infty)$, i.e. the highest order terms in y in the equations have no common zero. Therefore even when $b(a, y)$ drops degree at some value of $x = a$, $c(a, y)$ doesn't, so the resultant is the product of the values of $b(a, y)$ at the roots y of $c(a, y)$, vanishing precisely at the values of x where the curves have an intersection. In the proof of proposition 14.10 on page 124, we found the degree of the resultant to be at most $\deg b \deg c$ and nonzero. There are at most a number of intersection points given by the degree of the resultant as a polynomial. Hence the sum of the intersection numbers over the intersections in the affine plane is the sum of the degree of root over all roots of the resultant, so at most the degree of the resultant. Our triangle is very generic just when this sum covers all intersection points, as there are none on the line at infinity.

We still need to see that for our very generic triangle, the degree of the resultant is precisely $\deg B \deg C$. The assumption of being very generic is expressed algebraically as saying that the homogeneous polynomials $b(x, y, z), c(x, y, z)$ have no common roots along $z = 0$. In other words, $b(x, y, 0)$ and $c(x, y, 0)$ have no common roots on $\mathbb{P}^1$. So the resultant of $b(x, y, 0), c(x, y, 0)$ in y is nonzero, and has degree $\deg B \deg C$. But the degree of resultant of $b(x, y, z), c(x, y, z)$ can be no less for variable z as for the fixed value $z = 0$. Therefore the resultant has the required degree. $\qquad\square$

Theorem 22.6. *To any two projective algebraic plane curves B and C over a field k and any point p of the projective plane defined over the algebraic closure $\bar{k}$ of k, there is a unique quantity BC_p so that*

a. *$BC_p = CB_p$ and*

b. *$BC_p = 0$ just when p does not lie on any common point of B and C defined over $\bar{k}$ and*

c. *$BC_p = \infty$ just when p lies on a common component of B and C defined over $\bar{k}$ and*

d. *BC_p is a positive integer otherwise and*

e. *two distinct lines meet with intersection multiplicity 1 at their unique point of intersection and*

f. *if B splits into components (perhaps with multiplicities) B_1 and B_2, then $BC_p = B_1C_p + B_2C_p$ and*

g. *if B and C are defined by homogeneous polynomials $b(x, y, z)$ and $c(x, y, z)$ and E is the curve defined by $bh + c$ where $h(x, y, z)$ has degree $\deg h = \deg b - \deg c$, then $BC_p = BE_p$.*

Moreover, BC_p can be computed as defined above using any generic triangle.

Proof. In this proof, we write bc_p instead of BC_p if B is cut out by the equation $0 = b(x, y, z)$ and C by $0 = c(x, y, z)$. First we prove that the conditions a.–g. determine the multiplicity uniquely. Since they are independent of choice of affine chart, this ensures that the multiplicity is also independent. We then check that our definition above satisfies these, so must be determined by these conditions independent of the choice of affine chart made in the definition above.

Any two notions of multiplicity agree on common components of B and C, and on points not belonging to B or C, by the conditions above. So we only need to check points of positive finite multiplicity. We can assume by induction that B and

C are both irreducible, and cut out by irreducible homogeneous polynomials b and c. Suppose that we have two different methods of calculating an intersection number satisfying our various conditions, one of which we can take to be the definition by the resultant above. By induction suppose that they agree wherever they both assign a value less than the larger of the two values that they assign as bc_p.

Take affine coordinates in which $p = (0,0)$. If $\deg b(x,0) = 0$ then $b(x,0) = 0$ so $b(x,y)$ is divisible by y, not irreducible, so our result holds by induction. The same holds if $\deg c(x,0) = 0$. So both b and c have positive degree in each of x and y when the other variable is set to zero. Rescale to get both to have unit coefficient in x:

$$b(x,y) = x^{\beta} + \dots,$$
$$c(x,y) = x^{\gamma} + \dots.$$

Suppose (after perhaps swapping the names of b and c) that $\beta \leq \gamma$ and let

$$h(x,y) := c(x,y) - x^{\gamma - \beta} b(x,y).$$

By construction, $h(x,0)$ has degree in x at most $\gamma - 1$. Our last property of intersection multiplicity above demands that $bc_p = bh_p$. Therefore we can replace c by h and repeat, lowering the degree in x by induction. This proves uniqueness of intersection multiplicities satisfying our axioms.

We want to prove existence, i.e. that the recipe we defined above satisfies our axioms above.

a. Resultants change sign when we swap polynomials, so $\operatorname{res}_{b,c} = -\operatorname{res}_{c,b}$.

b. Follows from theorem 22.5 on page 190.

c. Follows from theorem 22.5 on page 190.

d. Follows from theorem 22.5 on page 190.

e. Follows by direct computation of the resultant.

f. Follows from lemma 13.5 on page 105.

g. Follows from proposition 13.4 on page 105, and the existence of splitting fields.

$\square$

We restate Bézout's theorem without reference to generic triangles, since the previous theorem proves that they are irrelevant.

Theorem 22.7 (Bézout)**.** *Over any field k, take two projective algebraic plane curves B and C not sharing a common component. Over the algebraic closure $\bar{k}$ of the field, the intersection numbers of the intersection points sum to*

$$\sum_{p \in \mathbb{P}^2(\bar{k})} BC_p = \deg B \deg C.$$

Sage

Pieter Belmans wrote the following sage code to find the intersection number of two algebraic curves at a point. First we find the minimum degree of any term in a polynomial $f(x)$ of one variable.

```
def ldegree(f):
    minimum = infinity
    for (n, m) in f.dict():
        minimum = min(minimum, n)
    return minimum
```

Given any two polynomials b, c in two variables, we want to figure out what the two variables are:

```
def determine_variables(b, c):
    if len(b.variables()) == 2:
        return b.variables()
    if len(c.variables()) == 2:
        return c.variables()
    if len(b.variables()) == len(c.variables()) == 1:
        if b.variables() == c.variables():
            return (b.variable(0), 0)
        else:
            return (c.variable(0), b.variable(0))
    return (0,0)
```

Finally, we use our definition of intersection number, and induction, to compute intersection numbers recursively.

```
def intersection_number(b, c, point = (0,0)):
    (x,y) = determine_variables(b, c)
    # translate both curves to origin and calculate it there
    b = b.subs({x:x + point[0], y:y + point[1]})
    c = c.subs({x:x + point[0], y:y + point[1]})
    # if $b(0,0)\neq 0$ or $c(0,0)\neq 0$ they don't intersect in the origin
    if b.subs({x:0, y:0}) != 0 or c.subs({x:0, y:0}) != 0:
        return 0
    # if $b$ or $c$ are zero they don't intersect properly
    if b == 0 or c == 0:
        return Infinity
    # we only look at factors of $x$
    f = b.subs({y:0})
    g = c.subs({y:0})
    # $b$ contains a component $y=0$
    if f == 0:
        # $c$ contains a component $y=0$ too, no proper intersection
        if c == 0:
            return infinity
        # remove common $y^n$ in $b$, count degree of $x$ in $c$ and recurse
        else:
            f = b.quo_rem(y)[0]
            return ldegree(g) + intersection_number(f, c)
    # $b$ does not contain a component $y=0$
    else:
        # $c$ *does* contain a component $y=0$
        if g == 0:
```

```
        g = c.quo_rem(y)[0]
        return ldegree(f) + intersection_number(b, g)
    # we recurse by removing factors of $x$
    else:
        p, q = f.lc(), g.lc()
        r, s = f.degree(), g.degree()
        # we drop the highest degree term
        if r <= s:
            return intersection_number(b, p*c - q*x^(s-r)*b)
        else:
            return intersection_number(q*b - p*x^(r-s)*c, c)
```

A test run:

```
P.<x,y> = PolynomialRing(QQ)
b = P(x^2+y^2)^2+3*x^2*y-y^3
c = P(x^2+y^2)^3-4*x^2*y^2
intersection_number(b,c)
```

yields 14, the intersection number at the origin of the curves:
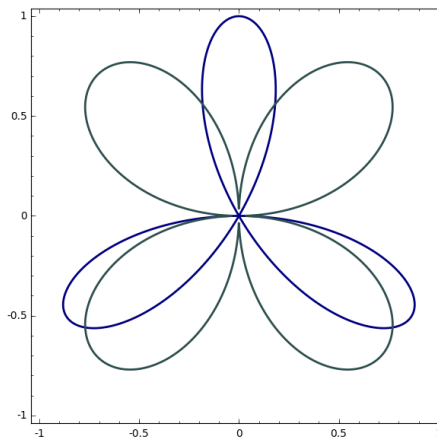
Chapter 23

Families of plane curves

All happy families are alike; each unhappy family is unhappy in its own way.

— Leo Tolstoy
ANNA KARENINA

Pencils of curves

Every curve C of degree n has a homogeneous equation $c(x, y, z) = 0$ of degree n, unique up to rescaling, by Study's theorem (theorem 22.3 on page 187). Take two curves B and C of the same degree n, with equations $b(x, y, z) = 0$ and $c(x, y, z) = 0$. The level sets

$$\frac{b(x, y, z)}{c(x, y, z)} = r$$

of the ratio form a collection of curves C_r, called the *pencil* of curves containing B and C. The curve B is C_0 while C is C_∞. Every point $[x, y, z]$ in the plane lies on a unique one of these curves, with value r given by taking the value of $p(x, y, z)/q(x, y, z)$ at that point, except for the points in $B \cap C$, for which the ratio is never defined. To be more precise, each curve C_r can also be written as

$$p(x, y, z) = rq(x, y, z),$$

(except for $C_\infty = C$). The points of $B \cap C$ belong to *all* curves C_r of the pencil. Moreover the degree of C_r is the same as that of B and C.

> The lines through a point form a pencil: if $B = (x = 0)$ and $C = (y = 0)$ then $C_r = (x = ry)$.
>
>
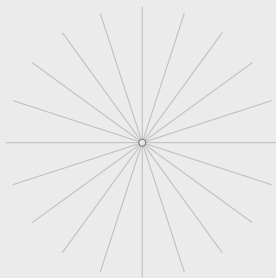
The pencil of conics containing the circle $x^2 + y^2 = 1$ and the reducible conic $x^2 = y^2$ (which is a pair of lines $x = \pm y$):
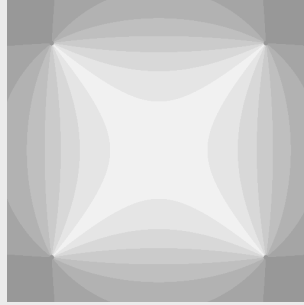
Look at the 4 points where the circle intersects the two lines; these points lie in every conic in the pencil.

The pencil of cubics containing $y^2 = x + x^3$ and $xy^2 = 0$:
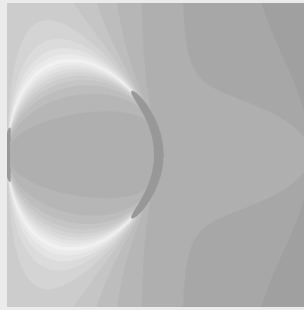
Danger: for some values of r, C_r might actually be of lower order than B or C. For example, if $B = (y^2 = x)$ and $C = (-y^2 = x)$, two conics, the pencil has curves C_r with equation

$$(1 - r)\, y^2 = (1 + r)x,$$

which drops order at $r = 1$, i.e. C_1 is a line. But note that if we homogenize,

$$(1 - r)\, y^2 = (1 + r)zx,$$

the curve C_1 is actually a pair of lines. In a pencil C_r for which C_0 and C_∞ are irreducible, we rarely but occasional find reducible curves, as in this example, and we might picture that the moving curve C_r "bubbles" into reducible components.

Creative Commons license, Attribution NonCommercial Unported 3.0, Fir0002/Flagstaffotos

Lemma 23.1. *Suppose that two projective plane curves B and C of order n intersect at exactly n^2 points (the maximum possible by Bézout's theorem). Write n as $n = p+q$. If exactly pn of these points lie on a curve E of degree p (again the maximum possible) then the remaining points, qn of them, lie on a curve F of degree q (again the maximum possible).*

Proof. Write B and C as C_0 and C_∞ in a pencil C_r. Pick a point s different from the intersection points of C_0 and C_∞. We can pick such a point, because we can replace our field by its algebraic closure if needed, or more simply by any infinite field containing. The numbers of intersection points can't increase when we work in a larger field, since Bézout's theorem tells us that they are maximal already. The point s then belongs to a unique element of the pencil, say to C_{r_0}.

In particular, we can pick s to belong to E. But then s sits in C_{r_0} along with the points of $B \cap C$ and s also sits in E along with pn points of $B \cap C$. So E and C_{r_0} contain $1 + pn$ points of $B \cap C$. By Bézout's theorem, the degrees of E and C_{r_0} are p and n, they contain a common irreducible component. Since E is irreducible by assumption, E is that component of C_{r_0}. So $C_{r_0} = E \cup F$, for some curve F of degree at most q, i.e. the pencil bubbles into two components. The remaining intersection points of $B \cap C$ don't lie in E, but all lie in C_{r_0}, so they lie in F. $\qquad\square$

Polygons

In a projective plane, we have a concept of line, but not of line segment, or of angle or distance. In a projective plane a *triangle* is a triple of points not all on the same line, together with the lines that connect them. So a triangle looks like
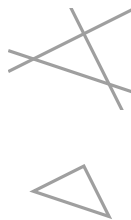
rather than

Similarly, a *quadrilateral* is a quadruple of points, called the *vertices*, no three on the same line, with a chosen ordering defined up to cyclically permuting, so there is no first point or second point, but if you pick a point from the quadruple, there is a next point, and a next, and so on until you come back to the point you started with. In the same way, we define a *polygon*. The lines through subsequent vertices are the *edges*. We can't define the notion of a square or rectangle since there is no notion of length or angle available.

Pascal's mystic hexagon

Draw a hexagon with vertices sitting inside a conic.
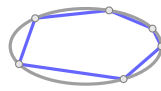
Since there is an even number of sides, each side has an "opposite" side, half way around the hexagon. Pick a side and its opposite, and draw lines joining them.
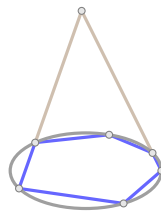
The same for the next pair of opposite sides.
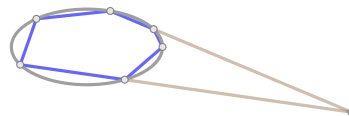
And finally for the third pair of opposite sides.
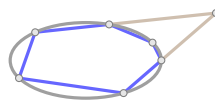
Draw all of these lines together in one picture.
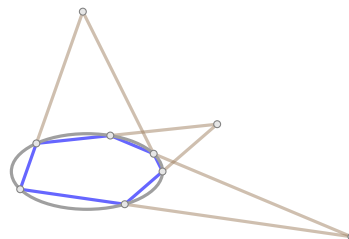
Mysteriously, all three intersection points lie on a line.
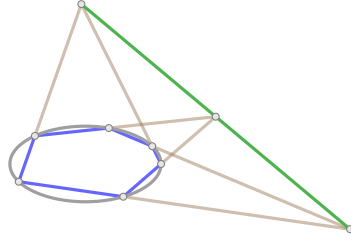
Theorem 23.2. *Given a hexagon in the projective plane over a field, the intersection points of opposite sides lie on a line.*

Proof. Let B be the reducible cubic curve consisting of three of the lines, no two of them in succession (for example, the first, third and fifth lines). Let C be the reducible cubic curve consisting of the remaining three lines of the hexagon. Let E be the quadric. The result follows immediately from lemma 23.1 on page 197. $\qquad\square$

The space of curves of given degree

Consider the vector space V_d of homogeneous polynomials of degree d in 3 variables. Each such polynomial has a coefficient of $x^a y^b z^c$ as long as $a + b + c = d$. Imagine that we take $d + 2$ sticks sitting in a row. Colour 2 of them white and the rest black. Take a to be the number of black sticks before the first white one, b the number of black sticks between the two white ones, and c the number of black sticks after the second white one. Clearly $a + b + c = d$. So the dimension of V_d as a vector space is

$$\binom{d+2}{2}.$$

But since the equation $p_C = 0$ of a curve C is defined only up to rescaling, the curves of degree d are identified with points of $\mathbb{P}^{d^*}(k)$ where

$$d^* = -1 + \binom{d+2}{2} = \frac{d(d+3)}{2}.$$

d	d^*
1	2
2	5
3	9
4	14
5	20
6	27
7	35
8	44
9	54

For example, the set of lines in $\mathbb{P}^2(k)$ is a 2-dimensional projective space, a projective plane, the *dual projective plane*. More generally, our projective space $\mathbb{P}^{d^*}(k)$ contains all curves of degree d, reducible or irreducible.

A pencil $C_r = (p = rq)$ of two curves B and C is just a line inside $\mathbb{P}^{d^*}(k)$. More generally, a *linear system* is collection of plane curves whose equations form a projective linear subspace in $\mathbb{P}^{d^*}(k)$.

The condition that a point $q_0 = (x_0, y_0, z_0)$ lies in a curve $C = (c(x, y, z) = 0)$ is a linear equation $c(x_0, y_0, z_0) = 0$ on the coefficients of the polynomial $c(x, y, z)$, so is satisfied by a linear projective hyperplane inside $\mathbb{P}^{d^*}(k)$. For example, the condition that a point lie in a line is one equation, so determines a line in the dual projective plane. Similarly, the condition that a curve has order at least k at a point q_0 of the plane is a collection of $k(k+1)/2$ linearly independent equations, killing the terms in the Taylor series up to order k (if we arrange that q_0 is the origin, for example). A collection of points $p_1, p_2, \ldots, p_n$ in the projective plane is in *general position* for curves of degree d if the linear system of curves passing through them has either (1) the minimal dimension possible $d(d+3)/2 - n$ if this quantity is not negative and (2) is empty otherwise.

Lemma 23.3. *The definition of general position makes sense, in that if we fix n points, then the linear system consisting of all degree d curves through those points has dimension at least*

$$d^* - n = \frac{d(d+3)}{2} - n.$$

After perhaps replacing our field with an infinite extension field, if $d^* - n \geq 0$ then equality is acheived for some collection of n points, while if $d^* - n < 0$ then there is a collection of n points not lying on any curve of degree d. In particular, if $d^* = n$, then (after perhaps an infinite field extension) there is a set of n points for which there is a unique curve of degree d passing through those n points.

Proof. The proof is by induction. For $d = 1$, there is a pencil of lines through a point $(n = 1)$ and a unique line through any 2 points $(n = 2)$ and no line through 3 or more points, if we pick our 3 points correctly. Suppose that the result is true for all values of n and d less than some particular choices of n and $d \geq 2$. Note that

$$d^* = \frac{d(d+3)}{2} = d + 1 + \frac{(d-1)(d-1+3)}{2} = d + 1 + (d-1)^*.$$

Pick $d + 1$ points on a line L, and then pick $(d-1)^*$ points in general position, so that they lie on a unique curve C of degree $d - 1$. By Bézout's theorem, the line L is a component of any curve D of degree d through the first $d + 1$ points. Therefore such a curve D is just $C \cup L$. Hence D is the unique curve of its degree through the given points. So through some collection of $d(d+3)/2$ points, there is a unique curve of degree d. If we add any more points, choosing them not to lie on the that curve, then there is no curve of degree d through those points. If we take away any points, say down to some number n of point, then the dimension goes up by at most one for each removed point, giving us our result by induction. □

Through any 2 points there is precisely one line: $\circ\!\!-\!\!\!-\!\!\!-\!\!\circ$

Through any 5 points in general position, there is precisely one conic.
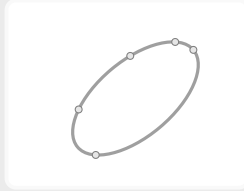
Through any 9 points in general position, there is precisely one cubic.
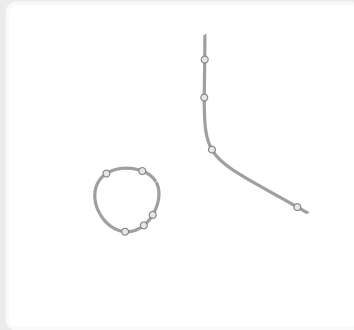
Through any 14 points in general position, there is precisely one quartic.

Through any 20 points in general position, there is precisely one quintic.

