844: Characterizing polynomials that can be solved

844 I: Rings, Factorization, and the fundamental theorem of Galois theory

844 II: Identifying (and solving) "solvable" polynomials, eg. solution formulas for cubics and quartics over Q

Math 844, Fundamental Theorem of Galois Theory; application to solution formulas for "solvable" polynomials.

§1) The problem of finding maximal ideals, with application to constructing "universal" splitting fields.

One new topic we want to cover is a technical digression into set theory and logic, not completely essential to do most of field theory in my opinion, but useful, and almost universally accepted as part of the tool kit of mathematicians today. It has however an unconstructive flavor that makes it challengingly abstract and therefore unpleasant to some people. This is the use of Zorn's Lemma to deduce the "existence" of some things which cannot be explicitly constructed in a finite, or even a countable number of steps, indeed sometimes for which no explicit construction can be given at all Questions of this type include whether every field can be embedded in an algebraically closed field, and whether every commutative ring contains maximal ideals. The difficulty seems to arise mainly in cases where the sets involved are uncountably infinite. [Recall that a set S is (at most) countable if there is an injection $S\to \mathbf{Z}$, and countably infinite if there is a bijection $S\to\mathbf{Z}$. We may be careless and say "countable" sometimes to mean "at most countable", and sometimes to mean "countably infinite".]

Let's consider how some of these problems arise from the questions we have been studying. The problem of "solving" a polynomial f requires understanding the relationship between the coefficients and the roots. Following Galois one can translate this into studying the structure of the field extension $k\subset F$ where k is the field generated by the coefficients and F is the field generated over k by the roots, i.e. where F is a splitting field for f. In particular the existence of splitting fields is fundamentally important Recall that for polynomials over $\mathbf{Q}$ it was easy to produce splitting fields since there is an inclusion $\mathbf{Q}\subset\mathbf{C}$, where $\mathbf{C}$ is "algebraically closed".

Definition: A field F is "algebraically closed" iff every polynomial of positive degree with coefficients in F has a root in F, equivalently iff every polynomial of positive degree over F factors completely into linear factors in F[X], equivalently iff the only irreducible polynomials over F are of degree one, iff there are no proper algebraic extension fields of F.

An algebraically closed extension of a field k contains splitting fields of any polynomial in k[X]

Eg. if f is any $\mathbb{Q}$-polynomial of positive degree n, since $\mathbb{C}$ is algebraically closed, there exist complex numbers numbers $\alpha_1,\ldots,\alpha_n$ such that, in the ring $\mathbb{C}[X]$, f factors as $f(X) = c\prod_{j=1,\ldots,n}(X-\alpha_j) = c(X-\alpha_1)(X-\alpha_2)\ldots(X-\alpha_n)$ A splitting field for f is thus given by the subfield $F = \mathbb{Q}(\alpha_1,\ldots,\alpha_n)\subset\mathbb{C}$. Note that this is not an explicit "construction" since no way is given of actually finding the numbers $\alpha_1,\ldots,\alpha_n$, and that we know nothing about the field F, neither its dimension over $\mathbb{Q}$, nor (since we do not know their real and imaginary parts) even do we even know how to explicitly multiply elements such as $\alpha_1\alpha_2$ in it. In that sense, our "abstract" construction (forming the quotient of the polynomial ring $\mathbb{Q}[X]$ by a maximal ideal generated by an irreducible factor g of f, and continuing), is in fact more concrete. I.e. one can point to at least one explicit root of f, namely X+(g), in the first extension field $\mathbb{Q}[X]/(g)$, and one can also write down explicitly all the elements of this field and its addition and multiplication tables. By iterating this procedure one can give a complete description of the whole splitting field. Let's review the construction.

Lemma: Let k be any field and f any positive degree polynomial over k. Then there is a field K containing k, in which f has a root. If f is irreducible, there is a unique minimal such field, up to k-isomorphism

Proof: Let g be any irreducible factor of f in k[X], so that $(g)\subset k[X]$ is a maximal ideal, and consider the field $K = k[X]/(g)$. Since $k\cap(g) = 0$, the natural map $k\rightarrow k[X]/(g)$ is injective. Hence we can replace the image of k in $K = k[X]/(g)$ by k itself, so that K then contains k. If we denote by [h] the coset of h, since $h\rightarrow[h]$ is a homomorphism we have $[0] = [g(X)] = g([X])$ in K, so the coset $[X] = X+(g)$, is a root of g in K, hence also a root of f. Note that since f is a multiple of g, f belongs to the maximal ideal (g), and hence the natural map $k[X]\rightarrow k[X]/(g)$, sends f to zero The uniqueness is an exercise (whose ingredients we have seen many times before). QED.

Exercise #79) Prove that if f is an irreducible polynomial over a field k, then every field extension of k in which f has a root contains an isomorphic copy of the field $k[X]/(f)$, [which is thus the unique "minimal" field extension in which f has a root]

Exercise #80) Describe explicitly the splitting field F for X^3-2 over $\mathbb{Q}$ i.e. prove that the elements of F can be written uniquely in the form $a+b\alpha+c\alpha^2+d\beta+e\alpha\beta+f\alpha^2\beta$, where a,b,c,d,e,f, are in $\mathbb{Q}$, and where $\alpha^3=2$, and $\beta^2 = -\alpha\beta-\alpha^2$ (in particular prove $\dim_{\mathbb{Q}}(F) = 6$), and give the rule for multiplication of two general elements of F:
$(a+b\alpha+c\alpha^2+d\beta+e\alpha\beta+f\alpha^2\beta)(a'+b'\alpha+c'\alpha^2+d'\beta+e'\alpha\beta+f'\alpha^2\beta) = ?$
[We already know the Galois group of this field is isomorphic to S_3, and now we know the field itself, up to isomorphism.]

The fact that $\mathbb{C}$ is algebraically closed means that $\mathbb{C}$ is a super splitting field, one so large that every polynomial over any subfield splits in it. Moreover the uniqueness of splitting fields implies $\mathbb{C}$ contains an isomorphic copy of every abstract splitting field $\mathbb{Q}\subset F$ of any $\mathbb{Q}$-polynomial. The proof you recall is by the extension theorem. if $\mathbb{Q}\subset F$ is any finite algebraic extension of $\mathbb{Q}$ (such as a splitting field) the inclusion map $\mathbb{Q}\subset\mathbb{C}$ can be extended to a $\mathbb{Q}$ embedding $F\rightarrow\mathbb{C}$ whose image is $\mathbb{Q}$- isomorphic to F. Thus for most purposes, we can restrict attention to $\mathbb{Q}\subset\mathbb{C}$ when studying splitting fields of $\mathbb{Q}$-polynomials or any other finite algebraic extensions of $\mathbb{Q}$.
We might ask whether the word "finite" is necessary in the previous remark, or whether in fact $\mathbb{C}$ contains an isomorphic copy also of every infinite algebraic extension of $\mathbb{Q}$.

An algebraically closed extension of a field k contains at least every countably generated algebraic extension of k, (up to isomorphism).

Suppose that $\mathbb{Q}\subset\mathbb{Q}(\alpha_1,\dots,\alpha_j,\dots) = F$ is a countably generated algebraic field extension of $\mathbb{Q}$; since $F = \cup F_j$ where $F_j = \mathbb{Q}(\alpha_1,\dots,\alpha_j)$, we can construct a $\mathbb{Q}$ embedding of $F\rightarrow\mathbb{C}$, simply by extending the $\mathbb{Q}$-homomorphism to each finite extension $\mathbb{Q}(\alpha_1,\dots,\alpha_j)$ in turn, for every j. Although it would take an infinitely "long time" to make this entire extension if we think in terms of doing "one extension per minute", it is plausible to say that this extension is well defined since if any particular element α of F is given then it lies in one of the finite subextensions $\mathbb{Q}(\alpha_1,\dots,\alpha_j)$, and thus our iterative procedure does give a finite process which defines the homomorphism at the element α. Hence the embedding is defined at every α. This solves the problem since every algebraic field extension of $\mathbb{Q}$ is countably generated, in fact every such field is countable. But suppose that k is an uncountably infinite field such as $\mathbb{C}(X)$ (the quotient field of the polynomial ring $\mathbb{C}[X]$), and $k\subset k(S) = F$ is an algebraic extension generated by an uncountable set S of algebraic elements over k: if $k\subset L$ where L is algebraically closed, would it be possible to extend the inclusion map $k\subset L$ to a k-embedding of $F\to L$? How would we proceed if the elements of S are too numerous to list in a sequence?

Q_1: Does an algebraically closed extension of a field k contain *every* algebraic extension of k (up to isomorphism), even uncountably generated ones?

In the uncountable case, this is the sort of question dealt with by Zorn's Lemma, i.e how to extend a construction which makes sense for finitely many or countably many objects, to a situation involving uncountably many objects. Zorn's Lemma says that the same argument by which we define a homomorphism on $\mathbb{Q}(\alpha_1,\ldots,\alpha_j,\ldots\ldots)$, extending it step by step from the subfields $\mathbb{Q}(\alpha_1,\ldots,\alpha_j)$, can be generalized to the case where there are uncountably infinitely many α's.

Exercise #81) Prove that every algebraic field extension of a countable field is again countable. [Hint: Prove first that if k is a countable field, then the set of polynomials over k is a countable set.]

Remark: As in the text above, this implies that every algebraic extension of $\mathbb{Q}$ is isomorphic to a subfield of $\mathbb{C}$, without using Zorn's Lemma. I.e. any algebraically closed extension of a countable field k contains an isomorphic copy of every algebraic extension of k.]

Exercise #82) Exhibit an uncountably infinite number of elements algebraic over the field $\mathbb{C}(X)$ (and which are not in $\mathbb{C}(X)$ of course)

Let's return to the existence problem by proving the following: **Given a sequence $\{f_j\}_{1,\ldots,\infty}$ of non constant polynomials over a field k, there is an extension of k in which every f_j has a root.**

We know how to construct a field extension $k\subset F$ in which a single non constant polynomial f has a root. By iterating this construction we can construct a field extension $k\subset F$ in which every one of the non constant polynomials in a finite collection $f_1,....,f_m$, f_j has a root. To see that the same argument still works for an infinite sequence $\{f_1,....,f_m,....\}$ is only a matter of inductive logic.

Lemma: If k is a field and $\{f_1,....,f_m,....\}$ is a sequence of non constant polynomials in k[X], there is a field extension $k\subset E$ such that every polynomial f_j in the sequence has a root in E
Proof: We first construct $k\subset E_1$, a field where f_1 has a root and then $E_1\subset E_2$, a field where f_2 also has a root, and so on: $k\subset E_1\subset E_2\subset .....\subset E_n\subset$ .... Since there is a procedure for constructing E_{n+1} from E_n and f_{n+1} for all n, most people would agree that this sequence of fields $k\subset E_1\subset E_2\subset .....\subset E_n\subset$ .... is well defined. Since the sequence is well defined, the union of the sequence is also well defined, and this union $E = \cup E_j$ is a field (see exercise 83 below) in which every f_j has a root. **QED**.

Remark: If k is countable, then k[X]/(g) is also countable for any g in k[X], (for non constant g it is isomorphic to a finite cartesian product of copies of k, and in general it is the image of a surjection from the countable set k[X], as proved in ex #81) Hence the construction in the Lemma can be done so that each E_j is countable, and thus (since a countable union of countable sets is countable) E is also countable.

Q_2: If S is an arbitrary set of non constant k-polynomials, is there an extension $k\subset F$ in which every polynomial in S has a root?

Exercise #83) Prove that if $E_1\subset E_2\subset .... \subset E_n\subset$ ...., is an increasing sequence of fields, then their union $E = \cup E_j$ is also a field. Also, if R is a ring and $I_1\subset I_2\subset .....\subset I_n\subset ....\subset R$ is an increasing sequence of ideals in R, prove the union $I = \cup I_j$ is an ideal in R.

Definition: If k is a field, an "algebraic closure" F, of k is an algebraic field extension $k\subset F$, where F is algebraically closed.

