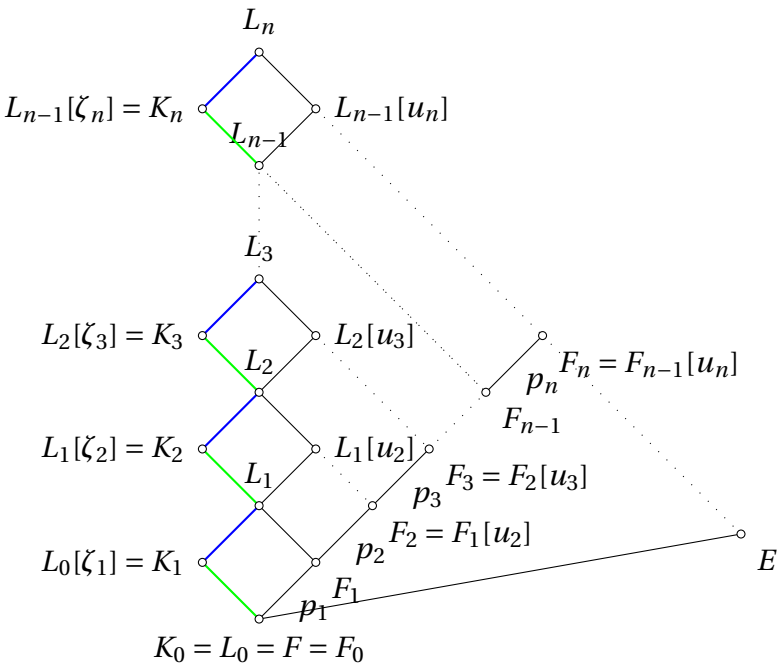
GEORGE MCNULTY

Groups and Fields

First Year Graduate Algebra
Part II

DRAWINGS BY THE AUTHOR



UNIVERSITY OF SOUTH CAROLINA

2015

PREFACE

The first part of this account of first-year graduate algebra was devoted to the essentials of rings and modules. There the central ambition was to give an account of the theory of unique factorization and the fundemental structure theorem for finitely generated modules over a principal ideal domain.

This second part begins with a presentation of the basics of the theory of groups, a presentation which is rapid in view of the sophistication gained by the graduate students in the first part of the course. But the real focus here will be effort to lay our hands on the roots of polynomials whose coefficients all lie in some given field. In short, our most important goal will be a development of Galois theory. Along the way, we will see how devise algebraically closed fields, we will have in hand a proof of the Fundamental Theorem of Algebra, a proof of the transcendence of the numbers π and e, as well as a proof of Hilbert's Nullstellen Satz, which illuminates the Galois connection between algebraic geometry and the ring of polynomials in several variables over an algebraically closed field.

Amongst this rich array of material are the solutions to open problems of long standing (some for thousands of years) devised by mathematicians in the 19^{th} century. You will see why it is impossible to trisect arbitrary angles, duplicate the cube, or square the circle with only straightedge and compass—and why there is no formula similar to the quadratic formula for describing roots of polynomials of degree 5 or higher.

Once more, as you proceed through these pages you will find many places where the details and sometimes whole proofs of theorems will be left in your hands. The way to get the most from this presentation is to take it on with paper and pencil in hand and do this work as you go. There are also weekly problem sets. Most of the problems have appeared on Ph.D. examinations at various universities. In a real sense, the problems sets are the real heart of this presentation.

George F. McNulty
Columbia, SC
2015

CONTENTS

Preface ii

LECTURE 1 **Concrete Groups** 1
1.1 Problem Set 12 8

LECTURE 2 **The theory of abstract groups: getting off the ground** 9
2.1 Defining the class of groups by equations 9
2.2 Homomorphisms and their kernels—and an extraordinary property of subgroups 12
2.3 Problems Set 13 16

LECTURE 3 **Isomorphism Theorems: the Group Versions** 17
3.1 Problem Set 14 20

LECTURE 4 **Useful Facts About Cyclic Groups** 21

LECTURE 5 **Group representation: Groups acting on sets** 26
5.1 Problem Set 17 31

LECTURE 6 **When Does a Finite Group Have a Subgroup of Size n?** 32
6.1 Problems Set 15 37

LECTURE 7 **Decomposing Finite Groups** 38
7.1 Direct products of groups 38
7.2 Decomposing a group using a chain of subgroups 40
7.3 Addendum: A notion related to solvability 46
7.4 Problem Set 16 48

LECTURE 8 **Where to Find the Roots of a Polynomial** 49

LECTURE 9 **Algebraically closed fields** 57

LECTURE 10 **Constructions by straightedge and compass** 61

LECTURE 11 **Galois Connections** 66
11.1 Abstract Galois Connections 66
11.2 The Connection of Galois 68
11.3 Problem Set 18 70

LECTURE 12 **The Field Side of Galois' Connection** 71
12.1 Perfect Fields 72
12.2 Galois Extensions 74
12.3 Problem Set 19 76

LECTURE 13 **The Group Side of Galois' Connection and the Fundamental Theorem** 77
13.1 Closed subgroups of a Galois group 77
13.2 The Fundamental Theorem of Galois Theory 80

LECTURE 14 **Galois' Criteria for Solvability by Radicals** 82

LECTURE 15 **Polynomials and Their Galois Groups** 88
15.1 Problem Set 20 95

LECTURE 16 **Algebraic Closures of Real-Closed Fields** 96
16.1 Problem Set 21 101

LECTURE 17 **Gauss on Constructing Regular Polygons by Straightedge and Compass** 102
17.1 Problem Set 22 103

LECTURE 18 **Algebraic Integers** 104

LECTURE 19 **The Lindemann-Weierstrass Theorem on Transcendental Numbers** 106
19.1 Problem Set 23 108

LECTURE 20 **The Galois Connection between Rings of Polynomials and Points in Affine Spaces** 109

LECTURE 1

CONCRETE GROUPS

Consider the Euclidean plane $\mathbb{P}$. As a mathematical system we might construe $\mathbb{P}$ as $\langle P, B, E\rangle$ where P is the set of points on the plane, B is the three-place betweeness relation among points (we want $B(a,b,c)$ to mean that he point b is on the line segmentjoining the points a and c) and E is the four-place equibistance relation among points (we want $E(a,b,c,d)$ to means that the distance between points a and b is the same as the distance between the points c and d, that is the line segment joining a and b is congruent to the line segment joining c and d). Were this a course in Euclidean geometry we would consider in detail the maps from plane into the plane that preserved the relations B and E. A bit of thought should lead hard-working graduate students to the conclusion that among these maps are the maps that preserve distance. That is σ is such a map provided for all points a and b the distance from a to b is the same as the distance from $\sigma(a)$ to $\sigma(b)$. The fancy word from distance-preserving maps is *isometry*. More plain spoken folks call these rigid motions.

There are lots of isometries. For example, translating every point 2 units of distance to the northwest is an isometry. You could pick an arbitrary point as a center, and rotate the whole plane about that point by some angle θ. You could pick an arbitrary line and reflect the plane across the line.

The rigid motion easiest to understand is the one that does nothing: the identity map. Rigid motions can be composed and the result is again a rigid motion—one might first perform a translation and follow that by a reflection, for example. Each rigid motion is plainly one-to-one. It takes a bit a thought to see that they must map P onto P. This means that each rigid motion can be inverted. The hard-working graduate students can see that the inverse is again a rigid motion. In this way, more complex rigid motions can be devised by repeatedly composing and inverting the translations, rotations, and reflections. An interesting exercise, well within the grasp of hard-working graduate students, is to determine all of the isometries of the plane. Let $\mathbb{I}$ denote the set of all isometries of the plane.

There are some other maps the preserve the relations B and E. Here is one example. Fix a particular point $a \in \mathbb{P}$. Let σ be the map that sends any $q \in \mathbb{P}$ to the midpoint of the line segment joining p and q (and sending p to itself). This map is a contraction toward the point p. There

are, of course, other contraction, to say nothing of expansions. Let $\mathrm{Aut}\mathbb{P}$ be the collection of all automorphism of the plane—that is all the one-to-one maps from $\mathbb{P}$ onto $\mathbb{P}$ that preserve both the relations B and E. This collection includes all the isometries but is larger since it also includes all the expansions and contractions, as well as all the maps that arise from them by way of composition. These maps are sometimes called similarities. The hardy graduate student may try to classify all the maps that belong to $\mathrm{Aut}\mathbb{P}$.

Here is a similar situation. Let $\mathbb{R}_2$ denote the two-dimensional vector space over the field of real numbers. The automorphisms of $\mathbb{R}_2$ are just the invertible linear operators on this vector space. While we may identify the vectors with points on the plane, the vector space $\mathbb{R}_2$ and the Euclidean plane $\mathbb{P}$ are not the same. For example, $\mathbb{R}_2$ gives a special role to the origin whereas any point of $\mathbb{P}$ is like any other point. Also $\mathrm{Aut}\mathbb{R}_2$ and $\mathrm{Aut}\mathbb{P}$ are different as well. Each automorphism of $\mathbb{R}_2$ fix the origin. So there are no nontrivial translations in $\mathrm{Aut}\mathbb{R}_2$, the only rotations must use the origin as their centers, and the only reflections must be reflections across lines through the origin. So a lot of maps in $\mathrm{Aut}\mathbb{P}$ seem to be missing from $\mathrm{Aut}\mathbb{R}_2$. On the other hand, $\mathrm{Aut}\mathbb{R}_2$ has maps that are not rigid motions. For example, in $\mathrm{Aut}\mathbb{R}_2$ one can find scalings, that is maps which stretch or shrink vectors. This is effected by multiplying by a fixed nonzero scalar. At any rate, $\mathrm{Aut}\mathbb{R}_2$ contains the identity map, it is closed under composition of linear operators, and it is also closed under the formation of inverse of linear operators.

Here is a related situation. Let us consider just a part of $\mathrm{Aut}\mathbb{R}_2$, namely those linear operators with determinant 1. (It may hlep to think of each linear operator as a 2×2 matrix.) Let S denote this collection of more specialized invertible linear operators. The only scalings that remain in S are multiplication by 1 (namely, the identity map) and multiplication be -1. However, S is still pretty rich. To all intents and purposes, it is the collection of 2×2 matrices with real entries that have determinant 1. As with the other cases, the identity map belongs to S and S is closed under composition of operators and under the formation of inverses of operators.

We have four examples: $\mathbb{I}$, $\mathrm{Aut}\mathbb{P}$, $\mathrm{Aut}\mathbb{R}_2$, and S. Each of these is a collection of one-to-one function from some set onto itself. Each of these collections includes the identity map and is closed under composition of functions and under the formation of inverse functions. In a sense, each of these are collections of second order objects that are functions on some (first-order) mathematical system. Evidently, we could derive such collections of the second-order from a wide assortment of mathematical systems. We can convert these three sets, and any others that arise in a similar way, into algebraic systems (algebras, for short) as

$$\langle \mathrm{Aut}\mathbb{P}, \circ, ^{-1}, \mathbf{1} \rangle \qquad \langle \mathrm{Aut}\mathbb{R}_2, \circ, ^{-1}, \mathbf{1} \rangle \qquad \langle S, \circ, ^{-1}, \mathbf{1} \rangle.$$

Here we use $\mathbf{1}$ to denote the identity map, $\circ$ to denote the composition of functions, and $^{-1}$ to denote the formation of inverses. These are algebras whose signature provides one two-place operation symbol to desginate the composition, a single one-place operation symbol to designate the formation of inverses, and an operation symbol of rank 0 to designate the identity.

The general situation, of which these are three special cases, starts with a set X. In our first example X is the Euclidean plane. We consider a set G of one-to-one maps from X onto X that includes the identity map $\mathbf{1}_X$ on X and that is closed with respect to both the composition of functions and the formation of inverse functions. The resulting algebra

$$\langle G, \circ, ^{-1}, \mathbf{1}_X \rangle$$

is called a **concrete group**. Since X can be any set and the selection of a particular G given X is unrestrained, apart from the closure conditions, there is quite a rich assortment of concrete

groups. Even so, we see that for concrete groups we know very well how the operations of functional composition and the formation of inverse functions work and we have a firm grip on the identity map.

A **group** is any algebra that is isomorphic to a concrete group. It is interesting to note that the the concept of a group arises by the process of abstraction from its concrete instances. Loosely speaking, an (abstract) group is a mathematical system that shares all its "algebraic" properties with some concrete group. This process differs from the process of generalization which prompted the concept of a ring. There, the idea was to extract from many particular instances, like the integers or the system of 2×2 matrices, a set of common properties. A ring was any algebra that had the selected common properties. One should notice that the properties we selected in coming to the notion of a ring were conventional and practical—that is, they were convenient properties like the distributive law, which arose again and again in practice. The theory of rings, in some sense, is the working out of the logical consequences of these selected properties. While these properties of plus and times are fairly natural in that they arose in the course of millennia of mathematical practice, there does not appear to be anything absolutely inevitable about them. The notion of a group, on the other hand, did not arise through the selection of a set of properties but rather through the selection of a class of concrete instances.

Groups of Permutations

Before turning to the theory of abstract groups, we will develop the first facts about concrete groups.

Let X be any set. A **permutation** on X is just a one-to-one function from X onto X. We use $\operatorname{Sym} X$ to the denote the set of all permutations on X and $\mathbf{Sym}\, X$ to denote the (concrete) group $\langle \operatorname{Sym} X, \circ, ^{-1}, \mathbf{1}_X \rangle$. We refer to this group as the **symmetric group** on X or sometimes as the groups of symmetries of X.

If Y is a set such that $|Y| = |X|$, that is if X and Y have the same cardinality, then $\mathbf{Sym}\, X$ and $\mathbf{Sym}\, Y$ will be isomorphic. Indeed, suppose that f is a one-to-one correspondence from X to Y. Then the map

$$\sigma \mapsto f \circ \sigma \circ f^{-1}$$

for all $\sigma \in \operatorname{Sym} X$, turns out to be an isomorphism from $\mathbf{Sym}\, X$ to $\mathbf{Sym}\, Y$, as the hard-working graduate can check.

For many purposes the thing about X that really matters for $\mathbf{Sym}\, X$ is the cardinality of X. This being the case, for a cardinal κ we use $\mathbf{S}_\kappa$ to denote the concrete group of all permutations of κ. Here we take κ to be the set of all ordinals strictly smaller than κ. When κ is finite, this means that $\kappa = \{0, 1, 2, \ldots, \kappa - 1\}$. For example, $6 = \{0, 1, 2, 3, 4, 5\}$. So $\mathbf{S}_6 = \mathbf{Sym}\{0, 1, 2, 3, 4, 5\}$.

Let $x \in X$ and $\sigma \in \operatorname{Sym} X$. The set

$$\{\sigma^k(x) \mid k \in \mathbb{Z}\}$$

is called the **orbit** of x under the action of σ. For a fixed permutation σ, the set X is actually partitioned into orbits. The equivalence relation that lies behind this partition makes elements $x, y \in X$ equivalent if and only if $\sigma^k(x) = y$ for some integer k. Notice that an orbit is either countably infinite or finite. The countably infinite orbits arise when $\sigma^k(x) \neq \sigma^j(x)$ whenever $k \neq j$. These orbits can the arranged like the integers:

$$(\ldots, \sigma^{-3}(x), \sigma^{-2}(x), \sigma^{-1}(x), \sigma^0(x), \sigma^1(x), \sigma^2(x), \sigma^3(x), \ldots)$$

where, of course, $\sigma^0(x) = x$. Suppose, on the other hand, that $\sigma^k(x) = \sigma^j(x)$ where $j < k$. Then some fiddling reveals that $\sigma^{k-j}(x) = x$. Let n be the smallest positive integer such that $\sigma^n(x) = x$. Then the orbit of x under the action of σ turns out to be $\{x, \sigma(x), \ldots, \sigma^{n-1}(x)\}$, as checked by every one of the hard-working graduate students. We can also regard this orbit as a kind of arranged list:

$$(x, \sigma(x), \sigma^2(x), \sigma^3(x), \ldots, \sigma^{n-1}(x))$$

so long as we think of it as a linear representation of a circular arrangement—that is we think of $\sigma^{n-1}(x)$ the (unrepresented) predecessor of x. We could represent the permutation σ by simply listing all these arranged orbits. For instance, here is such a representation of one member of S_6:

$$\sigma = (0,2,4)(1)(5,3)$$

This is a compact way that writing

$$\begin{aligned}
\sigma(0) &= 2 \\
\sigma(2) &= 4 \\
\sigma(4) &= 0 \\
\sigma(1) &= 1 \\
\sigma(5) &= 3 \\
\sigma(3) &= 5.
\end{aligned}$$

The natural number 1 is a fixed point of σ. By convention, fixed points are omitted from the representation. So we arrive at

$$\sigma = (0,2,4)(5,3).$$

The two parts in this representation are called **cycles**. They are cyclic representations of the two nontrivial (here that means have at least two elements) orbits into which σ partitions $\{0,1,2,3,4,5\}$. These cycles have lengths: $(0,2,4)$ is a three-cycle, while $(5,3)$ is a two-cycle. The orbits are, of course, disjoint. So we have decomposed σ into disjoint cycles.

Something interesting emerges here. Let τ be the permutation in S_6 represented by $(0,2,4)$ and ρ be the permutation represented by $(5,3)$. Then, as the hard-working graduate student can check, $\sigma = \tau \circ \rho$. This suggests that we can capture composition of permutations by juxtaposing a bunch of cycles. Suppose μ is a permutation on $\{0,1,2,3,4,5\}$ represented by $(0,1)(2,3)(4,5)$. We would like the represent $\sigma \circ \mu$ by

$$(0,2,4)(5,3)(0,1)(2,3)(4,5).$$

Observe that the listed cycles are no longer disjoint, so we should not think of this as a list of orbits. Rather, let us see what happens to the natural number 2 when we apply $\sigma \circ \mu$ to it. We know $\mu(2) = 3$ and $\sigma(3) = 5$ so that $\sigma \circ \mu(2) = 5$. Now consider the following

$$\begin{aligned}
\sigma \circ \mu(2) &= (0,2,4)(5,3)(0,1)(2,3)(4,5)2 \\
&= (0,2,4)(5,3)(0,1)(2,3)2 \text{ since } (4,5) \text{ fixes } 2 \\
&= (0,2,4)(5,3)(0,1)3 \\
&= (0,2,4)(5,3)3 \\
&= (0,2,4)5 \\
&= 5.
\end{aligned}$$

A more compact way to display the same information is

$$\begin{array}{ccccccccccc} & & (0,2,4) & & (5,3) & & (0,1) & & (2,3) & & (4,5) & \\ 5 & \longleftarrow & 5 & \longleftarrow & 3 & \longleftarrow & 3 & \longleftarrow & 2 & \longleftarrow & 2 \end{array}$$

In fact, we could extend this to display the whole effect of the composite permutation.

$$\begin{array}{ccccccccccc} & & (0,2,4) & & (5,3) & & (0,1) & & (2,3) & & (4,5) & \\ 1 & \longleftarrow & 1 & \longleftarrow & 1 & \longleftarrow & 0 & \longleftarrow & 0 & \longleftarrow & 0 \\ 2 & \longleftarrow & 0 & \longleftarrow & 0 & \longleftarrow & 1 & \longleftarrow & 1 & \longleftarrow & 1 \\ 5 & \longleftarrow & 5 & \longleftarrow & 3 & \longleftarrow & 3 & \longleftarrow & 2 & \longleftarrow & 2 \\ 4 & \longleftarrow & 2 & \longleftarrow & 2 & \longleftarrow & 2 & \longleftarrow & 3 & \longleftarrow & 3 \\ 3 & \longleftarrow & 3 & \longleftarrow & 5 & \longleftarrow & 5 & \longleftarrow & 5 & \longleftarrow & 4 \\ 0 & \longleftarrow & 4 & \longleftarrow & 4 & \longleftarrow & 4 & \longleftarrow & 4 & \longleftarrow & 5 \end{array}$$

Inspecting this array we find

$$\sigma \circ \mu = (0,1,2,5)(3,4),$$

which is the decomposition of $\sigma \circ \mu$ into a product of disjoint cycles.

Representing the inverse of a permutation is easy. For example, the inverse of $(0,1,2,5)(3,4)$ is just $(4,3)(5,2,1,0)$. We just write everything in reverse order.

We could pursue the same strategy of notation for representing the permutations and their products and inverses for any set X. For infinite sets, this becomes tricky, but when X is finite the strategy can be carried through without trouble.

Our first Fact is clear.

Fact. Every permutation on a finite set can be decomposed as a product of disjoint cycles. This decomposition is unique, up to rearranging the cycles. Also, any two disjoint cycles commute.

Permuations, even or odd

Let X be a set. A permutation σ of X is a **transposition** provided there are distinct elements $x, y \in X$ such that σ exchanges x and y and leave every other element of X fixed—that is $\sigma(x) = y$ and $\sigma(y) = x$ and $\sigma(w) = w$ for all $w \in X \setminus \{x, y\}$. In the notation above, this means $\sigma = (x, y)$. Evidently, σ is its own inverse: $\sigma \circ \sigma = \mathbf{1}_X$.

Fact. Every permutation on a finite set can decomposed as a product of transpositions.

Proof. The identity permutation is the product of the empty systems of transpositions. Since every permutation is a product of cycles, we need only prove that every cycle is a product of transpositions. Let $a_0, a_1, \ldots, a_{k-1}$ be distinct elements of our finite set. Just check

$$(a_0, a_1, \ldots, a_{k-1}) = (a_1, a_2)(a_0, a_{k_1}) \ldots (a_0, a_3)(a_0, a_2).$$

□

The decomposition of a permutation into transpositions is not unique. For example, $(0,1,2) = (1,2)(0,2) = (1,0)(1,2) = (0,2)(2,1)(0,1)(0,2)$. However, a shred of uniqueness remains.

Fact. Let $\sigma_0,\dots,\sigma_{k-1}$ and $\tau_0,\dots,\tau_{\ell-1}$ be sequences of transpositions such that

$$\sigma_0\circ\sigma_1\circ\cdots\circ\sigma_{k-1}=\tau_0\circ\tau_1\circ\cdots\circ\tau_{\ell-1}.$$

Then k and ℓ have the same parity, that is either both k and ℓ are even or both k and ℓ are odd.

Proof. Let X be the underlying set. Let us assume to the contrary that k is even and ℓ is odd. Since every transposition is its own inverse, we are led to

$$\tau_{\ell-1}\circ\cdots\circ\tau_0\circ\sigma_0\circ\sigma_1\circ\cdots\circ\sigma_{k-1}=\mathbf{1}_X.$$

So we see that the identity $\mathbf{1}_X$ can be written as the product of a sequence of transposition of length $k+\ell$, which is odd. So our proof will be completed by the following contention, since the identity permutation fixes every element of X.

Contention. Suppose σ is a permutation of X so that

$$\sigma=\tau_0\circ\tau_1\circ\cdots\circ\tau_{m-1}$$

where m is odd and each τ_j is a transposition. Then σ moves some element of X.

We assume, without loss of generality, that m be the smallest odd number so that σ is the product of a sequence of length m of transpositions. Let $\tau_{m-1}=(a,e)$ where a and e are distinct elements of X. (The hard working graduate students must figure out what to do when X has fewer than two elements.) We will actually prove that σ moves a. We could do the same for any element of X that is moved by any of the transpositions $\tau_0,\tau_1,\dots,\tau_{m-1}$. We achieve this by rewriting the factorization of σ in $m-1$ steps. We start by letting $\rho_{m-1}=\tau_{m-1}$. So our initial factorization is

$$\sigma=\tau_0\circ\dots\tau_k\circ\tau_{k+1}\circ\tau_{k+2}\circ\cdots\circ\tau_{m-2}\circ\rho_{m-1}.$$

It has the property that no transpostion to the *right* of ρ_{m-1} moves a. After a number of steps, we will have the factorization

$$\sigma=\tau_0\circ\dots\tau_k\circ\rho_{k+1}\circ\tau'_{k+2}\circ\cdots\circ\tau'_{m-2}\circ\tau'_{m-1},$$

where ρ_{k+1} and τ'_j for $k+1<j<m$ are transpositions and a is moved by ρ_{k+1} but fixed by all the transpositions to its right in the factorization.

The rewriting happens in this way. We will replace $\tau_k\circ\rho_{k+1}$ by $\rho_k\circ\tau'_{k+1}$ so that a is moved by ρ_k but fixed by τ'_{k+1}. First we observe that $\tau_k\neq\rho_{k+1}$ since that $\tau_k\circ\rho_{k+1}=\mathbf{1}_X$ and we could delete these two factors resulting in a factorization of σ of smaller odd length—a violation of the minimality of m. Let us say that $\rho_{k+1}=(a,b)$. Then there are only three alternatives for τ_k: it is disjoint from (a,b) or it moves b but not a or it moves a but not b. So $\tau_k\circ\rho_{k+1}$ has one the the following forms

$$(c,d)(a,b)$$
$$(b,c)(a,b)$$
$$(a,c)(a,b)$$

where all the letters in each line stand for distinct elements of X. It is easy to check that each of the equations below holds in **Sym** X.

$$(c,d)(a,b) = (a,b)(c,d)$$
$$(b,c)(a,b) = (a,c)(b,c)$$
$$(a,c)(a,b) = (a,b)(b,c)$$

So to obtain the next factorization of σ we simply replace the left side of the appropriate equation by its right side. For example, if $\tau_k = (b,c)$, then we take $\rho_k = (a,c)$ and $\tau'_{k+1} = (b,c)$. Here is what happens in detail:

$$\sigma = \tau_0 \circ \cdots \circ \ \tau_k \ \circ \rho_{k+1} \circ \tau'_{k+2} \circ \cdots \circ \tau'_{m-1}$$
$$\sigma = \tau_0 \circ \cdots \circ (b,c) \circ (a,b) \circ \tau'_{k+2} \circ \cdots \circ \tau'_{m-1}$$
$$\sigma = \tau_0 \circ \cdots \circ (a,c) \circ (b,c) \circ \tau'_{k+2} \circ \cdots \circ \tau'_{m-1}$$
$$\sigma = \tau_0 \circ \cdots \circ \ \rho_k \ \circ \tau'_{k+1} \circ \tau'_{k+2} \circ \cdots \circ \tau'_{m-1}$$

After $m-1$ rewrite steps of this kind we obtain

$$\sigma = \rho_0 \circ \tau'_1 \circ \cdots \circ \tau'_{m-1},$$

a factorization of σ into transpositions. But now observe that a is moved by ρ_0 but fixed by all the τ'_k's. Hence, σ moves a, as desired. □

We will call a permutation σ of a set X **even** if it can be decomposed as a product of a sequence transpositions, the sequence being of even length. We call σ **odd** if it can be decomposed as a product of a sequence of transpositions, the sequence being of odd length. If the number of elements of X moved by σ is finite, we see from the two facts above, that these are mutually exclusive and exhaustive alternatives. If the number of elements of X moved by σ is infinite, then σ cannot be written as the product of an finite sequence of transpositions, since any permutation that can be written as a product of a finite sequence of transpositions must fix all members of X that do not arise in the transpositions. Any permutation which moves infinitely many members of X cannot be either even or odd. For example, the permutation of $\mathbb{Z}$ such that $k \mapsto k+1$ for all $k \in \mathbb{Z}$ is of this kind.

In any case, the set $\mathrm{Alt}\, X$ of even permutations of X, contains $\mathbf{1}_X$ and is closed under composition and the formation of inverses. So

$$\mathbf{Alt}\, X := \langle \mathrm{Alt}\, X, \circ, ^{-1}, \mathbf{1}_X \rangle$$

is another example of a concrete group. It is called the **alternating group** on X. When $X = n = \{0,1,\ldots,n-1\}$ we adopt the notation $\mathbf{A}_n$.

1.1 Problem Set 12

Algebra Homework, Edition 12
Rings and modules one more time

Problem 37.
Let **F** be the subring of the field of complex numbers consisting of those numbers of the form $a+ib$ where a and b are rational. Let $\mathbb{G}$ be the subring of the field of complex numbers consisting of those numbers of the form $m+ni$ where m and n are integers.

(a) Describe all the units of $\mathbb{G}$.

(b) Prove that **F** is (isomorphic to) the field of fractions of $\mathbb{G}$.

(c) Prove that $\mathbb{G}$ is a principal ideal domain.

[Hint: In this problem it is helpful to consider the function that sends each complex number z to $z\bar{z} = |z|^2$.]

Problem 38.
Let **R** and **S** be commutative Noetherian rings. Prove that $\mathbf{R} \times \mathbf{S}$ is also Noetherian.

Problem 39.
Let **F** and **M** be modules over the same ring and let **F** be a free module. Let $h : \mathbf{M} \twoheadrightarrow \mathbf{F}$ be a homomorphism from **M** onto **F**. Prove each of the following.

(a) There is an embedding $g : \mathbf{F} \rightarrowtail \mathbf{M}$ of **F** into **M** such that $h \circ g = \mathrm{id}_F$. (Here id_F denotes the identity map of the set F.)

(b) $\mathbf{M} = \ker h \oplus \mathbf{F}'$, where $\mathbf{F}'$ is the image of **F** with respect to g.

Problem 40.
Let **M** and **N** by finitely generated modules over the same principal ideal domain. Prove that if $\mathbf{M} \times \mathbf{N} \times \mathbf{M} \cong \mathbf{N} \times \mathbf{M} \times \mathbf{N}$, then $\mathbf{M} \cong \mathbf{N}$.

Problem 41.
Give an example of two **dissimilar** matrices A and B with real entries that have all the following properties:

(a) A and B have the same minimal polynomial,

(b) A and B have the same characteristic polynomial, and

(c) The common minimal polynomial has no real roots.

LECTURE 2

THE THEORY OF ABSTRACT GROUPS: GETTING OFF THE GROUND

2.1 DEFINING THE CLASS OF GROUPS BY EQUATIONS

In a concrete group the operations are easy to understand: composition of functions, the formation of inverse functions, and the identity function as distinguished function. In an abstract group we lose this tight grip on the basic operations. Nevertheless, since an isomorphism ties each abstract group to a concrete group many properties of the basic operations of the concrete group must also hold in the abstract case. In the midst of the 19^{th} century Arthur Cayley made the breakthrough to the class of abstract groups with the following theorem.

Cayley's Equational Axiomatization Theorem. *Let* $\mathbf{G} = \langle G, \cdot, ^{-1}, 1\rangle$ *be an algebra of the signature of groups.* $\mathbf{G}$ *is a group if and only if the following equations are true in* $\mathbf{G}$*:*

$$x^{-1} \cdot x = 1 \qquad x \cdot (y \cdot z) = (x \cdot y) \cdot z \qquad 1 \cdot x = x$$
$$x \cdot x^{-1} = 1 \qquad\qquad\qquad\qquad x \cdot 1 = x$$

Proof. Suppose first that $\mathbf{G}$ is a group. This means that it is isomorphic to a concrete group. The five equations above hold about composition of functions, the formation of inverse functions, and the identity function. So they must also hold in $\mathbf{G}$.

Now let us suppose that $\mathbf{G}$ is an algebra in which these five equations happen to be true. To prove that $\mathbf{G}$ is a group we will devise a concrete group and show that $\mathbf{G}$ can be embedded into it. Having nothing else at hand we will take $\mathbf{Sym}\, G$ as our concrete group. Our embedding $\Phi : \mathbf{G} \to \mathbf{Sym}\, G$ can be given explicitily. We see that for each $g \in G$ that $\Phi(g)$ is supposed to be a permutation of G. Since writing things like $\Phi(g)(h)$ is clumsy, we use Φ_g in place of $\Phi(g)$. For each $g \in G$ we define $\Phi_g : G \to G$ in the following way. For any $h \in G$, put

$$\Phi_g(h) = g \cdot h.$$

We need to verify that each Φ_g is indeed a permutation of G, that the map Φ is one-to-one, and that Φ is a homomorphism. The five equations above allow us to succeed.

Contention. Φ_g is a permutation of G, for each $g \in G$.

It is evident from the definition of Φ_g that it is a function from G to G. Since the permutations of G are just the invertible functions from G to G, we will prove here that Φ_g and $\Phi_{g^{-1}}$ are inverses of each other. We will need this anyway to see that Φ is a homomorphism. That Φ_g and $\Phi_{g^{-1}}$ are inverses is equivalent to the assertion that for all $h, k \in G$

$$\Phi_g(h) = k \text{ if and only if } \Phi_{g^{-1}}(k) = h.$$

Here is the argument:

$$\begin{aligned}
\Phi_g(h) = k &\Rightarrow g \cdot h = k \\
&\Rightarrow g^{-1} \cdot (g \cdot h) = g^{-1} \cdot k \\
&\Rightarrow (g^{-1} \cdot g) \cdot h = \Phi_{g^{-1}}(k) \quad \text{by associativity} \\
&\Rightarrow 1 \cdot h = \Phi_{g^{-1}}(k) \quad \text{by } x^{-1} \cdot x = 1 \\
&\Rightarrow h = \Phi_{g^{-1}}(k) \quad \text{by } 1 \cdot x = x \\
&\Rightarrow h = g^{-1} \cdot k \\
&\Rightarrow g \cdot h = g \cdot (g^{-1} \cdot k) \\
&\Rightarrow \Phi_g(h) = (g \cdot g^{-1}) \cdot k \quad \text{by associativity} \\
&\Rightarrow \Phi_g(h) = 1 \cdot k \quad \text{by } x \cdot x^{-1} = 1 \\
&\Rightarrow \Phi_g(h) = k \quad \text{by } 1 \cdot x = x
\end{aligned}$$

This argument is a bit pedantic but it does show where four of our five equations come into play.

Contention. Φ is one-to-one.

Let us suppose that $\Phi_g = \Phi_h$. Then in particular, $\Phi_g(1) = \Phi_h(1)$. So we find $g = g \cdot 1 = \Phi_g(1) = \Phi_h(1) = h \cdot 1 = h$. Here we have appealled to our fifth equation $x \cdot 1 = x$, to conclude that Φ is one-to-one.

Here is our last contention:

Contention. Φ is a homomorphism.

So we must establish the following:

$$\begin{aligned}
\Phi_{g \cdot h} &= \Phi_g \circ \Phi_h \\
\Phi_{g^{-1}} &= \left(\Phi_g\right)^{-1} \\
\Phi_1 &= \mathbf{1}_G
\end{aligned}$$

Let k be any element of G. Here is the demonstration of the first piece.

$$\Phi_{g \cdot h}(k) = (g \cdot h) \cdot k = g \cdot (h \cdot k) = g \cdot \Phi_h(k) = \Phi_g\left(|\Phi_h(k)\right) = \Phi_g \circ \Phi_h)(k).$$

We already established the second piece on the way to showing that Φ_g is a permutation. The last is easiest of all:

$$\Phi_1(k) = 1 \cdot k = k, \text{ for all } k \in G.$$

So Φ_1 is the identity function, as desired.

Notice that what we showed is that Φ is a one-to-one homomorphism from **G** into **Sym** G. We did not claim that Φ is onto Sym G. Rather that concrete group isomorphic to **G** is a subalgebra of **Sym** G. □

Corollary 2.1.1. *Every subalgebra of a group is again a group. Every homomorphic image of a group is again a group. The direct product of any system of groups is again a group.*

These facts all follow since the truth equations is preserved under all these constructions.

In our proof of Cayley's Theorem each of the five equations came into play, suggesting that perhaps all of them are needed. On the other hand, may be a slicker proof would avoid the use of some of those equations. Actually, two of the equations can be omitted, as indicated in one of the problem sets.

While starting with concrete groups and then forming the class of all abstract groups and then seeing that this latter class can be described by a set of simple equations reflects the actual historical development, most expositors of group theory have chosen to start with the set of five equations (or something like them) as a way to define the notion of a group. From this perspective Cayley's Theorem is called the Cayley Representation Theorem because it shows that every group (i.e. an algebra satisfying those equations) is isomorphic to (can be represented as) a group of permutations.

A large number of authors conceive a group as an algebra $\langle G, \cdot \rangle$ with the follow properties:

(a) $\cdot$ is an associative operation on G.

(b) There is an element $1 \in G$ such that $1 \cdot a = a \cdot 1 = a$ for all $a \in G$.

(c) For every $a \in G$ there is $b \in G$ so that $a \cdot b = b \cdot a = 1$.

They then go on to show that only one element of a group can play the role of 1 and that every element of a group has exactly one inverse. This approach has the advantage that the algebras involved have only one operation. But it carries with it some annoyance as well. To see why, here is a more formalized presentation of this axiomatic approach.

(a) $\forall x, y, z \left[x \cdot (y \cdot z) \approx (x \cdot y) \cdot z \right]$.

(b) $\exists u \forall x \left[u \cdot x \approx x \cdot u \approx x \right]$.

(c) $\forall x \exists y \forall z \left[(x \cdot y) \cdot z \approx z \approx z \cdot (x \cdot y) \& (y \cdot x) \cdot z \approx z \approx z \cdot (y \cdot x) \right]$.

By formalizing these statements we can see easily that they have more involved logical forms than equations. While it takes more work, one can prove that truth sentences of these forms are preserved under the formation of homomorphic images and direct products. The same does not apply to subalgebras. For example $\langle \mathbb{R}^+, \cdot \rangle$ the positive reals under multiplication, is a group in this sense. It is easy to see that the positive reals strictly less than 1 constitute a subalgebra. But this subalgebra has neither a multiplicative unit (1 is just not in there) nor does any element have a multiplicative inverse. So this sublagebra is not a group. This annoyance in minor, of course. When group theory is developed from this starting point it very soon (say within five pages) gets to the point where one as named the multlipicative unit and begins using some notation for the formation of inverses. From that point on the development is carried forward just as if these operations were given at the very beginning.

One might wonder if group theory could be developed using a different choice of basic operations. Indeed, it can. Since $x \cdot x^{-1} = 1$ we see that the distinguished element can be defined by an equation. Indeed, we could regard 1 as an abbreviation for $x \cdot x^{-1}$ provided we add the equation $x \cdot x^{-1} = y \cdot y^{-1}$ to our list. The we could dispense with 1 and make do with just $\cdot$ and $^{-1}$. A more radical step is to use the operation

$$x \mid y := x^{-1} \cdot y.$$

It takes some doing, but the hard working graduate students filled up with interest about this point should be able to write down a short list of equations just in terms of the one two-place operation symbol $\mid$ which entirely captures the notion of group. One way to start this project is to try to devise a term in x and y and $\mid$ to recapture $\cdot$. A peculiar feature of this approach to group theory is the discovery of a single (rather long) equation in $\mid$ that defines the class of all groups.

2.2 Homomorphisms and their kernels—and an extraordinary property of subgroups

Just as they did in the case of rings, homomorphisms and their kernels will play a central role in the development of group theory.

Fact. Let **G** and **H** be groups and let $f : G \to H$. Then f is a homomorphism if and only if $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in G$.

Of course the direction from left to right is immediate. For the converse, we need to show that f preserves the other operations.

$$f(1) \cdot 1 = f(1) = f(1 \cdot 1) = f(1) \cdot f(1)$$

Holds because both **G** and **H** are groups and because f preserves the product. So we see $f(1) \cdot 1 = f(1) \cdot f(1)$ holds in **H**. Because **H** is a group we can cancel $f(1)$ from both sides, leaving $1 = f(1)$, which we need if f is going to be a homomorphism. So what about inverses? Well

$$1 = f(1) = f(a \cdot a^{-1}) = f(a) \cdot f(a^{-1}).$$

So $1 = f(a) \cdot f(a^{-1})$ holds in **H**. Since **H** is a group, we can multiply both sides of this equation by $\big(f(a)\big)^{-1}$ to arrive at $\big(f(a)\big)^1 = f(a^{-1})$, as desired. So the Fact is established.