Lemma 23.4. *Through any 5 points in the plane, there is a conic.*

- *No three of the points lie on a line just when the conic is irreducible.*
- *No four of the points lie on a line just when the conic is unique.*
- *Four but not five of the points lie on a line just when the conics through all five form a pencil, with each curve in the pencil being a pair of lines: the line through the four colinear points, and any line through the fifth point.*
- *All five points lie on a line just when the conics form a 2-dimensional linear system, with each curve in the system being a pair of lines: the line through the five points, and any other line.*

Proof. The space of conics through four points has dimension at least $d^* - 4 = 2^* - 4 = 5 - 4 = 1$. Indeed there is a pair of lines covering those four points. Add a fifth point, not colinear with any three of the other four. The condition of containing the fifth point is an additional linear equation on the quadratic polynomial of the conic, independent of those for the first four points, as it is not satisfied for that pair of lines. So the linear system of conics through the five points is a smaller dimensional projective space, but still not empty.

If no three of the points are colinear, then repeating the argument above, each point adds an independent condition, not a linear combination of the others, so there is a unique conic through the five points. If that conic is reducible, it is a pair of lines, and then all five points lie on two lines, so at least three points lie on one of the lines. Since some line contains three of the points, by Bézout's theorem, that line is a component of the conic, so the conic must be a pair of lines, one of which is determined.

If no line contains four of the points, then each line is determined by the two points it contains which are not on the other line, and the conic is uniquely determined. $\square$

A points $p_1, p_2, \ldots, p_n$ in the plane *impose independent conditions on curves of degree d* if the linear system of curves through those points has dimension $d^* - n$ (or is empty if $d^* - n < 0$). Lemma 23.3 on page 200 says that (after perhaps a field extension) failure to impose independent conditions on curves of degree d occurs just when the linear system of curves through those points has dimension more than $d^* - n$ (or is not empty, if $d^* - n < 0$), and also says that we can choose n points that impose independent conditions. We want to examine the failure of small numbers of points to impose independent conditions.

Imagine we try to prove that points $p_1, p_2, \ldots, p_n$ impose independent conditions on curves of degree d. Suppose we prove that there is a curve of degree d passing through $p_1, p_2, \ldots, p_{n-1}$ but not p_n. In the vector space of degree d polynomials, the condition of passing through p_n as well is an additional linear equation not always satisfied, so cuts out a linear subspace of one lower dimension. If the choice of which point is p_n is arbitrary, then the same will hold for any of the points p_i in place of p_n, giving n linearly independent equations, i.e. a linear subspace of n lower dimensions, so the points impose independent conditions on curves of degree d.

Theorem 23.5. *Suppose that $p_1, p_2, \ldots, p_n$ are distinct points in the plane, and $n \leq 2d + 2$. Then these points fail to impose independent conditions on curves of degree d just when either*

- *$d + 2$ of the points are colinear or*

- *$n = 2d + 2$ and the points $p_1, p_2, \ldots, p_n$ lie on a conic.*

Proof. Suppose that $d + 2$ of the points are colinear, say on a line L, and say they are $p_1, p_2, \ldots, p_{d+2}$. By Bézout's theorem, any curve of degree d containing those points contains L so has equation factoring into a product of the equation of L and an equation of degree $d - 1$. The set of curves of degree d containing L is a projective space of dimension $(d-1)^*$, and so imposing that our curve has degree d and contains L is imposing $d^* - (d-1)^* = d + 1$ conditions. The remaining points are $p_{d+3}, \ldots, p_n$, so $n - (d+2)$ points, and so together impose at most $n - (d+2)$ conditions, i.e. $p_1, \ldots, p_n$ together impose at most $d + 1 + n - (d+2) = n - 1$ conditions at most, so the points fail to impose independent conditions.

Suppose that $n = 2d+2$ and that the points $p_1, p_2, \ldots, p_n$ lie on a conic. Construct a curve of degree d through the points by constructing a conic through the points, and then constructing any curve of degree $d-2$, i.e. $(d-2)^*$ dimensions of curves at least. If the points were to impose independent conditions, we would have $d^* - n = d^* - (2d+2)$ dimensions of curves. But

$$(d-2)^* = \frac{d^2 - d - 2}{2} > \frac{d^2 - d - 4}{2} = d^* - n.$$

Now we suppose that $p_1, p_2, \ldots, p_n$ fail to impose independent conditions, and that $n \leq 2d + 2$. Some special cases to get started:

- $n = 0$: no conditions, so failure is impossible, i.e. d^* dimensions of curves.

- $n = 1$: one point cannot fail to impose independent conditions, as there is a curve of any degree $d > 0$ not passing through that point (at least after a field extension).

- $n = 2$: for any two points, take d lines through one point none of which contain the other (at least after a field extension). Their union is a degree d curve.

- $n = 3$: if the points are not colinear, a line through two of the points and not the third, taken d times, is a degree d curve. If the points are colinear, and $d = 1$, they fail to impose independent conditions, but satisfy our theorem. So we can suppose that the points are colinear, and $d \geq 2$. Pick three more points q_1, q_2, q_3 so that the five points p_1, p_2, q_1, q_2, q_3 contain no three colinear points. (We can do this after perhaps a field extension.) The conic through these five points is irreducible, so does not contain the line through p_1, p_2, p_3, by Bézout's theorem. Add some lines to that conic, none of them through p_3, to get a degree d curve through p_1, p_2 missing p_3.

- $d = 1$: degree d curves are lines. We have $n \leq 2d + 2 = 4$ points. Our n points fail to impose independent conditions on lines just when either

    - $n \geq 3$ and the points are colinear or

    - $n = 4$: all 4 points lie on a conic, for example on a pair of lines.

- $d = 2$: degree d curves are conics. Suppose that no $d + 2 = 4$ points lie on a line. By hypothesis, $n \leq 2d + 2 = 6$, so we need to consider $n = 4, 5, 6$. The conic through any 5 of the points is unique, by lemma 23.4 on page 201.

    - $n = 5$: the points impose independent conditions because $d^* - n = 0$ and the space of conics through all 5 points is zero dimensional.

    - $n = 4$: add a fifth point not on any line through the four points, and get a unique conic through all five, so the five points impose independent conditions, and so the four points do.

    - $n = 6$: any five of the points lie on a conic, so the sixth point fails to give an independent condition just when all six lie on the same conic.

- $n \leq d + 1$: pick lines $L_1, L_2, \ldots, L_{n-1}$, each line L_i passing through the point p_i and no other of the points. Then $C := L_1 \cup L_2 \cup \cdots \cup L_{n-1}$ has degree $n - 1$. If $n - 1 < d$ then pick some curve D of degree $d - n + 1$, not passing through p_n, and we get a curve $C \cup D$ of degree d through all but one point. To get this to work, we need $n - 1 \leq d$, i.e. $n \leq d + 1$.

By induction, suppose we have proven our result for all smaller values of d, and, for the given d, for all smaller values of n. Suppose that among $p_1, \ldots, p_n$, the points $p_1, \ldots, p_{d+1}$ lie on a line L. If any more of our points lie on that line, the result is proven. Let $d' := d - 1$ and $n' := n - (d + 1)$ and let

$$p_1', \ldots, p_{n'}'$$

be the points

$$p_{d+2}, \ldots, p_n.$$

If $p_1', \ldots, p_{n'}'$ impose independent conditions on curves of degree d', then we find a curve C of degree $d' = d - 1$ through all but one of $p_{d+2}, \ldots, p_n$, and then $L \cup C$ has degree d and passes through all but one of $p_1, \ldots, p_n$, so $p_1, \ldots, p_n$ are independent conditions. Hence $p_1', \ldots, p_{n'}'$ fail to impose independent conditions.

By induction, either

- $d' + 2$ of the points $p_1', \ldots, p_{n'}'$ are colinear or
- $n' = 2d' + 2$ and the points $p_1', \ldots, p_{n'}'$ lie in a conic.

The second possibility, $n' = 2d' + 2$, expands out to become $n = 3d + 1$. But $n \le 2d + 2$, which forces $d \le 1$, not possible. Hence $d' + 2$ of the points $p_1', \ldots, p_{n'}'$ are colinear, i.e. $d + 1$ of the points $p_{d+2}, \ldots, p_n$ are colinear, say on a line L'. So the conic $L \cup L'$ contains $2d + 2$ of the points $p_1, \ldots, p_n$. But $n \le 2d + 2$, so $n = 2d + 2$. So we have a contradiction: no $d + 1$ of our points are colinear.

Suppose that some number ℓ of our points are colinear, say $p_1, p_2, \ldots, p_\ell$. Let $d' := d - 1$ and $n' := n - \ell$ and let

$$p_1', \ldots, p_{n'}'$$

be the points

$$p_{\ell+1}, \ldots, p_n.$$

If $p_1', \ldots, p_{n'}'$ impose independent conditions on curves of degree d', then we find a curve C of degree d' through all but one of $p_1', \ldots, p_{n'}'$. Then $L \cup C$ has degree d and passes through all but one of $p_1, \ldots, p_n$, so $p_1, \ldots, p_n$ are independent conditions, a contradiction. Hence $p_1', \ldots, p_{n'}'$ fail to impose independent conditions.

By induction, either

- $d' + 2$ of the points $p_1', \ldots, p_{n'}'$ are colinear or
- $n' = 2d' + 2$ and the points $p_1', \ldots, p_{n'}'$ lie in a conic.

The first possibility, $d + 1$ of the points are colinear, we have already ruled out. So $n' = 2d' + 2$, expands out to become $n = 2d + 2 + \ell - 2$. But $n \le 2d + 2$, which forces $\ell \le 2$. So no 3 of our points are colinear.

Suppose that $p_3, \ldots, p_n$ impose independent conditions on curves of degree $d - 1$. Then there is a curve C of degree $d - 1$ through all of $p_4, \ldots, p_n$ and not through p_3. But then $p_1 p_2 \cup C$ has degree d and misses p_3, a contradiction. So any $n - 2$ of our points fail to impose independent conditions on curves of degree $d - 1$. We let $n' = n - 2$ and $d' = d - 1$, and apply induction to see that $n' \le 2d' + 2$, so either

- $d' + 2$ points are colinear, or
- $n' = 2d' + 2$ and $p_3, \ldots, p_n$ lie on a conic.

The first case expands out to $d + 1$ points are colinear, a contradiction to the results above. The second case expands out to $n = 2d + 2$ and $p_3, \ldots, p_n$ lie on a conic. But this is true for any reordering of our points: any set of $n - 2$ of our points lie on a conic. If $d = 1$ or $d = 2$, we have already checked our result above. For $d \ge 3$, since we have found that $n = 2d + 3 \ge 9$, so $n - 2 \ge 7$. If you choose $n - 2$ points and I choose $n - 2$ points, they must have at least $n - 4 \ge 5$ points in common. Any 5 of our points lie on at most one conic, since no three are colinear, by lemma 23.4 on page 201. So all n points lie on a conic. □

Chapter 24

Elliptic curves

It is possible to write endlessly on elliptic curves. (This is not a threat.)

— Serge Lang
ELLIPTIC CURVES: DIOPHANTINE ANALYSIS

Cubic curves

Theorem 24.1 (Chasles)**.** *Suppose that C_1 and C_2 are plane cubic curves intersecting at 9 points over the algebraic closure of the field. Every plane cubic curve passing through any 8 of those points passes through all 9. The plane cubic curves passing through those 9 points are precisely the pencil through C_1 and C_2.*

Proof. By theorem , with $n = 8$ and $d = 3$, either

  a. 8 of the points impose independent conditions or

  b. 5 of the points are colinear or

  c. all 8 of the points lie on a conic.

If 7 of the points lie on a conic, then by Bézout's theorem that conic lies in both C_1 and C_2, so C_1 and C_2 have infinitely many points in common, a contradiction. If 4 of the points lie on a line, then by Bézout's theorem that line lies in both C_1 and C_2, so C_1 and C_2 have infinitely many points in common, a contradiction. So all 8 points inmpose independent conditions, i.e. the projective space of cubics through all 8 points has dimension $3^* - 8 = 1$. Since $C_1 \neq C_2$, the projective space of cubics through all 9 points has dimension at least 1: the pencil through C_1 and C_2 is 1-dimensional. But the family through all 9 points is a projective subspace of the family through all 8 points. Both projective spaces have the same dimension, hence these two projective spaces are equal. □
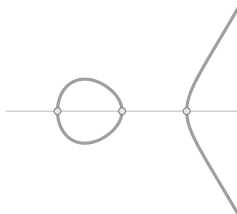
For any two distinct points p, q of a cubic curve C, the line through p and q intersects C in three points; let r be the third point. If p and q are defined in some field k, then so is the line through p and q. Suppose that C is also defined over k. On that line, parameterized over k as $y = mx + b$, the cubic equation of C (having coefficients over k) restricts to have at least two roots in k, corresponding to the points p and q. So that cubic polynomial on the line factors into linear factors over k. That third root gives the location of r on that line: the point r is also defined over k.

We can extend this definition to allow $p = q$, by taking r to be the point of C at which the tangent line to ℓ to C at p touches C again.
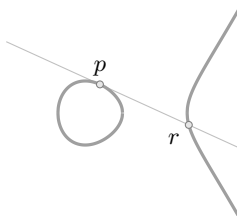
With a little algebra by hand, and some sage code to solve the mess, we can find the intersection points of tangent lines. Consider the curve

$$C = (y^2 - (x + 1)x(x - 1))$$

and the point p with $p = (x, y)$ with $x = -35/100$:

```
X=-35/100
Y=sqrt((X+1)*X*(X-1))
# The slope of the tangent line is given by implicit differentation.
M=(3*X^2-1)/(2*Y)
# Write the equation of the curve restricted to the tangent line.
x=var('x')
f=(Y+M*(x-X))^2-(x+1)*x*(x-1)
print("The tangent line strikes the curve at ",solve(f,x))
XX=(201601/196560)
YY=-sqrt((XX+1)*XX*(XX-1))
print("The other intersection point is ",XX.n(),YY.n())
plot(f,x,-2,2)
```

Elliptic curves

A *elliptic curve* over a field k is a smooth plane cubic curve C defined over k with a chosen point o of C, also defined over k, which we think of as the "origin" of C. Given any point p of C, the line op contains three points of C (counting with multiplicity if needed), say $p, o, \bar{p}$. Define the *addition* of points p, q by taking the line pq, which then intersects C at 3 points, say p, q, r, and then defining $p + q$ to be $\bar{r}$, so $p + q, r, o$ are the three points at which a line intersects C. Danger: note that this addition is *not* the same as adding the coordinates of the points, as you would do in linear algebra.

What if $p = q$? The tangent line to C at p intersects C at some point not equal to p, say r, and define $p + p = \bar{r}$.

Theorem 24.2. *Any elliptic curve is an abelian group under the addition operation.*

Proof. If p, q, r are the points at which a line intersects a cubic C, then so are q, p, r, and so $p + q = \bar{r} = q + p$. By definition, $p, o, \bar{p}$ are on a line, so $\bar{\bar{p}} = p$ and so $p + o = \bar{\bar{p}} = p$.

Consider the point $\bar{o}$, i.e. so that the line $\bar{o}o$ is the tangent line to C at o. Let $-p = \overline{p + \bar{o}}$, i.e. $-p, p, \bar{o}$ are on a line. So then $p + (-p) = \bar{\bar{o}} = o$.

Any equation like $r + q = p + q$ forces the points $r, q, \overline{r + q}$ to be $r, q, \overline{p + q}$, and hence to be $p, q, \overline{p + q}$, so $p = r$, we can cancel.

By definition of addition, for any two points p, q of C,

$$p, \overline{p + q}, \overline{p + \overline{p + q}}$$

lie in a line, while

$$p, q, \overline{p + q}$$

do as well, so

$$q = \overline{p + \overline{p + q}},$$

i.e.

$$\bar{q} = p + \overline{p + q},$$

a "funny cancellation".

Take three points p, q, r of C. We want to prove that $p + (q + r) = (p + q) + r$. Suppose that $p = o$ or $q = o$ or $r = o$ or $p = r$. The equation $p + (q + r) = (p + q) + r$ follows from any of these. So we can suppose that none of these is satisfied.

Define lines

$$L_1 = p(q + r),$$
$$L_2 = rq,$$
$$L_3 = o(p + q),$$
$$L_1' = r(p + q),$$
$$L_2' = pq,$$
$$L_3' = o(r + q).$$

Note that $p, q + r$ lie on L_1, and so therefore does $\overline{p + (q + r)}$. Similarly,

$$p, q + r, \overline{p + (q + r)} \in L_1,$$
$$r, q, \overline{q + r} \in L_2,$$
$$o, p + q, \overline{p + q} \in L_3,$$
$$r, p + q, \overline{(p + q) + r} \in L_1',$$
$$p, q, \overline{p + q} \in L_2',$$
$$o, q + r, \overline{q + r} \in L_3'.$$

Suppose that $L_1 = L_1'$. So the points $p, q + r, \overline{p + (q + r)}$ are some permutation of the points $r, p + q, \overline{(p + q) + r}$. This forces $p = r$ or $p = p + q$ or $q + r = r$ or $q + r = p + q$, and after cancellation these becomes $p = r$ or $q = o$, contradicting our hypotheses.

So $L_1 \neq L_1'$. Take the point s where L_1 and L_1' intersect. If s lies on C, then s lies on $p, q + r, s$, so $\bar{s} = p + (q + r)$, but also s lies on $r, p + q, s$, so $\bar{s} = (p + q) + r$. So s lies on C just when $p + (q + r) = (p + q) + r$.

Consider the 8 points $o, p, q, r, q+r, p+q$, and a ninth point s. The cubic $L_1 \cup L_2 \cup L_3$ contains all 9 points, as does the cubic $L_1' \cup L_2' \cup L_3'$. Suppose that these cubics have no component in common. Then by Chasles's theorem (theorem 24.1 on page 205), the plane cubics containing the 8 points contain all 9, so s belongs to C.

Suppose that these cubics share a component, say $L_i = L_j'$. The 3 points of C on L_i and the 3 points on L_j' must be the same 3 points. No other point of the cubic C can lie on that line, by Bézout's theorem. So the remaining 5 points lie on the other lines. Each pair of remaining lines is a conic. By lemma 23.4 on page 201, these conics are equal or share a common line through 4 of the points. But we can't have 4 points of C on the same line, by Bézout's theorem. So these conics are equal, i.e. all three lines L_1, L_2, L_3 are the same as L_1', L_2', L_3' up to reordering.

We know that $L_1 \neq L_1'$, so $L_1 = L_2'$ or $L_1 = L_3'$. Swapping p and r swaps primed and unprimed lines.

If $L_2 = L_2'$, the points $r, q, \overline{q+r}$ are the same as $p, q, \overline{p+q}$, up to reordering. If this forces $r = p$, we have seen that $p + (q+r) = (p+q) + r$ follows. Otherwise, we find $\overline{q+r} = p$ and $r = \overline{p+q}$, and then

$$p + (q+r) = p + \bar{p} = \bar{o},$$

and

$$(p+q) + r = \bar{r} + r = \bar{o}.$$

So we can assume that $L_2 \neq L_2'$.

If $L_3 = L_3'$, the points $o, p+q, \overline{p+q}$ are the same as $o, r+q, \overline{r+q}$. Since we suppose that $p \neq r$, $r + q = \overline{p+q}$. By funny cancellation,

$$p + (q+r) = p + \overline{p+q} = \bar{q},$$

while

$$(p+q) + r = \overline{q+r} + r = r + \overline{r+q} = \bar{q},$$

so again $p + (q+r) = (p+q) + r$.

If $L_2 = L_3'$ then $r = o$ or $q = o$, a contradiction.

Hence L_1, L_2, L_3 must be some permutation of L_1', L_2', L_3', but we can't have $L_1 = L_1'$ or $L_2 = L_2'$ or $L_2 = L_3'$ or $L_3 = L_3'$. So $L_1 = L_3'$. Permutation of p and r then forces $L_3 = L_1'$, so $L_2 = L_3'$, a contradiction. $\square$

We always take an elliptic curve to be a smooth cubic C with a chosen point $o \in C$, giving an addition operation as above. If we pick some other point o' of our cubic curve C to be the "origin" of an elliptic curve, then we get another addition operation, say $+'$. The map

$$p \in C \mapsto p + (o' - o),$$

using the operations defined above is an isomorphism of groups from C with $+$ to C with $+'$. So the choice of point o is not very important.