Every countable field has a (countable) algebraic closure.
Theorem: If k is a countable field, then k is contained in some countable algebraically closed field.

proof (E. Artin): By the assertion of the hint in ex. #81, the set S of polynomials of positive degree over k is countable. Hence we have proved there is a countable field extension $k \subset E_1$ in which every non constant polynomial in k[X] has a root. Now repeat the construction to get a countable field $E_1 \subset E_2$ such that every non constant polynomial in $E_1[X]$ has a root in E_2. Continue to construct a sequence of countable fields $k \subset E_1 \subset E_2 \subset E_3 \subset \ldots\ldots \subset E_n \subset E_{n+1} \subset \ldots\ldots$ such that every non constant polynomial in $E_n[X]$ has a root in E_{n+1}, for all n. Then let $L = \cup E_n$ be the union of all these fields. If f is a non constant polynomial in L[X] then f has only a finite number of coefficients, each of which belongs to one of the fields E_j. Hence if E_n is the largest of these E_j, f itself belongs to $E_n[X]$. Consequently, by our construction, f has a root in E_{n+1}, and hence in L. Thus L is algebraically closed. Since L is a countable union of countable fields, L is countable. QED.

Exercise #84) If $k \subset L$ is a field extension where L is algebraically closed, and if $F \subseteq L$ is the subset of all elements of L which are algebraic over k, then F is an algebraic closure of k.

Exercise #85) If k is a countable field, then k has a countable algebraic closure $k \subset \overline{k}$, and any two algebraic closures of k are k-isomorphic. [As a result of this uniqueness theorem, "the" algebraic closure of k is sometimes denoted by $\overline{k}$.]

Exercise #86) (i) Prove that there are an infinite number of irreducible polynomials over any field. [Hint: You might imitate Euclid's proof that there are an infinite number of prime integers.] (ii) Deduce from (i) that a finite field is never algebraically closed.

Q3: Does every field (even an uncountable one), have an algebraic closure?

Remark: The proof of the previous theorem shows that if Q_2 has the answer "yes", then Q_3 has the answer "yes" also.

To see how to generalize some of these results to the uncountable case, let's take another look at the construction of a field containing a root of each of two non constant polynomials. We know how to do this by iterating the quotient construction, but let's look at it slightly differently. Suppose we use X for the variable in f and Y for the variable in g. Then $f(X)$ and $g(Y)$ are non constant elements of the polynomial ring $k[X,Y]$ in two variables. We claim there must be a maximal ideal M in $k[X,Y]$ containing both $f(X)$ and $g(Y)$. To see this, let $k \subset F$ be a field extension in which both f, g have roots, say $f(\alpha) = 0$, $g(\beta) = 0$ for α, β in F, and consider the evaluation map $k[X] \rightarrow F$ taking a polynomial $h(X)$ to $h(\alpha)$. We know from the quotient construction above that the image of this map is the subfield $F_1 = k(\alpha) \subset F$ and that the kernel is the maximal ideal in $k[X]$ generated by the minimal k polynomial of α, which must be one of the irreducible factors of $f(X)$ in $k[X]$. Then let β in F be a root of $g(Y)$ and consider the evaluation map $F_1[Y] \rightarrow F$ taking $h(Y)$ to $h(\beta)$. The image of this map is the subfield $F_2 = F_1(\beta) = k(\alpha)(\beta) = k(\alpha,\beta) \subset F$. Now consider the map $k[X,Y] = k[X][Y] \rightarrow F_2$ taking $h(X,Y)$ to $h(\alpha,\beta)$. Since $k[X] \rightarrow F_1$ is surjective, so is $k[X][Y] \rightarrow F_1[Y]$. And since $F_1[Y] \rightarrow F_2$ is surjective, so is the composition $k[X,Y] \rightarrow F_2 = k(\alpha,\beta)$. This composition is just the evaluation map taking $h(X,Y)$ to $h(\alpha,\beta)$. Since $k(\alpha,\beta)$ is a field, the kernel of this surjection is a maximal ideal $M \subset k[X,Y]$, where M = those polynomials $h(X,Y)$ which equal zero when we substitute $X = \alpha$, $Y = \beta$. Since $f(\alpha) = 0 = g(\beta)$, both $f(X)$ and $g(Y)$ are contained in M as claimed.

Conversely, if M is any ideal in $k[X,Y]$ containing both $f(X)$ and $g(Y)$, then $F = k[X,Y]/M$ is a field and the natural map $k[X,Y] \rightarrow F$ sends both f, g to zero. If we write [X] and [Y] for the cosets X+M, Y+M, then this means that $f([X]) = [f(X)] = [0] = [g(Y)] = g([Y])$. Hence [X] is a root of f in F and [Y] is a root of g in F. Hence another way to construct fields in which polynomials have roots is to be able to produce maximal ideals containing those polynomials.

Here is a sketch of the construction. For each non constant polynomial f in $k[X]$, we need to introduce a root of that polynomial. Hence for each f choose a symbol X_f to represent the root of f we will adjoin In order to make the symbol X_f a root of f, first we simply adjoin X_f to k, and then we force X_f to be a root of f. To do this, first simply adjoin X_f as a new variable, and then force $f(X_f)$ to equal zero by modding out the relation $f(X_f)$ To do this all at once, we first form the huge polynomial ring $k[\ldots,X_f,\ldots]$ in the infinite collection of variables $\{...X_f...\}$, one variable for each non constant polynomial f in k[X]. Then we need to mod out by all the relations $f(X_f)$, and we also need to obtain a field. Hence we need to mod out by a maximal ideal containing all the relations $f(X_f)$. We need Zorn's Lemma to prove that such a maximal ideal exists. [End of sketch.]

Q4: Given a proper ideal $I\subset R$ in a commutative ring, does there exist a maximal (proper) ideal $M\subset R$ with $I\subset M\subset R$?

Again this is not obvious, but Zorn's Lemma shows that the argument showing the union of an increasing sequence of ideals is an ideal, implies the existence of ideals which are maximal among the family of all the ideals in a ring.

Theorem: If Q4 has answer "yes", then Q2 and Q3 also have answer "yes". I.e. if every proper ideal in every ring is contained in a maximal ideal, then every field has an algebraic closure.
Proof:
Digression: Polynomial rings can be defined with any number of variables, even infinitely many, as follows:
Definition: If S is a set of distinct "letters", a "monomial" in those letters is a product of non negative powers of a finite subset of those letters, i.e. an expression of form $X_1^{r_1}...X_n^{r_n}$, with $X_1,...,X_n$ in S and all r_j non negative integers. The order of the letters in a monomial is unimportant. We identify the trivial monomial having all $r_j = 0$ with the element 1 in k. If S is any set of letters such that $S\cap k$ is empty, then the polynomial ring k[S] is by definition the set of all finite k-linear combinations of all monomials in the letters in S. Multiplication of two monomials is, as usual, done by adding exponents of the same letters. Eg $(X^2YZ)(XY^3W^2) = X^3Y^4ZW^2$. Two polynomials are equal iff they involve the same monomials with the same coefficients. Thus distinct monomials are by definition linearly independent. This says that our construction yields a domain k[S] in which the elements of S are "independent transcendentals" over k.
End of digression.

We have seen that the Theorem follows from the next Lemma.
Lemma: If k is any field, and S is any collection of polynomials of positive degree in k[X], and if we assume proper ideals are always

contained in maximal ones, then there is a field extension $k \subset F$ in which every polynomial in S has a root.
proof: Choose a set (also called S) of letters X_f (more precisely, a set of independent transcendentals over k), one for each polynomial f in S, and form the polynomial ring k[S] (in possibly uncountably infinitely many variables) whose variables are the letters X_f.
Claim: The ideal I generated by all the polynomials $f(X_f)$ (where each f is written with its own variable X_f) is a proper ideal.
proof: If not, then 1 lies in the ideal, which would mean 1 can be written as a finite linear combination of the generators $f(X_f)$ with coefficients from k[S], i.e. $1 = \Sigma g_i f_i(X_i)$, finite sum, where all g_i are in k[S] and where X_i = the letter associated to f_i. Since a polynomial is a finite expression the finite sum $\Sigma g_i f_i(X_i)$ will involve altogether only a finite number of the variables, and hence belongs to a subring of k[S] of the form $k[X_1,.....,X_n]$ for some n. Then the equation $1 = \Sigma g_i f_i(X_i)$ would hold in that polynomial ring. However, since every f in S has positive degree, we know there is a field extension $k \subset E$ in which each of the finitely many polynomials $f_i(X_i)$ in that sum has a root, say α_i is a root in E of f_i. Then the evaluation map $k[X_1,.....,X_n] \to E$ substituting α_i for X_i takes each $f_i(X_i)$ to zero, but takes 1 to 1. Hence the right side of the equation $1 = \Sigma g_i f_i(X_i)$ goes to zero under the evaluation map while the left goes to 1, a contradiction. **QED for the claim.**
Assuming the answer to Q_4 is yes, there is a maximal ideal M in k[S] which contains I, and we can form F = k[S]/M, a field extension of k such that the natural map $k[S] \to F$ taking h to [h] = h+M, sends every one of the polynomials $f(X_f)$ to zero in F. That means as usual that $[X_f] = X_f + M$ is a root of f in F. I.e. F is a field in which all the polynomials in S have a root. **QED for the Lemma.**

It again follows that we can construct a sequence of fields $k = E_0 \subset E_1 \subset E_2 \subset E_3 \subset .....\subset E_n \subset E_{n+1} \subset .....$ such that every non constant polynomial in $E_n[X]$ has a root in E_{n+1}, for all $n \geq 0$ Then $L = \cup E_n$ is algebraically closed and contains k, and the set $\overline{k}$ of elements of L which are algebraic over k, form an algebraic closure of k.
QED. for the theorem.

Remark: Uniqueness (up to isomorphism) of the algebraic closure just constructed would follow from an affirmative answer to Q_1.

§2) Zorn's Lemma replies "yes" to questions Q_1 – Q_4 above

Let's learn the statement of Zorn's Lemma and see how to use it to settle questions Q_1–Q_4 raised above. Zorn's Lemma gives a criterion for a collection of subsets to contain a maximal subset. Of course the union of a collection of subsets is a subset containing all of them, hence maximal *if* an admissible member of the collection, but Zorn is usually applied to collections in which the union of everything in the collection is not itself in the collection. For example take the collection of all ideals in a ring. The union of even two ideals is usually not an ideal. But we have seen the union of an increasing sequence of ideals is an ideal. Zorn says essentially if you have a collection $\mathfrak{C}$ of subsets of a given set, and if for every "increasing" family of sets in $\mathfrak{C}$, $\mathfrak{C}$ contains their union, then $\mathfrak{C}$ contains maximal subsets. More generally the Lemma is stated for "partially ordered sets" which is sort of an abstract version of a collection of subsets. Since it is difficult to stay awake until the end of the definitions, we begin with a hopefully intelligible statement of the Lemma.

Zorn's Lemma: A non empty partially ordered set in which totally ordered subsets have upper bounds, contains maximal elements

Definition: A partially ordered set is a set S plus a binary relation which might only be defined for some pairs of elements of S, i.e. it is given by a subset $P \subset S \times S$, and we say $x \leq y$ for x,y in S iff the pair (x,y) belongs to P We require that if $x \leq y$ and $y \leq z$ then $x \leq z$, and that ($x \leq y$ and $y \leq x$) iff ($x = y$). We write ($x < y$) iff ($x \leq y$ and $x \neq y$). Elements x, y are called "comparable" if either $x \leq y$ or $y \leq x$.

Definition: A "linearly ordered" or "totally ordered" set S is a partially ordered set such that for every x, y in S, either $x \leq y$, or $y \leq x$.

Definition: An "upper bound" for a subset $T \subset S$ of a partially ordered set is an element b of S such that for every x in T, $x \leq b$. A "least upper bound" (sometimes written l.u.b. or lub) for T is an upper bound b for T such that if $c < b$ then c is not an upper bound for T.

Definition: A partially ordered set S is "inductively ordered" iff every totally ordered subset of S has an upper bound in S, and S is "strictly inductively ordered" iff every totally ordered subset of S has a *least* upper bound in S.