In using this fact to prove that some map is a homomorphism it is important to prove in advance that both **G** and **H** are groups.

Now suppose **G** is a group and h is a homomorphism from **G** onto **H**. Let

$$\theta := \{(a, b) \mid a, b \in G \text{ and } h(a) = h(b)\}.$$

So θ is the functional kernel of h. It was convenient in the theory of rings to replace θ with the congruence class of 0. We can do something along these lines in groups as well. Observe for all $a, b \in G$ we have

$$a \,\theta\, b \Leftrightarrow (b^{-1} \cdot a) \,\theta\, (b^{-1} \cdot b) = 1.$$

Another way to formulate this is $b \in a/\theta \Leftrightarrow (b^{-1} \cdot a) \in 1/\theta$. The upshot is that the congruence class $1/\theta$ completely determines the whole congruence relation. (Just as in ring theory congruence class of 0—an ideal—completely determined the congruence relation.) What properties of $1/\theta$ make it so special?

- $1/\theta$ is a subgroup of **G**.
- If $a \in 1/\theta$ and $b \in G$, then $(b^{-1} \cdot a \cdot b) \in 1/\theta$.

The hard-working graduate student will have no trouble verifying these points. For instance, the last follows since if $h(a) = 1$, then $h(b^{-1} \cdot a \cdot b) = h(b)^{-1} \cdot h(a) \cdot h(b) = h(b)^{-1} \cdot 1 \cdot h(b) = 1$.

A subgroup **N** of a group **G** is said to be **normal** provided for all a and b, if $a \in N$ and $b \in G$, then $(b^{-1} \cdot a \cdot b) \in N$. We use $\mathbf{N} \lhd \mathbf{G}$ to symbolize that **N** is a normal subgroup of **G**.

Theorem on Congruences and Normal Subgroups. *Let* **G** *be a group. The following are equivalent.*

(a) **N** *is a normal subgroup of* **G**.

(b) $N = 1/\theta$ *for some congruence* θ *of* **G**.

(c) $N = \{a \mid a \in G$ *and* $h(a) = 1\}$ *for some homomorphism from* **B**.

(d) $\{(a,b) \mid a, b \in G$ *and* $(b^{-1} \cdot a) \in N\}$ *is a congruence of* **G**.

Proof. Based on the discussion above and on our general understanding of the connection between homomorphisms and congruence relations, the only implication that really calls for further attention is $(a) \Rightarrow (b)$.

To establish this implication, let **N** be a normal subgroup of **G** and put

$$\theta := \{(a,b) mid a, b \in G \text{ and } (b^{-1} \cdot a) \in N\}.$$

Contention. θ is an equivalence relation on G.

For every $a \in G$ we see that $a^{-1} \cdot a = 1 \in N$ since **N** is a subgroup of **G**. This means that $(a,a) \in \theta$, so θ is reflexive.

To see symmetry, suppose $(a,b) \in \theta$. This means $(b^{-1} \cdot a) \in N$. Since **N** is a subgroup of **G** we know that $(b^{-1} \cdot a)^{-1} \in N$. But because **N** is a group we know that $a^{-1} \cdot b = (b^{-1} \cdot a)^{-1}$. Hence $(a^{-1} \cdot b) \in N$. But this entails $(b,a) \in \theta$ and the symmetry of θ is proved.

Finally, for transitivity suppose $(a,b), (b,c) \in \theta$. This means

$$(b^{-1} \cdot a) \in N$$
$$(c^{-1} \cdot b) \in N.$$

Since **N** is a subgroup of **G** we get

$$(c^{-1} \cdot b) \cdot (b^{-1} \cdot a) \in N.$$

Just a bit of fiddling gives $(c^{-1} \cdot a) \in N$. This entails $(a,c) \in \theta$ and establishes the transitivity ofθ, concluding the proof that θ is an equivalence relation on G. It is useful to point out that only the fact that **N** is a *subgroup* of **G** was used here.

Contention. θ is a congruence relation of **G**.

We need to establish two things:

$$a\,\theta\,b \text{ and } c\,\theta\,d \Rightarrow a\cdot c\;b\cdot d$$
$$a\,\theta\,b \Rightarrow a^{-1}\,\theta\,b^{-1}.$$

Using the definition of θ, these become

$$(b^{-1}\cdot a)\in N \text{ and } (d^{-1}\cdot c)\in N \Rightarrow \left((b\cdot d)^{-1}\cdot(a\cdot c)\right)\in N$$
$$(b^{-1}\cdot a)\in N \Rightarrow \left((b^{-1})^{-1}\cdot a^{-1}\right)\in N.$$

After some minor fiddling this gives

$$(b^{-1}\cdot a)\in N \text{ and } (d^{-1}\cdot c)\in N \Rightarrow \left(d^{-1}\cdot b^{-1}\cdot a\cdot c\right)\in N$$
$$(b^{-1}\cdot a)\in N \Rightarrow \left(b\cdot a^{-1}\right)\in N.$$

Let's tackle the top implication. Suppose $(b^{-1}\cdot a),(d^{-1}\cdot c)\in N$. Because **N** is a normal subgroup we see that $(d^{-1}\cdot b^{-1}\cdot a\cdot d)\in N$. Because **N** is a subgroup $d^{-1}\cdot b^{-1}\cdot a\cdot c=(d^{-1}\cdot b^{-1}\cdot a\cdot d\cdot d^{-1}\cdot c)\in N$ as desired. For the remaining implication, suppose $(b^{-1}\cdot a)\in N$. Since **N** is a subgroup we can apply the inverse to get $(a^{-1}\cdot b)\in N$. Now invoking the normality of **N** we see $b\cdot a^{-1}=b\cdot(a^{-1}\cdot b)\cdot b^{-1}\in N$ as desired. □

The import of this theorem is that we can use normal subgroups and congruences interchangeably, just as we used ideals and congruences interchangeably in ring theory. When h is a homomorphism from the group **G** we will call $\{a \mid a\in G \text{ and } h(a)=1\}$ the **kernel** of h. Of course, it is a normal subgroup and in fact, the normal subgroups of **G** are exactly the kernels of homomorphisms from **G**.

Let **N** be a normal subgroup of the group **G**. What are the congruence classes associated with **N**? We know that N itself is the congruence class containing the element 1. What about the others? Let $a\in G$. The congruence class containing a is evidently

$$\{b \mid b\in G \text{ and } (a^{-1}\cdot b)\in N\}.$$

Let $aN=\{a\cdot c \mid c\in N\}$. Then it is clear that $(a^{-1}\cdot b)\in N$ if and only if $b\in aN$. This means that the congruence class containing a is

$$\{b \mid b\in G \text{ and } (a^{-1}\cdot b)\in N\}=\{b \mid b\in G \text{ and } b\in aN\}=aN.$$

We observe that this little line of reasoning remains true even if **N** is only a subgroup of **G** (of course we should only say "equivalence class" in this case). Sets of the form aN are called **(left) cosets** of **N**. Right cosets I leave to your imagination. Here is the remarkable thing discovered by Lagrange.

Lagrange's Theorem. *Let* **H** *be a subgroup of the group* **G**. *All the cosets of* **H** *have the same cardinality. In particular,* $|H|$ *divides* $|G|$. *Moreover, if* **K** *is a homomorphic image of* **G**, *then* $|K|$ *also divides* $|G|$.

Proof. Let a be an arbitrary element of G. We need only exhibit a one-to-one correspondence from H onto aH. Just define $\Phi: H\to aH$ via

$$\Phi(b)=ab \text{ for all } b\in H.$$

This map is onto aH by the very definition of aH. It is one-to-one since $a \cdot b = a \cdot c \Rightarrow b = c$, since the cancellation law works in all groups. The first divisibility statement works because we know that G is partitioned into the cosets of H—that is, G is the disjoint union of some number of sets all of size $|H|$. The second divisibility statement follows since if **H** is the kernel of a homomorphism from **G** onto **K**, then by the Homomorphism Theorem $|K|$ must be the number of distinct cosets of **H**. □

Lagrange's Theorem (well this is only one of his theorems...) holds for all groups, finite or infinite, but it was the first key tool for dealing with finite groups. So a group of size 21 cannot have a subgroup of size 6 and the sizes of its homomorphic images can be found only among $1, 3, 7$, and 21.

Let **G** be a group and let **H** be a subgroup. We put

$$[\mathbf{G} : \mathbf{H}] = |\{aH \mid a \in G\}|.$$

That is $[\mathbf{G} : \mathbf{H}]$ is the number of left cosets of **H** in **G**. It is called the **index** of **H** in **G**. Another way to frame Lagrange's Theorem is to assert, for **H** a subgroup of **G** and $h : \mathbf{G} \twoheadrightarrow \mathbf{L}$, that

$$|G| = [\mathbf{G} : \mathbf{H}]|H| = |L||\ker h|.$$

It is evident that everywhere in the above discussion, if we were to replace *left* coset by *right* coset the result would be entirely similar. In particular, The number of left cosets of **H** is the same as the number of right cosets of **H**. This does not mean that every left coset is a right coset (and vice versa). The demonstration of the fact below is left to the hard-working graduate students.

Fact. Let **H** be a subgroup of the group **G**. Then **H** is a normal subgroup of **G** if and only if every left coset of **H** is **G** is a right coset of **H** in **G**.

2.3 Problems Set 13

Algebra Homework, Edition 13
Some little problems about groups

Problem 42.
Derive a list of equations that follow from the equations axiomatizing the theory of groups. This is rather open ended, but see if you can get a handful of useful looking equations.

Problem 43.
The five equations used to axiomatize groups are not all needed. Find a simpler set of equations that will serve.

Problem 44.
Prove that the additive group of all polynomials in x with integer coefficients is isomorphic to the multiplicative group of all postive rational numbers.

Problem 45.
Let **A** and **B** be groups and let $f : A \to B$. Prove that f is a homomoprhism if and only if $f(aa') = f(a)f(a')$ for all $a, a' \in A$.

Problem 46.
Let **A**, **B**, and **C** be groups. Let h be a homomorphism from **A** onto **B** and let g be a homomorphism from **A** onto **C** such the $\ker h = \ker g$.

Prove that there is an isomorphism f from **B** onto **C**.

LECTURE 3

ISOMORPHISM THEOREMS: THE GROUP VERSIONS

Last semester we learned a collection of theorems in the general context. Seeing in the previous lecture that, in group theory, we can replace congruences by normal subgroups, in this lecture we simply present the corresponding specializations without further proof. However we conclude with a more involved theorem for groups that is due to Hans Zassenhaus. We will use this theorem later.

The Homomorphism Theorem (Group Version). *Let* $\mathbf{A}$ *be a group, let* $f : \mathbf{A} \twoheadrightarrow \mathbf{B}$ *be a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{B}$*, and let* $\mathbf{N}$ *be a normal subgroup of* $\mathbf{A}$*. All of the following hold.*

(a) *The kernel of* f *is a normal subgroup of* A*.*

(b) $\mathbf{A}/\mathbf{N}$ *is group.*

(c) *The map* η *that assigns to each* $a \in A$ *the left coset* aN *is a homomorphism from* $\mathbf{A}$ *onto* $\mathbf{A}/\mathbf{N}$ *and its kernel is* $\mathbf{N}$*.*

(d) *If* $\mathbf{N}$ *is the kernel of* f*, then there is an isomorphism* g *from* $\mathbf{A}/\mathbf{N}$ *to* $\mathbf{B}$ *such that* $f = g \circ \eta$*.*

The Second Isomorphism Theorem (Group Version). *Let* $\mathbf{A}$ *be a group, let* $\mathbf{N}$ *be a normal subgroup of* $\mathbf{A}$*, and let* $\mathbf{B}$ *be a subgroup of* $\mathbf{A}$*. Then each of the following hold.*

(a) $\mathbf{N} \cap \mathbf{B}$ *is a normal subgroup of* $\mathbf{B}$*.*

(b) $\mathbf{NB}$ *is a subgroup of* $\mathbf{A}$ *and* $\mathbf{N}$ *is a normal subgroup of* $\mathbf{NB}$*.*

(c) $\mathbf{NB}/\mathbf{N} \cong \mathbf{B}/\mathbf{N} \cap \mathbf{B}$*.*

The Third Isomorphism Theorem (Group Version). *Let* **A** *be a group and let* **N** *and* **K** *be normal subgroups of* **A** *with* $N \subseteq K$. *Then*

(a) $\mathbf{K}/\mathbf{N}$ *is a normal subgroup of* $\mathbf{A}/\mathbf{N}$, *and*

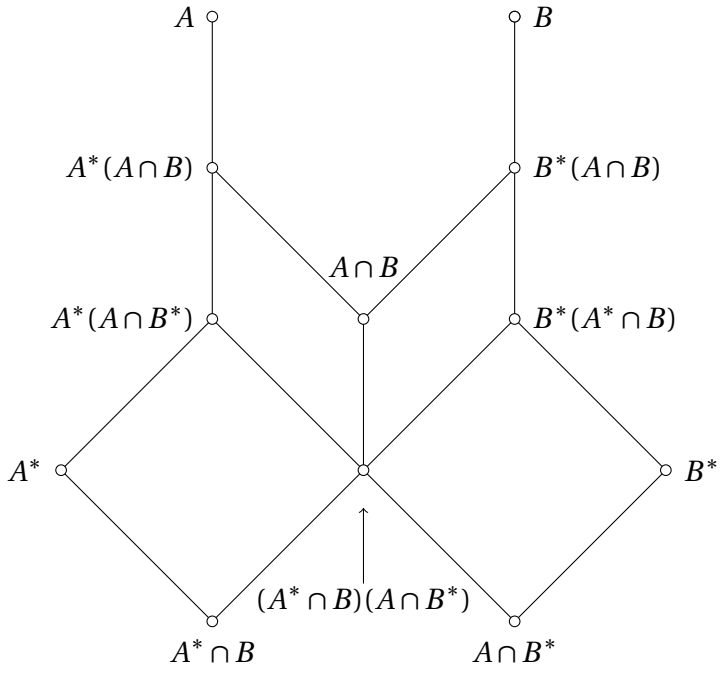
(b) $(\mathbf{A}/\mathbf{N})/(\mathbf{K}/\mathbf{N}) \cong \mathbf{A}/\mathbf{K}$.

The Correspondence Theorem (Group Version). *Let* **A** *be a group and let* **N** *be a normal subgroup of* **A**. *Let* $P = \{\mathbf{K} \mid \mathbf{K} \lhd \mathbf{A}$ *and* $\mathbf{N} \subseteq \mathbf{K}\}$. *Then the map from* P *to set of normal subgroups of* $\mathbf{A}/\mathbf{N}$ *that sends each* $\mathbf{K} \in P$ *to* $\mathbf{K}/\mathbf{N}$ *is an isomorphism between the ordered set* $\langle P, \subseteq \rangle$ *and the set of normal subgroups of* $\mathbf{A}/\mathbf{N}$ *ordered by set inclusion.*

Just as the corresponding special cases for rings were handy in the development of the theory of rings, these special cases for groups will play a similar role. Here is an interesting conclusion that comes from putting Lagrange's Theorem and the Second Isomorphism Theorem together.

Corollary 3.0.1. *Let* **A** *be a finite group and let* **B** *be a subgroup of* **A** *and let* **N** *be a normal subgroup of* **A**. *Then* $|NB||N \cap B| = |N||B|$. *In particular if* **N** *and* **B** *have only the identity element in common, that* $|NB| = |N||B|$.

Zassenhuas's Butterfly Lemma is a more involved statement. The name is suggested by the following picture, which displays part of the lattice of subgroups of some group **G**.



The Butterfly of Hans Zassenhaus

The Butterfly Lemma of Hans Zasssenhaus. *Let* $\mathbf{G}$ *be a group with subgroups* $\mathbf{A},\mathbf{A}^*,\mathbf{B}$ *and* $\mathbf{B}^*$*, where* $\mathbf{A}^*$ *is a normal subgroup of* $\mathbf{A}$ *and* $\mathbf{B}^*$ *is a normal subgroup of* $\mathbf{B}$*. Then all of the following hold.*

(a) $\mathbf{A}^*(\mathbf{A}\cap\mathbf{B}^*)\lhd\mathbf{A}^*(\mathbf{A}\cap\mathbf{B})$.

(b) $\mathbf{B}^*(\mathbf{A}^*\cap\mathbf{B})\lhd\mathbf{B}^*(\mathbf{A}\cap\mathbf{B})$.

(c) $\mathbf{A}^*(\mathbf{A}\cap\mathbf{B})/\mathbf{A}^*(\mathbf{A}\cap\mathbf{B}^*)\cong\mathbf{B}^*(\mathbf{A}\cap\mathbf{B})/\mathbf{B}^*(\mathbf{A}^*\cap\mathbf{B})$.

Proof. Because $\mathbf{B}^*$ is a normal subgroup of $\mathbf{B}$ it is easy to see that $\mathbf{A}\cap\mathbf{B}^*\lhd\mathbf{A}\cap\mathbf{B}$. Likewise, $\mathbf{A}^*\cap\mathbf{B}\lhd\mathbf{A}\cap\mathbf{B}$. So we can also conclude that $(\mathbf{A}^*\cap\mathbf{B})(\mathbf{A}\cap\mathbf{B}^*)\lhd\mathbf{A}\cap\mathbf{B}$. Here is why. First we know that the product of a normal subgroup and a subgroup is again a subgroup. (This point came up in the Second Isomorphism Theorem—hard working graduate students checked it then....) So we find that $(\mathbf{A}^*\cap\mathbf{B})(\mathbf{A}\cap\mathbf{B}^*)$ is a subgroup of $\mathbf{A}\cap\mathbf{B}$. To normality, pick $c\in A^*\cap B$, $d\in A\cap B^*$, and $a\in A\cap B$. Then notice

$$a(cd)a^{-1}=aca^{-1}ada^{-1}=(aca^{-1})(ada^{-1}).$$

Since $aca^{-1}\in A^*\cap B$ by normality of $\mathbf{A}^*\cap\mathbf{B}$ is $\mathbf{A}\cap\mathbf{B}$ and likewise $ada^{-1}\in A\cap B^*$, we see that $a(cd)a^{-1}\in(\mathbf{A}^*\cap\mathbf{B})(\mathbf{A}\cap\mathbf{B}^*)$. Let $\mathbf{D}$ denote $\mathbf{A}^*\cap\mathbf{B})(\mathbf{A}\cap\mathbf{B}^*$. So we have $\mathbf{D}\lhd\mathbf{A}\cap\mathbf{B}$.

Now define the map $f:A^*(A\cap B)\to(A\cap B)/D$ in the following way. Let $a\in A^*$ and $c\in A\cap B$. Put $f(ac)=cD$. First we need to see that we can get away with this. Suppose $a_o\in A^*$ and $c_o\in A\cap B$ so that $ac=a_oc_o$. We need $cD=c_oD$ or what is the same $c_oc^{-1}\in D$. We certainly get $a_o^{-1}a=c_0c^{-1}$. The left side is in A^* and the right in $A\cap B$. Since they are equal, we see that $c_0c^{-1}\in A^*\cap A\cap B=A^*\cap B\subseteq(A^*\cap B)(A\cap B^*)=D$. So our definition of the map f is sound.

It is evident from the definition that f maps onto $(A\cap B)/D$. We aim to show that f is a homomorphism with kernel $\mathbf{A}^*(A\cap B^*)$. Then we can appeal to the Homomorphism Theorem to obtain an isomorphism. Reversing the roles of A^* and B^* we can obtain a second isomorphism. Composing on with the inverse of the other gets us the isomorphism we desired in the statement of the lemma.

We need to see that f preserves products. Let $a,a_o\in A^*$ and $c,c_o\in A\cap B$. Then

$$f((ac)(a_oc_o))=f(aca_oc^{-1}cc_o)=f(aa'_occ_o)=cc_oD=cDc_oD=f(ac)f(a_oc_o).$$

Observe the appeal to normality of $\mathbf{A}^*$ in $\mathbf{A}$. So we find that f is a homomorphism.

Last, we need to understand the kernel of f.

$$ac\in\ker f\Leftrightarrow f(ac)=1D\Leftrightarrow cD=D\Leftrightarrow c\in D\Leftrightarrow c=de\text{ for some }d\in A^*\cap B\text{ and some }e\in A\cap B^*.$$

This means

$$ac\in\ker f\Leftrightarrow ac=(ad)e\text{ for some }d\in A^*\cap B\text{ and some }e\in A\cap B^*\Leftrightarrow ac\in A^*(A\cap B^*).$$

Hence, we find $\ker f=\mathbf{A}^*(\mathbf{A}\cap\mathbf{B}^*)$ and our proof of the Butterfly Lemma is complete. □

3.1 Problem Set 14

Algebra Homework, Edition 14
Subgroups and homomorphic images of groups

Problem 47.
Prove that every group that has a proper subgroup of finite index must have a proper normal subgroup of finite index.

Problem 48.
Let $\mathbf{G}$ be a group. Prove that $\mathbf{G}$ cannot have four distinct proper normal subgroups $\mathbf{N}_0, \mathbf{N}_1, \mathbf{N}_2$, and $\mathbf{N}_3$ so that $\mathbf{N}_0 \leq \mathbf{N}_1 \leq \mathbf{N}_2 \leq \mathbf{G}$ and so that $N_1 N_3 = G$ and $N_2 \cap N_3 = N_0$.

Problem 49.
Let $\mathbf{H}$ and $\mathbf{K}$ be subgroups of the group $\mathbf{G}$ each of finite index in $\mathbf{G}$. Prove that $\mathbf{H} \cap \mathbf{K}$ is also a subgroup of finite index in $\mathbf{G}$.

Problem 50.
Prove that there is no group $\mathbf{G}$ such that $\mathbf{G}/\mathbf{Z}(\mathbf{G}) \cong \mathbb{Z}$, where $\mathbb{Z}$ denotes the group of integers under addition.

LECTURE 4

USEFUL FACTS ABOUT CYCLIC GROUPS

Let **G** be a group and let $X \subseteq G$. We use $\langle X \rangle$ to denote, ambiguously, both the subgroup of **G** generated by X and the underlying universe of that subgroup. So we can construe $\langle X \rangle$ has the intersection of all the subgroups of **G** that include X (the shrinkwrap viewpoint) or as the set of all elements of G that can be built from the elements of X by iteratively applying the basic operations of the group **G**.

The group **G** is **cyclic** provided there is $a \in G$ so that $G = \langle \{a\} \rangle$. That is, **G** is generated by some single element. To save notation, we write $\langle a \rangle$ for $\langle \{a\} \rangle$. Taking $a^{-n} := (a^{-1})^n$ for every natural number, we see that

$$\langle a \rangle = \{a^r \mid r \in \mathbb{Z}\}.$$

It is easy to see that in any group, the equation $x^r x^s = x^{r+s}$ must be true, where r and s are any integers. From this we get

Fact. Every cyclic group is Abelian.

For any group **G** and any $a \in G$, we let **order** of a, denoted by $o(a)$, be $|\langle a \rangle$.

Fact. The order of any element of a group is either the countably infinite or it is finite and not 0.

Fact. Let **G** be a group and $a \in G$. The element a has finite order n if and only if n is the smallest positive natural number such that $a^n = 1$ in **G**.

Proof. Let us first consider the case when $a = 1$. Then $\langle a \rangle = \{1\}$, a set with 1 element and $n = 1$ is also the least positive natural number so that $1^n = 1$. So in the remainder of this proof we consider the case when $a \neq 1$.

First suppose that a has finite order n. Then $\{a^r \mid r \in \mathbb{Z}\}$ is finite. So pick integers $k < \ell$ so that $a^k = a^\ell$. It follows that $a^{\ell-k} = 1$ and $\ell - k > 0$. Pick m to be the least positive natural number so that $a^m = 1$. So the elements $1 = a^0, a^1, \ldots, a^{m-1}$ must all be distinct. This set evidently contains 1 and it is closed under inverses since $a^k a^{m-k} = a^{k+m-k} = a^m = 1$, for all $k < m$. It is also close

under products since for $k, \ell < m$ we have $a^k a^\ell = a^{k+\ell} = a^{mq+r} = (a^m)^q a^r = 1^q a^r = a^r$, where q and r are the unique integers such that

$$k + \ell = mq + r \qquad \text{where } 0 \leq r < m.$$

So $\langle a \rangle = \{1, a, a^2, \ldots, a^{m-1}\}$. In this way we see that $n = m$, as desired.

For the converse, suppose that n is the least positive integer so that $a^n = 1$. Then we have just shown that $\langle a \rangle = \{1, a, a^2, \ldots, a^{n-1}\}$ and that the elements listed are distinct. So n is the order of a. □

The proof above carries a bit more information.

Fact. A finite cyclic group of order n is isomorphic to the group $\mathbb{Z}_n = \langle \{0, 1, \ldots, n-1\}, +_n, -_n.0 \rangle$ where the operations work modulo n.

Proof. Let $\langle a \rangle$ be the cyclic group of order n. The elements of this group and $a^0, a^1, a^2, \ldots, a^{n-1}$. As shown above, the operations work like this for all natural numbers $k, \ell < n$

$$\begin{aligned} (a^k)^{-1} &= a^{n-k} \\ a^k a^\ell &= a^r \qquad \text{where } 0 \leq r < n \text{ and } k + \ell \equiv r \pmod{n}. \end{aligned}$$

Now just observe that the "logarithm" that sends $a^k \mapsto k$ for all natural numbers $k < n$ is an isomorphism from $\langle a \rangle$ onto $\mathbb{Z}_n$. □

The same sort of logarithm function applies to infinite cyclic groups, giving the following

Fact. Every infinite cyclic group is isomorphic to the additive group of integers, that is to $\mathbb{Z} := \langle \mathbb{Z}, +, -, 0 \rangle$.

Fact. Let **G** be a group and let $a \in G$ have finite order n. If $a^m = 1$ in **G** then $n \mid m$.

Proof. Let q and r be the unique integers such that

$$m = nq + r \qquad \text{where } 0 \leq r < n.$$

Then $1 = a^m = a^{nq+r} = (a^n)^q a^r = 1^q a^r = a^r$. Since n is the order of a and $0 \leq r < n$, we must have $r = 0$. Thus $m = nq$ and $n \mid m$. □

Fact. Every subgroup of a cyclic group is cyclic.

Proof. Let **H** be a subgroup of the cyclic group **G** and let a be an element of G which generates **G**. As the trivial group is cyclic, we will consider the remain case that **H** is nontrivial. Let k be the least positive natural number so that $a^k \in H$. We see that k must be strictly smaller than $o(a)$. Our contention is that a^k generates **H**. So let a^ℓ be an arbitrary element of H. Let q and r be the unique integers so that

$$\ell = kq + r \qquad \text{where } 0 \leq r < k.$$

Then $a^\ell = (a^k)^q a^r$. Now since $a^k \in H$, then so is $(a^k)^{-q}$. But $a^\ell \in H$. Hence $a^r = a^\ell (a^k)^{-q} \in H$. Since $0 \leq r < k$, we see that $r = 0$ by the minimality of k. Hence $a^\ell = (a^k)^q$ and so a^ℓ is in the subgroup generated by a^k. Since a^ℓ was an arbitrary element of H, we see that **H** is generated by a^k and therefore that **H** is cyclic. □

The next fact provides a remarkable property that cyclic groups possess that is not common even among Abelian groups.

Fact. Let **G** be a cyclic group of finite order n and let k be a natural number so that $k \mid n$. Then **G** has exactly one subgroup of order k. Moreover, **G** has no other subgroups.

Proof. Let a generate **G** and let m be the natural number so that $km = n$. Then the order of a^m is the least ℓ so that $a^{m\ell} = 1$. Since $mk = n$, we see that $\ell \leq k$. But also $mk = n \leq m\ell$. Hence $k \leq \ell$. Thus $k = \ell$ is the order of a^m. This means that **G** has at least one subgroup of order k, namely $\langle a^m \rangle$. Since every subgroup of **G** is cyclic, let us suppose that a^j generates a subgroup of order k. That is, a^j has order k. Pick the integers q and r so that

$$j = mq + r \qquad \text{where } 0 \leq r < m.$$

Now we know that k is the least positive integer so that $a^{jk} = 1$ So we see

$$1 = a^{jk} = a^{mkq+rk} = (a^{mk})^q a^{rk} = (a^n)^q a^{rk} = 1^q a^{rk} = a^{rk}.$$

But since $0 \leq r < m$ we have $0 \leq rk < mk = n$. Since n is the order of a, we see that $r = 0$. But then $j = mq$. Hence $a^j = (a^m)^q$. This means that $a^j \in \langle a^m \rangle$. Hence $\langle a^j \rangle \subseteq \langle a^m \rangle$. But both of these sets are finite and have the same cardinality. So the must be equal, as desired.

That **G** has no other subgroups is immediate by Lagrange. □

The next couple of facts deal with Abelian groups and will help us distinguish which Abelian groups are actually cyclic.

Fact. Let **G** be an Abelian group and let $a, b \in G$. If the orders of a and b are finite and relatively prime, then $o(ab) = o(a)o(b)$.

Proof. Suppose $(ab)^k = 1$. Then $a^k = b^{-k} \in \langle a \rangle \cap \langle b \rangle$. So according to Lagrange, the order of a^k must divide the order of a and also the order of b. These orders are relatively prime, so the order of a^k must be 1. Than is $a^k = (a^k)^1 = 1$. A similar argument gives that $b^k = 1$. So we have both $o(a) \mid k$ and $o(b) \mid k$. Since $o(a)$ and $o(b)$ are relatively prime we see $o(a)o(b) \mid k$. But it is easy to verify (as hard working graduate students will) that $(ab)^{o(a)o(b)} = 1$. So we see $o(ab) = o(a)o(b)$. □

Let **G** be a group the **exponent of G** is the least natural number e so that $a^e = 1$ for all $a \in G$. Every finite group has an exponent. Certain infinite groups also have exponents, but most do not.

Fact. Let **G** be an Abelian group and suppose that a is an element of largest order and that order in finite. Then the exponent of **G** is the order of a.

Proof. Let b be an arbitrary element of G and n be the order of a. We need only show that $b^n = 1$. Now we know that the order of b is bounded above by n, so in particular it is finite. Let it be m. Now factor n and m:

$$n = p_0^{e_0} p_1^{e_1} \dots p_k^{e_k}$$
$$m = p_0^{f_0} p_1^{f_1} \dots p_k^{f_k}$$

where $p_0, \ldots, p_k$ are distinct primes and the e_j's and f_j's are natural numbers. In the event that $m|n$ we have $b^n = 1$ as desired.

So consider the case when $m \nmid n$. Then there is some $j \leq k$ so that $f_j > e_j$. Without loss (and to simplify notation) let $j = 0$.

Now put $c = a^{p_0^{e_0}}$ and $d = b^{p_1^{f_1} \cdots p_k^{f_k}}$. Then

$$o(c) = p_1^{e_1} \ldots p_k^{e_k}$$
$$o(d) = p_0^{f_0}$$

This means that the orders of c and d are relatively prime. It follows that the order of cd is $p_0^{f_0} p_1^{e_1} \ldots p_k^{e_k}$. But this is larger than n contrary to the maximality of the order of a. So we must reject this case. □

Fact. Let **G** be a finite Abelian group. **G** is cyclic if and only if $|G|$ is the exponent of **G**.

The proof is immediate from the two preceding facts.

Let $\varphi(n)$ be Euler's totient function. That is, $\varphi(n)$ is the number of natural numbers less than n that are relatively prime to n.

Fact. Let **G** be a finite cyclic group of order n. Then **G** has precisely $\varphi(n)$ elements of order n and they are those of the form a^m where a generates **G** and m is relatively prime to n with $0 \leq r < n$.

Proof. First, suppose that m satisfies the listed conditions. Let k be the order of a^m. Then k is the least positive natural number such that $1 = (a^m)^k = a^{mk}$. So $n \mid mk$. But since m and n are relatively prime, we find that $n \mid k$. On the other hand, Lagrange tells us that $k \mid n$. Thus $k = n$. So a^m indeed has order n.

Now suppose a^m has order n and $0 \leq m < n$. Let d be the greatest common divisor of m and n. Let s be the natural number so that $ds = n$ and let t be the natural number so that $dt = m$. Then $(a^m)^s = a^{dts} = (a^{ds})^t = (a^n)^t = 1^t = 1$. Since n is the order of a^m we find that $n \mid s$. On the other hand, $n = ds$. So $n = s$ and $d = 1$. Since $d = 1$ we conclude that m and n are relatively prime. □

Fact. Let **G** and **H** be finite cyclic groups of order n. Then $\varphi(n)$ is the number of isomorphism from **G** onto **H**.

Proof. We already observed that each of these groups is isomorpic to $\mathbb{Z}_n$, so there are certainly isomorpisms between them. Let a generate **G**. Now any isomorphism must preserve the order of elements and so it must take a to a generator of **G**. Suppose b is a generator of **H**. Now in our proof that these cyclic groups were isomorphic to $\mathbb{Z}_n$ we use logarithm maps. Composing the logarithm map from **G** to $\mathbb{Z}_n$ with the inverse of the logarithm map from **H** to $\mathbb{Z}_n$ we obtain an isomorphism from **G** onto **H** that sends $a^k \mapsto b^k$ for all natural numbers $k < n$. Since there are $\varphi(n)$ choices for b, we have found $\varphi(n)$ distinct isomorpisms from **G** onto **H**. Are there anymore?

Suppose Φ is an isomorphism from **G** onto **H**. Then $\Phi(a) = b$ is a generator of **H**. Moreover, we have $\Phi(a^k) = (\Phi(a))^k = b^k$ for all natural numbers $k < n$. So Φ is one of the isomorphisms counted in the previous paragraph. □

Let $U_n = \{m \mid m \text{ is a natural number relatively prime to } n \text{ and } m < n\}$. Notice that $1 \in U_n$. By imposing multiplication modulo n and the correct inversion we make a group $\mathbf{U}_n$. (In finding the inverse, the hardworking graduate students should consider that the relative primeness of m and n leads to $1 = ms + nq$ for appropriate integers s and t.) We leave the following, which is a corollary of the fact above as a challenge to the graduate students.

Fact. Let $\mathbf{G}$ be a finite cyclic group of order n. Then $\mathrm{Aut}\,\mathbf{G} \cong \mathbf{U}_n$.

Finally, here is a theorem of Euler.

Fact. For all postive natural numbers m and n that are relatively prime, we have $m^{\varphi(n)} \equiv 1 \pmod{n}$.

Proof. First of all, we may insist that $m < n$. The reason is that the modulo n residue map is a homomorphism from the ring of integers onto the ring of integers modulo n. So $m \in U_n$. By Lagrange, the order of m divides the cardinality of U_n which is $\varphi(n)$. So the desired conclusion follows. □

LECTURE 5

GROUP REPRESENTATION: GROUPS ACTING ON SETS

Let **G** be a group and let X be a set. By an **action** of **G** on X we just mean a homomorphism $\Phi : \mathbf{G} \to \mathbf{Sym}\, X$.

Cayley's Theorem gave us one example of **G** acting on G. Recall that there $\Phi_a(b) := ab$ for all $a, b \in G$. This action is sometimes called to action of **G** on G by translation (on the left). We showed this Φ is one-to-one. One-to-one actions are said to be **faithful**.

Lagrange's Theorem also suggests an action. Let **G** be a group and let **H** be a subgroup of **G**. Let X be the collection of left cosets of **H** in **G**. We can have the action Φ defined so that $\Phi_a(bH) := (ab)H$ for all $a, b \in G$.

Here is another action. Let **G** be a group. Let X be G and define Φ so that $\Phi_a(b) := a^{-1}ba$ for all $a, b \in G$. Of course, one must actually show that this Φ is a homomorphism from **G** into **Sym** G. Of course, this will be verified by the hard-working.... This is the action of **G** on G by conjugation.

Roughly speaking, the idea of respresentations is that by exploring (a number of different) concrete homomorphic images of a group we might find out more about the group itself. By analogy, think of the homomorphic images has shadows onto a two-dimensional screen of some three-dimensional object. By understanding enough of the shadows we might be able to reconstruct what the object looks like.

The language using the homomorphism Φ can be streamlined. This streamlining can sometimes be ambiguous, but usually it is not. What Φ does is associate with each $g \in G$ a permutation Φ_g of X. The first step in the streamlining is to just regard g as a name for this permutation—in other words, drop the Φ and raise the g. This means we get things like the following for all $x \in X$ and all $g, h \in G$

$$1(x) = \mathbf{1}_X(x) = x$$
$$(gh)(x) = (g \circ h)(x) = g(h(x))$$

The last step in the streamlining process is to drop a set of parentheses—that is, to write gx in place of $g(x)$. Then the equations above become, for all $x \in X$ and all $g, h \in G$,

$$1x = x$$
$$(gh)x = g(hx)$$

You might notice that the formation of inverses is not mentioned above. This is legitimate since we know that both **G** and **Sym** X are groups. Sometimes authors say that an action of **G** on X is a kind of "scalar" multiplication of group elements by elements of X that satisfies the two equations above. This amounts to the same thing since the map Φ can be recovered from this information and it can be shown to be a homomorphism.

Let Φ be an action of **G** on a set X and let $x \in X$. The **orbit** of x under the action is the set

$$\mathcal{O}_x := \{\Phi_g(x) \mid g \in G\}.$$

In the streamlined notation this becomes

$$\mathcal{O}_x := \{gx \mid g \in G\}.$$

Observe that $x \in \mathcal{O}_x$ since $1 \in G$. Also notice that If $y \in \mathcal{O}_x$, then $\mathcal{O}_x = \mathcal{O}_y$. Here is why. Let $g \in G$ so that $gx = y$. Now observe

$$\begin{aligned} z \in \mathcal{O}_y &\Leftrightarrow hy = z \text{ for some } h \in G \\ &\Leftrightarrow h(gx) = z \text{ for some } h \in G \\ &\Leftrightarrow (hg)x = z \text{ for some } h \in G \\ &\Leftrightarrow kx = z \text{ for some } k \in G \text{(careful!)} \\ &\Leftrightarrow z \in \mathcal{O}_x. \end{aligned}$$

Hence $\mathcal{O}_y = \mathcal{O}_x$. Thus the orbits of any element $x, y \in X$ either coincide or they are disjoint. This means that X is partitioned into orbits by the action of **G**.

Now let $x \in X$. The **stablizer** of x with respect to the action Φ is the following set

$$\operatorname{Stab} x := \{g \mid g \in G \text{ and } gx = x\}.$$

That is, the stablizer of x consists of all the elements of G that leave x fixed under the action. It is easy to see that $\operatorname{Stab} x$ is closed under the group operations:

$$\begin{aligned} 1x = x \text{ so } 1 \in \operatorname{Stab} x \\ gx = x \text{ and } hx = x \Rightarrow (gh)x = x \text{ so } \operatorname{Stab} x \text{ is closed under products} \\ gx = x \Rightarrow x = g^{-1}x \text{ so } \operatorname{Stab} x \text{ is closed under inverses.} \end{aligned}$$

In this way we arrive at the group **Stab** x, which is a subgroup of **G**.