Lines crossing a curve

Proposition 25.1. *Pick an irreducible plane algebraic curve C of degree d and a point p of the plane not lying on C. Among all of the lines through p, every one of them strikes C in exactly d distinct points, except for at most $d(d-1)$ lines.*

Proof. First assume that C is irreducible. Arrange that the point p lies at the origin in an affine chart, so that the lines through p are those of the form $y = mx$ or $x = 0$. Also arrange that C is not entirely on the line at infinity, so contains some finite points. From the equation $q(x, y) = 0$ of the curve C, the intersections lie on $q(x, mx) = 0$, a polynomial of degree at most d in x. This polynomial is not the zero polynomial, because the origin does not lie on C. Therefore the number of intersections is d. The multiplicities will all be 1 except at values of m where the discriminant (as a polynomial in x) of $q(x, mx) = 0$ vanishes. The discriminant is an equation of degree at most $d(d-1)$ in m. If the discriminant vanishes as a polynomial in m, it vanishes as a rational function of m, and so $q(x, mx)$ has a zero of higher multiplicity as a polynomial in x with rational coefficients in m: $q(x, mx) = (x - f(m))^2 Q(x, m)$. But replacing m by y/x we find that $q(x, y)$ is reducible as a polynomial in x with coefficients rational in y, and so by the Gauss' lemma $q(x, y)$ is reducible as a polynomial in x, y, a contradiction. So the discriminant is a nonzero polynomial, and if it is constant we can homogenise and dehomogenise to get it to be nonconstant. $\square$

Given a curve in the projective plane, we can take any of the three homogeneous coordinate variables $[x, y, z]$ and rescale to 1, say to $x = 1$, and see the curve as lying on the affine plane. As we have seen, changing the choice of variable, say to $y = 1$ or to $z = 1$, involves rehomogenising the equation and then plugging in $y = 1$ or $z = 1$, giving us a birational curve.

Looking at our curve in the affine plane, by setting $x = 1$, the curve "homogenises" into its cone in k^3, by homogenising the equation. The tangent line in the affine plane $x = 1$ homogenises into a plane through the origin, the tangent plane to the cone. When we set $y = 1$ or $z = 1$, or more generally we look at how our homogenised cone surface passes through some plane not containing the origin, the tangent plane to the cone becomes a tangent line to the resulting curve. Therefore we can speak of a *tangent line* to an algebraic plane curve in the projective plane, by which we mean the projective line which corresponds to the tangent plane to the cone.

To calculate that tangent line, we can either operate directly in some affine plane, say $x = 1$, and just differentiate implicitly as usual, or we can homogenise, and then differentiate the equation of the surface via partial derivatives.

The curve $y^2 = x^2 + x^3$ has tangent line given by differentiating implicitly:

$$2yy' = 2x + 3x^2,$$

so that at the point $(x, y) = (3, 6)$, we have

$$2(6)y' = 2(3) + 3(3)^2,$$

i.e. the slope of the tangent line is

$$y' = \frac{11}{4},$$

giving the tangent line

$$y - 6 = \frac{11}{4}(x - 3),$$

at the point $(x, y) = (3, 6)$.

For the same curve, we could instead homogenise to the cone

$$y^2 z = x^2 z + x^3,$$

and then treat any one of the variables as a function of the other two, say treating $z = z(x, y)$, so take partial derivatives:

$$y^2 \frac{\partial z}{\partial x} = 2xz + x^2 \frac{\partial z}{\partial x} + 3x^2,$$

$$2yz + y^2 \frac{\partial z}{\partial y} = x^2 \frac{\partial z}{\partial y}.$$

We are interested in the point $(x, y) = (3, 6)$ in the affine plane $z = 1$, so the point $(x, y, z) = (3, 6, 1)$. We plug in these values to our equations:

$$6^2 \frac{\partial z}{\partial x} = 2(3)(1) + 3^2 \frac{\partial z}{\partial x} + 3(3)^2,$$

$$2(6)(1) + 6^2 \frac{\partial z}{\partial y} = 3^2 \frac{\partial z}{\partial y}.$$

Solve these to find

$$\frac{\partial z}{\partial x} = \frac{11}{9},$$

$$\frac{\partial z}{\partial y} = -\frac{4}{9},$$

giving an equation of the tangent plane:

$$z - 1 = \frac{11}{9}(x - 3) - \frac{4}{9}(y - 6),$$

which simplifies to

$$z = \frac{11}{9}x - \frac{4}{9}y.$$

This is the same solution, because along the affine plane $z = 1$ it gives

$$1 = \frac{11}{9}x - \frac{4}{9}y,$$

which, solving for y, is

$$y - 6 = \frac{11}{4}(x - 3),$$

exactly the same solution we had above.

At the point $(x, y) = (0, 0)$, the process above fails to yield a tangent line to the curve $y^2 = x^2 + x^3$ since differentiating implicitly:

$$2yy' = 2x + 3x^2,$$

and plugging in the point $(x, y) = (3, 6)$, we have $0 = 0$, no equation at all. In general, our procedure yields a tangent line to an irreducible curve precisely at the points (x, y) where the polynomial $p(x, y)$ in the equation $p(x, y) = 0$ of the curve has at least one nonzero partial derivative, so that we can solve for $y'(x)$ or for $x'(y)$ implicitly. In this example,

there are clearly 2 different lines tangent to the curve. We find these by taking the homogeneous terms of lowest degree in x, y, $y^2 = x^2$, dropping the x^3 because it has higher order, and then factoring into $y = \pm x$. We can justify this over the real numbers (or the complex numbers), by saying that our equation $y^2 = x^2 + x^3$ has solutions $y = \pm\sqrt{x^2 + x^3} \cong \pm x^2$, asymptotically as $x \to 0$.

Singular points

> *That's what love is. It's a recognition of singularity.*
>
> — Robert Graves
> CONVERSATIONS WITH ROBERT GRAVES

Take any algebraic curve C in the projective plane, and any point p_0 of C. Write out an affine chart in which p_0 is the origin $(x, y) = (0, 0)$, and write out the equation $0 = p(x, y)$ of the curve. We can take $p(x, y)$ to be a product of irreducibles $p_i(x, y)$. Since the origin belongs to C, at least one of the factors $p_i(x, y)$ vanishes there. Expand each $p_i(x, y)$ into sums of constants times monomials. Without loss of generality, we

can work over an infinite field. Rescale (x, y) by a nonzero constant, and then let that constant "get small", i.e. look at the powers of that constant, thinking of higher powers as "smaller". Each monomial term rescales by some power of that factor. Rescale both sides of the equation $0 = p(x, y)$ to get rid of the lowest power. For example if C is $0 = x(y^2 - x^2 - x^3)$, rescaling by a factor λ gives

$$0 = \lambda x \left(\lambda^2 y^2 - \lambda^2 x^2 - \lambda^3 x^3 \right).$$

Divide out the lowest power of λ, in this case λ^2, to get

$$0 = x \left(y^2 - x^2 - \lambda^3 x^3 \right).$$

The invariant terms give a homogeneous polynomial $x \left(y^2 - x^2 \right)$. The zeroes of this polynomial form the tangent lines to the curve. By homogeneity, the number of tangent lines is the degree of the homogeneous polynomial, which is at most the degree of the curve. The *order* of a point p_0 on a curve C is the number of tangent lines, counted with multiplicity. A *regular point* (also called a *smooth point*) of an algebraic curve is a point of order 1; any other point is a *singular point*. A curve without singular points (over the algebraic closure of the field) is *smooth* or *regular*.

> At any point of order 1, make a projective change of variables to put the point at the origin, and the line to be the horizontal line $y = 0$. So our curve has equation $p(x, y) = 0$ with $\frac{\partial p}{\partial y} \neq 0$ and with $\frac{\partial p}{\partial x} = 0$ at the origin. After a constant rescaling, our equation is $y = ax^2 + bxy + cy^2 + \dots$. Note that we cannot rid our equation of these higher order terms in y.

> The curve $y^3 = x^4$ over the real numbers is the graph of a differentiable (but not polynomial) function $y = x^{4/3}$. Its tangent lines at the origin arise from keeping the lowest degree homogeneous terms: $y^3 = 0$, i.e. there is a unique tangent line $y = 0$. Nonetheless, over the complex numbers this curve is *not* the graph of a differentiable complex function of a complex variable.

> The singular points of an affine algebraic curve C with equation $0 = p(x, y)$ are precisely the points where
>
> $$0 = p(x, y) = \frac{\partial p}{\partial x} = \frac{\partial p}{\partial y},$$
>
> by definition. Similarly, the singular points of a projective algebraic curve are the singular points of one of its representatives in an affine chart.

> Take a reducible curve and write it as a union of irreducible component curves. Correspondingly, write its equation as a product of two polynomials. Any two of the irreducible components intersect somewhere (at least in the points defined over the algebraic closure of the field), say at a point p. Take affine coordinates in which p is the origin. Each of the two components has equation given by a polynomial vanishing at the origin, so the original curve is a

product of such polynomials. Therefore the equation of the reducible curve vanishes to at least second order. Therefore every reducible curve is singular. Consequently, every nonsingular curve is irreducible.

The curve $y = x^n$ in the affine chart $z = 1$ has no singularities in that chart, if $n \geq 1$ is an integer. In the chart $x = 1$ it becomes $z^{n-1}y = 1$ which has singularities just when

$$z^{n-1}y = 1 \text{ and } z^{n-1} = 0 \text{ and } (n-1)z^{n-2}y = 0,$$

which never happen simultaneously. Therefore over any field, the curve $y = x^n$ is regular for every $n = 1, 2, \ldots$. In particular, there are regular curves of all degrees.

25.1 Suppose that every intersection point of two projective algebraic plane curves B and C, over an algebraically closed field, has multiplicity one. Prove that B and C are regular at these points.

25.2 Prove that every regular projective algebraic plane curve is irreducible. Give an example of reducible regular affine algebraic plane curve.

Lemma 25.2. *The number of singular points of an irreducible algebraic curve of degree d is at most $d(d-1)$.*

Proof. The singular points are the solutions of

$$0 = p(x, y) = \frac{\partial p}{\partial x} = \frac{\partial p}{\partial y}.$$

If one or both of these partial derivatives vanishes, then $p(x, y)$ is a function of one variable only, and the result is trivial to prove. If the two functions

$$p(x, y), \frac{\partial p}{\partial x}$$

have a common factor, then since $p(x, y)$ is irreducible, this factor must be $p(x, y)$. But $\frac{\partial p}{\partial x}$ is a polynomial of lower degree in x than $p(x, y)$, a contradiction. $\qquad\square$

25.3 What are the singularities, and their orders, of $y^p = x^{p+q}$ for $1 \leq p, q$?

Theorem 25.3. *The tangent line to a projective plane algebraic curve at a regular point is the unique line that has intersection number 2 or more at that point. More generally, the intersection number of a line ℓ and a curve $C = (0 = f)$ at a point p is the multiplicity of the zero of p as a zero of the restriction of f to ℓ.*

Proof. We can arrange that our coordinates axes form a generic triangle as usual, and that our intersection point is at $(x, y) = (1, 1)$. We can assume that our projective plane algebraic curve is

$$0 = f(x, y) = \sum_j f_j(x)y^j.$$

We can arrange that our line is $y = 1$. The intersection number is the multiplicity of $x = 1$ as a zero of the resultant

$$\det \begin{pmatrix} -1 & & & & & & f_0(x) \\ 1 & -1 & & & & & f_1(x) \\ & 1 & -1 & & & & f_2(x) \\ & & & \ddots & \ddots & & \vdots \\ & & & & 1 & -1 & f_{n-1}(x) \\ & & & & & 1 & f_n(x) \end{pmatrix},$$

and we add each row to the next

$$= \det \begin{pmatrix} -1 & & & & & f_0(x) \\ & -1 & & & & f_0(x) + f_1(x) \\ & & -1 & & & f_0(x) + f_1(x) + f_2(x) \\ & & & \ddots & & \vdots \\ & & & & -1 & f_0(x) + f_1(x) + \cdots + f_{n-1}(x) \\ & & & & & f_0(x) + f_1(x) + \cdots + f_n(x) \end{pmatrix},$$

$$= (-1)^n \left(f_0(x) + f_1(x) + \cdots + f_n(x) \right),$$
$$= (-1)^n f(x, 1).$$

So the intersection number of the line and the curve is the multiplicity of vanishing of $f(x, 1)$ at $x = 1$. Since $f(1, 1) = 1$, the intersection number is positive, and is two or more just when

$$\frac{d}{dx}\bigg|_{x=1} f(x, 1) = 0,$$

i.e. just when $y = 1$ is a tangent line. $\qquad\square$

Theorem 25.4. *Denote by C_p the order of a projective algebraic plane curve C at a point p. Then for any point p and projective algebraic plane curves C, D:*

$$C_p D_p \le C D_p,$$

with equality just when no line defined over any extension field is tangent to both curves at p.

Proof. If C or D is a line, our result is precisely theorem 25.3 on the preceding page. If either C or D is reducible, our result follows from induction on the degrees of the components and adding up multiplicities.

After perhaps replacing our field by a finite degree extension, pick a generic triangle for the two curves, and arrange that its vertices are $(0, 0), (\infty, 0), (0, \infty)$ as usual. It is easier to compute tangent lines if we shift the axes by adding constants to x and y to arrange that the vertices are at $(-1, -1), (\infty, -1), (-1, \infty)$. We can then arrange that our point p is $(x, y) = (0, 0)$. Write the equations of our curves as $C = (f = 0)$ and $D = (g = 0)$. Then $C D_p$ is the order of $x = 0$ as a zero of the resultant $r(x) = \text{res}_{f,g}(x)$ in y. Let $r := C_p$, $s := D_p$. The lowest order homogeneous polynomial terms have order equal to the order of the point on the curve, so we can write out

$$f(x, y) = f_0(x)x^r + f_1(x)x^{r-1}y + \cdots + f_r(x)y^r$$
$$+ f_{r+1}(x)y^{r+1} + \cdots + f_m(x)y^m,$$
$$g(x, y) = g_0(x)x^s + g_1(x)x^{s-1}y + \cdots + g_s(x)y^s$$
$$+ g_{s+1}(x)y^{s+1} + \cdots + g_n(x)y^n.$$

The tangent lines are the zero lines of the homogeneous terms of lowest order in x, y:

$$0 = f_0(0)x^r + f_1(0)x^{r-1}y + \cdots + f_r(0)y^r,$$
$$0 = g_0(0)x^s + g_1(0)x^{s-1}y + \cdots + g_s(0)y^s.$$

We can assume that our axes are chosen so that the horizontal and vertical axes are not tangent to either curve at the origin, i.e. to arrange that neither of these equations are satisfied by $(x, y) = (1, 0)$ or by $(x, y) = (0, 1)$, i.e. none of

$$f_0(0), g_0(0), f_r(0), g_s(0)$$

vanishes. Since these tangent line equations are homogeneous, to find tangent lines we can restrict to $y = 1$:

$$0 = f_0(0)x^r + f_1(0)x^{r-1} + \cdots + f_r(0),$$
$$0 = g_0(0)x^s + g_1(0)x^{s-1} + \cdots + g_s(0).$$

The resultant of these two polynomials doesn't vanish just when these polynomials have no common factor, i.e. just when the curves C and D have no common tangent line; called this resultant R_1.

   The point $(x, y) = (0, 0)$ is by assumption an intersection point of C and D. We can arrange, making our triangle generic enough, again perhaps using a field extension, that the lines $x = 0$ and $y = 0$ do not contain any other intersection points of C and D other than $(x, y) = (0, 0)$. This is precisely saying that our polynomials

$$f(x, y) = f_0(x)x^r + f_1(x)x^{r-1}y + \cdots + f_r(x)y^r$$
$$+ f_{r+1}(x)y^{r+1} + \cdots + f_m(x)y^m,$$
$$g(x, y) = g_0(x)x^s + g_1(x)x^{s-1}y + \cdots + g_s(x)y^s$$
$$+ g_{s+1}(x)y^{s+1} + \cdots + g_n(x)y^n.$$

do not both vanish when we substitute $x = 0$, except at $y = 0$, and vice versa. Plugging in $x = 0$:

$$f(0, y) = f_r(0)y^r + f_{r+1}(0)y^{r+1} + \cdots + f_m(0)y^m,$$
$$g(0, y) = g_s(0)y^s + g_{s+1}(0)y^{s+1} + \cdots + g_n(0)y^n.$$

Dividing out common factors of y:

$$\frac{f(0, y)}{y^r} = f_r(0) + f_{r+1}(0)y + \cdots + f_m(0)y^{m-r},$$
$$\frac{g(0, y)}{y^s} = g_s(0) + g_{s+1}(0)y + \cdots + g_n(0)y^{n-s}.$$

These are both polynomials and don't vanish at $y = 0$, since f and g vanish exactly to orders r and s at $y = 0$. These two polynomials also can't both vanish anywhere else, since there are no common roots of f and g on $x = 0$ except at $y = 0$. Hence the resultant of these two doesn't vanish; call this resultant R_2.

   We next claim that $\text{res}_{f,g}(x) = x^{rs}R_1R_2 + \ldots$ has lowest order term given by these two resultants. Before we prove our claim, note that if our claim is true, then CD_p is given by the order of this resultant, so is at least rs, proving the theorem. Moreover

this order is rs just when $R_1 \neq 0$ (since we know that $R_2 \neq 0$), i.e. just when there are no common tangent lines.

We are left to justify this claim, i.e. to compute out the lowest order term of

$$
\operatorname{res}_{f,g}(x) = \det \begin{pmatrix}
f_0(x)x^r & & g_0(x)x^s & \\
f_1(x)x^{r-1} & \ddots & g_1(x)x^{s-1} & \ddots \\
\vdots & \ddots & \vdots & \ddots \\
f_m(x) & \ddots & g_n(x) & \ddots \\
0 & \ddots & 0 & \ddots
\end{pmatrix}.
$$

The lowest order term in x is the same as the lowest order term of what happens when we plug $x = 0$ in all of the polynomials:

$$
\det \begin{pmatrix}
f_0(0)x^r & & g_0(0)x^s & \\
f_1(0)x^{r-1} & \ddots & g_1(0)x^{s-1} & \ddots \\
\vdots & \ddots & \vdots & \ddots \\
f_m(0) & \ddots & g_n(0) & \ddots \\
0 & \ddots & 0 & \ddots
\end{pmatrix}.
$$

So it suffices to prove our result for the special case that all $f_j(x)$ and $g_j(x)$ are constants, say:

$$
\begin{aligned}
f(x,y) &= a_0 x^r + a_1 x^{r-1} y + \cdots + a_r y^r \\
&\quad + a_{r+1} y^{r+1} + \cdots + a_m y^m, \\
g(x,y) &= b_0 x^s + b_1 x^{s-1} y + \cdots + b_s y^s \\
&\quad + b_{s+1} y^{s+1} + \cdots + b_n y^n.
\end{aligned}
$$

Then the resultant is

$$
\operatorname{res}_{f,g}(x) = \det \begin{pmatrix}
a_0 x^r & & b_0 x^s & \\
a_1 x^{r-1} & \ddots & b_1 x^{s-1} & \ddots \\
\vdots & \ddots & \vdots & \ddots \\
a_m & \ddots & b_n & \ddots \\
0 & \ddots & 0 & \ddots
\end{pmatrix}.
$$

We need to prove that this is $x^{rs} R_1 R_2 + \ldots$. Write out

$$
R_1 = \det \begin{pmatrix}
a_0 & & b_0 & \\
a_1 & \ddots & b_1 & \ddots \\
\vdots & \ddots & \vdots & \ddots \\
a_r & \ddots & b_s & \ddots \\
0 & \ddots & 0 & \ddots
\end{pmatrix}
$$

and

$$R_2 = \det \begin{pmatrix} a_r & & & b_s & & \\ a_{r+1} & \ddots & & b_{s+1} & \ddots & \\ \vdots & \ddots & \vdots & & \ddots & \\ a_m & \ddots & & b_n & \ddots & \\ 0 & \ddots & & 0 & \ddots & \end{pmatrix}.$$

Let's try a simple case: if $r = 2, s = 1, m = 3, n = 2$, the resultant is

$$\operatorname{res}_{f,g}(x) = \det \begin{pmatrix} a_0 x^2 & 0 & b_0 x & 0 & 0 \\ a_1 x & a_0 x^2 & b_1 & b_0 x & 0 \\ a_2 & a_1 x & b_2 & b_1 & b_0 x \\ a_3 & a_2 & 0 & b_2 & b_1 \\ 0 & a_3 & 0 & 0 & b_2 \end{pmatrix}.$$

Sadly, when we expand out this determinant, we get a huge mess:

$$\operatorname{res}_{f,g}(x) = x^2 R_1 R_2$$
$$+ x^3 (a_3^2 b_0^3 - 2\, a_1 a_3 b_0^2 b_2 + 3\, a_0 a_3 b_0 b_1 b_2 + a_1^2 b_0 b_2^2 - 2\, a_0 a_2 b_0 b_2^2 - a_0 a_1 b_1 b_2^2)$$
$$+ x^4 a_0^2 b_2^3$$

where

$$R_1 = a_2 b_0^2 - a_1 b_0 b_1 + a_0 b_1^2$$

and

$$R_2 = a_2 b_2 - a_3 b_1.$$

Nevertheless, the lowest term (as expected) is $\operatorname{res}_{f,g}(x) = x^2 R_1 R_2 + \ldots$.