Definition: A "maximal element" of a partially ordered set S is an element b of S such that there is no element y of S with $b < y$. For example if b is not comparable with any other element of S then b is maximal.

With this language we can state:
Zorn's Lemma: Every non empty inductively ordered set contains maximal elements.

Now let's assume Zorn's lemma and use it to deduce the answer to question Q_1:
Theorem: If k is a field and $k\subset L$ is an algebraically closed extension, then L contains every algebraic extension of k, up to isomorphism.
Proof: If $k\subset F$ is any algebraic extension of k, we want to embed L into L over k. We just consider the set of all partial embeddings and use Zorn to find a maximal one. I.e. let S = the set of all pairs (E,φ) where $k\subset E\subset F$ and $\varphi:E\to L$ is an embedding over k. Since $k\subset L$ at least S contains the pair $(k,\subset)$, so S is not empty. Moreover, S is partially ordered by the relation that says $(E_1,\varphi_1) \le (E_2,\varphi_2)$ provided $E_1\subset E_2$, and φ_2 extends φ_1. Now if $T\subset S$ is a totally ordered subcollection of these pairs, we claim S contains an upper bound for T. Namely let E be the union of the subfields E_α occurring in T, and let φ be the union of the corresponding embeddings φ_α, (i.e. there is a unique k-embedding $\varphi\ E\to L$ which extends all the φ_α, and in fact the graph of φ is the union of the graphs of the φ_α). Then it is easy to see that (E,φ) is an upper bound for T, in fact a least upper bound.
Thus by Zorn there is some maximal element (E,φ) of S. We claim E = F, and therefore φ is the desired k-embedding of F into L. For if E is not equal to F and x is any element of F not in E, we can extend the map $\varphi.E\to L$ to $\psi:E(x)\to L$, by our usual extension theorem for finite extensions, since x is algebraic over k, hence also over E, so E(x) is finite over E. This says $(E,\varphi) < (E(x), \psi)$, hence (E,φ) is not maximal in S after all, a contradiction. **QED.**

Now we could answer Q4, but it will be uncountably better if you do so as an exercise.
Exercise #87) (i) Prove, assuming Zorn, that if $I\subset R$ is a proper ideal in a commutative ring, then I is contained in some maximal ideal $M\subset R$. [Hint: Obviously you must produce a maximal element

in the set S of ideals of R which contain I.]
(ii) Prove every vector space V over a field k has a basis. [Hint: Use Zorn to prove there exist maximal elements in the set S of all independent subets of V. Then prove such a maximal independent set spans V.]

Challenge problem: If $I \subset k[X_1,\ldots,X_n]$ is a proper ideal, and $M \supset I$ is a maximal ideal containing I, we know $F = k[X_1,\ldots,X_n]/M$ is a field extension of k in which all polynomials of I have a common zero, but is there an algebraic such extension?
Exercise #88) Solve the challenge problem if $n = 2$, $I = (f(X,Y))$.

§3) Hilbert's methods in the study of polynomial rings, (arguments which do not use Zorn's Lemma)

In the late nineteenth century, David Hilbert grappled with the problem of constructing finite generators for certain sets of polynomials associated to group actions on polynomials rings, (i.e. sets of polynomials with values which are essentially constant on orbits of the action, the so - called "invariants" of the action). At that time, the explicit computational methods in use were overwhelmed by the magnitude of the calculations which arose, threatening a halt in progress in the subject. Hilbert responded with new ideas, which were designed to ignore the question of "constructibility", and focus instead on the question of mere existence of solutions. This approach was very reluctantly accepted by some of Hilbert's contemporaries, and one was said to have called it "not mathematics but theology". Hilbert later made some attempt to deal with these criticisms, but his powerful non constructive methods may well have been responsible for ushering in a whole new era of abstract approaches to mathematics, in which existence theorems are proved by methods which do not allow any calculation or even estimation of the the nature of the solution which are claimed to "exist". Like it or not, this was progress, when progress by the old approaches seemed impossible.

The abstract approach to algebra which resulted was developed by Emil Artin, Emma Noether, and B.L. Van der Waerden and others in the early 20th century, and now forms the core material at the base of this course and other similar courses on "abstract algebra".

The original title "Modern Algebra" of Van der Waerden's book acknowledged the new flavor of the methods it taught. Today most of these books are titled simply "Algebra" as we have seen, but I regret the loss of even an awareness by most people that advanced algebra too once had a more computational side, which is now being rediscovered and explored with the aid of computers. In my opinion, the term "abstract algebra" might be appropriate not so much for topics which are far from everyday experience, but for those arguments in algebra which do not involve computation. I urge you to learn such methods for their value, but to always try to see how explicit computational methods can be introduced wherever possible, in order to deal with concrete questions, and to find explicit answers. As an example, the apparently abstract proof of the following theorem of Hilbert, which proves the existence of a finite basis for an ideal in a polynomial ring, but seems to give no way to construct a finite basis, turns out when analyzed appropriately to contain the germ of the idea of how to make sense out of a division algorithm for polynomials in several variables. The point is that it leads to a definition of a particularly good type of ideal basis, a "Groebner" basis This is discussed after the proof of the theorem below.

Theorem (Hilbert): If k is a field, then every ideal in $k[X_1,\dots,X_n]$ has a finite number of generators.

Corollary: *Every* non empty collection of ideals (not just inductively ordered ones) in $k[X_1,\dots,X_n]$ contains maximal elements, in particular if I is a proper ideal of $k[X_1,\dots,X_n]$, the collection of all proper ideals containing I has maximal elements Thus I is contained in some maximal ideal.

proof of corollary: If S is a non empty collection of ideals in which none are maximal, then there is an ideal I_1 in S which is not maximal in S. Thus there is an ideal I_2 in S different from I_1 and with $I_1 \subset I_2$. Again there is another ideal I_3 in S different from these with $I_1 \subset I_2 \subset I_3$. Continuing we produce a sequence of distinct ideals in S, $I_1 \subset I_2 \subset I_3 \dots\dots\dots$ But this leads to a contradiction, using the theorem above, since the union $I = \cup I_j$ is also an ideal, and is thus finitely generated. Then some one of the ideals I_n must contain all the generators [why?], whence I_n also contains I, and thus $I_m = I_n$ for all $m > n$, a contradiction to the choice of the ideals as all different.

QED.

Let's remind ourselves of the key concept of ideal generators.

Definition: If $S\subset R$ is a subset, the ideal (S) "generated by" S is the intersection of all those ideals of R which contain S. Since R is an ideal containing S, (S) is an ideal in R but need not be a *proper* ideal. Thus (S) is the smallest (not nec. proper) ideal in R containing S.

Remark: It is easy to check that (S) = the set of all (finite) R-linear combinations of elements of S, but if S is empty we have to include also 0, i.e. $(\emptyset) = (0)$. Thus for S non empty, a typical element of (S) has form $\Sigma x_i a_i$ with x_i in R and a_i in S.

Definition: A ring R is "noetherian" iff every ideal is *finitely generated*..

Exercise #89) Prove the **equivalence** of the following three properties about a ring R:

(i) R is noetherian, i.e. every ideal is finitely generated.

(ii) Every non empty collection $\mathfrak{C}$ of ideals of R has maximal elements, (i.e. ideals not contained in other ideals of $\mathfrak{C}$).

(iii) Every increasing sequence of ideals of R "stabilizes", i.e. if $I_1\subset I_2\subset I_3\subset$ .... is an infinite sequence of ideals of R with $I_j\subset I_{j+1}$ for every j, then for some n we have $I_n = I_{n+j}$ for all $j\geq 0$.

[Hint: Reread the proof of the previous corollary, and prove (i) implies (iii) implies (ii) implies (i).]

proof of the theorem: Since $k[X_1,...,X_n] \cong k[X_1,...,X_{n-1}][X_n]$, and since the only ideals in k are (0) and (1), it suffices by induction to prove the following one variable version:

Hilbert's Basis Theorem: If R is a noetherian ring then the polynomial ring R[X] is also noetherian.

Proof: Let $I\subset R[X]$ be any R[X] - ideal. We must produce a finite number of polynomials in I such that all the rest can be written as R - linear combinations of these. Since we can only use the fact that ideals in R are finitely generated, we must cook up an ideal in R from our ideal in R[X].

The first trick is to look at the "leading coefficients" of elements of I That is, let $J\subset R$ be the set of leading coefficients (the coefficients of the non zero terms of highest degree) of all polynomials in I.

Claim: $J \subset R$ is an R - ideal

proof of claim: Since 0 is in I, 0 is in J also. If a is leading coefficient of an element f of degree n in I, and b the leading coefficient of an element g of degree $m > n$, then a is also leading coefficient of the element fx^{m-n} in I. Thus a-b belong to J if non zero, since it is then leading coefficient of the element f-g in I. If a-b = 0, it still belongs to J as noted above. If r is any element of R, and ra $\neq 0$, then ra is leading coefficient of the polynomial rf in I, hence ra belongs to J. QED.

Now by hypothesis, J is finitely generated as R - ideal, so there is a finite set of elements a,b, ..,c which generate J over R. By definition of J there are finitely many corresponding elements f,g,...h of I and having a,b,....,c as leading coefficients, and we can ask the following:

Question: Do the elements f,g,.....h generate I over R[X]?

When I encountered this proof on an exam myself as a student I could only remember the "first trick" above, and so I tried to prove the answer to this question was yes, unsuccessfully as we shall see.

Let F be any element of I and try to write it as an R[X] linear combination of the elements f,g,....h. Since you can at least get the leading coefficient of F from f,g,...,h you should be able to subtract and reduce the degree of F, hopefully finishing by induction. But look what happens. First assume the degree of F is greater than n = maximum of the degrees of the polynomials f,g, ..h, and that A is the leading coefficient of F. If we multiply by appropriate non negative powers of X, we can boost up the degrees of the polynomials f,g,....h, to get polynomials f',g',....,h' in f,g,....h)⊂I which all have the same degree as F, and which still have the same leading coefficients a,b,....c as before. Since a,b,....,c generate the ideal J of all leading coefficients, we can write $A = \alpha a+\beta b+ \dots+\gamma c$, for some $\alpha,\beta,\dots,\gamma$ in R. Then $(\alpha f'+\beta g'+\dots+\gamma g')$ belongs to (f,g,....,h)⊂I, has the same leading coefficient as F, and thus $F - (\alpha f'+\beta g'+\dots+\gamma g')$ is a polynomial in I of lower degree than the degree of F. If we could continue subtracting off elements of (f,g,....h) until we get the zero polynomial, we would have proved that F belongs to (f,g,....,h).

Unfortunately we can only continue this algorithm as long as the degree of the polynomial remains larger than n (= max of degrees of all f,g,....,h). For instance if degree of $F - (\alpha f'+\beta g'+\dots+\gamma g') < N$, we would have to stop after the first step. What do we do now?

Obviously we just need some more polynomials of degrees less than n, which we can use to lower the degree even further. Use the **second trick**: Let $J_d \subset R$ be the set of all coefficients of polynomials in I having degree exactly equal to d, plus 0.

Claim: J_d is an ideal of R.

proof of claim: If u,v, are leading coefficients of polynomials φ,ψ of degree d, then u-v is either zero or the leading coefficient of the polynomial $\varphi+\psi$, which also has degree d. If r is any element of R, then ru is either zero or the leading coeffcient of $r\varphi$, also of degree d. QED.

Now for each $0 \le d \le n-1$, there is a finite set $a_d, b_d, \ldots, c_d$ of elements of J_d which generate J_d over R, and a corresponding finite set $f_d, g_d, \ldots, h_d$ of polynomials in J_d, with leading coefficients $a_d, b_d, \ldots, c_d$.