Here is the

Key Fact About Group Actions. *Let the group* **G** *act on the set* X*. Then for all* $x \in X$ *we have* $|\mathcal{O}_x| = [\mathbf{G} : \mathbf{Stab}\, x]$.

Proof. Notice that $[\mathbf{G} : \mathbf{Stab}\, x]$ is the number of left cosets of $\mathrm{Stab}\, x$ in $\mathbf{G}$. To prove the Key Fact we present a one-to-one correspondence between $\mathcal{O}_x$ and the collection of left cosets of $\mathrm{Stab}\, x$. As preparation, suppose $y \in \mathcal{O}_x$. Then there is at least one $g \in G$ so that $gx = y$. Suppose also $h \in G$ and $hx = y$. Then, of course $gx = hx$ and so $(h^{-1}g)x = x$. This means that $h^{-1}g \in \mathrm{Stab}\, x$ or, what is the same, g and h are in the same left coset of $\mathrm{Stab}\, x$. This allows us to define Ψ from $\mathcal{O}_x$ to the collection of left cosets of $\mathrm{Stab}\, x$ as follows:

$$\Psi(y) := g\,\mathrm{Stab}\, x \text{ where } gx = y.$$

This definition works for all $y \in \mathcal{O}_x$. It remains to show that Ψ is the desired one-to-one correspondence.

To see one-to-oneness, let $y, z \in \mathcal{O}_x$ with $\Psi(y) = \Psi(z)$. Pick $g, h \in G$ so that $gx = y$ and $hx = z$. So we get $g\,\mathrm{Stab}\, x = h\,\mathrm{Stab}\, x$. This means $h^{-1}g \in \mathrm{Stab}\, x$. So $(h^{-1}g)x = x$. But then $gx = hx$. So we find $y = gx = hx = z$, and conclude that Ψ is one-to-one.

To see that Ψ maps $\mathcal{O}_x$ onto the collection of cosets of $\mathrm{Stab}\, x$, let $g \in G$. We must find $y \in \mathcal{O}_x$ so that $\Psi(y) = g\,\mathrm{Stab}\, x$. Let us try $y = gx$. It works, enough said. □

Let the group $\mathbf{G}$ act on the set X. By a **transversal** for this action we mean a set $T \subseteq X$ so that $|T \cap \mathcal{O}_x| = 1$ for each $x \in X$. This means that T is constituted by picking one "representative" element from each orbit. The next fact is a corollary of the Key Fact.

Fact. Let the group $\mathbf{G}$ act of the set X and let T be a transversal for this action. Then

$$|X| = \sum_{t \in T} [\mathbf{G} : \mathbf{Stab}\, t].$$

There are some interesting consequences when these notions are applied to the action of $\mathbf{G}$ on G by conjugation. Under this action

$$\Phi_g(h) := g^{-1}hg$$

for all $g, h \in G$. (This is one instance where our streamlining is unreasonable.) Our first remark is that conjugation by g is not only a permutation of G, but is, in fact, an automorphism of $\mathbf{G}$. Just observe that $g^{-1}(hk)g = g^{-1}hgg^{-1}kg = (g^{-1}hg)(g^{-1}kg)$. This means that $\Phi : \mathbf{G} \to \mathrm{Aut}\,\mathbf{G}$. Automorphisms of $\mathbf{G}$ that are conjugations by some fixed element g are called **inner automorphisms**. We see that since they constitute the image of $\mathbf{G}$ under the homomorphism Φ, the inner automorphisms of $\mathbf{G}$ form a subgroup of $\mathbf{AUT}\,\mathbf{G}$, which is in turn a subgroup of $\mathbf{Sym}\, G$. The group of inner automorphisms of $\mathbf{G}$ is denoted by $\mathbf{Inn}\,\mathbf{G}$. What is the kernel of Φ? Well, $g \in \ker\Phi$ if and only if Φ_g is $\mathbf{1}_G$ if and only if $g^{-1}hg = h$ for all $h \in G$ if and only if $hg = gh$ for all $h \in G$. This means

$$\ker\Phi = \{g \mid hg = gh \text{ for all } h \in G\}.$$

This group is call the **center** of $\mathbf{G}$ and consist of all elements of G that commute with every element of G. We use $\mathbf{Z}(\mathbf{G})$ to denote the center of the group $\mathbf{G}$. The next fact merely gathers together these findings.

Fact. Let $\mathbf{G}$ be a group. Then the center $\mathbf{Z}(\mathbf{G})$ is a normal subgroup of $\mathbf{G}$ and $\mathbf{G}/\mathbf{Z}(\mathbf{G}) \cong \mathbf{Inn}(\mathbf{G})$.

Now consider the corollary of the Key Fact, applied to the action by conjugation. We get

$$|G| = \sum_{t \in T} [\mathbf{G} : \mathbf{Stab}\, t].$$

To understand this a little better, look at **Stab** $t = \{g \mid g \in G \text{ and } g^{-1}tg = t\} = \{g \mid g \in G \text{ and } tg = gt\}$. So under this action Stab t turns out to be the set of those elements of G which commute with t. This set is called the **centralizer** of the element t and is denoted by $C(t)$. So we have

$$C(t) := \{g \mid g \in G \text{ and } tg = gt\}.$$

We know it is a subgroup of $\mathbf{G}$, as all stablizeres must be. The subgroup in denoted by $\mathbf{C}(t)$. Notice that $C(t) = G$ is equivalent to $t \in Z(\mathbf{G})$ and also to $[\mathbf{G} : \mathbf{C}(t)] = 1$. Now break the transveral T into two disjoint pieces T_0 and T_1, where $t \in T_0$ if and only if $C(t) = G$. Then we get

$$|G| = \sum_{t \in T_0} [\mathbf{G} : \mathbf{C}(t)] + \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)].$$

Now apply the Key Fact to the first sum.

$$|G| = \sum_{t \in T_0} |\mathcal{O}_t| + \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)].$$

Since the orbits are disjoint we get

$$|G| = |\bigcup_{t \in T_0} \mathcal{O}_t| + \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)].$$

But actually $T_0 = Z(\mathbf{G})$ and each $\mathcal{O}_t = \{t\}$ for $t \in T_0$. This means

$$|G| = |Z(\mathbf{G})| + \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)].$$

This equation is called the **Conjugacy Class Equaiton** or sometimes just the **Class Equation**.

Here is a useful consequence of the Conjugacy Class Equation.

Fact. Every nontrivial finite group of prime power order has a nontrivial center.

Proof. Let p be a prime number and suppose $\mathbf{G}$ is a group of order p^n where n is a positive natural number. The indices $[\mathbf{G} : C(t)]$ where $t \in T_1$ that occur in the Conjugacy Class Equation are larger than 1 and so by Lagrange each of them is some positive power of p. Thus $p \mid \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)]$ and $p \mid |G|$. Therefore $p \mid |Z(\mathbf{G})|$. So the center of $\mathbf{G}$ must have at least p elements. It is nontrivial. □

The Key Fact tells us something about the sizes of the orbits induced by the action of a group on a set. What about the number of orbits? To get at this, we need a companion to the notion of stablizer. Let $\mathbf{G}$ act on X and let $g \in G$. The **fixed set** of g is

$$\operatorname{Fix} g := \{x \mid x \in X \text{ and } gx = x\}.$$

The Cauchy-Frobenius Formula. *Let the group* $\mathbf{G}$ *act on the set* X *and let* κ *be the number of orbits of this action. Then*

$$\kappa|G| = \sum_{g \in G} |\operatorname{Fix} g|.$$

Proof. Let

$$P := \{(g,x) \mid g \in G \text{ and } x \in X \text{ with } gx = x\}.$$

Observe that for each $x \in X$ we have $(g,x) \in P$ if and only if $g \in \operatorname{Stab} x$. This gives us

$$|P| = \sum_{x \in X} |\operatorname{Stab} x|.$$

Now do the same with the other coordinate. For each $g \in G$ we have $(g,x) \in P$ if and only if $x \in \operatorname{Fix} g$. So we find

$$|P| = \sum_{g \in G} |\operatorname{Fix} g|.$$

So we find $\sum_{g \in G} |\operatorname{Fix} g| = \sum_{x \in X} |\operatorname{Stab} x|$. Now let T be a transversal of the orbits of this action. So $|T| = \kappa$, the number of orbits. But X is the disjoint union of the orbits. So we can rearrange the right-hand sum as follows:

$$\sum_{x \in X} |\operatorname{Stab} x| = \sum_{t \in T} \sum_{x \in \mathcal{O}_t} |\operatorname{Stab} x|.$$

Let $x \in \mathcal{O}_t$. Pick $h \in G$ so that $ht = x$. Then observe that for all $g \in G$ we get $gx = x \Leftrightarrow ght = ht \Leftrightarrow h^{-1}ght = t$. This means that $g \in \operatorname{Stab} x \Leftrightarrow h^{-1}gh \in \operatorname{Stab} t$. But conjugation by h is an automorphism of $\mathbf{G}$, so in particular it follows the subgroups $\mathbf{Stab}\, x$ and $\mathbf{Stab}\, t$ are isomorphic. But we only want that if $x \in \mathcal{O}_t$ then $|\operatorname{Stab} x| = |\operatorname{Stab} t|$. In this way we arrive at

$$\begin{aligned}
\sum_{g \in G} |\operatorname{Fix} g| &= \sum_{x \in X} |\operatorname{Stab} x| \\
&= \sum_{t \in T} \sum_{x \in \mathcal{O}_t} |\operatorname{Stab} x| \\
&= \sum_{t \in T} \sum_{x \in \mathcal{O}_t} |\operatorname{Stab} t| \\
&= \sum_{t \in T} |\operatorname{Stab} t| \sum_{x \in \mathcal{O}_t} 1 \\
&= \sum_{t \in T} |\operatorname{Stab} t||\mathcal{O}_t| \\
&= \sum_{t \in T} |G| \\
&= |G| \sum_{t \in T} 1 \\
&= |G||T| = \kappa|G|.
\end{aligned}$$

In the above chain of reasoning we use $|G| = |\operatorname{Stab} t||\mathcal{O}_t|$. This is just another way to state the Key Fact. □

5.1 Problem Set 17

Algebra Homework, Edition 17
Automorphisms of groups

Problem 51.
Prove that $\mathrm{Aut}(\mathbf{S}_n) \cong \mathbf{S}_n$, for ever natural number n, except when $n = 2$ or $n = 6$. You can use, without proof, that if $n \neq 6$ then, in $\mathbf{S}_n$, the image, under any automorphism, of any transposition is again a transposition.

Problem 52.
Let p be a prime number. Prove that if a and b are elements of the symmetric group S_p, where a has order p and b is a transposition, then $\{a, b\}$ generates S_p.

Problem 53.
Let $H \leq G$. Prove that $N_G(H)/C_G(H)$ is embeddable into $\mathrm{Aut}(H)$.

Problem 54.
Let G be a group of order n. Define $\varphi : G \to G$ by $\varphi(a) = a^{n^2+3n+1}$ for all $a \in G$. Prove that φ is an automorphism of G.

LECTURE 6

WHEN DOES A FINITE GROUP HAVE A SUBGROUP OF SIZE n?

Let **G** be a finite group and **H** be a subgroup of **G**. Lagrange tells us that $|H|$ must divide $|G|$. What about the converse? If n divides $|G|$ must **G** have a subgroup of order n? How many such subgroups? If not for all such n then for which?

Example. The alternating group $\mathbf{A}_4$, which has cardinality 12, has no subgroup of order 6.

Proof. By writing down the disjoint cycle decompositions one can see that in addition to the identity permutation, $\mathbf{A}_4$ has 3 elements of order 2 and 8 elements of order 3 making altogether 12 elements. The elements of order 3 are the 3-cycles and the elements of order 2 are the product of disjoint transpositions.

Let us try to make a subgroup **H** of order 6. There not being enough elements of order 2, we see that H must have an element of order 3. It does not harm to suppose that $(0,1,2) \in H$. Then the square of this element (which is also its inverse) $(0,2,1)$ also belong to H. With the identity permutation this gives us 3 of the 6 elements. We must also leave out of H the permutations $(0,2,3),(0,3,1)$, and $(1,3,2)$, since putting them in would also force in their inverses are well as their products with $(0,1,2)$. After a bit of computation we find that H would then have to have more than 6 elements. Next we see that we cannot put any of the element of order 2 into H since the product of $(0,1,2)$ with any element of order 2 yields one of the three 3-cycles we just threw out of H. This leaves 3 other 3-cycles to consider. But the product of $(0,1,2)$ with any of them yields an element of order 2. So we cannot put together six of the element of A_4 to form a subgroup. □

We cannot have the full-blown converse to Lagrange's Theorem. In the example above, while we didn't get a subgroup of order 6, we certainly saw subgroups of order 2 and 3, the primes that divide 12. Of course that is for just the one group $\mathbf{A}_4$. But it suggests a starting point. Cauchy noticed the following fact.

Fact. Let **G** be a finite group and let p be a prime number. If p divides $|G|$, then **G** has a subgroup of order p.

Proof. Since p is prime, having a subgroup of order p is the same as having an element of order p.

For the sake of contradiction, suppose it were not so. Then let **G** be a smallest finite group witnessing the failure. Consider the Conjugacy Class Equation

$$|G| = |Z(G)| + \sum_{t \in T_1} [\mathbf{G} : \mathbf{C}(t)]$$

recalling that for $t \in T_1$ we have that centralizer $\mathbf{C}(t)$ is a proper subgroup of **G**. By the minimality of **G** we find that $p \nmid |C(t)|$. Since p does divide $|G|$ we have by Lagrange that p must divide $[\mathbf{G} : \mathbf{C}(t)]$ for each $t \in T_1$. This forces the conclusion that p divides $|Z(\mathbf{G})|$. By the minimality of $|G|$ this entails that $G = Z(\mathbf{G})$. So **G** is Abelian. Now let **H** be any subgroup of **G**. Because **G** is Abelian **H** is a normal subgroup of **G**. So $\mathbf{G}/H$ is again a group and Lagrange tells us that $|G| = |G/H||H|$. So p divides $|G/H|$ or p divides $|H|$. Suppose **H** is a proper nontrivial subgroup of **G**. Then both $\mathbf{G}/H$ and **H** are smaller groups. So if p divides one of these order then it must have an element of order p. This is impossible in the case of **H** since then the same element would be an element of order p in **G**. What about if $\mathbf{G}/H$ has an element of order p? This is the same as asserting that there is $a \in G$ with $a \notin H$ but $a^p \in H$. Let r be the order of a^p in **H**. So p and r are relatively prime since p does not divide $|H|$ and $a^{pr} = 1$. But then $(a^r)^p = 1$. Since **G** has no elements of order p, it must be that $a^r = 1$. Since p and r are relatively prime, pick integers u and v so that $1 = pu + rv$. Then $a = a^1 = (a^p)^u (a^r)^v = (a^p)^u 1^v = (a^p)^u$. But $a^p \in H$. Therefore, $a \in H$ contrary to the way a was chosen.

What all that means is that **G** is a finite Abelian group with no proper nontrivial subgroups. That is, it is a finite cyclic group. But we already know that every finite cyclic group has elements of every order that divide the order of the group. That is the desired final contradiction. □

Here is a second slicker (but perhaps not as revealing a) proof published by J. H. McKay in 1959.

Proof. Let

$$X := \{(g_0, \ldots, g_{p-1}) \mid g_i \in G \text{ for all } i < p \text{ and } g_0 g_1 \ldots g_{p-1} = 1\} - \{(1, \ldots, 1)\}.$$

Let **A** be the cyclic group of order p and let a be a generator of **A**. Let **A** act on X in such a way that $a(g_0, g_1, \ldots, g_{p-1}) = (g_{p-1}, g_0, g_1, \ldots, g_{p-2})$ for any $(g_0, \ldots, g_{p-1}) \in X$. So the action of the generator of **A** is to rotate the coordinates of members of X by one notch. The hard-working graduate students should verify that the resulting rotated tuple belongs again to X and that this really does describe a group action. Now Lagrange and the Key Fact tell us that the size of each orbit must divide $|A| = p$. Because p is prime this means that an orbit must have size 1 or size p. An orbit of size one is just what we are looking for: a p-tuple that multiplies to 1 but which is the same tuple under all those rotation (i.e. every entry in the tuple is identical with every other entry).

So what if all the orbits were of size p. Since X is the disjoint union of the orbits, this would mean that $|X|$ would be divisible by p. But we can figure out the size of X. To make a p-tuple belonging to X we can make free and independent choices for the first $p-1$ entries. But then the last entry is forced to be the inverse of the product of the first $p-1$ entries. This means

$$|\{(g_0, \ldots, g_{p-1}) \mid g_i \in G \text{ for all } i < p \text{ and } g_0 g_1 \ldots g_{p-1} = 1\}| = |G|^{p-1}$$

So $|X| = |G|^{p-1} - 1$. But p divides $|G|$ so p cannot divide the cardinality of X. This means that not all the orbits can have cardinality p. So one of them must have cardinality 1. □

Let p be a prime number. A group is said to be a **p-group** provided each of its elements has an order which is a power of p. We get the following fact from Lagrange and Cauchy.

Fact. Let $\mathbf{G}$ be a finite group and p be a prime number. $\mathbf{G}$ is a p-group if and only if $|G|$ is a power of p.

Proof. First, suppose $\mathbf{G}$ is a p-group and that the order of $\mathbf{G}$ is n. By Cauchy's Theorem the only prime that can divide n is p. So n must be a power of p.

Conversely, if $|G|$ is a power of p, then by Lagrange the order of any subgroup must also be a power of p. This applies to the cyclic subgroups, so any element must have order a power of p. This means $\mathbf{G}$ is a p-group. □

Using the Correspondence Theorem, the Theorem of Lagrange, and the Theorem of Cauchy as a base, we can even get a considerable extension of Cauchy's Theorem in the case of p-groups.

Fact. Let $\mathbf{G}$ be a group, let p be a prime number, and let k be a natural number such that $|G| = p^k$. Then there is a sequence $\mathbf{G}_0 \lhd \mathbf{G}_1 \lhd \cdots \lhd \mathbf{G}_k$ of normal subgroups of $\mathbf{G}$ such that $|G_j| = p^j$ for all $j \leq k$.

Proof. Of course $\mathbf{G}_0$ will be the trivial subgroup of $\mathbf{G}$ and $\mathbf{G}_k = \mathbf{G}$.

We will prove our fact by induction on k. In the base step of the induction we have $k = 0$ and the sequence we desire has just one group in it: $\mathbf{G}$ itself, which is a trivial group. So consider the inductive step. Here we take as an hypothesis that our fact is true for k and prove it for $k+1$. So let $\mathbf{G}$ be a group of order p^{k+1}. We know that the center $\mathbf{Z}(\mathbf{G})$ is nontrivial. Cauchy tells us it must have an element of order p. Let $\mathbf{G}_1$ be the subgroup generated by such an element. Since $\mathbf{G}_1$ is a subgroup of the center, each of the elements of G_1 commutes with all the elements of G. This ensures that $\mathbf{G}_1 \lhd \mathbf{G}$. Now according to Lagrange, $\mathbf{G}/G_1$ of order p^k. By our inductive hypothesis it has a sequence

$$\mathbf{H}_0 \lhd \mathbf{H}_1 \lhd \cdots \lhd \mathbf{H}_k = \mathbf{G}/G_1$$

of normal subgroups so that $|H_j| = p^j$ for all $j \leq k$. According to the Correspondence Theorem there is a corresponding sequence

$$\mathbf{G}_1 \lhd \mathbf{G}_2 \lhd \cdots \lhd \mathbf{G}_{k+1} = \mathbf{G}$$

of normal subgroups of $\mathbf{G}$ so that $\mathbf{H}_j = \mathbf{G}_{j+1}/G_1$ for all $j \leq k$. So according to Lagrange, $|G_{j+1}| = p^{j+1}$ for all $j \leq k$. So the sequence

$$\mathbf{G}_0 \lhd \mathbf{G}_1 \lhd \mathbf{G}_2 \lhd \cdots \lhd \mathbf{G}_{k+1} = \mathbf{G}$$

is just what we desire. □

A generation after Cauchy, the great Norwegian mathematician Peter L. M. Sylow (a high school teacher for most of his life) made the breakthrough that launched a century and more of vigorous development of the rich theory of finite groups.

Let $\mathbf{G}$ be a finite group and p be a prime number. Then there must be a natural number m such that p^m divides $|G|$ but p^{m+1} does not divide $|G|$. Any subgroup of $\mathbf{G}$ of order p^m is said to be a **Sylow p-subgroup** of $\mathbf{G}$. Of course, if p does not divide $|G|$ then $m = 0$ and the Sylow p-subgroup is the trivial group (and not of much interest).

The First Sylow Theorem. *Let* **G** *be a finite group and p be a prime number. If p^k divides $|G|$, then* **G** *has a subgroup of order p^k. In particular,* **G** *has a Sylow p-subgroup.*

Proof. There is really nothing to prove unless p divides $|G|$. So we take that to be the case.

Once we prove that **G** has a Sylow p-subgroup we can appeal to a fact above to get the other subgroups we desire.

Our proof is by induction of $|G|$. As the base step is trivial, we consider just the inductive step.

Suppose **H** is a proper subgroup so that p does not divide $[\mathbf{G}:\mathbf{H}]$. Lagrange tells us that $|G| = [\mathbf{G}:\mathbf{H}]|H|$. So we see that p divides $|H|$ and that any Sylow p-subgroup of **H** is a Sylow p-subgroup of **G**. So we could appeal to the induction hypothesis to get a Sylow p-subgroup of **G**.

So it remains to consider the case that for every proper subgroup **H** of **G** we have that p divides $[\mathbf{G}:\mathbf{H}]$. Recall the Conjugacy Class Equation:

$$|G| = |Z(\mathbf{G})| + \sum_{t \in T_1} [\mathbf{G}:\mathbf{C}(t)].$$

In the sum each of the centralizers $\mathbf{C}(t)$ is a proper subgroup of **G**. So it follows from the Conjugacy Class Equation that p divides $|Z(\mathbf{G})|$. According to Cauchy, $\mathbf{Z}(\mathbf{G})$ has a subgroup **N** of order p. Since $N \subseteq Z(\mathbf{G})$ we know that **N** is a normal subgroup of **G**. But $\mathbf{G}/N$ is smaller than **G** , so by the inductive hypothesis it must have a Sylow p-subgroup $\mathbf{P}_o$. Let m be the natural number so that p^m divides $|G|$ but p^{m+1} does not divide $|G|$. Because $|N| = p$, in view of Lagrange's Theorem, we see that $|P_o| = p^{m-1}$. Now let $P = \{a \mid a \in G \text{ and } aN \in P_o\}$. It is easy to check that P is closed under the group operations (the inverse image under any homomorphism of a subgroup of the range is a subgroup of the domain). So we have a subgroup **P** of **G** and Lagrnage tells us that $|P| = |P_o||N| = p^{m-1}p = p^m$. This means that **P** is a Sylow p-subgroup of **G**, as desired. □

Our proof of the Second Sylow Theorem uses the notion of the normalizer of a subgroup. Suppose **H** is a subgroup of the group **G**. Let $N_{\mathbf{G}}\mathbf{H} := \{a \mid a \in G \text{ and } aH = Ha\}$. We see and $H \subseteq N_{\mathbf{G}}\mathbf{H} \subseteq G$. Hard-working graduates can check that $N_{\mathbf{G}}\mathbf{H}$ is closed under the group operations. So we have the subgroup $\mathbf{N_G H}$. It is called the **normalizer** of **H** in **G**. Evidently, **H** is a normal subgroup of $\mathbf{N_G H}$, and indeed the normalizer is the largest subgroup of **G** in which **H** is normal.

There is another way to get at the normalizer. Let **G** be a group and let X be the collection of all subgroups of **G**. Let **G** act on X by conjugation. Then a little work shows, for any subgroup **H**, that $\mathbf{Stab}\,\mathbf{H} = \mathbf{N_G H}$. The orbit of **H** under this action is just all the subgroups of **G** that are conjugate to **H**. The Key Fact tells us, in this setting, that the number of subgroups conjugate with **H** is $[\mathbf{G}:\mathbf{N_G H}]$.

The Second Sylow Theorem. *Let* **G** *be a finite group and let p be a prime number. Let* **P** *be a Sylow p-subgroup of* **G** *and let* **H** *be a p-subgroup of* **G**. *Then* **H** *is a subgroup of some conjugate of* **P**. *In particular, any two Sylow p-subgroups of* **G** *are conjugates.*

Proof. Pick m so that $|P| = p^m$.

Let X be the collection of subgroups of **G** that are conjugates of **P**. Let **H** act on X by conjugation. Consider one of the orbits $\mathcal{O}^H$ and let $\mathbf{P}_o$ be a member of $\mathcal{O}$. We have superscripted this orbit with H since later in this proof we will use a second action and consider on of its orbits. The Key Fact tells us

$$|\mathcal{O}^H| = [\mathbf{H}:\mathbf{Stab}\,P_o].$$

Now $\operatorname{Stab} P_o = \{h \mid h \in H \text{ and } h^{-1}P_o h = P_o\} = H \cap N_{\mathbf{G}}P_o$. Let $H_1 = H \cap N_{\mathbf{G}}P_o$. Now $\mathbf{H}_1$ is a subgroup of $\mathbf{N_G P}_o$ and $\mathbf{P}_o$ is a normal subgroup of $\mathbf{N_G P}_o$. Working inside $\mathbf{N_G P}_o$ we apply the Second Isomorphism Theorem:

$$\mathbf{H}_1\mathbf{P}_o/P_o \cong \mathbf{H}_1/H_1 \cap P_o.$$

We have $[\mathbf{H}_1\mathbf{P}_o : \mathbf{P}_o] = [\mathbf{H}_1 : \mathbf{H}_1 \cap \mathbf{P}_o]$. Since $\mathbf{H}$ is a p-group so is $\mathbf{H}_1$. So pick ℓ so that $[\mathbf{H}_1\mathbf{P}_o : \mathbf{P}_o] = p^\ell$. By Lagrange $|H_1P_o| = [\mathbf{H}_1\mathbf{P}_o : \mathbf{P}_o]|P_o| = p^\ell p^m = p^{\ell+m}$. Since $\mathbf{P}_o$ is a Sylow p-subgroup of $\mathbf{G}$, it must be that $\ell = 0$ and so $|H_1P_o| = |P_o|$. But $P_o \subseteq H_1P_o$ and these sets are finite. The means $P_o = H_1P_o$. In turn we have $H_1 \subseteq H_1P_o = P_o$. Recalling the definition of H_1, we get $H \cap N_{\mathbf{G}}P_o \subseteq P_o$. So intersecting H on both sides of this inclusion we get $H \cap N_{\mathbf{G}}P_o \subseteq H \cap P_o$. But since $P_o \subseteq N_{\mathbf{G}}P_o$ we find

$$\mathbf{Stab}\, P_o = H \cap N_{\mathbf{G}}P_o = H \cap P_o.$$

So the size of our arbitrary orbit $\mathcal{O}^H$ is $[\mathbf{H} : \mathbf{H} \cap \mathbf{P}_o]$. Notice that this must be a power of p.

On the other hand, if we let $\mathbf{G}$ act on X by conjugation then, as noted above the statement of the theorem, $|\mathcal{O}_P^G| = [\mathbf{G} : \mathbf{N_G P}]$. Since $\mathbf{P}$ is a Sylow p-subgroup of $\mathbf{G}$ we see that p cannot divide $|\mathcal{O}_P^G|$. Since $\mathcal{O}_P^G$ is a disjoint union of H-orbits and each H-orbit has cardinality a power of p, there must be at least one orbit whose size is $p^0 = 1$. Let $P_o = a^{-1}Pa$ be the element of this orbit. Then $[\mathbf{H} : \mathbf{H} \cap \mathbf{P}_o] = 1$, the size of the orbit. This means $H = H \cap P_o$. Hence $H \subseteq P_o = a^{-1}Pa$, as desired. □

Here is a useful corollary of the Second Sylow Theorem.

Fact. If $\mathbf{G}$ is a finite group and p is a prime number. A Sylow p-subgroup of $\mathbf{G}$ is normal if and only if $\mathbf{G}$ has exactly one Sylow p-subgroup.

So can we get a handle on the number of Sylow p-subgroups a finite group might have?

The Third Sylow Theorem. *Let $\mathbf{G}$ be a finite group and let p be a prime number. Then the number of distinct Sylow p-subgroups of $\mathbf{G}$ is congruent to 1 modulo p and divides $|G|$.*

Proof. Let X be the collection of all Sylow p-subgroups of $\mathbf{G}$. Let $\mathbf{P}$ be a Sylow p subgroup of $\mathbf{G}$ and let $\mathbf{P}$ act on X by conjugation. Consider any orbit $\mathcal{O}$ of this action and let $\mathbf{P}_o \in \mathcal{O}$. Now just as in the proof for the Second Sylow Theorem (letting $\mathbf{P}$ play the role here that $\mathbf{H}$ played there), we find

$$|\mathcal{O}| = [\mathbf{P} : \mathbf{P} \cap \mathbf{P}_o].$$

Again we find that each orbit has cardinality a power of p Observe that $\{\mathbf{P}\}$ is an orbit of this action and it is of size 1. For any other orbit $\mathcal{O}$ we have for $\mathbf{P}_o \in \mathcal{O}$ that $P \neq P_o$ so that $P \cap P_o$ is strictly smaller that P. This entails, by the equation displayed above, that p divides the size of every orbit different from $\{\mathbf{P}\}$. But X is a disjoint union of the orbits, so $|X|$ is the sum of the size of the orbits. So we get that $|X|$, the number of Sylow p-subgroups of $\mathbf{G}$, is congruent to 1 modulo p.

On the other hand, by letting $\mathbf{G}$ act on X by conjugation we get only one orbit, according to the Second Sylow Theorem. So letting $\mathbf{P}$ be any Sylow p-subgroup (we have one by the First Sylow Theorem), the Key Fact tells us

$$|X| = [\mathbf{G} : \mathbf{Stab}\, \mathbf{P}].$$

By Lagrange we have $|G| = [\mathbf{G} : \mathbf{Stab}\, \mathbf{P}]|\operatorname{Stab} \mathbf{P}| = |X||\operatorname{Stab} \mathbf{P}|$. So the number $|X|$ is Sylow p-subgeroups of $\mathbf{G}$ divides the order of $\mathbf{G}$. □

6.1 Problems Set 15

ALGEBRA HOMEWORK, EDITION 15
ASK SYLOW

PROBLEM 55.
Let p be the smallest prime that divides the cardinality of the finite group **G**. Prove that any subgroup of **G** of index p must be normal.

PROBLEM 56.
How many elements of order 7 are there in a simple group of order 168?

PROBLEM 57.
Let **N** be a normal subgroup of the finite group **G** and let **K** be a p-Sylow subgroup of **N** for some prime p. Prove that $\mathbf{G} = \mathbf{N}_G(K)\mathbf{N}$.

PROBLEM 58.
Prove that there is no simple group of order 56.

LECTURE 7

DECOMPOSING FINITE GROUPS

We have seen the Structure Theorem for Finitely Generated Modules over a Principal Ideal Domain. That theorem said there was a way to assemble each such module from indecomposable pieces in a way that was essentially unique. Recall that it had three aspects: an existence statement, a uniqueness statement, and a description of the indecomposable modules. Roughly speaking, such a theorem opens a way to tackle many problems: first figure out what happens to the indecomposable pieces and then figure out what goes on when you put the indecomposable pieces together to form more complicate modules. Another very useful consequence was the association with each such finitely generated module a sequence of numbers that determines the module up to isomorphism.

Of course, that theorem gave us an excellent structure theorem for finitely generated Abelian groups. Here we want to address the question of whether there is a similar result that applies to all finite groups or at least some way to pull a complicated finite group apart into less intricate pieces.

7.1 DIRECT PRODUCTS OF GROUPS

Since we know how to form direct products of any system of algebras all of the same signature, we know how to form direct products of any system of groups and, as we observed after Cayley's Theorem, such direct products will again be groups.

Just as for modules, so for groups we can give a nice internal representation of the direct product of two groups $\mathbf{N}$ and $\mathbf{H}$.

Indeed, notice that in $\mathbf{N}\times\mathbf{H}$ we have that $N^* := \{(a,1) \mid a \in N\}$ is the kernel of the projection from $\mathbf{N}\times\mathbf{H}$ onto $\mathbf{H}$ and that $H* := \{(1,b) \mid b \in H\}$ is the kernel of the other projection function. Observe that we have the following properties:

(a) $\mathbf{N}^* \lhd \mathbf{N}\times\mathbf{H}$.

(b) $\mathbf{H}^* \lhd \mathbf{N}\times\mathbf{H}$.

(c) $\mathbf{N}^*\mathbf{H}^* = \mathbf{N}\times\mathbf{H}$.

(d) $N^* \cap H^*$ is trivial.

(e) $\mathbf{N}^* \cong \mathbf{N}$.

(f) $\mathbf{H}^* \cong \mathbf{H}$.

On the other hand, let us start with a group $\mathbf{G}$ and subgroups $\mathbf{N}$ and $\mathbf{H}$ such that

(a) $\mathbf{N} \lhd \mathbf{G}$.

(b) $\mathbf{H} \lhd \mathbf{G}$.

(c) $\mathbf{NH} = \mathbf{G}$.

(d) $N \cap H$ is trivial.

Then it is an enjoyable task for hard-working graduate students to verify that $\mathbf{G} \cong \mathbf{N} \times \mathbf{H}$.

So we will say that $\mathbf{G}$ is the **(internal) direct product** of its subgroups $\mathbf{N}$ and $\mathbf{H}$ provided

(a) $\mathbf{N} \lhd \mathbf{G}$.

(b) $\mathbf{H} \lhd \mathbf{G}$.

(c) $\mathbf{NH} = \mathbf{G}$.

(d) $N \cap H$ is trivial.

We write $\mathbf{G} = \mathbf{N} \otimes \mathbf{H}$ to mean that $\mathbf{G}$ is the internal direct product of $\mathbf{N}$ and $\mathbf{H}$. Of course, $\mathbf{N} \otimes \mathbf{H} \cong \mathbf{N} \times \mathbf{H}$.

Here is a fact that hard-working graduate students should enjoy proving.

Fact. Let $\mathbf{G}$ be a finite group so that each of is Sylow subgroups is normal. Then $\mathbf{G}$ is the (internal) direct product of its Sylow subgroups.

Recall that we should say that a group $\mathbf{G}$ is directly indecomposable provided

- $\mathbf{G}$ is nontrivial and,
- if $\mathbf{G} = \mathbf{N} \otimes \mathbf{H}$, then either $\mathbf{N}$ or $\mathbf{H}$ is trivial.

The Krull-Schmidt Theorem. *Any finite group can be decomposed as a direct product of directly indecomposable groups. Any two such decompositions of the same finite group must have the same number (counting multiplicity) of direct factors and, after some rearranging of the factors, the corresponding direct factors in each decomposition are isomorphic.*

Thus a finite group has **unique direct factorization property**: it can be directly decomposed into directly indecomposable factors and the decomposition is unique (in the sense expressed in the theorem).

Even though I called this the Krull-Schmidt Theorem (as it is commonly called in the literature) it was known to J. H. M. Wedderburn and R. Remak is the early years of the 20^{th} century.

This theorem is somewhat more troublesome to prove than the Structure Theorem for Finitely Generated Modules over a PID. Moreover, a description of the directly indecomposable finite groups seems currently out of reach (even though a century has passed since this theorem was

first proved). The lack a such a description limits some of the usefulness of the Krull-Schmidt Theorem.

No proof is included here (but there are a number of accessible proofs in the literature).

The Krull-Schmidt Theorem has been extended a number of ways. It remains true (and is still called the Krull-Schmidt Theorem) when the finite group is expanded by one-place operations that are endomorphisms of the original group. These kinds of expanded groups are called **groups with operators**. It also remains true, even in the expanded form, when the finiteness restriction is weakened to the restriction that the congruence lattice of the group (with operators) satisfies the finite chain condition. This is what Krull and Schmidt did in the 1920's. There is also a Krull-Schmidt Theorem for modules satisfying the finite chain condition on their lattices of submodules.

There are more difficult and more far-reaching theorems that extend the Krull-Schmidt Theorem that are due to Garrett Birkhoff and to Bjarni Jónsson. These theorems depend on properties of the congruences of the algebras and of their congruence lattices and will not be formulated here.

In another direction there is a really striking extension of the Krull-Schmidt Theorem due to Bjarni Jónsson and Alfred Tarski. An algebra **A** is said to be an **algebra with a zero** provided **A** has among its basic operations an element designated by 0 and a two-pace operation + satisfying the following properties:

(a) The set $\{0\}$ is closed under all the basic operations.

(b) The equations $x+0=0+x=x$ hold in the algebra **A**.

The Jónsson-Tarski Theorem. *Every finite algebra with a zero is uniquely factorable.*

Algebras with a zero retain just a whiff of groupness: a two-place operation with a two-sided identity element so that the identity element constitutes a one-element subuniverse. No associativity is assumed nor any inverses. The other basic operations can be completely unrestricted, apart from the stipulation that if 0 is plugged into each input position, then the output is also 0. This whiff is enough!

7.2 Decomposing a group using a chain of subgroups

We saw another way to take a group apart. When **G** is a finite p-group, where p is a prime number, we saw that there was a sequence

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_s$$

of normal subgroups of **G** such that $\mathbf{G}_s$ is trivial and each $\mathbf{G}_k/G_{k+1}$ has a cyclic group of order p. So we conceive **G** has a sort of increasing union where the steps $\mathbf{G}_k/G_{k+1}$ are especially simple.

We weaken this in a couple of ways to reach the notion of a solvable group. We say a group **G** is **solvable** provided there is a finite sequence of subgroups of **G** such that

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_s$$

where $\mathbf{G}_s$ is trivial and the factor groups $\mathbf{G}_k/G_{k+1}$ are Abelian for all $k < s$. Here we did not that insist that each $\mathbf{G}_k$ was a normal subgroup of **G**. We also only required the factor groups to be Abelian rather than the more stringent requirement that they be cyclic.

Sequences like the one appearing in the definition of solvable, but without the stipulation about the factor groups, are called **normal series**. Some authors call them *subnormal* series since the groups involved may not actually be normal subgroups of **G**. Since this label might bear a demeaning psychological connotation, other authors use *normal series*.

The following fact just records an obvious point and restates a previous Fact.

Fact. Each Abelian group and each finite p-group, where p is a prime number, is solvable.

With this definition in hand, the hard-working graduate student should also be in a position to prove that both $\mathbf{S}_3$ and $\mathbf{S}_4$ are solvable.

Recall that a group **G** is said to be **simple** provided it has exactly two normal subgroups. This is equivalent to saying **G** is nontrivial and its only normal subgroups are the trivial subgroup and **G** itself.

A **composition series** of a group is a normal series in which every factor group is simple. In view of the Correspondence Theorem, another way to say this is that for any link $\mathbf{G}_i \rhd \mathbf{G}_{i+1}$ in the series there is no group **H** properly between $\mathbf{G}_i$ and $\mathbf{G}_{i+1}$ so that $\mathbf{G}_i \rhd \mathbf{H} \rhd \mathbf{G}_{i+1}$. In other words, the normal series cannot be made longer by inserting additonal groups in the series. The series we devised for finite p-groups was a composition series since each factor was a cyclic group of prime order and such groups are simple.

It is clear that every normal series for a finite group can be enhanced by the insertion of additional groups until a composition series is obtained. In particular, every finite group has at least one composition series.

