Polynomials and Fields:

A course at the 2015 Ross Mathematics Program

Lectures by Daniel Shapiro
Notes by George Hauser

Preface

Here are notes from a course at *Ross Mathematics Program* during the summer of 2015. George Hauser attended that class and took careful notes. This document is based on George's work.

The audience included people at several levels, from some students still in high school to others who had recently graduated from college as math majors. Consequently, the material in these lectures might jump from elementary to advanced at various points. The consensus was that the lectures were far too fast near the end of the course. Interest did increase for the final lecture (on $G - G$ fields) because of its connections with Prof. Bergelson's class.

Thanks to all the students who showed up faithfully on all those afternoons!

- D. Shapiro, September, 2015.

Contents

1 Fields, and the Degree of a Field Extension

Think deeply about simple things. - A.E. Ross

This topic could involve a lot of abstract algebra, and we don't want to just teach all the abstract algebra in the standard way (standard topics include groups, rings, fields, vector spaces, modules, etc.)

What is a field?

Definition. A *commutative ring* is a set of elements together with two binary operations, addition (written $+$), and multiplication (written $\times$ or $\cdot$), each of which is associative and commutative, along with special elements 0 and 1 that serve as identity elements for the addition and multiplication respectively. Additionally, every element a has an additive inverse, denoted $-a$. Last, addition is related to multiplication by a distributive law.

Examples. Here are a few (hopefully) familiar commutative rings:

- $\mathbb{Q}$, the rational numbers
- $\mathbb{R}$, the real numbers
- $\mathbb{Z}$, the integers
- $\mathbb{Z}[i]$, the Gaussian integers
- $\mathbb{Z}/m\mathbb{Z}$, the integers modulo m
- $\mathbb{Z}[X]$, the ring of polynomials with integer coefficients in the variable X
- $\mathbb{R}[X, Y]$, the ring of polynomials with real coefficients in two variables X and Y

Definition. An *integral domain* (or just a *domain*) is a commutative ring that has the *zero product property*: If $ab = 0$, then either $a = 0$ or $b = 0$.

We know that $\mathbb{Z}/6\mathbb{Z}$ is NOT a domain, because $2 \cdot 3 \equiv 0\ (6)$, but $2 \not\equiv 0\ (6)$ and $3 \not\equiv 0\ (6)$.

Definition. A *field* is an commutative ring in which every nonzero element has a multiplicative inverse: If $a \neq 0$, then there exists b such that $ab = 1$. Also, if a has such an inverse, we say a is a *unit*.

Certainly every field is a domain. (Why?) There are many domains that are not fields (like $\mathbb{Z}$ or $\mathbb{R}X$). But a finite domain is a field.

Why? Well, if $a \in D$ define the map $\mu_a : D \to D$ by $\mu_a(x) = ax$. The zero product property says: $\mu_a(b) = 0 \Rightarrow b = 0$. Deduce that μ_a is an injective map. If D is finite then μ_a must also be surjective. In particular, there exists $c \in D$ with $\mu_a(c) = 1$. Then $ac = 1$ and c is the

multiplicative inverse of a. So if D is a finite domain, every nonzero element is a unit: D is a field.

Finite Fields

If p is a prime number, then $\mathbb{Z}/p\mathbb{Z}$ is a finite field. (Why?) It is often denoted $\mathbb{F}_p$.

Are there any other finite fields? For instance, is there a finite field with 4 elements?

Claim A finite field of order q exists whenever q is a prime power.

The rest of this lecture might involve proving this.

First observation. Suppose a K is a finite field. Then K has a multiplicative identity 1. What happens when we add 1 to itself repeatedly? Eventually $1 + \cdots + 1 = 0$. (Why?) Let k be the smallest number such that $\underbrace{1 + \cdots + 1}_{k} = 0$. Then k must be prime (think about what would happen if k is composite ...), and let's call it p instead. This prime p is called the *characteristic* of the field K.

Then the subring generated by 1 is a copy of $\mathbb{F}_p$ embedded as a subfield of K. We may view K as a vector space over $\mathbb{F}_p$ (by forgetting most of the multiplication in K).

General facts about vector spaces: If V is a vector space over $\mathbb{F}_p$, then it has a basis $v_1, \ldots v_n$. Every element of V can be expressed uniquely as $c_1v_1 + \cdots + c_nv_n$ for scalars $c_j \in \mathbb{F}_p$. Therefore, if $\dim(V) = n$ then $\#(V) = p^n$. This is basic linear algebra!

Consequently, if K is a finite field of q elements then $q = p^n$ is a prime power, where p is the characteristic of K.

It is interesting to look for other proofs of this basic observation, perhaps using ring theory or group theory, rather than linear algebra.

Algebraic Extensions

Suppose $K \subseteq L$ are fields, one a subfield of the other. If $\theta \in L$ let $K[\theta]$ be the smallest subring of L that contains K and θ. Then $K[\theta]$ consists of all expressions of the form:

$$c_0 + c_1\theta + c_2\theta^2 + \cdots + c_n\theta^n \quad \text{with} \quad c_j \in K.$$

Those are just the values $f(\theta)$ for polynomials $f(X) \in K[X]$ (the polynomial ring in variable X over field K.) Moreover, the set of all such polynomial expressions is closed under the ring operations, and we deduce that every element of $K[\theta]$ has this polynomial form.

So we have an inclusion of rings:

$$K \subseteq K[\theta] \subseteq L.$$

For example, consider a square root of two, and get

$$\mathbb{Q} \subseteq \mathbb{Q}[\sqrt{2}] \subseteq \mathbb{R}.$$

Check that $\mathbb{Q}[\sqrt{2}]$ is a field, because you can always rationalize denominators. For instance, $1/(3+\sqrt{2}) = (3-\sqrt{2})/5)$. In general, we ask whether $K[\theta]$ always a field?

Definition $\theta \in L$ is said to be *algebraic* over K if there is a nonzero polynomial $f(X) \in K[X]$ such that $f(\theta) = 0$. Since the coefficients of f must be in K, this property depends on the base field K.

Remark: another way to say θ is algebraic over K is to say that the set of powers of θ, namely $\{1, \theta, \theta^2, \theta^3, \dots\}$, is linearly dependent over K (linear algebra strikes again).

Examples

- $\sqrt{2}$ is algebraic over $\mathbb{Q}$, since it is the root of $X^2 - 2$.
- $\sqrt{3}$ is algebraic over $\mathbb{Q}$, since it is a root of $X^2 - 3$.
- $\sqrt{2} + \sqrt{3}$ is algebraic over $\mathbb{Q}$ since it is a root of $X^4 - 10X^2 + 1$ (where did that come from?)

Note $\sqrt{2}$ is also a root of other polynomials like $X^4 + 8X^3 - X^2 - 16X - 1$. but it turns out that $X^2 - 2$ divides this polynomial. If $f(X)$ is a polynomial having $\sqrt{2}$ as a root, then must $X^2 - 2$ be a factor of $f(X)$? Of course not: let $f(X) = X - \sqrt{2}$. But if $f(X) \in \mathbb{Q}[X]$ and $f(\sqrt{2}) = 0$ then we can show that $(X^2 - 2) \mid f(X)$.

In a domain D there are two related notions for an element: *irreducible* and *prime*. An element c is *irreducible* if: Whenever $c = ab$ in D then a or b is a unit. An element $p \in D$ is prime if: whenever p divides a product ab, then it divides a or it divides b.

Every prime is irreducible. (Why?) Conversely, there can be some irreducible that is not prime? Can you think of an example in some domain D?

But in many domains, the notions of prime and irreducible do coincide. For instance, an element of $\mathbb{Z}$ is prime if and only if it is irreducible. Another such ring is $K[X]$.

The analogy between $\mathbb{Z}$ and $K[X]$ is a common theme in algebra. Another aspect of this analogy is between modular arithmetic in $\mathbb{Z}$ and in $K[X]$. When m is an integer, we can consider the ring $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m. Similarly, when $f(X)$ is a polynomial, we can consider the ring $K[X]/(f(X))$ of polynomials modulo $f(X)$. How far do these analogies extend? Do they extend to the ring $K[X, Y]$, polynomials in two variables?

Minimal Polynomial

Suppose θ is algebraic over K. That means there exists at least one nonzero polynomial $f(X) \in K[X]$ such that $f(\theta) = 0$. There are many others, since we can take one such $f(X)$ and multiply it by any polynomial to get another example. For given θ among all the $f \in K[X]$ with $f(\theta) = 0$, there is one of minimal degree (by WOP). Let $m(X)$ be such a minimal one. We may scale $m(X)$ to assume it is monic

If $g(X)$ is some polynomial in $K[X]$ that kills θ (i.e. with $g(\theta) = 0$), we claim that $m(X) \mid g(X)$. To see this, use polynomial division to write

$$g(X) = m(X)q(X) + r(X)$$

where q and r are in $K[X]$ and either r is identically 0 or $\deg r < \deg m$. Evaluate at θ and deduce that $r(\theta) = 0$. If r is not zero, then it is a polynomial that kills θ and has degree smaller than the degree of m. This contradiction implies that $r(X)$ is the zero polynomial, so that $m(X)|g(X)$, as claimed.

This property is an analogue to the notion of the order of an element in $U_m = (\mathbb{Z}/m\mathbb{Z})^*$. If $d = o(a)$ and n is a number such that $a^n = 1$, then d divides n.

That division result shows that $m(X)$ is unique. All polynomials in $K[X]$ that kill θ are multiples $m(X)$. We call this $m(X)$ the *minimal polynomial* of θ over K, and sometimes write it as $m_\theta(X)$, or $m_{\theta,K}(X)$.

The minimal polynomial $m_\theta(X)$ is irreducible in $K[X]$. For suppose $m_\theta(X) = a(X)b(X)$ in $K[X]$. Evaluate at θ to find that $a(\theta)b(\theta) = 0$. Since K is has no zero divisors, either $a(\theta) = 0$ or $b(\theta) = 0$. We may suppose the first case holds. Minimality of degree implies that $\deg a = \deg f$, and therefore $\deg b = 0$. Then b must be a constant polynomial, hence a unit. This shows that $m_\theta(X)$ is irreducible.

Exercise. Suppose $f \in K[X]$ is an irreducible monic polynomial that kills θ (that is, $f(\theta) = 0$). Then $f(X) = m_\theta(X)$.

Evaluation Homomorphism

We've used the phrase "evaluate at θ" a few times so far. What does this really mean? How do we know that evaluating a product (or sum) of polynomials at a given θ yields the same answer no matter which operations are done first? We can plug θ into each term and then multiply the results, or we can multiply the polynomials, and then plug in θ. Can you prove that those methods always yield the same result?
We'll assume that you can do this exercise and evaluation at θ really is well-defined.

Another way to say this is: Evaluation at θ is a *ring homomorphism*:

$$\text{Eval}_\theta : K[X] \to L \quad \text{given by} \quad f(X) \mapsto f(\theta)$$

This homomorphism fixes each element of the field K, since constant polynomials are, well, constant.

What is the image of this map? It is precisely $K[\theta]$.
What is its kernel? It is precisely $(m_\theta(X))$, the ideal generated by the minimal polynomial $m_\theta(x)$. An "Isomorphism Theorem" from algebra then implies:

$$K[\theta] \cong K[X]/(m_\theta(X)).$$

For example, let's compare $\mathbb{Q}[\sqrt{2}]$ with $\mathbb{Q}[X]/(X^2-2)$. Reducing $1, X, X^2, \ldots \pmod{X^2-2}$:

$$\begin{aligned} &1 \\ &X \\ &X^2 \equiv 2 \pmod{X^2-2} \\ &X^3 \equiv 2X \pmod{X^2-2} \\ &X^4 \equiv 4 \pmod{X^2-2} \end{aligned}$$

The elements of the factor ring are the equivalence class of polynomials mod X^2-2. (Just as we do in $\mathbb{Z}/n\mathbb{Z}$, whose elements are the equivalence classes of integers mod n).

Let α denote the class of X in $\mathbb{Q}[X]/(X^2-2)$. Then α is not really a number, but there is an injective homomorphism $\mathbb{Q}[\alpha] \longrightarrow \mathbb{R}$ sending $\alpha \mapsto \sqrt{2}$. Are there other homomorphisms $\mathbb{Q}[\alpha]$ to $\mathbb{R}$? Well since $\alpha^2 = 2$ there are two places to send α resulting in two embeddings:

$$\begin{aligned} \alpha &\mapsto \sqrt{2} \\ \alpha &\mapsto -\sqrt{2} \end{aligned}$$

So $\mathbb{Q}[\alpha]$ embeds into $\mathbb{R}$ in exactly two different ways.

That is, we have two isomorphisms

$$\mathbb{Q}(\sqrt{2}) \xleftarrow{\sim} \frac{\mathbb{Q}[X]}{(X^2-2)} \xrightarrow{\sim} \mathbb{Q}(-\sqrt{2}).$$

Combining those from left to right yields an isomorphism $\mathbb{Q}(\sqrt{2}) \xrightarrow{\sim} \mathbb{Q}(\sqrt{2})$ sending $\sqrt{2} \mapsto -\sqrt{2}$. This is the usual "bar" map sending $a + b\sqrt{2} \mapsto a - b\sqrt{2}$. The fact that the maps above are ring isomorphisms implies (without any calculations!) that $\overline{\alpha\beta} = \overline{\alpha}\overline{\beta}$.

Exercise. If $f(X) \in K[X]$ has degree n, then:
The ring $K[X]/(f(X))$ is a K-vector space of dimension n.

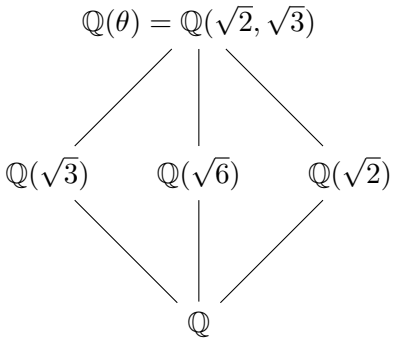
Degree of an Extension

A concrete example: Let $\theta = \sqrt{2} + \sqrt{3}$. Then $(\theta - \sqrt{2})^2 = 3$ so that $\theta^2 + 1 = 2\sqrt{2}\theta$ and $(\theta^2 - 1)^2 = 8\theta^2$ and therefore $\theta^4 - 10\theta^2 + 1 = 0$.

Exercise. The polynomial $g(X) = X^4 - 10x^2 + 1$ is irreducible in $\mathbb{Q}[X]$. Then $1, \theta, \theta^2, \theta^3$ are linearly independent over $\mathbb{Q}$, and $\mathbb{Q}[\theta]$ is a four-dimensional vector space over $\mathbb{Q}$. Therefore: $\mathbb{Q}(\theta) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$.

Hint. $g(X)$ has no roots in $\mathbb{Q}$, (Why?) it has no linear factors in $\mathbb{Q}[X]$. If it is a product of two quadratics, Gauss's Lemma implies that the factors must be in $\mathbb{Z}[X]$. Why is that impossible?

Here is a picture, where the lines represent inclusion between fields. The higher up in the diagram, the bigger the field.



Does this picture capture all of the subfields of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$? We will develop enough algebraic tools to answer this question.

Definition. If $K \subseteq L$ are fields, define the *degree* of the extension L/K to be the dimension of L as a K-vector space. This degree is written as $[L : K]$.

When $L = K[\theta]$ for an algebraic element θ, then $[K[\theta] : K]$ is equals the degree of the the minimal polynomial of θ over K. (This helps justify use of the word "degree in this context.) This claim follows from the earlier exercise about the dimension of the ring $K[X]/(f(X))$.

Here's another example. Let $\theta = \sqrt[3]{5}$, and consider $\mathbb{Q}[\theta]$. Every element of $\mathbb{Q}[\theta]$ is a polynomial in θ, say $f(\theta)$. Since $\theta^3 = 5$ we can eliminate every θ^n for $n \geq 3$, so that every element of $\mathbb{Q}[\theta]$ is of the form $a + b\theta + c\theta^2$ for some $a, b, c \in Q$. Said a different way: divide $f(X)$ by $X^3 - 5$, yielding $f(X) = (X^3 - 5)Q(X) + R(X)$, where R is identically zero or $\deg R < 3$. Evaluating at θ, we find that $f(\theta) = R(\theta)$ is a linear combination of $1, \theta, \theta^2$. This implies $[\mathbb{Q}[\theta] : \mathbb{Q}] \leq 3$. If we show that $X^3 - 5$ is irreducible, then this degree is $[\mathbb{Q}[\theta] : \mathbb{Q}] = 3$.

Is $X^3 - 5$ is irreducible? If it factors nontrivially in $\mathbb{Q}[X]$, then one of the factors must be linear (since the degrees add up to 3). But then $X^3 - 5$ would have a root in $\mathbb{Q}$ and it doesn't, since $\sqrt[3]{5}$ is irrational (proved using unique factorization in $\mathbb{Z}$).

Back to finite fields

Let's use the ideas above to build a field with 4 elements. It's not hard to list all polynomials of small degree in $\mathbb{F}_2[X]$ to see that $X^2 + X + 1$ is the only irreducible of degree 2. Then $K = \mathbb{F}_2[X]/(X^2 + X + 1)$ is a field with 4 elements: $0, 1, \alpha, \alpha + 1$, where $\alpha^2 = \alpha + 1$. This field is denoted $\mathbb{F}_4$.
Exercise. Any field of 4 elements is isomorphic to this one.

Does $\mathbb{F}_{25}$ exist and it is also unique? There are several irreducible quadratics over $\mathbb{F}_5$, for instance $X^2 - 2$ and $Y^2 - 3$. This yields two fields of 25 elements. If there is a uniqueness result, then there must be an isomorphism

$$\frac{\mathbb{F}_5[X]}{(X^2 - 2)} \cong \frac{\mathbb{F}_5[Y]}{(Y^2 - 3)}?$$

If such an isomorphism exists, then the field $\frac{\mathbb{F}_5[X]}{(X^2-2)}$ contains $\sqrt{3}$. Is there some simple reason why this must happen?

The polynomial $X^{25} - X$ will turn out to be key. For any nonzero $\alpha \in \mathbb{F}_{25}$, then $\alpha^{24} = 1$ (since the multiplicative group has order 24, and some version of Fermat applies). We can include the case $\alpha = 0$ by saying: $\alpha^{25} = \alpha$ for every α. Therefore that polynomial splits into linear factors in $\mathbb{F}_{25}$:

$$X^{25} - X = \prod_{\alpha \in \mathbb{F}_{25}} (X - \alpha)$$

If we can show that $X^2 - 3$ divides $X^{25} - X$ (over $\mathbb{F}_5$, don't forget), then $X^2 - 3$ must also split in $\mathbb{F}_{25}$.

Let's state the theorem now.

Theorem Let K be a finite field with q elements. Then $q = p^m$, where p is a prime and m is a positive integer. For any prime power q there is a field with q elements, and any two such fields are isomorphic.

Let's restart for a moment. If K is a finite field with of q elements, then $\mathrm{char}(K) = p$ for some prime p. From there we find that $\mathbb{F}_p$ embeds as a subfield of K. Also, the unit group K^* of K has order $q - 1$, and is cyclic (this is on the sets!). If γ is a generator of K^*, then $K = \mathbb{F}_p[\gamma]$. Certainly γ is algebraic over $\mathbb{F}_p$, so it has a minimal polynomial $m_\gamma(X) \in \mathbb{F}_p[X]$, say of degree n. Then

$$K \cong \frac{\mathbb{F}_p[X]}{(m_\gamma(X))}$$

Then $[K : \mathbb{F}_p] = \deg(m_\gamma(X)) = n$ and it quickly follows (?) that $|K| = p^n$.

This is a first step toward proving the theorem. It remains to show that for every n there exists an irreducible polynomial of degree n in $\mathbb{F}_p[X]$, and of course, there is the question of the uniqueness of K.

Other themes of this course will be complex numbers and, geometry, more variables, and sums of squares.

Digression about formal power series

Extending the definitions of $K[X]$, we may also define the ring of formal power series $K[[X]]$. Its elements are like polynomials, except with possibly infinitely many terms. We don't worry about convergence (we are studying algebra, not analysis!) so we just deal with their formal algebraic properties. Evaluations like $f(\theta)$ make sense only when $f(X)$ is a polynomial.

What are the units in the ring $\mathbb{C}[[X]]$? Well, any nonzero constant is a unit, just as in the case of $\mathbb{C}[X]$. But also $1 - X$ is a unit in $\mathbb{C}[[X]]$: its inverse is the geometric series $1 + X + X^2 + \cdots$. Push this idea further so see that any power series with nonzero constant term is a unit.

What are the primes in $\mathbb{C}[[X]]$? Certainly (X) is a prime (since $\mathbb{C}[[X]]/(X) \cong \mathbb{C}$ is a field). Are there any other primes? Well, we just showed that any power series that has nonzero constant term (i.e. that is not a multiple of X) is a unit. So the answer is no! (X) is the only prime. Contrast this with the situation in $\mathbb{C}[X]$, in which there are infinitely many primes, each of the form $(X - c)$ for some $c \in \mathbb{C}$. By extending $\mathbb{C}[X]$ to $\mathbb{C}[[X]]$ we kill off all primes other than (X).

We could have killed off all the primes except say $(X - c)$ instead, by taking $\mathbb{C}[[X - c]]$, the ring of power series about the number c.

Can we do something similar in $\mathbb{Z}$, to kill off all primes other than 3? One way to do so would be to adjoin to $\mathbb{Z}$ the reciprocals of every prime number other 3. This is the subring of $\mathbb{Q}$ consisting of those fractions whose denominators are prime to 3.

But could we also consider formal power series expansions of numbers, about p? Just food for thought.

2 Chevalley-Warning Theorem

Now, I give you fair warning, either you or your head must be off! - Red Queen

If $f \in K[x_1, x_2, \ldots, x_n]$, then we can evaluate $f(c)$ for any $c \in K^n$. In the one variable case, if $g(x) \in K[x]$, then we know

$$g(c) = 0 \iff (x - c) \mid g(x)$$

We proved this last time, by using the division algorithm to divide $g(x)$ by $x - c$. We found that

$$g(x) = (x - c)Q(x) + g(c)$$

where $Q(x) \in K[x]$. We also talked about the evaluation homomorphism:

$$\text{Eval}_c : K[x] \to K \text{ given by } f(x) \mapsto f(c)$$

The fact that this map is a homomorphism guarantees that evaluation at c "makes sense," meaning that it does not matter whether we add/multiply two polynomials, and then evaluate, or evaluate the two polynomials individually, and then add/multiply the results.

So if $g \in K[x]$, and $c \in K$, we have that $g(c) = 0$ if and only if $g \in (x - c)$ (the ideal generated by $x - c$). [Does this condition still hold if K is some commutative ring?]

If S is a subset of K, we define

$$\mathscr{I}(S) = \{f \in K[x] : f(s) = 0 \ \forall s \in S\}$$

Note that $\mathscr{I}(S)$ is an ideal of $K[x]$ (it's closed under addition, and under multiplication by elements of $K[x]$). If S is an infinite subset of K, then $\mathscr{I}(S) = (0)$, since a nonzero polynomial has only finitely many roots. If S is finite, then $\mathscr{I}(S) = (g)$, where $g \in K[x]$ is the polynomial $g(x) = \prod_{s \in S} (x - s)$.

We want to generalize those results to polynomials in two variables. If $f(0, 0) = 0$, then $f(x, y) = xA(x, y) + yB(x, y)$, that is: $f(x, y)$ (the ideal of $K[x, y]$ generated by x and y). Similarly, $f(a, b) = 0 \iff f(x, y) \in (x - a, y - b)$, the ideal generated by $x - a$ and $y - b$.

Here's a more complicated example. Suppose $f \in \mathbb{R}[x, y]$ vanishes on the unit circle. For instance $x^2 + y^2 - 1$ has this property. We might expect that $\mathscr{I}(\text{unit circle}) = (x^2 + y^2 - 1)$. Let us see if we can prove this. Suppose $g(x, y) \in \mathbb{R}[x, y]$ vanishes on the unit circle: that is, $g(a, b) = 0$ whenever $a^2 + b^2 = 1$. To see whether $g(x, y)$ is in the ideal $(x^2 + y^2 - 1)$, we divide $g(x, y)$ by $x^2 + y^2 - 1$. But how do we divide polynomials in two variables? Well, we consider $g(x, y)$ as a polynomial in y with coefficients in $\mathbb{R}[x]$:

$$g(x, y) = (y^2 + (x^2 - 1))Q(x, y) + R(x, y)$$

where $Q(x, y), R(x, y) \in \mathbb{R}[x][y]$, where $\deg_y R(y) < 2$. (We can include the zero polynomial by setting its degree to be $-\infty$.) Then $R(y) = a(x) + b(x)y$, for some $a, b \in \mathbb{R}[x]$. These manipulations are justified, since division by monic polynomials in $A[x]$ works well for any commutative ring A.

Then:

$$g(x,y) = (x^2+y^2-1)Q(x,y) + a(x) + b(x)y$$

Evaluate this at a point (r,s) on the unit circle to obtain:

$$0 = g(r,s) = 0 + a(r) + b(r)s.$$

That is, $a(r) + b(r)s = 0$ for every $r, s \in \mathbb{R}$ such that $r^2 + s^2 = 1$. Suppose (r,s) is on that circle and $r, s > 0$. Then $(r,-s)$ is also on the circle and we find:

$$a(r) + b(r)s = 0 \quad \text{and} \quad a(r) - b(r)s = 0.$$

Then $a(r) = b(r) = 0$, so that $a(x)$ and $b(x)$ vanish at every $r \in (0,1)$. But $a(x)$ and $b(x)$ are plain old single variable polynomials, so they must be identically 0. This proves that $g(x,y) = (x^2+y^2-1)Q(x,y)$ and: $g(x,y) \in (x^2+y^2-1)K[x,y]$.

Now for the general case. If $X = (x_1, \ldots, x_n)$ is a system of indeterminates, consider polynomials $f_1(X), \ldots, f_k(X)$ and define the set of common zeros:

$$\mathscr{Z}(f_1, \ldots, f_k) = \{c \in K^n : f_i(c) = 0 \text{ whenever } 1 \leq i \leq k\}$$

A subset $S \subset K^n$ is called an *algebraic set* if S is the zero set of some system of polynomials in $K[X]$.