Claim: The ideal $I = (f, g, \ldots, h, f_{n-1}, g_{n-1}, \ldots, h_{n-1}, \ldots\ldots, f_0, g_0, \ldots, h_0)$.

proof of claim: We have already seen that if F is any element of I, we can find an element G of $(f, g, \ldots, h)$ such that F-G belongs to I and has degree less than n. If F-G = 0, then F belongs to $(f, g, \ldots h)$. If F-G has degree d where $0 \le d < n$, there is an R-linear combination G_d of the polynomials $\{f_d, g_d, \ldots, h_d\}$ such that $F-G-G_d$ has still lower degree Continuing, we obtain a polynomial $G+\Sigma G_d$ in the ideal $(f, g, \ldots, h, f_{n-1}, g_{n-1}, \ldots, h_{n-1}, \ldots\ldots, f_0, g_0, \ldots, h_0)$ such that $F - (G+\Sigma G_d) = 0$.

QED for Claim and for Hilbert's theorem.

Remark: If you look at it again you will see that our proof of Hilbert's theorem resembled a division process where the arbitrary element F in the ideal I, was "divided by" the generators $(f, g, \ldots, h, f_{n-1}, g_{n-1}, \ldots, h_{n-1}, \ldots\ldots, f_0, g_0, \ldots, h_0)$; i.e. multiples of these generators were subtracted from F until we got zero [remember "division" by something is subtraction of a multiple of that something]. Moreover we chose the generators so this division process would succeed when the only multipliers we used were monomials; i.e. we only needed to multiply by monomials in the proof above in order to "boost up" the degrees of the generators appropriately. The reason it all worked, is that we constructed generators with the special property that the "initial forms" of the generators were sufficient to generate the "initial forms" of the elements of I. It is conceivable that there are generators for the ideal which do not have this property, and which cannot be shown to be generators by following this same division process. This is what happens in a polynomial ring of several variables such as k[X,Y] if we are careless with the ordering of the variables or of the generators. For example we could try to divide xy^2-x by $\{xy+1, y^2-1\}$ and get $xy^2-x = y(xy+1) +0(y^2-1) + (-x-y)$, where $-x-y$ appears to be the "remainder", since it cannot be divided by either $xy+1$ or y^2-1. However, in the other order we get $xy^2-x = x(y^2-1) +0(xy+1) +0$, with remainder zero. In the second case we have proved that xy^2-x belongs to the ideal $(xy+1, y^2-1)$, but not in the first case. The problem here is that the "initial forms" xy and y^2 of the generators do not generate the initial form ($-x$ or $-y$ according to ordering) of the element $-x-y$ of the ideal. An ideal basis whose initial forms do generate all initial forms is now called a "Groebner basis" for the ideal, and it can be proved that the remainder on division is independent of the ordering of the divisors, and more important, the remainder is zero iff the element being divided does belong to the ideal generated by the divisors. Thus there is a division algorithm of sorts to determine if a polynomial of several variables belongs to the ideal generated by a given Groebner basis. More over, there is an algorithm to transform any ideal basis into a Groebner basis for the same ideal. Thus it is computationally possible to determine whether a given polynomial in $k[X_1,\ldots,X_n]$ belongs to the ideal generated by a given finite set of generators. In particular it is possible to decide whether 1 belongs to the ideal, and hence whether the ideal is a proper ideal.

The book Ideals, Varieties, and Algorithms, an introduction to computational algebraic geometry and commutative algebra, by Cox, Little, and O'Shea, discusses (at an advanced undergraduate level) Groebner bases, division algorithms, and the use of computer algebra programs to make explicit calculations in polynomial rings of several variables. It is fascinating to me that the basis turned up in our proof (which is substantially Hilbert's own proof) is a Groebner basis, and thus exactly the sort needed to make explicit division calculations in rings such as k[X,Y], and yet apparently little attention was paid to such questions until the mid 1960's when Hironaka, and then Buchberger introduced these ideas independently. The ideas in this proof thus had lain around for 50 or 60 years apparently without their computational usefulness being noticed Take heed, all aspirants to mathematical research

Exercise #90) (i) If R is a noetherian ring, then any quotient ring R/I where I⊂R is any ideal is also noetherian.
(ii) If R is a noetherian ring and $f:R\rightarrow S$ is a surjective ring homomorphism, then S is also noetherian.

§4) Unique Factorization in $\mathbf{Z}[X]$

We have already used Eisenstein's criterion, without proof, to produce irreducible polynomials over $\mathbf{Q}$. We will fill the gap in our logic by proving that result now. The essential point is Gauss' theory of the content of a polynomial, and of primitive polynomials. These concepts allow us to compare factorization in $\mathbf{Q}[X]$ with that in $\mathbf{Z}[X]$, and to deduce that an integral polynomial which is irreducible in $\mathbf{Z}[X]$ remains irreducible in $\mathbf{Q}[X]$. The contrapositive statement that an integral polynomial which is reducible in $\mathbf{Q}[X]$ is also reducible in $\mathbf{Z}[X]$ allows us to obtain (unique) factorization of polynomials in $\mathbf{Z}[X]$, and more generally also in $k[X_1,\dots,X_n]$.

Theorem (Gauss): If R is a ufd, then R[X] is a ufd also.

This is the most general statement we shall prove, in the next section, but we shall proceed to the proof in stages, first proving that $\mathbf{Z}[X]$ is a ufd. This proof contains all the essential ideas. It is simple in principle, but the details are tedious to do completely We will attempt to make the main ideas clear, and we will also try to present essentially all the details. First of all, think back over your own experience, factoring things like $x^2+5x+6 = (x+2)(x+3)$. Notice that when the coefficients of the original polynomial are integers, then the coefficients of the factors are also integers. To be sure, you *can* factor $x^2-2 = (x-2^{1/2})(x+2^{1/2})$ with *irrational* numbers. But if an integral poynomial factors with rationa numbers, then it already factors with integers. This is one of the first results we shall prove using Gauss' idea of "content".
Very briefly then, to factor an integral polynomial f over $\mathbf{Z}[X]$, for example $f = 6x^2-30x+36$, just remove the gcd of the coefficients (this is the "content"), here $f = 6(x^2+5x+6)$, then factor separately the content and the remaining polynomial $f = (2)(3)(x+2)(x+3)$, and these are the irreducible factors of f over $\mathbf{Z}[x]$. Note there are four irreducible factors here since 2, 3 are not units, but primes in $\mathbf{Z}[x]$.

The most important concept is the following one:
Definition: The "content" of a polynomial f in **Z**[X] is the gcd of the coefficients of f. If f=0, the content is 0. We denote content(f) = c_f.

Thus c_f is a well defined non negative integer which is zero iff f =0

Definition: A polynomial f in **Z**[X] is "primitive" iff c_f = 1, iff the coefficients of f have no common prime integral factor.

Lemma: Let f,g,h, be non zero polynomials in **Z**[X].
(i) c in $\mathbf{Z}^+$ is the content of f iff f = cg where g is primitive.
(ii) If c,d are in $\mathbf{Z}^+$, g,h in **Z**[X] are primitive, and cg = dh, then c=d and g=h.
(iii) Every non zero f in **Z**[X] has an unique associated primitive polynomial f_0 such that f = $c_f(f_0)$. [Or if f=0, take f_0=1.]
(iv) If f≠0 in **Z**[X], and f = cg where c is in $\mathbf{Z}^+$ and g is primitive, then c = c_f and g = f_0.
proof: Exercise. QED.

The main property of the content is that it is multiplicative. We prove this in the following steps.
Lemma: If g,h are primitive in **Z**[X] and f = gh, then f is primitive
proof: If p is any prime integer, it suffices to prove that some coefficient of f is not divisible by p Since this is true for both g and h, among the coefficients of g which are not divisible by p there is a highest one say a_r, and similarly a highest one among the coefficient of h not divisible by p, say b_s. Then the coefficient c of X^{r+s} in f is a sum of terms, of which one is a_rb_s and the others are of form a_kb_l where k+l = r+s. Hence except when k = r, l = s, we must have k > r or l > s. In these cases, either p divides a_k or b_l and thus their product. Hence p divides every term but one in the coefficient c of X^{r+s}, and hence p does not divide c. QED.

[**Digression**: Another very nice proof of this lemma is possible using two natural auxiliary results. We give it as well

Lemma: If R is a domain, so is R[X].
proof: The key point is that although the coefficients of a product are in general a *sum* of products of various coefficients of the

factors, both the *highest* coefficient and the *lowest* coefficients of fg are a simple product. We can use either one for this argument. Eg. the highest coefficient of fg is the product of the highest coefficient of f and the highest coefficient of g. If f,g are both non zero, then their highest coefficients are both non zero, hence their product is non zero. Since fg thus has non zero highest coefficient, it is non zero as well. QED.

Lemma: The construction of polynomial rings is a functor from rings to rings. In particular, if f: $R \to S$ is a ring map, then there is a unique induced ring map $f^*:R[X] \to S[X]$ such that $f^*(\Sigma a_i X^i) = \Sigma f(a_i)X^i$.
proof: This is easy. QED.

Lemma: If g,h are primitive in $\mathbf{Z}[X]$ and f = gh, then f is primitive.
second proof: Again it suffices to show no prime integer p divides every coefficient of f, i.e. that the reduced polynomial [f] is not zero in the ring $\mathbf{Z}_p[X]$. The hypothesis says neither [g] nor [h] is zero in $\mathbf{Z}_p[X]$. Moreover, by the previous two lemmas, since $\mathbf{Z}_p$ is a domain, so is $\mathbf{Z}_p[X]$, hence $[f] = [g][h] \neq [0]$. QED.

Remark: This second proof is really the same as the first proof, since in the first proof the coefficient c of X^{r+s} which we showed was not divisible by p, becomes precisely the highest non vanishing coefficient of the reduced polynomial mod p. **End of Digression**.]

Lemma: If f,g,h are in $\mathbf{Z}[X]$ and f = gh, then $c_f = (c_g)(c_h)$
proof: In the notation introduced above we have $g = c_g(g_0)$, and $h = c_h(h_0)$, whence $f = gh = c_g c_h(g_0 h_0)$, where $g_0 h_0$ is primitive. Thus by the properties given above for content, $c_f = c_g c_h$. QED.

Lemma: If a, b, c, d are in $\mathbf{Z}^+$, and g, h are primitive (in $\mathbf{Z}[X]$), and if (a/b)g = (c/d)h, then a/b = c/d, and g = h.
proof: Multiplying by bd, we conclude that adg = bch, whence the properties above of content imply ad = bc, hence a/b = c/d. Dividing through by a/b = c/d, then g = h. QED.

Remark: If f is non zero in $\mathbf{Q}[X]$, there exist a,b in $\mathbf{Z}^+$ and a primitive g in $\mathbf{Z}[X]$ such that f = (a/b)g, since we may take b as a positive common multiple of the denominators of the coefficients of f, and a = content(bf), where bf is in **Z**[X]. By the previous lemma, a/b and g are unique.

Definition: For any non zero f in Q[X] the content is the unique positive element c_f of Q such that f = $c_f(f_0)$ where f_0 is primitive in **Z**[X]. The unique such f_0 is called the "primitive form" of f.

Multiplicativity holds also for rational contents.
Lemma: For any g,h in Q[X], if f=gh then $c_f = c_g c_h$.
proof: The proof is the same as for integral contents. QED.

It follows that the "primitive form" is also multiplicative:
Lemma: For any g,h in Q[X], $(gh)_0 = (g_0)(h_0)$.
proof: The lemmas imply that $(gh)_0$ is the unique primitive polynomial P such that gh is a positive rational multiple of P. But gh = $c_g(g_0)c_h(h_0) = c_g c_h(g_0 h_0)$, where c_g, c_h are positive and rational and $(g_0 h_0)$ is primitive. QED.

Now we can go through the proof that **Z**[X] is a ufd, by replacing every polynomial by its primitive form whenever possible. The point is that the primitive polynomials have the same divisibility properties in Q[X] as in **Z**[X], allowing us to bring unique factorization down from Q[X] to **Z**[X]. (We emphasize that primitive polynomials are always elements of **Z**[X], and the only constant primitive polynomials are 1, -1.) More precisely.