Returning to the general problem: luckily, we don't really need to compute any of these resultants at all. Suppose that $r \leq s$. After rescaling f, x and y we can assume that $a_r = 1$ and that $a_m = 1$. Every term in $g - b_s x^{s-r} f$ contains a factor of y, so write it as $g - b_s x^{s-r} f = y\tilde{g}$. We will see that replacing g by $\tilde{g}$, the three resultants we have to calculate then all compute the same values. By problem 13.6 on page 106:

$$\operatorname{res}_{f(x,y),g(x,y)}(x) = \operatorname{res}_{f(x,y),y\tilde{g}(x,y)}(x),$$
$$= \operatorname{res}_{f(x,y),y}(x)\operatorname{res}_{f(x,y),\tilde{g}(x,y)}(x),$$
$$= f(x,0)\operatorname{res}_{f(x,y),\tilde{g}(x,y)}(x),$$
$$= x^r \operatorname{res}_{f(x,y),\tilde{g}(x,y)}(x)$$

while, if we let F be the lowest order part of f, G the lowest order part of g, and $\tilde{G}$ the lowest order part of $\tilde{g}$,

$$R_1 = \operatorname{res}_{F(x,1),G(x,1)} = \operatorname{res}_{F(x,1),\tilde{G}(x,1)}$$

and

$$R_2 = \operatorname{res}_{\frac{f(0,y)}{y^r}, \frac{y\tilde{g}(0,y)}{y^s}} = \operatorname{res}_{\frac{f(0,y)}{y^r}, \frac{\tilde{g}(0,y)}{y^{s-1}}}.$$

Hence our claim is true for f, g just when it is true for $f, \tilde{g}$. Similarly, if $r < s$, swap the roles of f and g and repeat. Apply induction on r and s, driving at least one of them to become smaller at each step, with the other staying the same. The induction stops when one of them is zero, and C and D cease to intersect at p. □

Chapter 26

Inflection points

Traditionally, an *inflection point* meant a point where the graph of a function changes from being convex down to being convex up, or vice versa. Clearly at any inflection point, the derivative goes from decreasing to increasing, or vice versa, so the second derivative vanishes. Notions of convexity depend very strongly on having some idea of positive and negative numbers, so they don't apply in the context of algebraic curves. The closest we can come to this is to *define* an *inflection point* or *flex* of an algebraic plane curve to be a point at which the tangent line meets the curve to order 3 or more.

To find such points, we define the *Hessian* $\det b''(x, y, z)$ of a homogeneous polynomial $b(x, y, z)$ by

$$\det b''(x, y, z) = \det \begin{pmatrix} b_{xx} & b_{xy} & b_{xz} \\ b_{xy} & b_{yy} & b_{yz} \\ b_{xz} & b_{yz} & b_{zz} \end{pmatrix}$$

where

$$b_{xy} := \frac{\partial^2 b}{\partial x \, \partial y}$$

and so on. So if $b(x, y, z)$ is homogeneous of degree n, then b_{xy} is homogeneous of degree $n - 2$, so $\det b''(x, y, z)$ is homogeneous of degree $3(n - 2)$. If $b(x, y, z)$ has degree 2, so that the associated curve is a conic, then $\det b''(x, y, z)$ is a constant, the determinant of the symmetric matrix of coefficients of $b(x, y, z)$. So $\det b''(x, y, z)$ is a nonzero constant, i.e. H is empty, just when $b(x, y, z)$ is a smooth conic. Similarly, $\det b''(x, y, z)$ vanishes everywhere for a line or a singular conic (a pair of lines).

If the Hessian is not constant, the *Hessian curve* of the curve $B = (0 = b(x, y, z))$ is the curve $H = (0 = \det b''(x, y, z))$.

Sage can compute Hessians for us:

```
P.<x,y,z> = PolynomialRing(QQ)
b = x^3-y^2*z+z^3
Hb = matrix([[diff(b,x,x),diff(b,x,y),diff(b,x,z)],
             [diff(b,x,y),diff(b,y,y),diff(b,y,z)],
             [diff(b,x,z),diff(b,y,z),diff(b,z,z)]])
factor(Hb.det())
```

yields $(-24) \cdot x \cdot (y^2 + 3z^2)$.

> Working over any field of characteristic 3, the polynomial $b(x, y, z) := x^3 + y^3 - z^3$ has derivatives $0 = b_x = b_y = b_z$, and so has Hessian $\det b''(x, y, z) = 0$.

Lemma 26.1 (Euler)**.** *Every homogeneous polynomial $b(x, y, z)$, say of degree n, over any commutative ring with identity, satisfies*

$$nb(x, y, z) = x\frac{\partial b}{\partial x} + y\frac{\partial b}{\partial y} + z\frac{\partial b}{\partial z}.$$

Proof. Since $b(x, y, z)$ is a sum of homogeneous terms, it is enough to check one such term, which is easy for the reader to check. □

Lemma 26.2. *The Hessian $\det b''(x, y, z)$ vanishes at a smooth point of a plane algebraic curve $B = (0 = b(x, y, z))$ just when the tangent line at that point meets the curve to order 3 or more.*

Proof. Euler's lemma:

$$nb(x, y, z) = x\frac{\partial b}{\partial x} + y\frac{\partial b}{\partial y} + z\frac{\partial b}{\partial z}$$

applies to the homogeneous polynomials $b_x := \frac{\partial b}{\partial x}$ and so on to yield

$$(n-1)b_x = xb_{xx} + yb_{xy} + zb_{xz},$$
$$(n-1)b_y = xb_{xy} + yb_{yy} + zb_{yz},$$
$$(n-1)b_z = xb_{xz} + yb_{yz} + zb_{zz}.$$

Multiply row 1 of the Hessian determinant by x, row 2 by y, and add to the third row multiplied by z to get

$$z \det b''(x, y, z) = (n-1)\det\begin{pmatrix} b_{xx} & b_{xy} & b_{xz} \\ b_{xy} & b_{yy} & b_{yz} \\ b_x & b_y & b_z \end{pmatrix}.$$

Apply the same trick across the columns instead of the rows:

$$z^2 \det b''(x, y, z) = (n-1)^2 \det\begin{pmatrix} b_{xx} & b_{xy} & b_x \\ b_{xy} & b_{yy} & b_y \\ b_x & b_y & \frac{nb}{n-1} \end{pmatrix}.$$

At singular points of the curve $B = (0 = b(x, y, z))$, the final row vanishes, so the determinant vanishes. At any point where the curve $B = (0 = b(x, y, z))$ is not singular, which we can arrange that point to be $(x, y, z) = (0, 0, 1)$. Arrange that the tangent line of B at that point is the x-axis, so $b_y = 1$ while $b_x = 0$. Euler's lemma gives $b_z = 0$. Plug into equation 26 to get $\det b''(0, 0, 1) = (n-1)^2 b_{xx}(0, 0, 1)$. The multiplicity of b restricted to $x = 0$ is then given by expanding out $b(0, y) = b(0, 0) + b_y(0, 0)y + b_{yy}(0, 0)y^2/2 + \ldots$, clearly at least 3.

By implicit differentiation, at any smooth point, the slope of the tangent line is

$$0 = b_x + b_y y'$$

and so differentiating again gives

$$0 = b_{xx} + 2b_{xy}y' + b_{yy}(y')^2 + b_y y'',$$

as the derivative of y'. So

$$y' = -\frac{b_x}{b_y},$$

while

$$y'' = -\frac{b_{xx} + 2b_{xy}y' + b_{yy}(y')^2}{b_y}.$$

Expanding out,

$$y'' = \frac{-b_{xx}b_y^2 + 2b_{xy}b_xb_y - b_{yy}b_x^2}{b_y^3},$$

$$= \frac{1}{b_y^3} \begin{pmatrix} b_{xx} & b_{xy} & b_x \\ b_{xy} & b_{yy} & b_y \\ b_x & b_y & 0 \end{pmatrix}$$

So at any point where $b = 0$, this is

$$y'' = \frac{\det b''(x, y, 1)}{(n-1)^2 b_y^3}.$$

If we take a point (x_0, y_0) at which $b_y \neq 0$, let y_0' be y' at that point, and compute out that on the tangent line,

$$\begin{aligned} b(x, y) &= b(x_0 + \Delta x, y_0 + y_0'\Delta x), \\ &= b(x_0, y_0) + b_x(x_0, y_0)\Delta x + \dots, \\ &= -b_y(x_0, y_0)y_0'\Delta x - b_y(x_0, y_0)y_0''\frac{\Delta x}{2} + \dots \end{aligned}$$

So the Hessian vanishes just where the tangent line meets the curve to order 3 or more, and this occurs just where $y'' = 0$. □

Lemma 26.3. *Take a plane algebraic curve $B = (0 = b(x, y, z))$. Suppose that the characteristic of the field is coprime to the exponents appearing in all terms of $b(x, y, z)$. Then $b(x, y, z)$ has vanishing Hessian just when B is a union of lines. In other words, B has Hessian curve H defined just when B is not a union of lines.*

Proof. Continuing the previous proof, if we suppose that the Hessian vanishes everywhere, we see that y'' is everywhere zero, i.e. we see that at every point where $b = 0$,

$$0 = y'' = (y')_x + (y')_y y'.$$

Plug in $y' = -b_x/b_y$ to get

$$\left(\frac{b_x}{b_y}\right)_x = \left(\frac{b_x}{b_y}\right)\left(\frac{b_x}{b_y}\right)_y.$$

This last equation turns a derivative in x into a product, with the first factor having no derivative in x. Note that it only holds at points of the curve, so holds modulo b. Hence, when we differentiate both sides, the result hold modulo b, b_x, and so on. By induction, at a point where $0 = b = b_x$ and where $0 \neq b_y$, if we differentiate both sides repeatedly, we find a higher order x derivative of b vanishing each time. So at any point where $0 = b = b_x$ we also have $0 = b = b_x = b_{xx} = \dots$. Because the exponents arising in the terms of $b(x, y, z)$ are coprime to the characteristic of the field, the vanishing of derivatives $b_x, b_{xx}, \dots$ implies that the coefficients of $x, x^2, \dots$ in $b(x, y, 1)$ are zero. Every term in b contains a factor of y, so $b = 0$ along the line $y = 0$. □

Lemma 26.4. *Take a plane algebraic curve $B = (0 = b(x, y, z))$, which has Hessian curve H. Then H passes through all singular points of B, intersecting B at $3n(n-2)$ points at most, counting with multiplicity. The curves H and B meet at finitely many points. Moreover, H meets B at a smooth point of B with multiplicity r just when the tangent line at that point meets B with multiplicity $r + 2$.*

Proof. We can assume that the point in question is the origin in the affine plane, and that the tangent line is $y = 0$. Factoring out all copies of x from the terms with no y in them, the equation of B is now

$$0 = b(x, y) = y\, f(x, y) + x^{r+2} g(x),$$

with $f(0, 0) \neq 0$ and $g(0) \neq 0$. We can rescale x, y to arrange $f(0, 0) = 1$ and $g(0) = 1$.

The associated homogeneous polynomial is

$$b(x, y, z) = y\, f(x, y, z) + x^{r+2} g(x, z).$$

Note that the affine equations $f(0, 0) = 1$ and $g(0) = 1$ become homogeneous equations $f(0, 0, z) = z^{n-1}$ and $g(0, z) = z^{n-r-2}$. The Hessian $\det b''(x, y, z)$ is extremely complicated in terms of f, g. The first derivatives are

$$b_x = y f_x + (r+2)x^{r+1} g + x^{r+2} g_x,$$
$$b_y = f + y f_y + x^{r+2} g_y,$$
$$b_z = y f_z + x^{r+2} g_z,$$

and the second derivatives are

$$b_{xx} = y f_{xx} + (r+2)(r+1)x^r g + 2(r+2)x^{r+1} g_x + x^{r+2} g_{xx},$$
$$b_{xy} = f_x + y f_{xy} + (r+2)x^{r+1} g_y + x^{r+2} g_{xy},$$
$$b_{xz} = y f_{xz} + (r+2)x^{r+1} g_z + x^{r+2} g_{xz},$$
$$b_{yy} = 2 f_y + y f_{yy} + x^{r+2} g_{yy},$$
$$b_{yz} = f_z + y f_{yz} + x^{r+2} g_{yz},$$
$$b_{zz} = y f_{zz} + x^{r+2} g_{zz}.$$

Modulo y, x^r, i.e. dropping terms with y or x^r in them, we find

$$\det b''(x, y, z) = \det \begin{pmatrix} 0 & f_x & 0 \\ f_x & 2 f_y & f_z \\ 0 & f_z & 0 \end{pmatrix} = 0.$$

So every term in $\det b''(x, y, z)$ contains either y or x^r:

$$\det b''(x, y, z) = y F(x, y, z) + x^r G(x, y, z)$$

for some polynomials F, G. We can absorb any y terms in G into F, so

$$\det b''(x, y, z) = y F(x, y, z) + x^r G(x, z)$$

for some polynomials F, G. Compute $\det b''(x, y, z)$ modulo y, and divide out x^r from the first row, and then set $x = 0$, to find

$$G(0, 1) = \frac{\det b''}{x^r}(0, 0, 1) = -(r+2)(r+1)g\,(0, 1)\, f_z\,(0, 0, 1)^2\,.$$

We can actually get sage to compute this for us:

```
y=var('y')
z=var('z')
r=var('r')
f = function('f')(x,y,z)
g = function('g')(x,z)
b(x,y,z) = y * f(x,y,z) + x^(r+2)*g(x,z)
m=matrix([
        [diff(b(x,y,z),x,x),diff(b(x,y,z),x,y),diff(b(x,y,z),x,z)],
        [diff(b(x,y,z),x,y),diff(b(x,y,z),y,y),diff(b(x,y,z),y,z)],
        [diff(b(x,y,z),x,z),diff(b(x,y,z),y,z),diff(b(x,y,z),z,z)]])
h=m.det()
factor(simplify(expand((h.subs(y=0)/x^r).subs(x=0))))
```

prints out $-(r+2)(r+1)g\,(0,z)\,\mathrm{D}_2\,(f)\,(0,0,z)^2$. (The sage notation for derivatives is peculiar, and we leave the reader to work out why D_2 means derivative in the third variable.) Since we know that $f(0,0,z) = z^{n-1}$ and $g(0,0,1) = 1$, we can compute out

$$G(0,1) = \frac{\det b''}{x^r}(0,0,1) = -(r+2)(r+1)(n-1)^2.$$

Returning to affine coordinates, at the origin $(x,y) = (0,0)$ in the (x,y) affine plane,

$$G(0) = -(r+2)(r+1)(n-1)^2 \neq 0.$$

We are looking for $H \cap B$ near the origin $(0,0,1)$, i.e. we are looking for common zeroes, with multiplicity, of

$$0 = yf(x,y,z) + x^{r+2}g(x,z),$$
$$0 = yF(x,y,z) + x^r G(x,z).$$

Dehomogenize by setting $z = 1$:

$$0 = yf(x,y) + x^{r+2}g(x),$$
$$0 = yF(x,y) + x^r G(x).$$

If $r = 0$, then the second equation doesn't vanish near $(x,y) = (0,0)$, so the origin is not an intersection point, contradicting our assumptions. So $r \geq 1$.

Suppose that B and H share a common component, and suppose that the origin belongs to that component. If $y = 0$ on some intersection point, then plug in to get that either $x = 0$ or $g(x) = 0$. But $g(0) = 1$, so $g(x)$ is zero at only finitely many points, none near the origin. The same holds over any algebraic extension of our field. So $(y = 0)$ is not a component of $B \cap H$, i.e. if $B \cap H$ contains a curve, it is not inside the line $(y = 0)$. Solve for x^r as

$$x^r = \frac{y}{(r+1)(r+2)(1+\dots)},$$

and plug in to get

$$\frac{x^2 y}{(r+1)(r+2)(1+\dots)} = -y\frac{1+\dots}{1+\dots}.$$

Cancel the y, since $(y = 0)$ is not in $B \cap H$:

$$x^2 = -(r+1)(r+2)\frac{1+\dots}{1+\dots},$$

and clear denominators:

$$x^2(1 + \dots) = -(r+1)(r+2)(1 + \dots).$$

on any common component of B and H. But this is not satisfied at the origin, so the common component does not meet the origin. Recall that we picked any point of $H \cap B$ and made it the origin. Hence B and H share no common component.

We will compute the multiplicity of intersection BH_p at any smooth point p of B. We can assume that we have chosen very generic affine coordinates, so that the line through any two intersection points of B and H does not touch $(0, 1)$ (even after an algebraic extension of the field). We can still arrange that p is the origin $(0, 0)$ and that the tangent line to B at $(0, 0)$ is $y = 0$. Take the resultant $R(x)$ of $yf(x, y) + x^{r+2}g(x)$ and $yF(x, y) + x^r G(x)$. By theorem 22.6 on page 191, BH_p is the order of vanishing of $R(x)$ at the origin: if we write

$$f(x, y) = \sum y^j f_j(x),$$
$$F(x, y) = \sum y^j F_j(x),$$

then

$$R(x) = \det \begin{pmatrix} x^{r+2}g(x) & & & & x^r G(x) & & & \\ f_1(x) & x^{r+2}g(x) & & & F_1(x) & x^r G(x) & & \\ \vdots & f_1(x) & \ddots & & \vdots & F_1(x) & \ddots & \\ f_k(x) & \ddots & \ddots & x^{r+2}g(x) & F_\ell(x) & \ddots & \ddots & x^r G(x) \\ & f_k(x) & \ddots & f_1(x) & & F_\ell(x) & \ddots & F_1(x) \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & f_k(x) & & & & F_\ell(x) \end{pmatrix}.$$

Expand across the top row to see that $R(x)$ is a multiple of x^r. We want to prove that $R(x)$ has order exactly x^r, no something higher. The x^r term in $R(x)$ doesn't notice terms of order x^{r+1} or higher, so we can assume that $g(x)$ and $G(x)$ are constants, both equal to 1. Terms of higher order than x^r make no contribution to the order r term in $R(x)$, so we only need to find the resultant of $yf(x, y), yF(x, y) + x^r$. If we write dots to indicate terms of higher order than x^r,

$$R(x) = \operatorname{res}_{yf(x,y), yF(x,y)+x^r} + \dots,$$
$$= \operatorname{res}_{y, yF(x,y)+x^r} \operatorname{res}_{f(x,y), yF(x,y)+x^r} + \dots,$$
$$= x^r \operatorname{res}_{f(x,y), yF(x,y)+x^r} + \dots.$$

Suppose that the x^r term vanishes. Then $f(0, y)$ and $yF(0, y)$ have a common root (in some algebraic extension), say $y = y_0$. Then $(0, y_0)$ is a common root of $yf(x, y) + x^{r+2}g(x)$ and $yF(x, y) + x^r G(x)$, i.e. an intersection point $(0, y_0) \in B \cap H$. This intersection point is not at the origin, since $f(0, 0) = 1$. This intersection point is on the line between $(0, 1)$ and $(0, 0)$, a contradiction. So the x^r term in $R(x)$ doesn't vanish, i.e. the order of intersection BH_p is r. $\qquad\square$

Corollary 26.5. *A smooth plane algebraic curve of degree n with a defined Hessian curve has exactly $3n(n-2)$ inflection points, counting multiplicities, in some algebraic extension field.*

Proof. Follows from by Bézout's theorem (theorem 22.7 on page 192) and the degree of the Hessian being $3(n-2)$ as above. $\square$

Chapter 27

Conics and quadratic forms

> *I like mathematics because it is not human and has nothing particular to do with this planet or with the whole accidental universe—because, like Spinoza's god, it won't love us in return.*

> — Bertrand Russell

Quadratic forms

A *quadratic form* is a homogeneous polynomial of degree 2.