Given a set $S \subset K^n$, we form the ideal of S:

$$\mathscr{I}(S) = \{g \in K[X] : g \text{ vanishes on } S\}$$

We leave it as an exercise to verify that $\mathscr{I}(S)$ is really an ideal of $K[X]$.

We get a correspondence between algebraic sets in K^n and ideals in the polynomial ring $K[X]$. $\mathscr{I}$ maps subsets of K^n to ideals, and $\mathscr{Z}$ maps ideals in $K[X]$ to algebraic sets. Some basic facts (read: exercise):

1. If $S = \mathscr{Z}(J)$, then $\mathscr{Z}(\mathscr{I}(S)) = S$
2. If $J = \mathscr{I}(S)$, then $\mathscr{I}(\mathscr{Z}(J)) = J$

Exercise: Figure out how the $\mathscr{Z}$ and $\mathscr{I}$ operators act in the linear case.

For one non-constant polynomial f we hope $\mathscr{Z}(f)$ will be an $(n-1)$-dimensional hypersurface in K^n. There are many counterexamples to this wish, even over $\mathbb{R}$. For instance if $f(x,y) = x^2 + y^2 + 17$, then $\mathscr{Z}(f(x,y)) = \varnothing$. The empty set is certainly not 1-dimensional, in any sense of the term!

Likewise, over $\mathbb{R}$ we see that $\mathscr{Z}(x^2+y^2) = \{(0,0)\}$ is a single point, when it should intuitively be a curve in 2-space. The problem is that $\mathbb{R}$ is not algebraically closed, so some polynomials

have fewer zeros than expected. That's why algebraic geometry works better over $\mathbb{C}$. Convince yourself that when considered over $\mathbb{C}$, the set $\mathscr{Z}(x^2+y^2+17)$ is in fact a curve (1-dimensional over $\mathbb{C}$).

What do we really mean when we talk about dimension? We will skip all details and just believe for now that a meaningful idea of "dimension" can be defined for algebraic sets over $\mathbb{C}$. This should have properties similar to the simpler case studied in linear (degree 1) algebra. The space $\mathbb{C}^n$ has dimension n, since there are n unknown entries in a vector. Each equation in the system should allow us to solve for one of the variables, so that equation reduces the dimension by 1.

Oops! This doesn't work even for linear equations. For instance the system

$$x+y=3$$
$$x+y=4$$

has no solutions! This inconsistency arises because these equations are not *homogeneous*: they have nonzero constant terms. If we stick to systems of homogeneous equations, things work out more nicely. In a system of r homogeneous equations, the solution space arises from the original n-dimensional space by imposing r conditions, reducing the dimension by r. But that doesn't work either! The list of equations might have the same one twice. What looks like a system of 2 equations might "actually" be only 1 equation. To get a correct dimension-count we need to have "independent" equations. Or we can salvage it another way: A system of r homogeneous linear equations in n variables has solution set of dimension $\geq n-r$. Equality holds if those equations are independent.

This motivation from systems of linear equations leads us to guess the behavior of polynomial equations. A polynomial $f \in K[X]$ in n variables is called *homogeneous of degree d* if every monomial occurring in f has degree d. We also call such f a *form* of degree d.

Remark. Any polynomial $f(x_1,\ldots,x_n)$ in n variables of degree d, can *homogenized*. There is an associated homogeneous polynomial of degree d provided we add another variable. For instance, if $f(x,y)=x^3+2xy+6y$, its homogenized version is: $\widehat{f}(x,y,z)=x^3+2xyz+6yz^2$. Check that $f(x,y)=\widehat{f}(x,y,1)$.

Theorem. ($\mathbb{C}$-geometry) If $f_1,\ldots,f_r$ are forms in $\mathbb{C}[x_1,\ldots x_n]$, and $r<n$, then $\mathscr{Z}(f_1,\ldots f_r)$ is not trivial.

Emil Artin (the father of all this field theory) asked whether similar sorts of theorems might hold true over other types of fields K. It turns out that the degrees of the polynomials come into play when the field is not algebraically closed. For instance, over a finite field, any quadratic form in at least 3 variables has nontrivial zero. (This is on the Ross sets: For any nonzero a,b,c in $\mathbb{Z}_p$ there is a solution to $ax^2+by^2=c$. The homogenized version is: The form $ax^2+by^2-cz^2$ has a nontrivial zero.) Does this generalize to a theorem stating that any system of r quadratic forms in n variables over K admits a nontrivial solution, provided n is large compared to r? Does this generalize to systems of degree d forms?

In 1935 Artin conjectured: If $\varphi(X)$ is a form of n variables of degree d over a finite field K and if $n > d$, then φ must have nontrivial zeros in K^n.

Artin posed this problem to Warning for his PhD thesis. Artin later mentioned the problem to Chevalley over lunch and Chevalley soon published a paper proving it. Luckily Warning was able to strengthen Chevalley's result, so he still earned his PhD.

Chevalley-Warning Theorem. Suppose K is a finite field and f is a form of degree d in n variables over K. If $n > d$ then f has a nontrivial zero in K^n.
More generally, suppose $f_1, \ldots, f_r$ is a system of forms each of degree d. If $n > rd$ then that system has a nontrivial common zero in K^n.

To begin an attack on this result, we consider polynomial functions on K. Every polynomial $g(x) \in K[x]$ has an associated function $\widetilde{g} : K \to K$ defined by evaluation: $\widetilde{g}(c) = g(c)$ is the evaluation of g at c. A function $h : K \to K$ is called a *polynomial function*, if $h = \widetilde{g}$ for some polynomial g.

Over $\mathbb{Q}$, polynomial functions are rare. For instance $x \mapsto |x|$ and $x \mapsto \sin x$ are not polynomial functions. [Proof?] Perhaps a more basic example is provided by the indicator function δ_a for a constant a:

$$\delta_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise.} \end{cases}$$

δ_a is not a polynomial function over $\mathbb{Q}$: It has infinitely many zeros, but is not the zero function.

However, over a finite field K every δ_a is a polynomial function! If K has q elements, then K^* has $q-1$ elements, and Fermat says that $a^{q-1} = 1$ for every nonzero $a \in K$. For the polynomial $f(x) = 1 - (x-a)^{q-1}$ we see that $\widetilde{f} = \delta_a$.

It quickly follows that *every* $h : K \to K$ is a polynomial function. To see this, check: $h = \sum_{c \in K} h(c)\delta_c$. This is a finite sum since K is finite. Since each δ_c is polynomial, so is h. Moreover, if $|K| = q$ then we have found a polynomial of degree $d < q$ that represents the given function h. Moreover, that polynomial is unique: The set of functions $K \to K$ has the same cardinality as the set of polynomials of degree $< q$ in $K[x]$.

Let's move to several variables. A polynomial $g \in K[x_1, \ldots, x_n]$ yields a function $\widetilde{g} : K^n \to K$. Define polynomial functions $K^n \to K$ in the expected way, and claim that any $\alpha : K^n \to K$ is a polynomial function. The proof is the same. For $c = (c_1, \ldots, c_n) \in K^n$, check that:

$$f_c(x_1, \ldots, x_n) = \prod_{j=1}^{n} \left(1 - (x_j - c_j)^{q-1}\right) \quad \text{has} \quad \widetilde{f_c} = \delta_c.$$

Consequently, every $\alpha : K^n \to K$ is a polynomial function, and is represented by a "reduced polynomial," meaning a polynomial that has degree $< q$ in each variable. Then the total degree is $\leq n(q-1)$.

It will be useful to analyze indicator functions δ_S for subsets $S \subseteq K^n$. When does the polynomial degree equal that maximum? Suppose $S = \{c^{(1)}, \ldots, c^{(m)}\}$ for points $c^{(i)} \in K^n$. Then $\delta_S = \sum_{i=1}^{m} \delta_{c^{(i)}}$. The explicit formula above for f_c shows that each polynomial $f_{c^{(i)}}$ is a polynomial with top degree term $(-1)^n (x_1 \cdots x_n)^{q-1}$. The polynomial for δ_S is obtained by adding m of those terms, so if $m \neq 0$ in K then the top degree monomial $(x_1 \cdots x_n)^{q-1}$ has nonzero coefficient. When the field K has characteristic p then $m \neq 0$ in K is the same as: $m \not\equiv 0 \pmod{p}$. This discussion shows:

If $S \subseteq K^n$ and $|S| \not\equiv 0 \pmod{p}$ then the reduced polynomial representing δ_S has degree equal to $n(q-1)$.

We are now in a position to prove a general form of the Chevalley-Warning Theorem.

Theorem. Suppose K is a finite field of characteristic p, and $f_1, \ldots, f_r \in K[X]$ are polynomials in n variables. Suppose $\deg(f_j) = d_j$.

If $n > d_1 + \cdots + d_r$, then $|\mathscr{Z}(f_1, \ldots, f_r)| \equiv 0 \pmod{p}$.

If, in addition, each f_j is homogeneous, then the system $f_1, \ldots, f_r$ has a nontrivial common zero in K^n.

The second statement follows from the first, because the system has a trivial zero $0 \in K^n$, so $\mathscr{Z}(f_1, \ldots, f_r)$ is non-empty. The congruence shows that that zero set contains at least p elements.

First we need a lemma.

Lemma. Let K be a finite field of q elements and let $f \in K[x_1, \ldots, x_n]$. Then there exists a unique polynomial $\widetilde{f} \in K[x_1, \ldots, x_n]$ such that the degree of $\widetilde{f}$ in each variable is $\leq q-1$. Moreover, the total degree of $\widetilde{f}(X)$ is $\leq$ the total degree of $f(X)$. We call this polynomial $\widetilde{f}$ the reduced version of f.

Proof. Exercise. □

Now we can prove the Chevalley-Warning theorem. Let $S = \mathscr{Z}(f_1, \ldots, f_r)$ and define

$$F(X) = (1 - f_1(X)^{q-1}) \cdots (1 - f_i(X)^{q-1})$$

Observe that for any $c \in K^n$, we have

$$F(c) = \begin{cases} 1 & \text{if } c \in S \\ 0 & \text{if } c \notin S \end{cases}$$

After squinting, notice that the total degree of $F(X)$ is $\leq (d_1 + \cdots d_r)(q-1) < n(q-1)$. Next, use the lemma to get a reduced version of $F(X)$, yielding a form $\widetilde{F}(X)$. Note that the total degree of $\widetilde{F}(X)$ is no greater than the total degree of $F(X)$.

Also, for each $c \in K^n$ define $\delta_c(X) = \prod_{1 \leq i \leq n}(1 - (x_i - c_i))^{q-1}$, and set $\delta_S(X) = \sum_{c \in S} \delta_c$. Then:

$$\delta_S(c) = \begin{cases} 1 & \text{if } c \in S \\ 0 & \text{if } c \notin S \end{cases}$$

Therefore both $\widetilde{F}(X)$ and $\delta_S(X)$ induce the same function on K^n. Since they are both in reduced form, they must coincide.

Consider the coefficient of the monomial $(x_1 \cdots x_n)^{q-1}$ in $\delta_S(X)$. A bit of thought shows that it is $(-1)^n|S|$ (where $|S|$ is the number of elements of S). Therefore if $|S| \not\equiv 0 \ (p)$, then $\delta_S(X)$ has total degree equal to $n(q-1)$. This implies that $\delta_S(X)$ and $F(X)$ are not the same polynomial, a contradiction. $\square$

3 Building Algebraic Extensions.

Math is like chess, only without the dice.

Suppose L is an extension field of K. This means that K is a subfield of L. We also write "L/K" or "L over K," in this situation. We can view L as a *vector space* over K, simply by forgetting the multiplicative structure on L. (The only structure we remember is addition in L and multiplication of elements of K with elements of L.) Write $[L : K]$ to denote the dimension of L as a vector space over K. (Recall that "dimension" is defined as the number of elements in any basis.)

As an example, consider $\mathbb{C}/\mathbb{R}$. An $\mathbb{R}$-basis of $\mathbb{C}$ is $\{1, i\}$, so that $[\mathbb{C} : \mathbb{R}] = 2$. For another example: Let $\mathbb{Q}(\sqrt{2})$ be the set of all complex numbers of the form $a + b\sqrt{2}$ where a and b are rational. Then $\mathbb{Q}(\sqrt{2})$ is an extension field of $\mathbb{Q}$, and $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.

The method we used above to construct a field extension of $\mathbb{Q}$ was to look at subfield of a big field $\mathbb{C}$ which we know exists. Can we construct extension fields even if we did not know that a larger field (like $\mathbb{C}$) existed? In order to get around this, we make new fields from old ones using modular arithmetic in the polynomial ring.

Let K be a field. We have defined the polynomial ring $K[x]$. Then if $f(x)$ is in $K[x]$, we can mod out by $f(x)$ to form $K[x]/(f(x)$, in the same way we mod out by an integer n in $\mathbb{Z}$ to form $\mathbb{Z}/n\mathbb{Z}$. Recall that in $\mathbb{Z}$, if p is a prime, then $\mathbb{Z}/p\mathbb{Z}$ is a field. The analog of a prime in $K[x]$ is an irreducible polynomial. As an exercise, verify that if $f(x)$ is an irreducible polynomial in $K[x]$, then $K[x]/(f(x))$ is a field. For example, check that $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, and that $\mathbb{Q}[x]/(x^2 - 2) \cong \mathbb{Q}(\sqrt{2})$.

Let $L = K[x]/(f(x))$ where $f(x)$ is an irreducible polynomial in $K[x]$. Then L is an extension field of K (the constant polynomials in $K[x]$ correspond to the elements of K). Let θ be the class of x in L and note that $f(\theta) = 0$ in L since we forced this to happen! Exercise: $\{1, \theta, \theta^2, \ldots, \theta^{n-1}\}$ is a K-basis of L, and hence that $[L : K] = n$. To prove this we need to show that $\{1, \theta, \theta^2, \ldots, \theta^{n-1}\}$ spans L, and is linearly independent over K. That is, every $\alpha \in L$ can be expressed uniquely as

$$\alpha = a_0 + a_1\theta + \cdots + a_{n-1}\theta^{n-1}$$

for some $a_i \in K$.

Suppose that $[L : K] = n$, for a field extension L/K. Let V be an L-vector space, and suppose $\{v_1, \ldots, v_m\}$ is a basis of V over L. Then we can view V as a K-vector space, simply by forgetting about scalar multiplication by elements of L that are not in K. The dimension will probably increase. For instance, if V is a vector space over $\mathbb{C}$ with $\mathbb{C}$-basis $\{v_1, \ldots, v_m\}$, then V is also a vector space over $\mathbb{R}$ with $\mathbb{R}$-basis $\{v_1, iv_1, \ldots, v_m, iv_m\}$. Then $\dim_{\mathbb{R}}(V) = 2\dim_{\mathbb{C}}(V)$.

Lemma. Let L/K be a field extension with degree $[L : K] = n$. Suppose V is an L-vector space. Then V is also a K-vector space, and $\dim_K(V) = n\dim_L(V)$.

Proof. For simplicity, we consider only the case that $L = K(\theta)$. Then $\{1, \theta, \ldots, \theta^{n-1}\}$ is a K-

basis of L. Suppose V has an L-basis $\{v_1, \ldots, v_m\}$. Then any $v \in V$ can be expressed uniquely as $v = \sum_{j=1}^{m} c_j v_j$ for some $c_j \in L$. But then each c_j can be expressed uniquely as $c_j = \sum_{k=0}^{n-1} g_{j,k}\theta^k$ where $g_{j,k} \in K$. Putting the two together, we see that

$$v = \sum_{j=1}^{m} \sum_{k=0}^{n-1} g_{j,k}\theta^k v_j$$

and hence $\{\theta^k v_j\}$ spans V over K.

Moreover, the expression above is unique. To prove this, it suffices to show that 0 has a unique expression of this type. We leave the details to you. □

Corollary. (Tower theorem) Suppose $K \subseteq L \subseteq F$ is a tower of field extensions. Then

$$[F : K] = [F : L][L : K].$$

Proof. View F as a vector space over L and apply the lemma. □

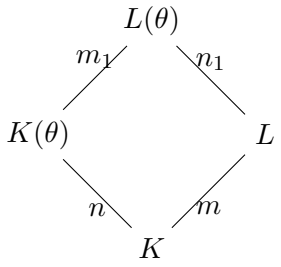
Example. Let $F = \mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$, and $L = \mathbb{Q}(\sqrt{2})$, and $E = \mathbb{Q}(\sqrt[3]{5})$, and $K = \mathbb{Q}$. From earlier work we know that $[L : \mathbb{Q}] = 2$ and $[E : \mathbb{Q}] = 3$. Suppose $[F : L] = n \leq 3$ and $[F : E] = m \leq 2$. By the corollary, $2n = 3m$, and it follows that $n = 3$ and $m = 2$. We just proved that $\sqrt{2} \notin \mathbb{Q}(\sqrt[3]{5})$ and vice versa! We also deduce that $x^3 - 5$ is irreducible over $\mathbb{Q}(\sqrt{2})$. All without any computation!

General Question. Let $f \in K[X]$ be an irreducible polynomial over K. When does f stay irreducible when considered as a polynomial over an extension field L? That is, when does f factor in $L[x]$?

Irreducibles can factor in extension fields. For instance, if $L = K(\theta)$ where θ is algebraic, let $m_\theta(x) \in K[x]$ be its minimal polynomial. Then $m_\theta(x) = (x - \theta)q(x)$ in $L[x]$, for some quotient polynomial $q(x)$.

Exercise. If $L = K(t)$ is the field of rational functions (so that t is not algebraic over K), show that every irreducible in $K[x]$ remains irreducible in $L[x]$.

In the example with $\mathbb{Q}(\sqrt{2}, \sqrt[3]{5})$, we exploited the fact that 2 and 3 are coprime. Generally, suppose $f(x) \in K[x]$ is irreducible and L/K is a finite extension. If $\deg(f)$ and $[L : K]$ are coprime, then must f stay irreducible in $L[x]$? Let θ be a zero of an irreducible polynomial $f \in K[X]$, and let $n = [K(\theta) : K] = \deg f$. Then we have a picture:



Here $m_1 = [L(\theta) : K(\theta)]$ and $n_1 = [L(\theta) : L]$. Note that $n_1 \leq n$, since θ is the root of a polynomial with coefficients in L of degree n. Furthermore: $n_1 = n$ if and only if f is irreducible in $L[X]$. By the tower theorem, $nm_1 = mn_1$. Knowing that m and n are relatively prime, it follows that $n|n_1$. Therefore $n_1 = n$ and f is irreducible over L.

Proposition. Let L/K be a finite extension of fields. Then every element of L is algebraic over K. That is, if $\alpha \in L$, then α is the root of a nonzero polynomial with coefficients in K.

Proof. For $\alpha \in L$ consider the infinite list $1, \alpha, \alpha^2, \ldots$. Since L is finite dimensional over K, these elements can't be linearly independent over L: they must admit a K-dependence relation. That relation is a nonzero polynomial in $K[x]$ that has α as a root. □

A field extension L/K is *algebraic* if every element of L is algebraic over K. We see that every finite extension is algebraic. Can you provide some examples of infinite algebraic extensions?

You may have noticed that it is convenient to work with fields of the form $L = K(\theta)$ for some $\theta \in L$. In this case, θ is called a *primitive element* of L over K. We would hope that every finite extension is primitive. This theorem is known as the primitive element theorem, and it is true in characteristic 0. In fact, it is true for every finite "separable" extension, but we won't pursue this further.

Now we introduce some terminology. We say a field K *embeds* into another field L if K is isomorphic to a subfield of L. Check that this is equivalent to saying: There is a homomorphism from K to L. That is, any ring homomorphism from K to L is an embedding. (Recall that such a homomorphism sends 1 to 1, by definition.)

Exercise. There is an analogy between injective functions of sets, and embeddings of fields. The Schröder-Bernstein theorem states that if A and B are sets, and there exists injective function $\alpha : A \to B$ and $\beta : B \to A$, then there exists a bijective function $\gamma : A \to B$. Is the analogous result true for embeddings of fields? If K embeds in L, and L embeds in K, then are K and L isomorphic?
See if you can come up with a counterexample.

If K embeds in L then certainly K and L have the same characteristic. (Why?) If K is a field of characteristic zero then $\mathbb{Q}$ embeds in K, and we usually consider $\mathbb{Q}$ as a subfield of K. (In fact, $\mathbb{Q}$ is isomorphic to the subfield of K generated by 1.)

Must every such K embed in $\mathbb{C}$?
Not always, for silly reasons: If K has cardinality bigger than that of $\mathbb{C}$, then K can't even inject into $\mathbb{C}$ as a set! But it turns out that any field of characteristic 0 and with cardinality at most that of $\mathbb{C}$ does embed into $\mathbb{C}$. A weaker version of this is: Any algebraic extension of $\mathbb{Q}$ embeds into $\mathbb{C}$. This is useful since it lets us view algebraic extensions of $\mathbb{Q}$ as subfields of a single large field. As of now, we can't say the same thing for extension fields of, say $\mathbb{Z}/p\mathbb{Z}$.

This leads to an important topic: "algebraic closure." We want to construct an extension field over K that is large enough to contain all algebraic extensions of K. (In the same sense that $\mathbb{C}$ contains all algebraic extensions of $\mathbb{Q}$.)

Definition. A field K is *algebraically closed* if every non-constant polynomial in $K[x]$ has a root in K.

Exercise. (1) Prove: A field F is algebraically closed $\iff$ every irreducible in $F[x]$ has degree 1. Then every non-constant $f(x) \in F[x]$ is a product of linear factors in $F[x]$.
(2) Explain why an algebraically closed field must be infinite.

A given field K can be embedded in an algebraically closed field. Here is the first step in this construction.

Lemma. If K is a field and $f(x) \in K[x]$ is irreducible, then there is an extension field L/K such that f has a root in L.

Proof. As usual, define $L = K[x]/(f(x))$ and let $\theta \in L$ be the class of x. Then $f(\theta) = 0$. □

To make an algebraically closed extension field, repeat this process over and over to get a huge tower of field extensions

$$K \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_\infty = \text{union of all previous } K_i.$$

Then every irreducible in $K[x]$ has a root in K_∞, but possibly there are some non-linear irreducibles in $K_\infty[x]$. Let $K^{(1)} = K_\infty$ and do that whole process again to build $K^{(2)}$. Repeat that construction infinitely many times, until at last we reach a great big field Ω with the property that there are no irreducible polynomials in $\Omega[x]$ except for the linear ones. Then every polynomial $f \in \Omega[x]$ splits as a product of linear factors, so that that Ω is algebraically closed.

The fundamental theorem of algebra says that $\mathbb{C}$ is algebraically closed. Let $\mathbb{Q}^{\text{alg}}$ be the set of all $\alpha \in \mathbb{C}$ such that α is algebraic over $\mathbb{Q}$. Check that $\mathbb{Q}^{\text{alg}}$ is algebraically closed. (This follows from the definition, and the fact that $\mathbb{C}$ is algebraically closed.)

As an exercise, show that $\mathbb{Q}^{\text{alg}}$ is countable.

4 Constructible Numbers

We re-invent the wheel not because we want new wheels, but because we want new inventors.

We proved if $K \leq L \leq F$ are fields, then $[F : K] = [F : L][L : K]$. Recall $[L : K]$ is the dimension of L as a K-(vector space).

Ancient Greeks. In the middle ages Euclidean Geometry was considered a perfection, an insight into the mind of god. The ancient Greeks discovered everything there was to know about math, and the European thinkers were simply rediscovering lost knowledge. This was apparently how people in the middle ages saw mathematics. But then in the 1540s Cardano published the cubic formula, and there was no indication the Greeks were aware of that. But then people revised their view: the ancient Greeks knew the holy grail of Geometry, if not of all mathematical ideas.

In the 1800s results in geometry were discovered that the ancient Greeks definitely did not know. Several hard problems were stated but not solved in ancient times:

- Trisect an angle:
  Given an arbitrary angle θ, construct the angle $\frac{1}{3}\theta$.
- Duplicate a cube:
  Given a unit cube, construct a cube of volume 2. (That is: construct the length $\sqrt[3]{2}$).
- Square a circle:
  Given a unit circle, construct a square with the same area.
- Construct a regular heptagon.
  The ancients were able to construct regular n-gons for $n = 3, 4, 5, 6$, but 7 was not done.

Euclidean tools for construction: compass and straight edge. Why those tools in particular? (Euclid did have access to more complicated tools, so why did he choose only those two?) Answer: These tools arise directly from Euclid's axioms for geometry.

One axiom is the existence of lines through two given points. The straight-edge captures this notion. Another axiom is the existence a circle of a given radius centered at a given point. The compass captures this idea.

"Squaring a rectangle" is something that the ancient Greeks knew how to do. Given a rectangle of side lengths a and b, they could make a square of side length s, where $s = \sqrt{ab}$. This s is the "geometric mean" or "mean proportional", arising in terms of ratios: $\frac{a}{x} = \frac{x}{b}$.

Exercise. Explain how to do this, using only Euclidean tools. Given segments of lengths 1, a, and b, explain how to construct segments of length $a + b$, $a - b$, ab, and a/b, and $\sqrt{a}$.

Definition. A number r is said to be constructible if it is possible to create a segment of length $|r|$ given a unit segment 1 and Euclidean tools. Define Co to be the set of constructible numbers.

Since 1 is given, the exercise above shows: Co is a subfield of $\mathbb{R}$, every positive number in Co is a square.

We will use coordinates (analytic geometry) to develop a relationship between constructions and geometry and the algebra of fields.

Proposition. Suppose $\alpha \in Co$. Then α is algebraic over $\mathbb{Q}$. In fact, $\alpha \in K$ for some field K that is at the top of a tower of quadratic extensions starting from $\mathbb{Q}$.

To illustrate this idea, let $\alpha = \sqrt{4-\sqrt{2}} - \sqrt{3}$. Define $K_1 = \mathbb{Q}(\sqrt{2})$, $K_2 = K_1(\sqrt{4-\sqrt{2}})$, and $K_3 = K_2(\sqrt{3})$. Then $\mathbb{Q} \subseteq K_1 \subseteq K_2 \subseteq K_3$ is a tower of quadratic extensions, and $\alpha \in K_3$.

Proof sketch. Each step of geometric construction involves drawing lines and circles between points that have been already constructed, and finding the intersections of lines and circles drawn in this way. Intersections of lines and circles are given by solutions to linear or quadratic equations whose coefficients are previously constructed numbers. Several cases need to be checked. For instance, for a circle and a line we need to solve the system:

$$(x-u)^2 + (y-v)^2 = r^2 \quad \text{and} \quad y = mx + b$$

where $u, v, r, m, b \in Co$.