Fact. Let **G** be a finite group. **G** is solvable if and only if **G** has a composition series in which each factor group is a finite cyclic group of prime order.

Proof. Since every composition series is a normal series and since every cyclic group is Abelian, we see that the condition about composition series implies that **G** is solvable.

For the converse, suppose **G** is solvable. Let

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_s$$

witness the solvability of **G**. Obtain from this series a composition series by inserting additional subgroups along the series. So a part of this composition series would be

$$\mathbf{G}_i \rhd \mathbf{H}_1 \rhd \cdots \rhd \mathbf{H}_k \rhd \mathbf{G}_{i+1}.$$

Now consider the situation where we have three groups so that $\mathbf{K} \rhd \mathbf{L} \rhd \mathbf{M}$ and we know that $\mathbf{K}/M$ is Abelian. By the Third Isomorphism Theorem we have

$$(\mathbf{K}/M) \,/\, (L/M) \cong \mathbf{K}/L.$$

Since $\mathbf{K}/M$ is Abelian and $\mathbf{K}/L$ is a homomorphic image of $\mathbf{K}/M$, and every homomorphic image of an Abelian group is Abelian, we find that $\mathbf{K}/L$ is Abelian. We also see that $\mathbf{L}/M$ is a subgroup of $\mathbf{K}/M$. So $\mathbf{L}/M$ is also Abelian. This means each time we insert a new subgroup in our normal series we get a longer normal series for which all the factor groups are still Abelian. This means that the composition series we ultimately obtain has the property that each of its factor groups is a finite simple Abelian group. But the hard-working graduate students will have no trouble convincing themselves that the finite simple Abelian groups are exactly the (cyclic) groups of prime order. Of course many different primes might be associated in this way with our composition series. □

There is another characterization of the notion of solvable group which we will find useful. It involves the notion of the commutator of two normal subgroups of a group. We start by devising a new two-place operation on a group. Let $\mathbf{G}$ be a group and $a, b \in G$. We put $[a,b] := a^{-1}b^{-1}ab$ and call it the **commutator** of a and b. Notice that if a and b commute with each other in $\mathbf{G}$, then $[a,b] = 1$. Also it proves convenient to know that $[a,b]^{-1} = [b^{-1}ab, b^{-1}]$, as can be verified by hard-working graduate students with a bit of calculation.

Now let $\mathbf{N}$ and $\mathbf{K}$ be normal subgroups of the group $\mathbf{G}$. We let $[\mathbf{N},\mathbf{K}]$ denote the subgroup of $\mathbf{G}$ that is generated by the set $\{[a,b] \mid a \in N \text{ and } b \in K\}$. We call $[\mathbf{N},\mathbf{K}]$ the **commutator** of $\mathbf{N}$ and $\mathbf{K}$.

Fact. Let $\mathbf{G}$ be a group with normal subgroups $\mathbf{N}$ and $\mathbf{K}$. Then the elements of $[\mathbf{N},\mathbf{K}]$ are exactly the elements of G of the form

$$[a_0,b_0][a_1,b_1]\dots[a_k,b_k]$$

where k is some natural number and $a_i \in N$ and $b_i \in K$ for each $i \leq k$.

This Fact just depends of the normality of $\mathbf{N}$ and the fact that $\mathbf{H}$ is a subgroup of $\mathbf{G}$, view of the description of $[a,b]^{-1}$ given above. Some crucial properties of the commutator are gathered in the next Fact. Its proof requires a straightforward pleasant effort from the hard-working graduate students.

Fact. Let $\mathbf{G}$ be a group with normal subgroups $\mathbf{N}$ and $\mathbf{K}$ and subgroup $\mathbf{H}$. Then

(a) $[\mathbf{N},\mathbf{K}] \lhd \mathbf{G}$.

(b) $[\mathbf{N},\mathbf{K}] \subseteq \mathbf{N} \cap \mathbf{K}$.

(c) $[\mathbf{H},\mathbf{H}] \leq [\mathbf{G},\mathbf{G}]$.

(d) $[\mathbf{H}/N, \mathbf{H}/N] = [\mathbf{H},\mathbf{H}]/N$.

In conclusion (d) above we mean by $\mathbf{H}/N$ and $[\mathbf{H},\mathbf{H}]/N$ the subgroups of $\mathbf{G}/N$ that are the images of $\mathbf{H}$ and of $[\mathbf{H},\mathbf{H}]$ respectively under the quotient map. Some care is needed in noting this, because it may well be that $\mathbf{N}$ is neither a subgroup of $\mathbf{H}$ nor of $[\mathbf{H},\mathbf{H}]$. We also observe that $[\mathbf{H},\mathbf{H}]$ and $[\mathbf{H}/N, \mathbf{H}/N]$ are to be understood for the commutator in th groups $\mathbf{H}$ and $\mathbf{H}/N$ respectively.

Let $\mathbf{G}$ be a group. The **derived group** of $\mathbf{G}$ is the group $\mathbf{G}' := [\mathbf{G},\mathbf{G}]$. We can iterate this formation of derived groups by the following recursion.

$$\mathbf{G}^{(0)} := \mathbf{G}$$
$$\mathbf{G}^{(k+1)} := [\mathbf{G}^{(k)}, \mathbf{G}^{(k)}] \text{ for all natural numbers } k$$

So $\mathbf{G}' = \mathbf{G}^{(1)}$, $(\mathbf{G}')' = \mathbf{G}^{(2)}$, and so on.

Perhaps the next Fact gives more support to the label "commutator".

Fact. Let $\mathbf{G}$ be a group. Then $\mathbf{G}/[\mathbf{G},\mathbf{G}]$ is Abelian; moreover, if $\mathbf{N} \lhd \mathbf{G}$ and $\mathbf{G}/N$ is Abelian, then $[\mathbf{G},\mathbf{G}] \subseteq N$.

Proof. Let $a, b \in G$. The cosets $a[\mathbf{G},\mathbf{G}]$ and $b[\mathbf{G},\mathbf{G}]$ are arbitrary elements of $\mathbf{G}/[\mathbf{G},\mathbf{G}]$. To say that they commute is just to assert $ab[\mathbf{G},\mathbf{G}] = ba[\mathbf{G},\mathbf{G}]$. But this is evidently the same as asserting $[a,b] = (ba)^{-1}(ab) \in [\mathbf{G},\mathbf{G}]$. This assertion is certainly true since we took the elements of the form $[a,b]$ as generators of $[\mathbf{G},\mathbf{G}]$.

Now suppose $\mathbf{N} \lhd \mathbf{G}$ and $\mathbf{G}/N$ is Abelian. Let $a, b \in G$. So we see that $abN = baN$. But this means $[a,b] \in N$. So all the generators of $[\mathbf{G},\mathbf{G}]$ belong to N. Since $\mathbf{N}$ is a subgroup, this entails that $[\mathbf{G},\mathbf{G}] \subseteq N$. □

So we see that for an arbitrary group $\mathbf{G}$ we have

$$\mathbf{G} = \mathbf{G}^{(0)} \rhd \mathbf{G}^{(1)} \rhd \mathbf{G}^{(2)} \rhd \cdots \rhd \mathbf{G}^{(k)} \rhd \mathbf{G}^{(k+1)} \rhd \ldots$$

As far as it goes it is a normal series (and moreover each of the groups in even a normal subgroup of $\mathbf{G}$) and each factor group is Abelian. Here is our characterization of solvability.

Fact. Let $\mathbf{G}$ be a group. The group $\mathbf{G}$ is solvable if and only if $\mathbf{G}^{(n)}$ is the trivial group, for some natural number n.

Proof. In the event that $\mathbf{G}^{(n)}$ is trivial, the series

$$\mathbf{G} = \mathbf{G}^{(0)} \rhd \mathbf{G}^{(1)} \rhd \mathbf{G}^{(2)} \rhd \cdots \rhd \mathbf{G}^{(n)}$$

witnesses that $\mathbf{G}$ is solvable.

For the converse, suppose that $\mathbf{G}$ is solvable and let

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_n$$

be a normal series that witnesses the solvability of $\mathbf{G}$. So $\mathbf{G}_n$ is trivial and all the factor groups are Abelian. Consider the first link $\mathbf{G} \rhd \mathbf{G}_1$. We certainly get $\mathbf{G}_1 \rhd [\mathbf{G},\mathbf{G}] = \mathbf{G}^{(1)}$. Similarly, at the next link we see $\mathbf{G}_2 \rhd [\mathbf{G}_1,\mathbf{G}_1]$. But we already know $\mathbf{G}_1 \rhd \mathbf{G}^{(1)}$. Since we know the commutator respects the inclusion ordering, we get $[\mathbf{G}_1,\mathbf{G}_1] \rhd [\mathbf{G}^{(1)},\mathbf{G}^{(1)}] = \mathbf{G}^{(2)}$. Putting things together, we get $\mathbf{G}_2 \rhd \mathbf{G}^{(2)}$. Continuing in this way, we get $\mathbf{G}_k \rhd \mathbf{G}^{(k)}$ in general. So at the end we have $\mathbf{G}_n \rhd \mathbf{G}^{(n)}$. Since $\mathbf{G}_n$ is trivial, we find that $\mathbf{G}^{(n)}$ is also trivial, as desired. □

Fact. Every subgroup of a solvable group is solvable. Every homomorphic image of a solvable group is solvable. Let $\mathbf{N}$ be a normal subgroup of the group $\mathbf{G}$. $\mathbf{G}$ is solvable if and only if both $\mathbf{N}$ and $\mathbf{G}/N$ are solvable.

Proof. Let $\mathbf{G}$ be a solvable group.

For any normal subgroup $\mathbf{N}$ of $\mathbf{G}$ we know $[\mathbf{G}/N,\mathbf{G}/N] = [\mathbf{G},\mathbf{G}]/N$. Another way to write this is $(\mathbf{G}/N)^{(1)} = \mathbf{G}^{(1)}/N$. Using this equality, we also see

$$(\mathbf{G}/N)^{(2)} = [(\mathbf{G}/N)^{(1)}),(\mathbf{G}/N)^{(1)})] = [\mathbf{G}^{(1)}/N, \mathbf{G}^{(1)}/N] = [\mathbf{G}^{(1)},\mathbf{G}^{(1)}]/N = \mathbf{G}^{(2)}/N$$

Proceeding in this way we find $(\mathbf{G}/N)^{(k)} = \mathbf{G}^{(k)}/N$. So if $\mathbf{G}^{(n)}$ turns out to be the trivial group, then so will $(\mathbf{G}/N)^{(n)}$. This means that if $\mathbf{G}$ is solvable, then so is its homomorphic image $\mathbf{G}/N$.

For any subgroup $\mathbf{H}$ of $\mathbf{G}$ we know that $\mathbf{H}^{(1)} = [\mathbf{H},\mathbf{H}] \leq [\mathbf{G},\mathbf{G}] = \mathbf{G}^{(1)}$. An easy induction argument shows that $\mathbf{H}^{(n)} \leq \mathbf{G}^{(n)}$ for all natural numbers n. So if $\mathbf{G}$ is solvable so must its subgroup $\mathbf{H}$ be solvable.

Now suppose that $\mathbf{G}$ is an arbitrary group and that $\mathbf{N}$ is a normal subgroup such that both $\mathbf{N}$ and $\mathbf{G}/N$ are solvable. Pick a natural number so that $(\mathbf{G}/N)^{(n)}$ is trivial. Since we now know that $(\mathbf{G}/N)^{(n)} = \mathbf{G}^{(n)}/N$ it follows that $\mathbf{G}^{(n)} \leq \mathbf{N}$. But $\mathbf{N}$ is solvable, so we know all its subgroups are solvable. This means we can pick a natural number m so that $(\mathbf{G}^{(n)})^{(m)}$ is trivial. But it is easy to discover that $(\mathbf{G}^{(n)})^{(m)} = \mathbf{G}^{(n+m)}$, which must be trivial. So $\mathbf{G}$ is solvable. □

A somewhat different proof could be mounted that involves manipulating the normal series witnessing the various solvability constraint. Those proofs make heavy use of the isomorphism theorems.

So far this approach to decomposing a group using a normal series has concentrated on existence. We have seen that at least every finite group has a composition series (where the factor groups are all simple). For solvable groups we even got the existence of a composition series where the factor groups were cyclic groups of prime order. What about uniqueness?

Even for finite Abelian groups it easy to find examples where there are several different composition series. This is something like the situation with direct decompositions—one could get a different decomposition just by rearranging the direct factors in the direct product and swapping out some factors with isomorphic copies. So we aim to prove a kind of uniqueness theorem for composition series.

Let $\mathbf{G}$ be a group. We will say that two normal series for $\mathbf{G}$

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd G_n$$
$$\mathbf{G} = \mathbf{H}_0 \rhd \mathbf{H}_1 \rhd \cdots \rhd H_m$$

are **equivalent** provided $n = m$ and for some permutation σ of $\{0, 1, \ldots, n\}$ we have

$$\mathbf{G}_i / G_{i+1} \cong \mathbf{H}_{\sigma(i)} / H_{\sigma(i)+1} \text{ for all } i < n$$

That is, the series are the same length and the sequence of factor groups along one of the normal series can be rearranged to obtain, within isomorphism, the sequence of factor groups along the other normal series.

Our aim is to prove

The Jordan-Hölder Theorem. *Any composition series of a group is equivalent to any other compositions series.*

We will be able to obtain this theorem as an immediate consequence of another theorem.

Let $\mathbf{G}$ be a group. We say one normal series for $\mathbf{G}$ is a **refinement** of another if the first can be obtained from the second by inserting some finite number of additional subgroups along the series. The Jordan-Hölder Theorem is an immediate consequence of

Schreier's Refinement Theorem. *Any two normal series for a group have refinements that are equivalent to each other.*

Proof. Let the group $\mathbf{G}$ have two normal series

$$\mathbf{G} = \mathbf{A}_0 \rhd \mathbf{A}_1 \rhd \cdots \rhd \mathbf{A}_n$$
$$\mathbf{G} = \mathbf{B}_0 \rhd \mathbf{B}_1 \rhd \cdots \rhd \mathbf{B}_m.$$

So we know that both $\mathbf{A}_n$ and $\mathbf{B}_m$ are the trivial subgroup of $\mathbf{G}$.

We will invoke the Zassenhaus Butterfly Lemma to construct the two refinements we require. From the coarsest view, that lemma allows us to insert in

$$\mathbf{A} \rhd \mathbf{A}^*$$
$$\mathbf{B} \rhd \mathbf{B}^*$$

two additional groups each so that

$$\mathbf{A}\rhd \geq \mathbf{A}_u \rhd \mathbf{A}_\ell \geq \mathbf{A}^*$$
$$\mathbf{B}\rhd \geq \mathbf{B}_u \rhd \mathbf{B}_\ell \geq \mathbf{B}^*$$
$$\mathbf{A}_u / A_\ell \cong \mathbf{B}_u / B_\ell$$
$$.$$

This coarse view is not adequate for our purposes because some of the subgroup relations are not normal. But the Butterfly Lemma is certainly tempting due to that isomorphism. Fortunately, the actual Butterfly Lemma carries more detail.

Here is what works. Let

$$\mathbf{C}_{i,j} := \mathbf{A}_{i+1}(\mathbf{A}_i \cap \mathbf{B}_j) \text{ and } \mathbf{D}_{i,j} := \mathbf{B}_{j+1}(\mathbf{B}_j \cap \mathbf{A}_i)$$

These are the groups that come up in full detail in the Butterfly Lemma. What the Butterfly Lemma say about them is

$$\mathbf{C}_{i,j} \rhd \mathbf{C}_{i,j+1}$$
$$\mathbf{D}_{i,j} \rhd \mathbf{D}_{i+1,j}$$
$$\mathbf{C}_{i,j} / C_{i,j+1} \cong \mathbf{D}_{i,j} / D_{i+1,j}$$

To understand better what is going on, fix a value of i. Then we see

$$\mathbf{C}_{i,0} = \mathbf{A}_{i+1}(\mathbf{A}_i \cap \mathbf{B}_0) = \mathbf{A}_{i+1}(\mathbf{A}_i \cap \mathbf{G}) = \mathbf{A}_i$$
$$\mathbf{C}_{i,1} = \mathbf{A}_{i+1}(\mathbf{A}_i \cap \mathbf{B}_1)$$
$$\vdots$$
$$\mathbf{C}_{i,m} = \mathbf{A}_{i+1}(\mathbf{A}_i \cap \mathbf{B}_m) = \mathbf{A}_{i+1}$$

where the last line comes about since $\mathbf{B}_m$ is the trivial subgroup. Moreover, the Butterfly Lemma tells us

$$\mathbf{A}_i = \mathbf{C}_{i,0} \rhd \mathbf{C}_{i,1} \rhd \cdots \rhd \mathbf{C}_{i,m} = \mathbf{A}_{i+1}$$

In this way, we see that the $\mathbf{C}_{i,j}$'s, once they are arranged in the proper order, form a normal series for $\mathbf{G}$ that refines the series of the $\mathbf{A}_i$'s. The proper order is the lexicographic order on $\{(i,j) \mid i < n \text{ and } j < m\}$ given by

$$(i,j) \geq (\ell,k) \Leftrightarrow i \geq \ell \text{ or else } i = \ell \text{ and } j \geq k.$$

Of course an entirely similar analysis leads us to the conclusion that the $\mathbf{D}_{i,j}$'s, arranged properly, form a normal series that refines the series of the $\mathbf{B}_j$'s.

But these two refinements are equivalent since for all $i \leq n$ and all $j \leq m$ we have

$$\mathbf{C}_{i,j} / C_{i,j+1} \cong \mathbf{D}_{i,j} / D_{i+1,j}.$$

□

You should notice that we did not that insist that the normal subgroups along a normal series be proper subgroups. It would be sensible to insist on this since it gives a cleaner connotation to the length of a series. Then in the proof above one must systematically delete one of the groups when

$$\mathbf{C}_{i,j} = \mathbf{C}_{i,j+1} \text{ or } \mathbf{D}_{i,j} = \mathbf{D}_{i+1,j}.$$

Since we know $\mathbf{C}_{i,j}/C_{i,j+1} \cong \mathbf{D}_{i,j}/D_{i+1,j}$ every deletion from the series of $\mathbf{C}_{i,j}$'s must be accompanied by a deletiion from the series of $\mathbf{D}_{i,j}$'s, and vice versa. So after all the deletions, the resulting refinements will still have the same length.

The Jordan-Hölder Theorem was proved, in some form, late in the 19$^{\text{th}}$ century when Otto Hölder put the finishing touches on a proof of Camille Jordan. There are a number of different ways to prove this. For finite groups, it is possible to devise a proof by induction of the size of the group. It is also possible to make a proof by induction of the length of the composition series involved. Otto Schreier's proof of the Jordan-Hölder Theorem using the Refinement Theorem was published in 1928. Hölder wa still alive but Jordan had died six years earlier. Hans Zassenhaus gave a new proof of Shreier's Refinement Theorem using his own Butterfly Lemma in the early 1930's while he was still a hard-working graduate student under the direction of Emil Artin. Zassenhaus himself became a prolific mathematician with over 200 papers and 41 PhD students. He died in 1991.

The theorem is sometimes called the Jordan-Hólder-Schreier Theorem or even the Jordan-Hölder-Schreier-Zassenhaus Theorem. It attaches to every finite group a sequence of finite simple groups that provides some structural information about the finite group.

7.3 Addendum: A notion related to solvability

It is an easy observation that every finite Abelian group can be decomposed as a direct product of its Sylow subgroups. In fact one path to the Fundamental Structure Theorem for Finite Abelian Groups starts from this observation. One could consider the class of all finite groups that can the decomposed as a direct product of their Sylow subgroups. This proves to be a class that is wider than the class of finite Abelian groups but narrower than the class of finite solvable groups. Finite groups that are the direct product of their Sylow subgroups are called *nilpotent*. Just as the class of solvable groups can be characterized using the commutator of normal subgroups, so can the class of nilpotent groups.

Let $\mathbf{G}$ be a group. The $\mathbf{G}^{[n]}$ is defined by the following recursion:

$$\mathbf{G}^{[0]} = \mathbf{G}$$
$$\mathbf{G}^{[k+1]} = [\mathbf{G}^{[k]}, \mathbf{G}] \text{ for all natural numbers } k.$$

An easy induction shows the $\mathbf{G}(k)$ is always a subgroup of $\mathbf{G}^{[k]}$.

The group $\mathbf{G}$ is said to be **nilpotent** provided $\mathbf{G}^{[n]}$ is trivial for some natural number n. This definition works for groups that might not be finite. So it is true that every nilpotent group is solvable, although the converse fails. The class of nilpotent groups in a proper subclass of the class of solvable groups. It is also a theorem that a finite group is nilpotent if and only if it is the direct product of its Sylow subgroups. The theories of nilpotent and of solvable groups have elaborate developments, exhibiting many parallels and interconnections.

It was, of course, Galois that first noticed the significance of the class of (finite) solvable groups in his investigations of the solvability of polynomial equations of degree n by means of radicals.

One of the most notable theorems about solvable groups is

The Feit-Thomspon Theorem. *Every group of odd order is solvable.*

The proof extends to hundreds of pages.

7.4 Problem Set 16

ALGEBRA HOMEWORK, EDITION 16
MR. SYLOW DEALS WITH FINITE ABELIAN GROUPS

PROBLEM 59.
Prove that if **G**, **H**, and **K** are finite Abelian groups and $\mathbf{G} \times \mathbf{H} \cong \mathbf{G} \times \mathbf{K}$, then $\mathbf{H} \cong \mathbf{K}$.

PROBLEM 60.
Prove that every group of order 35 is cyclic.

PROBLEM 61.
Describe, up to isomorphism, all groups of order 1225.

PROBLEM 62.
Let **G** be a finite Abelian group. Prove that if $|G|$ is not divisible by k^2 for any $k > 1$, then **G** is cyclic.

LECTURE 8

WHERE TO FIND THE ROOTS OF A POLYNOMIAL

A leading motivation for the rest of the semester is the project: to describe all the roots of a given polynomial in one variable with coefficients from some field.

Let **F** be a field and $f(x) \in F[x]$ be a polynomial with coefficients from F. For instance, **F** might be the field $\mathbb{Q}$ of rational numbers and $f(x)$ might be $x^2 - 2$. This polynomial has no roots in $\mathbb{Q}$, but on the other hand, $f(x)$ is also a polynomial over the field $\mathbb{R}$ real numbers and in this larger field we find two roots of $f(x)$, namely $\sqrt{2}$ and $-\sqrt{2}$. After a bit of reflection, observing that $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable, we see that on the one hand there is quite a gap between $\mathbb{Q}$ and $\mathbb{R}$, while on the other hand $\mathbb{R}$ is not really adequately supplied with roots—the polynomial $x^2 + 1$, has no roots in $\mathbb{R}$.

Wanting to describe the roots of the polynomials from the ring $\mathbf{F}[x]$, we see that we might well consider fields **K** that extend **F**. There is an unlimited supply of these. The principle of parsimony leads us to look for the ones that are some way or another close to **F** but still rich enough to allow us to have a full complement of roots of $f(x)$ or, what is the same, to be able to factor $f(x)$ into a product of polynomials of degree 1.

There are two key insights that are the starting point of our efforts. The first is that if **F** is a subfield of **K**, then **K** can be construed as a vector space over **F**. This allows us to use one of the most well-understood and thoroughly developed branchs of mathematics, the theory of vector spaces. We might even hope that the most interesting extension fields **K** will turn out to be finite dimensional over **F**. We use $[\mathbf{K} : \mathbf{F}]$ to denote the dimension of **K** as a vector space over **F**. We also refer to this dimension as the **degree** of the extension. It may be an infinite cardinal number.

The second insight is that, when **K** has a full complement of roots of $f(x)$, then every automorphism of **K** that has all the coefficients of $f(x)$ as fixed points must permute the roots of $f(x)$ that are in **K**. The set of roots of $f(x)$ is a finite subset of K. So we see emerging a finite subgroup of the concrete group of all permutations of this set of roots. This allows us to bring in the theory of (finite) groups.

So we see our enterprise as a mixture of ring theory (to understand the rings like $\mathbf{F}[x]$ and $\mathbf{K}[x]$),

the theory of fields, the theory of vector spaces, and group theory.

The first step we will take is to lay our hands on the minimal extension **K** of **F** that has a complete set of roots of $f(x)$.

We say that $f(x)$ **splits** over the field **K** provided $f(x) \in K[x]$ and the irreducible factors of $f(x)$ in $K[x]$ all have degree 1. We start looking for at least one root.

Kronecker's Theorem, 1882. *Let* **F** *be a field and let* $f(x) \in \mathbf{F}[x]$ *be irreducible. Then there is a field* **K** *extending* **F** *such that* $f(x)$ *has a root* $s \in K$ *and if* **L** *is any field extending* **F** *such that* $f(x)$ *has a root* $r \in L$*, then there is an embedding of* **K** *into* **L** *that fixes each element of* F *and sends* s *to* t*. Moreover, the dimension of* **K** *as a vector space over* **F** *is the degree of* $f(x)$*.*

Proof. Because $\mathbf{F}[x]$ is a principal ideal domain we know that irreducible and prime elements coincide and that so do the prime ideals and the maximal ideals. So $(f(x))$ is a maximal ideal of $\mathbf{F}[x]$. Consequently, $\mathbf{F}[x]/(f(x))$ is a field. Essentially, this is the field we desire and the element $x/(f(x)$ is the root s. A bothersome point is that it does not actually extend the field **F**, but rather has a subfield (with underlying set equal to $\{a/(f(x)) \mid a \in F\}$) easily seen to be isomorphic to **F**. So one must do some set theoretic surgery, snipping out the isomorphic copy and stitching in its place the field **F** itself. The result is the field **K**.

Now let **L** be any field extending **F** that has a root r of $f(x)$. We know that we can map $\mathbf{F}[x]$ into **K** via a homomorphism Ψ that extends the identity map on **F** and so that $\Psi(x) = r$. We see that $\Psi(f(x)) = f(r) = 0$ since Ψ is a homomorphism and r is a root of $f(x)$ in **L**. This means that $f(x) \in \ker\Psi$. On the other hand, if $g(x) \in \mathbf{F}[x]$ and $g(r) = 0$ in **L**, then $f(x)$ and $g(x)$ have a common fact $x - r$ in $\mathbf{L}[x]$. So they are not relatively prime. This means they cannot be relatively prime in $\mathbf{F}[x]$ either. Since $f(x)$ is prime in $\mathbf{F}[x]$ we find that $f(x) \mid g(x)$. Hence every polynomial belonging to the kernel of Ψ is a multiple of $f(x)$. So $(f(x)) = \ker\Psi$. This means, according to the Homomorphism Theorem, that $\mathbf{F}[x]/(f(x)) \cong \mathbf{L}'$ where $\mathbf{L}'$ is the image of $\mathbf{F}[x]$ under Ψ. But this means that $\mathbf{K} \cong \mathbf{L}'$ (so $\mathbf{L}'$ is actually a field) and we see that **K** embeds into **L** by a map that fixes each element of F and send s to r.

Finally, suppose s is a root of $f(x)$ in K and suppose that n is the degree of $f(x)$. Let Φ be a homomorphism from $\mathbf{F}[x]$ onto **K** that fixes every element of F and maps x to s. So every element of **K** is the image of some $h(x) \in F[x]$ under Φ. But in $\mathbf{F}[x]$ we can pick (uniquely) polynomials $q(x)$ and $r(x)$ so that $h(x) = q(x)f(x) + r(x)$ such that either $r(x)$ is the zero polynomial or else the degree of $r(x)$ is strictly less than the degree of $f(x) = n$. So we find $h(s) = q(s)f(s) + r(s) = r(s)$. But $r(s)$ is a linear combination with scalars from F of $\{1, s, \ldots, s^{n-1}\}$. So the latter set spans **K** as a vector space over **F**. But our contention is that this set is also linearly independent. Were it otherwise, we would have a nontrivial linear combination of these that would be 0. This would give us a nonzero polynomial $g(x)$ in $\mathbf{F}[x]$ that has s as a root. So we would see that $f(x)$ and $g(x)$ are not relatively prime. But $f(x)$ is prime (in $\mathbf{F}[x]$) and so $f(x) \mid g(x)$ in $\mathbf{F}[x]$, which is impossible since $g(x)$ is a nonzero polynomial of degree strictly less than the degree of $f(x)$. So the degree of $f(x)$ is the dimension of **K** as a vector space of **F**. □

There is a bit more mileage to be had from the proof of Kronecker's Theorem. Let **K** be a field extending the field **F**. We say that an element $s \in K$ is **algebraic** over **F** provided s is a root of some polynomial of positive degree from $\mathbf{F}[x]$. Of course, since every such polynomial can be factored into irreducible polynomials, and since **K** is an integral domain, we must have that every algebraic element of K actually is the root on an irreducible monic polynomial from $\mathbf{F}[x]$. This monic irreducible polynomial is called the **minimal polynomial** of s. An element of K that is not

algebraic over **F** is said to be **transcendental** over **F**. So here are some corollaries of our proof of Kronecker's Theorem.

Corollary 8.0.1. *Let the field* **K** *extend the field* **F** *and let* $s \in K$ *be algebraic over* **F**. *Then the smallest subring of* **K** *that includes* $F \cup \{s\}$ *is, in fact, a subfield of* **K**.

In general, when **K** is a field extending the field **F** and $s \in K$, we use the notation $\mathbf{F}[s]$ for the subring of **K** generated by $F \cup \{s\}$ and the notation $\mathbf{F}(s)$ for the subfield of **K** generated by $F \cup \{s\}$. The corollary above says that if s is algebraic over **F**, then $\mathbf{F}(s) = \mathbf{F}[s]$.

Corollary 8.0.2. *Let the field* **K** *extend the field* **F** *and let* $s \in K$ *be algebraic over* **F**. *Then every element of* $\mathbf{F}[s]$ *is algebraic of* **F**.

In general, we say that **K** is an **algebraic extension** of **F** provided every element of K is the root of some polynomial in $\mathbf{F}[x]$ that has positive degree. So this corollary asserts that if $s \in K$ is algebraic over **F**, then $\mathbf{F}[s]$ is an algebraic extension of **F**. If **K** is not an algebraic extension of **F** we call it a **transcendental extension** of **F**.

Now the field **K** given to us in Kronecker's Theorem provides us with an essentially unique extension of **F** that contains at least one root r of our irreducible polynomial $f(x) \in \mathbf{F}[x]$. So in $\mathbf{K}[x]$ we can factor $f(x)$ at least a little bit: there is $q(x) \in \mathbf{K}[x]$ so that $f(x) = (x-r)q(x)$. But we are not assured that $q(x)$, which still might have large degree, can be factored any further. So while **K** has at least one root of $f(x)$ it may not have a full complements of roots. Of course, the remedy is obvious. The degree of $q(x)$ is smaller than the degree of $f(x)$, so first we factor $q(x)$ into irreducibles in $\mathbf{K}[x]$ and then we invoke Kronecker on each of these, doing this again and again until some field **L** is reached in which $f(x)$ splits. While each step in this envisioned construction yields an essentially unique way to get to the next field extension, there are lots of arbitrary choices that have to be made along the way. The question of which irreducible polynomial to address next can be resolved at any stage in a number of ways. So maybe there are lots of different fields like **L** in which $f(x)$ splits, with any one of them as minimal as it can be. Fortunately, the whole business works out better than that.

Let **F** be a field and let $\mathscr{S}$ be a collection of of polynomials of positive degree, all drawn from $\mathbf{F}[x]$. A field **K** that extends **F** is said to be a **splitting field** of $\mathscr{S}$ over **F** provided

- in $\mathbf{K}[x]$ every polynomial in $\mathscr{S}$ factors as a product of polynomials of degree 1, and
- **K** is generated by $F \cup \{r \mid r \in K$ and r is a root of some polynomial in $\mathscr{S}\}$.

We say that **K** is a splitting field of $f(x)$ over **F** instead of that **K** is a splitting field of $\{f(x)\}$ over **F**. Then the step-by-step, recursive extension of Kronecker's Theorem outlined above gives us

Corollary 8.0.3. *Let* **F** *be a field and* $f(x)$ *be a polynomial of positive degree that belongs to* $F[x]$. *Then* $f(x)$ *has a splitting field over* **F**.

We would like to see that the splitting field is essentially unique, that it is an algebraic extension of **F**, and that it is finite dimensional as a vector space over **F** (and even more, we would like to lay hands on this dimension).

The Dimension Formula. *Let the field* **L** *be an extension of the field* **K** *that is in turn an extension of the field* **F**. *Then* $[\mathbf{L}:\mathbf{F}] = [\mathbf{L}:\mathbf{K}][\mathbf{K}:\mathbf{F}]$.

Proof. Let B be a basis for the vector space **K** over the field **F** and let C be a basis for the vector space **L** over the field **K**. Put

$$BC := \{bc \mid b \in B \text{ and } c \in C\}.$$

Our contention is that the set BC is a basis for the vector space **L** over the field **F** and that $|BC| = |B||C|$. So we must prove that BC spans **L** (with scalars chosen from F), that BC is linearly independent, and that there is a one-to-one correspondence between $B \times C$ and BC.

Contention. BC spans **L** as a vector space over **F**.

Let $w \in L$. Since C spans L over **K**, pick $c_0, c_1, \ldots, c_{n-1} \in C$ and $d_0, d_1, \ldots, d_{n-1} \in K$ so that

$$w = \sum_{i<n} d_i c_i.$$

Now consider any $i < n$. We have $d_i \in K$. Since B spans K over **F**, pick $b_{i,0}, bi, 1, \ldots . b_{i,m_i-1} \in B$ and $a_{i,0}, a_{i,}, \ldots, a_{i,m_i-1} \in F$ so that

$$d_i = \sum_{j<m_i} a_{i,j} b_{i,j}.$$

Putting these two pieces together, we get

$$w = \sum_{i<n} d_i c_i = \sum_{i<n} \left(\sum_{j<m_i} a_{i,j} b_{i,j} \right) c_i.$$

In this way, we see

$$w = \sum_{i<n, j<m_i} a_{i,j} (b_{i,j} c_i),$$

which is a linear combination of elements of BC using scalars from F.

Contention. The set BC is a linearly independent subset of the vector space **L** over the field **F**.

Let us suppose that $a_0, a_1, \ldots, a_{n-1} \in F$, $b_0, b_1, \ldots, b_{n-1} \in B$, and $c_0, c_1, \ldots, c_{n-1} \in C$ have been chosen so that

$$\sum_{i<n} a_i (b_i c_i) = 0$$

and that $b_0 c_0, b_1 c_1, \ldots, b_{n-1} c_{n-1}$ are distinct. Now perhaps not all the c_i's are distinct. However, by rearranging the indices we may suppose that $c_0, \ldots, c_{\ell-1}$ are all distinct but that any c with a later index is equal to one of these ℓ disitinct c's. For each $k < \ell$ we put $I_k = \{i \mid c_i = c_k\}$. Then we can reorganize the sum above as

$$\begin{aligned} 0 &= \sum_{i \in I_0} (a_i b_i) c_0 + \sum_{i \in I_1} (a_i b_i) c_1 + \cdots + \sum_{i \in I_{\ell-1}} (a_i b_i) c_{\ell-1} \\ &= \left(\sum_{i \in I_0} a_i b_i \right) c_0 + \left(\sum_{i \in I_1} a_i b_i \right) c_1 + \cdots + \left(\sum_{i \in I_{\ell-1}} a_i b_i \right) c_{\ell-1}. \end{aligned}$$

Because C is linearly independent (over **K**) and because $c_0, \ldots, c_{\ell-1}$ are distinct, we find, for each $k < \ell$, that

$$0 = \sum_{i \in I_k} a_i b_i.$$

Now suppose $i, j \in I_k$ and $i \neq j$. So we know that $b_i c_i \neq b_j c_j$ but also that $c_i = c_j = c_k \neq 0$, with the last $\neq$ following because 0 cannot be in any linearly independent set like C. So we see that $b_i c_i \neq b_j c_i$ and $c_i \neq 0$. Dividing away the c_i, we conclude that $b_i \neq b_j$. This means that the b_i's occurring in $0 = \sum_{i \in I_k} a_i b_i$ are all distinct. Since B is linearly independent (over **F**) we find that $a_i = 0$ for all $i \in I_k$ and for all $k < \ell$. This means that $a_i = 0$ for all $i < n$, and the set BC is linearly independent, as desired.

Contention. The map from $B \times C$ to BC that sends (b, c) to bc for all $(b, c) \in B \times C$ is a one-to-one correspondence.

According to the definition of BC, this map is onto BC. So it remains to show that it is one-to-one. So pick $b_0, b_1 \in B$ and $c_0, c_1 \in C$ so that $b_0 c_0 = b_1 c_1$. We need to show that $b_0 = b_1$ and $c_0 = c_1$. We note that none of b_0, b_1, c_0, and c_1 can be 0 since 0 belongs to no linearly independent set. There are two cases: etiher $c_0 = c_1$ or else $c_0 \neq c_1$. In the first case we can cancel the c's from $b_0 c_0 = b_1 c_1$ to obtain as well that $b_0 = b_1$, our desire. In the second case we see that $b_0 c_0 - b_1 c_1 = 0$. Since in this case c_0 and c_1 are distinct and linearly independent, we find that $b_0 = -b_1 = 0$, which we have already observed is impossible. So we must reject the second case.

This establishes the Dimension Formula. □

Notice that since the product on any two infinite cardinals is always an infinite cardinal (in fact, the larger of the two), we see that in the Dimension Formula, $[\mathbf{L} : \mathbf{F}]$ is infinite if and only if at least one of $[\mathbf{L} : \mathbf{K}]$ and $[\mathbf{K} : \mathbf{F}]$ is infinite.

The following fact will be useful in our ensuing work.

Fact. Let the field **K** extend the field **F** and suppose that $[\mathbf{K} : \mathbf{F}]$ is finite. Then **K** is an algebraic extension of **F**.

Proof. Let $s \in K$. Since **K** is a finite dimensional vector space over **F** it cannot happen that all the elements on the list below are distinct and linearly independent:

$$1, s, s^2, s^3, s^4, \ldots.$$

This means that there must be elements $a_0, a_1, \ldots, a_n \in F$ that are not all 0 such that

$$a_0 + a_1 s^1 + a_2 s^2 + \cdots + a_n s^n = 0$$

It does no harm to suppose that $a_n \neq 0$. Notice that $n \neq 0$. So we see that s is a root of the polynomial $a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \in F[x]$ of positive degree. Therefore s is algebraic over **F**. □

Exrtension like the on in this Fact are called **finite extensions**. Another way to frame the Fact is "Finite extensions are algebraic extensions".

The Algebraic Extension Theorem. *Let the field* **L** *be an algebraic extension of the field* **K** *and let* **K** *be an algebraic extension of the field* **F**. *Then* **L** *is an algebraic extension of* **F**.