Lemma: If f is primitive, then f is reducible in **Z**[X] iff f is reducible in Q[X]. In fact if f = gh, with g,h non units in Q[X], then also f = $(g_0)(h_0)$, where g_0, h_0 are the primitive forms of g,h.
proof: If f is reducible in **Z**[X], f = gh, then both g,h have degree ≥1 since f is primitive, hence g,h are non units in Q[X] and f is reducible in Q[X]. If f = gh, with g,h non units in Q[X], by multiplicativity of primitive forms we have f = $f_0 = (g_0)(h_0)$, so f is reducible in **Z**[X]. QED.

Remark: The previous lemma fails in one direction for non primitive polynomials; eg. 3X+3 is reducible in **Z**[X] but not in Q[X] It still holds in the other direction, as the next lemma shows.

Lemma: If f in $\mathbf{Z}[X]$ is reducible in $\mathbf{Q}[X]$, f is also reducible in $\mathbf{Z}[X]$ and can be factored into factors of degree ≥ 1 in $\mathbf{Z}[X]$.
proof: If f = gh, with f in $\mathbf{Z}[X]$ and g,h of degree ≥ 1 in $\mathbf{Q}[X]$, then $c_f(f_0) = f = gh = c_g c_h(g_0 h_0)$. Thus $c_g c_h = c_f$ is an integer, and f = $c_f(g_0 h_0)$ is a factorization of f over $\mathbf{Z}[X]$ with degrees $g_0, h_0 \geq 1$. , QED.

The previous result allows us to prove Eisenstein's criterion.
Eisenstein criterion: If $f = a_0 + a_1X + a_2X^2 + \ldots + a_nX^n$, where all a_i are in $\mathbf{Z}$, and if there exists a prime integer p such that $p|a_i$ for all $i < n$, but p does not divide a_n, and if p^2 does not divide a_0, then f is irreducible in $\mathbf{Q}[X]$.
proof: If not, then f = gh, where g, h are in $\mathbf{Z}[X]$ of degree ≥ 1. Thus $f = (a_0 + a_1X + \ldots + a_nX^n) = gh =$
$(b_0 + b_1X + \ldots + b_rX^r)(c_0 + c_1X + \ldots + c_sX^s)$, where r+s = n = degree(f), and reduce all polynomials in $\mathbf{Z}_p[X]$. By hypothesis, $[f] = [a_n]X^n$, where $[a_n] \neq [0]$ in $\mathbf{Z}_p$. Then we have $[g][h] = [f] =[a_n]X^n$, and since the only prime factor of this element is X, then by uniqueness of prime factorization in $\mathbf{Z}_p[X]$, X is the only prime factor that can occur in [g], [h]. Hence we must have $[g] = [b_r]X^r$, $[h] = [c_s]X^s$where r+s = n. Since $r,s \geq 1$, $[b_0] = [c_0] = [0]$. But then p divides both b_0 and c_0, hence p^2 divides a_0, contrary to hypothesis. QED.

Remark: Note that since $\mathbf{Z}$ is a domain, deg(fg) = deg(f) + deg(g), so constants in $\mathbf{Z}$ can have only constant factors, hence prime integers in $\mathbf{Z}$ are also irreducible in $\mathbf{Z}[X]$.

Now we can prove existence of factorization into irreducibles in $\mathbf{Z}[X]$.
Lemma: Every non zero, non unit element of $\mathbf{Z}[X]$ can be factored into irreducible elements.
proof: Let f be non zero, non unit in $\mathbf{Z}[X]$. If f is in $\mathbf{Z}$, then the previous remark shows the prime factorization in $\mathbf{Z}$ gives a factorization into irreducibles in $\mathbf{Z}[X]$. If $\deg(f) \geq 1$, factor it as $f = c_f(f_0)$. Then $f_0 = \prod g_i$, with g_i irreducible in $\mathbf{Q}[X]$. By the previous lemmas, then $f_0 = \prod (g_i)_0$, where the $(g_i)_0$ are primitive forms of the g_i. Then each $(g_i)_0$ is a non zero rational multiple of g_i, hence still irreducible in $\mathbf{Q}[X]$ and also primitive, hence irreducible in $\mathbf{Z}[X]$, by our lemma above. Factoring $c_f = \prod p_i$ into primes in $\mathbf{Z}$ gives us the

factorization of $f = \prod p_i \prod (g_i)_0$ into irreducibles in $\mathbb{Z}[X]$. QED.

Remark: Existence of irreducible factorizations is not really the hard part of the theory in this case, since we already know $\mathbb{Z}[X]$ is a noetherian domain, and it can be proved easily that factorization into irreducibles is always possible in any noetherian domain. The uniqueness however is not always true in a noetherian domain. The proof just given of existence of factorizations in $\mathbb{Z}[X]$ will also work in R[X] where R is a non noetherian ufd.

We need one more technical property of primitive polynomials
Lemma: If f in $\mathbb{Z}[X]$ is primitive, and g is in $\mathbb{Z}[X]$, then f|g in $\mathbb{Z}[X]$ iff f|g in $\mathbb{Q}[X]$.
proof: If f|g in $\mathbb{Z}[X]$ then $g = fh$, for h in $\mathbb{Z}[X] \subset \mathbb{Q}[X]$, hence f|g in $\mathbb{Q}[X]$. And if f|g in $\mathbb{Q}[X]$, then $g = fh$, for h in $\mathbb{Q}[X]$. Then $c_g = c_f c_h = c_h$, so $c_h = c_g$ is an integer. Then $h = c_h (h_0)$ is in $\mathbb{Z}[X]$, so f divides g in $\mathbb{Z}[X]$. QED.

Remark: Again one direction fails for non primitive polynomials, since 3X+3 divides X+1 in $\mathbb{Q}[X]$, but not in $\mathbb{Z}[X]$.

The next property is the key to proving uniqueness of factorization.
Lemma: If f,g,h are in $\mathbb{Z}[X]$, f is irreducible, and f|gh, then f|g or f|h.
proof: First note that an integer c divides a polynomial F in $\mathbb{Z}[X]$ iff c divides all the coefficients of F, iff $c|c_F$. Hence if f is irreducible in $\mathbb{Z}[X]$ and an integer, f=p is prime in $\mathbb{Z}$. Then if p divides gh, p divides $c_{gh} = c_g c_h$, so p divides either c_g or c_h by the corresponding lemma in $\mathbb{Z}$. Hence f=p divides either g or h. That settles this case.
If $\deg(f) \geq 1$, then f irreducible implies f is primitive, hence f is also irreducible in $\mathbb{Q}[X]$ and divides gh also in $\mathbb{Q}[X]$ Since the present lemma holds in the Euclidean domain $\mathbb{Q}[X]$, f divides either g or h in $\mathbb{Q}[X]$. Since f is primitive, then f divides either g or h also in $\mathbb{Z}[X]$
QED.

Lemma: Factorization into irreducibles is unique in $\mathbb{Z}[X]$, up to order of factors and sign.
proof(same proof as in $\mathbb{Z}$): If $\prod g_i = \prod h_j$ where all g_i, h_j are irreducible in $\mathbb{Z}[X]$, then g_1 divides the left side, hence also the right, so by the previous lemma g_1 divides some h_j which we may

renumber as h_1. Since h_1 is irreducible, and the only units in $\mathbb{Z}[X]$ are ± 1, then $h_1 = \pm g_1$. Hence we may cancel g_1 from both sides leaving $(g_2)(\ldots..)(g_n) = \pm(h_2)(\ldots.)(h_m)$. Continuing with $g_2, \ldots..$, we eventually cancel all terms. I.e. there are the same number of g's and h's and after renumbering the indices, for every i, $g_i = \pm h_i$. If you want the proof to appear more rigorous, use induction on the number n of factors g_i. If there is only one g_i there can be only one h_j since g_i is irreducible. This proves the result for $n=1$. Assuming the theorem for n-1 factors g_i, we are done after we cancel g_1 from both sides as above, since then by induction $n-1 = m-1$, hence $n=m$ and the factors $g_2, \ldots., g_n$ must agree with the factors $\pm h_2, \ldots., h_n$ up to order and multiplication by units. QED.

§5) Proof that if R is any ufd, then R[X] is a ufd also.

This proof is exactly analogous to the previous one for $\mathbb{Z}[X]$, except that we must define the gcd in an arbitrary ufd. It can be done, but the fact that there may be many units prevents us from defining a unique gcd. Hence the gcd, the content, and the primitive form of a polynomial, will only be defined up to multiplication by units. Nothing in the proof will be affected essentially by this since every statement is merely that some element f divides some other element g. Multiplying either f or g, or both, by units does not affect the divisibility of g by f. Hence we can just go back and copy the whole proof. We only sketch the highlights and recall the definitions of the main concepts. We also make a few remarks about how to take advantage of the noetherian property when it is present. Remember all our rings are commutative.

The most important subset of a ring R is the group R^* of "units".
Definition: An element x of a ring is a <u>unit</u> iff it has a multiplicative inverse y in R, i.e. iff there is an element y in R such that $xy = 1$.

Definition: A non zero element x of a domain R is called <u>irreducible</u> iff x is not a unit and whenever $x = yz$, with y,z in R, then either y or z is a unit. Equivalently, x is irreducible iff x is not a unit and when $x = yz$, then either x divides y or x divides z.

Remark: An element $x \neq 0$ is irreducible in R iff the ideal $(x) \subset R$ is not contained in any other (proper) principal ideal, i.e. iff (x) is maximal among (proper) principal ideals.

Definition: A non zero element x of a domain R is <u>prime</u> iff x is not a unit and whenever $x|yz$, with y,z in R, then either $x|y$ or $x|z$. Equivalently x is prime iff R/(x) is a domain.

It is obvious (from the last sentence in the definition of an irreducible element) that (in a domain) every prime element is irreducible

Definition: A <u>unique factorization domain</u>, or ufd for short, is a domain in which every non-zero, non-unit element can be expressed as a (finite) product of irreducible elements in a way which is unique except for order of elements and multiplication by units. I.e. if $\prod_I p_i = \prod_J q_j$, where all p's and all q's are irreducible, then there must be a bijection between the index sets $\sigma: I \rightarrow J$, such that for every i in I, if $j = \sigma(i)$, then $p_i = u_i q_j$ for some unit u_i.

Lemma: If R is a domain in which every non-zero, non-unit element can be expressed as a product of irreducible elements, then R is a ufd iff every irreducible element is prime.
proof: A moment's thought, or a review of the proof of uniqueness of factorization in **Z**, or in **Z**[X], will show that if irreducible elements satisfy the definition just given of prime elements, then factorization into irreducibles is unique up to order and units.

Conversely, irreducibles are prime in any ufd, since if a is irreducible and divides bc, then we can write $ax = bc$, for some x. Since x has a factorization into irreducibles $x = \prod p_i$ we get $a\prod p_i = bc$, and since b,c also have such factorizations $b = \prod q_j$, $c = \prod r_k$, we get $a\prod p_i = \prod q_j \prod r_k$. Since all factors are irreducible, by uniqueness a must be a unit times one of the irreducible factors q_j, or r_k. Thus a divides either b or c. QED.

We pause for a result that implies that *existence* of factorization into irreducibles is very often true.

Definition: In a ring R, two elements a,b are called "associates" iff a = ub for some unit u in R^*.

Remark: In a domain R, two elements a,b are associates iff they generate the same principal ideal.
proof: If (a) = (b), then a = bx for some x, and b = ay for some y. Hence a = bx = ayx = a1, so yx = 1, and both x,y, are units. The converse is even more trivial. QED.

Lemma: If R is any noetherian domain, then every non zero, non unit can be expressed as a product of irreducibles.
proof: Suppose not, and that x is a non zero, non unit element of R which does not factor into irreducibles. Then x is not irreducible, so x factors into $x = a_1b_1$, with neither factor a unit. If both a_1, b_1 factor into irreducibles then so does x, hence at least one of them does not, and hence that one, say b_1 can be factored into two new non unit factors $b_1 = a_2b_2$, such that at least one of these new factors does not itself factor into irreducibles. Suppose b_2 does not. Then again we have $b_2 = a_3b_3$, and we can continue,........ We obtain an infinite sequence of elements a_n, b_n, where no a_i nor b_i is a unit. Since for every n, $b_{n-1} = a_nb_n$, we see that b_n divides b_{n-1} for every n, but the quotient a_n is not a unit Thus the ideals (b_n) are all distinct. This yields an increasing sequence of distinct ideals $(b_1) \subset (b_2) \subset ....\subset (b_n) \subset ....$, in contradiction to the assumption that the ring is noetherian. QED.