Exercise. Complete the proof of the Proposition.

Corollary. If $\alpha \in Co$, then $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ is a power of 2.

This follows from the multiplicativity of the degree of field extensions.

Corollary. Let $\gamma = \sqrt[3]{2}$ has degree 3 over $\mathbb{Q}$ (Proof?). Therefore γ is not constructible, so it is impossible to "duplicate the cube" with Euclidean tools.

Exercise. Is the converse of the Proposition true? That is, if $\beta \in \mathbb{R}$ has degree 2^k over $\mathbb{Q}$, then must β be constructible? [Answer: No. Proved via Galois theory.]

These ideas also apply to angle trisection. To trisect a 60° angle, we need to construct one of 20°. This is equivalent to showing: $\cos 20^\circ \in Co$. Exercise: Prove $\cos 20^\circ$ has degree 3 over $\mathbb{Q}$.

For regular polygons, check that:

$$\text{can construct a regular } n\text{-gon} \iff \cos(2\pi/n) \in Co.$$

Let $\beta = \cos 2\pi/n$ and notice that $2\beta = \zeta + \zeta^{-1}$, where $\zeta = \zeta_n = e^{2\pi i/n} = \cos 2\pi/n + i \sin 2\pi/n$. The degree of ζ and the degree of β are therefore related.

Theorem. ζ_n has degree $\varphi(n)$ over $\mathbb{Q}$.

This is equivalent to proving that the n-th cyclotomic polynomial $\Phi_n(X)$, is irreducible in $\mathbb{Q}[X]$. Those polynomials are discussed on a "Special Set" in the Ross Number Theory course. We skip further discussion of this famous result.

We have a tower $\mathbb{Q} \subseteq \mathbb{Q}(\beta) \subseteq \mathbb{Q}(\zeta)$ and the degree of the top over the bottom is $\varphi(n)$. The degree of the top over the middle is 2. (Proof?) Therefore $[\mathbb{Q}(\beta) : \mathbb{Q}] = \varphi(n)/2$.

For example, the degree of $\cos 2\pi/5$ is 2, which verifies the fact that regular pentagons are constructible (as the ancient Greeks proved). On the other hand, the degree of $\cos 2\pi/7$ is 3, and therefore a regular 7-gon is not constructible.

When is a regular n-gon constructible? At least we must have $\varphi(n)$ is a power of 2. The formula for $\varphi(n)$ shows that for every odd prime factor $p \mid n$, then $p - 1 = \varphi(p)$ is a 2-power. Then $p = 2^k + 1$ and it follows that $p = 2^{2^m} + 1$, for some m. Such primes are called Fermat primes. Are there infinitely many primes of this type?

The converse is true as well: If p is a Fermat prime, then a regular p-gon is constructible. Gauss earliest major result was his proof that a regular 17-gon is constructible.
(He was born in 1777 and was probably 17 at the time.)

One ancient problem in our list requires more work. Squaring a circle is equivalent to showing that π is not constructible. This is much harder to prove than the other examples because π is not algebraic. The proof that π is transcendental was first proved by Lindemann (using ideas of Hermite). It requires some ideas from calculus, a topic that is not part of this course.

5 Norm and Trace

The path of the norm is the path of least resistance. - Melanie Joy

For Gaussian integers we defined the norm $N : \mathbb{Z}[i] \to \mathbb{Z}$ by $N(a+bi) = a^2 + b^2$. It turns out that $N(\alpha) = \alpha\overline{\alpha}$, where: $\alpha = a + bi \Rightarrow \overline{\alpha} = a - bi$. Multiplicativity $N(\alpha\beta) = N(\alpha)N(\beta)$ follows from the similar property for "bar": $\overline{\alpha\beta} = \overline{\alpha}\,\overline{\beta}$. This says: "bar" is a ring isomorphism.

We could prove this mechanically, but there is an abstract proof. Recall that evaluation at i gives a homomorphism $\mathbb{Z}[X] \to \mathbb{Z}[i]$. The kernel of this map is precisely (X^2+1) so we have an isomorphism $\mathbb{Z}[X]/(X^2+1) \xrightarrow{\sim} \mathbb{Z}[i]$ with the class of X mapping to i. Evaluation at $-i$ provides a similar isomorphism, and the two together yield:

$$\mathbb{Z}[i] \xleftarrow{\sim} \frac{\mathbb{Z}[X]}{(X^2+1)} \xrightarrow{\sim} \mathbb{Z}[-i].$$

Combining those isomorphisms from left to right, we obtain a map from $\mathbb{Z}[i]$ to itself, sending i to $-i$. That is: the "bar" map is a ring isomorphism. Proved without any explicit calculation!

If K is a field and $f \in K[X]$ irreducible, suppose θ is a root of f in some larger field. Then "evaluation at θ" yields a K-isomorphism $K[X]/(f) \xrightarrow{\sim} K(\theta)$ with θ corresponding to the class of X. (A K-isomorphism is a bijective ring homomorphism that fixes elements of K.) The idea above then shows that if θ and θ' are two roots of f in some larger field, then there is a K-isomorphism $K[\theta] \xrightarrow{\sim} K[\theta']$ sending $\theta \mapsto \theta'$.

Let's back up a step and review what we mean by an *ideal* in a ring. If R is a commutative ring, and J is a subset of R such that $J + J \subseteq J$, and $R \cdot J \subseteq J$, then we say J is an ideal of R. The natural thing to do with an ideal is to mod out by it: forming a quotient ring R/J. For instance, $5\mathbb{Z}$ is an ideal of $\mathbb{Z}$, and we form the quotient ring $\mathbb{Z}/5\mathbb{Z}$, (called $\mathbb{Z}_5$ on the Ross sets).

Evaluation $g(X) \mapsto g(\theta)$ induces a ring homomorphism: $K[X] \to K[\theta]$ sending X to θ. Its kernel is $\{g(X) \in K[X] : g(\theta) = 0\} = (m_\theta(X))$, the principal ideal generated by the minimal polynomial $m_\theta(X)$ in $K[X]$. If $f(X)$ is irreducible over K and θ is a root in some larger field, then $f(X) = m_\theta(X)$. Then we get an isomorphism from $K[X]/(f(X))$ to $K[\theta]$ sending X to θ.

Exercise. Explain why $\mathbb{Q}[X]/(X^2 - 5X + 6)$ is isomorphic to the direct product ring $\mathbb{Q} \times \mathbb{Q}$. What ordered pair corresponds to the class of X? [Hint: Chinese remainder theorem.]

To summarize: If $f(X)$ is irreducible over K and θ and θ' are two roots of f in some larger field, then there is an isomorphism from $K[\theta]$ to $K[\theta']$ that fixes K and sends θ to θ'.

For example, suppose $d \in K$ is not a square in the field K. Then $K(\sqrt{d})$ is a field, and a typical element α is of the form $\alpha = a + b\sqrt{d}$ for $a, b \in K$. Define $\overline{\alpha} = a - b\sqrt{d}$. The observations above show that this "bar" map is a K-automorphism of K. Therefore the norm $N\alpha = \alpha\overline{\alpha} = a^2 - db^2$ is multiplicative: $N\alpha\beta) = N(\alpha)N(\beta)$. Analogously, we define the trace $\mathrm{Tr}(\alpha) = \alpha + \overline{\alpha} = 2a$.

To generalize the ideas of norm and trace to extensions of larger degree, we note that those ideas

are similar to the notions of determinant and trace of matrices. Let $\mathbb{M}_n(K)$ be the ring of all $n \times n$ matrices over K. Then $\det(A)$ and $\text{trace}(A)$ are defined, providing maps

$$\det, \text{trace} : \mathbb{M}_n(K) \to K.$$

Standard linear algebra tells us that $\det(A)$ is the sum of the eigenvalues of A, so it is the constant term of the characteristic polynomial $p_A(x) = \det(xI_n - A)$. Its key property is: $\det(AB) = \det(A)\det(B)$. Similarly trace is the sum of the eigenvalues, it is the sum of the diagonal entries of the matrix, and it is the coefficient of x^{n-1} in the characteristic polynomial is $-\text{trace}(A)$. The trace is a K-linear map.

The connection between field extensions and linear algebra comes from multiplication maps.

Definition. Suppose L/K is a field extension of degree n, and $\alpha \in L$. Define $\mu_\alpha : L \to L$ to be "multiplication by α." Note that L as an n-dimensional K-vector space and μ_α is a K-linear map $L \to L$. Define

$$N_K^L(\alpha) = \det(\mu_\alpha) \quad \text{and } Tr_K^L(\alpha) = \text{trace}(\mu_\alpha).$$

Then N_K^L and Tr_K^L are mappings $L \to K$.
When the fields L, K are held fixed, we often abbreviate notations to just N and Tr.

Properties of determinants quickly imply: $\quad N(\alpha\beta) = N(\alpha)N(\beta) \quad$ for every $\alpha, \beta \in L$.

Suppose $L = \mathbb{Q}(\sqrt{d})$ over $K = \mathbb{Q}$ with basis $\{1, \sqrt{d}\}$. Then $\alpha = r + s\sqrt{d}$ in L is represented as a column-vector $\left[\begin{smallmatrix} r \\ s \end{smallmatrix}\right]$ and the matrix of μ_α is

$$[\mu_\alpha] = \begin{bmatrix} r & s\sqrt{d} \\ s & r \end{bmatrix}.$$

Exercise. Check that: $N(\alpha) = \det(\mu_\alpha) = r^2 - ds^2$ and $Tr(\alpha) = 2r$.
The map μ_α has characteristic polynomial $X^2 - 2rX + (r^2 - ds^2)$ and eigenvalues α and $\overline{\alpha}$. Then $N(\alpha) = \det(\mu_\alpha)$ is the product of the eigenvalues: $N(\alpha) = \alpha\overline{\alpha}$.

Exercise. (1) Let $K = \mathbb{Q}$ and $L = Q[\theta]$ where $\theta = \sqrt[3]{2}$. Show that $m_\theta(X) = X^3 - 2$ so that $[\mathbb{Q}(\theta) : \mathbb{Q}] = 3$ and every element has the form $\alpha = a + b\theta + c\theta^2$ for some $a, b, c \in \mathbb{Q}$.

(2) Express $(1 + 2\theta)^{-1}$ in that form. Can this be done by "rationalizing the denominator"?

(3) Factor $X^3 - 2 = (X - \theta)(X - \omega\theta)(X - \omega^2\theta)$, where $\omega = (-1 + \sqrt{3})/2$, so that $\omega^3 = 1$ in $\mathbb{C}$. Then $\theta, \omega\theta, \omega^2\theta$ are the *conjugates* of θ: the roots of its minimal polynomial.
If $\alpha = a + b\theta + c\theta^2$ is not in $\mathbb{Q}$ then $m_\alpha(X)$ has degree 3. Use matrices to show that $\text{Tr}(\alpha) = 3a$ and $N(\alpha) = a^3 + 2b^3 + 4c^3 - 6abc$.
Show that α, $\alpha' = a + b\omega\theta + c\omega^2\theta^2$ and $\alpha'' = a + b\omega^2\theta + c\omega\theta^2$ are the conjugates and verify that $Tr(\alpha) = \alpha + \alpha' + \alpha''$ and $N(\alpha) = \alpha\alpha'\alpha''$.

Suppose L/K is a finite extension of fields and $\alpha \in L$. To analyze the norm $N_K^L(\alpha)$ and trace $Tr_K^L(\alpha)$, we first consider the case that $L = K(\alpha)$.

Suppose $m_\alpha(X) = X^n + c_1X^{n-1} + \cdots + c_{n-1}X + c_n$ is the minimal polynomial of α over K.

Then $(1, \alpha, \alpha^2, \ldots, \alpha^{n-1})$ is a K=basis of L. The map μ_α acts as on this basis as follows:

$$\begin{aligned}
&1 \mapsto \alpha \\
&\alpha \mapsto \alpha^2 \\
&\alpha^2 \mapsto \alpha^3 \\
&\quad \vdots \\
&\alpha^{n-2} \mapsto \alpha^{n-1} \\
&\alpha^{n-1} \mapsto -c_1\alpha^{n-1} - \cdots - c_{n-1}\alpha - c_n
\end{aligned}$$

Therefore the matrix for μ_α in this basis is:

$$C_\alpha = [\mu_\alpha] = \begin{bmatrix} 0 & 0 & \cdots & 0 & -c_n \\ 1 & 0 & \cdots & 0 & -c_{n-1} \\ 0 & 1 & \cdots & 0 & -c_{n-2} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -c_1 \end{bmatrix}$$

This is often called the "companion matrix" of the polynomial $m_\alpha(X)$.

Exercise. The characteristic polynomial of C_α is $m_\alpha(X)$. Therefore:

$$N_K^L(\alpha) = \det(C_\alpha) = (-1)^n c_n \quad \text{and} \quad Tr(\alpha) = \operatorname{trace}(C_\alpha) = -c_1.$$

Verify that $N(\alpha)$ is the product of the conjugates of α, and $\mathrm{Tr}(\alpha)$ is the sum of those conjugates.

Now let's return to the general case of L/K where $[L : K] = n$, and $\alpha \in L$. Suppose α has degree d over K. Then $K(\alpha) : K] = d$ and $n = dk$ where $k = [L : K(\alpha)] = k$. Let $m_\alpha(X)$ be the minimal polynomial of α over K: $m_\alpha(X) = X^d + c_1X^{d-1} + \cdots + c_{n-1}X + c_d$.
Then $(1, \alpha, \ldots, \alpha^{d-1})$ is a K-basis of $K(\alpha)$.

Since L is a vector space over $K(\alpha)$ we may choose a basis $\beta_1, \ldots, \beta_k$ for L over $K(\alpha)$. Then every $\lambda \in L$ has a unique expression as:

$$\lambda = r_1\beta_1 + \cdots r_k\beta_k \text{ for some } r_j \in K(\alpha).$$

In turn, each r_j can be expanded in terms of the basis $(1, \alpha, \ldots, \alpha^{d-1})$. Therefore $\{\alpha^i\beta_j\}$ forms a K-basis of L. How does the map μ_α act on this basis? For a fixed j, notice that

$$\begin{aligned}
\mu_\alpha(\alpha^i\beta_j) &= \alpha^{i+1}\beta_j \quad \text{for every } i \leq d-2 \\
\mu_\alpha(\alpha^{d-1}\beta_j) &= -c_1\alpha^{d-1}\beta_j - \cdots - c_{d-1}\alpha\beta_j - c_d\beta_j
\end{aligned}$$

Therefore, $\mu_\alpha : L \to L$ is represented by a $k \times k$ block diagonal matrix, with $d \times d$ blocks on the diagonal:

$$[\mu_\alpha] = \begin{bmatrix} C & 0 & \cdots & 0 \\ 0 & C & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & C \end{bmatrix}.$$

Here C is the d-by-d companion matrix of α, and the whole matrix has size $n \times n$, since $dk = n$.

From this matrix we deduce:

$$N_K^L(\alpha) = \det(\mu_a) = (\det(C))^k = (N_K^{K(\alpha)}(\alpha))^k,$$

$$Tr(\alpha) = \operatorname{trace}(\mu_\alpha) = k \operatorname{trace}(C) = kTr_K^{K(\alpha)}(\alpha).$$

Suppose $K \subseteq L \subseteq F$ is a tower of finite extension fields. We know that the degrees behave well: $[F : K] = [F : L][L : K]$. The claim is that the norms and traces also work well: If $\alpha \in L$, then we can compute $N_K^F(\alpha)$ by first computing the norm from F to L, and then computing its norm from L to K.

Proposition. If $F/L/K$ is a tower of finite extensions of fields, then

$$N_K^F(\alpha) = N_K^L\big(N_L^F(\alpha)\big) \quad \text{for every } \alpha \in L.$$

That is: $N_K^F = N_K^L \circ N_L^F$.

The work above with companion matrices verifies this formula holds true when $L = K(\alpha)$.

We omit the proof of the Proposition. One of its proofs uses matrices. That method requires a fact about determinants of block matrices. Here's the idea when $k = 2$.
If A, B, C, D are commuting $d \times d$ matrices, then:

$$\det_{2d} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det_d(AD - BC).$$

Proof idea: If A is invertible, a "block row operation" alters that matrix to $\begin{bmatrix} A & B \\ 0 & D - CA^{-1}B \end{bmatrix}$.
Therefore $\det_{2d} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det_d(A) \det_d(D - CA^{-1}B) = \det_d(AD - ACA^{-1}B)$.
If those matrices commute, the stated formula follows. What if A is not invertible?

6 Hilbert's 17^{th}, and Orderings of Fields

I have yet to see any problem, however complicated, which, when you looked at it in the right way, did not become still more complicated. - Poul Anderson

Let L/K be an finite extension of fields. We defined the *norm* and *trace* from L to K. For any given $\alpha \in L$, the multiplication map $\mu_\alpha : L \to L$ (defined as $\mu_\alpha(x) = \alpha x$) is K-linear. We defined $N_K^L : L \to K$ as $N_K^L(\alpha) = \det(\mu_\alpha)$, and $\mathrm{Tr}_K^L : L \to K$ as $\mathrm{Tr}_K^L(\alpha) = \mathrm{trace}(\mu_\alpha)$. (We assume you know what the determinant and trace are in linear maps). Note that the norm is a multiplicative homomorphism from L^* to K^*, and the trace is an additive homomorphism from L to K. The norm and trace are useful tools. For instance, the trace helps prove that $\sqrt{2}$ and $\sqrt{3}$ are linearly independent over $\mathbb{Q}$.

Let $L = K(\sqrt{a})$ have degree 2 over K. Then the norm $N : L \to K$ is given by $N(r + s\sqrt{a}) = r^2 - as^2$. As mentioned above, the norm is a multiplicative homomorphism of L^* to K^*. That is, for all $\alpha, \beta \in L$, $N(\alpha\beta) = N(\alpha)N(\beta)$. What is its kernel? That is, can we characterize the $\alpha \in L$ of norm 1 in some useful way? For instance, $N(\pm 1) = 1$. In addition, for any $\theta \in L^*$, we have $N(\theta) = N(\bar{\theta})$, and hence $N(\bar{\theta}/\theta) = 1$ as well.

Lemma. The kernel of N is precisely the set of $\bar{\theta}/\theta$ for some nonzero $\theta \in L$.

Proof. If $N(\gamma) = 1$ let $\theta = 1 + \overline{\gamma}$ and note: $\gamma\theta = \gamma + 1 = \overline{\theta}$, so $\gamma = \overline{\theta}/\theta$. At least if $\theta \neq 0$. If $\theta = 0$, then $\gamma = -1$ and we want nonzero α such that $\overline{\alpha} = -\alpha$. That's easy: just take $\sqrt{a}$. $\square$

By the way, this is a special case of a result known as Hilbert's theorem 90.

Exercise. When is $N\alpha$ a square in K? Show: If $N\alpha \in K^2$, then $\alpha = c\beta^2$ for some $c \in K$ and some $\beta \in L$.

Now we turn our attention to Hilbert's 17th Problem. Let $X = (x_1, \ldots, x_n)$ be a system of indeterminate, and let $f(X) \in \mathbb{R}[X]$ be a polynomial in the variables $x_1, \ldots, x_n$. We say f is *positive semidefinite*, or PSD, if $f(c) \geq 0$ for every $c \in \mathbb{R}^n$. The story is that in the 1880's, Hilbert attended Minkowski's PhD oral exam, and the following question arose. If $f \in \mathbb{R}[X]$ is PSD over $\mathbb{R}$, must $f(X)$ be a sum of squares? Hilbert proved several cases, and the problem became known as Hilbert's 17th problem.

Exercise. (1) If $f(x) = ax^2 + bx + c \in \mathbb{R}[x]$ (one variable), and $f(c) \geq 0$ for all $c \in \mathbb{R}$, then $f(x) = g_1(x)^2 + g_2(x)^2$ for some polynomials $g_i(x) \in \mathbb{R}[x]$.

(2) **Lemma.** If $f(x) \in \mathbb{R}[x]$ is PSD then $f(x) = g_1(x)^2 + g_2(x)^2$ for some $g_i(x) \in \mathbb{R}[x]$.
Outline: If $f(x)$ is PSD, then certainly $\deg f(x)$ is even. Any nonzero polynomial $f(x) \in \mathbb{R}[x]$ splits as a product of linear factors in $\mathbb{C}[x]$, so it is the product of some irreducible linear and quadratic factors in $\mathbb{R}[x]$:

$$f(x) = c(x - r_1)^{n_1} \cdots (x - r_k)^{n_k} Q_1(x) \cdots Q_s(x),$$

where $c \in \mathbb{R}^*$, r_i are the distinct real roots of $f(x)$ with multiplicities n_i, and each $Q_i(x)$ is a

monic irreducible quadratic.

Now suppose that $f(x)$ is PSD. Then each n_i is even. [n_i odd $\Rightarrow f(x)$ changes sign at r_i]. By part (1) each $Q_j(x)$ is a sum of two squares in $\mathbb{R}[x]$. Deduce that $f(x)$ is a sum of two squares. $\square$

What about if we allow two variables? Hilbert showed in 1888 that there exists a PSD polynomial in $\mathbb{R}[x, y]$ not expressible as a sum of squares in $\mathbb{R}[x, y]$. Motzkin's polynomial

$$M(x, y) = 1 + x^2y^4 + x^4y^2 - 3x^2y^2$$

is an example.

Exercise. Prove that $M(x, y)$ is PSD (by GM $\leq$ AM). Prove $M(x, y)$ is NOT a sum of squares in $\mathbb{R}[x, y]$. With a bit more work, you might find an explicit formula expressing $M(x, y)$ as a sum of 4 squares of rational functions. (Of course, some denominators must be involved.)

It is known that Motzkin's polynomial is not a sum of 3 squares in $\mathbb{R}(x, y)$, but this is more subtle. That result is due to Cassels, Ellison, and Pfister.

Hilbert's 17th Problem If $f(X) \in \mathbb{R}[X]$ is a positive semidefinite polynomial in several variables, then can $f(X)$ be expressed as a sum of squares of rational functions? That is:
Do there exist rational functions $a_i(X) \in \mathbb{R}(X)$ such that $f(X) = \sum_{i=1}^{k} (a_i(X))^2$?

In 1927 Artin published a solution to Hilbert's 17th problem. Artin and Schreier together built up a theory of *ordered fields* as machinery to attack Hilbert's 17th problem. We will describe some of those beautiful ideas.

Definition. An *ordering* on a field K is a proper subset $P \subset K$ such that
$$P + P \subseteq P, \quad P \cdot P \subseteq P, \quad 0 \in P, \text{ and } P \cup (-P) = K.$$

If K is a field and P is an ordering of K, we can use P to define a total order relation $\leq_P$ on K, by defining $a \leq_P b$ if and only if $b - a \in P$. Then P is the set of "non-negative elements" with respect to this order. Check that the properties of P required in the definition above are necessary and sufficient for P to induce a total order $\leq_P$ that respects the ring operations in K in the usual ways.

Let P be an ordering of a field K. Here some simple consequences of the definitions. Let $\sum K^2$ denote the set of elements of K that are expressible as a sum of squares in K. Then

$$\sum K^2 \subseteq P.$$

Also, $-1 \notin P$, since otherwise P would be equal to K. (Since every number is some $u^2 - v^2$). But then P is not a proper subset of K.

Some fields do not have any orderings. For instance, $\mathbb{C}$ can't be ordered, since -1 is a square.

In addition, no field of positive characteristic can be ordered. For if K has characteristic p then -1 is a sum of $p-1$ ones, so -1 is a sum of squares and would be in P.

Some fields have exactly one ordering. For instance $\mathbb{R}$ has an ordering $P = \mathbb{R}^+$ (the set of positive reals). That ordering is unique since P is the set of squares.

Definition. A field K is *formally real* if $-1 \notin \sum K^2$.

Proposition. (Artin) A field K is formally real if and only if K has an ordering.

To prove this, we introduce the notion of a *preorder*.

Definition. Let K be a field. A *preorder* of K is a subset T of K such that
$$T + T \subseteq T, \quad T \cdot T \subseteq T, \quad K^2 \subseteq T, \text{ and } \quad -1 \notin T.$$

Note that if K has a preordering $-1 \notin \sum K^2$, i.e. K is formally real. Conversely, if K is formally real, then $\sum K^2$ is a preordering of K. If T is a preordering, and $a \in K$, define $T[a] = T + aT$.

Exercise. $T[a]$ is a preordering if and only if $a \notin -T^*$.
(Here T^* is the set of nonzero elements of T. It is a multiplicative group.)
[Key step: $-1 \notin T[a]$. If not, then $-1 = t + at'$ for some $t, t' \in T^*$, and $a = -(1+t)/t' \in -T^*$.]

In order to prove Artin's proposition, note that if K is formally real then $\sum K^2$ is a preordering. Now apply Zorn to the set of preorderings of K. This yields a maximal preordering, say P. **Claim.** P is an ordering of K. Since it's a preordering, we need only show: $P \cup (-P) = K$. Suppose $a \in K$ but $a \notin -P$. The Exercise implies $P[a]$ is also a preordering, and maximality forces $P = P[a]$. Therefore $a \in P$. □

Exercise. Strengthen Artin's result as follows:
Proposition. If T is a preordering of a field K then T is contained in some ordering of K.

This implies the following (philosophically stunning!) result, also due to Artin:

Corollary. Let K be a field, and let $a \in K$. Then
$$a \in \sum K^2 \quad \iff \quad a \in P, \text{ for every ordering } P \text{ of } K.$$
In other words (symbols):
$$\sum K^2 = \bigcap_{\substack{P \text{ ordering} \\ \text{of } K}} P.$$

Then to check whether an element a of K^* is a sum of squares, we need only verify that a is positive relative to every ordering of K.

7 Embedding into K^{alg}

Mathematicians strive to confuse their audiences; where there is no confusion there is no prestige.
- Carl Linderholm, *Mathematics Made Difficult*

Some review: Let L/K be an extension of fields (i.e. K is a subfield of the field L). What does it mean for L/K to be an "algebraic" extension? Recall that we defined what it means for an element $\alpha \in L$ to be algebraic over K: it means that α is the root of a nonzero polynomial with coefficients in K. Here are a few equivalent ways of formulating this idea:

Proposition. Let L/K be an extension of fields, and let $\alpha \in L$. The following are equivalent:
(1) α is algebraic over K.
(2) $K[\alpha]$ is a finite dimensional vector space over K.
(3) There exists a field L containing K such that $\alpha \in L$ and $[L : K]$ is finite.
(4) $1/\alpha \in K[\alpha]$.
(5) $K[\alpha]$ is a field.