Proposition 27.1 (Sylvester's Law of Inertia)**.** *Every quadratic form in n variables over the real numbers is identified, by some linear change of variables, with precisely one of*

$$x_1^2 + x_2^2 + \cdots + x_p^2 - \left(y_1^2 + y_2^2 + \cdots + y_q^2 \right)$$

where $p + q \leq n$. The triple of numbers (p, q, r) so that $p + q + r = n$ is called the signature *of the quadratic form.*

Proof. Write the quadratic form as

$$q(x) = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{nn}x_n^2,$$

or in other words as $q(x) = \langle Ax, x \rangle$ for a symmetric matrix A. Apply an orthogonal change of variables which diagonalizes the matrix A. Then rescale each variable to get the eigenvalues of the matrix to be ± 1 or zero. □

Lemma 27.2. *Every quadratic form in n variables over a field k of characteristic not 2 can be brought to the "diagonal" form*

$$q(x) = \sum a_i x_i^2$$

by a linear change of variables. The a_i are uniquely determined up to multiplying by squares $a_i b_i^2$ of nonzero elements b_i of k.

Proof. Take a vector x for which $q(x) \neq 0$. Make a linear change of variables to get $x = e_1$, so that $q(e_1) \neq 0$. Hence after rescaling

$$q(x) = c_1 x_1^2 + \dots,$$

say

$$q(x) = c_1 x_1^2 + c_2 x_1 x_2 + \cdots + c_n x_1 x_n + \dots$$

where x_1 does not appear in the remaining terms. Replace $x_1, x_2, \ldots, x_n$ with $X_1, x_2, \ldots, x_n$ where

$$X_1 := x_1 - \frac{c_2}{2c_1} x_2 - \cdots - \frac{c_n}{2c_1} x_2$$

and check that now

$$q(x) = c_1 X_1^2 + \ldots$$

where the ... don't involve X_1, so we can apply induction. □

For the real numbers, any positive real number has a square root, so we recover the previous lemma.

For $k = \mathbb{C}$, any complex number has a square root, so every quadratic form in n complex variables becomes

$$q(z) = z_1^2 + z_2^2 + \cdots + z_p^2,$$

for a unique $p \le q$.

For $k = \mathbb{Q}$, any rational number, up to multiplying by a nonzero square of a rational number, is a *square-free* integer, i.e. has the form

$$\alpha = \pm p_1 p_2 \ldots p_s$$

where all of the primes

$$p_1, p_2, \ldots, p_s$$

are distinct. So every quadratic form over k has the form

$$q(x) = \sum \alpha_i x_i^2$$

for square-free integers α_i. In the associated conic curve, we can rescale the coefficients, so we can assume that the square-free integers α_i have no common factor. We can also assume, by changing all signs if needed, that there are at least as many positive α_i as negative.

Since every element of any field obtains a square root in some extension, the quadratic forms over any algebraically closed field have the form

$$q(z) = z_1^2 + z_2^2 + \cdots + z_p^2,$$

for a unique $p \le q$.

Theorem 27.3. *Over a field k, take a quadratic form q, say in n variables. Then after perhaps repeatedly replacing k by a quadratic extension of k, finitely many times, and a linear change of variables, we can arrange that either*

$$q(x_1, \ldots, x_n) = x_1 x_2 + x_3 x_4 + x_5 x_6 + \cdots + x_{m-1} x_m$$

or

$$q(x_1, \ldots, x_n) = x_1^2 + x_2 x_3 + x_4 x_5 + \cdots + x_{m-1} x_m,$$

for some integer m.

Proof. A *null vector* of q is a vector v so that $q(v + x) = q(x)$, i.e. translating q by a null vector doesn't change the value of q. The *null space* of q is the set of all null vectors. The null space is a linear subspace of k^n since if v_1 and v_2 belong to the null space, then $q(v_1 + v_2 + x) = q(v_1 + x) = q(x) = q(v_2 + x)$, and if v belongs to the null space, and c is any constant in k, then $q(cv + x) = c^2 \, q(v + x/c) = c^2 \, q(x/c) = q(x)$.

Make a linear change of variables to get various of the variables to parameterize this subspace, i.e. the subspace is given by setting the first j variables $x_1, x_2, \ldots, x_j$ to zero and the last $n - j$ variables $x_{j+1}, \ldots, x_n$ to any values. So q is a function of those first j variables only. We can reduce our problem to the study of those first j variables, i.e. we can assume that the null space is zero.

If $n = 1$, $q = ax^2$, we only need to ensure that a has a square root in k. If $n = 2$, $q = ax^2 + bxy + cy^2$, we only need to ensure that $ax^2 + bx + c$ has a root in k, and then q factors into linear factors, and we change variables to arrange that $q = xy$. So suppose that $n \geq 3$.

A *plane* is a 2-dimensional linear subspace of k^n. On any plane P, we can arrange that $q = 0$ or $q = x_1 x_2$ or $q = x_1^2$, as above, after some linear change of variables. For the moment, suppose that there is a plane P on which we can arrange that $q = x_1 x_2$, after perhaps a quadratic extension of k.

Consider the subset Q of k^n consisting of those vectors w so that $q(v + w) = q(v) + q(w)$ for all v in P. Expand out this equation to see that this is a linear equation in w, so Q is a linear subspace. Moreover, we only need to check the equation $q(v_i + w) = q(v_i) + q(w)$ for any two vectors v_i forming a basis of P, so Q is cut out by two linear equations, so has dimension at least $n - 2$. Check that $P \cap Q = \{0\}$, because on P, $q = x_1 x_2$. So Q is a linear subspace of k^n of dimension $n - 2$ exactly.

Make a linear change of variables so that P is spanned by the first two basis vectors (i.e. by setting any values to x_1 and x_2 while setting all remaining variables equal to zero) and Q by the remaining $n - 2$ basis vectors: $k^n = P \oplus Q = k^2 \oplus k^{n-2}$. Change the first two variables if needed to arrange that $q = x_1 x_2$ on P, as we saw is possible. By induction, we can arrange that on Q, either

$$q(0, 0, x_3, \ldots, x_n) = x_3 x_4 + x_5 x_6 + \cdots + x_{n-1} x_n$$

or

$$q(0, 0, x_3, \ldots, x_n) = x_3 x_4 + x_5 x_6 + \cdots + x_{n-2} x_{n-1} + x_n^2.$$

and so

$$q(x_1, \ldots, x_n) = x_1 x_2 + x_3 x_4 + x_5 x_6 + \cdots + x_{n-1} x_n$$

or

$$q(x_1, \ldots, x_n) = x_1 x_2 + x_3 x_4 + x_5 x_6 + \cdots + x_{n-2} x_{n-1} + x_n^2.$$

Now suppose that there is no plane P in k^n on which we can arrange that q has the form $x_1 x_2$. Suppose that this remains true even after taking finitely many quadratic extensions. So on any plane P in k^n, either $q = 0$ or we can arrange that $q = x_1^2$.

Let U be the set of all vectors in k^n on which q vanishes. Take two linearly independent vectors on which q vanishes, i.e. two points of U. On the plane they

span, q vanishes on at least two lines, so q vanishes everywhere on that plane. So U is a linear subspace of k^n. Arrange by linear change of variables that U is given by the linear equations $0 = x_{s+1} = \cdots = x_n$. If $s \geq 2$, then the plane parameterized by $(x_1, x_2, 0, \ldots, 0)$ intersects U only at the origin, i.e. $q = 0$ on that plane only at the origin. After a quadratic extension, as we have seen, we can arrange that $q = x_1 x_2$, a contradiction. Hence $s = 0$ or $s = 1$. But $s = 0$ just when $U = k^n$, i.e. just when $q = 0$ everywhere. So we can assume that $s = 1$, i.e. $0 = q(0, x_2, \ldots, x_n)$, i.e. $q = 0$ whenever $x_1 = 0$, i.e. q has a factor of x_1, and so splits into a product of linear factors. $\square$

Plane conics

> Look at the curves $x^2 = y^2$ and the curves $x^2 + y^2 = 0$, defined over $k = \mathbb{R}$; each is a singular plane conic. The first is a pair of lines $x = \pm y$ while the second, a circle of radius zero, has the origin as its only real point, but splits into $x = \pm iy$ over $\bar{k} = \mathbb{C}$.

Lemma 27.4. *If a plane conic over a field k has a singular point with coordinates in k then it is either*

   a. *a pair of lines defined over k or else*

   b. *a pair of lines defined over a quadratic extension $k(\sqrt{\alpha})$ meeting at a single point defined over k, the singular point, and no other point of the curve is defined over k.*

Proof. At a singular point of a plane conic, in an affine chart in which the singular point is the origin, the equation of the conic becomes a quadratic form in x, y. This factors over k or over a quadratic extension of k, i.e. the conic is a pair of lines in a quadratic extension. If any point (except the origin) on one of those lines has coordinates in k, then the line does, and so does the associated linear factor, and then by dividing that linear factor in, we see that the other linear factor is also defined over k. Suppose on the other hand that the linear factors don't exist over k, only over the quadratic extension. Then no other point of the affine plane with coordinates in k belongs to the curve, except the singular point. Write out the equation of the curve as $ax^2 + bxy + cy^2 = 0$, say. Switch to the affine chart $y = 1$ where the curve is $ax^2 + bx + c$. Since the homogeneous equation has no roots in $\mathbb{P}^1(k)$, this quadratic equation has no root in k, so there is no point in this affine chart $y = 1$. Therefore the projective conic has precisely one point with coordinates in k. $\square$

Proposition 27.5. *If a conic in the projective plane over a field k has a regular point with coordinates in k, it is either a pair of distinct lines, and taken to the conic $xy = 0$ by some projective automorphism, or is irreducible and taken to the conic $y = x^2$ by some projective automorphism.*

Proof. Suppose that our conic is not a pair of lines. It has a regular point with coordinates in k, and is not a pair of lines, so it has no singular points, and so is

irreducible. Arrange by projective automorphism that the regular point lies at the origin of the affine chart $z = 1$ with tangent line $y = 0$, so the curve has the form

$$y = ax^2 + bxy + cy^2.$$

If $a = 0$, we get a factor of y on both sides, so the equation is reducible. Rescale y to arrange $a = 1$, and write the equation as $x^2 = y\,(1 - bx - cy)$. Change variables by a projective automorphism $X = x, Y = y, Z = z - bx - cy$ to get to $x^2 = y$. $\qquad\square$

Corollary 27.6. *If an irreducible conic in the projective plane over a field k has a regular point with coordinates in k then it is birational to a line.*

Proof. Suppose that the conic, call it C, has equation $x^2 = y$, i.e. $x^2 = yz$ homogeneously. Map

$$[s, t] \in \mathbb{P}^1(k) \mapsto \left[st, s^2, t^2\right] \in C \subset \mathbb{P}^2(k).$$

The inverse map is

$$[x, y, z] \in C \mapsto \begin{cases} [x, y], & \text{if } y \neq 0, \\ [z, y], & \text{if } z \neq 0. \end{cases}$$

Clearly this identifies the rational functions.

More geometrically, pick a regular point p_0 on the curve, and arrange that it is the origin in an affine chart. Then take a line ℓ in the plane, and identify each point $p \in \ell$ of that line with the intersection point of the line pp_0 with C. We leave the reader to check that this gives the same map. $\qquad\square$

Chapter 28

Projective duality

There does not exist a category of science to which one can give the name applied science. There are sciences and the applications of science, bound together as the fruit of the tree which bears it.

— Louis Pasteur
Revue Scientifique (1871)

The set of lines in the projective plane $P = \mathbb{P}^2(k)$ is called the *dual plane* and denoted $P^* = \mathbb{P}^{2*}(k)$. Each line has a homogeneous linear equation $0 = ax + by + cz$, unique up to rescaling, and so the point $[a, b, c]$ is determined, allowing us to identify P^* with $\mathbb{P}^2(k)$ itself. The points of P^* correspond to lines in P. Through each point of P, there is a pencil of lines; for example if the point is (x_0, y_0) in the affine chart $z = 1$, then the lines are $a(x - x_0) + b(y - y0) = 0$, or expanding out: $ax + by + c = 0$ where $c = -ax_0 - by_0$. Hence in P^* the pencil is the line consisting of those points $[a, b, c]$ so that $ax_0 + y_0 + c = 0$. So each point p of P is identified with a line p^* in P^*, consisting of those lines in P passing through p. Each line λ of P is identified with a point λ^* in P^*. Given two points p, q of P and the line pq through them, the corresponding lines p^* and q^* in P^* intersect at the point corresponding to $(pq)^*$. We can drop all of this notation and just say that each point p of P is a line in P^*, and vice versa.

Dual curves

Take a projective algebraic curve C in $P = \mathbb{P}^2(k)$ over a field k with algebraic closure $\bar{k}$, with equation $0 = p(x, y)$ in some affine chart.

28.1 Denote partial derivatives as $p_x := \frac{\partial p}{\partial x}$, etc. Show that at any point (x_0, y_0) of C at which $p_x \neq 0$ or $p_y \neq 0$, the equation of the tangent line is

$$0 = p_x(x_0, y_0)(x - x_0) + p_y(x_0, y_0)(y - y_0).$$

Lemma 28.1. *In homogeneous coordinates $[x, y, z]$ on $P = \mathbb{P}^2$, if the equation of a curve C is $0 = p(x, y, z)$ with $p(x, y, z)$ homogeneous, then the tangent line at a point $[x_0, y_0, z_0]$ of C has homogeneous equation*

$$0 = xp_x(x_0, y_0, z_0) + yp_y(x_0, y_0, z_0) + zp_z(x_0, y_0, z_0).$$

Proof. Differentiate to get tangent line

$$0 = p_x(x_0, y_0, z_0)(x - x_0) + p_y(x_0, y_0, z_0)(y - y_0) + p_z(x_0, y_0, z_0)(z - z_0).$$

Take the equation that says that $p(x, y, z)$ is homogeneous of degree d:

$$p_x(\lambda x, \lambda y, \lambda z) = \lambda^d p(x, y, z),$$

233

and differentiate it in λ to get

$$xp_x(\lambda x, \lambda y, \lambda z) + yp_y(\lambda x, \lambda y, \lambda z) + zp_z(\lambda x, \lambda y, \lambda z) = d\lambda^{d-1}p(x, y, z).$$

Now set $\lambda = 1$ to get

$$xp_x + yp_y + zp_z = dp.$$

Plug into the homogeneous equation for the tangent line to get

$$0 = xp_x(x_0, y_0, z_0) + yp_y(x_0, y_0, z_0) + zp_z(x_0, y_0, z_0) - dp(x_0, y_0, z_0).$$

But at $[x_0, y_0, z_0]$ we know that $p(x_0, y_0, z_0) = 0$. $\qquad\qquad\square$

Each line $[a, b, c]$ in P^* has homogeneous equation $ax + by + cz = 0$, so the equation of the tangent line at each point $[x_0, y_0, z_0]$ of a curve with homogeneous equation $0 = p(x, y, z)$ is

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}.$$

The *dual curve* of C is the smallest algebraic curve C^* in P^* containing the tangent lines of C with coordinates in $\bar{k}$.

Corollary 28.2. *Every curve has a dual curve, i.e. the map*

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}$$

taking the smooth points of C to P^* has image inside an algebraic curve.

Proof. Assume that the field is algebraically closed. Suppose that the image contains finitely many points. The original curve has finitely many tangent lines, so is a finite union of lines. Following Study's theorem (theorem 22.3 on page 187) we can assume that C is irreducible, say $C = (p(x, y, z) = 0)$. Take the four equations $p = 0$, $a = p_x$, $b = p_y$, $c = p_z$, rescaling out any one of the x, y, z variables by homogeneity (or some linear combination of them), and compute resultants for them in one of the x, y, z variables, treating them as having coefficients rational in the other two x, y, z variables and in a, b, c. Repeat until we get rid of x, y, z entirely. Since we have polynomial expressions to begin with in all of our variables, each resultant is also polynomial in those variables. By corollary 18.5 on page 157, perhaps making a linear change of x, y, z variables before taking our first resultant, the resultants we end up are not all zero polynomials. So the image of the duality map lies in an algebraic curve in the dual plane. But the resultant vanishes just precisely on the points in the image of the duality map, since these equations specify precisely those points $[a, b, c]$. $\qquad\square$

For any irreducible conic C, given by an equation $0 = \langle Ax, x \rangle$ for a symmetric matrix A, with homogeneous coordinates

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}$$

and dual plane homogeneous coordinates

$$a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

differentiate to get that the partial derivatives of $q(x) = \langle Ax, x \rangle$ are $a = q'(x) = 2Ax$. Solve for x: $x = 2A^{-1}a$. Careful to note that we cannot solve for x over a field of a characteristic 2. Plug in to the equation of the conic: $0 = \langle a, A^{-1}a \rangle$. So, for any conic over any field not of characteristic 2, the dual conic has the inverse matrix.

Each singular point of C is tangent to all lines through it, so by definition that collection of lines belongs to the dual curve. But that collection of lines forms a line in the dual plane. So the dual curve is a union of irreducible algebraic curves, finitely many of which are lines, precisely one line for each singular point of C.

Proposition 28.3. *Suppose that C is an irreducible plane algebraic curve and the degree of C is denoted d. The dual curve has degree $d(d-1)$, containing no lines if C is not singular, and containing a line for each singular point of C in the closure of the field of definition, with multiplicity of that line given by the multiplicity of the singularity.*

Proof. We can assume the field is algebraically closed. The degree of C^* is the number of intersections of C^* with the generic line L in the plane. Each line in the dual plane is the dual $L = p^*$ of a point p in the original plane, i.e. the set of points in the dual plane corresponding to lines through the point p of the plane. The intersection of L^* with C^* occurs at the points of C^* corresponding to the lines tangent to C passing through p. There are $d(d-1)$ tangent lines through p, counting multiplicity, by proposition 25.1 on page 209. Among these lines are all lines connecting p to any singular point of C, corresponding to lines in C^*. □

Chapter 29

Polynomial equations have solutions

If the followers of the prophets err in their proofs, have no fear; the foundation of their knowledge is the authority of the prophets. They provide proofs only for added strength. Following the authority of the prophets is sufficient for them, unlike the philosophers. The philosophers stray from the authority of the prophets and instead rely on their proofs to establish an argument. Misguided, they misguide.

— Ahmad al-Fārūqī al-Sirhindī
Maktubat

To argue with a man who has renounced the use and authority of reason, and whose philosophy consists in holding humanity in contempt, is like administering medicine to the dead.

— Thomas Paine
The Crisis V

Affine space

For any field k, *affine space k^n* is the set of all points

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

with coordinates $c_1, c_2, \ldots, c_n$ in k.

We want to say: if you have infinitely many polynomial equations, in a finite set of variables, you can throw away all but finitely many without changing the solutions. Let us be more precise. Recall that an *ideal I* in $k[x]$ is a collection of polynomials so that we can add and multiply polynomials in I without leaving I, and I is *sticky*: if $f \in I$ and $g \in k[x]$ then $fg \in I$. Any set S of polynomials generates an ideal (S): the ideal is the set of all expressions

$$g_1 f_1 + g_2 f_2 + \cdots + g_s f_s$$

where $f_1, f_2, \ldots, f_s$ is any finite collection of polynomials in S and $g_1, g_2, \ldots, g_s$ are any polynomials at all in $k[x]$. Given any set of polynomials, we can replace that set

by the ideal of polynomials it generates inside $k[x]$ without changing the roots of the polynomials.

Theorem 29.1 (Hilbert basis theorem). *For any field k and any variables $x = (x_1, x_2, \ldots, x_n)$, every ideal I in $k[x]$ is finitely generated. In other words, there are finitely many polynomials*

$$f_1(x), f_2(x), \ldots, f_s(x)$$

in I so that the polynomials in I are precisely those of the form

$$g_1(x)f_1(x) + g_2(x)f_2(x) + \cdots + g_s(x)f_s(x),$$

for any polynomials $g_1(x), g_2(x), \ldots, g_s(x)$ in $k[x]$.