Proof. Let $s \in L$. Since $\mathbf{L}$ is an algebraic extension of $\mathbf{K}$, we pick $c_0, c_1, \ldots, c_n \in K$ with $n > 0$ and $c_n \neq 0$ such that s is a root of $c_0 + c_1 x + \cdots + c_n x^n$. Now let

$$\begin{aligned} \mathbf{K}_0 &= \mathbf{F}[c_0] \\ \mathbf{K}_1 &= \mathbf{K}_0[c_1] \\ &\vdots \\ \mathbf{K}_n &= \mathbf{K}_{n-1}[c_n]. \end{aligned}$$

Using the Dimension Formula repeatedly we find

$$[\mathbf{K}_n : \mathbf{F}] = [\mathbf{K}_n : \mathbf{K}_{n-1}][\mathbf{K}_{n-1} : \mathbf{K}_{n-2}] \cdots [\mathbf{K}_1 : \mathbf{K}_0].$$

But we know each of the dimensions on the right is finite. So $[\mathbf{K}_n : \mathbf{F}]$ is finite. But s is a root of a polynomial in $\mathbf{K}_n[x]$. So $\mathbf{K}_n[s] : \mathbf{K}_n]$ is finite. Invoking the Dimension Formula one more time yields that $[\mathbf{K}_n[s] : \mathbf{F}]$ is finite. Since $s \in K_n[s]$, we see that s is algebraic over $\mathbf{F}$, which is just what we want. □

There is one more thing to do. Tackle the uniqueness of splitting fields.

The Basic Fact About Extending Isomorphisms. *Let* $\mathbf{F}$ *and* $\mathbf{F}^*$ *be fields. Let* Φ *be an isomorphism from* $\mathbf{F}$ *onto* $\mathbf{F}^*$. *Let* $\mathbf{L}$ *be a field extending* $\mathbf{F}$ *and let* $\mathbf{L}^*$ *be a field extending* $\mathbf{F}^*$. *Let* $s \in L$ *be algebraic over* $\mathbf{F}$ *with minimimal polynomial* $f(x)$. *Then* Φ *can be extended to an embedding of* $\mathbf{F}[s]$ *into* $\mathbf{L}^*$ *if and only if* $f^*(x)$ *has a root in* $\mathbf{L}^*$, *in which case the number of distinct extensions of* Φ *is the same as the number of distinct roots of* $f^*(x)$ *in* $\mathbf{L}^*$. *(Here* $f^*(x) \in F*[x]$ *is obtained from* $f(x)$ *by applying* Φ *to each of its coefficients.)*

Proof. Suppose first that Φ has an extension Ψ. Applying Ψ to the equation $f(r) = 0$, which is true in $\mathbf{L}$, gives $f^*(\Psi(s)) = 0$, which is true in $\mathbf{L}^*$. Hence $\Psi(s) \in L^*$ is a root of $f^*(x)$.

On the hand, let r be any root of $f^*(x)$ that belongs to L^*. We apply Kronecker's Theorem twice. So we have isomorphisms

$$\begin{aligned} \Lambda : \mathbf{F}[x]/(f(x)) &\twoheadrightarrow \mathbf{F}[s] \text{ with } \Lambda(x/(f(x))) = s \\ \Theta : \mathbf{F}^*[x]/(f^*(x)) &\twoheadrightarrow \mathbf{F}^*[r] \text{ with } \Theta(x/(f^*(x))) = r. \end{aligned}$$

But the isomorphism Φ from $\mathbf{F}$ to $\mathbf{F}^*$ induces an isomorphism $\Pi : \mathbf{F}[x]/(f(x)) \twoheadrightarrow \mathbf{F}^*[x]/(f^*(x))$. Putting things together we get an isomorphism $\Theta \circ \Pi \circ \Lambda^{-1}$ from $\mathbf{F}[s]$ onto $\mathbf{F}[r]$ that send s to r and extends Φ. Since every embedding from $\mathbf{F}[s]$ into $\mathbf{L}^*$ that extends Φ is determined by what image it gives to s, we see that there are precisely as many extensions of Φ as there are distinct roots of $f^*(x)$ in L^*. □

Existence and Uniqueness of Splitting Fields. *Let* $\mathbf{F}$ *be a field and* $f(x) \in F[x]$ *be a polynomial of degree* $n > 0$. *Then there is a field* $\mathbf{E}$ *extending* $\mathbf{F}$ *such that*

(a) $\mathbf{E}$ *is a splitting field of* $f(x)$ *over* $\mathbf{F}$,

(b) $[\mathbf{E} : \mathbf{F}] \leq n!$, *and*

Moreover, suppose that $\mathbf{E}$ *and* $\mathbf{E}^*$ *are splitting fields of* $f(x)$ *over* $\mathbf{F}$. *Then*

(c) $\mathbf{E}$ *and* $\mathbf{E}^*$ *are isomorphic via an isomorphism that fixes each element of* F*, and*

(d) *The number of isomorphisms for* $\mathbf{E}$ *onto* $\mathbf{E}^*$ *that fix each element of* F *is no greater than* $[\mathbf{E}:\mathbf{F}]$ *and it is equal to* $[\mathbf{E}:\mathbf{F}]$ *if* $f(x)$ *has* n *distinct roots in* E*.*

Proof. Let us prove the existence part by induction of n. The base step is that $f(x) = ax+b$ where $a, b \in F$ and $a \neq 0$. So $\frac{b}{a} \in F$ is a root of $f(x)$. So we take $\mathbf{E} = \mathbf{F}$.

For the induction step we take $f(x)$ to be a polynomial of degree $k+1$ and we assume the (existence parts of) the theorem *over arbitrary fields* for $n < k+1$. Let $p(x) \in F[x]$ be an irreducible factor of $f(x)$. According to Kronecker's Theorem there is an extension $\mathbf{E}$ of $\mathbf{F}$ and an $s \in K$ so that s is a root of $f(x)$. So in $\mathbf{F}[s]$ we can factor $f(x) = (x-s)g(x)$ where $g(x) \in F[s][x]$ has degree k. Using the induction hypothesis we obtain a splitting field $\mathbf{E}$ of $g(x)$ over $\mathbf{F}[s]$ such that $[\mathbf{E}:\mathbf{F}[s]] \leq k!$.

In $\mathbf{E}$ we see that $f(x)$ factors into a product of polynomials of degree one and that the roots of $f(x)$ in E consist of the element s and the roots of $g(x)$. Because $\mathbf{E}$ is a splitting field of $g(x)$ over $\mathbf{F}[s]$ we know that it is generated by $F[s] \cup R$, where R is the set of all roots of $g(x)$ in E. But $F[s]$ is generated by $F \cup \{s\}$. So $F \cup \{s\} \cup R$ generates $\mathbf{E}$. In this way we see that $\mathbf{E}$ is a splitting field of $f(x)$ over $\mathbf{F}$. So condition (a) is met in the inductive step.

We know, by Kronecker, that $[\mathbf{F}[s]:\mathbf{F}]$ is the degree of the irreducible factor $p(x)$ of $f(x)$. So the degree of $p(x) \leq k+1$, which is the degree of $f(x)$. By the Dimension Formula, we see

$$\begin{aligned}[\mathbf{E}:\mathbf{F}] &= [\mathbf{E}:\mathbf{F}[s]][\mathbf{F}[s]:\mathbf{F}] \\ &\leq k!(k+1) = (k+1)!\end{aligned}$$

So condition (b) is met in the inductive step.

For the rest, it proves more convenient to prove something a bit stronger.

Let $\mathbf{F}^*$ be a field and $\Phi : \mathbf{F} \rightarrowtail\!\!\!\!\!\rightarrow \mathbf{F}^*$. Let $f^*(x)$ be the polynomial over $\mathbf{F}^*$ obtained from $f(x)$ by applying Φ to each coefficient.

Now suppose $\mathbf{E}$ and $\mathbf{E}^*$ are splitting fields of $f(x)$ over $\mathbf{F}$ and of $f^*(x)$ over $\mathbf{F}^*$ respectively. Instead of considering maps the fix each element of F (i.e. those extending the identity map on F) we consider maps extending Φ.

We proceed by induction on $[\mathbf{E}:\mathbf{F}]$.

For the base step, we will have $F = E$ and $f(x)$ factors into polynomials of degree 1 over $\mathbf{F}$. So all the roots of $f(x)$ lie in F. Likewise for $f^*(x) \in \mathbf{F}^*[x]$. Since the roots of $f(x)$ together with F itself generate $\mathbf{E}$ and likewise for $f^*(x)$ and $\mathbf{E}^*$, we see that $\mathbf{E} = \mathbf{F} \overset{\Phi}{\rightarrowtail\!\!\!\!\!\rightarrow} \mathbf{F}^* = \mathbf{E}^*$ and there is only one isomorphism between $\mathbf{E}$ and $\mathbf{E}^*$ that extends Φ, namely the map Φ itself. So the appropriately modiifications of conditions (c) and (d) both hold in the base step.

Now we turn to the inductive step, where we have $[\mathbf{E}:\mathbf{F}] > 1$. So there must be a root $r \in E$ that does not belong to F. Let $p(x)$ be the minimal polynomial of r_0. The degree of $p(x)$ must be at least 2. Since we see that $p(x)$ is a factor of $f(x)$ so the corresponding $p^*(x)$ is a factor of $f^*(x)$ and $p^*(x)$ must split in $\mathbf{E}^*$. Let us say that it has $m > 1$ roots in $\mathbf{E}^*$. Then by our basic fact about extending isomorphisms there are exactly m distinct extensions of Φ to embeddings of $\mathbf{F}[r]$ into $\mathbf{E}^*$. Consider one of them Φ' and let $s = \Phi'(r)$. Now $[\mathbf{E}:\mathbf{F}] = [\mathbf{E}:\mathbf{F}[r]][\mathbf{F}[r]:\mathbf{F}]$ since $[\mathbf{F}[r]:\mathbf{F}]$ is the degree of $p(x)$, which is at least 2, we see that $[\mathbf{F}:\mathbf{F}[r]] < [\mathbf{E}:\mathbf{F}]$. But $\mathbf{E}$ is a splitting field of $f(x)$ over $\mathbf{F}[r]$ and $\mathbf{E}^*$ is a splitting field of $f^*(x)$ over $\mathbf{F}*[s]]$. So we can appeal to the induction hypothesis to get at least one extension of Φ' to an isomorphism between $\mathbf{E}$ and $\mathbf{E}^*$. Evidently, such an

isomorphism also extends Φ and we obtain, in the inductive step, the appropriate modification of condition (c).

The induction hypothesis also tells us that the number of such extensions of Φ' is not greater than $[\mathbf{E} : \mathbf{F}[r]]$ and is equal to $[\mathbf{E} : \mathbf{F}[r]]$ if the number of distinct roots of $f(x)$ in E coincides with the degree of $f(x)$. Recall that Φ' was one of the m extensions of Φ that embed $\mathbf{F}[r]$ into $\mathbf{E}^*$. So The number of extension of Φ to isomorphism between $\mathbf{E}$ and $\mathbf{E}^*$ is no greater than the product of $[\mathbf{E} : \mathbf{F}[r]]m$. But m, the number of distinct roots of $p^*(x)$ in $\mathbf{E}^*$ can be no greater than the degree of $p(x)$, which we know is $[\mathbf{F}[r] : \mathbf{F}]$. So $[\mathbf{E} : \mathbf{F}] = [\mathbf{E} : \mathbf{F}[r]][\mathbf{F}[r] : \mathbf{F}]$ is an upper bound on the number of ways Φ can be extended to an isomorphism between $\mathbf{E}$ and $\mathbf{E}^*$. Last suppose that $f(x)$ has distinct roots. Then so must $p(x)$. This means that $m = [\mathbf{F}[r] : \mathbf{F}]$. In this case we know there are precisely $[\mathbf{F}[r] : \mathbf{F}]$ ways to extend Φ to an embedding of $\mathbf{F}[r]$ into $\mathbf{E}^*$ and, for each such extension, there are precisely $[\mathbf{E} : \mathbf{F}[r]]$ ways to extend it to an isomorphism between $\mathbf{E}$ and $\mathbf{E}^*$. So the number of extensions of Φ to an isomorphism from $\mathbf{E}$ onto $\mathbf{E}^*$ is precisely $[\mathbf{E} : \mathbf{F}]$ by the Dimension Formula. So condition (d) holds in the inductive step.

The proof is complete. □

LECTURE 9

ALGEBRAICALLY CLOSED FIELDS

A field **F** is said to be **algebraically closed** provided irreducible polynomials in $\mathbf{F}[x]$ coincide with the polynomials of degree 1. This is the same as saying that every polynomial in $\mathbf{F}[x]$ of positive degree has a root in **F**. It is also evidently equivalent to the requirement at **F** has no proper algebraic extensions.

Neither the field $\mathbb{Q}$ of rational numbers nor the field $\mathbb{R}$ of real numbers is algebraically closed. It is a nontrivial fact (which we will prove later in the semester) that the field $\mathbb{C}$ of complex numbers is algebraically closed.

An extension **K** of the field **F** is an **algebraic closure** of **F** provided **K** is an algebraically closed algebraic extension of **F**.

Consider how we might arrive at an algebraic closure of the field $\mathbb{Q}$ of rational numbers. We might begin by making a list of all the polynomials of positive degree in $\mathbb{Q}[x]$. This list is countably infinite and it takes a bit of work to arrange these polynomials like the natural numbers are arranged. But image we have made this list: $f_0(x), f_1(x), f_2(x), \ldots$. Now we could proceed by letting $\mathbf{F}_0$ be the splitting field of $f_0(x)$ over $\mathbb{Q}$. Next, we let $\mathbf{F}_1$ be the splitting field of $f_1(x)$ over $\mathbf{F}_0$. We continue in this way to split, one after another, all the polynomials on our list. We get in this way a chain of fields, each extending the one before. Fearlessly, we form the union of this chain of fields to arrive at $\mathbf{K}_0$. A little thought shows us that $\mathbf{K}_0$ is an algebraic extension of $\mathbb{Q}$, that it is generated over $\mathbb{Q}$ by the roots of all those polynomials, and that all those polynomials split in $\mathbf{K}_0$. Unfortunately, along the way we have added a lot of new elements and these new elements can be coefficients of polynomials in $\mathbf{K}_0[x]$ that haven't yet been addressed. So now we must list all of these polynomials, build another infinite chain of splitting fields, and finally arrive at the union $\mathbf{K}_1$. Now many more polynomials have been split but many more new elements have also been introduced. But we continue anyhow to construct $\mathbf{K}_2$, then $\mathbf{K}_3, \ldots$. Finally, we take one last union of this chain to obtain the field $\mathbb{A}$. We would be able to show that $\mathbb{A}$ is an algebraically closed algebraic extension of $\mathbb{Q}$ and even that $\mathbb{A}$ is countably infinite. The idea behind this proof sketch could be made to work starting with any field (although the countability of the algebraic closure has to be modified if the field we start with is uncountable). In general, this construction requires making a lot of choices along the way (particularly, choices about how to order the polynomials

at each step).

We can avoid the complexity of this construction by invoking Zorn's Lemma.

The Existence Theorem for Algebraic Closures. *Every field has an algebraic closure.*

Proof. The basic idea is that we will take $\mathscr{F}$ to be the collection of all algebraic extensions of the given field **F**. Using Zorn's Lemma we will extract a maximal member of $\mathscr{F}$ that will turn out to be an algebraic extension of **F** that is algebraically closed. There is one stumbling block to this scheme: the collection $\mathscr{F}$ turns out to be too large and wild to be a set.

Let **F** be a field. Let U be an uncountably infinite set of cardinality properly larger than the cardinality of F so that $F \subseteq U$. Take $\mathscr{F}$ to be the collection of all algebraic extensions **K** of **F** so that $K \subseteq U$.

To invoke Zorn's Lemma, consider any nonempty chain $\mathscr{C} \subseteq \mathscr{F}$. We contend $\mathscr{C}$ has an upper bound in $\mathscr{F}$. Indeed, let $L = \bigcup\{K \mid \mathbf{K} \in \mathscr{C}\}$. Of course $0, 1 \in L$. We impose $+$ and $\cdot$ on L is the natural way: for $a, b \in L$, using the fact that $\mathscr{C}$ is a chain, pick $\mathbf{K} \in \mathscr{C}$ so that $a, b \in K$. Take $a + b$ and $a \cdot b$ in **L** just as ther are understood in **K**. (The hard-working graduate students should confirm that the particular **K** chosen works no bad idiosyncratic influence here.) This, of course, is just another case of the union of a chain of algebraic entities resulting in another entity of the same kind. It is straightforward to provide the details showing that **L** is a field that extends **F** and of course $l \subseteq U$. It is also clear the **L** is an upper bound of $\mathscr{C}$. To conclude that $\mathbf{L} \in \mathscr{F}$, we need to show that **L** is also an algebraic extension of **F**. But this is clear: let $a \in L$ and pick $\mathbf{K} \in \mathscr{C} \subseteq \mathscr{F}$ so that $a \in K$. Since $\mathbf{K} \in \mathscr{F}$ it is an algebraic extension of **F**. So a is a root of a polynomial in $\mathbf{F}[x]$. That is, a is algebraic over **F**, as desired.

So Zorn's Lemma provides us with a field **M** that is a maximal element of $\mathscr{F}$. In particular, **M** is an algebraic extension of **F**. Now the idea is to take any irreducible polynomial $p(x) \in \mathbf{M}[x]$. Applying Kronecker's Theorem, we obtain an algebraic extension $\mathbf{M}[r]$ of **M** so that r is a root of $p(x)$. We know that an algebraic extension of **M** must be an algebraic extension of **F**, since **M** is an algebraic extension of **F**. Were we able to appeal to be maximality of **M**, we would conclude that $M = M[r]$, so that $r \in M$ and the arbitrary irreducible polynomial $p(x)$ has a root in **M**. Thus **M** would be algebraically closed.

The point of difficulty is that $M[r]$ might not be contained in U.

This would present no trouble if there were enough room in $U \setminus M$ to fit in a copy of $M[r] \setminus M$. So what is the size of $M \setminus M[r]$? Well, we know that $\mathbf{M}[r]$ is a vector space over **M** with dimension equal to the degree d of the minimal polynomial of r. So, as with any finite dimensional vector space, we find $|M[r]| = |M|^d$. So we see that $|M[r] \setminus M| \leq |M[r]| = |M|^d$. We could argue (maybe the curious graduate student will do it) that M must be infinite. We take $\kappa = |M|$ if M is infinite (as it is) and otherwise take κ to be the smallest infinite cardinal (namely $|\mathbb{N}|$. One useful fact from the arithmetic of infinite cardinals is that $\kappa \cdot \kappa = \kappa$. So a touch of induction shows that $|M[r] \setminus M| \leq \kappa$.

How big is $|U \setminus M|$? Every element of M is a root of some irreducible polynomial in $\mathbf{F}[x]$ and each such polynomial has only finitely many roots. How many polynomials are there. The zero polynomial together with the polynomials of degree 0 make up F. So $|F| = |F|^1$ is an upper bound on this collection. The polynomials of degree 1 each have two coefficients. So $|F| \cdot |F| = |F|^2$ is an upper bound on the number of these polynomials. In general, $|F|^d$ bounds the number of polynomials of degree d. Now each polynomial of degree d can have at most d distinct roots in M. Altogether, we set that

$$\sum_{1 \leq d < \omega} d|F|^d$$

is an upper bound on the number of elements of M. Let $\mu = |F|$ if F is infinite and let μ be the least infinite cardinal ω otherwise. Then

$$\sum_{1\le d<\omega} d|F|^d \le \sum_{1\le d<\omega} d\mu^d = \sum_{1\le d<\omega} \mu \le \omega\cdot\mu = \mu.$$

Our choice of the size of U at the beginning of the proof ensures that $|U| > \mu$ and hence that $|U \setminus M| = |U| > \mu \ge \kappa \ge |M[r] \setminus M|$.

This means that there is enough room left over in U, after M is in hand, to construct a copy of $\mathbf{M}[r]$.

Now we can really appeal to the maximality of $\mathbf{M}$ to complete the proof. □

The Uniqueness Theorem for Algebraic Closures. *Let* $\mathbf{F}$ *be a field and let* $\mathbf{A}$ *and* $\mathbf{K}$ *be algebraic extensions of* $\mathbf{F}$ *which are algebraically closed. Then there is an isomorphism from* $\mathbf{A}$ *onto* $\mathbf{K}$ *which fixes each element of* $\mathbf{F}$.

Proof. Let $\mathscr{I}$ be the set of all isomorphisms with domains which are subfields of $\mathbf{A}$ that extend $\mathbf{F}$, whose images are subfields of $\mathbf{K}$ that extend $\mathbf{F}$, and which fix every element of $\mathbf{F}$.

Recalling that each function is a set of ordered pairs, we see that $\mathscr{I}$ is partially ordered by $\subseteq$. It is easy to see that this ordering is the same as the ordering by extension of functions.

To invoke Zorn's Lemma, we need to see that any chain $\mathscr{C}$ in $\mathscr{I}$ has an upper bound. If $\mathscr{C}$ is empty, then the identity function of $\mathbf{F}$ is an upper bound of $\mathscr{C}$ and it belongs to $\mathscr{I}$. Consider the case when $\mathscr{C}$ is not empty. Let $\Phi = \bigcup\mathscr{C}$.

Claim. Φ is a function.

Proof. Suppose $(a,b),(a,c) \in \Phi$. Pick $\varphi,\psi \in \mathscr{C}$ so that $(a,b) \in \varphi$ and $(a,c) \in \psi$. Since $\mathscr{C}$ is a chain, either $\varphi \subseteq \psi$ or $\psi \subseteq \varphi$. Without loss of generality, let us suppose that $\psi \subseteq \varphi$. Then $(a,b),(a,c) \in \varphi$. But φ is a function, so $b = c$. Consequently, Φ is a function. □

Claim. The domain of Φ is a subfield of $\mathbf{A}$ which extends $\mathbf{F}$.

Proof. A routine argument shows that $\operatorname{dom}\Phi = \bigcup\{\operatorname{dom}\varphi \mid \varphi \in \mathscr{C}\}$. Since $\mathscr{C}$ is not empty and $\mathbf{F} \subseteq \operatorname{dom}\varphi$ for each $\varphi \in \mathscr{C}$, we see that $\mathbf{F} \subseteq \operatorname{dom}\Phi$. Let $a,b \in \operatorname{dom}\Phi$. As above, pick $\varphi \in \mathscr{C}$ so that $a,b \in \operatorname{dom}\varphi$. Since $\operatorname{dom}\varphi$ is a subfield of $\mathbf{A}$, we see that $a+b, ab \in \operatorname{dom}\varphi \subseteq \operatorname{dom}\Phi$ and also that $a^{-1} \in \operatorname{dom}\Phi$ if $a \neq 0$. This means that $\operatorname{dom}\Phi$ is a subfield of $\mathbf{A}$. □

Claim. Φ is a homomorphism.

Proof. Let $a,b \in \operatorname{dom}\Phi$. Pick $\varphi,\psi \in \mathscr{C}$ so that $a \in \operatorname{dom}\varphi$ and $b \in \operatorname{dom}\psi$. As above, without loss of generality we suppose that $a,b \in \operatorname{dom}\varphi$. Now φ is a homomorphism, so $(a+b,\varphi(a)+\varphi(b))$ and $(ab,\varphi(a)\varphi(b))$ both belong to φ, as do $(a,\varphi(a))$ and $(b,\varphi(b))$. But $\varphi \subseteq \Phi$. So those four ordered pairs belong to the function Φ. Translated into usual usage we have

$$\Phi(a+b) = \varphi(a+b) = \varphi(a)+\varphi(b) = \Phi(a)+\Phi(b)$$
$$\Phi(ab) = \varphi(ab) = \varphi(a)\varphi(b) = \Phi(a)\Phi(b).$$

We see, even more easily, that $\Phi(0) = 0$ and $\Phi(1) = 1$. So Φ is a homomorphism. □

Claim. Φ fixes each element of $\mathbf{F}$.

Proof. This is too easy to prove. □

Since Φ is a homomorphism from one field into another and it fixes each element of the subfield **F**, we see it cannot collapse everything to one value. So it must be one-to-one (after all, fields are simple). So Φ is an isomorphism. All this, taken together, means that $\Phi \in \mathscr{I}$. So Φ is an upper bound of $\mathscr{C}$ as desired.

Invoking Zorn's Lemma, we see that $\mathscr{I}$ must have a maximal member. Let Ψ be such a maximal member. It remains to show that **A** is the domain of Ψ and that **K** is the image of Ψ.

The field **A** is an algebraic extension of $\operatorname{dom}\Psi$, because any element of **A** is a root of some polynomial of positive degree with coefficients in $\mathbf{F} \subseteq \operatorname{dom}\Psi$. Let $u \in \mathbf{A}$. Let $p(x) \in \operatorname{dom}\Psi[x]$ be an irreducible polynomial with root u. Say $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ with $a_n \neq 0$. Then the polynomial $\Psi(a_0) + \Psi(a_1)x + \cdots + \Psi(a_n)x^n$ is irreducible over the field that is the image of Ψ. Denote the image of Ψ by **B**. But **K** is algebraically closed, so this polynomial must have a root v in **K**. Consider the possibility that the degree of $p(x)$ is bigger than 1. In that event, $u \notin \operatorname{dom}\Psi$ and $v \notin \mathbf{B}$. Then we can extend Ψ to an isomorphism from $\operatorname{dom}\Psi[u]$ onto $\mathbf{B}[v]$. This extension also belongs to $\mathscr{I}$. In this way the maximality of Ψ is violated. So $p(x)$ must have degree 1. But this entails that $u \in \operatorname{dom}\Psi$. Since u was an arbitrary element of **A**, we see that $\mathbf{A} = \operatorname{dom}\Psi$.

It remains to see that the image **B** of Ψ is **K**. Since we have seen, by this point, that **B** is isomorphic to **A**, and we know that **A** is algebraically closed, we conclude that **B** is also algebraically closed. Now **K** is an algebraic extension of **B**. But algebraically closed fields cannot have proper algebraic extensions. So **K** is not a proper extension of **B**. This means $\mathbf{B} = \mathbf{K}$, concluding our proof. □

In one of the problem sets you will be asked to prove that no algebraically closed field can be finite. In the proof of existence of algebraic closure we was that the algebraic closure of a finite field in countably infinite. For any infinite field, our argument shows that the algebraic closure is the same cardinality as the original field. So the algebraic closure of the field of rational numbers is countably infinite.

The field of complex numbers turns out to be algebraically closed, a fact customarily referred to as the Fundamental Theorem of Algebra. Proofs of this were offered in the 18^{th} century, notably by Euler, Lagrange, and Laplace. These proofs all had gaps. Roughly speaking, these gaps are filled by Kronecker's Theorem. In his 1799 doctoral dissertation, Gauss devoted a lot of space to picking out the flaws in these proofs, and then supplied a flawed proof of his own. (This proof by Gauss had an ingenious geometric turn—the gap was finally filled by Ostrowski in 1920). The first complete proof was given in 1806 by Argand. Gauss later gave two further proofs.

The Fundamental Theorem of Algebra no longer plays a fundamental role in algebra. At some level, it is basically an analytical rather than an algebraic theorem, although we will give toward the end of these lectures a largely algebraic account. The quickest modern proofs appeal to theorems in complex analysis, for example to Liouville's Theorem.

As a consequence, the algebraic closure of the rationals can be construed as a subfield of the complex numbers.

LECTURE 10

CONSTRUCTIONS BY STRAIGHTEDGE AND COMPASS

Euclid's book, sadly now fallen from the mathematician's bookshelf, should properly still be the property of every mathematician. It is filled with theorems proved by means of constructions using straightedge and compass. Loosely speaking, these constructions started with a given finite configuration of points on the plane and then proceeded in a step-by-step fashion to construct further points. The new points could only arise in one of three ways:

(a) As the point of intersection of two line segments, each drawn with the help of the straightedge through two distinct points already at hand. Here the endpoints of the segments, and indeed almost all the points on the segment, need not be among the constructed points.

(b) As points of intersection between a line segment and a circle, where the line segment arises as above and the circle is drawn with the help of the compass by placing the the foot and the drawing points of the compass on points constructed at some prior step. Again the only points on the line segment and the circle that qualify as constructed are the point of intersection.

(c) As points of intersection between two circles, each circle drawn as described above.

Among the problems left as unsolved by the geometers of this classical period were the following:

The Trisection of Angles. *Given an arbitrary angle, to trisect it by means of straightedge and compass.*

The Duplication of the Cube or the Delian Problem. *Given an arbitrary cube, to construct a cube of twice the volume by means of straightedge and compass.*

A legend behind this problem concerns a serious plague. The advice of the great oracle of Apollo at Delphi was sought. The Altar of Apollo was an impressive cube. The oracle advised that Apollo would intercede once the altar had been exactly doubled in volume. Apollo never interceded.

Squaring the Circle. *Given a circle, to construct, by means of straightedge and compass, a square with the area.*

The Construction of Regular Polygons. *Given a line segment construct, by means of straightedge and compass, a regular polygon of n sides each of length the same as the given line segment.*

We are now in a position to present the solutions to most of these problems.

If we identify Euclid's plane with $\mathbb{R} \times \mathbb{R}$, we can convert these geometric problems into algebraic problems. To set a unit length, we start our analysis with two points $(0,0)$ and $(1,0)$. Let $\mathscr{C}$ be the totality of all points that can be constructed by straightedge and compass from these first two points. Since at any stage only finitely many new points are constructed and since any constructible point is reached after some finite sequence of steps, we see that $\mathscr{C}$ is countable. We say a real number r is **constructible** provided r is one of the coordinates of a point that belongs to $\mathscr{C}$. It is an informative exercise in straightedge and compass construction to show that r is constructible if and only if $|r|$ is the length of a line segment joining to constructible points (including degenerate segments of length 0). Let E be the set of constructible real numbers.

[The leading sentence of the last paragraph might well have given you pause. Is it really permissible to identify Euclid's plane with $\mathbb{R} \times \mathbb{R}$? What would you have to do the prove this statement?]

It is clear that $0, 1 \in E$. We will show next E is closed under addition and multiplication, and that every nonzero constructible real has a constructible multiplicative inverse. In this way, we will arrive at **E**, the field of constructible reals. Actually, **E** has very special properties. Let us say that a subfield **K** of $\mathbb{R}$ is **closed under the extraction of square roots** provided a real number r belongs to K whenever $r^2 \in K$. This property holds for $\mathbb{R}$, essentially by default, but not for $\mathbb{Q}$.

Let **F** be a field. By a **square root tower** over **F** we mean a finite sequence

$$\mathbf{F} = \mathbf{F}_0 \leq \mathbf{F}_1 \leq \cdots \leq \mathbf{F}_n$$

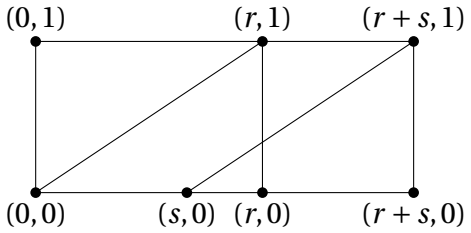
of field extensions such that for each $j < n$ there is u_j so that $u_j^2 \in F_j$ and $\mathbf{F}_j[u_j] = \mathbf{F}_{j+1}$. That is, we obtain the next field up the tower $\mathbf{F}_{j+1}$ by adjoining to $\mathbf{F}_j$ a square root of an element belonging to F_j. Let **K** be a field extending **F** and let $r \in K$. We say r is **captured in a square root tower** over **F** provided $r \in F_n$ for some square root tower $\mathbf{F} = \mathbf{F}_0 \leq \cdots \leq \mathbf{F}_n$.

The Basic Theorem for the Field of Constructible Reals. *The constructible real numbers constitute the smallest subfield* **E** *of* $\mathbb{R}$ *that is closed under the extraction of square roots. Moreover,* $r \in E$ *if and only if* r *captured in a square root tower over* $\mathbb{Q}$*. In particular,* **E** *is algebraic over* $\mathbb{Q}$ *and* $[\mathbb{Q}[r] : \mathbb{Q}]$ *is a power of* 2*, for all* $r \in E$*.*

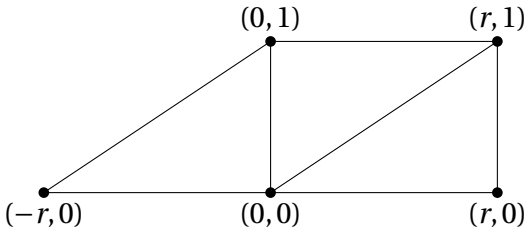
Proof. That E is closed under addition, multiplication, inversion of nonzero element and square roots of positive elements will follow from a series of diagrams. The intention of the diagrams is to display how the construction by straightedge and compass should proceed. Some familiarity with the use of straightedge and compass is needed to interpret the diagrams. For instance,

- given a line segment (i.e. its endpoints) and a point on the line segment, there is a straightedge and compass construction of a second line segment perpendicular to the first at the given point;

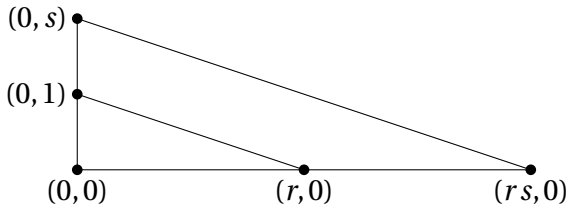
- given a line segment ℓ and a point $\mathbf{p}$ not collinear with ℓ, there is a straightedge and compass construction of a point $\mathbf{q}$ collinear with ℓ so that the segment joining $\mathbf{p}$ and $\mathbf{q}$ is perpendicular to ℓ (extended as required to include the point $\mathbf{q}$);
- given a line segment ℓ and a point $\mathbf{p}$ not collinear with ℓ, there is a straightedge and compass construction of a line segment through $\mathbf{p}$ that is parallel to ℓ.



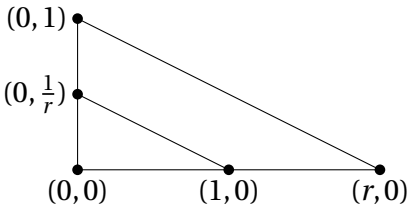
Construction of $r+s$ when $0 \le s \le r$



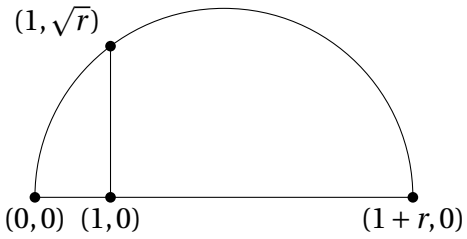
Construction of $-r$ when $0 < r$



Construction of rs when $1 \le s$ and $0 \le r$



Construction of $\frac{1}{r}$ when $0 < r$



Construction of $\sqrt{r}$ when $0 < r$

At this point we see that $\mathbf{E}$ is indeed a subfield of the field $\mathbb{R}$ of real numbers and it is closed under the extraction of square roots. Of course, it is also an extension of the field $\mathbb{Q}$ of rationals.

Now suppose that $r \in E$. Pick $s \in E$ so that $(r, s) \in \mathscr{C}$. There is a finite sequence

$$\mathbf{p}_0, \ldots, \mathbf{p}_n = (r, s)$$

of points so that $\mathbf{p}_0$ is constructible from the points $(0,0)$ and $(1,0)$ in one step by straightedge and compass, and for each $k < n$ the point $\mathbf{p}_{k+1}$ is constructible in one step from points in $\{(0,0),(1,0),\mathbf{p}_0,\ldots,\mathbf{p}_k\}$. For each $j < n$ put $\mathbf{p}_j = (r_j, s_j)$. Let $\mathbf{K} = \mathbb{Q}[r_0, s_0, r_1, s_1, \ldots, r_{n-1}, s_{n-1}]$.

Contention. $r \in K[u]$ for some $u \in \mathbb{R}$ so that $u^2 \in K$.

There are three kinds of straightedge and compass steps.

The first produces a point of intersection of two lines. Each of the lines is determined by two distinct points. Construing this algebraically, we have a system of two linear equations in two unknowns, which are the coordinates of the point of intersection. We can solve this system just using the field operations. This yields that $r \in K$ and so $r \in K[1]$.

The second produces points of intersection of a line and a circle. The line is determined by two distinct points (a,b) and (c,d), and the circle is determined by its center (a',b') and one point (c',d') on the circle. Construing this algebraically we arrive at a system of two equations:

$$(c-a)(y-b) = (d-b)(x-a)$$
$$(x-a')^2 + (y-b')^2 = (c'-a')^2 + (d'-b')^2$$

The point (r,s) is a root of this system. Using just the field operations solve the first equation for one of the unknowns in terms of the other (taking care not to divide by 0). Substituting the result into the second equation yields a quadratic equation with coefficients in K. Invoking the quadratic formula will produce values for the unknown. The formula involves the extraction of a square root of a nonnegative number $u \in K$. The value of the other unknown can be determined just using the field operations. So $r \in K[u]$ where $u^2 \in K$ in this case.

The last kind of straightedge and compass step produces the points of intersection of two circles, each determined by its center and a point of the circle. Algebraically, this yields the system

$$(x-a)^2 + (y-b)^2 = (c-a)^2 + (d-b)^2$$
$$(x-a')^2 + (y-b')^2 = (c'-a')^2 + (d'-b')^2$$

Subtracting these equations eliminates the x^2's and y^2's. The resulting equation is of the form $Ax + By = C$. This equation can be solved for one of the unknowns in terms of the other and the result substituted into the first displayed equation. From this point the argument proceeds as before.

In this way the contention is established.

Evidently, square root towers can be extended by square root towers to obtain longer square root towers. So we see, inductively, that $r_0, s_0, r_1, s_1, \ldots, r_n, s_n$ are all contained in a single square root tower over $\mathbb{Q}$.

So every element of E is captured in some square root tower over $\mathbb{Q}$.

Conversely, since square roots of nonnegative constructible numbers are themselves constructible, we see, via induction, that any real captured in a square root tower over $\mathbb{Q}$ must be constructible.

Now observe that if $\mathbf{K}$ is any subfield of $\mathbb{R}$ that is closed under the extraction of square roots, then every square root tower over $\mathbb{Q}$ is included in $\mathbf{K}$. Therefore, $\mathbf{E}$ is a subfield of $\mathbf{K}$. Since $\mathbf{E}$ is itself closed under the extraction of square roots, we see that indeed it must be the smallest subfield of $\mathbb{R}$ that is closed under the extraction of square roots.

Finally, let $r \in E$ and let $\mathbb{Q} = \mathbf{F}_0 \leq \mathbf{F}_1 \leq \cdots \leq \mathbf{F}_n$ be a square root tower that captures r. We know that $1 \leq [\mathbf{F}_{k+1} : \mathbf{F}_k] \leq 2$. It follows from the Dimension Formula that $[\mathbf{F}_n : \mathbb{Q}]$ must be a power of 2.

But $\big[\mathbf{F}_n : \mathbb{Q}\big] = \big[\mathbf{F}_n : \mathbb{Q}[r]\big]\big[\mathbb{Q}[r] : \mathbb{Q}\big]$. So we conclude that $[\mathbb{Q}[r] : \mathbb{Q}]$ must also be a power of 2, and also that r is algebraic of $\mathbb{Q}$. □

The Impossibility of General Angle Trisection. *The angle $\frac{\pi}{3}$ radians (whose construction was given in Euclid's First Proposition), cannot be trisected with straightedge and compass.*

Proof. This is the angle in an equilateral triangle. Its trisection results in an angle of $\frac{\pi}{9}$ radians. The construction of such an angle entails the constructibility of a right triangle with hypotenuse 1 and legs of length $\cos(\frac{\pi}{9})$ and $\sin(\frac{\pi}{9})$. In turn, this entails that $\cos(\frac{\pi}{9})$ would be a constructible real number. We will see that this is not the case.