The proof of the equivalence of these statements is left as an exercise. Here, as usual, $K[\alpha]$ denotes the smallest ring containing both K and α, and $K(\alpha)$ denotes the smallest field containing both K and α. We may restate (5) as: α is algebraic over $K \iff K[\alpha] = K(\alpha)$.

Definition. L/K is an *algebraic extension* if every element of L is algebraic over K.

The Proposition above implies that if $[L : K]$ is finite, then L/K is an algebraic extension. (Find an example of an algebraic extension with $[L : K]$ infinite.)

Proposition. Let L/K be an extension of fields, and suppose $\alpha, \beta \in L$ are algebraic over K. Then $\alpha + \beta$, $\alpha\beta$, and (if $\beta \neq 0$) α/β are also algebraic over K.

This proposition is not at all obvious from the original definitions. Given polynomials that kill α and β individually, it seems computationally challenging to produce a polynomial that kills $\alpha + \beta$, for instance.
Item (3) of the earlier Proposition above makes theis proof much easier. □

Let us consider the polynomial $X^n - a \in \mathbb{Q}[x]$ ($a \in \mathbb{Q}$). Is it irreducible in $\mathbb{Q}[x]$? Gauss's Lemma says that it suffices to work in $\mathbb{Z}[X]$.
One famous trick is **Eisenstein's Irreducibility Criterion** for a prime number p:
Suppose $f(X) \in \mathbb{Z}[X]$ has degree n and $f(X) \equiv X^n \pmod{p}$, but $f(0) \not\equiv 0 \pmod{p^2}$.
Then $f(X)$ is irreducible in $\mathbb{Q}[X]$.

Proof. Suppose f factors as $f = ab$ in $\mathbb{Q}[X]$. According to Gauss's lemma, we may assume that a, b lie in $\mathbb{Z}[X]$. Then $f(X) \equiv X^n \equiv a(X)b(X) \pmod{p}$. Unique factorization in $\mathbb{Z}/p\mathbb{Z}[X]$ implies $a(X) \equiv X^i \pmod{p}$ and $b(X) \equiv X^j \pmod{p}$ for some i and j such that $i + j = n$. If both $i, j > 0$ then $a(0)$ and $b(0)$ are each divisible by p and $f(0)$ is divisible by p^2, contrary to the hypothesis. Therefore $i = 0$ or $j = 0$ so that one of $a(X)$ or $b(X)$ is a unit. This proves $f(x)$ is irreducible. □

We were wondering for which numbers a is $X^n - a$ irreducible in $\mathbb{Q}[X]$. The question is subtle: check that $X^2 - 2$ is irreducible, but in fact $X^4 + 4$ factors in $\mathbb{Q}[X]$! It is an interesting exercise to try to figure out for which $a \in \mathbb{Q}$ and which n the polynomial $X^n - a$ is irreducible over $\mathbb{Q}$. But now we return to general theory.

Suppose $f(X) \in K[X]$ is non-constant. Then there is a field $L \supseteq K$ in which f has a root. To construct L, choose an irreducible factor $\pi(X)$ of $f(X)$ in $K[X]$, and let $L = K[\theta] = K[X]/(\pi(X))$, where θ is the class of X. Then L/K is a finite extension of K and L contains a root of f. We may repeat this process as needed to create a finite sequence of field extensions $K \subseteq L \subseteq L' \subseteq \cdots \subseteq F$ such that $f(X)$ splits into linear factors in $F[X]$.

Suppose F is a field such that $f(X) = c\prod_j (X - \theta_j)$ where $c \in K$ and each $\theta_j \in F$. Then $E = K(\theta_1, \ldots, \theta_n)$ is called a *splitting field* for f over K. The remarks above show that spitting fields exist.

Claim. A splitting field of f over K is unique, up to K-isomorphism.

Proof Outline. Suppose f has degree n over K and $E, E' \supseteq K$ are splitting fields. Then $E = K(\theta_1, \ldots, \theta_n)$ where $\theta_1, \ldots, \theta_n \in E$ are roots of f. Then $K(\theta_1) \cong K[X]/(m_1(X))$ where $m_1(X)$ is the minimal polynomial of θ_1 over K. Since $m_1(X)$ is a factor of $f(X)$ and $f(X)$ also splits in $E'[X]$, then $m_1(x)$ must split in $E'[X]$. Choose a root $\gamma_1 \in E'$, so that $K(\gamma_1) \cong K[x]/(m_1(x))$ as well. This provides a K-isomorphism $\varphi_1 : K(\theta_1) \to K(\gamma_1)$ sending θ_1 to γ_1.

Now let $m_2(X)$ be the minimal polynomial of θ_2 over $K(\theta_1)$, look at $\varphi_1\big(m_2(X)\big)$ in $K(\gamma_1)[X]$. Choose root $\gamma_2 \in E'$, repeat ideas and extend φ_1 to a K-isomorphism $\varphi_2 : K(\theta_1, \theta_2) \to K(\gamma_1, \gamma_2)$ sending θ_j to γ_j. Continue this process until we have chosen roots γ_j in E' for $f(X)$ and a K-isomorphism $\varphi_n : E = K(\theta_1, \ldots, \theta)n) \to K(\gamma_1, \ldots, \gamma_n) = E'$ sending each θ_j to γ_j. □

Uniqueness of splitting fields has an immediate application to finite fields.

Proposition. Two finite fields of the same cardinality are isomorphic.

Proof. Suppose E is a finite fields of size q. We saw previously that $q = p^m$ for some prime p, and that $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ is a subfield of E. Fermat says that every $\theta \in E$ satisfies $\theta^q = \theta$. Consequently, E is a splitting field for $X^q - X$ over $\mathbb{F}_p$. The uniqueness of splitting fields then implies that all fields of q elements are isomorphic. □

We don't need to stop with one polynomial. Define the notion of a splitting field of a family of polynomials $f_j(X)$ as j varies in some index set. Imagine some huge field A that contains a splitting field of every polynomial in $K[X]$, or even for every polynomial in $A[x]$. Is there such a field extension in which every polynomial splits? (Why not just use Zorn to get a maximal element in the set of all algebraic extensions of K?)

Definition. Let K be a field. An *algebraic closure* of K is a field A containing K such that
(1) A/K is an algebraic extension, and
(2) A is algebraically closed.

Condition (2) means: Every irreducible in $A[X]$ has degree 1, or equivalently: A admits no proper algebraic extension field.

For example, $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$, but not of $\mathbb{Q}$. An algebraic closure of $\mathbb{Q}$ does exist inside $\mathbb{C}$. The set $A = \{\alpha \in \mathbb{C} : \alpha \text{ is algebraic over } \mathbb{Q}\}$ of all algebraic numbers is algebraically closed and algebraic over $\mathbb{Q}$.

Existence of algebraic closures can be proved, but the logic is somewhat trickier that a standard Zorn's Lemma argument. Furthermore, any two algebraic closures of a field K are K-isomorphic. This uniqueness justifies the terminology "*the* algebraic closure of K" and use the notation K^{alg} to refer to that field.
Here is a key property of algebraic closures.

Lemma. For any algebraic extension L/K, there is K-embedding of L into K^{alg}.

Proof idea. More generally, suppose $\sigma : K \to E$ is a homomorphism of fields and E is algebraically closed.

To prove: L/K is algebraic there is a homomorphism $\widehat{\sigma} : L \to E$ extending σ.
Consider the special case that $L = K(\theta)$ so that $L \cong K[X]/(m_\theta(X))$. Homomorphism σ extends $\sigma : K[X] \to E[X]$ by fixing X. That is, $f = \sum_j c_j X^j$ is sent to $\sigma(f) = \sum_j \sigma(c_j) X^j$. Since E is algebraically closed, $\sigma(m_\theta) \in E[X]$ has a root $\alpha \in E$. Compose $\sigma : K[X] \to E[X]$ with the evaluation $E[X] \to E$ sending $X \to \alpha$. This yields a homomorphism $\sigma' : K[X] \to E$ whose kernel is the ideal (m_θ). This induces the extension $\widehat{\sigma} : K(\theta) \cong K[X]/(m_\theta(X)) \to E$ sending $\theta \mapsto \alpha$.
For general algebraic extensions L/K, express L as a tower of simple extensions $K = K_0 \subset K_1 \subset \cdots \subset L$ where $K_n = K_{n-1}(\theta_n)$. Then we can apply the special case at each step. This probably requires a Zorn's Lemma argument since $[L : K]$ might be infinite. □

If $[L : K] = n$, how many K-embeddings $L \to K^{\text{alg}}$ are there?
For example, when $L = K(\theta)$ the minimal polynomial $m_\theta(X) \in K[X]$ factors in K^{alg} as: $m_\theta(x) = \prod_{j=1}^{n}(x - \gamma_j)$. The proof above shows that each γ_j yields an K-embedding $L \to K^{\text{alg}}$. Conversely, any K-embedding σ is determined by the value $\sigma(\theta)$, and that is a root of $\_th(X)$. Therefore: the number of K-embeddings $L = K(\theta) \to K^{\text{alg}}$ equals the number of distinct roots of $m_t h(X)$ in K^{alg}.

Repeating this for a tower of simple extensions, we conclude that:

$$\#\{K\text{-embeddings } L \to K^{\text{alg}}\} \leq [L : K].$$

Is it possible for the number of embeddings to be strictly less than the degree? That would arise if there is an irreducible $m(X) \in K[X]$ such that $m(x)$ has fewer distinct roots in K^{alg} than its degree. That is: $m(X)$ has repeated roots. This is impossible over $\mathbb{Q}$ but it can happen when the characteristic is $p > 0$. Repeated factors of a polynomial $f(x)$ can be detected with its derivative.

Definition. The formal derivative $d : K[X] \to K[X]$ to is the K-linear map with $d(x^n) = nx^{n-1}$ for every n. We often write f' instead of $d(f)$.

Exercise. (1) Deduce the product rule $d(fg) = d(f)g + fd(g)$.
(2) Does the chain rule hold true in $K[X]$?
(3) Extend d to rational functions $K(X)$ and verify the quotient rule.
(4) Does these ideas extend to $K[[X]]$ and $K((X))$ (the field of formal Laurent series)?
(5) **Lemma.** If $f \in K[X]$ then:

$$f \text{ has a repeated factor in } K^{\text{alg}} \iff \gcd(f, f') \neq 1.$$

Suppose $\pi(X)$ is monic irreducible in $K[X]$ with a repeated root in K^{alg}. Then $(\pi, \pi') \neq 1$. Irreducibility implies π divides π' and therefore $\pi' = 0$ since $\deg \pi > \deg \pi'$. Expressing $\pi(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$ we find that every $ja_j = 0$ in K. Then $a_j = 0$ whenever $j \neq 0$ in K. In characteristic zero, this holds for every $j > 0$, and $\pi(X)$ is a constant, contrary to irreducibility. Then K has characteristic p and: if a_j is non-zero, then $p|j$. We have proved:

Lemma. Suppose $\pi(X) \in K[X]$ is irreducible. Then:
π has a repeated root in $K^{\text{alg}} \iff \text{char}(K) = p$ and $\pi(X) = g(X^p)$ for some $g \in K[X]$.

Suppose K is a field with characteristic p. Then $\sigma : \alpha \mapsto \alpha^p$ is a field automorphism. (Exercise: Show that $(a + b)^p = a^p + b^p$.) Consequently the image $\sigma(K) = K^p = \{\alpha^p : \alpha \in K\}$ is actually a subfield of K!

Exercise. (1) If K is finite then $K^p = K$.
(2) Let $K = \mathbb{F}_p(X)$ be the field of rational functions over $\mathbb{F}_p$. Then $X \notin K^p$. Find the degree $K : K^p]$ in this case.
[Note: $X \in K^p$ implies existence of monic polynomials $f, g \in \mathbb{F}_p[X]$ with $Xg(X^p) = f(X^p)$.]

These ideas lead to the theory of "inseparable extensions." A polynomial $f \in K[X]$ is *separable* if it has distinct roots in K^{alg}. An element of an extension field of K is *separable* if its minimal polynomial is separable. An algebraic extension is *separable* if all of its elements are separable. In characteristic zero, every algebraic extension is separable.

If a finite extension L/K is separable then:

$$\#\{K\text{-embeddings } L \to K^{\text{alg}}\} = [L : K].$$

Now let's return to ordered fields for a few minutes, to lead toward ideas in the next lecture. Suppose (K, P) an ordered field.

Exercise. (1) Suppose (L, Q) is an extension of (K, P) in the category of ordered fields. That is, $K \subseteq L$ is a subfield, and a positive element of K remains positive in L. That is, $P \subseteq Q$. Prove: $P = Q \cap K$.
(2) Let $K = \mathbb{Q}(\sqrt{2})$ a subfield of $\mathbb{R}$. Then $P^+ = \mathbb{R}^2 \cap K$ is an ordering of K. Define $P^- = \overline{P^+}$, the set of $\alpha \in K$ such that $\overline{\alpha} \geq 0$ in $\mathbb{R}$. Then K has exactly two orderings, namely P^+ and P^-.

8 Extensions of Ordered Fields; Real Closed Fields

It's not denial. I'm just very selective about what I accept as reality.
- Calvin ("Calvin and Hobbes")

Let K be an ordered field. Artin showed that $\sum K^2$ is precisely the intersection of all orderings P of K. This means that an element $a \in K$ is expressible as a sum of squares in K if and only if a is positive at every ordering of K. The proof of this result uses the idea of a *preordering* (a concept introduced by Serre many years after Artin's paper).

Definition Let K be a field. A *preordering* of K is a subset T such that:
$K^2 \subseteq T, \quad T+T \subseteq T, \quad T \cdot T \subseteq T, \quad \text{and} \quad -1 \notin T.$

Note that K is formally real if and only if K has a preordering. In this case, K has characteristic zero so $\mathbb{Q} \subseteq K$.

Lemma. Every preordering of K is contained inside of some ordering of K. In particular, any formally real field has an ordering.

Proof idea. Apply Zorn to the set of preorderings of K that contain T, to get a maximal element P. To prove: If $a \in K$ then either $a \in P$ or $-a \in P$. We did in a previous class. □

Then a preordering T is the intersection of all orderings containing it.

In 1888 Hilbert showed that some PSD polynomials have no expression as a sum of squares of polynomials. (This was illustrated explicitly in 1967 by Motzkin's polynomial.)

Exercise. Define Motzkin's polynomial $M(x,y) = 1 - 3x^2y^2 + x^2y^4 + x^4y^2$. Prove the following:
(1) M is PSD. [Hint. Geom Mean $\leq$ Arith Mean.]
(2) M cannot be expressed as a sum of squares in $\mathbb{R}[x,y]$.
(3) M is a sum of squares of rational functions.
[Note: $(x^2+y^2)^2 \cdot M(x,y) = x^2y^2(x^2+y^2+1)(x^2+y^2-2)^2 + (x^2-y^2)^2$. Deduce M is a sum of 4 squares.]

In 1893 Hilbert showed that every PSD polynomial in $\mathbb{R}[x,y)$ (2 variables) is a sum of squares of in $\mathbb{R}(x,y)$, the field of rational functions (quotients of polynomials). This led him to conjecture the result generally, as formulated in 1900 for his list of problems for the next century:

Hilbert's 17th Problem. If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is PSD, (i.e. $f(c) \geq 0$ for all $c \in \mathbb{R}^n$), then is f a sum of squares in $\mathbb{R}(x_1, \ldots, x_n)$?

Artin reformulated Hilbert's problem: If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is PSD, show that f is positive at every ordering of $\mathbb{R}(x_1, \ldots, x_n)$.

Exercise: Orderings of $\mathbb{R}(x)$. (one variable)

(1) Define P^+ to be the ordering of $\mathbb{R}(x)$ generated by polynomials with positive leading coefficient. [Clarify that definition and check that P^+ is an ordering.]
In this ordering, x is positive infinite relative to $\mathbb{R}$ (i.e. $x > c$ for every $c \in \mathbb{R}$).

(2) Find an ordering P^- for which x is negative infinite. Find one where x is positive infinitesimal (i.e. $0 < a$ in $\mathbb{R}$ implies $0 < x < a$).
Given $c \in \mathbb{R}$, find two orderings for which x is infinitesimally close to c.

(3) **Lemma.** If P is an ordering of field K and if $\sigma \in \mathrm{Aut}(K)$ is a field automorphism, then $\sigma(P)$ is also an ordering of K.
Apply this to automorphisms σ of $\mathbb{R}(x)$ and the ordering P^+ to generate the orderings mentioned in part (2).

(4) If P is an ordering of $\mathbb{R}(x)$ then it must be one of the ones constructed above. That is, P extends the usual ordering of $\mathbb{R}$ and x is "centered" at some $c \in \mathbb{R} \cup \{\infty\}$.

The two orderings on $\mathbb{Q}(\sqrt{2})$ are related by the "bar" automorphism. Viewing things a bit differently, the two orderings of $\mathbb{Q}(\sqrt{2})$ are explained by the two embeddings of that field into $\mathbb{R}$. We will prove later that if $K/\mathbb{Q}$ is a finite extension then the orderings on K all arise from real embeddings $\sigma : K \to \mathbb{R}$.

Exercise. Suppose (R, Q) is an ordered field and $\sigma : K \to R$ is a homomorphism of fields. Define the induced ordering P on K.
When σ is an inclusion map $K \subseteq R$, this is just $P = K \cap Q$.

When does an ordering on K extend to an extension field?

Lemma. If (K, P) is an ordered field and L/K is an extension, then:
P extends to an ordering of L if and only if $-1 \notin \sum L^2 P$.

Proof. By our work on preorderings, we know that P is a subset of some ordering of L if and only if $(\sum L^2)P$ is a preordering of L. This happens exactly when $-1 \notin \sum L^2 P$. □

This situation is not hard to analyze for quadratic extensions.

Proposition. Suppose (K, P) is an ordered field and $a \in K$ is a non-square.
(1) P extends to $K(\sqrt{a}) \iff a \in P$.
(2) $K(\sqrt{a})$ is formally real $\iff -a \notin \sum K^2$.

Proof. (1) By the Lemma we need to determine when $-1 \in \sum L^2 P$. That says there exist $r_i, s_i \in K$ and $p_i \in P$ such that

$$-1 = \sum (r_i + s_i\sqrt{a})^2 p_i.$$

A bit of algebra yields: $\underbrace{1 + \sum r_i^2 p_i}_{\in P} = -a \underbrace{\sum s_i^2 p_i}_{\in P}$, and we conclude that $-a \in P$. Conversely,

$-a \in P$ implies $-1 = \frac{1}{a}\big(-a\big) \in L^2P$. We have proved: P does not extend to L iff $-a \in P$.

(2) By Artin's work, $L = K(\sqrt{a})$ is not formally real if and only if no ordering extends to L. By (1) this says that $-a \in P$ for every P. This is equivalent to saying $-a \in \sum K^2$.
Exercise. Prove (2) directly from the definition of "formally real, without using (1). □

The next proposition is one of our favorites! It yields extensions for extensions of odd degree. The result was proved by Artin and Schreier in the context of formally real fields, and was extended by T. A. Springer for extensions of quadratic forms.

Proposition Let (K, P) be an ordered field, and let L/K be an odd degree extension. Then P extends to an ordering of L. In particular, if K is formally real and L/K is an odd degree extension, then L is formally real.

We will prove the "in particular" part here. The more general result is similar, and you can fill in the details.

Proof. Consider all odd degree extensions of formally real fields. If the proposition fails, there is a counterexample L/K of smallest degree. Minimality implies that L/K has no proper intermediate fields. (If $K \subset E \subset L$ then E/K or L/E is a smaller counterexample.) In particular, if $L = K(\theta)$ for some θ.

Let $g(x)$ be the minimal polynomial of θ over K, so that $L = K(\theta) \cong K[x]/(g(x))$, where $n = \deg g(x) = [L : K]$ is odd. Since L is not formally real there exist some $\alpha_j \in L$ such that $-1 = \sum \alpha_j^2$. Express $\alpha_j = f_j(\theta)$ for some $f_j(x) \in K[x]$ of degree $< n$. Then for some $Q(x) \in K[x]$,

$$-1 \;=\; \underbrace{\sum_{j=1}^{s} f_j(x)^2}_{\deg \,\le\, 2(n-1)} \;+\; \underbrace{g(x)}_{\deg \,=\, n}\; Q(x). \qquad (*)$$

Then $\deg Q(x) \le n-2$.
Claim. $\sum f_j(x)^2$ has even degree. Let $d = \max_j \big(\deg(f_j)\big)$ and let c_j be the x^d-coefficient of $f_j(x)$. Then the x^{2d}-coefficient of $\sum_j f_j(x)^2$ equals $\sum_j c_j^2$. This is nonzero, since K is real, proving the Claim.

That claim, and equation (*) imply that $Q(x)$ has odd degree $\le n-2$. Let $\pi(x)$ be an irreducible odd degree factor of $Q(x)$ in $K[x]$. Then

$$-1 \equiv \sum_{i=1}^{s} f_i(x)^2 \qquad (\mathrm{mod}\ \pi(x)).$$

This shows that the field $E = K[x]/(\pi(x))$ is not formally real. But $[E : K] < n$, contradicting the minimality of $[L : K]$. □

As an exercise, prove the more general version of this result: Suppose L/K is a finite extension

of odd degree, then every ordering P of K can be extended to L.

Recall that an algebraic closure of K is a maximal algebraic extension of K. Artin and Schreier introduced an analogous idea: The "real closure" of an ordered field (K, P). That is defined as a maximal algebraic extension of K to which P extends.

Definition A field R is said to be *real-closed* if R is formally real and every proper algebraic extension of R is not formally real.

Choose an ordering P of such a field R. If $a \in P$, a proposition above implies that P extends to $R(\sqrt{a})$, so by maximality: $\sqrt{a} \in R$. Therefore $P = R^2$, and R has a unique ordering $R^2 = \sum R^2$.

If $\pi(x) \in R[x]$ is irreducible of odd degree, then $L = R[x]/(\pi(x))$ is a field with $[L : R]$ is odd. By the odd degree result above, L is formally real and maximality implies $L = R$. This says that $\pi(x)$ is a constant and R has no proper odd degree extensions. We summarize these properties shared by any real closed field R:

1. Every proper algebraic extension of R is not formally real.
2. R^2 is the unique ordering on R.
3. Every odd degree polynomial in $R[x]$ has a root in R.
4. R has no proper odd degree extensions.

This field R is beginning to look a lot like $\mathbb{R}$, the field of real numbers!

Exercise. Prove the intermediate value theorem (usually done in calculus) for polynomial functions in $R[x]$. That is, if $f(x) \in R[x]$ and $f(a) < 0 < f(b)$ in R, show that there exists $c \in R$ with $a < c < b$, such that $f(c) = 0$.
Is this statement still true for rational functions, in $R(x)$?

We now investigate algebraic extensions of a real closed field R. Let $i = \sqrt{-1}$ and consider $C = R(i)$. We know that $[C : R] = 2$, and we will show that C extension is the unique proper algebraic extension of R. First we settle the quadratic case.

Lemma. Suppose R is a real closed field and $C = R(\sqrt{-}1)$. Then C has no quadratic extensions.

Proof. We want to prove that every $\alpha \in C$ is a square. Express $\alpha = a + bi$ for $a, b \in R$ and search for $x, y \in R$ such that $\alpha = (x + yi)^2$.
Method #1. One approach is to show directly that the equations $a = (x^2 - y^2)$ and $b = 2xy$ have a solution $x, yinR$. We leave this as an exercise.

Method #2. We use the norm N, recalling an earlier lemma: $N(\delta) = 1 \iff \delta = \frac{\gamma}{=}{\gamma}$ for some

γ. In our case, $R^2 = \sum R^2$ so there exists $c \in R$ with $N(\alpha) = c^2$. Then $N(\alpha/c) = 1$ and the Lemma provides a $\gamma \in C$ with $\alpha/c = \gamma/\overline{\gamma}$. Then $\alpha = c\gamma^2/N(\gamma)$ which is a square in C. □

Theorem. Let R be a real closed field, then $C = R(i)$ is algebraically closed. (!)

Proof idea. This proof requires some "black boxes": Galois Theory, the Sylow Theorem, and basic properties of finite 2-groups. To show that C has no proper algebraic extensions, suppose E/C is a finite extension field. We hope to prove $E = C$.

Then E/R as a finite extension, and we can adjoin all conjugates of elements of E to build a possibly bigger field F that is a Galois extension of R. So we have a tower of intermediate fields

$$R \subseteq C \subseteq E \subseteq F.$$

Let $G = \mathrm{Gal}(F/R)$ be the Galois group and let S be a Sylow 2-subgroup of G. Then $|S|$ is a 2-power and the index $(G : S)$ is odd. (Such an S exists by Sylow's Theorem.) Let K fixed field of the subgroup S, so that $R \subseteq K \subseteq F$. The Fundamental Theorem of Galois theory implies that $[K : R] = (G : S)$ is odd. Since R is real closed, it has no proper odd degree extensions, so $K = R$, and $S = G$. Then $|G| = 2^m$ and $[F : R] = 2^m$.

A theorem from group theory says that if p is a prime and group H has order $|H| = p^m > 1$, then H has a subgroup on index p. In our situation, Let $H = Gal(F/C)$. Then H is a 2-group (since it is a subgroup of G). If H is non-trivial it has a subgroup A of index 2 and the fixed field of A is an extension of C if degree 2, contrary to the Lemma. Therefore H is trivial and $F = C$ and therefore $E = C$. This proves that C is algebraically closed. □

This argument (due to Artin) provides an algebraic way to prove a version of the Fundamental Theorem of Algebra.

9 Uniqueness of Real Closure

You are unique, just like everybody else.

Last class we defined a real closed field as a maximal formally real field:
Definition. A field K is *real closed* if it is formally real, and no proper algebraic extension of K is formally real.

Proposition: Let K be a field. The following are equivalent.
(1) K is real closed.
(2) K^2 is an ordering of K, and K has no proper odd degree extension.
(3) K is formally real and $K(\sqrt{-1})$ is algebraically closed.
(4) K is formally real and a quadratic extension of K is algebraically closed.