Proof. The result is obvious if $n = 0$ since the only ideals in k are k and 0. By induction, suppose that the result is true for any number of variables less than or equal to n. Take an ideal $I \subset k[x, y]$ where $x = (x_1, x_2, \ldots, x_n)$ and y is a variable. Write each element $f \in I$ as

$$f(x, y) = f_0(x) + f_1(x)y + \cdots + f_d(x)y^d,$$

so that $f_d(x)$ is the *leading coefficient*. Let J be the set of all leading coefficients of elements of I. Clearly J is an ideal in $k[x]$, so finitely generated by induction. So we can pick polynomials $f_1(x, y), f_2(x, y), \ldots, f_s(x, y)$ in I whose leading coefficients generate the ideal J of all leading coefficients.

Let N be the maximum degree in y of any of the $f_j(x, y)$. Take some polynomial $p(x, y)$ from I of degree greater than N in y, say with leading coefficient $\check{p}(x)$. Express that leading coefficient in terms of the leading coefficients $\check{f}_j(x)$ of the polynomials $f_j(x, y)$ as

$$\check{p}(x) = \sum p_j(x)\check{f}_j(x).$$

If we raise the y degree of $f_j(x, y)$ suitably to match the y degree of $p(x, y)$, say by multiplying by y^{k_j}, some positive integer k_j, then

$$\sum p_j(x)y^{k_j}f_j(x, y)$$

has leading coefficient the same as $p(x, y)$, so

$$p(x, y) = \sum y^{k_j}p_j(x)f_j(x, y) + \ldots$$

up to lower order in y. Inductively we can lower the order in y of p, replacing p by the $\ldots$, until we get to order in y less than N. At each step, we modify p by something in the ideal generated by the f_j. So every polynomial in I of degree N or less in y is in the ideal generated by the f_j. Add to our list of $f_j(x, y)$ some additional $f_j(x, y) \in I$ to span all of the polynomials in I of degree up to N in y. By induction we have generated all of the polynomials in I. $\qquad\square$

Theorem 29.2 (Nullstellensatz). *Given a collection of polynomial equations in variables*

$$x = (x_1, x_2, \ldots, x_n).$$

over a field k, either there is a point $x = c$ defined in a finite algebraic extension field of k where all of these polynomials vanish or else the ideal generated by the polynomials is all of $k[x]$.

Proof. Follows from theorem 29.1 on the facing page and corollary 18.6 on page 157. $\square$

29.1 Given a point $c \in k^n$, prove that the ideal

$$I_c := \{ f(x) \mid f(c) = 0 \}$$

of functions vanishing at $x = c$ is a maximal ideal. Prove that if $x = p$ and $x = q$ are distinct points of k^n, then $I_p \neq I_q$.

Corollary 29.3. *Take variables*

$$x = (x_1, x_2, \ldots, x_n) .$$

over an algebraically closed field k. Every maximal ideal in $k[x]$ is the set of all polynomial functions vanishing at some point c, for a unique point c.

Proof. Take any ideal $I \subset k[x]$ Take a point c so that all polynomials in I vanish there, using theorem 29.2 on the facing page. Then $I \subset I_c$. $\square$

Closed sets

An *affine variety* $X = X_S$ is the collection of zeroes of a set S of polynomials in variables

$$x = (x_1, x_2, \ldots, x_n) .$$

A *closed set* in affine space k^n (over a field k) is just another name for an affine variety.

> The entire set k^n is closed: it is the zero locus of the polynomial function $p(x) = 0$.

> Every finite set in k is closed: the set $\{ c_1, c_2, \ldots, c_n \}$ is the zero locus of
> $$p(x) = (x - c_1)(x - c_2) \ldots (x - c_n) .$$

> Any infinite subset of k, except perhaps k itself, is *not* closed. We know that every ideal I in $k[x]$ is generated by a single polynomial (the greatest common divisor of the polynomials in I), say $I = p(x)$, so the associated closed set is the finite set of roots of $p(x)$.

> A subset of k^2 is closed just when it is either (a) all of k^2 or (b) a finite union of algebraic curves and points.

An *open set* is the complement of a closed set. An open set containing a point c of k^n is a *neighborhood* of c. The intersection of any collection of closed sets is a closed set: just put together all of the equations of each set into one set of equations. In particular, the intersection of all closed sets containing some collection X of points

of k^n is a closed set, the *closure* of X. A set S is dense in a closed set X if X is the closure of S.

If the set X in k^n is the zero locus of some polynomials $f_i(x)$ and the set Y in k^n is the zero locus of some polynomials $g_j(x)$, then $X \cup Y$ in k^n is the zero locus of the polynomials $f_i(x)g_j(x)$. Hence the union of finitely many closed sets is closed.

29.2 Give an example of a field k and a polynomial $b(x, y)$ over k so that the points in the image of the polynomial (i.e. the points of the form $z = b(x, y)$ for any x and y in k) do *not* form an affine subvariety of k.

朱漫屠於支益,
千金之家,
三年技成,
而所用其巧。

There once lived a man
who learned how to slay dragons
and gave all he possessed
to mastering the art.

After three years
he was fully prepared but,
alas, he found no opportunity
to practise his skills.

       — 庄周 (Zhuang Zhou)
         子 (ZHUANGZI), 列御寇 (LIE YU-KOU)

As a result he began
to teach how to slay dragons.

       — René Thom
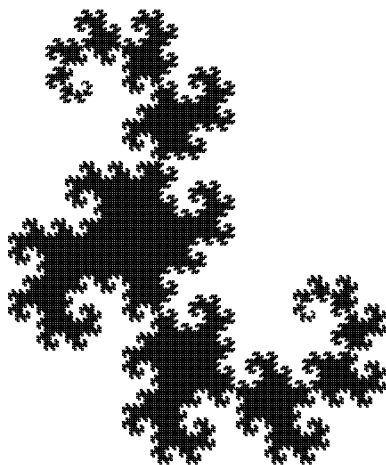
Hints

    a. distributive

    b. associative multiplication

    c. commutative multiplication

1.2. By the commutative law of multiplication, $(a+b)c = c(a+b)$. By the distributive law, $c(a + b) = ca + cb$. By two applications of the commutative law, $ca + cb = ac + cb = ac + bc$.

1.3. For any integer a: $a + 0 = a$. Pick a to be $a = 0$.

1.4.

$$
\begin{aligned}
0 &= 0 + 0 && \text{by problem 1.3,} \\
0 \cdot 0 &= 0 \cdot (0 + 0) && \text{by multiplying by } 0, \\
0 \cdot 0 &= 0 \cdot 0 + 0 \cdot 0 && \text{by the distributive law,} \\
\text{Let } b &= -(0 \cdot 0), \\
0 \cdot 0 + b &= (0 \cdot 0 + 0 \cdot 0) + b && \text{by adding } b \text{ to both sides,} \\
0 \cdot 0 + b &= 0 \cdot 0 + (0 \cdot 0 + b) && \text{by the associative law for addition,} \\
0 &= 0 \cdot 0 + 0 && \text{by the definition of } b, \\
0 &= 0 \cdot 0 && \text{by the definition of } 0.
\end{aligned}
$$

1.5.

$$
\begin{aligned}
0 &= 0 + 0 && \text{by problem 1.3,} \\
a \cdot 0 &= a \cdot (0 + 0) && \text{by multiplying by } a, \\
a \cdot 0 &= a \cdot 0 + a \cdot 0 && \text{by the distributive law,} \\
\text{Let } b &= -(a \cdot 0), \\
a \cdot 0 + b &= (a \cdot 0 + a \cdot 0) + b && \text{by adding } b \text{ to both sides,} \\
a \cdot 0 + b &= a \cdot 0 + (a \cdot 0 + b) && \text{by the associative law for addition,} \\
0 &= a \cdot 0 + 0 && \text{by the definition of } b, \\
0 &= a \cdot 0 && \text{by the definition of } 0.
\end{aligned}
$$

1.6. There is at least one such b, by the existence of negatives. Suppose that $a+b=0$ and that $a+c=0$.

$$
\begin{aligned}
(a+b)+c &= 0+c, \\
&= c+0 && \text{by commutativity of addition,} \\
&= c && \text{by the definition of } 0, \\
&= (a+b)+c && \text{by returning to the start again,} \\
&= a+(b+c) && \text{by associativity of addition,} \\
&= a+(c+b) && \text{by commutativity of addition,} \\
&= (a+c)+b && \text{by associativity of addition,} \\
&= 0+b && \text{by the definition of } c, \\
&= b+0 && \text{by commutativity of addition,} \\
&= b && \text{by the definition of } 0.
\end{aligned}
$$

Hence $b=c$.

1.7.

$$
\begin{aligned}
a+(-1)a &= a+a(-1) && \text{by commutativity of multiplication,} \\
&= a(1+(-1)) && \text{by the distributive law on the right hand side,} \\
&= a(0) && \text{by the definition of } -, \\
&= 0 && \text{by the definition of } 0.
\end{aligned}
$$

So $(-1)a$ fits the definition of $-a$. By problem 1.6 on page 3, $(-1)a = -a$.

1.8. By problem 1.7 on page 3, $(-1)(-1) = -(-1)$ is the unique integer which, added to -1, gives zero, and we know that this integer is 1.

1.12. We need to see that $(-b)+(-c)$ has the defining property of $-(b+c)$: adding to $b+c$ to give zero.

$$
\begin{aligned}
(b+c)+((-b)+(-c)) &= b+c+(-b)+(-c) && \text{by as above: parentheses not needed,} \\
&= b+(-b)+c+(-c) && \text{by commutativity of addition,} \\
&= 0+0 && \text{by existence of negatives,} \\
&= 0 && \text{by the definition of } 0.
\end{aligned}
$$

1.13. If $b=c+1=d+1$ for two integers c,d, then $c+1=d+1$ so $(c+1)+(-1) = (d+1)+(-1)$, where the existence of negatives ensures that there is a negative -1 of 1. So $c+(1+(-1)) = d+(1+(-1))$ by the associative law for addition. So $c+0 = d+0$, by the existence of negatives law. So $c=d$, by the identity law for addition.

1.19. Suppose that there are positive integers b,c with $b+c$ not positive. By the law of well ordering, there is a least possible choice of b, and, for that given b, a least possible choice of c. If $b=1$ then $b+c=1+c=c+1$ is positive by the succession law. If $b \neq 1$, then $b=d+1$ for some positive d by the succession law, and then $b+c=(d+1)+c=(d+c)+1$. This is not positive, so by the succession law, $d+c$ is not positive. But then d is smaller than b, so b is not the least possible choice.

1.19. Suppose that there are positive integers b, c with bc not positive. By the law of well ordering, there is a least possible choice of b, and, for that given b, a least possible choice of c. If $b = 1$ then $bc = c$ is positive. If $b \neq 1$, then $b = d + 1$ for some positive d by the succession law, and then $bc = (d + 1)c = dc + c$. This is not positive, so by problem 1.19 on page 3, dc is not positive. But then d is smaller than b, so b is not the least possible choice.

1.21. If $b, c > 0$ or if $b, c < 0$, we know that $bc > 0$. If $b > 0$ and $c < 0$, say $c = -d$ with $d > 0$, then $bc = b(-d)$. If we add $bd + b(-d)) = b(d + (-d)) = b0 = 0$. So $b(-d) = -(bd)$ is negative. Similarly, if $b < 0$ and $c > 0$, $bc = cb$ is negative.

1.23.

$$
\begin{aligned}
(b + c) - (a + c) &= b + c + (-(a + c)) && \text{by } d - e = d + (-e) \text{ as above,} \\
&= b + c + (-a) + (-c) && \text{by } -(d + e) = (-d) + (-e) \text{ as above,} \\
&= b + c + (-c) + (-a) && \text{by commutativity of addition,} \\
&= b + 0 + (-a) && \text{by existence of negatives,} \\
&= b + (-a) && \text{by definition of zero,} \\
&= b - a && \text{by } d + (-e) = d - e \text{ as above.}
\end{aligned}
$$

So $a + c < b + c$ just when $a < b$.

1.24. By distributivity of subtraction, proven in problem 1.16 on page 3, $bc - ac = (b - c)a$. Apply the solution of problem 1.21 on page 4.

1.27. Dividing a nonzero integer a by zero would mean find an integer c so that $a = 0 c$. But we have seen that $0 c = 0$, so $a = 0$, a contradiction. Dividing zero by zero would mean find an integer c so that $0 = 0 c$. But any integer c satisfies this equation, so there is no way to pick out one value c to be $0/0$.

1.31. Pick any nonzero number, say m, from the collection. By proposition 1.3 on page 5, m is not divisible by any integer larger than $|m|$. So all of the positive integer divisors are between 1 and $|m|$. Put these integers in a bag. Repeatedly throw out the largest integer in our bag which is not a common divisor of the collection. But 1 is a common divisor, so we can't throw them all out. By induction, eventually we get a greatest common divisor, unique because it is the greatest.

1.32.

   a. 17

   b. 1

   c. 1

   d. 12

   e. 4

   f. 49

   g. 11111.

1.33. By repeated addition, you can build any positive integer multiple of any integer from your collection. By subtracting an integer from your collection from itself, you can build zero. By repeated subtraction, you can build any negative integer multiple of any integer from your collection. So by adding up, you can build any linear combination $s_1 m_1 + s_2 m_2 + \cdots + s_n m_n$ of integer multiples $s_1, s_2, \ldots, s_n$ of

any elements $m_1, m_2, \ldots, m_n$ from the collection, with any integers $s_1, s_2, \ldots, s_n$. If all integers in our collection are divisible by some integer d, then clearly so are any integer multiples or finite sums of integer multiples. So the greatest common divisor divides all such sums. Hence the greatest common divisor is unchanged if we replace the collection by the collection of all such sums. So we can suppose that any such sum is already in our collection. Our collection contains some positive integer, because if m_1 is in our collection, then taking s_1 to be 1 if $m_1 > 0$ and -1 if $m_1 < 0$, $s_1 m_1$ is positive. By well ordering, there is a least positive element c in our collection. Take quotient and remainder of any element m in our collection by c: $m = qc + r$. The remainder $r = m - qc$ is also in the collection, not negative, but smaller than c, and therefore is zero. So c is a common divisor of every element in the collection. Since d is the greatest common divisor, $c \le d$. But c is in the collection, and d divides everyhing in the collection, so d divides c, so $d \le c$. Hence $c = d$, i.e. the greatest common divisor is the smallest positive integer we can build.

2.8. When we expand out $(1 + x)^n$, each term in the expansion arises from choosing either 1 or x from each factor $1 + x$, and multiplying out the choices to produce the term. So terms with x^k arise when we choose x from k of the factors $1 + x$, and 1 from the other $n - k$. So in $(1 + x)^{n+1}$, the x^k terms arise from choosing x from k of the first n factors $1 + x$, and choosing 1 from the last one, or from choosing x from $k - 1$ of the first n factors, and also choosing x from the last one.

2.9. For $n = 1, 2, 3$, we can check by hand:

n	$p_n(x)$	$p_n(1 + x)$
1	1	1
2	$1 + x$	$2 + x = \binom{2}{1}x^0 + \binom{2}{2}x^1$
3	$1 + x + x^2$	$3 + 3x + x^2 = \binom{3}{1}x^0 + \binom{3}{2}x^1 + \binom{3}{3}x^2$

From then on, we will need to use induction:

$$p_{n+1}(x) = p_n(x) + x^n,$$

so

$$p_{n+1}(1 + x) = p_n(1 + x) + (1 + x)^n$$
$$= \sum_{k=0}^{n-1} \binom{n}{k+1} x^k + \sum_{k=0}^{n} \binom{n}{k} x^k,$$
$$= \sum_{k=0}^{n-1} \left(\binom{n}{k+1} + \binom{n}{k} \right) x^k + x^n,$$
$$= \sum_{k=0}^{n-1} \binom{n+1}{k+1} x^k + x^n,$$
$$= \sum_{k=0}^{n} \binom{n+1}{k+1} x^k.$$

3.1.

a.

$$\begin{pmatrix} 1 & 0 & 2468 \\ 0 & 1 & 180 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 180 \\ 1 & -13 & 128 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -13 & 128 \\ -1 & 14 & 52 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 14 & 52 \\ 3 & -41 & 24 \end{pmatrix}$$

$$\begin{pmatrix} 3 & -41 & 24 \\ -7 & 96 & 4 \end{pmatrix}$$

$$\begin{pmatrix} -7 & 96 & 4 \\ 45 & -617 & 0 \end{pmatrix}$$

$$(-7)(2468) + (96)(180) = 4$$

b.

$$\begin{pmatrix} 1 & 0 & 79 \\ 0 & 1 & -22 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & -22 \\ 1 & 4 & -9 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 & -9 \\ -2 & -7 & -4 \end{pmatrix}$$

$$\begin{pmatrix} -2 & -7 & -4 \\ 5 & 18 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 5 & 18 & -1 \\ -22 & -79 & 0 \end{pmatrix}$$

$$(5)(79) + (18)(-22) = -1$$

c.

$$\begin{pmatrix} 1 & 0 & 45 \\ 0 & 1 & 16 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & 16 \\ 1 & -2 & 13 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -2 & 13 \\ -1 & 3 & 3 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 3 & 3 \\ 5 & -14 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 1 \\ -16 & 45 & 0 \end{pmatrix}$$

$$(5)(45) + (-14)(16) = 1$$

d.

$$\begin{pmatrix} 1 & 0 & -1000 \\ 0 & 1 & 2002 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & -1000 \\ 3 & 1 & -998 \end{pmatrix}$$

$$\begin{pmatrix} 3 & 1 & -998 \\ -2 & -1 & -2 \end{pmatrix}$$

$$\begin{pmatrix} -2 & -1 & -2 \\ 1001 & 500 & 0 \end{pmatrix}$$

$$(-2)(-1000) + (-1)(2002) = -2$$

4.2. $2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113$

4.4.

$$
\begin{aligned}
\gcd\{4864, 3458\} &= 2\gcd\{2432, 1729\}, \\
&= 2\gcd\{1216, 1729\}, \\
&= 2\gcd\{608, 1729\}, \\
&= 2\gcd\{304, 1729\}, \\
&= 2\gcd\{152, 1729\}, \\
&= 2\gcd\{76, 1729\}, \\
&= 2\gcd\{38, 1729\}, \\
&= 2\gcd\{19, 1729\}, \\
&= 2\gcd\{19, 1710\}, \\
&= 2\gcd\{19, 855\}, \\
&= 2\gcd\{19, 836\}, \\
&= 2\gcd\{19, 418\}, \\
&= 2\gcd\{19, 209\}, \\
&= 2\gcd\{19, 190\}, \\
&= 2\gcd\{19, 95\}, \\
&= 2\gcd\{19, 76\}, \\
&= 2\gcd\{19, 38\}, \\
&= 2\gcd\{19, 19\}, \\
&= 2\gcd\{19, 0\}, \\
&= 38
\end{aligned}
$$

4.5. Let d be the greatest common divisor of some collection of integers. So each of the integers in the collection is a multiple of d. If some integer d' divides into d, i.e. d is a multiple of d', then each of the integers in the collection is a multiple of d, so a multiple of d'. On the other hand, suppose that d' is an integer which divides into all integers in our collection. Let $g := \gcd\{d, d'\}$, so $d' = D'g$ and $d = Dg$. Clearly D, D' are coprime. Note: d' divides d just when D' divides D. Moreover, all integers in our collection are divisible by g. So we make a new collection of these integers, dividing each integer in the old collection by g. Replacing the old collection by the new, we can assume that d, d' are coprime. So d' divides all integers in our collection, but is coprime to d. Each integer in the collection has the form dm, so d' m, say $dm = dd'm'$. So dd' is a positive common divisor, larger than d if $|d'| > 1$, a contradiction, so $d' = \pm 1$ divides into d.

4.6. Let $d := \gcd\{b, c\}$, so $b = Bd$ and $c = Cd$. If B, C have a positive common divisor k, say $B = kB'$ and $C = kC'$, then $b = B'kd$ and $c = C'kd$ so kd divides b and c, but d is the greatest common divisor so $k = 1$. Therefore $\gcd\{B, C\} = 1$. Multiple by m: $mb = B(md)$ and $mc = C(md)$, so md divides mb and mc. By problem 4.5 on page 25, md divides D, say $D = md\ell$. Then $mb = md\ell b'$ and $mc = md\ell c'$ so $b = d\ell b' = Bd$ so $\ell b' = B$ and similarly $\ell c' = C$. But B and C are coprime, so $\ell = 1$.