Corollary: A noetherian domain is a ufd iff every irreducible element is prime.

We already "know" that a domain in which the division algorithm holds (see below for a definition of the relevant division algorithm) is both a pid and a ufd. We show now that in fact every pid is a ufd.

Theorem: If R is any pid, then R is a ufd.
proof: Certainly every pid is noetherian, so we just need to show every irreducible element in a pid is prime. If x is irreducible, we must show (x) is a prime ideal i.e. that R/(x) is a domain. We have already observed that x irreducible means (x) is maximal among principal proper ideals. Since all ideals are principal, (x) is maximal,

hence $R/(x)$ is a field and thus a domain. QED.

Definition: A domain is called a Euclidean domain if the division algorithm holds in the following form: to each non zero element a of R there is associated a non negative integer $d(a)$, such that
(i) for a,b non zero in R, we have $d(a) \le d(ab)$,
(ii) for a,b non zero in R, there exist t,r in R such that $a = bt + r$, and either $r=0$ or $d(r) < d(b)$. [Note: uniqueness of t,r, is not required.]

Corollary: For any domain R, Euclidean ⇒ PID ⇒ UFD.
proof: To prove that a Euclidean domain is a pid, imitate the proof that $k[X]$ is a pid. I.e. given any ideal I, consider the element a of I with $d(a)$ as small as possible. Then for any b in I, divide b by a to get $b = sa + r$, with either $r=0$ or $d(r) < d(a)$. Since $r = b-sa$ belongs to I, $d(a) \le d(r)$, so we must have $r=0$. QED.

Remark: If $R = k[X,Y,Z,W]/(XY-ZW)$, then R is a noetherian domain, hence factorization into irreducibles is always possible in R. However $[X][Y] = [Z][W]$ in R, seems to give two distinct factorizations of the same element into irreducibles, suggesting R is a noetherian domain which is not a ufd. On the other hand it seems $k[X_1,....,X_n,....]$ (infinitely many variables), is a non noetherian ufd.

Now we return to the proof of Gauss' theorem. The main point is to clarify the concept of gcd. The definition is the same as before.
Definition: In any domain R, d is a gcd of a,b iff $d|a$, $d|b$ and whenever $c|a$ and $c|b$, then $c|d$. I.e. d is a common divisor and the only other common divisors are its factors. A gcd is not unique, but is determined up to multiplication by units in R

Lemma: In any Euclidean domain R, c is a gcd of a, b iff there exist elements x,y of R such that $c = xa+by$, and if among all such c, $d(c)$ is minimal. Such an element c can be calculated by Euclid's algorithm. In any pid R, c is a gcd of a,b iff the ideals $(c) = (a,b)$ are equal. In any ufd, let $a = \prod p_i^{r_i}$, $b = \prod p_i^{s_i}$ be prime factorizations of a,b, where we allow some r_i, s_i to be 0 (in order to use the same set of primes in both products). Then $c = \prod p_i^{\min(r_i,s_i)}$ is a gcd of a,b, and all gcd's arise this way.
proof: Exercise. QED.

Definition: If R is a ufd and f is in R[X], the content c_f of f is any gcd of the coefficients of f. If f = 0 then $c_f = 0$. Thus c_f is not uniquely defined, but any two values of c_f have ratio in R^*.

Definition: If R is a ufd, a polynomial f in R[X] is "primitive" iff $c_f = 1$, (i.e. iff 1 is a gcd of the coefficients), iff every c_f is a unit.

Lemma: If f is in R[X], c_f is a content of f, c is in R, then $c = uc_f$ for some unit u, iff f = cg where g is primitive.
proof: This follows from the fact that c is a gcd of a set of elements of R iff c is a common factor and after it is factored out, it leaves the numbers relatively prime. QED.

Cor: If c,d are in R and g,h, are primitive and cg = dh, then c = ud, $g = u^{-1}h$, for some unit u in $R^* \subset R$.
proof: Both c, d are contents of f, hence equal up to unit multiples. QED.

Cor: For any non zero element f of R[X] there are associated primitive forms f_0, determined up to unit multiple by the equation $f = c_f(f_0)$

Lemma: If f = gh in R[X] with contents c_f, c_g, c_h, then $c_f = uc_gc_h$, i.e. the content is multiplicative up to units. Similarly $f_0 = u^{-1}g_0h_0$.
proof: As before. QED.

Let F be the quotient field of R, **F** = q.f.(R). Then we can define as before, c_f in F for any f in F[X], unique up to unit multiples in R^*, and all the properties hold as for content of polynomials in R[X].

We get lemmas analogous to those above:
Lemma: If f is primitive, f is reducible in R[X] iff reducible in F[X].
proof: If f = gh, f in R[X], g,h in F[X], then $f = f_0 = c_gc_h(g_0h_0)$, whence $c_f = 1 = uc_gc_h$, so $c_gc_h = u^{-1}$ belongs to R. Thus $f = (u^{-1}g_0)(h_0)$ is a factorization in R[X]. QED.

Just as for **Z**[X], the previous lemma gives factorization of elements of R[X] into irreducibles, and the next lemmas give uniqueness.

Lemma: If f is primitive, g is in R[X], then f|g in R[X] iff f|g in F[X].
proof: Suppose f|g in F[X] so that g = fh with h in F[X]. Then $c_g = uc_fc_h$ for u,cf in R*, so $c_h = c_gu^{-1}c_f^{-1}$ is in R. Hence so is $h = c_hh_0$. The other direction is easier. **QED.**

Lemma: If f is irreducible in R[X] then f is prime.
proof: Assume degree(f) ≥ 1, then f is primitive. If f|gh in R[X] then in F[X] the same holds whence f divides either g or h in F[X]. Then f divides one of them also in R[X]. If degree(f) = 0, and f divides gh then f divides c_gc_h, and f is irreducible in R. Then f is prime in R, so f divides either c_g or c_h, hence f divides either g or h. **QED.**

We see that the irreducible elements of R[X] are exactly the prime elements, and consist of precisely the primes of R plus the irreducible primitive polynomials of R[X], the latter being the primitive forms of irreducible polynomials of F[X].

Corollary: If k is a field then $k[X_1,...,X_n]$ is a ufd.

Exercise #91) (i) Prove R = k[X,Y,Z,W]/(XY-ZW), is a noetherian domain, [hence factorization into irreducibles is always possible].
(ii) Prove [X] is a non unit, and irreducible in R.
(iii) Prove [Z] is not a multiple of [X] in R
(iv) Deduce that [X] is not prime in R, hence R is not a ufd.

Exercise #92) Prove $k[X_1,...,X_n,...]$ (polynomial ring in infinitely many variables), is a non noetherian ufd.

Exercise #93) Prove $k[X_1,...,X_n]$ is not a pid if n≥2. [Thus there are many ufd's that are not pid's.]

Definition: A proper ideal I⊂R is "prime" iff whenever xy is in I, for x,y, in R, then either x or y is in I, iff R/I is a domain. In particular (0) is prime iff R is a domain.

Definition: Let R be a domain. A prime ideal I≠(0) in R has "height one" iff (0) is the only prime ideal strictly contained in I. A prime ideal I⊂R has "height r" iff there is a sequence of r-1 distinct prime ideals $I_1 \subset I_2 \subset ... \subset I_{r-1}$ properly contained in I, where I_1 has height one, but there is no such sequence of r distinct prime ideals properly contained in I.

Exercise #94) Let R be a ufd.
(i) Prove that every irreducible element α in R generates a principal height one prime ideal $(\alpha)\subset R$.
(ii) Prove that every height one prime ideal $I\subset R$ is principal and generated by an irreducible element α in R.

Remark: If R is a noetherian domain in which every height one prime ideal is principal, then R is a ufd, but my proof of this uses two big theorems we don't have yet, primary decomposition, and Krull's principal ideal theorem.

Exercise #95) (i) In a Euclidean domain R, prove c is a gcd of a,b iff there exist elements x,y of R such that c = xa+by, and if among all such c, d(c) is minimal. (ii) Prove such an element c can be calculated by Euclid's algorithm.

Exercise #96) Prove: In a pid R, c is a gcd of a,b iff the ideals (c) = (a,b) are equal.

Exercise #97) In a ufd R, let $a=\prod p_i^{r_i}$, $b=\prod p_i^{s_i}$ be prime factorizations of a,b, where we allow some r_i , s_i to be 0 (in order to use the same set of primes in both products). Prove $c = \prod p_i^{\min(r_i,s_i)}$ is a gcd of a,b, and all gcd's arise this way.

§6) A Diophantine Puzzle

We give an application of ring theoretic methods to study a classical problem: which primes p in **Z** are sums of two squares? Trial and error suggests that the answer is $p = 2 = 1^2+1^2$, and those p which are congruent to 1(mod 4), such as $5 = 1^2+2^2$, $13 = 2^2+3^2$, $17 = 1^1+4^2$, $29 = 5^2+2^2$, $37 = 6^2+1^2$, $41 = 5^2+4^2$, $53 = 2^2+7^2$,.......

Assume the elementary fact from number theory that x^2+1 has roots in $\mathbf{Z}_p$ for precisely such p. Can we make a link between these two problems? I.e. can we prove somehow that p is a sum of two squares iff X^2+1 has roots mod p? Consider the following argument:

Notice that if $p=a^2+b^2$, then using complex numbers we get $p = (a+bi)(a-bi)$, so p is no longer prime in the ring $\mathbf{Z}[i]$. Conversely, if $p = (a+bi)(c+di)$ in $\mathbf{Z}[i]$ then taking absolute values on both sides and squaring, we get $p^2 = (a^2+b^2)(c^2+d^2)$, and by uniqueness of prime factorization in $\mathbf{Z}$, there are only two prime factors on the right, both equal to p, hence $p = a^2+b^2$. Thus a prime p in $\mathbf{Z}$ is no longer prime in $\mathbf{Z}[i]$ iff p is a sum of two squares.

Now introduce $\mathbf{Z}_p[i]$ as follows: Since $\mathbf{Z}[i]/(p) \cong \mathbf{Z}_p[i] \cong \mathbf{Z}_p[X]/(X^2+1)$, then p is a sum of two squares iff p not prime in $\mathbf{Z}[i]$ iff $\mathbf{Z}_p[i]$ is not a domain iff X^2+1 is not prime in $\mathbf{Z}_p[X]$ iff X^2+1 has roots mod p iff $p = 2$ or $p \equiv 1 \pmod 4$. Ok? Now consider:

Puzzle: If we use these ideas to analyze the equation $X^2+5Y^2 = p$, we seem to get that $p = a^2+5b^2$ for some a,b iff p is not prime in $\mathbf{Z}[\sqrt{-5}]$ iff $\mathbf{Z}_p[\sqrt{-5}]$ not a domain iff X^2+5 not prime in $\mathbf{Z}_p[X]$ iff X^2+5 has roots mod p. But what about $p = 3$? Then $X = 1$ is a root of $X^2+5 \pmod 3$, but obviously $X^2+5Y^2 = 3$ has no integral solution. What gives?