The step (3) $\Rightarrow$ (4) is trivial. Last time we proved (1) $\Rightarrow$ (2) $\Rightarrow$ (3). Most of the hard work was done in the proof that (2) $\Rightarrow$ (3), using Theorems from Galois theory and group theory.

(4) $\Rightarrow$ (3): Suppose K is real and $L = K(\sqrt{d})$ is algebraically closed. We want to show: $\sqrt{-1} \in L$, for then $L = K(\sqrt{-1})$. By hypothesis, every element of L is a square. In particular *sqrtd* is a square in L, so its norm $N(\sqrt{d}) = -d$ is a square in K. Therefore -1 is a square in L, as claimed.

(3) $\Rightarrow$ (1): To show: K has no proper real extensions.
Suppose E/K is an algebraic extension and $E \neq K$. Since any algebraic extension embeds in $K^{\text{alg}} = L$ we may assume that $E \subseteq L$. But $[L : K] = 2$, so we must have $E = L$. Then E is not formally real. □

In summary, suppose an ordered field K has the properties:
every positive element of K is a square, and
every odd degree polynomial in $K[x]$ has a root in K.
Then $K(\sqrt{-1})$ is algebraically closed.

Remark These results of Artin and Schreier from the 1920s have been teased apart in various ways since then. For instance a field K is called *Pythagorean* if $\sum K^2 = K^2$. It is called *Euclidean* if K^2 is an ordering of K. In his investigations of the foundations of Euclidean Geometry, Hilbert considered what sorts of fields would be appropriate to parametrize the plane. Seeds of the ideas about these sorts of fields were planted in his work.

We mentioned previously that algebraic closures of K exist, and that they are unique up to K-isomorphism. We consider an analogous notion for real fields.

Definition Let (K, P) be an ordered field. A *real-closure* of (K, P) is an algebraic extension Δ that is real closed and whose ordering extends P.

Does a real closure of (K, P) always exist? Consider algebraic extensions in the category of ordered fields. For set-theoretical purposes, we fix an algebraic closure K^{alg} and consider exten-

sions (L, Q) such that $K \subseteq L \subseteq K^{\text{alg}}$. The term "extension" means that $K \subseteq L$ and Q extends P (i.e. $K \cap Q = P$). Zorn implies that a maximal element (K_1, P_1) exists. The next Lemma will help show that (K_1, P_1) is real-closed.

Lemma. Let (K, P) be an ordered field.
(1) P extends to an ordering of $K(\sqrt{a})$ if and only if $a \in P$.
(2) P extends to an ordering of L for any odd degree extension L/K.

Conveniently for us, we proved this last time! This Lemma and the maximality of (K_1, P_1) imply: Every $a \in P_1$ is a square in K_1, and K_1 has no proper odd degree extensions. By item (2) of the Proposition, we find. Thus real closures always exist.

Are real closures of (K, P) unique? Algebraic closures of a field K all always K-isomorphic. The analogue for real closures would be K-isomorphisms that preserve the orderings. If S is real closed, we often don't specify the ordering in the notations (since the unique ordering of S is S^2).

The key to proving that uniqueness is a (non-trivial) result on extensions.

Proposition. Suppose (K, P) is an ordered fields, and $\varphi : (K, P) \to S$ is an order-embedding into a real closed field S. If (L/Q) is a finite extension of (K, P), then φ extends to an order-embedding $\widehat{\varphi} : L \to S$.

The hypothesis is that L/K is a finite extension of fields and the orderings are compatible: $P = K \cap Q$.

To prove this Proposition, we need a way to detect (count) "real roots" of polynomials $f \in K[x]$ by working entirely over K. That way, the number of real roots is independent of which real closure is chosen. In the Nineteenth Century a lot of work was done on calculating roots of polynomials (long before any sort of machine calculations were conceived). For instance, if a polynomial $f(x)$ has a root α in some interval $[a, b]$ then approximation methods were devised to obtain more accuracy. For instance, Horner's method or Newton's method, come to mind.

But how can we calculate whether a given polynomial $f(x)$ has a root in $[a, b]$? One way to check is to compute the values $f(a)$ and $f(b)$ and see whether they have opposite sign. If both of those values are positive there might still be roots in $[a, b]$. (Will there always be an an even number of roots in that case?)

An early result (published in 1637) is **Descartes' Rule of Signs**:
Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ in $\mathbb{R}[x]$. Let $\text{Var}(f) = \text{Var}(a_0, \ldots, a_n)$ be the number of "variations in sign" for the sequence of coefficients $(a_0, a_1, \ldots, a_n)$. For example, the sequence $(1, 0, -3, 0, -5, 4)$ has 2 variations (ignore zeros and count adjacent pairs of opposite sign).

Descartes' Rule: #(positive real roots of f) $\leq$ Var(f). The difference is an even number.

To detect negative real roots, apply Decartes' Rule to $f(-x)$.

Exercise. Prove Descartes' Rule of Signs.

By the 19$^{\text{th}}$ century several mathematicians had considered the general problem of "separation of roots" :

If $f(x) \in \mathbb{R}[x]$, compute the number of roots of f in a given interval (a, b)?

The definitive result was proved by Sturm in 1829. His method involves the euclidean algorithm for f and its derivative f'. To simplify things we assume here f is separable, that is: f has no repeated factors. Equivalently, f and f' are relatively prime.

Given a separable polynomial $f \in \mathbb{R}[x]$, perform the Euclidean Algorithm as follows.

$$\begin{aligned} g_0 = f, \quad g_1 = f', \quad &\text{and:} \\ g_0 &= g_1 Q_2 - g_2 \\ g_1 &= g_2 Q_3 - g_3 \\ \vdots \quad &\quad \vdots \quad \vdots \\ g_{r-2} &= g_{r-1} Q_r - g_r \\ g_{r-1} &= g_r Q_{r+1} \end{aligned}$$

Here $\deg(g_k) < \deg(g_{k-1})$ for each $k = 1, 2, \ldots, r$. For $c \in \mathbb{R}$, define

$$W(c) = \text{Var}\big(g_1(c), \ldots g_r(c)\big).$$

Sturm's Theorem. If (a, b) is an interval in $\mathbb{R}$, then the number of roots of f in the interval (a, b) equals $W(a) - W(b)$.

Of course, the Theorem implies that if $a < b$ then $W(a) \geq W(b)$.

Proof idea: Since f is separable, no consecutive pair g_i, g_{i+1} has a common root. Separate (a, b) into subintervals by using all real roots of all the g_i's. No sign changes can occur within any sub-interval. Moreover, check that $W(t)$ doesn't change as t passes a root of g_k for $k > 0$.
[To see this, suppose $g_k(\alpha) = 0$. Then $g_{k-1}(\alpha)$ and $g_{k+1}(\alpha)$ have opposite sign. Examine $W(t)$ as t increases past α.]

On the other hand, suppose α is a root of $g_0 = f$. If $g_1(\alpha) = f'(\alpha) > 0$, then $f(t)$ increases near α, and we list sign possibilities.

	$f(t)$	$f'(t)$	$g_2(t)$	$\cdots$
$t < \alpha$	$-$	$+$	$\cdots$	
$t = \alpha$	0	$+$	$\cdots$	
$t > \alpha$	$+$	$+$	$\cdots$	

Therefore $W(t)$ decreases by 1 as t passes α. In the case $f'(\alpha) < 0$, a similar result follows. This completes the outline of the proof. $\square$

Artin's observed that Sturm's theorem holds true for any real closed field rather than just the field $\mathbb{R}$. In fact, if (K, P) an ordered field with real closure S, and $f \in K[x]$, then Sturm counts the number of roots f has in a given interval (a, b) in S.

Proof of ***Proposition*** *on extending embeddings.* Given an order-embedding $\varphi : (K, P) \to S$ where S is real closed, and given a finite extension $(L, Q)/K$, we want to extend φ to an order embedding $\widehat{\varphi} : (L, Q) \to R$. To simplify notations, let's replace K by $\varphi(K) \subseteq L$ and assume that φ is just the inclusion map $K \hookrightarrow L$. Then we are looking a K-embedding $L \to S$ that respects the orderings (sending Q into S^2).

As usual, it suffices to settle the case $L = K(\beta) \cong K[x]/(g(x))$, where $g(x)$ is irreducible and β corresponds to the class of x. Existence of a K-embedding $\widehat{\varphi} : L = K(\beta) \to S$ is equivalent to existence of a root of $g(x)$ in S. [Why does a root $\beta' \in S$ yields a K-embedding $\widehat{\varphi}$ with $\widehat{\varphi}(\beta) = \beta'$?]

Claim. $g(x)$ has root $\beta' \in S$.
Proof. Since β is algebraic over K, there exist a, b in K such that $a < \beta < b$. [Why?] Let R be a real closure of (L, Q). Then Sturm's theorem for R implies that $W(a) - W(b) > 0$. A key point: The value $W(a) - W(b)$ is computed entirely inside K. Sturm's theorem for S implies that $g(x)$ has a root β' in S, and proves the Claim.

Now we have a K-embedding $\widehat{\varphi} : K(\beta) \to S$, but it does not necessarily preserve orders. But Artin has a trick up his sleeve! Let $\sigma_1, \ldots, \sigma_s$ be the list of all K-embeddings $L \to S$. If one σ_i preserves the orderings (i.e. $\sigma_i(Q) \subseteq S^2$), then we are done. If that's false then for each j there exists $a_j \in Q$ such that $\sigma_j(a_j)$ is negative in S. To derive a contradiction, let $E = L(\sqrt{a_1}, \ldots, \sqrt{a_s})$. Since each a_j is in Q, that ordering on L extends to an ordering Q' on E. By the first part of this proof applied to E/K, there exists a K-embedding $\sigma : E \to S$. That σ restricts to an embedding $L \to S$, so that $\sigma|_L = \sigma_i$ for some index i. But a_i is a square in E, so that $\sigma_i(a_i)$ must be a square in S, hence positive. This contradicts our assumption that each $\sigma_j(a_j)$ is negative in S. □

Theorem. Real closures of an ordered field (K, P) exist. Any two of those real closures are equivalent, via an order preserving K-isomorphism.

10 Artin's Theorem Solving the 17^{th} Problem

If 90% of the ideas you generate aren't absolutely worthless,
then you're not generating enough ideas. - Mike Artin (son of Emil)

In this class we will outline a proof of Artin's theorem that solves Hilbert's 17^{th} Problem (but we'll leave a large step unproved!)

Before that. let's talk for a moment about real closures. Recall that a real closure of an ordered field (K, P) is a maximal element in the set of ordered fields (L, Q) such that L/K is algebraic and Q extends P. Such a maximal element R is a real closed field. An important result was the Theorem that all real closures of (K, P) are K-isomorphic.

At one point in that uniqueness proof we used the following Lemma stating that no element of R is infinitely large relative to K. Similarly, no element of R is infinitesimal compared to K.

Lemma. If R is a real closure of (K, P) and $\theta \in R$, then there exist $a, b \in K$ such that $a < \theta < b$.

Proof. It suffices to prove that $\theta < b$ for some $b \in K$. (To get a lower bound, apply this to $-\theta$.) We assume $\theta > 1$, for otherwise we just use $b = 1$.

Define absolute value $|\alpha|$ on R, and check that: $|\alpha\beta| = |\alpha||\beta|$ and $|\alpha + \beta| \leq |\alpha| + |\beta|$.

Since θ is algebraic there is a relation: $\theta^n + a_{n-1}\theta^{n-1} + \cdots + a_1\theta + a_0 = 0 \quad$ for some $a_j \in K$.
Then $\theta^n = \left|a_{n-1}\theta^{n-1} + \cdots + a_1\theta + a_0\right| \leq \big(|a_{n-1}| + \cdots + |a_1| + |a_0|\big)\theta^{n-1}$, since $\theta > 1$. Therefore $\theta < b$ where $b = 1 + |a_{n-1}| + \cdots + |a_1| + |a_0|$. □

The Lemma indicates that elements of the real closure R are mixed with elements of K. But that does **NOT** imply that K is dense in R.

An interval (α, β) in R might contain no elements of K!

Exercise. For $K = \mathbb{R}(x)$ let P be the ordering that contains all polynomials with positive leading coefficient. Check that this does define an ordering P of K and that x is infinitely large relative to $\mathbb{R}$. (We worked with this ordering before.)
Let $L = K(\sqrt{x})$ and choose an ordering Q of L extending P. (This exists since $x \in P$.)
Prove: The interval $(\sqrt{x}, \sqrt{x} + 1)$ in L contains no elements of K.
Since K is not dense in L, it is not dense in its real closure.

Now we return to ...

Hilbert's 17^{th} Problem. Let $K = \mathbb{R}(x_1, \ldots, x_n) = \mathbb{R}(X)$ be the field of rational functions in n variables. Then $f \in K$ is PSD if and only if f is a sum of squares in K.

This question was a major motivation for the development of the Artin-Schreier theory of ordered fields. We will outline a version of Artin's affirmative answer to Hilbert's question.

Artin's Theorem. (1927) Suppose (K, P) is an ordered field with real closure R. Write $K(X)$ for the rational function field over K in n variables. For polynomial $f \in K[X]$, the following are equivalent:

(1) f is PSD, that is: $f(a) \geq 0$ for every $a \in R^n$.

(2) $f \in \sum PK(X)^2$, that is: $f = \sum_{i=1}^{m} p_i f_i(X)^2$, for some $p_i \in P$ and $f_i(X) \in K(X)$.

If K has a unique ordering then $P = \sum K^2$, and condition (2) above says that $f \in \sum K(X)^2$.

Our proof uses the Artin-Lang Homomorphism Theorem (a hard result formulated long after Artin's original proof). That Theorem is an analog of Hilbert's Nullstellensatz but over real closed fields, rather than algebraically closed fields. To begin, let's recall the background of the Nullstellensatz.

Suppose $f_1, \ldots, f_s \in \mathbb{C}[X]$ are polynomials in n variable over a field K. Their set of common zeros is a set $\mathcal{Z}(f_1, \ldots, f_s) \subseteq K^n$, defined as:

$$\mathcal{Z}(f_1, \ldots, f_s) = \{a \in K^n : f_1(a) = \cdots = f_s(a) = 0\}.$$

A bit of thought shows that this zero-set depends only on the ideal J generated by the f_j. Any set of generators of that ideal yields the same zero set. So if $J \subseteq K[X]$ is an ideal we define

$$\mathcal{Z}(J) = \{a \in K^n : f(a) = 0 \text{ for every } f \in J\}.$$

Over the field $\mathbb{C}$ of complex numbers, and one polynomial $f(X)$, think of the zero set $\mathcal{Z}(f)$ as a hyper-surface (of dimension $n-1$) in $\mathbb{C}^n$. For example, the unit sphere S^2 is the zero set $\mathcal{Z}(f)$ for $f = x_1^2 + x_2^2 + x_3^2 - 1$. This is a 2-dimensional surface in 3-space.
Then $\mathcal{Z}(f_1, \ldots, f_k) = \mathcal{Z}(f_1) \cap \cdots \cap \mathcal{Z}(f_k)$ is the intersection of s hyper-surfaces in $\mathbb{C}^n$. Once a good definition of "dimension" for such sets is defined, we would hope that $\dim\big(\mathcal{Z}(f_1, \ldots, f_k)\big) \geq n - s$. Ideas of dimension don't make sense over general fields, since even for one polynomial f the set $\mathcal{Z}(f)$ can be empty. For instance, consider $1 + x_1^2 + \cdots + x_n^2$ over $\mathbb{R}$.

Aside Remark: A list of finitely many f_i's is sufficient to describe a set $\mathcal{Z}(J)$, because the Hilbert Basis Theorem says: Every ideal of $K[X]$ is finitely generated.

For the other direction, let S be a subset of K^n and consider the set of polynomials that kill S:

$$\mathcal{I}(S) = \{f \in K[X] : f \text{ vanishes on } S\}.$$

Check that this is an ideal in $K[X]$.

Exercise. If $J_1 \subseteq J_2$ then $\mathcal{I}(J_1) \supseteq \mathcal{I}(J_2)$. If $S_1 \subseteq S_2$ then $\mathcal{Z}(S_1) \supset \mathcal{Z}(S_2)$.
$J \subseteq \mathcal{I}\mathcal{Z}(J)$ and $S \subseteq \mathcal{Z}\mathcal{I}(S)$.
If $J = \mathcal{I}(S)$ for some S, then $J = \mathcal{I}\mathcal{Z}(J)$. If $S = \mathcal{Z}(J)$ for some J, then $S = \mathcal{Z}\mathcal{I}(S)$.

Exercise. (1) For the point $0 = (0, \ldots, 0)$ show that $\mathcal{I}(\{0\}) = (x_1, \ldots, x_n)$ is the ideal of all polynomials f with $f(0) = 0$.

(2) For $a = (a_1, \ldots, a_n) \in K^n$, define $M_a = (x_1 - a_1, \ldots, x_n - a_n)$. Then M_a is the kernel of the evaluation map $Eval_a : K[X] \to K$, so M_a is a maximal ideal. Check that $M_a = \mathcal{Z}(\{a\})$.

If J is the trivial ideal, with $1 \in J$, then obviously $\mathcal{Z}(J) = \varnothing$. The converse was proved by Hilbert, provided the coefficient field K is algebraically closed.

Hilbert's Nullstellensatz. (zero locus theorem) Let K be an algebraically closed field.
(a) If J is a proper ideal of $K[X]$ then $\mathcal{Z}(J) \neq \varnothing$.
(b) Every maximal ideal of $K[X]$ equals M_a for some $a \in K^n$.
(c) Any affine K-algebra A admits a K-homomorphism $A \to K$.
(d) The only affine K-algebra that is a field is K itself.

This theorem is hard to prove! Here we show only that those four statements are equivalent (for any field K). A lot more work is needed to prove that those statements hold true in case K is algebraically closed.

Terminology: A *K-algebra* is a ring containing K as a subring.
A *K-homomorphism* is a K-linear, ring homomorphism.
An *affine K-algebra* is a commutative, finitely generated, K-algebra. That is, $A \cong K[\theta_1, \ldots, \theta_n]$ for some commuting generators θ_j. Evaluation at $\theta = (\theta_1, \ldots, \theta_n)$ yields a surjective, K-homomorphism $Eval_\theta : K[X] \to A$. Setting $J = \ker(Eval_\theta)$ we find that $A \cong K[X]/J$.

Proof of Equivalence. $(a) \Rightarrow (b)$. For a maximal ideal M, part (a) provides a point $a \in \mathcal{Z}(M)$. Then $M \subseteq M_a$, and maximality implies $M = M_a$.

$(b) \Rightarrow (a)$. A proper ideal J is contained in a maximal ideal M (by Zorn). (b) implies $M = M_a$, so that $a \in \mathcal{Z}(M) \subseteq \mathcal{Z}(J)$.

$(a) \Rightarrow (c)$. Given $A \cong K[X]/J$ for ideal J of $K[X]$, there exists $a \in \mathcal{Z}(J)$, by (a). The evaluation map $Eval_a : K[X] \to K$ is a K-homomorphism, and $Eval_a$ kills J. Therefore it induces a K-homomorphism $A \cong K[X]/J \to K$, as claimed.

$(c) \Rightarrow (d)$. If A is an affine K-algebra, then (c) yields a K-homomorphism $\varphi : A \to K$. It is surjective since φ is K-linear. If A is a field then φ must be injective, hence an isomorphism.

$(d) \Rightarrow (b)$. If M is a maximal ideal of $K[X]$ then $A = K[X]/M$ is a affine K-algebra that is a field. Then (d) says that $A \cong K$, and that isomorphism lifts to a surjective K-homomorphism $\widehat{\varphi} : K[X] \to K$ with kernel M. Let $a = (a_1, \ldots, a_n)$ where $a_j = \widehat{\varphi}(x_j) \in K$. Check that $\widehat{\varphi}$ equals the evaluation map $Eval_a$. Then $M = \ker\left(\widehat{\varphi}\right) = M_a$. □

The Artin-Schreier theory of real-closed fields led to an analogue of the Nulstellensatz over a real-closed field R, extending the classical algebraically closed case. This theorem was formulated and extended by Serge Lang, who was one of Artin's students in the early 1950s.

Suppose the affine R-algebra $L = R[\theta_1, \ldots, \theta_n]$ is a field. Then $A = R[X]/M$ for some maximal ideal M of $R[X]$. If there is an R-homomorphism $\varphi : L \to R$ then certainly L must be formally

real. The converse is the big Theorem.

Artin-Lang Theorem. Suppose R is a real closed field. Then:
(1) An affine R-algebra A that is a subring of a formally real field, admits an R-homomorphism $A \to R$.
(2) If L is an affine R-algebra that is a formally real field, then $L \cong R$.

Assuming this hard theorem (a Black Box), we can immediately prove Artin's positive answer to Hilbert's Seventeenth Problem.

Proof of Artin's Theorem. We are given an ordered field (K, P) with real closure R, and a polynomial $f \in K[X]$ that is positive semi-definite on R.
Claim: $f \in \sum PK(X)^2$.

The set $T = \sum PK(X)^2$ is a preordering of $K(X)$, and equals the intersection of all orderings Q of $K(X)$ that extend P. Therefore, the Claim follows if we prove:
$f > 0$ for every such ordering Q of $K(X)$.

For contradiction, suppose $f < 0$ for such a Q and let S be a real closure of $(K(X), Q)$. Then $f = -w^2$ for some $w \in S$. Define the affine K-algebra $A = K[X, w, 1/w]$. Since $A \subseteq S$ the Artin-Lang theorem yields an R-homomorphism $\varphi : A \to R$. Let $a_i = \varphi(x_i) \in R$ and $a = (a_1, \ldots, a_n) \in R^n$. Then $\varphi(f(X)) = f(a)$. But this equals $\varphi(-w^2) = -\varphi(w)^2 < 0$ in R. (We know $\varphi(w) \neq 0$ since $\varphi(1/w) = 1/\varphi(w)$ is also in R.) That is $f(a) = \varphi(f) = -\varphi(w)^2 < 0$ in R, contrary to the hypothesis that f is PSD on R. This proves the Claim. □

As a special case of the proof we just outlined, if $f \in \mathbb{R}(X)$ is positive semidefinite over $\mathbb{R}$, then f is a sum of squares of rational functions. However, that proof is non-constructive. We deduced that f is a sum of squares by showing that it is positive in every ordering. That provides zero insight about the number of squares needed to represent f, or about bounds on the degrees (or on the coefficients) of the numerators and denominators in such an expression.

In the 1960s A. Pfister extended the theory of quadratic forms over fields to settle several old questions about sums of squares, and to raise interesting new questions. One of the triumphs of his work is a bound on the number of squares needed in Hilbert's 17^{th} Problem.

Pfister's Theorem. (1967) Suppose R is a real closed field and X is a system of n variables. Then any sum of squares in $R(X)$ is a sum of 2^n squares in $R(X)$.

Those ideas, and related topics from the algebraic theory of quadratic forms, might be discussed on another day.

11 Irreducibility of $X^n - a$

Three steps must be taken to solve this problem.
The first step: Decide on what the three steps should be. - Stephen Colbert

In the 1920s, Artin and Schreier worked together on real closed fields, the maximal ordered fields. With those examples in mind, they investigated a natural problem:

• Which fields K have the property that the algebraic closure K^{alg} is a finite extension of K? They obtained the wonderful result: We already know all fields of that type!

Recall that for a field R, the following statements are equivalent:

(1) R is real closed (i.e. R is formally real with no proper formally real extensions).

(2) R is an ordered field such that: every positive element of R is a square, and every odd degree polynomial has a root in R.

(3) R has a quadratic extension that is algebraically closed.

Theorem. (Artin-Schreier) Suppose L is algebraically closed and L/K is a finite extension. Then either $L = K$, or $L = K(\sqrt{-1})$ and K is real closed.

An equivalent statement is: If K is neither algebraically closed nor real closed, then there exist irreducible polynomials in $K[x]$ of arbitrarily large degree.

Exercise. If K is a finite field, prove that K has an extension of every degree n. Alternatively, for every n show: There is an irreducible polynomial of degree n polynomial in $K[x]$.

Knowing the result of that exercise, we concentrate on the case K is an infinite field. If p is a prime number, then Eisenstein's criterion (in $\mathbb{Q}[X]$) shows that $X^n - p$ is irreducible for every n. This method probably extends to $K[X]$ when K is the field of fractions of a principal ideal domain. However, there are fields not expressible that way. [Examples?]

Is there some way to test whether $X^n - a$ is irreducible in $K[X]$?

Our treatment (following Lang's *Algebra*) uses an 1898 result of Capelli, providing simple criteria to guarantee that $X^n - a$ is irreducible. We'll get to the statement after several preliminaries.

Lemma. If p is prime and $X^p - a$ has no root in K, then it is irreducible in $K[X]$.

Proof. Let α be a root of $X^p - a$. If β is also a root, then $(\beta/\alpha)^p = 1$, so we analyze roots of $X^p - 1$. If $\mathrm{char}K \neq p$, then $X^p - 1$ has no repeated factors. [Recall: f has no repeated factors iff f and f' are relatively prime.] Let $\zeta \neq 1$ be a root of $X^p - 1$ in some extension field. Then the roots of $X^p - 1$ are $1, \zeta, \cdots, \zeta^{p-1}$. If $\mathrm{char}K = p$, the polynomial $X^p - a$ has only one root: $X^p - a = (X - \alpha)^p$. In this case, set $\zeta = 1$.

Then (for any characteristic) our polynomial splits as: $\quad X^p - a \;=\; \prod_{j=0}^{p-1} (X - \zeta^j \alpha)$.

Let $g(X)$ be the minimal polynomial for α over K, so that $g(x)$ divides $X^p - a$. If $g(X) = X^p - a$ then we are done, so let's assume $g(X)$ is a proper divisor, of degree $r < p$. Then it is a product of a subset of those linear term:

$$g(X) = (X - \zeta^{j_1}\alpha) \cdots (X - \zeta^{j_r}\alpha) \text{ for } 1 \le j_1 < \cdots < j_r < p.$$

Let $(-1)^r c$ be the constant term, so that $c = \zeta^s \alpha^r \in K$ for some $s \in \mathbb{Z}^+$. Then $c^p = a^r$ in K.

Let $G = K^{*p}$ be the subgroup of p^{th} powers in K^*. Then $a \in K^*$ has $a^r \in G$ and $a^p \in G$. Since r, p are coprime it follows that $a \in G$. [For $a^{ru+pv} \in G$ for any $u, v \in \mathbb{Z}$. Choose u, v so that $ru + pv = 1$.] This contradicts the hypothesis that $X^p - a$ has no root in K. □

The next step is to extend those ideas to $X^{p^m} - a$, using results due to Capelli. If p is an odd prime, then there is a direct extension.