5.2. A 1-digit positive integer has the form $b = b_0$ with $0 \le b_0 \le 9$ its first digit. A 2-digit positive integer has the form $b = b_0 + 10\, b_1$ with $0 \le b_0, b_1 \le 9$. In general, an

n-digit positive integer has the form

$$b = b_0 + 10\,b_1 + 10^2\,b_2 + \cdots + 10^{n-1}b_{n-1}.$$

Modulo 9, $10 \equiv 1 \pmod 9$, so modulo 9

$$b \equiv b_0 + 10\,b_1 + 10^2\,b_2 + \cdots + 10^{n-1}b_{n-1} \pmod 9,$$
$$\equiv b_0 + b_1 + b_2 + \cdots + b_{n-1} \pmod 9.$$

5.12.

a.

$$\begin{pmatrix} 1 & 0 & 13 \\ 0 & 1 & 59 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 13 \\ -4 & 1 & 7 \end{pmatrix}$$

$$\begin{pmatrix} -4 & 1 & 7 \\ 5 & -1 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 5 & -1 & 6 \\ -9 & 2 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -9 & 2 & 1 \\ 59 & -13 & 0 \end{pmatrix}$$

$$(-9)(13) + (2)(59) = 1$$

Answer: $13^{-1} = 50$ modulo 59

b.

$$\begin{pmatrix} 1 & 0 & 10 \\ 0 & 1 & 11 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 10 \\ -1 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 1 & 1 \\ 11 & -10 & 0 \end{pmatrix}$$

$$(-1)(10) + (1)(11) = 1$$

Answer: $10^{-1} = 10$ modulo 11

c.

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 193 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 2 \\ -96 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -96 & 1 & 1 \\ 193 & -2 & 0 \end{pmatrix}$$

$$(-96)(2) + (1)(193) = 1$$

Answer:$2^{-1} = 97$ modulo 193

d.

$$\begin{pmatrix} 1 & 0 & 6003722857 \\ 0 & 1 & 77695236973 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 6003722857 \\ -12 & 1 & 5650562689 \end{pmatrix}$$

$$\begin{pmatrix} -12 & 1 & 5650562689 \\ 13 & -1 & 353160168 \end{pmatrix}$$

$$\begin{pmatrix} 13 & -1 & 353160168 \\ -220 & 17 & 1 \end{pmatrix}$$

$$\begin{pmatrix} -220 & 17 & 1 \\ 77695236973 & -6003722857 & 0 \end{pmatrix}$$

$$(-220)(6003722857) + (17)(77695236973) = 1$$

Answer:$6003722857^{-1} = 77695236753$ modulo 77695236973

5.17. We want to find remainders $(1, 2, 3)$ modulo $(3, 7, 19)$. We let

$$u_1 = 7 \cdot 19 = 133,$$
$$u_2 = 3 \cdot 19 = 57,$$
$$u_3 = 3 \cdot 7 = 21.$$

Modulo $3, 7, 19$ these are

$$u_1 = 1 \quad \mathrm{mod}\ 3,$$
$$u_2 = 1 \quad \mathrm{mod}\ 7,$$
$$u_3 = 2 \quad \mathrm{mod}\ 19.$$

The multiplicative inverses of these, modulo $3, 7, 19$, are obvious by reducing and inspection:

$$v_1 = 1,$$
$$v_2 = 1,$$
$$v_3 = 10.$$

By the Chinese remainder theorem, the answer, up to multiples of $3 \cdot 7 \cdot 19 = 399$, is

$$r_1 u_1 v_1 + r_2 u_2 v_2 + r_3 u_3 v_3 = 1 \cdot 133 \cdot 1 + 2 \cdot 57 \cdot 1 + 3 \cdot 21 \cdot 10,$$
$$= 133 + 114 + 630,$$
$$= 877,$$
$$= 79 + 2 \cdot 399.$$

So there are at least 79 soldiers in Han Xin's army.

5.18. We have $r_1, r_2 = 2, 8$, $m_1, m_2 = 22, 39$, so $u_1 = 39$, $u_2 = 22$. Modulo 22:

$$v_1 = 39^{-1} = (22 + 17)^{-1} = 17^{-1}.$$

Bézout:

$$\begin{pmatrix} 1 & 0 & 17 \\ 0 & 1 & 22 \end{pmatrix} \quad \text{add } -\text{row 1 to row 2},$$

$$\begin{pmatrix} 1 & 0 & 17 \\ -1 & 1 & 5 \end{pmatrix} \quad \text{add } -3\cdot\text{row 2 to row 1},$$

$$\begin{pmatrix} 4 & -3 & 2 \\ -1 & 1 & 5 \end{pmatrix} \quad \text{add } -2\cdot\text{row 1 to row 2},$$

$$\begin{pmatrix} 4 & -3 & 2 \\ -9 & 7 & 1 \end{pmatrix}$$

So $(-9)(17) + (7)(22) = 1$, so mod 22, $v_1 = -9 = 22 - 9 = 13$. Modulo 39, $v_2 = 22^{-1}$,

$$\begin{pmatrix} 1 & 0 & 22 \\ 0 & 1 & 39 \end{pmatrix} \quad \text{add } -\text{row 1 to row 2},$$

$$\begin{pmatrix} 1 & 0 & 22 \\ -1 & 1 & 17 \end{pmatrix} \quad \text{add } -\text{row 2 to row 1},$$

$$\begin{pmatrix} 2 & -1 & 5 \\ -1 & 1 & 17 \end{pmatrix} \quad \text{add } -3\cdot\text{row 1 to row 2},$$

$$\begin{pmatrix} 2 & -1 & 5 \\ -7 & 4 & 2 \end{pmatrix} \quad \text{add } -2\cdot\text{row 2 to row 1},$$

$$\begin{pmatrix} 16 & -9 & 1 \\ -7 & 4 & 2 \end{pmatrix}$$

So $v_2 = 16$. Modulo $39 \cdot 22 = 858$,

$$x = r_1 u_1 v_1 + r_2 u_2 v_2,$$
$$= (2)(39)(13) + (8)(22)(16),$$
$$= 1014 + 2816,$$
$$= 3830,$$
$$= (4)(858) + 398,$$
$$= 398.$$

5.22. Factor $120 = 2^3 \, 3^1 \, 5^1$. So Euler's totient is $\phi(120) = (2^3 - 2^2)(3^1 - 3^0)(5^1 - 5^0) = 32$. Since 127 is prime, it is coprime to 120. The theorem of Euler's totient function says that $127^{\phi(120)} = 1$ modulo 120. So $127^{32} = 1$ modulo 120. In other words, every time you multiply together 32 copies of 127, modulo 120, it is as if you multiplied together no copies. Check that $162 = 5 \cdot 32 + 2$. So modulo 120, $127^{162} = 127^{5 \cdot 32 + 2} = 127^2$. This is not quite the answer. But we know that $127 = 7$ modulo 120. So modulo 120, $127^{162} = 7^2 = 49$.

7.1. $224/82 = 112/41$, $324/-72 = -9/2$, $-1000/8800 = -5/44$.

7.2. We can start by defining, for any pair of integers (b, c) with $c \neq 0$, and any pair of integers (B, C) with $C \neq 0$, a rational number

$$\frac{\beta}{\gamma} = (Cb + cB, cC).$$

This is well defined, since $cC \neq 0$. If we replace (b, c) by (ab, ac) and replace (B, C) by (AB, AC), then we get rational number

$$\frac{ACab + acAB}{acAC} = \frac{(Aa)(Cb + cB)}{(Aa)(Cc)} = \frac{Cb + cB}{Cc}$$

unchanged. So the resulting β/γ is independent of the choices of pairs, depending only on the rational numbers b/c and B/C. The other proofs are very similar.

7.3. By hand, showing your work, simplify

$$\frac{2}{3} - \frac{1}{2} = \frac{1}{6}, \, \frac{324}{49} \cdot \frac{392}{81} = 32, \, \frac{4}{5} + \frac{7}{4} = \frac{51}{20}.$$

7.5. If $\sqrt{3} = x + y\sqrt{2}$, square both sides to get

$$3 = x^2 + 2xy\sqrt{2} + 2y^2.$$

If $y \neq 0$, we can solve for $\sqrt{2}$ as a rational expression in x, y, so a rational number, a contradiction. Hence $y = 0$ and $\sqrt{3} = x$ is rational, a contradiction.

7.6. By definition, every positive rational number is a ratio b/c with b and c positive. We can divide out their greatest common divisor to arrange that b, c are coprime. Take the prime factorization of b, and that of c. As they are coprime, no prime occurs in both factorizations.

8.2. Take the polynomial $b(x) = (1/2)x(x + 1) = (1/2)x^2 + (1/2)x$. If x is an odd integer, then $x + 1$ is even, and vice versa, so $b(x)$ is an integer for any integer x.

8.5. Keep in mind that, modulo 3, $-1 = 2$, $-2 = 1$, $1^{-1} = 1$, $2^{-1} = 2$, so

$$\begin{pmatrix} 1 & 0 & x^3 + x \\ 0 & 1 & x^4 + 1 \end{pmatrix} \text{ add } -x(\text{row 1}) \text{ to row 2,}$$

$$\begin{pmatrix} 1 & 0 & x^3 + x \\ -x & 1 & -x^2 + 1 \end{pmatrix} \text{ add } x(\text{row 2}) \text{ to row 1,}$$

$$\begin{pmatrix} 1 - x^2 & x & 2x \\ -x & 1 & -x^2 + 1 \end{pmatrix} \text{ add } 2x(\text{row 1}) \text{ to row 2,}$$

$$\begin{pmatrix} 1 - x^2 & x & 2x \\ x^3 + x & 1 + 2x^2 & 1 \end{pmatrix}$$

So the Bézout coefficients are $x^3 + x, 1 + 2x^2$ and the gcd is 1. We can check: expand out

$$(x^3 + x)(x^3 + x) + (1 + 2x^2)(x^4 + 1) = (x^6 + 2x^4 + x^2) + (2x^6 + x^4 + 2x^2 + 1) = 1.$$

8.12. roots: 0, 1, 3, 4; factorizations: $(x + 2)(x + 3) = x(x + 5)$.

9.3. To get you started: $p_0(x) = 4x^4 + 2x^2 - 1$, $p'(x) = 16x^3 + 4x = 4(4x^3 + x)$, so we can rescale by $1/4$ to simplify to $p_1(x) = 4x^3 + x$, $p_2(x) = 1 - x^2$, $p_3(x) = -5x$ which you can rescale by $1/5$ to get $p_3(x) = -x$, $p_4(x) = -1$. You should find that there is one root in that interval. You can check by letting $x^2 = t$, write as a quadratic equation in t, solve with the quadratic formula, and also look at the graph:
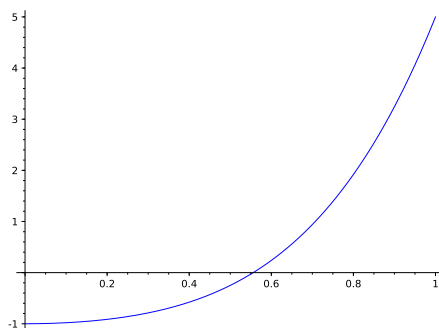
9.4. To get you started: $p_0(x) = x^4 + 4x^3 - 1$, $p'(x) = 4x^3 + 12x^2 = 4(x^3 + 3x^2)$, so we can rescale by $1/4$ to simplify to $p_1(x) = x^3 + 3x^2$, $p_2(x) = 3x^2 + 1$, $p_3(x) = \frac{4}{3}x + 4$, $p_4(x) = -28$. There is one root in that interval:
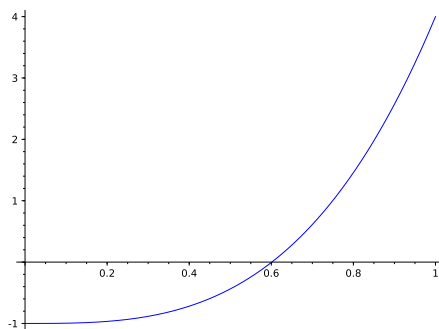
9.5. The sequence is $x^6 + 4x^3 - 2, 6x^5 + 12x^2, -2x^3 + 2, -18x^2, -2$. These have values at $x = 0$ of $-2, 0, 2, -2$, and at $x = 1$ they have $3, 18, -18, -2$. There is one root in that interval:
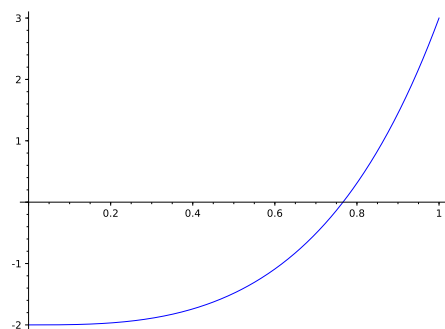
9.6. The sequence is $2x^4 + 4x^3 + 2x^2 - 1, 8x^3 + 12x^2 + 4x, (1/2)x^2 + (1/2)x + 1, 16x + 8, -7/8$. These have values at $x = -2$ of $7, -24, 2, -24, -7/8$, and at $x = 1$ they have $7, 24, 2, 24, -7/8$. There are two roots in that interval:
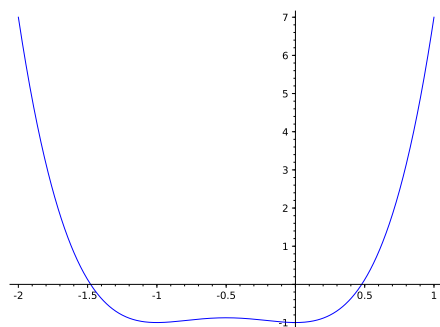
9.7. Each pattern of signs in the Sturm sequence, followed by a picture of parabolas with that pattern, indicating whether the point where we calculate the Sturm sequence gives a positive or negative value of the quadratic function, whether the point is to the left or right of the vertex of the parabola, and whether the discriminant is positive

or negative:
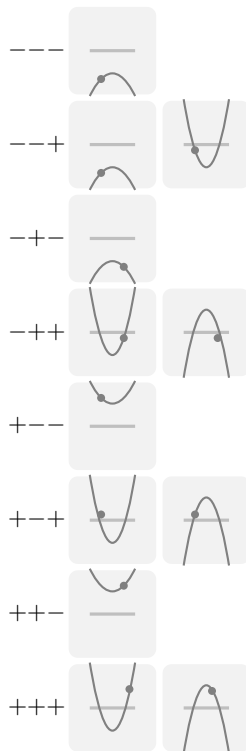
9.8. If $p(x) = 0$ then

$$a_n x^n + \cdots + a_0 = 0.$$

Subtract off all but the first term,

$$a_n x^n = -a_{n-1} x^{n-1} - \cdots - a_0.$$

Divide by a_n:

$$x^n = -\frac{a_{n-1} x^{n-1}}{a_n} - \cdots - \frac{a_0}{a_n}.$$

Divide by x^{n-1}:

$$x = -\frac{a_{n-1}}{a_n} - \frac{a_{n-2}}{a_n x} - \cdots - \frac{a_0}{a_n x^{n-1}}.$$

So

$$|x| = \left| -\frac{a_{n-1}}{a_n} - \frac{a_{n-2}}{a_n x} - \cdots - \frac{a_0}{a_n x^{n-1}} \right|,$$

$$\leq \left| \frac{a_{n-1}}{a_n} \right| + \left| \frac{a_{n-2}}{a_n x} \right| + \cdots + \left| \frac{a_0}{a_n x^{n-1}} \right|.$$

So if $|x| > 1$,

$$|x| \leq \left| \frac{a_{n-1}}{a_n} \right| + \left| \frac{a_{n-2}}{a_n} \right| + \cdots + \left| \frac{a_0}{a_n} \right|.$$

Hence every root lies in either $|x| \le 1$ or this interval, i.e. we can take c to be either $c = 1$ or

$$c = \frac{|a_{n-1}| + |a_{n-2}| + \cdots + |a_0|}{|a_n|},$$

whichever is larger.

9.9.
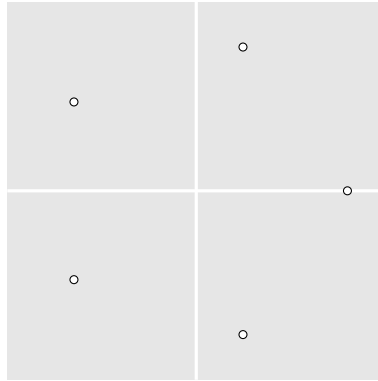
10.1. We can flip the sign to get $p(x)$ to have a positive leading coefficient. After x is made larger than any of the (finitely many) roots of $p(x)$, we know that $p(x) > 0$. Pick any integer N larger than any of the roots of $p(x)$. In particular, $p(N) > 0$. For any integer ℓ, if we set $x = N + \ell p(N)$, then expanding out $x^k = N^k + \ldots$, the $\ldots$ terms all contain a factor of $p(N)$, so $p(x) = p(N) + \ldots$, where again the $\ldots$ terms all contain a factor of $p(N)$. So $p(x) = 0$ modulo $p(N)$ for all of the infinitely many integers $x = N + \ell p(N)$.

10.6. Take $p(y) = y + 1$, clearly irreducible since linear and not constant. Note that $p(y)^2 = y^2 + 1$. So Eisenstein applies: $a(x, y)$ is irreducible.

10.7. By problem 2.9 on page 14,

$$c(x+1) = \binom{p}{1} + \binom{p}{2}x + \binom{p}{3}x^2 + \cdots + \binom{p}{p-1}x^{p-2} + \binom{p}{p}x^{p-1},$$

$$= p + \frac{p!}{2!(p-2)!}x + \frac{p!}{3!(p-3)!}x^2 + \cdots + \frac{p!}{(p-1)!1!}x^{p-2} + x^{p-1}.$$

Clearly there is a factor of p in each term, and no p^2 factor in the constant term.

10.8. Over the complex numbers,

$$b(x, y) = (x + iy)(x - iy).$$

Over all of the other fields, if reducible, it splits into linear factors

$$b(x, y) = (\alpha x + \beta y)(\gamma x + \delta y) = \alpha\gamma x^2 + (\alpha\delta + \beta\gamma)xy + \beta\delta y.$$

None of the coefficients can vanish, since $b(x, y)$ doesn't have a factor of x or y. For real coefficients, $b(x, y)$ vanishes along a line $\alpha x + \beta y = 0$. This is not possible because squares x^2, y^2 can't be negative, so if $x^2 + y^2 = 0$ then $x = y = 0$. Alternate proof: plug in $y = 1$ to get $x^2 + 1 = (\alpha x + \beta)(\gamma x + \delta$, vanishing at $x = -\beta/\alpha$, a square root

of -1. It follows that $b(x, y)$ is irreducible over the rationals, because rationals are real. The coefficients of $b(x, y)$ are coprime, so $b(x, y)$ is irreducible over the integers.

10.9. Eisenstein with $p = 3$, $111 = (3)(37)$, and 37 is not divisible by 3, so $q(x)$ is irreducible.

10.9. This $b(x, y)$ is irreducible, by Eisenstein's criterion, with with $p(y) = y^2 + y + 1$. We need to know that $p(y)$ is irreducible, and to see this: any reducible quadratic, having degree 2, splits into irreducibles of degrees 1 each, so linear. The only linear polynomials are the two linears $y, y + 1$. Their 3 possible products are $y^2, y(y+1), (y + 1)^2 = y^2 + 1$, not equal to $y^2 + y + 1$. So irreducible over remainders modulo 2. If it reduces over integers, quotient by 2 to get a reduction in remainders modulo 2, or to get zero. But it remains nonzero and irreducible.

10.13. Take two such, cross multiply, and use unique factorisation of polynomials.

12.1.