While you think about it, we prove the elementary fact from number theory used above

Theorem: The polynomial x^2+1 has roots in $\mathbf{Z}_p$ iff $p = 2$ or $p \equiv 1 \pmod 4$.

proof: Existence of roots: For any prime p, $\mathbf{Z}_p$ is a field, so there are exactly two roots of X^2-1 in $\mathbf{Z}_p$, namely $1, -1$. Thus for every other non zero element x of $\mathbf{Z}_p$, x and x^{-1} are distinct. Hence in the product of all non zero elements of $\mathbf{Z}_p$, these pairs cancel, so the whole product equals $(1)(-1)(xx^{-1})(\ldots)(yy^{-1}) = -1$. On the other hand if $p>2$, 2 is invertible mod p, and thus for all $x\neq 0$, we have $x+x=(1+1)x\neq 0$. Hence for all $x\neq 0$, $x\neq -x$, and the product of all non zero elements of $\mathbf{Z}_p$ has form $(-1) = \prod(x_i)(-x_i) = (-1)^{(p-1)/2}y^2$, where $y = \prod x_i$. Finally $(-1)^{(p-1)/2} = 1$ iff $(p-1)/2$ is even, iff $p-1$ is divisible by 4, iff $p\equiv 1 (\text{mod} 4)$. Thus X^2+1 has a root in $\mathbf{Z}_p$ if $p \equiv 1 (\text{mod} 4)$ Since $1^2+1 = 0 \bmod 2$, we have proved one direction. Conversely if there is a solution α of $x^2+1 = 0 \bmod p$, then $\alpha^2 = -1 \pmod p$. Since $\mathbf{Z}_p^*$ is a group of order $p-1$, any element raised to the power $p-1$ equals 1 [by LaGrange's theorem last fall], so $1 = \alpha^{p-1}$

$= (\alpha^2)^{(p-1)/2} = (-1)^{(p-1)/2}$ (mod p). Thus we must have either $1 = -1$, hence $p = 2$, or else $(p-1)/2$ must be even, whence $p \equiv 1$(mod 4). **QED.**

Now let's examine the false argument given above more closely. The essential point is to be careful about distinguishing between the properties "prime" and "irreducible", which are not always the same.

Theorem: The equation $x^2+y^2 = p$, has solutions x,y in **Z** iff $p = 2$ or $p \equiv 1$(mod 4).

proof: Notice that if $p = a^2+b^2$, then using complex numbers we get $p = (a+bi)(a-bi)$, so p is no longer *irreducible* in the ring **Z**[i]. Conversely, if $p = (a+bi)(c+di)$ in **Z**[i] then taking absolute values on both sides and squaring, we get $p^2 = (a^2+b^2)(c^2+d^2)$, and by uniqueness of prime factorization in **Z**, there are only two prime factors on the right, both equal to p, hence $p = a^2+b^2$. Thus a prime p in **Z** is no longer irreducible in **Z**[i] iff p is a sum of two squares.

Now recall $\mathbf{Z}_p(i) = \mathbf{Z}_p[i] = \{a+bi$, for a,b in $\mathbf{Z}_p$, where i is not in $\mathbf{Z}_p$ but $i^2 = -1$ in $\mathbf{Z}_p\}$, and consider the following argument:
Since $\mathbf{Z}_p[i] \cong \mathbf{Z}_p[X]/(X^2+1)$, and $\mathbf{Z}_p[X]$ is a ufd, it follows that $\mathbf{Z}_p[i]$ is not a domain, iff X^2+1 is not prime in $\mathbf{Z}_p[X]$, iff X^2+1 is reducible in $\mathbf{Z}_p[X]$, iff X^2+1 has roots mod p.
Since on the other hand $\mathbf{Z}_p[i] \cong \mathbf{Z}[i]/(p)$, then $\mathbf{Z}_p[i]$ is not a domain iff p is not prime in **Z**[i]. Thus X^2+1 has roots mod p iff p is not prime in **Z**[i], while (from above) p is a sum of two squares iff p is reducible in **Z**[i]. Thus we have not quite succeeded in equating the two problems. I.e. to equate the problem of p being a sum of two squares with that of X^2+1 having roots mod p, we must equate the property of p being prime in **Z**[i] with that of p being irreducible. This means we could finish if we only knew **Z**[i] is a ufd! More is true, **Z**[i] is actually a Euclidean domain (see Ex. 87 below). Assuming this result, we have proved that those primes p which are sums of two squares are precisely $p = 2$, and $p \equiv 1$(mod4). **QED.**

Remark: (i) One direction of the proof above is actually much easier by direct methods than by our approach. Namely if $a^2+b^2 =$

p, then neither a nor b can be divisible by p [if a is, then b is too, then the left side is divisible by p^2, contradiction], so if we reduce mod p, we have $[a]^2+[b]^2 = [0]$, with both $[a],[b] \neq [0]$. Dividing by $[b]^2$, thus $[a/b]^2+[1] = [0]$ in $\mathbf{Z}_p$, so X^2+1 has solution $X = [a/b]$ in $\mathbf{Z}_p$.
(ii) Since the analogous argument gave an incorrect result for the equation $X^2+5Y^2 = p$, presumably $\mathbf{Z}[\sqrt{-5}]$ is not a ufd.

Exercise #98) Define a size function δ on $\mathbf{Z}[i]$ by $\delta(a+bi) = a^2+b^2$.
(i) Prove $\delta(zw) = \delta(z)\delta(w)$, and $\delta(z) \geq 1$ for $z \neq 0$ in $\mathbf{Z}[i]$.
(ii) Given two numbers in $\mathbf{Z}[i]$, $z = a+bi$, and $w = c+di$, if $z \neq 0$, prove there exist numbers $q = e+fi$, and $r = s+ti$, in $\mathbf{Z}[i]$, such that $w = zq + r$, and $\delta(r) < \delta(z)$. More precisely, we have $c+di = (a+bi)(e+fi) + (s+ti)$, and $s^2+t^2 < a^2+b^2$.
[Hint: If $z = a+bi$ is in $\mathbf{Z}[i]$, then $1/z$ is in $\mathbf{Q}[i]$, so define q in $\mathbf{Z}[i]$ to be an element of $\mathbf{Z}[i]$ whose coordinates are those integers which are as close as possible to the coordinates of $w(1/z)$. For example, if $w(1/z) = (14)/3 + i\ 8/7$, then $q = 5 + i$.]

Exercise #99) (i) Prove that in $\mathbf{Z}[i]$, the prime elements (the "Gaussian primes"), are precisely those primes p in $\mathbf{Z}$ such that $p \equiv 3(\text{mod}4)$, plus those elements $a+bi$ such that $a^2+b^2 = q$ where q is prime in $\mathbf{Z}$ (whence $q=2$ or $q \equiv 1(\text{mod}4)$), plus associates.
(ii) For any ring map $f{:}R \to S$, prove that if $I \subset S$ is any prime ideal then $f^{-1}(I) \subset R$ is a prime ideal.
(iii) For the map $f{:}\mathbf{Z} \to \mathbf{Z}[i]$, and any prime ideal $J \subset \mathbf{Z}$, prove there is either one or two prime ideals $I \subset \mathbf{Z}[i]$ such that $f^{-1}(I) = J$.

Exercise #100) (i) For which primes $p \leq 19$, are the equations $X^2+2Y^2 = p$, and $X^2+2 \equiv 0$ (modp) either both solvable or both not solvable?
(ii) For which primes $p \leq 19$, are the equations $X^2+3Y^2 = p$, and $X^2+3 \equiv 0$ (modp) either both solvable or both not solvable?
(iii) For which primes $p \leq 19$, are the equations $X^2+5Y^2 = p$, and $X^2+5 \equiv 0$ (modp), either both solvable or both not solvable?
(iv) Prove that 2 is irreducible but not prime in $\mathbf{Z}[\sqrt{-5}]$. [Thus $\mathbf{Z}[\sqrt{-5}]$ is another example of a noetherian domain but not a ufd.]
(v) For which primes $p \leq 19$, are the equations $X^2+5Y^2 = 2p$, and $X^2+5 \equiv 0$ (modp), either both solvable or both not solvable?

(vi) Which of the rings $\mathbf{Z}[\sqrt{-2}]$, $\mathbf{Z}[\sqrt{-3}]$, if any, is a ufd?

It turns out that $\mathbf{Z}[\sqrt{-3}]$ is "very close" to a ufd, but that $\mathbf{Z}[\sqrt{-5}]$ is not quite so close. In a sense, the experimentation done in exercise 89 above suggests that $\mathbf{Z}[\sqrt{-3}]$ is a ufd "except for something to do with the prime 2", i.e. the solvability of the two equations $X^2+3Y^2 = p$, and $X^2+3 \equiv 0$ (modp) is apparently the same for all primes except 2. The equations $X^2+5Y^2 = p$, and $X^2+5 \equiv 0$ (modp) on the other hand are only both solvable or both not, for about half the primes. Hence in some sense $\mathbf{Z}[\sqrt{-5}]$ is only "halfway to being a ufd". It turns out that there is a reason to look at the ring "$O(\sqrt{-3})$" $=\{a+b(1+\sqrt{-3})/2$, a,b in $\mathbf{Z}\}$ instead of $\mathbf{Z}[\sqrt{-3}]$, and that $O(\sqrt{-3})$ is a ufd. This ring $O(\sqrt{-3})$ is a ufd which is very close to $\mathbf{Z}[\sqrt{-3}]$, only slightly larger, and in that sense $\mathbf{Z}[\sqrt{-3}]$ is "close to being a ufd". I.e. The subring of $O(\sqrt{-3})$ consisting of those elements $a+b(1+\sqrt{-3})/2$ with b *even* , is exactly $\mathbf{Z}[\sqrt{-3}]$. Now it turns out that one can deduce that if $X^2+3 \equiv 0$ (modp) has a solution in $\mathbf{Z}_p$, then p has a factorization in $O(\sqrt{-3})$, and then $X^2+3Y^2 = 4p$ has a solution, because you have to allow the possibility of a denominator of 2, which gets squared and then has to be multiplied out. However, even though $O(\sqrt{-3})$ is a "unique factorization domain", our problem is solved by the *failure* of strict uniqueness of factorization. I.e. recall that factorization, even in a ufd, is really not unique, but only unique up to order and multiplication by units. Now that means that every factorization yields more factorizations by multiplying one factor by a unit u, and the other factor by u^{-1}. Now in $O(\sqrt{-3})$, there are exactly six units: ± 1, and $\pm (1\pm\sqrt{-3})/2$. These are just the three cube roots of 1, and the three cube roots of -1. It turns out that if you find a factorization of p over $O(\sqrt{-3})$, then by multiplying by a judicious choice of these units you can find a factorization into elements $a+b(1+\sqrt{-3})/2$ in which the coefficients b are even. So there is also a factorization of p in $\mathbf{Z}[\sqrt{-3}]$, and hence a solution in $\mathbf{Z}$ of $X^2+3Y^2 = p$.

There is no such natural modification of $\mathbf{Z}[\sqrt{-5}]$ to look at, on the other hand. The distinction is that $(1+\sqrt{-3})/2$ satisfies a monic polynomial over $\mathbf{Z}$ but $(1+\sqrt{-5})/2$ does not, so one is stuck working with $\mathbf{Z}[\sqrt{-5}]$. To analyze it, Kummer figured out how to measure exactly how far $O(\sqrt{-3})$ is from being a ufd, by creating an equivalence relation on ideals in a ring, such that all the principal ideals form one equivalence class. The equivalence classes even turn out to form a group, under multiplication of ideals, called the "class group". Now for rings of integers in a number field, such as the ones we have been looking at above, it turns out that being a ufd is equivalent to being a pid Hence the size of the class group, which measures how far the ring is from being a pid, also measures exactly how far the ring is from being a ufd. Moreover in the case of $\mathbb{Z}[\sqrt{-5}]$ the class group is $\mathbb{Z}_2$, so this ring is sort of halfway to being a ufd, i.e. there are exactly two equivalence classes of ideals, the principal ones, and one other class. Finally, prime factorization is possible in $\mathbb{Z}[\sqrt{-5}]$ on the level of *ideals*, i.e. every ideal is a product of (not necessarily principal) prime ideals, and these properties of $\mathbb{Z}[\sqrt{-5}]$ turn out to be enough to analyze fully the equation $X^2+5Y^2 = p$. The results say that $X^2+2Y^2 = p$ has a solution in $\mathbb{Z}$ iff $p \equiv 1$ or 3, (mod 8), that $X^2+3Y^2 = p$ has a solution in $\mathbb{Z}$ iff $p \equiv 1$ (mod 6), and that $X^2+5Y^2 = p$ has a solution in $\mathbb{Z}$ iff $p \equiv 1$ (mod 4) and $p \equiv \pm 1$ (mod 5) [See M. Artin, Algebra, and I. Niven, Introd. to Number Theory.]