Proposition. Suppose K is a field, $a \in K$, and p is an odd prime.
If $a \notin K^p$, then for every $m \ge 0$, $\quad X^{p^m} - a$ is irreducible in $K[X]$.

Proof. If $m = 0$ the result is clear, and the Lemma proves the case $m = 1$. For an inductive (or WOP) proof, suppose $X^{p^m} - a$ is irreducible over K, for some $m \ge 1$. We will prove that $X^{p^{m+1}} - a$ is also irreducible.

Let θ be a root of $X^{p^m} - a$, and write N for the norm $N_K^{K(\theta)}$. Since $X^{p^m} - a$ is the minimal polynomial for θ, we know that $N(\theta) = a$. Since a is not a p^{th} power in K, it follows that θ is not a p^{th} power in $K(\theta)$. By the Lemma, $X^p - \theta$ is irreducible over $K(\theta)$. Let β be a root of $X^p - \theta$, so that $[K(\beta) : K(\theta)] = p$. Degrees of extensions multiply in towers, so $[K(\beta) : K] = p^{m+1}$. Since β is a root of $X^{p^{m+1}} - a$, we conclude that $X^{p^{m+1}} - a$ is irreducible over K. □

This result implies that $K[X]$ has irreducible polynomials of arbitrarily large degree, provided $K \ne K^p$ for some odd prime p.

Exercise. Find an example of K with $K = K^p$ for every odd prime p, but $K \ne K^2$.
[One idea: Let K be a subfield of $\mathbb{C}$, maximal such that $\sqrt{2} \notin K$. Why does such K exist? Why is every finite extension of 2-power degree? (Artin used this example to "dig holes" in an algebraic closure.)]

What about polynomials of the form $X^{2^m} - a$? Certainly if a is not a square in K then $X^2 - a$ is irreducible. What about $X^{2^m} - a$?
Example. $X^4 + 4$ factors in $\mathbb{Z}[X]$. (Look at $(X^2 - 2X + 2)(X^2 + 2X + 2)$.)
In fact, if $c \in K$ for a field K, then $X^4 + 4c^4$ factors in $K[X]$. So if we want $X^{2^m} - a$ to be irreducible for all m, we need to assume that both $X^2 - a$ and $X^4 - a$ are irreducible. The surprise is that this is sufficient.

Proposition. Let K be a field and $a \in K$.
If $a \notin K^2$ and $a \notin -4K^4$, then for every $m \ge 0$, $\quad X^{2^m} - a$ is irreducible in $K[X]$.

Proof. We follow the same path as before. Assume $m \ge 1$ and that the assertion holds true for m (over any field). In particular, $X^{2^m} - a$ is irreducible over K. Let θ be a root of $X^{2^m} - a$, so

that $[K(\theta):K] = 2^m$. Let β be a root of $X^2 - \theta$. Then $\beta^2 = \theta$, so that β is a root of $X^{2^{m+1}} - a$.

Claim, $X^2 - \theta$ is irreducible over $K(\theta)$. Equivalently, θ is not a square in $K(\theta)$.
If this is true, then the previous arguments prove the Proposition.

To prove the claim, suppose $\theta = \delta^2$ for some $\delta \in K(\theta)$. Let N be the norm from $K(\theta)$ to K and check that $-a = N(\theta) = b^2$ where $b = N(\delta) \in K$. Since $a \notin K^2$ we find that $-1 \notin K^2$. Setting $i = \sqrt{-1}$, then $K(i)$ is a quadratic extension of K. Now

$$X^{2^{m+1}} - a \;=\; X^{2^{m+1}} + b^2 \;=\; (X^{2^m} - ib)(X^{2^m} + ib).$$

We will show that ib and $-ib$ are not squares in $K(i)$. To do this, suppose $ib = (c + di)^2 = c^2 - d^2 + 2cdi$ for some $c, d \in K$. Then $c^2 = d^2$ and $b = 2cd$. Then $-a = b^2 = 4c^4$, contrary to the hypothesis $a \notin -4K^2$. The proof for $-ib$ is similar. Related arguments also show that ib and $-ib$ are not in$-4K(i)^4$.

The assumed result for m (over the field $K(i)$) implies that both $X^{2^m} + ib$ and $X^{2^m} - ib$ are irreducible over $K(i)$. Then any nontrivial factorization of $X^{2^{m+1}} - a$ over K must match those two irreducible factors when lifted to $K(i)$. But then $X^{2^m} \pm ib$ would be in $K[X]$ contrary to the fact that $i \notin K$. This proves that $X^{2^{m+1}} - a$ is irreducible over K. □

With those prime-power results, we can now tackle the general case of $X^n - a$.

Theorem. (Capelli 1898) Let K be a field, $n \in \mathbb{Z}^+$, and $a \in K$. For each prime p dividing n, suppose $a \notin K^p$. If $4|n$, suppose $a \notin -4K^4$. Then $X^n - a$ is irreducible in $K[X]$.

Proof. Working inductively (or by WOP) we assume that this result holds true for exponents less than n (over any field). Suppose $a \in K$ satisfies the hypotheses, and let β be a root of $X^n - a$. To show that $X^n - a$ is irreducible we will prove: $[K(\beta):K] = n$.

When n is a prime power we are done by the preceding work. So suppose n is not a prime power and express $n = p^r m$ where p is an odd prime, $r \geq 1$, $m > 1$, and $p \nmid m$. Let $\alpha = \beta^{p^r}$, so that is a root of $X^m - a$. The inductive hypothesis implies that $X^m - a$ is irreducible over K, and therefore $[K(\alpha):K] = m$. The result will follow if we prove $[K(\beta):K(\alpha)] = p^r$. Since β is a root of $X^{p^r} - \alpha$, it suffices to show that this polynomial is irreducible in $K(\alpha)[X]$. By the prime power Proposition above, this is equivalent to proving $\alpha \notin K(\alpha)^p$.

Let $N = N_K^{K(\alpha)}$ be the norm and check that $N(\alpha) = (-1)^m a$. For contradiction, suppose $\alpha = \delta^p$ for some $\delta \in K(\alpha)$. Then $N(\alpha) = N(\delta)^p$ in K, and we find that

$$a = (-1)^m N(\alpha) = (-1)^m N(\delta)^p = \big((-1)^m N(\delta)\big)^p.$$

Then $a \in K^p$, contrary to hypothesis. □

This Theorem established a result on algebraic closures first proved by Artin and Schreier:

Theorem. $[K^{\text{alg}}:K] < \infty$ if and only if K algebraically closed or real closed.

This result has been poked in various ways since the 1920s. We can consider other types of "closures." For a prime p define the *p-closure* $K^{(p)}$ of K as a minimal algebraic extension L/K

such that $L = L^p$. Or define a (possibly) different p-closure $K^{[p]}$, requiring that L has no extensions of degree p. We can define a *pythagorean closure* K^{pyth} to be the smallest algebraic extension L/K where $L^2 + L^2 = L^2$. For each type of closure we ask the analogous question:

- For what class of fields is the closure of K a finite extension of K?

12 Proof of the Nullstellensatz

If there are ways I can make this more confusing, please let me know.

Today we talk about some more algebraic geometry. For a set B of polynomials in $\mathbb{C}[X] = \mathbb{C}[x_1, \ldots, x_n]$, we defined $\mathcal{Z}(J)$ as the as thee set of common zeros of all elements of J:

$$\mathcal{Z}(B) = \{c \in \mathbb{C}^n : f(c) = 0 \text{ for every } f \in B\}.$$

If J is the ideal generated by set B, check that $\mathcal{Z}(B) = \mathcal{Z}(J)$. Then we will restrict attention to ideals of polynomials, rather than just sets of them. These sets can also be considered as the zero-set of a finite list of polynomials, because of Hilbert's basic result:

Theorem. (Hilbert Basis Theorem) Every ideal of $\mathbb{C}[X]$ is finitely generated.

This famous result that was important in mathematical history. It showed that the algorithmic techniques of classical 19^{th} century invariant theory could be replaced by abstract existence proofs. Emmy Noether generalized Hilbert's ideas to show that if R is a ring in which every ideal is finitely generated, then $R[x]$ also has that property. We will not discuss this theory of "noetherian rings."

Ring Theory Review.
All rings considered today are commutative (and associative), with 1. An *ideal* of ring R is a subset J such that $J + J \subseteq J$ and $RJ \subseteq J$. A key point: If $\varphi : R \to S$ is a ring homomorphism, then the kernel $J = \ker(\varphi)$ is an ideal of R. Conversely, every ideal is such a kernel. In fact, if J is an ideal define the factor ring R/J as "R mod J." The kernel of the natural map $\pi : R \to R/J$ is exactly J.

For subset $S \subseteq R$, the ideal (S) generated by S is the intersection of all ideals containing S. Check that a typical element of (S) is $r_1 s_1 + \cdots + r_m s_m$ for some $r_j \in R$ and $s_j \in S$. If $S = \{a_1, \ldots, a_k\}$ is a finite subset, then that ideal $(a_1, \ldots, a_n)$ equals $a_1 R + \cdots + a_n R$. A "principal ideal" is and ideal with one generator, so that $(a) = aR$ is principal. Note that the trivial ideals $(0) = \{0\}$ and $(1) = R$ are principal.

Combining two ideals.
Suppose I and J are ideals in R.
• Their *intersection* $I \cap J$ is also an ideal.
• Their *sum* $I + J$ is the ideal generated by $I \cup J$. A typical element of $I + J$ is just $i + j$ for $i \in I$ and $j \in J$.
• Their *product* IJ is the ideal generated by the set of products. A typical element of IJ is a finite sum $\sum i_k j_k$ where each $i_k \in I$ and $j_k \in J$.

Those ideals satisfy a chain of inclusions:

$$(0) \subseteq IJ \subseteq I \cap J \subseteq I, J \subseteq I + J \subseteq R.$$

Exercise. The ring $\mathbb{Z}$ is a *principal ideal domain*, meaning that it is an integral domain and every ideal is principal. [Proof: the $ax + by$ result on Ross sets.] Suppose I, J are nonzero ideals.

Then $I = (m) = m\mathbb{Z}$ and $J = (n) = n\mathbb{Z}$ for some $m, n \in \mathbb{Z}^+$. Show:

$IJ = (mn)$,

$I \cap J = (\ell)$ where $\ell = \text{lcm}(m, n)$, and

$I + J = (d)$, where $d = \gcd(m, n)$.

$J \subseteq I$, that is, $(n) \subseteq (m)$, if and only if $m \mid n$.

Exercise. If I, J are ideals of ring R and $I + J = (1)$, show that $IJ = I \cap J$.
[Analogue in $\mathbb{Z}$: If $\gcd(m, n) = 1$ then $\text{lcm}(m, n) = mn$.]

Definition. The *radical* $\sqrt{J}$ of an ideal J of R is:

$$\sqrt{J} = \{a \in R \mid a^n \in J \text{ for some } n \in \mathbb{Z}^+\}$$

The *nilradical* of R is $\text{nil}(R) = \sqrt{(0)}$, the set of all nilpotent elements of R.

Exercise. (1) Clearly $J \subseteq \sqrt{J}$. Why is $\sqrt{J}$ an ideal? [Idea: a, b are nilpotent $\Rightarrow$ $a + b$ is nilpotent.]
(2) If $m \in \mathbb{Z}^+$, determine $\sqrt{(m)}$ as an ideal in $\mathbb{Z}$. For which m does $(m) = \sqrt{(m)}$?
(3) An ideal I is a *radical ideal* if $I = \sqrt{I}$. For any J explain why $\sqrt{J}$ is a radical ideal.
(4) PODASIP: $\sqrt{I + J} = \sqrt{I} + \sqrt{J}$ and $\sqrt{IJ} = \sqrt{I \cap J} = \sqrt{I} \cap \sqrt{J}$.
(5) If $J \subseteq P$ for a prime ideal P, show that $\sqrt{J} \subseteq P$.
Somewhat trickier: $\sqrt{J} = \bigcap_{P \supseteq J} P$, the intersection of all prime ideals containing J.

Algebraic sets.
A subset A of $\mathbb{C}^n$ is an *algebraic set* if $A = \mathcal{Z}(J)$ for some ideal J in $C[X]$. For instance, when $n = 1$ a proper subset $A \subset \mathbb{C}$ is algebraic exactly if it is finite.

If subset $S \subseteq \mathbb{C}^n$ is an algebraic set, can we determine the corresponding ideal J? Is there a unique ideal J with $S = \mathcal{Z}(J)$?

As we mentioned in an earlier lecture, the natural candidate for the ideal for S is:

$$\mathcal{I}(S) \;=\; \{f \in \mathbb{C}[X] \mid f \text{ vanishes on } S\}.$$

Exercise. (1) If S is an algebraic set then $S = \mathcal{Z}\mathcal{I}(S)$.
(2) If ideal J is of the form $\mathcal{I}(S)$ for some set $S \subseteq \mathbb{C}^n$, then $J = \mathcal{I}\mathcal{Z}(J)$.

It's not immediately clear which ideals are of the form $\mathcal{I}(S)$. For instance, the ideal $J = (x^2, y^3)$ in $\mathbb{C}[x, y]$ is not of that form. It has zero set $\mathcal{Z}(J) = \{(0, 0)\}$, a single point in $\mathbb{C}^2$. Its corresponding ideal is $M = \mathcal{I}\mathcal{Z}(J) = (x, y)$ does not equal J.

It's not hard to see that every ideal of the form $\mathcal{I}(S)$ is a radical ideal. In fact, for any ideal J, check that: $\sqrt{J} \subseteq \mathcal{I}\mathcal{Z}(J)$. Hilbert was able to prove the converse: Every radical ideal is of the form $\mathcal{I}(S)$.

Hilbert's Nullstellensatz. If J is an ideal of $\mathbb{C}[X]$, then $\mathcal{I}\mathcal{Z}(J) = \sqrt{J}$.

This sets up a wonderful bijection:

$$\{ \text{ algebraic sets in } \mathbb{C}^n \} \longleftrightarrow \{ \text{ radical ideals in } \mathbb{C}[X] \}.$$

One hint of the power of this Theorem is to notice that the points of $\mathbb{C}^n$ are in bijective correspondence with the maximal ideals of $\mathbb{C}[X]$. This is an early indication that Geometry (e.g. points in space) can be replaced by Algebra (e.g. maximal ideals in a ring)! During the Twentieth Century mathematicians rewrote the foundations of algebraic geometry in more and more general ways.

Hilbert's result holds true for any algebraically closed field K in place of $\mathbb{C}$. Here is a weaker version of the Theorem, mentioned in an earlier lecture. Note that J is a proper ideal if $J \neq (0)$ and $J \neq (1)$.

Theorem (Weak Nullstellensatz) Suppose K is algebraically closed field. If J is a proper ideal in $K[X]$, then $\mathcal{Z}(J) \neq \varnothing$.

Here is an equivalent statement: If $f_1, \ldots, f_k \in K[X]$ have no common zero in K^n, then there exist polynomials Q_i such that $1 = f_1Q_1 + \cdots + f_kQ_k$.

Rabinowitsch (1929) discovered a trick to prove that this weak version implies the full theorem.

Proof of Nullstellensatz from its weak version. Suppose J is an ideal in $K[X]$ and $f \in \mathcal{IZ}(J)$, so that f vanishes on $\mathcal{Z}(J)$. Choose generators $g_1, \ldots, g_k$ for J (using the Basis Theorem). Then f vanishes on $\mathcal{Z}(g_1, \ldots, g_k)$.

To Prove: $f^m \in (g_1, \ldots, g_k)$ for some $m > 0$.

Let z be a new variable and consider $K[X, z]$. The ideal $J' = (g_1, \ldots, g_k, 1 - zf)$ has $\mathcal{Z}(J') = \varnothing$. For, if $(c, r) \in K^n \times K$ is a common zero then $c \in \mathcal{Z}(g_1, \ldots, g_k)$ so that $f(c) = 0$, yielding a contradiction. The Weak Nullstellensatz implies that there are polynomials $Q_j \in K[X, z]$ such that

$$1 = g_1Q_1 + \cdots + g_kQ_k + (1 - zf)Q_{k+1}.$$

Now substitute $z = 1/f$ to get an identity in $K[X][1/f]$:

$$1 = g_1(X)Q_1(X, 1/f) + \cdots + g_k(X)Q_k(X, 1/f) + 0.$$

Multiplying by a high enough power of f to clear denominators, we find that

$$f^m = g_1A_1 + \cdots g_kA_k,$$

for some $m > 0$ and some $A_k \in K[X]$. □

We still have not discussed any proof of the Weak Nullstellensatz ! There are several approaches. The standard method in modern times involves the study of integral extensions of rings. Today, let's try to describe the Nineteenth Century proof that uses Kronecker's method of elimination.

First we recall the **resultant**. Let $f = a_nx^n + \cdots + a_0$ and $g = b_mx^m + \cdots + b_0$ be two polynomials (in one variable x). How do we test whether f and g have a common factor? Of course one

way is to use Euclid's Algorithm to compute the GCD. But we use a different method, based in matrix theory.

Show (as an exercise) that f and g share a common factor if and only if $hf = kg$ for some polynomials h and k such that $\deg h < \deg g = m$ and $\deg k < \deg f = n$. We try to compute h and k by viewing their coefficients as unknowns and writing out the resulting system of equations. This turns out to be a linear system in $n + m$ unknowns, with the following matrix:

$$\mathrm{Syl}(f,g) = \begin{bmatrix} a_n & a_{n-1} & \cdots & a_1 & a_0 & 0 & \cdots & 0 \\ 0 & a_n & \cdots & a_2 & a_1 & a_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & a_n & \cdots & \cdots & \cdots & a_0 \\ b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & \cdots & 0 \\ 0 & b_m & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & b_m & \cdots & \cdots & b_0 \end{bmatrix}$$

This "Sylvester matrix" has m rows involving the a_i and n rows involving the b_j.

Definition. The *resultant* of f and g is $R(f,g) = \det Syl(f,g)$.

By construction, f and g have a common factor if and only if $R(f,g) = 0$.

If $f, g \in K[x]$, then $R(f,g) \in K$. In fact, the resultant is an integer polynomial expression involving the a_i and b_j. Writing $\vec{a} = (a_n, \ldots, a_0)$ and $\vec{b} = (b_n, \ldots, b_0)$, we can write this as: $R(f,g) \in \mathbb{Z}[\vec{a}, \vec{b}]$.

Lemma. $R(f,g)$ is in the ideal generated by f and g. That is,

$$R(f,g) = A(x)f(x) + B(x)g(x)$$

for some polynomials $A(x), B(x)$ with coefficients in $\mathbb{Z}[\vec{a}, \vec{b}]$.

Proof. Observe that $\mathrm{Syl}(f,g) \begin{bmatrix} x^{m+n-1} \\ x^{m+n-2} \\ \vdots \\ x^2 \\ x \\ 1 \end{bmatrix} = \begin{bmatrix} x^{m-1}f(x) \\ x^{m-2}f(x) \\ \vdots \\ f(x) \\ x^{n-1}g(x) \\ x^{n-2}g(x) \\ \vdots \\ g(x) \end{bmatrix}$. Multiply on the left by the classical adjoint (or adjugate) of $Syl(f,g)$ to show that $R(f,g)$ times the left-hand column vector equals that adjoint matrix times the right-hand column. The bottom entry yields an expression of $R(f,g)$ as a combination of the terms $x^i f(x), x^j g(x)$ for various i, j with coefficients in $\mathbb{Z}[\vec{a}, \vec{b}]$. (This method is essentially Cramer's Rule.) □

With a bit more work we can see that this $R(f,g)$ equals the resultant studied on the Ross sets. If $f(x) = a(x-\alpha_1)\cdots(x-\alpha_n)$ and $g(x) = b(x-\beta_1)\cdots(x-\beta_m)$, then

$$R(f,g) = a^m b^n \prod_{i,j}(\alpha_i - \beta_j).$$

Resultants can be used in studying polynomials in two variables. If $f, g \in \mathbb{C}[x,y]$ are nonzero then the zero sets $\mathcal{Z}(f)$, $\mathcal{Z}(g)$ are curves in the plane $\mathbb{C}^2$. Given those polynomials, along with pens and paper, how can we compute where they intersect? One method is to compute $R(x) = R_y(f,g)$, the resultant of f, g as polynomials in y (viewing x as a constant for now). That quantity is a polynomial in x alone, and we can determine its roots (perhaps by Sturm's Theorem and Newton's Method). Once we have $a \in \mathbb{C}$ with $R(a) = 0$ then we know $f(a,y)$ and $g(a,y)$ have a common zero, and a GCD calculation will yield all $b \in \mathbb{C}$ with $f(a,b) = g(a,b) = 0$.

How many points (a,b) are there in $\mathcal{Z}(f,g)$, the intersection of those two curves? If we know the degree of $R(x)$, then we know how many zeros a there are, and can count how many (a,b) arise. The exact number is found by analyzing those degrees more carefully. That root counting must be done "with multiplicity," meaning that double roots are counted twice, etc. For instance, $f(x) = x^2(x-4)^3$ has 5 roots. The roots are 0 and 4, but root 0 has multiplicity 2 and root 4 has multiplicity 3.

Bezout's Theorem. If $f, g \in \mathbb{C}[x,y]$ are homogeneous of degrees m, n respectively, then $\mathcal{Z}(f,g)$ contains exactly mn points, when counted with multiplicity.

Nullstellensatz proof. To study common zeros of polynomials of several variables, resultants will reduce the number of variables involved (even though the resulting polynomials might have higher degrees and larger coefficients). This sort of "elimination theory" provided an early proof of Hilbert's Nullstellensatz. We outline the basic idea, skipping all details. We will work over $\mathbb{C}$ (rather than a general algebraically closed field K) just to match the old flavor.

Suppose $f_1, \ldots, f_r \in \mathbb{C}[X]$ where $X = (x_1, \ldots, x_n)$ is a system of n variables. The Weak Nullstellensatz says: If those f_j have no common zero in $\mathbb{C}^n$, then the ideal $J = (f_1, \ldots f_r)$ contains 1.

The key step of the algorithm is to produce a "resultant system" $D_1, \ldots, D_s \in \mathbb{C}[x_2, \ldots, x_n]$ with the property that every common zero in $\mathbb{C}^{n-1}$ for $D_1, \ldots, D_s$ yields at least one common zero in in $\mathbb{C}^n$ for $f_1, \ldots, f_r$. Moreover, each D_i lies in the ideal J.
Repeat that step n times to form some complicated resultant system $g_1, \ldots, g_N$ consisting of **constants** in $\mathbb{C}$, and each of the g_j lies in the ideal J. Then a common zero for the ideal J exists in $\mathbb{C}^n$ if and only if those constants are all zero. Therefore, if J has no common zero then that final system must contain a nonzero constant. But that constant is a unit and lies in the ideal J. That proves the Weak Nullstellensatz.

Details of that key step are fairly tricky. Set $x = x_1$ and $K = \mathbb{C}(x_2, \ldots, x_n)$. Introduce new systems of indeterminates $u = (u_1, \ldots, u_s)$ and $v = (v_1, \ldots, v_s)$, and form polynomials $f_u =$

$\sum u_j f_j$ and $f_v = \sum v_j f_j$ in $K(u,v)[x]$. Eliminate x by taking the resultant $R(u,v) = R(f_u, f_v)$. Let $D_1, \ldots, D_s \in K$ be the list of coefficients of $R(u,v)$. Many further details need to be worked out to make this precise. All of these ideas about resultants, elimination theory, and the Nullstellensatz, are discussed in early editions of van der Waerden's 1931 book *Moderne Algebra*. That classic text was based on lectures by Emil Artin and Emmy Noether.

13 Quadratic Forms: Polynomials and Inner Products

Education is an admirable thing, but it is well to remember from time to time that nothing worth knowing can be taught. - Oscar Wilde

To continue investigating sums of squares we will talk about the basic ideas of quadratic and bilinear forms. One motivating question is: Can $x^2 + y^2 + z^2$ be expressed as a sum of two squares in $\mathbb{R}(x, y, z)$?

Polynomial Version.
Recall that a *quadratic form* Q is a homogeneous polynomial of degree 2. If the variables are $X = (x_1, \ldots, x_n)$ then $Q(X) = \sum_{1 \le i \le j \le n} c_{ij} x_i x_j$. The standard convention is to allow all pairs (i, j) (not restricting to $i \le j$) but to require the coefficients to be symmetric.

Exercise. If $Q(X)$ is a quadratic form as above, then $Q(X) = \sum_{i,j} a_{ij} x_i x_j$ where $M = (a_{ij})$ is a symmetric $n \times n$ matrix over K. For example, $Q = x^2 + xy + 3yz + 5y^2 - 8z^2$ has corresponding symmetric matrix $M = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 5 & \frac{3}{2} \\ 0 & \frac{3}{2} & -8 \end{bmatrix}$.

Note that $Q(x, y, z) = x^2 + y^2 + z^2$ can be viewed as the squared-length of the vector $v = (x, y, z)$. Using the standard dot product, we can express this as $Q(v) = |v|^2 = v \cdot v$.

We will generalize that connection between quadratic forms and the dot product. For indeterminates $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$, define a symmetric bilinear form $B(X, Y)$ to be a polynomial over K that is linear (homogeneous degree 1) in X, and separately linear in Y. Then we must have $B(X, Y) = \sum_{i,j} a_{ij} x_i y_j$ for some coefficients $a_{ij} \in K$. Moreover, as in the exercise, we may assume that those coefficients are symmetric. Check that $Q(X) = B(X, X)$ is a quadratic form. Moreover, every quadratic form Q arises in this way from an associated bilinear form B.

In matrix terms, let's express vectors as columns. If X is the column vector with entries x_j, its transpose X^T is the corresponding row vector. With notations above, let $M = (a_{ij})$ be the symmetric $n \times n$ matrix associated to Q or B. Then:

$$B(X, Y) = X^\mathsf{T} M Y \quad \text{and} \quad Q(X) = X^\mathsf{T} M X.$$

The theory of vector spaces and linear maps is an abstraction of the more concrete theory of column vectors and matrices. Suppose V and W are K-vector spaces of dimensions m and n, and $f : V \to W$ is a linear map. If bases are chosen for V and for W, a vector in V is represented by a column in K^m, a vector in W becomes a column in K^n, and f is expressed as an $n \times m$ matrix.