The hard-working graduate student can verfiy the following trigonometric identity

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha.$$

Since $\cos 3\frac{\pi}{9} = \frac{1}{2}$, we see that $\cos\frac{\pi}{9}$ is a root of $4x^3 - 3x - \frac{1}{2}$. This polynomial is irreducible over $\mathbb{Q}$ (the verification of this is left to the enjoyment of the graduate students). This means, according to Kronecker, that $[\mathbb{Q}[\frac{\pi}{9}] : \mathbb{Q}] = 3$. Since 3 is not a power of 2, we see that $\cos\frac{\pi}{9}$ is not constructible. □

The Impossibility of Duplicating the Unit Cube. *A line segment the cube of whose length is 2 cannot be constructed by straightedge and compass.*

Proof. Evidently, $\sqrt[3]{2}$ is a root of $x^3 - 2$. According to Eisenstein, this polynomial is irreducible over $\mathbb{Q}$ and according to Kronecker $[\mathbb{Q}[\sqrt[3]{2}] : \mathbb{Q}] = 3$. Since 3 is not a power of 2 we are done. □

The Impossibility of Squaring the Unit Circle. *A line segment the square of whose length is the area of the unit circle cannot be constructed by straightedge and compass.*

We are not yet in a position to prove this theorem. The area of the unit circle is π. A square of this area would have sides of length $\sqrt{\pi}$. The constructibility of this number would entail the constructibility of π. But it turns out that π is not even algebraic over $\mathbb{Q}$. We will prove this later, finally putting this old problem to rest.

This leaves the problem of constructing regular polygons. Here is what is known.

A prime number p is said to be a **Fermat prime** provided $p = 2^a + 1$ for some positive natural number number a. Here are the first five Fermat primes: $3, 5, 17, 257$, and 65537. These five were already known to Fermat—no further Fermat primes have been found in the ensuing years, even with the very considerable computational power now at our disposal. It was conjectured by Eisenstein that there are infinitely many Fermat primes. It is even conceivable that Eisenstein was wrong and that those we now know are all that there are.

Gauss's Theorem on Constructible Regular Polygons. *Let $n \geq 3$. It is possible to construct by straightedge and compass a regular n-gon if and only if n has the form*

$$n = 2^e p_1 p_2 \dots p_m$$

where e is a natural number and $p_1, \dots, p_m$ are distinct Fermat primes.

We also have to defer the proof of this theorem. Essentially, we can identify the vertices of a regular n-gon as the complex numbers that are the roots of $x^n - 1$. We will take up a detailed study of these roots of unity later. At that time we can provide a proof of this theorem of Gauss. Even then, our grasp of the situation is incomplete since our knowledge of Fermat primes is so sketchy.

LECTURE 11

GALOIS CONNECTIONS

In his investigation of the solvability of polynoimial equations by radicals, Galois came across a way to connect (the splitting field of) a polynomial with a finite combinatorial object (in fact a finite group) which proved more amenable to analysis than the splitting field, which was an infinite object. It turns out that the connection Galois discovered is a particular instance of a what has turned out to be a common phenomena. Because more general situation is not encumbered with all the details of Galois's particular conection, and because the idea requires hardly any mathematical background, I will present the general situation first.

11.1 ABSTRACT GALOIS CONNECTIONS

Consider any two classes A and B and any two-place relation $R \subseteq A \times B$. Two-place relations are ubiquitous in mathematics. Here are some examples:

- Take $A = \mathbb{Z} = B$ and let R be the divisibility relation.
- Take A and B two the set of rational numbers in the unit interval and let R be their usual ordering $\leq$.
- Let A and B both be the set of vertices of some graph and let R be the relation of adjacency.
- Let A and B both be the class of all groups and let R be the relation of one group being a homomorphic image of another.
- Let A be the ring of polynomials in 5 variables over the complex numbers, let B be the vector space of 5-tuples of complex numbers. Take R to be the relation between a polynomial and its solutions.
- Let A and B be the class of all sets and take R to be the membership relation.
- Let A be the points in the Euclidean plane and let B be the collection of 2-element subsets of A. Let R relate of point p two $\{a, b\}$ provided p is on the line segment joining a and b.

- Imagine two or three more examples.

We call a system (A, B, R), where $R \subseteq A \times B$, a **Galois connection**. Each Galois connection induces two maps $^{\rightarrow}$ and $^{\leftarrow}$, called the **polarities** of the Galois connection.

$$^{\rightarrow} : \mathscr{P}(A) \to \mathscr{P}(B)$$

where for each $X \subseteq A$

$$X^{\rightarrow} := \{y \mid y \in B \text{ and } (x, y) \in R \text{ for all } x \in X\}$$

and

$$^{\leftarrow} : \mathscr{P}(B) \to \mathscr{P}(A)$$

where for each $Y \subseteq B$

$$Y^{\leftarrow} := \{x \mid x \in A \text{ and } (x, y) \in R \text{ for all } y \in Y\}$$

I read $X^{\rightarrow}$ as "X going over" and $Y^{\leftarrow}$ as "Y coming back."

An example is in order. Let $A = \mathbb{Z} = B$ and let R be the divisibility relation. So

$$\{6, 9\}^{\rightarrow} = \{r \mid r \in \mathbb{Z} \text{ and } 6 \mid r \text{ and } 9 \mid r\}$$

That is, $\{6, 9\}^{\rightarrow}$ is just all the multiples of 18, since 18 is the least common multiple of 6 and 9. So even though we started with a finite set, by going over we got an infinite set—but it is a nice one, an ideal of the ring of integers. Now lets see what we get by coming back.

$$\{6, 9\}^{\rightarrow\leftarrow} = \{s \mid s \in \mathbb{Z} \text{ and } s \mid r \text{ for all } r \in \{6, 9\}^{\rightarrow}\}$$

With a little thought the industrious graduate students will find that

$$\{6, 9\}^{\rightarrow\leftarrow} = \{1, -1, 2, -2, 3, -3, 6, -6, 9, -9, 18, -18\}.$$

This is just the set of divisors of 18. So we started with a set with two elements and then by going over and coming back we arrive at a set with 12 elements that includes our original set. Now observe

$$\varnothing^{\rightarrow} = \mathbb{Z} \text{ and } \varnothing^{\leftarrow} = \mathbb{Z}.$$

Also observe

$$\mathbb{Z}^{\rightarrow} = \{0\} \text{ and } \mathbb{Z}^{\leftarrow} = \{1, -1\}.$$

Now $\mathscr{P}(A)$ and $\mathscr{P}(B)$ are partially ordered by the inclusion relation $\subseteq$. The basic properties of Galois connections concern how the polarities and this order relation interact.

The Polarity Theorem for Galois Connections. *Let (A, B, R) be any Galois connection. All of the following hold.*

(a) *If $X_0 \subseteq X_1 \subseteq A$, then $X_1^{\rightarrow} \subseteq X_0^{\rightarrow}$. If $Y_0 \subseteq Y_1 \subseteq B$, then $Y_1^{\leftarrow} \subseteq Y_0^{\leftarrow}$.*

(b) *If $X \subseteq A$, then $X \subseteq X^{\rightarrow\leftarrow}$. If $Y \subseteq B$, then $Y \subseteq Y^{\leftarrow\rightarrow}$.*

(c) *If $X \subseteq A$, then $X^{\rightarrow} = X^{\rightarrow\leftarrow\rightarrow}$. If $Y \subseteq B$, then $Y^{\leftarrow} = Y^{\leftarrow\rightarrow\leftarrow}$.*

(d) *If* $X_0, X_1 \subseteq A$ *and* $X_0^{\rightarrow\leftarrow} = X_1^{\rightarrow\leftarrow}$, *then* $X_0^{\rightarrow} = X_1^{\rightarrow}$.
If $Y_0, Y_1 \subseteq B$ *and* $Y_0^{\leftarrow\rightarrow} = Y_1^{\leftarrow\rightarrow}$, *then* $Y_0^{\leftarrow} = Y_1^{\leftarrow}$.

(e) *For all* $X \subseteq A$ *and all* $Y \subseteq B$,

$$X \subseteq Y^{\leftarrow} \text{ if and only if } Y \subseteq X^{\rightarrow}.$$

The proof of this theorem is left in the trustworthy hands of the graduate students. It is even part of an official problem set. Can you deduce the other parts of this theorem from (a) and (e)?

We say that subsets of A of the form $Y^{\leftarrow}$ are **closed**, as are the subsets of B of the form $X^{\rightarrow}$. Part of the content of this theorem about polarities is that restricted to the closed sets on each side of the Galois connection, the polarities are inverses of each other and they are order reversing. So viewed as ordered sets, the systems of closed sets on each side are *anti-isomorpic*: one looks like the upside down version of the other. The polarities are the anti-isomorphisms.

The intersection of any nonempty collection of closed sets from one side of a Galois connection will be again a closed set. There is a unique smallest closed set on each side. The least closed subset of A is, of course, $B^{\leftarrow}$. This means that any collection of closed subsets from one side of a Galois connection always has a greatest lower bound. It follows via the polarities (which are anit-isomorphisms) that every collection of closed sets from one side of a Galois connection always has a least upper bound. A partially ordered set with these properties is called a **complete lattice**. So the closed sets from any one side of a Galois connection always constitute a complete lattice. In the example we worked with, the integers with divisibility, the closed sets on the right side of the Galois connection turn out to be the ideals of the ring of integers.

11.2 The Connection of Galois

The Galois connection discovered by Evariste Galois was not listed among our examples in the section above. We describe it in this section.

Let **E** be a field that extends a field **F**. The set E will be the left side of Galois' connection. Let

$$\mathrm{Gal}\,\mathbf{E}/\mathbf{F} = \{\sigma \mid \sigma \text{ is an automorphism of } \mathbf{E} \text{ and } \sigma(a) = a \text{ for all } a \in F\}.$$

$\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$ is the group of automorphisms of **E** that fix each element of **F**. It is called the **Galois group** of **E** over **F**. It is the right side of Galois' connection. The relation that connects these two sides is

$$\{(a, \sigma) \mid a \in E \text{ and } \sigma \in \mathrm{Gal}\,\mathbf{E}/\mathbf{F} \text{ and } \sigma(a) = a\}.$$

The polarities of Galois' connection are given as follows, for any $X \subseteq E$ and any $Y \subseteq \mathrm{Gal}\,\mathbf{E}/\mathbf{F}$:

$$X^{\rightarrow} = \mathrm{Gal}\,X = \{\sigma \mid \sigma \in \mathrm{Gal}\,\mathbf{E}/\mathbf{F} \text{ and } \sigma(x) = x \text{ for all } x] \in X\}$$
$$Y^{\leftarrow} = \mathrm{Inv}\,Y = \{a \mid a \in E \text{ and } \sigma(a) = a \text{ for all } \sigma \in Y\}$$

We abandon the arrow notation in favor of Gal and Inv. We leave it in the trustworthy hands of the graduate students to work out the $\mathrm{Gal}\,X$ is always a subgroup of $\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$ and that $\mathrm{Inv}\,Y$ is always a subfield of **E** that extends **F**. $\mathrm{Inv}\,Y$ is called the **fixed field** of Y and $\mathrm{Gal}\,X$ is called the Galois group of X.

In Galois' investigations, the field **E** was the splitting field of some polynomial $f(x)$ with coefficients from **F**. Galois realized that an automorphism of **E** that leaves the coefficients of the

polynomial fixed must send roots of $f(x)$ to roots of $f(x)$. And as the roots of $f(x)$ determine the elements of the splitting field $\mathbf{E}$, this meant that $\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$ was, in essence, just some group consisting of certain permutations of the roots of $f(x)$. Such a group is finite since $f(x)$ can have only finitely many roots. In this way, Galois saw that it might be possible to understand the roots of $f(x)$ by understanding this finite group instead of trying to understand how the roots where situated in the (usually) infinite field $\mathbf{E}$. Galois succeeded.

The next two lectures are devoted to understanding the closed sets on each side of Galois' connection.

11.3 Problem Set 18

Algebra Homework, Edition 18
Galois Connections

In Problem 63 to Problem 67 below, let A and B be two classes and let R be a binary relation with $R \subseteq A \times B$. For $X \subseteq A$ and $Y \subseteq B$ put

$$X^{\rightarrow} = \{b \mid x\ R\ b \text{ for all } x \in X\}$$
$$Y^{\leftarrow} = \{a \mid a\ R\ y \text{ for all } y \in Y\}$$

Problem 63.
Prove that if $W \subseteq X \subseteq A$, then $X^{\rightarrow} \subseteq W^{\rightarrow}$. (Likewise if $V \subseteq Y \subseteq B$, then $Y^{\leftarrow} \subseteq V^{\leftarrow}$.)

Problem 64.
Prove that if $X \subseteq A$, then $X \subseteq X^{\rightarrow\leftarrow}$. (Likewise if $Y \subseteq B$, then $Y \subseteq Y^{\leftarrow\rightarrow}$.)

Problem 65.
Prove that $X^{\rightarrow\leftarrow\rightarrow} = X^{\rightarrow}$ for all $X \subseteq A$ (and likewise $Y^{\leftarrow\rightarrow\leftarrow} = Y^{\leftarrow}$ for all $Y \subseteq B$).

Problem 66.
Prove that the collection of subclasses of A of the form $Y^{\leftarrow}$ is closed under the formation of arbitrary intersections. (As is the collection of subclasses of B of the form $X^{\rightarrow}$.) We call classes of the form $Y^{\leftarrow}$ and the form $X^{\rightarrow}$ closed.

Problem 67.
Let $A = B = \{q \mid 0 < q < 1 \text{ and } q \text{ is rational}\}$. Let R be the usual ordering on this set. Identify the system of closed sets. How are they ordered with respect to inclusion?

LECTURE 12

THE FIELD SIDE OF GALOIS' CONNECTION

Suppose that **E** is the splitting field of a polynomial $f(x)$ over the field **F** and that **K** is a field intermediate between **F** and **E**. From general facts about Galois connections, we know that $\mathrm{Inv}\,\mathrm{Gal}\,E/K$ is a subfield of **E** that extends **K**. Our hope is that $\mathbf{K} = \mathrm{Inv}\,\mathrm{Gal}(E/K)$. While this hope cannot be realized in general, there is an important case in which it does hold.

Let us say that an irreducible polynomial $p(x) \in \mathbf{F}[x]$ is *separable* provided the number of roots it has in its splitting field over **F** is the same as its degree. This means that when $p(x)$ is completely factored over its splitting field, then all the factors are distinct, that is all the roots are distinct or separated. We say a polynomial $f(x) \in \mathbf{F}[x]$ is *separable* provided each of its irreducible factors is separable. Notice that a separable polynomial of degree n may have fewer than n distinct roots in its splitting field. For example $x^2+2x+1 = (x+1)^2$ has degree 2 but it has only one root (namely -1). In general, a polynomial $f(x) \in \mathbf{F}[x]$ factors over **F** as

$$f(x) = g_0(x)^{e_0} g_1(x)^{e_1} \dots g_{m-1}(x)^{e_{m-1}}$$

where each $g_k(x)$ is irreducible and disinct from the other g's and each e_k is a positive integer. This polynomial must have repeated roots in its splitting field as long as some $e_k > 1$. But consider the polynomial

$$h(x) = g_0(x) g_1(x) \dots g_{m-1}(x).$$

The polynomials $f(x)$ and $h(x)$ have the same splitting field over **F**, and $h(x)$ will have distinct roots, provided $f(x)$ (and hence $h(x)$) is separable.

The Galois Field Closure Theorem. *Let* **E** *be the splitting field of a separable polynomial over the field* **F**. *Then* $\mathrm{Inv}\,\mathrm{Gal}(\mathbf{E}/\mathbf{K}) = \mathbf{K}$, *for every field* **K** *intermediate between* **F** *and* **K**.

Proof. Let $f(x) \in \mathbf{F}[x]$ be a separable polynomial from $\mathbf{F}[x]$ and let **E** be the splitting field of $f(x)$ over **F**. Let **K** be a field intermediate between **F** and **E** and put $\mathbf{L} = \mathrm{Inv}\,\mathrm{Gal}(\mathbf{E}/\mathbf{K})$. For the general properties of polarities for Galois connections, we see that $K \subseteq L$ and that $\mathrm{Gal}(\mathbf{E}/\mathbf{K}) = \mathrm{Gal}(\mathbf{E}/\mathbf{L})$. But **E** is the splitting field of $f(x)$ over both **K** and **L**. By the Existence and Uniqueness Theorem for Splitting Fields, we see that $[\mathbf{E},\mathbf{K}] = |\mathrm{Gal}(\mathbf{E}/\mathbf{K})| = |\mathrm{Gal}(\mathbf{E}/\mathbf{L}| = [\mathbf{E},\mathbf{L}]$. But we know that $[\mathbf{E},\mathbf{K}] = [\mathbf{E},\mathbf{L}][\mathbf{L},\mathbf{K}]$. It follows that $[\mathbf{L},\mathbf{K}] = 1$. Hence, $\mathbf{K} = \mathbf{L}$, as desired. □

12.1 Perfect Fields

This leaves us with the question of when an irreducible polynomial is separable. It turns out that just a bit of formal calculus does the trick. Consider a polynomial

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

We can define the derivative of $f'(x)$ as follows

$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1}.$$

But we have to be careful. The exponents are natural numbers, but our field **F**, might not have in it any natural numbers. So, to be necessarily more fussy, we define

$$f'(x) = a_1 + ((1+1)a_2)x + ((1+1+1)a_3)x^2 + \cdots + ((\underbrace{1+\cdots+1}_{n\text{-times}})a_n)x^{n-1}.$$

But after this we will write it as we did at first. Notice that this definition always produces another polynomial, regardless of the field over which we are working. No limits or other notion of convergence enters here.

It is left to the eager graduate students to verify that the derivatives of sums and products (and even compositions) of polynomials work out just like they do in calculus.

Fact. Let $f(x)$ be an irreducible polynomial with coefficients in the field **F**. Then $f(x)$ is separable if and only if $f(x)$ and $f'(x)$ are relatively prime.

Proof. It is harmless to suppose that $f(x)$ is monic. Let **E** be the splitting field of $f(x)$ over **F**. Let $r_0, \ldots, r_{m-1}$ be the distinct roots of $f(x)$ in E. Then we see that

$$f(x) = (x - r_0)_0^e (x - r_1)^{e_1} \ldots (x - r_{m-1})^{e_{m-1}},$$

for certain positive integers $e_0, \ldots, e_{m-1}$.

Suppose first that $f(x)$ is separable. This only means that $e_0 = \cdots = e_{m-1} = 1$. Under this supposition, $f'(x)$ is, according to the product rule, just the sum of all terms made by deleting single factors from the factorization above. This entails (with the graduate students fiddling down the details) that none of the irreducible factors (over **E**) of $f(x)$ can divide $f'(x)$. This means that $f(x)$ and $f'(x)$ are relatively prime over **E**. Therefore, they must be relatively prime over **F**.

Now suppose that $f(x)$ is not separable. This means that some $e_k > 0$. Hence, $f(x) = (x-r)^2 g(x)$ is a factorization over **E** for some $r \in E$ and some $g(x) \in \mathbf{E}[x]$. The product rule now tells us that $f'(x) = 2(x-r)g(x) + (x-r)^2 g'(x)$. This means that $x - r$ is a common divisor of $f(x)$ and $f'(x)$. So $f(x)$ and $f'(x)$ are not relatively prime over **E**. Hence (why?) they are not relatively prime over **F**. □

Actually, the proof above does not make significant use of the irreducibility of $f(x)$. The graduate students should be able to reformulate the statement of this fact so as to remove the irreducibility condition.

Of course, from our perspective the best thing that can happen is for all polynomials of positve degree to turn out to be separable. A field **F** with this property is called *perfect*.

Here is an important corollary of the Fact above.

Corollary 12.1.1. *Every field of characteristic* 0 *is perfect.*

Proof. We only need to pay attention to irreducible polynomials. Observe that in a field of characteristic 0, if $f(x)$ has positve degree, then $f'(x)$ cannot be the zero polynomial and must have degree properly smaller than the degree of $f(x)$. (Over fields of prime characteristic it is possible for $f'(x)$ to be the zero polynomial.) Since we are taking $f(x)$ to be irreducible, we see that $f(x)$ and $f'(x)$ must be relatively prime, since $f(x)$ cannot divide $f'(x)$, the degree of $f'(x)$ being too small. □

So what happens for fields of prime characteristic?

Suppose that **F** is a field of characteristic p, where p is a prime number. There is an interesting thing that happens. According to the Binomial Theorem (that holds in every commutative ring) in $\mathbf{F}[x]$ we have

$$(a+b)^p = \sum_{k \le p} \binom{p}{k} a^k b^{p-k}$$

for all $a, b \in F$. Recall that $\binom{p}{0} = 1 = \binom{p}{p}$ but that $p \mid \binom{p}{k}$ when $0 < k < p$. Recalling the fussy point made above about positive integer multiples, we see that $\binom{p}{k}$ reduces to 0 in the characteristic p case, whenever $0 < k < p$. This means that

$$(a+b)^p = a^p + b^p$$

for all $a, b \in F$. But we also know that

$$(ab)^p = a^p b^p$$

for all $a, b \in F$. This means that the map $a \mapsto a^p$ for all $a \in F$, must be a homomorphism from **F** into **F**. Now fields are simple, that is they have just two ideals. So the kernel of this special map must either be $\{0\}$ (in which case the map is one-to-one) or F itself (in which case the map sends every element of F to 0). Since $1^p = 1 \neq 0$, we see that our map is actually one-to-one, that is it is an embedding of **F** into **F**. This map is known as the *Frobenius embedding*.

Theorem on Perfect Fields of Charactistic p**.** *Let* **F** *be a field of prime characteristic* p*. The field* **F** *is perfect if and only if every element of* F *has a* p^{th} *root in* F*.*

Proof. Suppose first that there is $a \in F$ so that a has no p^{th} root in F. We contend that the polynomial $x^p - a$ is irreducible. Suppose otherwise. So $x^p - a = g(x)h(x)$ where $g(x)$ is a monic polynomial of positive degree $k < p$. Let **E** be the splitting field of $x^p - a$ over **F** and let $b \in E$ be a root of $g(x)$. Now notice $b^p = a$ so $b \notin F$ and

$$x^p - a = x^p - b^p = (x-b)^p$$

in $\mathbf{E}[x]$. By unique factorization, $g(x) = (x-b)^k$. As b^k is the constant term of $g(x)$, we find that $b^k \in F$. But k and p are relatively prime integers (since p is prime and $0 < k < p$). Pick integers u and v so that $1 = uk + vp$. But then

$$b = b^{uk+vp} = (b^k)^u (b^p)^v = (b^k)^u a^v \in F.$$

This provides a contradiction to our supposition. So $x^p - a$ is irreducible. Its derivative is the zero polynomial. So we see that it is not separable. This means that **F** is not perfect.

For the converse, suppose that every element of F has a p^{th} root in F. Consider any irreducible polynomial $f(x)$. The only barrier to $f(x)$ being separable is that $f'(x)$ might be the zero polynomial. This can only happen when

$$f(x) = a_0 + a_p x^p + a_{2p} x^{2p} + \cdots + a_{np} x^{np}.$$

Since every element of F is a p^{th} power we can pick $b_0, \ldots, b_{np} \in F$ so that $a_{kp} = (b_{kp})^p$ for all $k \leq n$. This gives us

$$f(x) = b_0^p + b_p^p x^p + \cdots + b_{np}^p x^n p = (b_0 + b_p x + b_{2p} x^2 + \cdots + b_{np} x^n)^p.$$

But $f(x)$ is irreducible. So it cannot happen that $f'(x)$ is the zero polynomial. Consequently, $f(x)$ is separable and **F** must be a perfect field. □

A nice corollary of this theorem is

Corollary 12.1.2. *Every finite field is perfect.*

The reason is that the Frobenius map from a finite field to itself must be onto the field since it is one-to-one (and any one-to-one map of a finite set to itself must be onto). So we see that every element of a finite field is a p^{th} power of some other element, where p is the characteristic.

12.2 Galois Extensions

The key hypothesis of the Galois Flied Closure Theorem is that **E** should be the splitting field of a separable polynomial over **F**. In this case, we say that **E** is a **Galois extension** of **F**. There are several useful ways to characterize this kind of extension.

Theorem Characterizing Galois Extensions. *Let* **E** *be a finite extension of the field* **F**. *The following conditions are equivalent.*

(a) **E** *is a Galois extension of* **F**.

(b) *Every element of* E *is a root of a separable polynomial in* $\mathbf{F}[x]$ *and every irreducible polynomial in* $\mathbf{F}[x]$ *that has a root in* E *splits in* **E**.

(c) $\mathbf{F} = \text{Inv}\,\text{Gal}\,\mathbf{E}/\mathbf{F}$.

We say that **E** is a **separable extension** of **F** provided every element of E is a root of a separable polynomial in $\mathbf{F}[x]$. We say that **E** is a **normal extension** of **F** provided every polynomial of $\mathbf{F}[x]$ that has a root is E splits over **E**. So condition (b) in this theorem says that **E** is a normal separable extension of **F**.

Proof. **(a)⇒(c)**
According to the Galois Field Closure Theorem, every intermediate field between **F** and **E** is closed. In particular, **F** is closed. This entails that $\mathbf{F} = \text{Gal}(E\mathbf{E}/\mathbf{F})$.

(c)⇒(b)
Let $r \in E$. Since $[\mathbf{E} : \mathbf{F}]$ is finite we know that r is algebraic over **F**. Let $m(x)$ be the minimal polynomial of r over **F**. We need to show that $m(x)$ is separable and that it splits over **E**. Now

for each $\sigma \in \mathrm{Gal}\,\mathbf{E}/\mathbf{F}$ we know that $\sigma(r)$ is also a root of $m(x)$. Let $r_0, r_1, \ldots, r_{\ell-1}$ be a list of all the distinct images of r by automorphisms belonging to $\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$. (This is just the orbit of r under the acton of $\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$.) Let $f(x) = (x-r_0)(x-r_1)\ldots(x-r_{\ell-1})$. The coefficients of $f(x)$ are fixed by each automorphism belonging to $\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$. That is these coefficients belong to $\mathrm{Inv}\,\mathrm{Gal}\,\mathbf{E}/\mathbf{F}$. So by (c) we find that $f(x) \in F[x]$. So $m(x) \mid f(x)$. On the other hand, $(x-r_i) \mid m(x)$ for each $i < \ell$. This means $f(x) \mid m(x)$. Since both $m(x)$ and $f(x)$ are monic, we see $m(x) = f(x)$. So $m(x)$ is separable and splits over $\mathbf{E}$.

(b)⇒(a)
Since $[\mathbf{E} : \mathbf{F}]$ is finite, there are finitely many elements $s_0, \ldots, s_{n-1} \in E$ so that $\mathbf{E} = \mathbf{F}[s_0, \ldots, s_{n-1}]$. Let $m_i(x)$ be the minimal polynomial of s_i over $\mathbf{F}$, for each $i < n$. According to (b), each of these polynomial is separable and splits over $\mathbf{E}$. Let $f(x) = m_0(x)m_1(x)\ldots M_{n-1}(x)$. So $f(x)$ is a separable polynomial that splits over $\mathbf{E}$. Evidently, $\mathbf{E}$ is the splitting field of $f(x)$ over $\mathbf{F}$. So $\mathbf{E}$ is a Galois extension of $\mathbf{F}$. □

12.3 Problem Set 19

Algebra Homework, Edition 19
Field extensions

Problem 68.
Let **E** and **F** be fields. Prove that **E** is an algebraic closure of **F** if and only if **E** is an algebraic extension of **F** and for every algebraic extension **K** of **F** there is an embedding of **K** into **E** which fixes each element of **F**.

Problem 69.
Prove that if **E** extends the field **F** and $[\mathbf{E}:\mathbf{F}] = 2$, then **E** is a normal extension of **F**.

Problem 70.
Let **E** be a field extending the field **F**. Let **L** and **M** be intermediate fields such that **L** is the splitting field of a separable polynomial in $\mathbf{F}[x]$. Let $\mathbf{L} \vee \mathbf{M}$ denote the smallest subfield of **E** that extends both **L** and **M**. Prove that $\mathbf{L} \vee \mathbf{M}$ is a finite normal separable extension of **M** and that $\mathrm{Aut}_{\mathbf{M}}(\mathbf{L} \vee \mathbf{M}) \cong \mathrm{Aut}_{\mathbf{M} \cap \mathbf{L}} \mathbf{L}$.

Problem 71.
Let **L** and **M** be fields. Then the collection of functions from **L** into **M** can be regarded as a vector space over **M**. (Add functions like we do in calculus...). Prove that the collection of field embeddings from **L** into **M** is a linearly independent set in this vector space.

Problem 72.
Let **F** be a field. We use $\mathbf{F}^\times$ to denote the group of nonzero elements of **F** under multiplication and the formation of multiplicative inverses. Show that every finite subgroup of $\mathbf{F}^\times$ is a cyclic group.

LECTURE 13

THE GROUP SIDE OF GALOIS' CONNECTION AND THE FUNDAMENTAL THEOREM

13.1 CLOSED SUBGROUPS OF A GALOIS GROUP

Now we want to determine what the closed subgroups of a Galois group are. Since on the field side we found it convenient to look at Galois extensions, here we will focus on the case when $\mathbf{E}$ is a Galois extension of $\mathbf{F}$.

Our first step is to develop more information on the field side. We begin with a theorem of Ernst Steinitz

Theorem on Primitive Elements—Steinitz, 1910. *Let* $\mathbf{E}$ *be a finite extension of* $\mathbf{F}$. *The following are equivalent.*

(a) *There is an element* $r \in E$ *so that* $\mathbf{E} = \mathbf{F}[r]$.

(b) *There are only finitely many fields intermediate between* $\mathbf{F}$ *and* $\mathbf{E}$.

The element r mentioned in (a) is a **primitive** element of $\mathbf{E}$ with respect to $\mathbf{F}$.

Proof. Since $\mathbf{E}$ is a finite extension of $\mathbf{F}$, we observe that E is finite if and only if F is finite. Let us first dispose of the case when either (and hence both) of these fields is finite. Of course, we have that (b) holds in this case. So to see that (a) and (b) are equivalent, we must only prove that (a) is also true. Let $\mathbf{E}^\times$ denote the group of nonzero elements of E under multiplication. This is a finite subgroup of $\mathbf{E}^\times$ (of course). But we saw last semester that such finite subgroups must be cyclic. Let r be any generator of the group $\mathbf{E}^\times$. Evidently, $\mathbf{E} = \mathbf{F}[r]$ and we have found our primitive element.

So now we turn to the case when F is infinite.

Let $\mathscr{F} = \{\mathbf{K} \mid \mathbf{F} \leq \mathbf{K} \leq \mathbf{E}\}$. That is, $\mathscr{F}$ is be collection of intermediate fields.

(a) $\Longrightarrow$ (b).

Let $r \in E$ such that $\mathbf{E} = \mathbf{F}[r]$. Let $f(x)$ be the minimal polynomial of r over $\mathbf{F}$. Let $\mathscr{P} = \{g(x) \mid$

$g(x)$ is a monic polynomial in $\mathbf{E}[x]$ that divides $f(x)\}$. By unique factorization for $\mathbf{E}[x]$ we see that $\mathscr{P}$ is finite. So our proof of (a) $\Longrightarrow$ (b) will be complete when we show that $\mathscr{F}$ can be mapped into $\mathscr{P}$ by a one-to-one map.

So suppose $\mathbf{K} \in \mathscr{F}$. Let $g_{\mathbf{K}}(x)$ be the minimal polynomial of r over $\mathbf{K}$. Then $g_{\mathbf{K}}(x)$ is certainly monic and irreducible. Also $g_{\mathbf{K}}(x)$ must divide $f(x)$ since the set of polynomials over $\mathbf{K}$ that have r as a root is just the ideal generated by $g_{\mathbf{K}}(x)$ and $f(x)$ belongs to this ideal since $\mathbf{F} \leq \mathbf{K}$. So let $\Phi : \mathscr{F} \to \mathscr{P}$ be defined so that

$$\Phi(\mathbf{K}) := g_{\mathbf{K}}(x) \text{ for all } \mathbf{K} \in \mathscr{F}.$$

We need to prove that Φ is one-to-one, or, what is the same, that $\mathbf{K}$ can be recovered from $g_{\mathbf{K}}(x)$.

So let $g_{\mathbf{K}} = a_0 + a_1 x + \cdots + a_{m-1}x^{m-1} + x^m$. Let $\mathbf{L} = \mathbf{F}[a_0, \ldots, a_{m-1}]$. Evidently $\mathbf{F} \leq \mathbf{L} \leq \mathbf{K}$ and $g_{\mathbf{K}}(x)$ is irreducible over $\mathbf{L}$. This means that $g_{\mathbf{K}}(x)$ is the minimal polynomial of r over $\mathbf{L}$. From this and Kronecker we see

$$[\mathbf{E} : \mathbf{L}] = \deg g_{\mathbf{K}}(x) = [\mathbf{E} : \mathbf{K}].$$

But we also have $[\mathbf{E} : \mathbf{K}] = [\mathbf{E} : \mathbf{L}][\mathbf{L} : \mathbf{K}]$. So it follows that $[\mathbf{L} : \mathbf{K}] = 1$. This means that $\mathbf{K} = \mathbf{L} = \mathbf{F}[a_0, \ldots, a_{m-1}]$. Therefore, $\mathbf{K}$ can indeed be recovered from $g_{\mathbf{K}}(x)$. We conclude that Φ is one-to-one and that $\mathscr{F}$ is finite. This establishes (a) $\Longrightarrow$ (b).

(b) $\Longrightarrow$ (a)
So we assume that $\mathscr{F}$ is finite and we want to prove there is a primitive element. We proceed by induction of $[\mathbf{E} : \mathbf{F}]$.

The base step of the induction is the case when $\mathbf{E} = \mathbf{F} = \mathbf{F}[1]$, which almost proves itself.

For the induction step, let $s \in E \setminus F$. Then $[\mathbf{E} : \mathbf{F}[s]] < [\mathbf{E} : \mathbf{F}]$. As there are only finitely many fields between $\mathbf{F}[s]$ and $\mathbf{E}$ condition (b) holds. By the inductive hypothesis pick $t \in E$ so that $\mathbf{E} = \mathbf{F}[s][t] = \mathbf{F}[s, t]$. For each $a \in F$ let $\mathbf{K}_a = \mathbf{F}[s + at]$. Each of the fields $\mathbf{K}_a$ is between $\mathbf{F}$ and $\mathbf{E}$. There are only finitely many intermediate fields but there are infinitely many choices for a since F is infinite. As pigeons know, this means that there are $a, b \in F$ with $a \neq b$ but $\mathbf{K}_a = \mathbf{K}_b$. Now $s + at, s + bt \in K_a$. Subtracting, we find that $(a - b)t \in K_a$. But $a - b \neq 0$ and $a - b \in F \subseteq K_a$. So we can conclude that $t \in K_a$. But $a \in F \subseteq K_a$, so $at \in K_a$. But $s + at \in K_a$ so we arrive at $s \in K_a$. But this means

$$\mathbf{E} = \mathbf{F}[s, t] \leq \mathbf{K}_a = \mathbf{F}[s + at] \leq \mathbf{E}.$$

So we can take our primitive element to be $r = s + at$. □

Corollary: Artin's Primitive Element Theorem. *Let* $\mathbf{E}$ *be a finite separable extension of* $\mathbf{F}$. *Then* $\mathbf{E}$ *has a primitive element with respect to* $\mathbf{F}$.

Proof. Since $\mathbf{E}$ is a finite extension of $\mathbf{F}$, we pick $s_0, \ldots s_{m-1} \in E$ so that $\mathbf{E} = \mathbf{F}[s_0, \ldots, s_{m-1}]$. For each $j < m$ let $f_j(x)$ be the minimal polynomial of s_j over $\mathbf{F}$. Since $\mathbf{E}$ is a separable extension, each of these minimal polynomials is separable. Let $f(x)$ be the product of all the $f_j(x)$'s. Then $f(x)$ is also separable. Let $\mathbf{L}$ be a splitting field of $f(x)$ over $\mathbf{F}$. Since $\mathbf{E} = \mathbf{F}[s_0, \ldots, s_{m-1}]$ and each s_j is a root of $f(x)$, we can insist that $\mathbf{E} \leq \mathbf{L}$. Since $\mathbf{L}$ is the splitting field over $\mathbf{F}$ of a separable polynomial, we know that $\mathbf{L}$ is a Galois extension of $\mathbf{F}$. Now $\mathrm{Gal}(\mathbf{L}/\mathbf{F})$ is finite, it is even embeddable into the symmetric group on the set of all roots of $f(x)$ in $\mathbf{L}$, which is a finite set. In particular, $\mathrm{Gal}(\mathbf{L}/\mathbf{F})$ has only finitely many subgroups. We know our Galois connection sets up a one-to-one correspondence between the fields intermediate between $\mathbf{F}$ and $\mathbf{L}$ and certain subgroups of $\mathrm{Gal}(\mathbf{L}/\mathbf{F})$. So there can only be finitely many fields intermediate between $\mathbf{F}$ and $\mathbf{L}$, and hence, between $\mathbf{F}$

and **E**. By Steinitz' Theorem on Primitive Elements **E** must have a primitive element with respect to **F**. □

Fact. Let **E** be a finite extension of **F**. Each of the following statements holds.

(a) $|\mathrm{Gal}(\mathbf{E}/\mathbf{F})|$ divides $[\mathbf{E}:\mathbf{F}]$.

(b) $|\mathrm{Gal}(\mathbf{E}/\mathbf{F})| = [\mathbf{E}:\mathbf{F}]$ if and only if **E** is a Galois extension of **F**.

Proof. Let $\bar{\mathbf{F}} = \mathrm{Inv}\,\mathrm{Gal}(\mathbf{F}) = \mathrm{Inv}\,\mathrm{Gal}(\mathbf{E}/\mathbf{F})$. We see that **E** is a Galois extension of $\bar{\mathbf{F}}$. In particular, **E** is a finite separable extension of $\bar{\mathbf{F}}$ and so it has a primitive elements. Pick $r \in E$ so that $\mathbf{E} = \bar{\mathbf{F}}[r]$. Let $f(x)$ be the minimal polynomial of r over $\bar{\mathbf{F}}$.

Now $\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})$ acts on E. Let $\mathcal{O}$ be the orbit of r under this action. Of course, every automorphism in $\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})$ maps r to some other root of $f(x)$. By Kronecker, there are enough automorphisms in this Galois group to map r to each other root of $f(x)$. So the orbit $\mathcal{O}$ is exactly the set of all roots of $f(x)$. But since $f(x)$ is an irreducible separable polynomial the number of its roots is just its degree. This tells us

$$|\mathcal{O}| = \deg f(x) = [\bar{\mathbf{F}}[r]:\bar{\mathbf{F}}] = [\mathbf{E}:\bar{\mathbf{F}}].$$

But we can count the number of elements in $\mathcal{O}$ using the Key Fact about group actions.

$$|\mathcal{O}| = [\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}} : \mathbf{Stab}\, r].$$

Observe that $\mathbf{Stab}\, r = \{\sigma \mid \sigma \in \mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}}) \text{ and } \sigma(r) = r\}$. But $\mathbf{E} = \bar{\mathbf{F}}[r]$. So $\mathbf{Stab}\, r$ is just a one element group. This means that $|\mathcal{O}| = |\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})|$. Consequently,

$$[\mathbf{E}:\bar{\mathbf{F}}] = |\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})|.$$

Recalling that $[\mathbf{E}:\mathbf{F}] = [\mathbf{E}:\bar{\mathbf{F}}][\bar{\mathbf{F}}:\mathbf{F}]$ and $\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}}) = \mathrm{Gal}(\mathbf{E}/\mathbf{F})$ we obtain (a) and the right to left direction of (b).