	0	1	2	α	2α	$1 + \alpha$	$2 + \alpha$	$1 + 2\alpha$	$2 + 2\alpha$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	α	2α	$1 + \alpha$	$2 + \alpha$	$1 + 2\alpha$	$2 + 2\alpha$
2	0	2	1	2α	α	$2 + 2\alpha$	$1 + 2\alpha$	$2 + \alpha$	$1 + \alpha$
α	0	α	2α	2α	α	0	α	2α	0
2α	0	2α	α	α	0	2α	2α	α	0
$1 + \alpha$	0	$1 + \alpha$	$2 + 2\alpha$	0	2α	$1 + \alpha$	$2 + 2\alpha$	$1 + \alpha$	$2 + 2\alpha$
$2 + \alpha$	0	$2 + \alpha$	$1 + 2\alpha$	α	2α	$2 + 2\alpha$	1	2	$1 + \alpha$
$1 + 2\alpha$	0	$1 + 2\alpha$	$2 + \alpha$	2α	α	$1 + \alpha$	2	1	$2 + 2\alpha$
$2 + 2\alpha$	0	$2 + 2\alpha$	$1 + \alpha$	0	0	$2 + 2\alpha$	$1 + \alpha$	$2 + 2\alpha$	$1 + \alpha$

12.2. $2x^8 + x$

13.2. Either note the common factor of $x+1$: $b(x) = (x+1)(x^2+x+1)$, $c(x) = (x+1)^2 x$, or compute the determinant of

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

by expanding across the first row:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

and note that two rows are the same so the determinant (and the resultant) is zero.

13.3. The resultant is

$$\mathrm{res}_{x^2+1, x^2+3x} = \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 1 & 0 & 1 & 3 \\ 0 & 1 & 0 & 1 \end{pmatrix} = 10.$$

But 10 only has 2 and 5 as prime factors, so this resultant vanishes just exactly when $p = 2$ or $p = 5$. If $p = 2$ the factors are $x^2 + 1 = (x+1)^2$ and $x^2 + 3x = x^2 + x = x(x+1)$, so a common factor of $x + 1$. If $p = 5$ the factors are $x^2 + 1 = (x - 2)(x - 3)$ and $x^2 + 3x = x(x - 2)$, so a common factor of $x - 2$.

13.4. The resultant is the determinant of a matrix whose every column has an even number of nonzero entries, so sum equal to zero. So the sum of the rows is zero, a linear relation among rows, hence zero determinant.

13.7.

$$b(x) = x^2\, c(x) + 2x + 1,$$

so

$$\begin{aligned}
\mathrm{res}_{b(x),c(x)} &= \mathrm{res}_{2x+1,c(x)}, \\
&= \mathrm{res}_{2x+1,x^6+4}, \\
&= 2^6\, \mathrm{res}_{x+\frac{1}{2},x^6+4}, \\
&= 2^6\, \left(x^6 + 4\right)\big|_{x=-\frac{1}{2}}, \\
&= 2^6\, \left(\frac{1}{2^6} + 4\right), \\
&= 1 + 2^6 \cdot 4, \\
&= 257.
\end{aligned}$$

13.8.

a. The derivative of $ax^2 + bx + c$ is $2ax + b$. The discriminant of $ax^2 + bx + c$ is

$$\det \begin{pmatrix} c & b & 0 \\ b & 2a & b \\ a & 0 & 2a \end{pmatrix} = -a(b^2 - 4ac)$$

which is surprisingly *not* the usual expression. So a quadratic polynomial has a common factor with its derivative just when $b^2 = 4ac$.

b. For $p(x) = x^3 + x^2$, $p'(x) = 3x^2 + 2x$, so

$$\Delta_p = \det \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 1 & 0 & 3 & 2 & 0 \\ 1 & 1 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix} = 0$$

due to the row of zeroes, or to the double root at $x = 0$.

c. For $p(x) = x^3 + x^2 + 1$, $p'(x) = 3x^2 + 2x$, so

$$\Delta_p = \det \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 1 & 0 & 3 & 2 & 0 \\ 1 & 1 & 0 & 3 & 2 \\ 0 & 1 & 0 & 0 & 3 \end{pmatrix} = 31$$

so there is no common factor between $p(x)$ and $p'(x)$, and so no double roots among the complex numbers.

d. For $p(x) = x^3 + 2x - 1$, $p'(x) = 3x^2 + 2$ so

$$\Delta_p = \det \begin{pmatrix} -1 & 0 & 2 & 0 & 0 \\ 2 & -1 & 0 & 2 & 0 \\ 0 & 2 & 3 & 0 & 2 \\ 3 & 0 & 0 & 3 & 0 \\ 0 & 3 & 0 & 0 & 3 \end{pmatrix} = 27$$

so there is no common factor between $p(x)$ and $p'(x)$, and so no double roots among the complex numbers.

13.10. Write out

$$p(x) = (x - x_1)(x - x_2) \ldots (x - x_n).$$

Differentiate:

$$p'(x) = (x - x_2)(x - x_3) \ldots (x - x_{n-1})(x - x_n)$$
$$+ (x - x_1)(x - x_3) \ldots (x - x_{n-1})(x - x_n)$$

$$+ \vdots$$
$$+ (x - x_1)(x - x_2)(x - x_3) \ldots (x - x_n)$$
$$+ (x - x_1)(x - x_2)(x - x_3) \ldots (x - x_{n-1})$$

Every term but the first has a factor of $x - x_1$, so if we plug in $x = x_1$:

$$p'(x_1) = (x_1 - x_2)(x_1 - x_3) \ldots (x_1 - x_{n-1})(x_1 - x_n) = \prod_{j \neq 1}(x_1 - x_j).$$

By the same reasoning, replacing x_1 by any other root x_i:

$$p'(x_i) = \prod_{j \neq i}(x_i - x_j).$$

So the discriminant is

$$\Delta_{p(x)} = \mathrm{res}_{p(x),p'(x)},$$
$$= (-1)^{n(n-1)}\,\mathrm{res}_{p'(x),p(x)},$$

Note that either n is even or $n - 1$ is even, so $(-1)^{n(n-1)} = 1$:

$$\Delta_{p(x)} = \mathrm{res}_{p'(x),p(x)},$$
$$= \prod_i p'(x_i),$$
$$= \prod_i \prod_{j \neq i}(x_i - x_j),$$
$$= \prod_{i \neq j}(x_i - x_j).$$

So now, for example, $x_1 - x_2$ occurs here, as does $x_2 - x_1$, so putting those together into one factor $-(x_1 - x_2)$:

$$\Delta_{p(x)} = \prod_{i < j}(-1)(x_i - x_j)^2.$$

There are $n(n-1)/2$ choices of $i < j$, so finally

$$\Delta_{p(x)} = (-1)^{n(n-1)/2} \prod_{i<j} (x_i - x_j)^2.$$

13.11. Write these equations as polynomials in increasing powers of t, with coefficients rational in x, y:

$$0 = -x + 0t + t^2,$$
$$0 = -y - t + 0t^2 + t^3.$$

The resultant is

$$\det \begin{pmatrix} -x & 0 & 0 & -y & 0 \\ 0 & -x & 0 & -1 & -y \\ 1 & 0 & -x & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

You can simplify this determinant by adding x(row 3) to row 1, and similar tricks, to compute it out: $y^2 - (x-1)x(x+1)$. So the equation of the curve is $y^2 = (x-1)x(x+1)$.

13.13. $x^2 + (1-x)y^2$

13.14. The equations, in powers of t, are

$$0 = x + t + t^3,$$
$$0 = 1 + yt.$$

The resultant is

$$\begin{vmatrix} x & 1 & 0 & 0 \\ 1 & y & 1 & 0 \\ 0 & 0 & y & 1 \\ 1 & 0 & 0 & y \end{vmatrix} = xy^3 + y^2 + 1.$$

So $p(x, y) = xy^3 + y^2 + 1$.

14.1.

  a.

$$x^3 y^3 z + x^3 yz^3 + xy^3 z^3 = (xyz)(x^2 y^2 + x^2 z^2 + y^2 z^2),$$
$$= e_3(x^2 y^2 + \dots),$$
$$= e_3(e_2^2 + \dots).$$

  But $e_2 = xy + xz + yz$, so

$$e_2^2 = x^2 y^2 + x^2 z^2 + y^2 z^2 + 2x^2 yz + 2xy^2 z + 2xyz^2,$$

  so

$$x^3 y^3 z + x^3 yz^3 + xy^3 z^3 = e_3(e_2^2 - [2x^2 yz + 2xy^2 z + 2xyz^2]),$$
$$= e_3(e_2^2 - 2e_3[x + y + z]),$$
$$= e_3(e_2^2 - 2e_3 e_1)$$

b. $4xyz^3 + 4xzy^3 + 4yzx^3 = 4e_3(e_1^2 - 2e_2)$

c. $x^4 y^4 z^4 = e_3^4$

d. $xyz + x^2 y^2 z^2 = e_3 (1 + e_3)$

14.2.

$$
\begin{aligned}
\chi_A \lambda &= (t_1 - \lambda)(t_2 - \lambda)\dots(t_n - \lambda) \\
&= (-1)^n P_t(\lambda) \\
&= e_n(t) - e_{n-1}(t)\lambda + e_{n-2}(t)\lambda^2 + \cdots + (-1)^n \lambda^n.
\end{aligned}
$$

14.4. Cramer's rule says that the inverse A^{-1} has entries

$$
A_{ij}^{-1} = \frac{(-1)^{i+j} \det(A \text{ with row } j \text{ and column } i \text{ deleted})}{\det A},
$$

so rational functions of the entries of A.

14.6. If there were two such, cross multiply and apply unique factorisation of polynomials (theorem 10.9 on page 81). To see that there is one such, write out the function as $p(x_1,\dots,x_n)/q(x_1,\dots,x_n)$. Apply a permutation of the variables, and the solution of problem 10.13 on page 82, to see that each permutation alters the numerator by a nonzero constant multiple, and the denominator by the same multiple. Swapping two variables gives some multiple, and repeating that swap gives the same multiple squared, but returns to the original order of the variables. So any swap either changes sign or leaves both numerator and denominator alone. Multiple the numerator and denominator by the function $\prod_{i<j}(t_i - t_j)$ to ensure that the don't change when we transpose, and so they don't change under any permutation. Apply theorem 14.5 on page 118.

15.3. If K is the kernel, an element r of R belongs to K just when $f(r) = 0$. So if r_0, r_1 belong to K, then $r_0 + r_1$ has $f(r_0 + r_1) = f(r_0) + f(r_1) = 0 + 0 = 0$, so $r_0 + r_1$ belongs to K; similarly for $r_0 - r_1$ and for $r_0 r_1$. Zero belongs to K because $f(0) = 0$. So K is a subring.

15.9. If $I \subset \mathbb{Q}$ is an ideal, containing some nonzero element $b \in \mathbb{Q}$, then I also contains $(1/b)b = 1$, and so, for any rational number c, I contains $c \cdot 1 = c$, i.e. $I = \mathbb{Q}$.

16.2. If $z = x - y$ then $f(x) = f(y + z) = f(y) + f(z)$ so $f(z) = f(x) - f(y)$.

16.3. The elementary symmetric polynomials of the roots are the coefficients of $p(x)$. All symmetric functions are polynomials in the elementary symmetric polynomials. A different proof: all of $\alpha_1, \dots, \alpha_n$ belong to the splitting field of $p(x)$ inside K. So we can assume without loss of generality that K is the splitting field of $p(x)$. The Galois group of K/k acts as permutations of $\alpha_1, \dots, \alpha_n$, fixing k, so fixes $q(\alpha_1, \dots, \alpha_n)$. But K is a Galois extension of k, so the fixed elements under the Galois group are precisely the elements of k.

17.6.

$$2 \int \left(t^2 - 1 \right)^2 \, dt = 2 \int \left(t^4 - 2t^2 + 1 \right) \, dt,$$

$$= 2 \left(\frac{t^5}{5} - \frac{2}{3} t^3 + t \right),$$

$$= 2 \left(\frac{1}{5} \left(\frac{y}{x} \right)^5 - \frac{2}{3} \left(\frac{y}{x} \right)^3 + \frac{y}{x} \right),$$

$$= 2 \left(\frac{1}{5} \frac{y^5}{x^5} - \frac{2}{3} \frac{y^3}{x^3} + \frac{y}{x} \right),$$

$$= \frac{2y}{15x^5} \left(3 \left(y^2 \right)^2 - 10 x^2 y^2 + 15 x^4 \right),$$

$$= \frac{2y}{15x^5} \left(3 \left(x^3 + x^2 \right)^2 - 10 x^2 (x^3 + x^2) + 15 x^4 \right),$$

$$= \frac{2y}{15x} \left(3x^2 - 4x + 8 \right),$$

$$= \frac{2\sqrt{x^2 + x^3}}{15x} \left(3x^2 - 4x + 8 \right).$$

17.7. The substitution

$$\cos \theta = (1 - t^2)/(1 + t^2),$$
$$\sin \theta = 2t/(1 + t^2),$$

rewrites the integral in terms of t. (In fact, this substitution is just the famous substitution $t = \tan(\theta/2)$ which you can find in the calculus textbooks.)

$$\int f(\cos \theta, \sin \theta) d\theta = \int \frac{f(\cos \theta, \sin \theta)}{-\sin \theta} (-\sin \theta) d\theta,$$

$$= \int \frac{f(\cos \theta, \sin \theta)}{-\sin \theta} d\cos \theta,$$

$$= \int \frac{f(x, y)}{-y} dx,$$

$$= \int f \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right) \frac{1 + t^2}{-2t} d(\frac{1 - t^2}{1 + t^2}),$$

$$= \int f \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right) \frac{1 + t^2}{-2t} \frac{(-4t)}{(1 + t^2)^2} dt,$$

$$= \int f \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right) \frac{2}{1 + t^2} dt.$$

Take a partial fraction decomposition as in the proof of corollary 9.8 on page 72.

19.2. Let $k_1 \subset k_2 \subset k_3 \ldots$ be finite fields of orders $2, 4, 8$ and so on. Our result is clear if $k = k_1 = \mathbb{Z}/2\mathbb{Z}$. By induction, suppose that our result is true for $k_1, k_2, \ldots, k_{n-1}$. All of the elements α of k_n that are not elements of k_{n-1} satisfy $\alpha^2 + \alpha + c = 0$ for

some c in k_{n-1}, as we saw above. By induction, $c = b^2$ for some element b of k_{n-1}. So $\alpha^2 + \alpha + b^2 = 0$, i.e. $\alpha^2 + b^2 = \alpha$. Expand out $(\alpha + b)^2$ to find $(\alpha + b)^2 = \alpha$.

20.2. Take a basis consisting of one element $\alpha \in K$. Then $1 \in K$ is somehow a multiple $1 = a\alpha$ for some $a \in k$ so $\alpha = 1/a \in k$. Any element β of K is $\beta = c\alpha$ for some $c \in k$ so $\beta \in k$.

20.3. Let n be the degree of $p(x)$. We write each element $b(x)$ in $k[x]$ as a polynomial, but if the polynomial has degree n or more, rewrite it as $b(x) = q(x)p(x) + r(x)$, quotient and remainder. Then every element of K is written as the remainder term, i.e. as a polynomial of degree at most $n - 1$. Hence the elements $1, x, \ldots, x^{n-1}$ in $k[x]$ map to a spanning set inside K. If not a basis, then there must be some linear relation between them, i.e. a lower degree polynomial in x vanishing in K, i.e. vanishing modulo $p(x)$. But then taking quotient and remainder, we see that this is not possible. Hence K has a basis over k consisting of the images of $1, x, x^2, \ldots, x^{n-1}$ in $k[x]$.

22.3. Homogenize to $x^3 + x^2y + xy^2 + y^3 + x^2z = z^3$, and then set $z = 0$ to get the homogeneous equation $x^3 + x^2y + xy^2 + y^3 = 0$. Factor: $(x + y)(x^2 + y^2) = 0$. So $x + y = 0$, i.e. $[x, -x, 0] = [1, -1, 0]$, or $x^2 + y^2 = 0$, which factors over any field k which contains an element i so that $i^2 = -1$, as $(x - iy)(x + iy) = 0$, i.e. $y = \pm ix$, so $[x, ix, 0] = [1, i, 0]$ and $[x, -ix, 0] = [1, -i, 0]$. Hence the points of our curve that lie on the line at infinity are $[1, -1, 0], [1, i, 0], [1, -i, 0]$.

22.4. Start with the affine plane. Remember (from problem 8.7 on page 57) that you can use the quadratic formula:

$$y = \frac{-1 \pm \sqrt{1 - 4(-x^3 - x - 1)}}{2}.$$

Simplify, using $2^{-1} = 3$ and $4 = -1$, to:

$$y = 3(-1 \pm \sqrt{-x^3 - x})$$

Note that $0^2 = 0, 1^2 = 1, 2^2 = 4, 3^2 = 9 = 4, 4^2 = 16 = 1$, so only $0, 1, 4$ have square roots.

- $x = 0$: $-x^3 - x = 0$, $y = -3 = 2$, $[x, y, z] = [0, 2, 1]$.
- $x = 1$: $-x^3 - x = 3$, no square root of 3, no solution.
- $x = 2$: $-x^3 - x = -8 - 2 = 0$, $y = -3 = 2$, $[x, y, z] = [2, 2, 1]$.
- $x = 3$: $-x^3 - x = -27 - 3 = 0$, $y = -3 = 2$, $[x, y, z] = [3, 2, 1]$.
- $x = 4$: $-x^3 - x = -68 = 2$, no square root of 2 no solution.

We still need the points "at infinity". Homogenize: $y^2z + yz^2 = x^3 + xz^2 + z^3$, and set $z = 0$: $0 = x^3$, so $x = 0$, i.e. $[x, y, z] = [0, y, 0]$, and rescale y to $[0, 1, 0]$. Final answer: $[0, 2, 1], [2, 2, 1], [3, 2, 1], [0, 1, 0]$.

29.2. $p(x, y) = (1 - xy)^2 + x^2$ over the field $k = \mathbb{R}$.

Bibliography

[1] Lars V. Ahlfors, *Complex analysis*, third ed., McGraw-Hill Book Co., New York, 1978, An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics. MR 510197

[2] C. Herbert Clemens, *A scrapbook of complex curve theory*, second ed., Graduate Studies in Mathematics, vol. 55, American Mathematical Society, Providence, RI, 2003. MR 1946768 (2003m:14001)

[3] David S. Dummit and Richard M. Foote, *Abstract algebra*, third ed., John Wiley & Sons, Inc., Hoboken, NJ, 2004. MR 2286236 (2007h:00003)

[4] Pavel Etingof, Oleg Golberg, Sebastian Hensel, Tiankai Liu, Alex Schwendner, Dmitry Vaintrob, and Elena Yudovina, *Introduction to representation theory*, Student Mathematical Library, vol. 59, American Mathematical Society, Providence, RI, 2011, With historical interludes by Slava Gerovitch. MR 2808160

[5] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994, A foundation for computer science. MR 1397498 (97d:68003)

[6] Kenneth Hoffman and Ray Kunze, *Linear algebra*, Second edition, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1971. MR 0276251

[7] Thomas W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York-Berlin, 1980, Reprint of the 1974 original. MR 600654 (82a:00006)

[8] E. L. Ince, *Ordinary Differential Equations*, Dover Publications, New York, 1944. MR 0010757

[9] Tristan Needham, *Visual complex analysis*, The Clarendon Press, Oxford University Press, New York, 1997. MR 1446490

[10] Gian-Carlo Rota, *Combinatorial theory and invariant theory*, Bowdoin College, Maine, U.S.A., 1971, Notes by L. Guibas.

[11] Igor R. Shafarevich, *Basic notions of algebra*, Encyclopaedia of Mathematical Sciences, vol. 11, Springer-Verlag, Berlin, 2005, Translated from the 1986 Russian original by Miles Reid, Reprint of the 1997 English translation [MR1634541], Algebra, I. MR 2164721

[12] Michael Spivak, *Calculus*, Calculus, Cambridge University Press, 2006. 50, 63, 70

[13] The Sage Developers, *Sagemath, the Sage Mathematics Software System (Version 7.5.1)*, 2018, http://www.sagemath.org.

[14] E. B. Vinberg, *A course in algebra*, Graduate Studies in Mathematics, vol. 56, American Mathematical Society, Providence, RI, 2003, Translated from the 2001 Russian original by Alexander Retakh. MR 1974508 (2004h:00001)

[15] Judy L. Walker, *Codes and curves*, Student Mathematical Library, vol. 7, American Mathematical Society, Providence, RI; Institute for Advanced Study (IAS), Princeton, NJ, 2000, IAS/Park City Mathematical Subseries. MR 1768485 (2001f:14046)

[16] André Weil, *Number theory for beginners*, Springer-Verlag, New York-Heidelberg, 1979, With the collaboration of Maxwell Rosenlicht. MR 532370 (80e:10004)

List of notation

Index