Abstract version.
In the abstract theory, if V is a K vector space of dimension n, there there exists a basis $\mathscr{B} = \{e_1, \ldots, e_n\}$ for V. That is, every $v \in V$ is represented uniquely as a linear combination $v = c_1e_1 + \cdots + c_ne_n$ for some coefficients $c_j \in K$. Then the coordinate vector for v (relative to that basis) is $C_v = (c_1, \ldots, c_n)^\mathsf{T} \in K^n$. If $\mathscr{B}' = \{e'_1, \ldots, e'_n\}$ is another basis for V, the vector v is represented by a different coordinate vector $C'_v = (c'_1, \ldots, c'_n)^\mathsf{T}$. There is an invertible $n \times n$ matrix P that expresses the new basis in terms of the old. Then $C'_v = PC_v$ for every V.

Definition. A *bilinear form* is a map $b : V \times V \to K$ that is K-linear in each slot. That form is *symmetric* if $b(v, w) = b(w, v)$ for every $v, w \in V$.
The form b is *nonsingular* if for every nonzero $v \in V$, there exists $w \in V$ such that $b(v, w) \neq 0$. Equivalently, if $b(v, w) = 0$ for every w, then $v = 0$.
The *quadratic form* associated to b is $q : V \to K$ defined by: $q(v) = b(v, v)$.
For a basis $\mathscr{B} = \{e_1, \ldots, e_n\}$ for V, let $a_{ij} = b(e_i, e_j)$, and define $M = (a_{ij})$ to be the *Gram matrix* of b, relative to $\mathscr{B}$. Then M is a symmetric $n \times n$ matrix over K.

Exercise. Continue the notations above.
(1) If $v, w \in V$ have coordinate (column) vectors C_v, C_w for basis $\mathscr{B}$, then:

$$b(v, w) = C_v^\mathsf{T} M C_w, \text{ and } q(v) = C_v^\mathsf{T} M C_v.$$

Then if v is represented as the column vector $(x_1, \ldots, x_n)^\mathsf{T}$, then $q(v) = \sum_{ij} a_{ij}x_ix_j$.
(2) The form b is nonsingular $\iff$ M is nonsingular (i.e invertible) as a matrix.
(3) If $\mathscr{B}'$ is another basis, the new Gram matrix M' is related to the previous M by:

$$M' = P^\mathsf{T} M P \quad \text{for an invertible } n \times n \text{ matrix } P.$$

(4) A map $q : V \to K$ is a quadratic form if and only if
the map $b(v, w) = \frac{1}{2}\big(q(v + w) - q(v) - q(w)\big)$ is bilinear.

Exercise. (1) Quadratic form $Q_2 = 2xy$ has Gram matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. The form $Q'(x, y) = x^2 - y^2$ has Gram matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Explain why those polynomials represent the same abstract quadratic form, using two choices of basis for the underlying 2-dimensional space.
(That is, Q_2 and Q' are isometric.)
(2) Show that $Q_a(x, y) = axy$ is also isometric to Q', for every $a \in K^*$. That is: $aQ' \cong Q'$.

Quadratic forms are types of polynomials in n variables, and were initially studied purely in that aspect. The standard dot product on $\mathbb{R}^n$ motivates a link between those quadratic forms as polynomials and generalized dot products on vector spaces. For column vectors u, v in $\mathbb{R}^n$, their dot product is $u \cdot v = u^\mathsf{T} v = \sum_i u_i v_i$. As mentioned in freshman calculus (or physics) classes, dot products are related to angles. In fact,

$$u \cdot v = |u||v| \cos(\theta)$$

where θ is the angle between u and v.

Assumptions. Today assume that $\text{char}(K) \neq 2$ and every vector space is finite dimensional.

Suppose V is a K-vector space and $b : V \times V \to K$ is a symmetric bilinear form. We imitate the geometry of dot products in this more general context.

Definition. Define vectors $v, w \in V$ to be *orthogonal* or perpendicular if $b(v, w) = 0$.
For a subset S of V, define its *orthogonal complement*:

$$S^\perp = \big\{ v \in V : b(v, x) = 0 \text{ for every } x \in S \big\}.$$

As shorthand, "S is $\perp$ to T" means $T \subseteq S^\perp$, or equivalently, $S \subseteq T^\perp$.

Exercise. Continue the notations b, V, K as above.
(1) $S^\perp$ is a vector subspace of V, for any $S \subseteq V$. If $U = \text{span}(S)$, then $S^\perp = U^\perp$.
(2) If $U \subseteq W$, then $W^\perp \subseteq U^\perp$. Moreover $S \subseteq S^{\perp\perp}$ for any S.
If $U = S^\perp$ for some S, then $U^{\perp\perp} = U$.
(3) b is nonsingular $\iff$ $V^\perp = (0)$.
(4) If U, W are subspaces, does $(U + W)^\perp = U^\perp \cap W^\perp$? Does $(U \cap W)^\perp = U^\perp + W^\perp$?

Exercise. If (V, b) and (V', b') are quadratic spaces, and $f : V \to V'$, define what it means for f to be an isometry. If M and M' are the Gram matrices, show that $(V, b) \cong (V', b')$ if and only if $M' = P^{\mathsf{T}} M P$ for some invertible P.

Definition. A *symmetric bilinear space* or a *quadratic space* over K is a pair (V, b) where b is a nonsingular, symmetric, bilinear form on the K-vector space V.

Exercise. If (V, b) and (V', b') are quadratic spaces, and $f : V \to V'$, define what it means for f to be an isometry. If M and M' are the Gram matrices, show that $(V, b) \cong (V', b')$ if and only if $M' = P^{\mathsf{T}} M P$ for some invertible P.

This geometric viewpoint should not be taken too literally. There might be nonzero vectors whose "length" is zero! For instance, think of the standard 2-dimensional form $Q(x, y) = x^2 + y^2$. Over $\mathbb{R}$ is the only vector of length 0 is the 0-vector. But over $\mathbb{C}$ the vector $(1, i)$ has length $Q(1, i) = 0$.

Notation. A vector v in (V, b) is *isotropic* if $v \neq 0$ but $q(v) = 0$. The form q, or the space (V, b), is called isotropic if it contains an isotropic vector. Spaces that are not isotropic are usually called *anisotropic*.

Proposition. Suppose b is a nonsingular symmetric bilinear form on V, where $\dim(V) = n$. Then for any subspace U:
(1) $\dim(U^\perp) = n - \dim(U)$.
(2) $U^{\perp\perp} = U$.

The proof is best understood by using properties of the dual space $V^* = \text{Hom}_K(V, K)$. General theory (using "dual bases") shows that V and V^* have the same dimension. A bilinear form b induces a linear map $\theta_b : V \to V^*$, given by $v \mapsto b(v, \cdot)$. Check that b is nonsingular $\iff$ the map θ_b is bijective. We omit further details.

Exercise. A subspace U of (V, b) might be a subspace of its own complement! In this case, the space $U^\perp$ is not much of a "complement" of U.
(1) $U \subseteq U^\perp \iff b(U, U) = 0$. That is: Every vector in U is isotropic.

Such a subspace is said to be *totally isotropic*.
(2) If (V, b) is nonsingular and $U \subseteq V$ is totally isotropic, prove: $\dim(U) \leq \frac{1}{2}\dim(V)$.
[If $\dim(U) = r$ there is a basis of V whose Gram matrix M has an $r \times r$ block of zeros is the upper left. if $r > n - r$ show that $\det(M) = 0$.]

Exercise. Suppose (V, b) is a quadratic space and $U \subseteq V$ is a subspace. The restriction of b to $U \times U$ is a bilinear form we can call $b|_U$.
(1) Show: $(U, b|_U)$ is nonsingular $\iff$ $U \cap U^\perp = (0)$. In this case, $U^\perp$ is also a nonsingular subspace. Moreover, $V = U \oplus U^\perp$, a direct sum of subspaces.
(2) Since b is nonsingular, there exists $v_1 \in V$ with $q(v_1) = b(v_1, v_1) \neq 0$. (Why?)
Show that subspaces $U_1 = Kv_1$ and $U_1^\perp$ are nonsingular.
Repeat: Choose $v_2 \in U_1^\perp$ with $q(v_2) \neq 0$, define $U_2 = \text{span}(v_1, v_2)$, etc. Explain why this process yields a basis $\{v_1, \ldots, v_n\}$ such that $b(v_i, v_j) = 0$ whenever $i \neq j$.

Such a basis is called an *orthogonal basis*. Its Gram matrix is diagonal.
(3) If M is a symmetric $n \times n$ matrix over K, prove that there exists an invertible matrix P such that $P^\mathsf{T} M P$ is diagonal.

Such a basis is called an *orthogonal basis*. Its Gram matrix is diagonal.
(4) A nonsingular (V, b) over K might fail to have an orthogonal basis if $\text{char}(K) = 2$. Where does the proof above fail in that case?

Here's the polynomial version: A quadratic form (polynomial) is "diagonal" if is $a_1x_1^2 + \cdots + a_nx_n^2$ for some scalars a_j (with no "cross terms"). The Exercise shows that for any quadratic form $Q(X)$, there is a linear change of variables making it into a diagonal form. The geometric viewpoint symmetric bilinear forms is the most natural way to understand this result.

Notation. If $a_j \in K^*$, we write $\langle a_1, \ldots, a_n \rangle$ for the n-dimensional quadratic space (V, b) with an orthogonal basis $\{e_1, \ldots, e_n\}$ such that $q(e_j) = a_j$.

In this notation, the usual dot product on $\mathbb{R}^n$ is $n\langle 1 \rangle = \langle 1, 1, \ldots, 1 \rangle$.

Exercise. (α) Find an invertible 2×2 matrix P such that $P^\mathsf{T} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} P = \begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix}$.
In other terminology: $\langle 1, 2 \rangle \cong \langle 3, 6 \rangle$.
[Idea: In the space $\langle 1, 2 \rangle$ find vector v_1 with $q(v_1) = 3$, then search for $v_2 \in (v_1)^\perp$ with $q(v_2) = 6$.]

(β) If $c = a^2 + 2b^2$ in K, show: $\langle 1, 2 \rangle \cong \langle c, 2c \rangle$. Does your proof yield an invertible P_c such that:

$$P_c^\mathsf{T} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} P_c = \begin{bmatrix} c & 0 \\ 0 & 2c \end{bmatrix} ?$$

Definition. A quadratic form (V, q) *represents* a value $c \in K$ if there exists a nonzero vector $v \in V$ with $q(v) = c$. Note: q represents $0 \iff q$ is isotropic.

Exercise. (1) Which $c \in \mathbb{Q}$ are represented by the form $\langle 1, 1\rangle$? (The "2-square theorem.")
(2) If p is an odd prime, and $a, b, d \in \mathbb{F}_p^*$, show that $\langle a, b\rangle$ represents d.
Deduce that every 3-dimensional quadratic form over $\mathbb{F}_p$ is isotropic.
(3) Suppose quadratic space (V, q) represents $c \in K^*$. Then there is an orthogonal basis $\{v_1, \ldots, v_n\}$ with $q(v_1) = c$. That is, $q \cong \langle c, a_2, \ldots, a_n\rangle$, for some $a_j \in K^*$.

The quadratic form $\langle 1, -1\rangle$ is certainly isotropic. As a polynomial, this form is $x^2 - y^2$ and we know that every $c \in K$ can be represented this way. (Why does $c = x^2 - y^2$ always have a solution?) This fact is generalized as follows.

Lemma. If (V, q) is an isotropic quadratic space over K, then q represents every $c \in K$.
That is, for any c there exists $u \in V$ with $q(u) = c$.

Proof. There exists $v \neq 0$ in V with $q(v) = 0$. Nonsingularity yields some $w \in V$ with $b(v, w) \neq 0$, and we may scale w to assume $b(v, w) = 1$. Let $s = q(w)$. For $\alpha \in K$, then $q(\alpha v + w) = \alpha^2 q(v) + 2\alpha b(v, w) + q(w) = 2\alpha + s$. As α varies, that quantity ranges over all values in K. □

14 Cassels' Trick to Eliminate Denominators

If more people would try and reach up instead of going to the lowest common denominator, I think we'd be a lot better off. - Mary Badham

Around 1962, Cassels suspected that $x_1^2 + \cdots + x_n^2$ cannot be expressed as a sum of fewer than n squares in the field $\mathbb{R}(x_1, \ldots, x_n) = \mathbb{R}(X)$. He succeeded in proving this using a wonderful trick that helps eliminate denominators in certain expressions.

As usual, we assume the field K has characteristic $\neq 2$.

Cassels' Theorem. Let $\gamma = \langle a_1, \ldots, a_n \rangle$ be a nonsingular quadratic form over K, and let $p(x) \in K[x]$ be a polynomial of one variable.

If γ represents $p(x)$ over $K(x)$, then γ represents $p(x)$ over $K[x]$.

Proof. We start by eliminating a trivial case. Suppose γ is isotropic over K (i.e. there exists $v \neq 0$ with $\gamma(v) = 0$). Then γ represents every value $c \in K$. That proof (given at the end of Lecture 13) quickly shows that γ represents $p(x)$ in $K[x]$. Therefore, for this proof we may assume that γ is anisotropic.

Since $\gamma = \langle a_1, \ldots, a_n \rangle$ represents $p(x)$ over $K(x)$, there exist rational functions $r_j(x) \in K(x)$ such that $p(x) = a_1 r_1(x)^2 + \cdots + a_n r_n(x)^2$. Clearing denominators, this becomes

$$p(x) f_0(x)^2 \;=\; a_1 f_1(x)^2 + \cdots + a_n f_n(x)^2, \qquad (*)$$

for polynomials $f_j(x) \in K[x]$, with $f_0(x) \neq 0$. From all expressions of this form, choose one for which the degree of $f_0(x)$ is minimal. It follows that $\gcd(f_0, f_1, \ldots, f_n) = 1$. If $\deg f_0(x) = 0$, then $f_0(x)$ is a nonzero constant, and we are done! For the sake of contradiction, suppose $\deg f_0(x) > 0$.

Define the quadratic form φ of dimension $n+1$ over the field $K(x)$, by:

$$\varphi = \langle -p(x), a_1, \ldots, a_n \rangle.$$

Then $\varphi(\vec{f}) = 0$, for the vector $\vec{f} = (f_0, f_1, \ldots, f_n)$.

Here's the geometric motivation for the proof below: The zero set $\mathcal{Z}(\varphi)$ is a nonzero hypersurface in $(K(x))^{n+1}$. Since $\mathcal{Z}(\varphi)$ is quadratic, we expect that a line in $(K(x))^{n+1}$ that meets that surface will intersect $\mathcal{Z}(\varphi)$ in one other point. (Think of a line in 3-space that meets the unit sphere.) We will choose a particular line through $\vec{f}$, and find the other point where that line intersects $\mathcal{Z}(\varphi)$. We hope to choose things so that this new point yields an expression for $p(x)$ of smaller degree.

Here are the algebraic details: For each j, divide f_j by f_0, writing $f_j = f_0 Q_j + r_j$ where $\deg r_j < \deg f_0$. When $j = 0$, we get $Q_0 = 1$ and $r_0 = 0$. Put the quotients and remainders into tuples, writing $\vec{Q} = (Q_0, \ldots, Q_n)$ and $\vec{r} = (r_0, \ldots, r_n)$. In this notation, we have $\vec{f} = f_0 \vec{Q} + \vec{r}$.

Let $B(u, v)$ be the symmetric bilinear form associated with φ, so that $\varphi(u) = B(u, u)$, and

$2B(u,v) = \varphi(u+v) - \varphi(u) - \varphi(v)$. The line through $\vec{f}$ and $\vec{Q}$ will meet the surface $\mathcal{Z}(\varphi)$ is some point $\vec{h}$. Some calculations show that: $\quad \vec{h} = \varphi(\vec{Q})\vec{f} - 2B(\vec{f},\vec{Q})\vec{Q}$.

Write $\vec{h} = (h_0, h_1, \ldots, h_n)$. As an exercise, verify that $\vec{h} \in \mathcal{Z}(\varphi)$.

Claim. $\deg h_0 < \deg f_0$.

This inequality will complete the proof, since $(h_0, \ldots, h_n)$ satisfies the equation (*) and has smaller degree, contrary to the minimal choice made at the start of the proof.

To prove the Claim, we derive an explicit formula for h_0. By definitions, we know:

$$\varphi(\vec{Q}) = -pQ_0^2 + \sum_{j=1}^{n} a_j Q_j^2,$$

$$B(\vec{f},\vec{Q}) = -pf_0Q_0 + \sum_{j=1}^{n} a_j f_j Q_j.$$

Consider the first (zeroth?) components here, noting that $Q_0 = 1$, to get:

$$\begin{aligned} h_0 &= \varphi(\vec{Q})f_0 - 2B(\vec{f},\vec{Q})Q_0 \\ &= \big(-p + \sum a_j Q_j^2\big) f_0 - 2\big(-pf_0 + \sum_{1}^{n} a_j f_j Q_j\big) \\ &= pf_0 + \sum_{1}^{n} a_j (f_0 Q_j^2 - 2 f_j Q_j). \end{aligned}$$

Now recall that $pf_0^2 - \sum_{1}^{n} a_j f_j^2 = 0$ and $f_j = f_0 Q_j + r_j$ for each $j \geq 1$. Then:

$$\begin{aligned} f_0 h_0 = &= pf_0^2 + \sum a_j \big((f_0 Q_j)^2 - 2 f_0 Q_j f_j\big) \\ &= pf_0^2 + \sum a_j \big((f_0 Q_j - f_j)^2 - f_j^2\big) \\ &= pf_0^2 - \sum a_j f_j^2 \; + \; \sum a_j r_j^2 \\ &= \sum_{j=1}^{n} a_j r_j^2. \end{aligned}$$

Could $h_0 = 0$? If so then $\sum_1^n a_j r_j^2 = 0$. But we assumed that γ is anisotropic, and $f_0 h_0 = \gamma(r_1, \ldots, r_n)$. Then $h_0 = 0$ implies that every $r_j = 0$, but then f_0 would divide every f_j. This fails because $\gcd(f_0, f_1, \ldots, f_n) = 1$. Therefore, $h_0 \neq 0$.

By definition of the remainders r_j we know: $\deg r_j < \deg f_0$. Then $\deg\big(\sum a_j r_j^2\big) < 2 \deg f_0$ and we conclude that $\deg h_0 < \deg f_0$. This proves the Claim and completes the proof. $\square$

So we can get rid of denominators! (At least for one variable situations.) This next result is the inductive step used to settle Cassels's original question. Here if (V, φ) is a quadratic space, we write $\langle a \rangle \perp \varphi$ for the space $(\varphi', Kv' \oplus V)$ of one dimension larger, where the new line is perpendicular to V and is spanned by a vector v' with $\varphi'(v') = a$. In polynomial terms, this is the same the form $ay^2 + \varphi(X)$, where y is a new variable. For example, if $\varphi = \langle b_1, \ldots, b_n \rangle$, then $\langle a \rangle \perp \varphi = \langle a, b_1, \ldots, b_n \rangle$.

Proposition. Let φ be a nonsingular quadratic form over K, and $a, d \in K^*$. Suppose the form $\langle a \rangle \perp \varphi$ is anisotropic. Then:

φ represents $ax^2 + d$ over $K(x) \iff \varphi$ represents d over K.

To simplify notations, write $D_K(n)$ for the set $D_K(n\langle 1 \rangle)$, the set of all elements of K^* represented as a sum of n squares in K. If K is formally real then $n\langle 1 \rangle$ is anisotropic for every n. In that case, the Proposition says:

$$x^2 + d \in D_{K(x)}(n+1) \iff d \in D_K(n).$$

Corollary. $x_1^2 + \cdots + x_n^2$ is not the sum of $n-1$ squares in $\mathbb{R}(x_1, \ldots, x_n)$.

Proof that Proposition $\Rightarrow$ Corollary. First check it when $n = 2$: $x_1^2 + x_2^2$ is not a square in $K(x_1, x_2)$. If the Corollary is false, choose a counterexample with minimal $n > 2$. Then $x_1^2 + \cdots x_n^2 \in D_{K(x_1)}(n-1)$ where $K = \mathbb{R}(x_2, \ldots, x_n)$. Apply the Proposition with $x = x_1$ and $d = x_2^2 + \cdots + x_n^2 \in K$, to conclude that $d \in D_K(n-2)$. But this cannot happen by the minimality of n. □

Proof of the Proposition. Suppose $\varphi = \langle a_2, \ldots, a_n \rangle$ and set $a = a_1$. Then $\langle a \rangle \perp \varphi = \langle a_1, \ldots, a_n \rangle$ is anisotropic. We assume there exist $f_j(x) \in K(x)$ such that

$$a_1 x^2 + d = a_1 f_1(x)^2 + \cdots + a_n f_n(x)^2. \qquad (**)$$

By Cassels' Trick, we may assume that each $f_j \in K[x]$ is a polynomial. Now compare degrees. For $d = \max_j \{\deg f_j(x)\}$, then the right side of (**) has degree d. (Look at the x^d coefficient and recall that $\langle a_1, \ldots, a_n \rangle$ is anisotropic.) Then $d = 1$ and $f_j(x) = r_j x + s_j$ for some $r_j, s_j \in K$. Then (**) becomes:

$$a_1 x^2 + d = a_1(r_1 x + s_1)^2 + \cdots + a_n(r_n x + s_n)^2.$$

Why is this impossible? We might try setting $x = 0$, or setting $x = -r_1/s_1$, but those ideas fail. (Why?) Instead, we cleverly choose $c \in K$ that satisfies the equation $(r_1 c + s_1)^2 = c^2$. Plug in this $x = c$ to find an expression showing that d is represented by $\langle a_2, \ldots, a_n \rangle$. □

Let us review the big picture of quadratic forms. There are really two complementary versions of the idea of a quadratic form, the algebraic and the geometric. Algebraically, a quadratic form

is a homogeneous polynomial of degree 2 in n variables

$$Q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

Recall that by a linear change of variables, we can diagonalize:

$$Q'(x_1, \ldots, x_n) = b_1 x_1^2 + \cdots + b_n x_n^2.$$

The geometric version starts with symmetric bilinear forms. Let V be an n dimensional vector space over field K. Let B be a symmetric bilinear form on V. Then $Q(v) = B(v, v)$ is a quadratic form in the geometric sense.

We can translate back and forth between the algebraic and geometric perspectives. To recover the algebraic perspective from the geometric, first choose a basis $e_1, \ldots, e_n$ for V. If $v \in V$ then we can express v in terms of this basis, $v = \sum x_i e_i$, for some $x_i \in K$. Instead of thinking of these x_i as elements of K, though, we can also think of them as formal variables over K. Then $Q(v) = Q(x_1, \ldots, x_n)$ is a quadratic polynomial in the x_i's. If we chose the basis $e_1, \ldots, e_n$ to be an orthogonal basis with respect to B, then Q would end up being a diagonal form $b_1 x_1^2 + \cdots + b_n x_n^2$.

Given an algebraic quadratic form $Q(X)$, we can recover geometric intuition by checking that $2B(X, Y) = Q(X+Y) - Q(X) = Q(Y)$ defines a symmetric bilinear form with $Q(X) = B(X, X)$.

We introduce a partial order relation on the set of quadratic forms K, or, more accurately, on the set of isometry classes of nonsingular quadratic forms.

Definition. Suppose α, β are nonsingular quadratic forms over K, with corresponding quadratic spaces $(V, \alpha), (W, \beta)$. Then α is a *subform* of β, written $\alpha \preccurlyeq \beta$, if there exists an injective isometry $\sigma : (V, \alpha) \to (W, \beta)$. That is, $\sigma : V \to W$ is an injective linear map with $\beta(\sigma(v)) = \alpha(v)$ for every $v \in V$. This says that (V, α) is isometric (as a quadratic space) to a subspace of (W, β).

Exercise. Show: $\langle 1, 1, 1 \rangle \cong \langle 2, 3, 6 \rangle$. Conclude that $\langle 2, 3 \rangle \preccurlyeq \langle 1, 1, 1 \rangle$.
[Here K is a field where $6 \neq 0$. For $Q = \langle 1, 1, 1 \rangle$, the space (K^3, Q) has vector $u = (1, 1, 0)$ with $Q(u) = 2$. The subspace $(v)^\perp$ consists of vectors $v = (a, -a, b)$ with $Q(v) = 2a^2 + b^2$. Then $v = (1, -1, 1)$ has $Q(v) = 3$, so that $U = \text{span}(u, v)$ has form $\langle 2, 3 \rangle$ The last slot can be computed by determinants. (Define $det(V, Q)$ to be the determinant of a Gram matrix and show it is well-defined mod squares.) Alternatively, examine $U^\perp$ and find that $w = (1, 1, 2)$ and check $Q(w) = 6$.

Lemma. If $\alpha \preccurlyeq \beta$. Let $X = (x_1, \ldots, x_n)$ where $n = \dim(\alpha)$. Then β represents the value $\alpha(X)$ over $K(X)$.

Proof. The space (U, α) is isometric to a subspace of (V, β). The complement $U^\perp$ in V inherits a form γ, so that $\beta \cong \alpha \oplus \gamma$. Choosing a basis of V to match that splitting. In polynomial form this

says that there are variables $Y = (y_1, \ldots, y_k)$ where $\dim \beta = n+k$, and $\beta(X, Y) = \alpha(X) + \gamma(Y)$. Then $\alpha(X) = \beta(X, 0) \in D_{K(X)}(\beta)$. $\square$

The amazing thing is that the converse holds true.

Cassels-Pfister Subform Theorem. Let α, β be quadratic forms over K, and assume β is anisotropic. Let $n = \dim \alpha$, and let $X = (x_1, \ldots, x_n)$ be a system of indeterminates. Then

$$\alpha \preccurlyeq \beta \iff \alpha(X) \in D_{K(X)}(\beta).$$

That is: α is a subform of β over K iff β represents the element $\alpha(X)$ over the field $K(X)$.

The motivating example: Let $\alpha = n\langle 1\rangle = \langle 1, \ldots, 1\rangle$ over K, so that $\alpha(X) = x_1^2 + \cdots + x_n^2$. If $\alpha(X)$ is a sum of m squares in $K(X)$, says that the form $m\langle 1\rangle$ represents $\alpha(X)$ over $K(X)$. If K is formally real, the form $\beta = m\langle 1\rangle$ is anisotropic and the Cassels-Pfister Theorem implies $\alpha \preccurlyeq m\langle 1\rangle$, so (in particular) $n = \dim(\alpha) \leq m$. This is the main result of the first half of the lecture! Cassels' result on sums of squares is a special case!