To obtain the left to right direction of (b), we need to see that if $|\mathrm{Gal}(\mathbf{E}/\mathbf{F})| = [\mathbf{E}:\mathbf{F}]$ then $\bar{\mathbf{F}} = \mathbf{F}$. From general considerations about Galois connections we know that $\mathrm{Gal}(\mathbf{E}/\mathbf{F}) = \mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})$. But by what we saw above

$$[\mathbf{E}:\bar{\mathbf{F}}] = |\mathrm{Gal}(\mathbf{E}/\bar{\mathbf{F}})| = |\mathrm{Gal}(\mathbf{E}/\mathbf{F}) = [\mathbf{E}:\mathbf{F}].$$

Since $\mathbf{F} \leq \bar{\mathbf{F}}$ we draw the desired conclusion that $\mathbf{F} = \bar{\mathbf{F}}$. □

With this groundwork, we are prepared to examine the closed subgroups on the group side of the Galois connection.

The Galois Group Closure Theorem. *Let* **E** *be a Galois extension of* **F** *and let* **H** *be a subgroup of* $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$. *Then* $\mathrm{Gal}\,\mathrm{Inv}\,\mathbf{H} = \mathbf{H}$. *In other words, every subgroup of the Galois group is closed with respect to Galois'f connection.*

Proof. On general principles we know $\mathbf{H} \leq \mathrm{Gal}\,\mathrm{Inv}\,\mathbf{H}$. We need to reverse the order.

Let $\mathbf{K} = \mathrm{Inv}\,\mathbf{H}$. So **K** is the subfield of all elements of E fixed by every automorphism in H. Of course $\mathrm{Gal}\,\mathbf{K} = \mathrm{Gal}(\mathbf{E}/\mathbf{K})$.

Now **H** acts on E. For each $s \in E$ let $f_s(x)$ be the minimal polynomial of s over **K**. Since **E** is a Galois extension of **F** it must also be a Galois extension of **K**, so we know these polynomials are separable. Let $\mathcal{O}_s$ be the orbit of s under the action by **H**. Then we see

$$[\mathbf{K}[s]:\mathbf{K}] = \deg f_s(x) = |\mathcal{O}_s| = [\mathbf{H}:\mathbf{Stab}\, s].$$

But by Lagrange, we know that $[\mathbf{H} : \mathbf{Stab}\, s]$ divides the order of $\mathbf{H}$, which is finite. This means that $[\mathbf{K}[s] : \mathbf{K}]$ is bounded above by a finite number as s ranges through E. Pick $t \in E$ so that $[\mathbf{K}[t] : \mathbf{K}]$ is as large as possible. Now we also know that $\mathbf{E}$ has a primitive element r with respect to $\mathbf{K}$. So $\mathbf{E} = \mathbf{K}[r]$. As a consequence of the Dimension Formula, we find that $\mathbf{E} = \mathbf{K}[t]$ as well. Putting this together with the Fact proved just above, we get

$$|\operatorname{Gal}\operatorname{Inv}\mathbf{H}| = |\operatorname{Gal}(\mathbf{E}/\mathbf{K})| = [\mathbf{E} : \mathbf{K}] = [\mathbf{H} : \mathbf{Stab}\, t] \leq |H|.$$

But we know $\mathbf{H} \leq \operatorname{Gal}\operatorname{Inv}\mathbf{H}$ and that these groups are finite. Therefore $\mathbf{H} = \operatorname{Gal}\operatorname{Inv}\mathbf{H}$, as desired. □

13.2 The Fundamental Theorem of Galois Theory

It is traditional to gather together the bits and pieces up to this point and package them into one theorem. Here it is.

The Fundamental Theorem of Galois Theory. *Let* $\mathbf{E}$ *be a Galois extension of* $\mathbf{F}$. *Then the following hold.*

(a) *The closed sets on the field side of Galois' connection are exactly the fields intermediate between* $\mathbf{F}$ *and* $\mathbf{E}$.

(b) *The closed sets of the group side of Galois' connection are exactly the subgroups of the Galois group* $\operatorname{Gal}(\mathbf{E}/\mathbf{F})$.

(c) *Polarities of Galois' connection, namely* Inv *and* Gal, *are inverses of each other and establish an anti-isomorphism between the two lattices of closed sets.*

(d) $[\mathbf{E} : \mathbf{K}] = |\operatorname{Gal}\mathbf{K}|$ *and* $[\mathbf{K} : \mathbf{F}] = [\operatorname{Gal}\mathbf{F} : \operatorname{Gal}\mathbf{K}]$, *for each intermediate field* K. *In particular,* $|\operatorname{Gal}(\mathbf{E}/\mathbf{F})| = [\mathbf{E} : \mathbf{F}]$.

(e) *Let* $\mathbf{H}$ *be any subgroup of* $\operatorname{Gal}(\mathbf{E}/\mathbf{F})$. *Then* $\mathbf{H} \lhd \operatorname{Gal}(\mathbf{E}/\mathbf{F})$ *if and only if* $\operatorname{Inv}\mathbf{H}$ *is a normal extension of* $\mathbf{F}$. *In this case,* $\operatorname{Gal}(\operatorname{Inv}\mathbf{H}/\mathbf{F}) \cong \operatorname{Gal}(\mathbf{E}/\mathbf{F})/\mathbf{H}$.

Proof. The only parts that need attention, perhaps, are (d) and (e).

For (d), notice that $\operatorname{Gal}\mathbf{K} = \operatorname{Gal}(\mathbf{E}/\mathbf{K})$. We know that $\mathbf{E}$ is a Galois extension of $\mathbf{K}$, so by the Fact immediately preceding the Galois Group Closure Theorem, we see $|\operatorname{Gal}\mathbf{K}| = [\mathbf{E} : \mathbf{K}]$ as well as $|\operatorname{Gal}\mathbf{F})| = [\mathbf{E} : \mathbf{F}]$. The Dimension Formula tells us

$$[\mathbf{E} : \mathbf{F}] = [\mathbf{E} : \mathbf{K}][\mathbf{K} : \mathbf{F}]$$

and Lagrange tells us

$$|\operatorname{Gal}\mathbf{F}| = [\operatorname{Gal}\mathbf{F} : \operatorname{Gal}\mathbf{K}]|\operatorname{Gal}\mathbf{K}|.$$

A bit a twiddling extracts $[\mathbf{K} : \mathbf{F}] = [\operatorname{Gal}\mathbf{F} : \operatorname{Gal}\mathbf{K}]$ from these equations. This secures (d).

For (e), suppose first the $\mathbf{H}$ is a normal subgroup of $\operatorname{Gal}(\mathbf{E}/\mathbf{F})$. Let $s \in \operatorname{Inv}\mathbf{H}$. We need to see that the minimal polynomial $f(x)$ of s splits in $\operatorname{Inv}\mathbf{H}$. Now $f(x)$ certainly splits in $\mathbf{E}$ since $\mathbf{E}$ is a normal extension of $\mathbf{F}$. Let $r \in E$ be a root of $f(x)$. What we need is to show that $r \in \operatorname{Inv} H$. Relying on Kronecker, we pick $\sigma \in \operatorname{Gal}(\mathbf{E}/\mathbf{F})$ so that $\sigma(s) = r$. So we must show that $\tau(r) = r$ for

every $\tau \in H$. But $\sigma H \sigma^{-1} = H$ by normality of the subgroup. This means what we have to show is $\sigma \circ \tau \circ \sigma^{-1}(r) = r$. But this is immediate:

$$\sigma \circ \tau \circ \sigma^{-1}(r) = \sigma(\tau(\sigma^{-1}(r))) = \sigma(\tau(s)) = \sigma(s) = r.$$

So we see that Inv$\mathbf{H}$ is a normal extension of $\mathbf{F}$.

Now suppose that Inv$\mathbf{H}$ is a normal extension of $\mathbf{F}$. Let $\sigma \in \text{Gal}(\mathbf{E}/\mathbf{F})$ and let $r \in \text{Inv}\mathbf{H}$. Let $f(x)$ be the minimal polynomial of r over $\mathbf{F}$. Then $\sigma(r)$ must also be a root of $f(x)$. But $f(x)$ splits in Inv$\mathbf{H}$. So $\sigma(r) \in \text{Inv}\mathbf{H}$. This means that the restriction $\sigma \restriction_{\text{Inv}\mathbf{H}}$ belongs to Gal(Inv$\mathbf{H}/\mathbf{F}$). Now define $\Phi : \text{Gal}(\mathbf{E}/\mathbf{F}) \to \text{Gal}(\text{Inv}\mathbf{H}/\mathbf{F})$ via

$$\Phi(\sigma) = \sigma \restriction_{\text{Inv}\mathbf{H}}, \text{ for all } \sigma \in \text{Gal}(\mathbf{E}/\mathbf{F}).$$

The eager graduate students will find it easy to show that Φ is a homomorphism onto the group Gal(Inv$\mathbf{H}/\mathbf{F}$) and that its kernel is $\mathbf{H}$ (because $\mathbf{H}$ is closed). So we see that $\mathbf{H}$ is a normal subgroup of Gal($\mathbf{E}/\mathbf{F}$) and, by the Homomorphism Theorem, that

$$\text{Gal}(\mathbf{E}/\mathbf{F})/\mathbf{H} \cong \text{Gal}(\text{Inv}\mathbf{H}/\mathbf{F}),$$

as desired. □

LECTURE 14

GALOIS' CRITERIA FOR SOLVABILITY BY RADICALS

Given a field **F** and a polynomial $f(x) \in \mathbf{F}[x]$ the task of explicitly discribing, in some manner, all the roots of $f(x)$ is a project that is most appropriate carried forward in the splitting field of the polynomial. So let **E** be the splitting field of $f(x)$ over **F**. The set of all roots of $f(x)$ is a finite set, so it would be possible to simply make a list of all these elements. But it is not apparent, just given $f(x)$ how such a list might be made. Just trying to use the field operations and the coefficients of $f(x)$ is not even adequate for describing all the roots of x^-2. By permitting the extracting of square roots, we can resolve this case and that of all polynomials of degree no more than 2. By allowing the extraction of cube roots, fourth roots, and so on, one might hope to succeed, at least with some frequency. The problem of determining when success is possible is what Galois undertook.

Recall how we approached the notion of a constructible number. We envisioned a tower of field extensions so that later fields in the tower were obtained by adjoining a square root to an earlier field. We simply expand our horizons by allowing the adjunction of k^{th} roots for any positive integer k. More precisely, we say that

$$\mathbf{F} = \mathbf{F}_0 \leq \mathbf{F}_1 \leq \cdots \leq \mathbf{F}_{m-1}$$

is a **radical tower** over **F** provided for all $i < m$

$$\mathbf{F}_{i+1} = \mathbf{F}_i[r] \text{ for some } r \text{ such that } r^k \in F_i \text{ for some positive integer } k.$$

We will say that **K** is a **radical extension** of **F** provided **K** extends **F** and $\mathbf{K} \leq \mathbf{F}_{m-1}$ for some radical tower $\mathbf{F}_0 \leq \mathbf{F}_1 \leq \cdots \leq \mathbf{F}_{m-1}$ over **F**.

We say a polynomial $f(x) \in \mathbf{F}[x]$ is **solvable by radicals** over **F** exactly when the splitting field of $f(x)$ over **F** is a radical extension of **F**. By the **Galois group** of $f(x)$ we mean the group Gal(**E**/**F**), where **E** is the splitting field of $f(x)$ over **F**.

Galois's Criterion for Solvabiity by Radicals. *Let* **F** *be a field of characteristic* 0 *and let* $f(x) \in \mathbf{F}[x]$. *The polynomial* $f(x)$ *is solvable by radicals over* **F** *if and only if the Galois group of* $f(x)$ *over* **F** *is a solvable group.*

We need a few preliminaries to prepare the way.

Lemma 14.0.1. *Over any field* **F***, the polynomial* $x^p - a$*, with* p *a prime number and* $a \in F$*, either has a root in* F *or it is irreducible over* **F**.

Proof. This is clear if $a = 0$, so we consider that $a \neq 0$. There are two cases.

Case: p **is not the characteristic of F.**
In this case we know that $x^p - a$ must have distinct roots. Let **E** be the splitting field of $x^p - a$ over **F**. So

$$x^p - a = (x - r_0)(x - r_1)\dots(x - r_{p-1})$$

where $r_0, \dots, r_{p-1}$ are the p distinct roots of $x^p - a$. Notice that $r_0 \neq 0$. So

$$1 = \frac{r_0}{r_0}, \quad \frac{r_1}{r_0}, \dots, \frac{r_{p-1}}{r_0}$$

must be the p distinct p^{th} of unity. Let ζ be a primitive p^{th} of unity in E. This means that

$$r_0, \zeta r_0, \zeta^2 r_0, \dots, \zeta^{p-1} r_0$$

are the roots of $x^p - a$.

Now consider the possibility that $x^p - a$ is reducible in $\mathbf{F}[x]$. We desire to show that $x^p - a$ has a root in F. For some k with $1 \leq k < p$ we can render a factor of $x^p - a$ as a product of k factors, each of the form $x - \zeta^j r_0$. computing the constant term of this product, we find $\zeta^\ell r_0^k \in F$ for some ℓ. Put $b = \zeta^\ell r_0^k$. Now $b^p = \zeta^{p\ell} r_0^p k = a^k$. Since $1 \leq k < p$ and p is a prime number, we see that k and p are relatively prime. Pick integers s and t so that $1 = sk + tp$. We get

$$a = a^1 = a^{sk+tp} = (a^k)^s (a^t)^p = (b^p)^s (a^t)^p = (b^s a^t)^p.$$

This means that a has a p^{th} in F, namely $b^s a^t$. Hence $x^p - a$ has a root in F. This finishes the first case.

Case: F has characteristic p**.**
In the the splitting field **E** pick a root r of $x^p - a$. It follows that $x^p - a = x^p - r^p = (x - r)^p$. Consider the case that $x^p - a$ is reducible in $\mathbf{F}[x]$. This means that for some k with $1 \leq k < p$ we wil have $(x - r)^k \in \mathbf{F}[x]$. Computing the constant term, we find $r^k \in F$. Put $b = r^k$. Hence $b^p = (r^k)^p = (r^p)^k = a^k$. As above, we have integers s and t so that $1 = sk + tp$. Just as above, we have $a = (b^s a^t)^p$. This makes $b^s a^t$ a root of $x^p - a$, as desired. □

Theorem 14.0.2. *Let* p *be a prime that is not the characteristic of* **F** *and let* $a \in F$*. The Galois group of* $x^p - a$ *over* **F** *is solvable.*

Proof. Let **E** be the splitting field of $x^p - a$ over **F**. As in the proof of the lemma above, **E** has p distinct p^{th} roots of unity. Let ζ be a primitive one. Let r be any root of $x^p - a$. The we have seen that $\mathbf{E} = \mathbf{F}[\zeta, r]$. We also know that $x^p - a$ is separable. So the Fundamental Theorem of Galois Theory applies here.

The field $\mathbf{F}[\zeta]$ is the splitting field of the separable polynomial $x^p - 1$ over **F**. So $\mathbf{F}[\zeta]$ is a normal extension of **F**. By the Fundamental Theorem, Gal($\mathbf{E}/\mathbf{F}[\zeta]$ is a normal subgroup of Gal($\mathbf{E}/\mathbf{F}$) and

$$\text{Gal}(\mathbf{F}[\zeta]/\mathbf{F}0 \cong \text{Gal}(\mathbf{E}/\mathbf{F}/\text{ Gal}(\mathbf{E}/\mathbf{F}[\zeta]).$$

Observe that every automophism belonging to Gal($\mathbf{F}[\zeta]/\mathbf{F}$) is determined by what it does to ζ (which it must map to another root of unity). Thus restriction is actually an embedding of Gal($\mathbf{F}[\zeta]/\mathbf{F}$) into the group of automorphisms of the groups of p^{th} roots of unity. But the group of p^{th} roots of unity is just a copy of the cyclic group $\mathbb{Z}_p$. It is an exercise (to be carried out by the diligent graduate students) to prove that $\operatorname{Aut}\mathbb{Z}_p \cong \mathbb{Z}_{p-1}$. In this way we see that Gal($\mathbf{F}[\zeta]/\mathbf{F}$) is embeddable into the cyclic group $\mathbb{Z}_{p-1}$. But every subgroup of a cyclic group is cyclic, so Gal($\mathbf{F}[\zeta]/\mathbf{F}$) is cyclic.

Now consider the group Gal($\mathbf{E}.\mathbf{F}[\zeta]$). In case $x^p - a$ has a root is $\mathbf{F}[\zeta]$, then all its roots are in $\mathbf{F}[\zeta]$. This means $\mathbf{E} = \mathbf{F}[\zeta, r] = \mathbf{F}[\zeta]$. So Gal($\mathbf{E}/\mathbf{F}[\zeta]$) is a trivial group. In case $x^p - a$ has no root in $\mathbf{F}[\zeta]$ we know by the lemma that $x^p - a$ is irreducible over $\mathbf{F}[\zeta]$. So by Kronecker, $p = [\mathbf{F}[\zeta, r] : \mathbf{F}[\zeta]] = |\text{Gal}(\mathbf{E}/\mathbf{F}[\zeta])|$. This means that Gal($\mathbf{E}/\mathbf{F}[\zeta]$) is a cyclic group of order p.

So the normal series Gal($\mathbf{E}/\mathbf{F}$) $\rhd$ Gal($\mathbf{E}/\mathbf{F}[\zeta]$) $\rhd 1$ has cyclic factor groups. Therefore Gal($\mathbf{E}/\mathbf{F}$) is solvable. □

Lemma 14.0.3. *Let p be a prime number and suppose that the field $\mathbf{F}$ contains p distinct p^{th} roots of unity. Let $\mathbf{K}$ extend $\mathbf{F}$ so that $[\mathbf{K} : \mathbf{F}] = p$ and so that* Gal($\mathbf{K}/\mathbf{F}$) *is cyclic of order p. Then $\mathbf{K} = \mathbf{F}[d]$ for some d such that $d^p \in F$.*

Proof. Let η generate the cyclic group Gal($\mathbf{K}/\mathbf{F}$) and let ζ be a primitive p^{th} root of unity in F.

Begin by picking $c \in K \setminus F$. Then $\mathbf{K} = \mathbf{F}[c]$ because since p is prime, by the Dimension Formula there can be no fields properly intermediate between $\mathbf{F}$ and $\mathbf{K}$.

For each $i < p$, put $c_i = \eta^i(c)$. So we get

$$\begin{aligned} c_0 &= c \\ c_{i+1} &= \eta(c_i) \text{ for all } i < p-1 \\ c_0 &= \eta(c_{p-1}) \end{aligned}$$

Put

$$d_i = c_0 + c_1\zeta^i + c_2\zeta^{2i} + \cdots + c_{p-1}\zeta^{(p-1)i} \text{ for } i < p. \qquad (\star)$$

A straightforward computation shows $eta(d_i) = \zeta^{-i} d_i$ for all $i < p$. Hence $eta(d_i^p) = (\eta(d_i))^p = ((\zeta^{-i} d_i)^p = 1 \cdot d_i^p$ for all $i < p$. Since the generator of Gal($\mathbf{E}/\mathbf{F}$) fixes each d_i^p, we find that each d_i^p belongs to the fixed field, namely to F. (The fixed field must be $\mathbf{F}$, for lack of other intermediate fields.)

It remains to show that $d_i \notin F$ for some i, for then we can take that d_i to be our desired d. Let us render the system $(\star)$ of equations in matrix form.

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta & zeta^2 & \dots & \zeta^{p-1} \\ 1 & \zeta^2 & \zeta^4 & \dots & \zeta^{2(p-1)} \\ 1 & \zeta^3 & \zeta^6 & \dots & \zeta^{3(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta^{p-1} & \zeta^{2(p-1)} & \dots & \zeta^{(p-1)(p-1)} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{p-1} \end{pmatrix} = \begin{pmatrix} d_0 \\ d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_{p-1} \end{pmatrix}$$

The $p \times p$ matrix displayed above is invertible, since it is a Vandermonde matrix. This means that the column vector of c_i's can be obtained by multiplying the column vector of d_i's by the inverse of the Vandermonde matrix—which is a matrix over $\mathbf{F}$. In particular, this means that $c = c_0$ is an $\mathbf{F}$-linear combination of the d_i's. Since $c \notin F$, we see that at least one of the d_i's must also fail to be in F. This completes the proof of the lemma. □

We need on more lemma.

Lemma 14.0.4. *Let* $\mathbf{F}$ *be a field of characteristic* 0. *Any radical extension of* $\mathbf{F}$ *can be embedded into a separable normal radical extension of* $\mathbf{F}$ *that has a solvable Galois group over* $\mathbf{F}$.

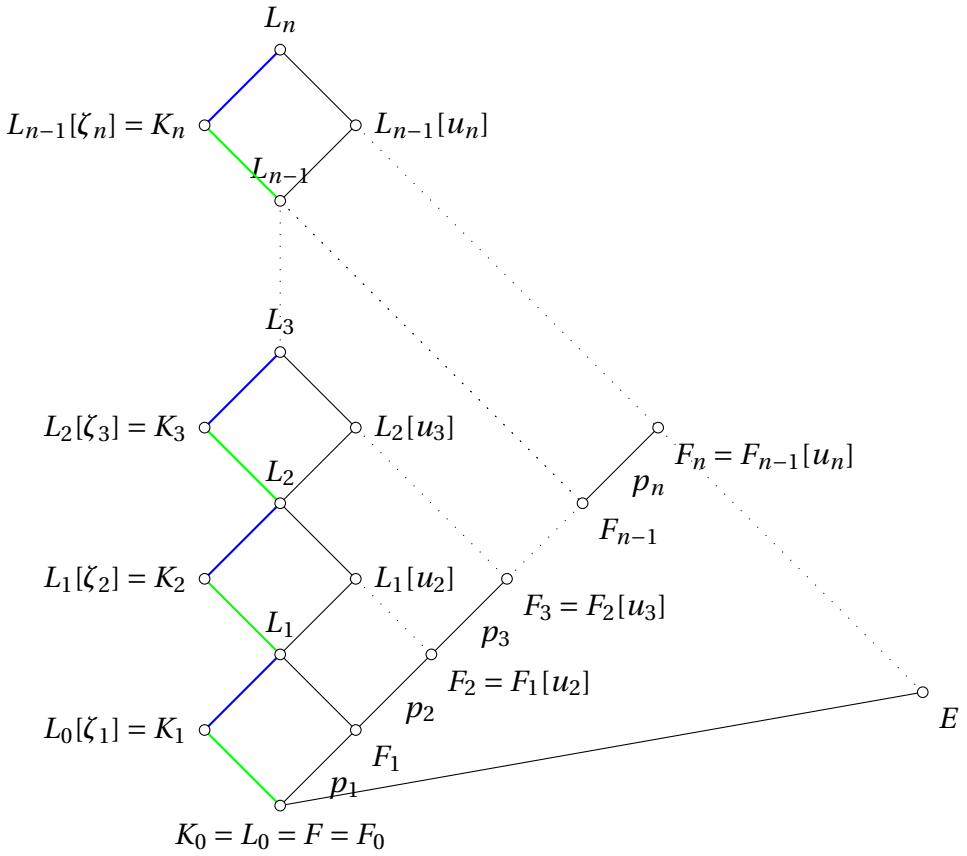
Proof. We do this by induction on the number of intermediate fields in the root tower leading to the radical extension. The base step (when the number of intermediate fields in 0) is evident. So consider the inductive step. Suppose the last step in the given radical tower in $\mathbf{F}_k < \mathbf{F}_k[u_k]$ where $u_k^p \in F_k$ for some prime p. By the inductive hypothesis, we suppose that we have in hand $\mathbf{L}_k$,which is a normal separable radical extension of $\mathbf{F}$ such that $\mathbf{L}_k$ extends $\mathbf{F}_k$, and that $\mathrm{Gal}(\mathbf{L}_k/\mathbf{F})$ is solvable. To conserve notation, let $\mathbf{G} = \mathrm{Gal}(\mathbf{L}_k/\mathbf{F})$ and let $G = \{\sigma_0, \ldots, \sigma_{m-1}\}$. Put $a = u_k^p$. Let

$$f(x) = \prod_{\sigma \in G} (x^p - \sigma(a)).$$

Observe that the coefficients of $f(x)$ are fixed by every $\tau \in G$. Since $\mathbf{F}$ is the fixed field of $\mathbf{G}$, we see that $f(x) \in F[x]$. Let $\mathbf{L}_{k+1}$ be the splitting field of $f(x)$ over $\mathbf{L}_k$. Now, as we saw several times above, each $x^p - \sigma(a)$ is a separable polynomial, so in L_{k+1} we must have a primitive p^{th} root ζ of unity. For each $i < m$ pick $r_i \in L_{k+1}$ that is a root of $x^p - \sigma_i(a)$. This entails that $\mathbf{L}_{k+1} = \mathbf{L}_k[\zeta, r_0, \ldots, r_{m-1}]$. Now consider the following root tower of fields:

$$\mathbf{F} \leq \mathbf{L}_k \leq \mathbf{L}_k[\zeta] \leq \mathbf{L}_k[\zeta, r_0] \leq \cdots \leq \mathbf{L}_k[\zeta, r_0, \ldots, r_{m-1}] = \mathbf{L}_{k+1}.$$

The field obtained at each step above $\mathbf{L}_k$ is a normal extension of the previous field. Moreover, the Galois groups associated to each step are cyclic. This entails that the associated series of subgroups of $\mathbf{G}$ above $\mathrm{Gal}(\mathbf{E}/\mathbf{L}_k)$ have cyclic factors. Since we also have that $\mathrm{Gal}(\mathbf{L}_k, \mathbf{F})$ is solvable, it follows that $\mathrm{Gal}(\mathbf{L}_{k+1}/\mathbf{F})$ is solvable. □



Normalizing a Root Tower over E

Here ζ_i is a primitive p_i^{th} root of unity and $u_i^{p_i} \in F_{i-1}$. The blue lines represent root towers in their own right along which only p_i^{th} roots are extracted, where p_i is the prime labelling the parallel edge in the original root tower. Each L_i is a separable normal extension of F. The zigzag path along the left edge of the diagram is a root tower.

Now we are ready.

Proof of Galois' Criterion. First, let us suppose that $f(x)$ is solvable by radicals. Let $\mathbf{E}$ be the splitting field of $f(x)$ over $\mathbf{F}$. We aim to show that $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$ is a solvable group. Use the lemma just above to obtain a field $\mathbf{L}$ that is a separable normal radical extension of $\mathbf{F}$ that also extends $\mathbf{E}$. Observe $\mathrm{Gal}(\mathbf{L}/\mathbf{E}) \lhd \mathrm{Gal}(\mathbf{L}/\mathbf{F})$, since $\mathbf{E}$ is a normal extension of $\mathbf{F}$. Moreover,

$$\mathrm{Gal}(\mathbf{E}/\mathbf{F}) \cong \mathrm{Gal}(\mathbf{L}/\mathbf{F}) / \mathrm{Gal}(\mathbf{L}/\mathbf{F}).$$

This means that $\mathrm{Gal}(\mathbf{E}/\mathbf{F}$ is a homomorphic image of the solvable group $\mathrm{Gal}(\mathbf{L}/\mathbf{F})$. Hence, $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$ is solvable.

For the converse, suppose that $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$ is solvable. Let $n = |\mathrm{Gal}(\mathbf{E}/\mathbf{F})| = [\mathbf{E} : \mathbf{F}]$. Let $r_0, \ldots, r_{m-1}$ be the roots of $f(x)$ is E. Let ζ be a primitive n^{th} root of unity. (It might not be in E.) Observe that $\mathbf{E}[\zeta]$ is a splitting field of $f(x)$ over $\mathbf{F}[\zeta]$. Let $\eta \in \mathrm{Gal}(\mathbf{E}[\zeta]/\mathbf{F}[\zeta])$. Then η will permute the elements of $\{r_0, \ldots, r_{m-1}\}$. The elements generate $\mathbf{E}$ over $\mathbf{F}$. This means that restricting η to E produces a member of $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$. Indeed, an argument routine by now shows that restriction to E

is an embedding of $\mathrm{Gal}(\mathbf{E}[\zeta]/\mathbf{F}[\zeta])$ into $\mathrm{Gal}(\mathbf{E}/\mathbf{F})$. This latter group is solvable. Since subgroups of solvable groups are themselves solvable, we find that $\mathrm{Gal}(\mathbf{E}[\zeta]/\mathbf{F}[\zeta])$ is solvable. Let

$$G(\mathbf{E}[\zeta]/\mathbf{F}[\zeta]) = \mathbf{H}_0 \rhd \mathbf{H}_1 \rhd \cdots \rhd \mathbf{H} + \ell - 1$$

be a composition series. So its factor groups must be cyclic of prime order. Let

$$\mathbf{F} \leq \mathbf{F}[\zeta] = \mathbf{K}_0 \leq \mathbf{K}_1 \leq \cdots \leq \mathbf{K}_{\ell-1} = \mathbf{E}[\zeta]$$

be the corresponding tower of fixed fields. By our lemmas, each step of this tower is made by adding a k^{th} root of an element of the previous field, for some k. This means that $\mathbf{E}[\zeta]$ is a radical extension of $\mathbf{F}$. Since $\mathbf{E} \leq \mathbf{E}[\zeta]$, we see that $f(x)$ is solvable by radicals. □

LECTURE 15

POLYNOMIALS AND THEIR GALOIS GROUPS

In order to take advantage of Galois's Criterion for Solvability by Radicals we need at least some way to start with a polynomial $f(x)$ and finds its Galois group. With the group in hand, we may be able to determine whether it is a solvable group. Such a group is, after all, finite and even if no more elegant approach is at hand it would be possible to undertake, with computational assistance the brute force examination of its subgroup lattice. At any rate we see a two step process:

- Given $f(x)$ construct its Galois group $\mathbf{G}$.
- Given a finite group $\mathbf{G}$ determine whether it is solvable.

Even over the field of rational numbers the situation has been the focus of very considerable mathematical effort and is still not well understood.

Here we will see just three results: The determination of which symmetric groups $\mathbf{S}_n$ are solvable and two conditions, each sufficient to ensure that the Galois group of $f(x)$ is not solvable.

The following theorem is due to Galois.

The Solvability of Symmetric Groups. *The groups* $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3$, *and* $\mathbf{S}_4$ *are solvable. The groups* $\mathbf{S}_n$ *where* $4 < n$ *are not solvable.*

Proof. The groups $\mathbf{S}_1$ and $\mathbf{S}_2$ have one and two elements respectively. They are evidently solvable. Observe the $\mathbf{S}_3 \rhd \mathbf{A}_3$ (since $[\mathbf{S}_3 : \mathbf{A}_3] = 2$) and that $\mathbf{A}_3$ is the three element group, which is simple and Abelian . This gives us a normal series with Abelian factors, witnessing the solvability of $\mathbf{S}_3$. We can try the same thing with $\mathbf{S}_4$:

$$\mathbf{S}_4 \rhd \mathbf{A}_4 \rhd \dots$$

Now $\mathbf{A}_4$ has twelve elements. Let $\mathbf{V}$ be the subgroup of $\mathbf{A}_4$ consisting of the identity permutation and the following three permutations:

$$(0,1)(2,3),\ (0,2)(1,3),\ \text{and}\ (0,3)(1,2).$$

Direct calculation reveals that there elements constitute an Abelian subgroup of $\mathbf{A}_4$ and that this subgroup is normal. (It is the Sylow 2-subgroup of $\mathbf{A}_4$. So we see

$$\mathbf{S}_4 \rhd \mathbf{A}_4 \rhd \mathbf{V} \rhd \mathbb{1}$$

is a normal series with Abelian factors, witnessing that $\mathbf{S}_4$ solvable.

To see, on the other hand, that $\mathbf{S}_n$ is not solvable when $n > 4$ we will show that $\mathbf{A}_n$ is simple: it has no proper nontrivial normal subgroups. This will mean that $\mathbf{S}_n \triangle \mathbf{A}_n \triangleright \mathbb{1}$ is a composition series. Since $\mathbf{A}_n$ is not Abelian this will demonstrate that $\mathbf{S}_n$ is not solvable.

So let $\mathbf{N} \triangleleft \mathbf{A}_n$ and suppose that $\mathbf{N}$ is nontrivial. We have to prove that $\mathbf{N} = \mathbf{A}_n$. We use the following fact:

Fact. $\mathbf{A}_n$ is generated by the set of all 3-cylces, if $3 \leq n$.

The verification of the fact is left as an entertainment for graduate students.

Let us first see that $\mathbf{N}$ must have at least one 3-cycle. Each element of $\mathbf{A}_n$ is a permutation of $\{0,1,2,3,4,5,\ldots,n-1\}$. A permutation might have fixed points. (The identity fixes all n points.) Let $\alpha \in A_n$ fix as many points as possible while still being different from the identity permutation. Consider the decomposition of α into a product of disjoint cycles. There are two cases.

Let us first suppose that α is a product of disjoint transpositions. It does no harm to suppose

$$\alpha = (0,1)(2,3)\cdots.$$

Let $\beta = (2,3,4) \in A_n$. Notice that $\beta\alpha\beta^{-1} \in N$ by normality. So $\beta\alpha\beta^{-1}\alpha^{-1} \in N$ as well. Direct computations show that 0 and 1 are fixed points of $\beta\alpha\beta^{-1}\alpha^{-1}$. The only points moved by β are 2, 3, and 4. Now any point that is fixed by both α and β is fixed by $\beta\alpha\beta^{-1}\alpha^{-1}$. It might be that 4 is a fixed point of α. In that event, it would not be fixed by $\beta\alpha\beta^{-1}\alpha^{-1}$. But in any event, $\beta\alpha\beta^{-1}\alpha^{-1}$ has at least one more fixed point than α. This is contrary to the choice of α, so we reject this case.

This means that the decomposition of $alpha$ into a product of disjoint cycles must contain a cycle of length exceeding 2. So we have, without loss of generality,

$$\alpha = (0,1,2,\ldots)\cdots.$$

If $\alpha = (0,1,2)$ then we have found the 3-cycle we desire. So suppose, to the contrary, that α moves some other point. Without loss of generality let it be 3. Now $(0,1,2,3)$ is odd, so it cannot be α. So it does no harm to suppose that $\alpha(3) = 4$. Notice that α also moves 4. Let $\beta = (2,3,4)$ and as above observe that $\beta\alpha\beta^{-1}\alpha^{-1} \in N$ and it fixes every point fixed by $alpha$. Direct computation shows that $\beta\alpha\beta^{-1}\alpha^{-1}$ fixes 1. This violates the choice of α. So we reject the notion that α moves any other points apart from 0, 1, and 2. That is $\alpha = (0,1,2)$ and we find that N contains a 3-cycle.

Now, suppose that i, j, k, ℓ, and m are distinct elements of $\{0,1,2,3,4,\ldots\}$. Let γ be the permutation so that

$$\begin{aligned}
\gamma(0) &= i \\
\gamma(1) &= j \\
\gamma(2) &= k \\
\gamma(3) &= \ell \\
\gamma(4) &= m \\
&\vdots
\end{aligned}$$

Now either γ is even or $(\ell, m)\gamma$ is even. Let λ be the one of these that is even. Then direct calculation shows $\lambda\alpha\lambda^{-1} = (i, j, k)$. This means that N contains all the 3-cylces.

So we conclude that $\mathbf{A}_n$ is simple and that $\mathbf{S}_n$ is not solvable. □

So how can we ensure that a polynomial $f(x)$ have Galois group $\mathbf{S}_n$?

The $\mathbf{S}_p$ Criterion over $\mathbb{Q}$. *Let p be a prime number and let $f(x) \in \mathbb{Q}[x]$. If $f(x)$ is irreducible and $f(x)$ has exactly two non-real roots in $\mathbb{C}$, then the Galois group of $f(x)$ over $\mathbb{Q}$ is $\mathbf{S}_p$ and $f(x)$ is not solvable by radicals over $\mathbb{Q}$.*

Proof. Let $r_0, \ldots, r_{p-3}$ be the real roots of $f(x)$ and let r_{p-2} and r_{p-1} be the non-real complex roots. Let $\mathbf{E} = \mathbb{Q}[r_0, \ldots, r_{p-1}]$ be the splitting field of $f(x)$. By Kronecker we know that $[\mathbb{Q}[r_0] : \mathbb{Q}] = p$. By the Dimension Formula, we see that $p \mid [\mathbf{E} : \mathbb{Q}]$. By the Fundamental Theorem of Galois Theory, this means $p \mid |\mathrm{Gal}(\mathbf{E}/\mathbb{Q})|$. By Cauchy, this Galois group must have an element of order p. On the other hand, complex conjugation is an automorphism of $\mathbb{C}$ that fixes every real. In particular, all the coefficients of $f(x)$ are fixed by conjugation. So conjugation must permute the roots of $f(x)$. This entails that r_{p-1} is the complex conjugate of r_{p-2}. By restricting our group to p-element set $\{r_0, \ldots, r_{p-1}\}$ we see that this subgroup of $\mathbf{S}_p$ has an element of order p and a transposition (inherited from complex conjugation. It is a fact for the entertainment of graduate students that for any prime p, that $\mathbf{S}_p$ is generated by any transposition and any element of order p. So the Galois group of $f(x)$ is isomorphic to $\mathbf{S}_p$. □

It is easy to devise polynomials of prime degree with integer coefficients that meet these criteria. For example, suppose we want $f(x)$ to be of degree 5. We want it to have 3 real roots and 2 non-real complex roots (which we know must be complex conjugates. So we could just pick any three real numbers r_0, r_1, and r_2 and any non-real complex number s and let

$$f(x) = (x - r_0)(x - r_1)(x - r_2)(x - s)(x - \bar{s}).$$

But we have two problems: the coefficients of $f(x)$ might not be rational and even if they were $f(x)$ might not be irreducible over $\mathbb{Q}$. For this approach to work, r_0, r_1, r_2, and s must at least be algebraic and $f(x)$ must be their common minimal polynomial. This is harder to arrange, but still possible.

But there is another approach advanced by Richard Brauer. The idea is to use the curve sketching techniques of freshman calculus. The graph of $f(x)$ should cross the X-axis exactly 3 times. This can be arranged if $f(x)$ has a unique relative maximum (and $f(x)$ as a positive value there) and a unique relative minimum (and $f(x)$ has a negative value there). This suggests looking at the derivative $f'(x)$. This derivative will have degree 4 and we want it to have 2 real roots. Lets try

$$f'(x) = 5x^4 - 5 \cdot 16 = 5(x^2 + 4)(x^2 - 4) = 5(x^2 + 4)(x - 2)(x + 2)$$

This means that $f(x) = x^5 - 5 \cdot 16x + c$ where c is the constant of integration but is subject to the following constraints:

$$\begin{aligned} 0 < f(-2) &= -2^5 + 5 \cdot 2^5 + c = 2^7 + c \\ 0 > f(2) &= 2^5 - 5 \cdot 2^5 + c = -2^7 + c \\ f(x) &= x^5 - 5 \cdot 16x + c \text{ is irreducible over } \mathbb{Q}. \end{aligned}$$

The first two constraints reduce to $-2^7 < c < 2^7$. This is a comfortably sized range. With Eisenstein's Criteria in mind, pick $c = 5$ (and there are other obvious choices). We find

$$f(x) = x^5 - 80x + 5$$

is a polynomial of degree 5 that cannot be solved by radicals.