Proof of Theorem. ($\Longrightarrow$) Observed in the Lemma.

($\Longleftarrow$) If the Theorem is false, choose a counterexample α, β with $\dim(\beta)$ minimal. The anisotropic form β represents the value $\alpha(X) = a_1x_1^2 + \cdots + a_nx_n^2$ over the field $K(X)$. Let $a_1 = \alpha(1, 0, \ldots, 0)$. Then β represents a_1 over K. (Why? This follows from Cassels' trick.) Then there is some $v \in (V, \beta)$ such that $\beta(v) = a_1$. Extend to orthogonal basis and express $\beta = \langle a_1\rangle \oplus \beta'$. Express $\alpha(X) = a_1x_1^2 + d$ where $d = a_2x_2^2 + \cdots + a_nx_n^2$. Express $K(X) = E(x_1)$ where $E = K(x_2, \ldots, x_n)$. Then $a_1x_1^2 + d \in D_{E(x_1)}(\langle a_1\rangle \oplus \beta')$ (given). The Proposition above then implies: $d \in D_E(\beta')$. Then β' is an anisotropic form over K, and β' represents $d = a_2x_2^2 + \cdots + a_nx_n^2$ over $E = K(x_2, \ldots, x_n)$. By minimality, the Theorem holds true in this case, and we have $\langle a_2, \ldots, a_n\rangle \preccurlyeq \beta'$. But then $\alpha = \langle a_1\rangle \oplus \langle a_2, \ldots, a_n\rangle \preccurlyeq \langle a_1\rangle \oplus \beta' = \beta$. This shows that the original α, β was not a counterexample after all, completing the proof. $\square$

This result was the starting point for Pfister's "algebraic theory of quadratic forms." He investigated "multiplicative" quadratic forms Q over K: Those where $D_L(Q)$ a group for every extension field L/K. One consequence of his work is that the form $n\langle 1\rangle$ is multiplicative whenever n is a 2-power. That is: For every m, $D_K(2^m)$ is a subgroup of K^*.

We discussed a matrix proof of this result earlier, but that elementary proof (due to Ernst Witt) was found only after Pfister had settled the question.

15 The Level $s(K)$

It's OK to fall asleep in lecture, but not to snore (because you might wake up your neighbor).
- V. Bergelson

Let's review the notion of the level[1] of a field K.

Definition. The level $s(K)$ is the smallest n such that -1 is a sum of n squares in K. If K is formally real, then $s(K) = \infty$. More generally, if $d \in K$, define: $\text{length}_K(d)$ to be the minimum n such that d is a sum of n squares in K. Equivalently,

$$\text{length}_K(d) = n \iff d \in D_K(n), \text{ but } d \notin D_K(n-1).$$

Then $s(K) = \text{length}_K(-1)$.

Exercise. Check that $s(\mathbb{Q}(\sqrt{-2})) = 1$, $s(\mathbb{Q}(\sqrt{-2})) = 2$, and $s(\mathbb{Q}(\sqrt{-7})) = 4$.
Use more sophisticated number theory to prove:
If K is a finite non-real extension of $\mathbb{Q}$ and K, then $s(K) = 1, 2$ or 4.

What levels are possible?

We observed long ago that $s(K) \neq 3$. Proof. If $-1 = a^2 + b^2 + c^2$, then $0 = (1 + a^2)\Big(1 + \frac{b^2+c^2}{1+a^2}\Big)$.
Then $-1 = \frac{b^2+c^2}{1+a^2}$ is a sum of 2 squares, so that $s(K) \leq 2$.
$\big($Explicitly, $-1 = u^2 + v^2$ where $u = \frac{b-ac}{1+a^2}$ and $v = u = \frac{ab+c}{1+a^2}$.$\big)$

The key to this proof is that $D_K(2)$ is a group. As part of his theory of multiplicative forms, Pfister proved that every $D_K(2^m)$ is a group. This led to his determination of possible levels.

Theorem. (Pfister) If K is a non-real field, then $s(K)$ is a power of 2.

Proof. We provided a matrix proof (in the Homework) showing that for every field K, the set $D_K(2^m)$ is a subgroup of K^*. Since K is non-real (i.e., not formally real) the level $s(K)$ is finite and $-1 = a_1^2 + \cdots + a_s^2$, for some $a_j \in K$. Suppose $2^m \leq s < 2^{m+1}$ for some m. If $2^m < s$ then the same trick as before shows that:

$$-1 \;=\; \frac{a_{2^m}^2 + \cdots + a_s^2}{1 + a_1^2 + \cdots a_{2^m-1}^2}.$$

This is in the group $D_K(2^m)$, contradicting $2^m < s$. Therefore, $s = 2^m$. □

Pfister considered the natural question: If m is given, does there exist a field K with level $s(K) = 2^m$.

It's interesting to investigate those sets $D_K(n)$, the set of all $cin K^*$ expressible as a sum of n squares in K. We stick to one field K and suppress the subscript for now. Since every $D(2^m)$ is a group, we know that, say, $D(2)D(3) \subseteq D(4)$. But those sets are actually equal!

[1]The audience provided additional 5-letter palindromes in English (no proper names). Our list so far is: civic, kayak, level, madam, minim, radar, refer, rotor, sagas, sexes, shahs, solos, stats, tenet. Less common words include: rever, semes, stets, stots. Any more?

For if $u \in D(4)$ then express

$$u = a^2 + b^2 + c^2 + d^2 \ = \ (a^2 + b^2)\Big(1 + \tfrac{c^2+d^2}{a^2+b^2}\Big),$$

and that second term is in $D(3)$.

Exercise. (1) Show that $D(4)D(5) = D(8)$.
(2) Explain why $D(3)D(5) \subseteq D(7)$. Show that those sets are equal.
[First part: $D(3)D(5) \subseteq D(3)\big(D(4) + D(1)\big) \subseteq D(3)D(4) + D(3)D(1) \subseteq D(4) + D(3) = D(7)$.]
(3) For any $r, s \in \mathbb{Z}^+$ there exists a number $r \circ s$ such that:

$$D_K(r)D_K(s) = D_K(r \circ s), \text{ for every field } K.$$

For instance, $2^m \circ 2^m = 2^m$, $\quad 4 \circ 5 = 8$, and $3 \circ 5 = 7$.
(4) Investigate this operation $\circ$. First step: Is this $r \circ s$ well-defined? This requires a Lemma:
If $m \neq n$ then $D_K(m)$ is not always equal to $D_K(n)$.
This Lemma follows from Cassels' trick: If $K = \mathbb{R}(x_1, \ldots, x_n)$ then $D_K(n) \neq D_K(n-1)$.

More information on that "circle" operation appears in Lecture 1 on Sums of Squares Identities, posted at Shapiro home page .

The natural question is: Does there exist a field of given level 2^m? Let's consider a generic sort of ring in which -1 is a sum of n squares.

$$A_n \ = \ \frac{\mathbb{R}[x_1, \ldots, x_n]}{(1 + x_1^2 + \cdots + x_n^2)}.$$

This A_n is an integral domain, and we define K_n to be its field of fractions. Then certainly $s(K_n) \leq n$. Check that

$$K_n \cong \mathbb{R}(x_1, \ldots, x_{n-1})(\sqrt{-d}) \ \text{ where } \ d = 1 + x_1^2 + \cdots + x_{n-1}^2.$$

Pfister was able to prove that this field answers the question:

Theorem. (Pfister) If $2^m \leq n < 2^{m+1}$ then $s(K_n) = 2^m$.

More generally, if L/K is a finite extension, say $L = K(\theta)$ where θ has minimal polynomial $f(x) \in K[x]$, how can we determine the level $s(L)$ from information about K and $f(x)$?

If we define the level of a ring, does $s(A_n) = n$?

But we don't have time left. You are invited to read further about these sorts of questions by searching in Wikipedia, or by reading about levels in one of Lam's books.

16 $x^3 + y^3$ and $G - G = K$

It is impossible to be a mathematician without being a poet in soul. - Sofia Kovalevskaya

A problem on the Ross sets asked for a proof that every element of $\mathbb{Z}/p\mathbb{Z}$ is a sum of two squares. If we replace "squares" with "cubes" is the claim still true? Certainly if $p \not\equiv 1 \pmod 3$ then every element of $\mathbb{Z}/p\mathbb{Z}$ is a cube that the result is trivial.

Exercise. If K is a field define $\sigma_m : K \to K$ by: $\sigma_m(a) = a^m$. If $K = \mathbb{F}_q$ is the field of q elements, then: σ_m is bijective if and only if m and $q-1$ are coprime.
In this case, every element of K is an m^{th} power.

For primes $p \equiv 1 \pmod 3$, the set of nonzero cubes in K is the unique subgroup of index 3 (and order $(p-1)/3$) in K^*. Checking the case $p = 7$ we see that the numbers 3 and 4 are not expressible as a sum of two cubes in $\mathbb{Z}/7\mathbb{Z}$. The surprising result is:

Proposition. For every prime $p \neq 7$, each element of $\mathbb{Z}/p\mathbb{Z}$ is the sum of two cubes.

Today we prove this, following ideas told to me by Pedro Berrizbeitia.

First, we can without loss of generality assume that $p = 1 \pmod 3$ and let G be the set of cubes in $U_p = (\mathbb{Z}/p\mathbb{Z})^*$. Then G is a subgroup of U_p of index 3. If $c \notin G$ (that is, if c is a non-cube), then G, cG, and c^2G are disjoint and

$$U_p = (G) \cup (cG) \cup (c^2G).$$

For those of you who have seen some group theory, this is obvious. If you are not familiar with group theory, it's is a good exercise.

Now let q be the smallest positive integer that is not a cube mod p. (Check that q is prime.) Any element of G is the sum of two cubes in $\mathbb{Z}/p\mathbb{Z}$ (since $x^3 + 0^3 = x^3$). So we hope to show that every element of qG and every element of q^2G is the sum of two cubes. For this, it suffices to show that each of q and q^2 is a sum of two cubes. That fact for q is easy: $q = 1 + (q-1)$, and $q-1$ is a cube since it is a positive integer smaller than q. So we just need to check that q^2 is the sum of two cubes. Suppose q is odd. Then $q^2 = 1 + (q^2 - 1) = 1 + (q-1)(q+1)$. If you squint you'll find that $q+1$ is a cube. To see why, note that 2 and $(q+1)/2$ are smaller than q. The last case is when q is even, so that $q = 2$. Then

$$U_p = (G) \cup (2G) \cup (4G)$$

We need to show that 4 is the sum of two cubes. Suppose not, that is: $4 \notin G + G$. Then $4G \cap (G + G) = \varnothing$ (easy exercise). We restate this in two more ways: $G \cap (2G + 2G) = \varnothing$, and $2G \cap (4G + 4G) = \varnothing$ (also an easy exercise). Now let's bash some cases. Since $p \equiv 1 \pmod 3$ we know $p > 5$.

Claim. $3 \in 4G$.
If $3 \in G$ then $4 = 3 + 1 \in G + G$, contradicting our assumption that $4 \notin G + G$. If $3 \in 2G$ then

$1 = 3 - 2 \in 2G + 2G$. But $1 \in G$ so this contradicts the fact that G and $2G + 2G$ are disjoint. The only remaining possibility is: $3 \in 4G$.

Claim. $5 \in 2G$.
If $5 \in G$ then $4 = 5 - 1 \in G + G$, a contradiction. If $5 \in 4G$ then $2 = 5 - 3 \in 4G + 4G$, contradicting the fact that $2G$ and $4G + 4G$ are disjoint. This forces $5 \in 2G$.

Where is 7? Check that $7 = 8 - 1 \in G + G$, and $7 = 5 + 2 \in 2G + 2G$, and $7 = 3 + 4 \in 4G + 4G$. By the hypotheses, it follows that 7 cannot be in any of those three cosets. That is, 7 is not in U_p forcing $7 = 0$ in $\mathbb{Z}/p\mathbb{Z}$. Then $p = 7$.

We have proved: if $4 \notin G + G$, then $p = 7$. Taking the contrapositive: If $p \neq 7$, then $4 \in G + G$. This concludes the proof of the Proposition. □

We have answered the question. Now let's question the answer.

(Q1) What if we don't allow zero as a cube? That is, we want every c expressed as $x^3 + y^3$ for *non-zero* x, y in $\mathbb{Z}/p\mathbb{Z}$.

In this case, $p = 7$ and $p = 13$ are immediate counterexamples. As an exercise, you can PODASIP the following claim: "If $p \neq 7, 13$ is prime, then $G + G = \mathbb{Z}/p\mathbb{Z}$."

(Q2) When $q = p^m$ is a prime power, is every $c \in \mathbb{F}_q$ expressible as $c = x^3 + y^3$ for $x, y \in \mathbb{F}_q$?

Again, we may without loss of generosity assume that $q \equiv 1 \pmod 3$. As before, some exceptions arise for small sizes:

Proposition. Suppose q is a prime power, and G is the set of nonzero cubes in $\mathbb{F}_q$. Then $G + G = \mathbb{F}_q$, except when $q = 4, 7, 13$ or 16.

(Q3) Another direction to poke: what about infinite fields?

For example, in the field $\mathbb{Q}$ the subgroup of cubes in $\mathbb{Q}^*$ has infinite index. So maybe we want forget about cubes and instead consider subgroups of index 3.

If $G \subseteq \mathbb{Q}^*$ is a subgroup index 3, then does $G + G = \mathbb{Q}$?

Wait! Are there any index 3 subgroups of $\mathbb{Q}^*$? Well, every rational number can be written uniquely as $\pm 2^{n_2} 3^{n_3} 5^{n_5} 7^{n_7} \cdots$ where $n_j \in \mathbb{Z}$ and all but finitely many $n_j = 0$. Given a prime p, let G_p be the set of rational numbers where the exponent of p is divisible by 3. Then G_p is not the subgroup of cubes, but it is the subgroup of "cubes-at-p." You can show easily that each G_p has index 3. Does every subgroup of index 3 equal one of the groups G_p? This is an exercise for you to figure out.

Anyway, if G has index 3 in $\mathbb{Q}^*$, does $G + G = \mathbb{Q}$? You can prove this with a version of the method used above. Those ideas do generalize to all fields:

Theorem. Let K be a field, and let G be an index 3 subgroup of K^*. Then $G+G=K$, except when $K=\mathbb{F}_7$, $\mathbb{F}_4$, $\mathbb{F}_{13}$, or $\mathbb{F}_{16}$.

(Q4)] The larger question is: What about $G+G$ for other values of 3? That is, if G is a subgroup of K^* with index 2 or 4 or n, does $G+G=K$?

Watch out! The $n=2$ case already fails! For instance, $\mathbb{R}^+$ (positive reals) is an index 2 subgroup of $\mathbb{R}^*$, but $\mathbb{R}^+ + \mathbb{R}^+ = \mathbb{R}^+$. (Positives are close under addition.)

Let's address this situation by changing the question! Berrizbeitia's insight was to ask about $G-G$ instead of $G+G$. Note that if the index n is odd, then -1 must be in G. (Why?) If $-1 \in G$ then $G+G=G-G$. So we can expect a more uniform result if we use $G-G$.

New Hope. If K is a field and G is a subgroup of K^* of finite index, then does $G-G=K$?

We analyze this question when K is an infinite field. For completeness, I'll remark that the question over finite fields can be settled using classical methods (Jacobi sums, etc). Suppose $q \equiv 1 \pmod{n}$ and G_n is the subgroup of n^{th} powers in $\mathbb{F}_q^*$. Let's say that q is "good for n" if $G_n - G_n = \mathbb{F}_q$. The theorem is: Given n, only finitely many q are not good for n.

Berrizbeitia was able to prove that if K is any field of characteristic zero and G is a finite index subgroup of K^*, then $G-G=K$. His proof used the fact that $\mathbb{Z} \subseteq K$ and he could apply van der Waerden's theorem about long arithmetic progressions in $\mathbb{Z}$. What a wonderful idea to apply van der Waerden to the problem! Unfortunately, this method fails for a field K of finite characteristic because $\mathbb{Z}$ is not inside such K.

Shapiro tried to prove it in general, but got stuck. So he asked ... Bergleson, who said, "I'll think about it." He solved it that night. In honor of Bergleson, we introduce a Black Box:

Ramsey Theory.

Let K_6 be the complete graph on 6 vertices and 2-color the edges of K_6 (each edge is red or blue). Then there exists a monochromatic triangle. This is easy to prove. A vertex v has 5 edges coming out of it. At least 3 of these are the same color, so assume those 3 are red. Look at the edges connecting the other ends of those 3 edges. There are 3 of those edges, and they form a triangle. If one of them is red, then there is a red triangle containing v. If none of them is red, then they are all blue, and so they form a blue triangle. Either way, there is a monochromatic triangle. This is motivation for Ramsey's Theorem.

For a set S let $[S]^2$ be the set of all the two element subsets of S. The edges of the graph K_6 correspond to $[S]^2$ for the set S of 6 vertices.

Ramsey's Theorem. Suppose S is an infinite set and $[S]^2$ is finitely colored. Then there is an infinite subset $A \subseteq S$ such that $[A]^2$ is monochromatic.

There are various finite versions of this theorem, but this infinite version is what we will use.

Let's state our theorem and start the proof.

THEOREM. Suppose K is an infinite field and G is a subgroup of K^* with finite index. Then $G - G = K$.

Proof. Decompose K^* into disjoint cosets: $K^* = (c_1G) \cup \cdots \cup (c_rG)$.
We want to define a coloring of $[K^*]^2$ using those cosets as the colors, assigning to an element $\{a, b\}$ of $[K^*]^2$ the "color" $(a - b)G$. But wait! Those are un-ordered sets, so we don't know whether to use color $(a - b)G$, or $(b - a)G$.

Let's begin the proof again. Choose an infinite indexed subset $S = \{a_1, a_2, \dots\} \subseteq K^*$. For $i > j$, the element $\{a_i, a_j\}$ in $[S]^2$ is assigned the color $(a_i - a_j)G$.

By Ramsey's Theorem there exists an infinite subset B of S such that $[B]^2$ is monochromatic, say with color dG. Writing $B = \{a_{k_1}, a_{k_2}, \dots\}$, where $k_1 < k_2 < \cdots$, then: $a_{k_i} - a_{k_j} \in dG$ whenever $i > j$. Re-index by setting $b_j = a_{k_j}$ for each j. That way, $B = \{b_1, b_2, \dots\}$ is an infinite subset of K^*, and $b_i - b_j \in dG$ whenever $i > j$.

CLAIM. Some coset cG has the following property:
For each $x \in K^*$, the sets $xb_1 + cG,\ xb_2 + cG, \dots$ are not all disjoint.
Assuming this holds, there exist $i > j$ such that $xb_i + cG$ and $xb_j + cG$ meet. That is, there exist $g, g' \in G$ such that $xb_i + cg = xb_j + cg'$. Then $x(b_i - b_j) = c(g' - g)$. But $b_i - b_j \in dG$, so there exists $g'' \in G$ such that $xdg'' = c(g' - g)$. Therefore $c^{-1}dx = (g - g')(g'')^{-1} \in G - G$. Since this holds for every $x \in K^*$, we see that $K^* \subseteq G - G$. It follows that $K = G - G$, as hoped.

To prove that Claim we invoke another Black Box:

Amenable Groups.

The Banach-Tarski paradox states that it's possible to cut a solid unit ball in $\mathbb{R}^3$ into a finite number of pieces, and then to rearrange those pieces (using translations and rotations) to produce two disjoint solid unit balls. This is certainly a violation of intuition, but is mathematically correct (the pieces involved are so complicated that the concept of "volume" does not apply to them). This famous result from the 1920s added flames to discussions of whether the Axiom of Choice is a reasonable assumption for the foundations of set theory.

The study of this apparent paradox led to the theory of amenable groups (groups that possess an "invariant mean"). For a set S let $\mathcal{P}(S)$ be the power-set of S, the set of all subsets of S.

Definition. A group G is *amenable* if there exists a finitely additive G-invariant probability measure μ on G. This means that there is a function $\mu : \mathcal{P}(G) \to [0, 1]$ satisfying:
$\mu(A \cup B) = \mu(A) + \mu(B)$ whenever A, B are disjoint subsets of G.

$\mu(gA) = \mu(A)$ for every $g \in G$ and every subset $A \in \mathcal{P}(G)$.
$\mu(G) = 1$.

Note: $\mu(A)$ is defined for every subset A. The contrasts with Lebesgue measure where "non-measurable sets" are standard fare.

In fancy language, the Banach-Tarski paradox is possible because the rotation group SO(3) is not amenable. Explanation of the relationship between that paradox and amenability is explained in Stan Wagon's wonderful book *The Banach-Tarski Paradox*.

The Black Box we need is the following result.

Theorem. Every commutative group is amenable.

Consequently, for our infinite field K, there exists a finitely additive probability measure $\mu : \mathcal{P}(K) \to [0, 1]$ that is additively-invariant:
$\mu(c + A) = \mu(A)$ for every $c \in K$ and every $A \subseteq K$.

For such a measure μ, every singleton set has measure 0. To prove this, note that for any $a, b \in K$, additive-invariance implies $\mu(\{a\}) = \mu(\{b\})$ (because $(b - a) + \{a\} = \{b\}$). If the measure of a singleton is positive, then (since K is infinite) measures of finite subsets of K can be arbitrarily large. This contradicts the fact that $\mu(K) = 1$.
Then $\mu(\{0\}) = 0$ and therefore: $\mu(K^*) = 1$.

Finally, we can prove the Claim.

Since K^* has measure 1 and K^* is a finite union of cosets aG, there is some $c \in K^*$ such that $\mu(cG) > 0$. All those sets $xb_i + cG$ have measure equal to $\mu(cG)$. If they are pairwise disjoint for $i = 1, 2, \ldots$ then the union $A_n = (xb_1 + cG) \cup (xb_2 + cG) \cup \cdots \cup (xc_n + cG)$ has measure $\mu(A_n) = n\mu(cG)$. This cannot be true for every n since $\mu(K^*) = 1$.

This completes the proof of the Theorem! □

Remark. Are there doubly-invariant measures on K? We are looking for a measure μ as above that is invariant under *affine transformations*: $\mu(A) = \mu(c + uA)$ for every $A \subseteq K$, every $c \in K$ and every $u \in K^*$. That affine group is not commutative, so the previous Black Box theorem fails. However, such measures do exist because the affine group is solvable, and there is a Bigger Blacker Box: **Theorem.** Every solvable group is amenable.

We can push this further by considering a division ring D, (an associative ring in which every nonzero element is a unit, but multiplication is not necessarily commutative). Check that the same proof shows: If D is an infinite division ring and $G \leq D^*$ is a subgroup of finite index,

then $G - G = D$.

Even more generally, define a ring R to be a "$G - G$ ring" if $G - G = R$ for every subgroup $G \leq R^*$ of finite index. Then every infinite division ring is a $G - G$ ring.

17 References

Hello! Welcome to the Department of Redundancy and Repetition Department. Hi!

Mathematical:

[1] E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, *Abh. math. Sem. Hamburg*, **5** (1927) 110 - 115.
Solution to Hilbert's 17^{th} problem.

[2] E. Artin and O. Schreier, Eine Kennzeichnung der reell abgeschlossenen Körper, *Abh. math. Sem. Univ. Hamburg*, **5** (1927) 225-231.
Where real closed fields were introduced.

[3] V. Bergelson and D. Shapiro, Multiplicative subgroups of finite index in a ring, *Proc. Amer. Math. Soc.* **166** (1992) 885-896.

[4] N. Jacobson, *Basic Algebra I, II*, W. H. Freeman, 1985, 1989.
Jacobson's books use somewhat non-standard notations but are packed with information. He includes many of the same topics as Lang, often with different proofs.

[5] T. Y. Lam, *The Algebraic Theory of Quadratic Forms*, W. A. Benjamin/Addison Wesley, 1973. Revised: 1980.
Excellent exposition of the early state of the theory. Lam is a wonderful writer.

[6] T. Y. Lam, *Introduction to Quadratic Forms over Fields*, Amer. Math. Soc., Graduate Studies in Math., vol. 67, 2004.
A more encyclopedic presentation.

[7] S. Lang, *Algebra*, 3rd ed., Springer, 2002.
Classic text from the 1960s is an updated and expanded version of van der Waerden. Terse but full of information. Includes Nullstellensatz, real closed fields, Artin-Lang Homomorphism Theorem, etc.

[8] A. Pfister, Zur Darstellung von -1 als Summe von Quadraten in einem Körper, *J.London Math. Soc.* **40** (1965) 159-165.
Pfister's first major paper, proving: If the level is finite it must be a 2-power.

[9] A. Pfister, *Quadratic Forms with Applications to Algebraic Geometry and Topology*, London Math. Soc. Lec. Notes 217, Cambridge University Press, 1995.
Well written text including Subform Theorem, Nullstellensatz, 17^{th} problem, levels of fields and rings, etc.

[10] W. Scharlau, *Quadratic and Hermitian Forms*, Grundlehren 270, Springer, 1985.
A unified treatment of forms, includes real closed fields, Hilbert's 17^{th} problem, C_i fields, and much more. Parallels Lam's books.

[11] D. Shapiro, *Compositions of Quadratic Forms*, W. de Gruyter, 2000.
Full of information but hard to read.

[12] B. L. van der Waerden, *Modern Algebra*, in two volumes, F. Ungar, 1950.
Translation of 1931 German edition. Caption: "Using lectures by E. Artin and E. Noether."
The most influential algebra text of the twentieth century.

Historical:

[13] F. Browder, ed., *Mathematical Developments Arising from Hilbert Problems*, Proc Sympos. Pure Math., vol 28, Amer. Math. Soc, 1976.
Contains chapter by Pfister on the 17^{th} Problem.

[14] D. Hilbert, Mathematical Problems,
Lecture delivered before the International Congress of Mathematics in Paris in 1900,
1902 English translation posted at: Hilbert address

[15] C. Reid, *Hilbert*, Springer, 1970.
Definitive biography of Hilbert, and a wonderful source for information about Noether, Artin, van der Waerden, etc.

[16] B. Yandell, *The Honors Class, Hilbert's Problems and Their Solvers*, A. K. Peters/CRC Press, 2002.
Many interesting stories and some math.