What about the possibility that the Galois group of our polynomial is only a subgroup of $\mathbf{S}_n$? To put it another way, what can we say about the solvable subgroups of $\mathbf{S}_n$ that are Galois groups? One thing that we can recall from Kronecker is that given any two roots of an irreducible polynomial, there will be an automorphism in the Galois group that takes one root to another. We could frame this property for any subgroup of $\mathbf{S}_n$. We will say that any subgroup $\mathbf{G}$ of $\mathbf{S}_n$ is **transitive** provided for any $i, j \in \{0, 1, 2, \ldots, n-1\}$ there is $\sigma \in G$ so that $\sigma(i) = j$. Then, according to Kronecker, the Galois group of any separable irreducible polynomial of degree n will be (isomorphic to) a transitive subgroup of $\mathbf{S}_n$. What can we say about the transitive solvable subgroups of $\mathbf{S}_n$?

Theorem on the Transitive Solvable Subgroups of $\mathbf{S}_p$. *Let p be a prime number and let $\mathbf{G}$ be a transitive solvable subgroup of $\mathbf{S}_p$. Then every $\sigma \in G$, except the identity, has no more than one fixed point.*

The proof of this theorem relies on two lemmas that are of some interest in their own right.

Lemma on Normal Subgroups of Transitive Groups. *Let p be a prime number, let $\mathbf{G}$ be a transitive subgroup of $\mathbf{S}_p$, and let $\mathbf{N}$ be a nontrivial normal subgroup of $\mathbf{G}$. Then $\mathbf{N}$ is transitive.*

Proof. The group $\mathbf{N}$ induces a partition of $\{0, 1, \ldots, p-1\}$ into orbits. I contend that all the orbits have the same size. To see this let $\mathcal{O}$ and $\mathcal{Q}$ be any two orbits. Pick elements $a \in \mathcal{O}$ and $b \in \mathcal{Q}$ and, since $\mathbf{G}$ is transitive, pick $\beta \in G$ so that $\beta(a) = b$. Let $c \in \mathcal{O}$. Pick $\sigma \in N$ so that $\sigma(a) = c$. Then observe that

$$\beta(c) = \beta(\sigma(a)) = \beta(\sigma(\beta^{-1}(b))) = (\beta \circ \sigma \circ \beta^{-1})(b).$$

Since $\mathbf{N}$ is a normal subgroup of $\mathbf{G}$, we see that $\beta(c) \in \mathcal{Q}$. So β induces a map from $\mathcal{O}$ into $\mathcal{Q}$. A similar argument shows that β^{-1} induces a map from $\mathcal{Q}$ into $\mathcal{O}$. There induced maps invert each other, so the two orbits are the same size.

Since $\mathbf{N}$ is nontrivial, there must be a nontrivial orbit. So all orbits have the same size $k > 1$. But our set with p elements is partitioned into sets of size k. So $k \mid p$. Since p is prime, we have $k = p$. That is, there is only one orbit. This means $\mathbf{N}$ is transitive. □

Lemma on p-cycles in Solvable Transitive Subgroups of $\mathbf{S}_p$. *Let p be a prime number and $\mathbf{G}$ be a nontrivial transitive solvable subgroup of $\mathbf{S}_p$. The last nontrivial group in any composition series of $\mathbf{G}$ is a cyclic group of order p and every p-cycle in G belongs to this cyclic group.*

Proof. Let $\mathbf{G}$ is a solvable transitive subgroup of $\mathbf{S}_p$, where p is prime. Consider a composition series for $\mathbf{G}$.

$$\mathbf{G} = \mathbf{G}_0 \rhd \mathbf{G}_1 \rhd \cdots \rhd \mathbf{G}_{n-1} \rhd \mathbf{G}_n$$

so that $\mathbf{G}_n$ is trivial, $\mathbf{G}_{n-1}$ is not trivial, and $\mathbf{G}_k/\mathbf{G}_{k+1}$ is of prime order for all $k < n$. By the lemma above, $\mathbf{G}_k$ is transitive for each $k < n$. Notice that $\mathbf{G}_{n-1}$ is a cyclic group of prime order q. Let σ generate this group. Write the permutation σ as a product of disjoint cycles. These cycles must all have length q since q is prime. Moreover, every power of σ is the product of the powers of these disjoint cycles. So if there were more than one cycle in the decomposition of σ we would have that $\mathbf{G}_{n-1}$ could not be transitive. So there is only one cycle and it is of length p. So $p = q$ and we infer, with Lagrange's help, that $p \mid |G_k|$ for all $k < n$.

Now let τ be any p-cycle that belongs to G. Since $\mathbf{G}_n$ is trivial, we see that $\tau \notin G_n$. Pick k as small as possible so that $\tau \in G_k$. So $\tau \notin G_{k+1}$. Let $\mathbf{H}$ be the subgroup of $\mathbf{G}_k$ generated by τ. Then $|H| = p$.

Every element of H other than the identity generates $\mathbf{H}$. So $H \cap G_{k+1}$ can contain only the identity element since $\tau \notin G_{k+1}$. Since $\mathbf{G}_k \rhd \mathbf{G}_{k+1}$ we see that $\mathbf{HG}_{k+1}$ is a subgroup of $\mathbf{G}_k$. We also know

$$|HG_{k+1}||H \cap G_{k+1}| = |H||G_{k+1}|.$$

So we find that $|HG_{k+1}| = p|G_{k+1}|$. Now $\mathbf{HG}_{k+1}$ is a subgroup of $\mathbf{S}_p$ and this last group has cardinality $p!$. So p^2 cannot divide the order of $\mathbf{HG}_{k+1}$. But this means that p cannot divide $|G_{k+1}|$. This forces $k+1 = n$. Therefore, $\tau \in G_{n-1}$, as desired. □

Proof of the Theorem on the Transitive Solvable Subgroups of $\mathbf{S}_p$. Let $\tau \in G$ and suppose that $a, b \in \{0, 1, \ldots, p-1\}$ with

$$\tau(a) = a$$
$$\tau(b) = b$$

Our aim is to show that $a = b$ or that τ is the identity. Let σ be a p-cycle that generates $\mathbf{G}_{n-1}$. Now $\tau\sigma\tau^{-1} \in G$ and it must also be a p-cycle (since it has order p and p is prime). So $\tau\sigma\tau^{-1} \in G_{n-1}$ by the last lemma. With this in view, pick a positive $k < p$ so that

$$\tau\sigma\tau^{-1} = \sigma^k.$$

Since $\mathbf{G}_{n-1}$ is transitive, we can pick a natural number $\ell < p$ so that $\sigma^\ell(a) = b$. From the displayed equation we see that $\tau\sigma(a) = \sigma^k\tau(a) = \sigma^k(a)$. An easy induction reveals that $\tau\sigma^j(a) = \sigma^{jk}(a)$ for every natural number j. In particular,

$$\sigma^\ell(a) = b = \tau(b) = \tau\sigma^\ell(a) = \sigma^{\ell k}(a).$$

So we find $a = \sigma^{\ell k - \ell}(a)$. But the only permutation in G_{n-1} with a fixed point is the identity permutation. This means $p \mid \ell(k-1)$. Since p is prime and $0 \le \ell, k < p$ and k is positive, we conclude that either $\ell = 0$ or $k = 1$. In the first alternative we have $b = \sigma^\ell(a) = \sigma^0(a) = a$, while in the second alternative we have that σ and τ commute. In that case,

$$\begin{aligned}
\tau(\sigma(a)) &= \sigma(\tau(a)) = sigma(a) \\
\tau(\sigma^2(a)) &= \sigma^2(\tau(a)) = \sigma^2(a) \\
&\vdots \\
\tau(\sigma^j(a)) &= \sigma^j(a) \\
&\vdots
\end{aligned}$$

In this way we see that τ must fix on the elements of $\{0, 1, \ldots, p-1\}$. That is τ is the identity. □

As a corollary we arrive at the following result.

Artin's Criterion for Unsolvability over Q. *Every irreducible polynomial of prime degree with coefficients in* $\mathbb{Q}$ *that has at least two real roots and at least one nonreal root in* $\mathbb{C}$ *is not solvable by radicals.*

Proof. Let **E** be a subfield of $\mathbb{C}$ that is a splitting field of our polynomial. Since the coefficients of our polynomial are real (even rational) we see that complex conjugation, restricted to E, is a member of the Galois group. Because the polynomial has a root that is not real, we see that the restriction of complex conjugation is not merely the identity map. On the other hand, complex conjugation fixes two of the roots (since two of them are real). So the Galois group of the polynomial cannot be solvable, by the Theorem on Transitive Solvable Subgroups of $\mathbf{S}_p$. So by Galois' characterization, our polynomial is not solvable by radicals. □

Notice that the hypotheses of this theorem are weaker than those laid out in the $\mathbf{S}_p$ Criterion over $\mathbb{Q}$. On the other hand, the conclusion is also weaker: that the Galois group is not solvable rather than that the Galois group is actually $\mathbf{S}_p$.

In applying this new criteria it is enough to show that the graph of $f(x)$ on the $X \times Y$ plane crosses the X-axis at least twice, but not p times (provide $f(x)$ is irreducible of of prime degree p).

Let us devise an example to which Artin's Criterion applies, but not the earlier criterion. We will find an irreducible polynomial $f(x)$ of degree 7 with 3 real roots and 4 nonreal roots. The graph of such a polynomial will cross the X-axis 3-times. One way to achieve this is to make sure the leading coefficient is positive and that the graph has one local maximum (where the function is positive) and one local minimum (where the function is negative). We hope to use Eisenstein to ensure that $f(x)$ is irreducible, so our polynomial will have integer coefficients. Given the curve-sketching nature of this idea, we will first create a suitable derivative $f'(x)$. This must have degree 6. Here is one to start with:

$$(x-3)(x+3)(x^2+3)(x^2+9) = x^6 + 3x^4 - 3^4x^2 - 3^5.$$

I have used a lot of 3's in the hope that this will make the use of eventual use Eisenstein easier. Were this the derivative of our polynomial, we would know that the graph is increasing on $(-\infty,-3)$, that it is decreasing on $(-3,3)$ and that it is increasing again on $(3,\infty)$. The next step would be the integrate this polynomial, but that would introduce some fractional coefficients. To ease this, why not multiply the thing by $7\cdot 5$? So take

$$f'(x) = 7\cdot 5x^6 + 7\cdot 5\cdot 3x^4 - 7\cdot 5\cdot 3^4x^2 - 7\cdot 5\cdot 3^5.$$

Integrating gets us

$$f(x) = 5x^7 + 7\cdot 3x^5 - 7\cdot 5\cdot 3^3x^3 - 7\cdot 5\cdot 3^5x + c$$

where c is the constant of integration. I hope to chose c so that the local maximum (it is at $x=-3$) is positive, that the local minimum (it is at $x=3$) is negative, and finally, so that Eisenstein will tell us that $f(x)$ is irreducible. Sheer computation shows

$$f(-3) = 3^6\cdot 48 + c \qquad \text{and } f(3) = -3^6\cdot 48 + c.$$

Given the desire for $f(-3) > 0$ and $f(3) < 0$, it turns out that c must be selected so that

$$-3^6\cdot 48 < c < 3^6\cdot 48.$$

This is a very commodious range of choices. I take $c = 3$ (but you might like 7 or even 21 better). This choice gives

$$f(x) = 5x^7 + 7\cdot 3x^5 - 7\cdot 5\cdot 3^3x^3 - 7\cdot 5\cdot 3^5x + 3.$$

So Eisenstein applies with 3 as the chosen prime. We could make this more mysterious by saying

$$f(x) = 5x^7 + 21x^5 - 945x^3 - 8,435x + 3.$$

This polynomial is not solvable by radicals.

15.1 Problem Set 20

Algebra Homework, Edition 20

Solvability by radicals and other things Galois may have known

Problem 73.
Let p be prime and let H be a subgroup of S_p. Prove that if H has a transposition and an element of order p, then $H = S_p$. Provide an explicit counterexample when p is not prime.

Problem 74.
Prove that $x^5 - 2x^3 - 8x + 2$ is not solvable by radicals over the field $\mathbb{Q}$ of rational numbers.

Problem 75.
Let F be a finite field. Prove that the product of all the nonzero elements of F is -1. Using this, prove Wilson's Theorem:

$$(p-1)! \equiv -1 \pmod{p}$$

for every prime number p.

Problem 76.
Let E be the splitting field of $x^5 - 2$ over the field $\mathbb{Q}$ of rationals. Find the lattice (draw a picture) of all fields intermediate between $\mathbb{Q}$ and E.

Problem 77.
Let F be a field of characteristic p, where p is a prime. Let E be a field extending F. Prove that E is a normal separable extension of F of dimension p if and only if E is the splitting field over F of an irreducible polynomial of the form $x^p - x - a$, for some $a \in F$.

LECTURE 16

ALGEBRAIC CLOSURES OF REAL-CLOSED FIELDS

Here we want to obtain the result that the field of complex numbers is the algebraic closure of the field of real numbers. This assertion, traditionally called the Fundamental Theorem of Algebra, has a storied past and many proofs—indeed there are whole monographs devoted to the exposition of an array of proofs of this theorem. Many point to the doctoral dissertation of Gauss for the first fully correct proof. Sadly, even the proof in Gauss's dissertation also has a gap—let the dedicated graduate students take note!

The shortest proofs come by way of complex analysis: Were $f(z)$ a rootless polynomial of positive degree then $\frac{1}{f(z)}$ would be analytic on the whole complex plane (i.e. it is holomorphic) and a simple argument shows it is bounded. So Liouville tells us that it must be constant—an impossibility. Of course, developing complex analysis to the point of Liouville's Theorem (or any of a number of other theorems of complex analysis the would serve) is not entirely immediate. These proofs have the added feature that they apply to complex functions other than polynomials of positive degree.

The approach will we take uses the apparatus of Galois theory, and has the advantage that it applies to fields other than the complex numbers.

The field of real numbers has the following three properties:

(a) Every polynomial of odd degree has a root.

(b) There is a set P of elements with the following properties:

  (i) $0 \notin P$.

  (ii) If $a, b \in P$, then $a+b, ab \in P$.

  (iii) For every nonzero element a of the field exactly one of $a \in P$ or $-a \in P$ holds.

(c) Every element of P has a square root in the field.

You should note that all these properties have an algebraic flavor. Property (a) follows by the familiar Intermediate Value Theorem of freshman calculus. The set P is just the set of positive reals. Property (c) again follows by the Intermediate Value Theorem.

Any field that has properties (a), (b), and (c) is called a **real closed field**. Of course, $\mathbb{R}$ is a real closed field, but there are other real closed fields, even ones that are countable.

The properties stipulated in (b) allow us to define a linear ordering of the field:

$$a < b \overset{\text{def}}{\iff} b - a \in P.$$

The demonstration that this defines a linear order that has the expected properties with respect to $+$ and $\cdot$ is left in the hands of the graduate students. We could, of course, reverse this process: start with a well-behaved linear order and take $P = \{a \mid a > 0\}$ and show that P has the attributes given in (c).

To get a better grip on this notion, the eager graduate students should try proving that in a real closed field the square of any nonzero element is positive, that 1 is postive, and that the field must have characteristic 0.

The notion of a real closed field was propounded by Emil Artin around 1924 as a means to bring algebraic methods into play in what had been a largely analytic enterprise: the investigation of the real and complex numbers. The theorems here are taken largely from two papers of Emil Artin and Otto Schreier which appeared in 1926 and 1927. Artin's famous solution to Hilbert's Seventeenth Problem, published also in 1927, was based on theory developed by Artin and Schreier in these two papers.

The proof I give below is the work of Artin and Schreier and uses Galois Theory and Sylow's Theorem. Artin and Schreier also provide a second argument that lifts a 1795 proof of Laplace of the Fundamental Theorem of Algebra to the case of real closed fields. Laplace's proof depended on Kronecker's Theorem, which was unknown at the time. In 1816 Gauss published a proof that filled this gap in Laplace's proof by an analysis of symmetric polynomials, circumventing the still unknown result of Kronecker.

The Artin-Schreier Fundamental Theorem of Algebra for Real Closed Fields. *If* $\mathbf{R}$ *is a real closed field, then* $\mathbf{R}\left[\sqrt{-1}\,\right]$ *is algebraically closed.*

Proof. First notice that $\left(\sqrt{-1}\right)^2 = -1$ and -1 is not positive. This means $\sqrt{-1} \notin R$. So x^2+1 is irreducible over $\mathbf{R}$ and $\mathbf{R}\left[\sqrt{-1}\,\right]$ is the splitting field of x^2+1. Let $\mathbf{C} = \mathbf{R}\left[\sqrt{-1}\,\right]$. According to Kronecker $[\mathbf{C}:\mathbf{R}] = 2$. Of course the members of C have the form

$$a + b\sqrt{-1}$$

where $a, b \in R$. Now conjugation has its usual definition and it is an automorphism of $\mathbf{C}$ that fixes each element of R.

Now let $f(x) \in \mathbf{C}[x]$. By $\bar{f}(x)$ we mean the polynomial obtained from $f(x)$ by applying conjugation to each of the coefficients. Then $f(x)\bar{f}(x) \in R[x]$ follows easily from the description of the coefficients of the product of two polynomials together with the fact that conjugation is an automorphism of $\mathbf{C}$.

Observe that $f(x)$ has a root in C if and only if $f(x)\bar{f}(x)$ has a root is C. So it is enough for us to prove that every monic polynomial in $\mathbf{R}[x]$ of positive degree has a root in C. We already know this for polynomials of odd degree—they even have roots in R.

We use the following fact.

Contention. Every element of C has a square root in C.

Proof. Let $a+b\sqrt{-1}$ be an arbitrary element of C. In case $b=0$, this element will belong to R and, since every positive element of R already has a square root in R, it is easy to see that every element of R has a square root in C. So we consider that $b\neq 0$. Then direct computation shows that $c+d\sqrt{-1}$ is a square root of $a+b\sqrt{-1}$, where

$$c=\frac{b}{2d} \text{ and } d^2=\frac{-a+\sqrt{a^2+b^2}}{2}.$$

Notice that $\frac{-a+\sqrt{a^2+b^2}}{2}$ is a positive member of R. These constraints came about by twiddling with the quadratic formula. □

This entails that there is no extension $\mathbf{E}$ of $\mathbf{C}$ with $[\mathbf{E}:\mathbf{C}]=2$ since every polynomial of degree 2 in $\mathbf{C}[x]$ is reducible using the quadratic formula.

Now let $f(x)\in\mathbf{R}[x]$ be a monic polynomial. Let $\mathbf{E}$ be the splitting field of $f(x)(x^2+1)$ over $\mathbf{R}$. We can suppose that $\mathbf{E}$ extends $\mathbf{C}$. Since our characteristic is 0, we know that $\mathbf{E}$ is a Galois extension of $\mathbf{R}$. Pick natural numbers ℓ and m, with m odd, so that

$$|\operatorname{Gal}(\mathbf{E}/\mathbf{R})|=2^\ell m.$$

By Sylow, $\operatorname{Gal}(\mathbf{E}/\mathbf{R})$ has a subgroup $\mathbf{H}$ with $|H|=2^\ell$. Let $\mathbf{K}=\operatorname{Inv}\mathbf{H}$. Then $[\mathbf{E}:\mathbf{K}]=2^\ell$ and $[\mathbf{K}:\mathbf{R}]=m$, by the Fundamental Theorem of Galois Theory. Since $\mathbf{R}$ has no proper extension of odd dimension (every polynomial of odd degree as a root—so you get a grumble out of Kronecker and the Dimension Formula), we must have $m=1$ and $\mathbf{K}=\mathbf{R}$. But then $[\mathbf{E}:\mathbf{R}]=2^\ell$. But recall that $\mathbf{E}$ extends $\mathbf{C}$. So

$$2^\ell=[\mathbf{E}:\mathbf{R}]=[\mathbf{E}:\mathbf{C}][\mathbf{C}:\mathbf{R}]=[\mathbf{E}:\mathbf{C}]2.$$

In this way we find $[\mathbf{E}:\mathbf{C}]=2^{\ell-1}$. If $\ell=1$ then we find $\mathbf{E}=\mathbf{C}$, and we have reached the conclusion we desire. On the other hand, if $\ell>1$, we see that $\operatorname{Gal}\mathbf{E}/\mathbf{C}$ is a group of cardinality $2^{\ell-1}$. By Sylow, there is a subgroup $\mathbf{N}$ of this Galois group so that $|N|=2^{\ell-2}$. So $[\operatorname{Gal}\mathbf{E}/\mathbf{C}:\mathbf{N}]=2$. Now every subgroup of index 2 must be a normal subgroup. The fixed field of $\mathbf{N}$ must be a (normal) extension of $\mathbf{C}$ dimension 2. But we know that $\mathbf{C}$ has no extensions of dimension 2. So we reject the possibility that $\ell>1$.

This means every polynomial over $\mathbf{R}$ of positive degree has a root is C. So our proof is complete.

The use of Sylow's Theorem (unknown until the late 1800's) above and of the Fundamental Theorem of Galois Theory to produce the fixed field of $\mathbf{N}$ can be avoided by following the line of reasoning proposed by Laplace in 1795. Here is how.

We still want to show that every polynomial $f(x)$ of positive degree with coefficients in R has a root in C. Let ℓ by the natural number so that the degree of $f(x)$ is $n=2^\ell m$ where m is odd. Call this number the 2-*index* of $f(x)$. Our proof is by induction on the 2-index. In the base step, $f(x)$ is a polynomial of odd degree, so it even as a root in R. For the inductive step, suppose $f(x)$ has 2-index $k+1$. Let $r_0,\ldots,r_{n-1}$ be the roots of $f(x)$ in $\mathbf{E}$. For each real number a define

$$g_a(x)=\prod_{i<j<n}\left(x-(r_i+r_j+ar_ir_j)\right)$$

Notice that the degree of $g_a(x)$ is $\binom{n}{2}$. But

$$\binom{n}{2}=\frac{n(n-1)}{2}=\frac{1}{2}2^{k+1}m(2^{k+1}-1)=2^km(2^{k+1}-1)$$

so that the 2-index of each $g_a(x)$ is k. But observe that the coefficients of $g_a(x)$ must be fixed by every automorphism in $\mathrm{Gal}\,\mathbf{F}/\mathbf{R}$. So each $g_a(x) \in \mathbf{R}[x]$. (This uses a little bit of Galois theory to say that the fixed field of $\mathbf{E}$ is actually $\mathbf{R}$.) So for each $a \in R$ we see by the induction hypothesis that

$$r_i + r_j + a r_i r_j \in C$$

for some choice of the natural numbers i and j with $i < j < n$. Now there are only finitely many ways to pick such i and j but infinitely choices for a. As every pigeon knows, we must have two distinct member of R, say a and b, so that for some choice of i and j

$$r_i + r_j + a r_i r_j \qquad \text{and} \qquad r_i + r_j + b r_i r_j \qquad \text{both belong to } C.$$

Subtracting these and dividing away the nonzero $b-a$, we find first that $r_i r_j$ and then $r_i + r_j$ also belong to C. But everyone can see that

$$(x - r_i)(x - r_j) = x^2 - (r_i + r_j) + r_i r_j,$$

which is a polynomial of degree 2 with coefficients in C. But all polynomials in $\mathbf{C}[x]$ of degree 2 have roots in C. So $r_i \in C$ and it is a root of $f(x)$.

Laplace still needed the (much delayed) aide of Kronecker to obtain the splitting field $\mathbf{E}$, but the little bit of Galois Theory used here can be finessed. □

Artin and Schreier also proved the converse.

Artin and Schreier's Characterization of Real Closed Fields. *A field* $\mathbf{R}$ *is a real closed field if and only if* x^2+1 *has no root in* $\mathbf{R}$ *and* $\mathbf{R}\big[\sqrt{-1}\big]$ *is algebraically closed.*

Proof. We only have to prove one direction. So suppose x^2+1 has no root in $\mathbf{R}$ and that $\mathbf{R}\big[\sqrt{-1}\big]$ is algebraically closed. First, observe that if $a, b \in R$ then there is $c \in R$ so that $c^2 = a^2 + b^2$. This follows since the analog of complex conjugation in $\mathbf{R}\big[\sqrt{-1}\big]$ is an automorphism whose set of fixed points is just R (an entertainment for graduate students!). Now using the algebraic closedness, pick $u \in \mathbf{R}\big[\sqrt{-1}\big]$ with $u^2 = a + bi$. Then

$$a^2 + b^2 = (a+bi)(a-bi) = (a+bi)\overline{(a+bi)} = u^2\overline{u^2} = (u\overline{u})^2.$$

But $u\bar{u} \in R$ since it is fixed by this analog of complex conjugation. So take $c = u\bar{u}$. So we see that in $\mathbf{R}$ the sum of two squares is again a square. It follows that the sum of any finite number of squares in again a square. Now -1 cannot be a square in $\mathbf{R}$ since x^2+1 has no root in R. This also means that 0 cannot be the sum of a finite number of nonzero squares. Let us take P to be the set of all those members of R that can be written as a sum of nonzero squares, which is the same as the set of those members of R that are themselves nonzero squares. In the definition of real closed fields there are four stipulations our set P must satisfy. They are all easy (aren't they?).

So it only remains to show that every polynomial in $\mathbf{R}[x]$ of odd degree has a root in R. Now every polynomial of odd degree must have an irreducible factor of odd degree. Such an irreducible polynomial must have a root r in $\mathbf{R}\big[\sqrt{-1}\big]$ since that field is algebraically closed. But this is a field of dimension 2 over $\mathbf{R}$. Consider the Dimension Formula and Kronecker's Theorem. The degree of our irreducible polynomial must divide 2. The only odd number that divides 2 is 1. So our irreducible polynomial has degree 1. That means it has a root in R.

In this way, we see that $\mathbf{R}$ is a real closed field. □

Actually, Artin and Schreier go on to prove that if **R** is any field so that [**A** : **R**] is finite, where **A** is algebraically closed, then **R** is a real closed field. This is an intriguing result: given a field **F** and its algebraic closure **A** there are only three possibilities: **A** is infinite dimensional over **F** or the dimension is just 2 (and **F** is a real closed field) or the dimension is just 1 (the field **F** is algebraically closed already). Why not look into this matter a bit on your own?

You can see from the proof of the Fundamental Theorem of Algebra for Real Closed Fields, that the properties of the set P were used only to establish that every polynomial in $\mathbf{C}[x]$ of degree 2 has a root in C. After some tampering, you can see that the following statement can also be proven:

> Every field of characteristic 0 in which every polynomial of degree 2 has a root and in which every polynomial of odd degree has a root, is algebraically closed.

In essence, this is the result of Artin and Schreier (and of Gauss) with the "real" part stripped out. In 2007, Joseph Shipman proved a wide ranging extension of this result, namely:

> Every field in which every polynomial of prime degree has a root is algebraically closed.

16.1 Problem Set 21

Algebra Homework, Editon 21
Norm and trace and real closed fields

Problem 78.
Let E be a finite separable extension of the field F. Prove that the trace maps E onto F.

Problem 79.
Let E be a finite extension of the finite field F. Prove that both the norm and the trace map E onto F.

Problem 80.
Let R be a real closed field and let $f(x) \in R[x]$. Suppose that $a < b$ in R and that $f(a)f(b) < 0$. Prove that there is $c \in R$ with $a < c < b$ such that c is a root of $f(x)$.

Problem 81.
Let R be a real closed field, let $a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n = f(x) \in R[x]$, and put $M = |a_0| + |a_1| + \cdots + |a_{n-1}| + 1$. Prove that every root of $f(x)$ which belongs to R belongs to the interval $[-M, M]$.

LECTURE 17

GAUSS ON CONSTRUCTING REGULAR POLYGONS BY STRAIGHTEDGE AND COMPASS

17.1 Problem Set 22

ALGEBRA HOMEWORK, EDITON 22

RANDOM FIELD PROBLEMS

PROBLEM 82.
Prove that every element of a finite field can be written as the sum of two squares.

PROBLEM 83.
Prove that every polynomial with rational coefficients whose splitting field over $\mathbb{Q}$ has dimension 1225 is solvable by radicals.

PROBLEM 84.
Let F be a field. Prove that the following are equivalent.

(a) F is not algebraically closed but there is a finite upper bound on the degrees of the irreducible polynomials in $F[x]$.

(b) F is a real closed field.

PROBLEM 85.
Let E be the splitting field over $\mathbb{Q}$ of $x^4 - 2$. Determined all the fields intermediate between E and $\mathbb{Q}$. Draw a diagram of the lattice of intermediate fields.

PROBLEM 86.
Prove that the field of real numbers has only one ordering that makes it into an ordered field. In contrast, prove that $\mathbb{Q}[\sqrt{2}]$ has exactly two such orderings.

LECTURE 18

ALGEBRAIC INTEGERS

Here is a question that may seem, at first glance, to be so obvious that it barely needs to be asked:

> "How can one pick out the subset of integers from the set of rationals?"

Of course, we saw how to *build* the rationals from the integers—we even know how to do this fraction field trick with any integral domain. However, the question above asks for a reversal of this trick. We could invent an infinite process that collects the integers: first throw in 0, then throw in 1 and -1, then $1+1$ and $(-1)+(-1)$, ... But is there another way, a finite elementary way?

We look at one ingenuous way here. Let **E** be a field and let **R** be a subring of **E**. We say that $u \in E$ is **integral** over **R** provided u is a root of a monic polynomial belonging to $\mathbf{R}[x]$. When **R** is actually a subfield of **E** the integral elements correspond to the elements that are algebraic over **R**. When $\mathbf{E} = \mathbb{C}$ and $\mathbf{R} = \mathbb{Z}$ we refer to the integral elements as **algebraic integers**.

Fact. A complex number u is an algebraic integer if and only if u is algebraic over $\mathbb{Q}$ and the minimal polynomial of u over $\mathbb{Q}$ actually belongs to $\mathbb{Z}[x]$.

Proof. Let $m(x)$ be the minimal polynomial of u over $\mathbb{Q}$. It is evident that if $m(x) \in \mathbb{Z}[x]$, then u is an algebraic integer.

So now suppose that u is an algebraic integer and pick $f(x) \in \mathbb{Z}[x]$ so that $f(x)$ is monic and u is a root of $f(x)$. Then we have that $m(x) \mid f(x)$ in $\mathbb{Q}[x]$. Now $f(x)$ factors (uniquely) into irreducible monic factors in $\mathbb{Z}[x]$. We also know that every irreducible in $\mathbb{Z}[x]$ is irreducible in $\mathbb{Q}[x]$. This means our factorization of $f(x)$ is $\mathbb{Z}[x]$ is also a factorization of $f(x)$ in $\mathbb{Q}[x]$ into irreducibles. So $m(x)$ must be an associate (over $\mathbb{Q}$) of one of the factors of $f(x)$. But the factors of $f(x)$ as well as $m(x)$ are monic. This forces the unit involved in the association to be 1 and so $m(x)$ must be one of the irreducible factors of $f(x)$, which were all in $\mathbb{Z}[x]$. So $m(x) \in \mathbb{Z}[x]$, as desired. □

Theorem on Algebraic Integers. *A rational number is an algebraic integer if and only if it is an integer. A complex number u is algebraic if and only if there is $b \in \mathbb{Z}$ such that bu is an algebraic integer.*

Proof. Evidently, every member of $\mathbb{Z}$ is an algebraic integer. For the converse, suppose that u is an algebraic integer that happens to be rational. Then $x - u$ is the minimal polynomial of u over $\mathbb{Q}$. By the fact above, it belongs to $\mathbb{Z}[x]$. Hence, $u \in \mathbb{Z}$.

Now let u be a complex number that is algebraic over $\mathbb{Q}$. Let $m(x) \in \mathbb{Q}[x]$ be the minimal polynomial of u. Let $b \in \mathbb{Z}$ be the product of the denominators of the coefficients of $m(x)$. Suppose

$$m(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + x^n.$$

Let

$$f(x) = b^n a_0 + b^{n-1} a_1 x + b^{n-2} a_2 x^2 + \cdots + b a_{n-1} x^{n-1} + x^n.$$

Then $f(x) \in \mathbb{Z}[x]$ and $f(x)$ is monic. It is routine to check that $f(bu) = b^n m(u) = 0$. So bu is an algebraic integer. For the converse, if bu is a root of a monic $g(x) \in \mathbb{Z}[x]$. Then u is a root of $g(bx)$, which is certainly a polynomial in $\mathbb{Z}[x] \subseteq \mathbb{Q}[x]$, even if it is not monic. So u is algebraic over $\mathbb{Q}$. □

LECTURE 19

THE LINDEMANN-WEIERSTRASS THEOREM ON TRANSCENDENTAL NUMBERS

There are only countably many real numbers that are algebraic over the field $\mathbb{Q}$ of rational numbers. So almost every real number is transcendental. On the other hand, the real numbers that we can describe, in one manner or another appear to be preponderantly algebraic. This is not too surprising, since even being generous with the notion of description, leads to only countably many descriptions (all of which can be typed in at a keyboard, say) and hence to only countably many describable reals.

In 1844, Liouville cooked up the earliest examples of describable reals that are transcendental. These designer reals are likely interesting only for this particular property. It wasn't until 1873 that Charles Hermite proved that e is transcendental. In 1882, Ferdinand Lindemann built on Hermite's methods to prove that π is transcendental (finally settling the millenia old problem about squaring the circl with straightedge and compass). In 1885, Karl Weierstrass reframed Lindemann's proof to obtain the general result that is the topic of this lecture.

Before we can formulate the theorem, we need a new notion. Let $\mathbf{K}$ be a field extending $\mathbf{F}$. Let $u_0, \ldots, u_{n-1} \in K$ be distinct. We say that these elements are **algebraically independent over F** provided that whenever $f(x_0, \ldots, x_{n-1}) \in \mathbf{F}[x_0, \ldots, x_{n-1}]$ such that $f(u_0, \ldots, u_{n-1}) = 0$ then it must be that $f(x_0, \ldots, x_{n-1})$ is the zero polynomial.

The Lindemann-Weierstrass Theorem. *If $u_0, \ldots, u_{n-1}$ are distinct complex numbers that are algebraic over the rationals and are also linearly independent over the rationals, then the complex exponentials $e^{u_0}, e^{u_1}, \ldots, e^{u_{n-1}}$ are algebraically independent over the field of complex numbers that are algebraic over $\mathbb{Q}$.*

Before launching into the proof, here are two corollaries.

Hermite says, "e is transcendental".

Proof. Let $u = 1$. It is immediate that 1 is algebraic and that $\{1\}$ is linearly independent over $\mathbb{Q}$. So the theorem says e^1 is algebraically independent over $\mathbb{Q}$. This means that e is not the root of any polynomial in $\mathbb{Q}[x]$ of positive degree. So e is transcendental. □

Lindemann says, "π is transcendental".

Proof. We all know that $e^{i\pi} = -1$. So $e^{i\pi}$ is algebraic. This means that $\{e^{i\pi}\}$ is not algebraically independent over $\mathbb{Q}$. By the theorem $i\pi$ cannot be algebraic since $\{i\pi\}$ is certainly linearly independent over $\mathbb{Q}$. But i is algebraic and we know that the product of algebraic numbers is algebraic, so π is not algebraic. That is, π is transcendental. □

Here is a closely related theorem.

The Lindemann-Weierstrass Theorem, Alternate Version. *If $u_0, \ldots, u_{n-1}$ are distinct algebraic numbers, then the complex exponentials $e^{u_0}, \ldots, e^{u_{n-1}}$ are linearly independent over the field of complex numbers that are algebraic over $\mathbb{Q}$.*

To see how close these to versions of the Lindemann Weierstrass Theorem are, here are the corollaries again.

Hermite says, "e is transcendental".

Proof. Let n be any positive integer. Let $u_0 = 0, u_1 = 1, \ldots, u_n = n$. These are distinct algebraic numbers. So by the alternate version $\{e^0, e^1, e^2, \ldots, e^n\}$ linearly independent over $\mathbb{Q}$. So we see that e cannot be a root of any polynomial in $\mathbb{Q}[x]$ of degree n. Since n was arbitrary, we see that e must be transcendental. □

Lindemann says, "π is transcendental".

Proof. We all know that $e^{i\pi} = -1$ and $e^0 = 1$. We know that $\{-1, 1\}$ is not linearly independent over $\mathbb{Q}$. Since $i\pi$ and 0 are certainly distinct and 0 is algebraic, we find that $i\pi$ cannot be algebraic. But i is algebraic and we know that the product of algebraic numbers is algebraic, so π is not algebraic. That is, π is transcendental. □

Now let us turn to the proof of the Lindemann-Weierstrass Theorem.

Proof of the Lindemann-Weierstrass Theorem from the Alternate Version. Let $u_0, \ldots, u_{n-1}$ be distinct complex algebraic numbers that are linearly independent over $\mathbb{Q}$. Let $\langle k_0, \ldots, k_{n-1}\rangle$ and $\langle \ell_0, \ldots, \ell_{n-1}\rangle$ be two different sequences of natural numbers. Then

$$\prod_{i<n} (e^{u_i})^{k_i} = e^{\sum_{i<n} k_i u_i} \text{ and } \prod_{i<n} (e^{u_i})^{\ell_i} = e^{\sum_{i<n} \ell_i u_i}.$$

Notice that $\sum_{i<n} k_i u_i \neq \sum_{i<n} \ell_i u_i$ by the linear independence of the u_i's. So the corresponding products are also different. Now suppose we are given m distinct sequences of natural numbers. This would result is m pairwise distinct products of the form above. By the Alternate Version, these products will be linearly independent over the algebraic numbers. But this is just another way of saying that the complex exponentials $e^{u_0}, \ldots, e^{u_{n-1}}$ are algebraically independent over the algebraic numbers. □

So what we need is a proof of the Alternate Version of the the Lindemann-Weierstrass Theorem.

19.1 Problem Set 23

ALGEBRA HOMEWORK, EDITON 23

TRANSCENDENTAL NUMBERS, CONSTRUCTIBLE NUMBER, AND OTHER PUZZLES

PROBLEM 87.
Prove that $\ln u$ and $\sin u$ are transcendental over the field of rational numbers, whenever u is a positive algebraic real number.

PROBLEM 88.
Let **F**, **K**, and **L** be fields so that **K** is a finite separable extension of **F** and **L** is a finite separable extension **K**. Prove that **L** is a finite separable extension of **F**.

PROBLEM 89.
Prove that no finite field is algebraically closed.

PROBLEM 90.
Archimedes studied cylinders circumscribed around spheres. Let us say that such a cylinder is **constructible** provided the radius of the sphere is a constructible real number. So the cylinder circumscribed around a sphere of radius 1 is constructible. Call this cylinder the **unit cylinder**. Let **C** be a cylinder circumscribed around a sphere so that the volume of **C** is twice as larger as the volume of the unit cylinder. Explain in detail why **C** is not constructible.

LECTURE 20

THE GALOIS CONNECTION BETWEEN RINGS OF POLYNOMIALS AND POINTS IN AFFINE SPACES