§7) Back to Galois theory: normal and separable extensions

We have proved that a solvable polynomial has a solvable Galois group, but we have not studied the converse question. If a polynomial has a solvable group, is the polynomial actually solvable? This is true, at least in characteristic zero, due also to Galois: If a polynomial over Q has a solvable Galois group, then the roots of the polynomial can be expressed in terms of radicals. So although we know most polynomials of degree ≥ 5 are not solvable by radicals, and all those of degree ≤4 are solvable, results known before Galois, this precise result tells us exactly which polynomials in every degree are solvable, and thus goes far beyond anything known before.
This converse direction of the result is deeper than the other, since it begins with information on a relatively simple object, the group, and gives back information about a more complex object, the field. I.e. recall that many different fields have the same group, so it is not so clear that knowledge of the group should yield any precise information on the field it came from.
In order to prove the converse of Galois' theorem, we will first prove the famous Fundamental Theorem of Galois Theory. This theorem

tells us that you can go back from the group to the field in at least a relative way. I.e. if you know both the field and its group, then from the subgroups of the group you can sometimes recover all the intermediate subfields of the field. Of course the group contains little information about non normal or non separable fields, because then the field does not contain enough distinct roots, and the group is after all isomorphic to a permutation group of some roots. Hence we must assume the field is normal and separable (i.e. "Galois"). For Galois extensions the fundamental theorem says there is a perfect correspondence between subgroups and subfields. A useful technical result, the "theorem of the primitive element", will make the proof easier. We will also use the corollary of Zorn's Lemma that every field can be embedded in an algebraically closed field, just to make some arguments more convenient. This is not essential but saves us the trouble of stopping repeatedly to enlarge our field by adding roots of any polynomials we might be working with. Our earlier proof of Galois' theorem that solvable polynomials have solvable groups was only given for the field $\mathbf{Q}$, but it holds for general fields. The converse direction we are aiming for now needs some restriction, eg. to fields characteristic zero, but again we will only prove it over $\mathbf{Q}$.

Terminology: A finite (dimensional) field extension $k \subset L$ is called "Galois" iff it is both normal and separable. The algebraic closure of a field k is an extension $\overline{k}$ which is both algebraic over k and algebraically closed. It exists and is unique up to k isomorphism.

Theorem (Fundamental Theorem of Galois Theory, "FTGT"):
(1) If $k \subset L$ is a Galois extension, with Galois group G, then the correspondence assigning to a subgroup $H \subset G$ the subfield $\varphi(H) = F \subset L$ of elements which are left fixed by all elements of H, defines a 1-1 correspondence between the set of all subgroups H of G and the set of all subfields F of L which contain k.
(2) Under this correspondence, the extension $F \subset L$ is normal for every subgroup H, and has Galois group $H = G_F(L)$.
(3) With the same notation, H is a normal subgroup of G iff $k \subset F$ is a normal extension, and then $G/H \cong G_k(F)$.

Let us review the concepts of normal and separable. Our original definition of normal was this:

Definition: A finite (hence algebraic) field extension $k \subset L$ is called "normal" provided for every inclusion $L \subset F$ in a larger field, every k-homomorphism $\varphi: L \to F$ maps L isomorphically onto itself.

There is another version of this definition of normality which may be easier to remember, in terms of the notion of "conjugacy".
Definition: In any field extension $k \subset F$, let two intermediate fields $k \subset L \subset F$, $k \subset K \subset F$ be given We say K, L are "conjugate" (in F) iff they are k-isomorphic. [To be conjugate, it is not enough for two extensions of k to be k-isomorphic, they must also be contained in a common extension of k.]

Theorem: A finite field extension $k \subset L$ is normal iff any of the following equivalent properties hold:
(i) In any further extension $k \subset L \subset F$, L is conjugate only to itself,
(ii) In the extension $L \subset \bar{k}$, L is conjugate only to itself,
(iii) In any further extension $k \subset L \subset F$, every k-automorphism of F restricts to a k-automorphism of L,
(iv) L is the splitting field over k of a polynomial in $k[X]$,
(v) For every element α of L, its minimal k-polynomial splits into linear factors in L.
(vi) There is a further normal (finite) extension, i.e. $k \subset L \subset F$, with $k \subset F$ normal, and with L conjugate only to itself in F.
proof: "normal" $\Rightarrow$ (i). If $k \subset L$ is normal and $k \subset L \subset F$ is a further extension, L cannot be conjugate to another subfield $K \subset F$ since that would imply the existence of a k isomorphism from L to K. Since $K \neq L$ that contradicts the definition of normal.
(i) $\Rightarrow$ (ii): This is trivial.
(ii) $\Rightarrow$ (iii): A k-automorphism of F restricts to a k-homomorphism $f: L \to F$. Let L_1 be the image of L in F. Since L is finitely generated over k, say by $\alpha_1, \dots, \alpha_n$, then L_1 is also finitely generated over k, by $f(\alpha_1), \dots, f(\alpha_n)$. Let K be the subfield of F generated over k by the union of these generators, i.e. $K = k(\alpha_1, \dots, \alpha_n, f(\alpha_1), \dots, f(\alpha_n))$. Since L is finite dimensional over k, all $\alpha_1, \dots, \alpha_n$ are algebraic, hence so are all $f(\alpha_1), \dots, f(\alpha_n)$. Thus K is algebraic over k and thus by our extension theorems, there is a k-embedding of K into $\bar{k}$. Under this embedding the images of L, L_1 are conjugate. We may replace the images of L, L_1 in $\bar{k}$ by L, L_1 thenselves. Therefore by (ii), these images in $\bar{k}$ are equal. Thus L, L_1 were already equal in F, hence

the restriction of the automorphism of F was indeed an automorphism of L.

(iii) → (v). First embed L into $\bar{k}$ by our extension theorem Suppose α is an element of L whose minimal polynomial φ does not split into linear factors in L. It does split in $\bar{k}$, and let β be a root of φ in $\bar{k}$ which does not lie in L Then our usual extension argument shows we can embed $k(\alpha)$ isomorphically onto $k(\beta)$ in $\bar{k}$, sending α to β. We can extend this to a k-embedding of L onto a subfield L_1 of $\bar{k}$. If we enlarge the set $\{\alpha=\alpha_1\}$ to a set of generators for L over k, say $\{\alpha_1,...,\alpha_n\}$, let $\{\beta_1,...,\beta_m\}$ be the larger set containing *all* roots of the minimal polynomials of the α's. If $F = k(\beta_1,...,\beta_m)$ then $k\subset L\subset F$, and we can extend the map $L\to\bar{k}$ to an embedding $F\to\bar{k}$. By our usual theory of extending embeddings we know every β_i goes to some β_j, so that the map $F\to\bar{k}$ is an automorphism of F. However since this map restricts to an isomorphism $L\to L_1$, and since β is in L_1 but not in L, this violates property (iii).

(v) ⇒ (iv): Let $L = k(\alpha_1,...,\alpha_n)$ and let $\{\beta_1,...,\beta_m\}$ again be the larger set containing *all* roots of the minimal polynomials of the α's. Then by (v) all the β's are in L. Hence $L = k(\beta_1,...,\beta_m)$ also. Thus if f is the product of the minimal polynomials of the α's, the roots of f are precisely the set $\{\beta_1,...,\beta_m\}$ of β's. Thus L is the splitting field of f.

(iv) ⇒ "normal": Let $k\subset L\subset F$ be an inclusion in a larger field, and $\varphi:L\to F$ be a k- homomorphism. If $L= k(\beta_1,...,\beta_m)$ where $\{\beta_1,...,\beta_m\}$ is the set of roots of the k-polynomial f, our usual extension theory shows each β_i must map to some β_j Thus φ maps L isomorphically onto itself.

(ii) → (vi): If we enlarge L to a finite normal extension $k\subset L\subset F$, and then embed $F\subset\bar{k}$, since L is conjugate only to itself in $\bar{k}$, F is also conjugate only to itself in L.

(vi) → (ii): Given the finite normal extension $k\subset F$, with $k\subset L\subset F$, embed $F\subset\bar{k}$, and consider any k-homomorphism $\varphi:L\to\bar{k}$. Extend φ to a k-homomorphism $\varphi:F\to\bar{k}$. Since $k\subset F$ is normal, we must have $\varphi(F)=F$. Thus $\varphi(L)\subset\varphi(F) = F$, and hence any conjugate $\varphi(L)$ of L in $\bar{k}$ is actually in F. Since L is conjugate only to itself in F, L is conjugate only to itself in $\bar{k}$. **QED**.

Now we recall the concept of "separable" field extension. (The terminology is due to Van der Waerden.)

Definition: A "separable" k-polynomial is one whose roots are all distinct, in any splitting field.

Remark: Since all splitting fields are k-isomorphic it does not matter which one we use, and since $\bar{k}$ contains a splitting field, we can simply say f is separable iff its roots are distinct in $\bar{k}$.

Definition: An element of an extension of k is separable over k iff its minimal k-polynomial is separable. A field extension $k\subset L$ is separable iff every element of L is separable over k.

Definition: If $L=k(\alpha_1,\ldots,\alpha_n)$ is a finite extension of k, the "separable degree" of L over k, $[L:k]_s$, is the number of k-homomorphisms $\varphi:L\to\bar{k}$, from L into an algebraic closure of k.

Although the terminology is new, the next result is the same as one proved last quarter, but it is worth reviewing the ideas
Lemma: Assume $L = k(\alpha_1,\ldots,\alpha_n)$ is a finite extension of k. Then the *degree* of L over k equals the *separable degree* iff all the generators $\{\alpha_1,\ldots,\alpha_n\}$ are separable over k. If any α_i is not separable over k, the separable degree is less than the degree.
proof: Let us compute the separable degree, by computing the number of k-homomorphisms $\varphi:L\to\bar{k}$. We can obtain every such homomorphism φ in stages, by starting with the identity map $k\to k\subset\bar{k}$, extending it to a map $k(\alpha_1)\to\bar{k}$, then extending it further to a map $k(\alpha_1,\alpha_2)\to\bar{k}$, etc until we get a map $k(\alpha_1,\ldots,\alpha_n) = L\to\bar{k}$. Moreover, if there are n_1 ways to extend the identity map $k\to k\subset\bar{k}$. to a map of $k(\alpha_1)\to\bar{k}$., and n_2 ways to extend further to a map $k(\alpha_1,\alpha_2)\to\bar{k}$, then for each choice of an image of α_1 there are n_2 choices for the image of α_2. Hence there are altogether n_1n_2 ways to define φ on α_1,α_2, hence n_1n_2 extensions of $k\to k$ to $k(\alpha_1,\alpha_2)\to\bar{k}$. Reasoning in this way, the number of k-homomorphisms of $L\to\bar{k}$ equals the product of the number of possible extensions at each stage. From our usual extension theory, we know the number of possible extensions to $k(\alpha_1)$ is just the number of distinct roots in $\bar{k}$ of the minimal polynomial g of α_1 over k. This number equals the

degree of g over k if g is separable, and otherwise is less than that degree. Moreover the degree of g over k equals the degree of the field extension $[k(\alpha_1):k]$ Hence the number of extensions to $k(\alpha_1)$ is at most the degree $[k(\alpha_1):k]$, and equals that degree iff α_1 is separable over k. Similarly, the number of further extensions to $k(\alpha_1,\alpha_2)$ is at most the degree of the field extension $[k(\alpha_1,\alpha_2):k(\alpha_1)]$, and equals that degree iff α_2 is separable over $k(\alpha_1)$. In particular, since the minimal polynomial of α_2 over $k(\alpha_1)$ is a factor of the minimal polynomial over k, if α_2 is separable over k, it is also separable over $k(\alpha_1)$. Consequently the number of extensions from $k = k \subset L$ to $k(\alpha_1,\alpha_2) \to L$ is at most the product of these degrees, $[k(\alpha_1):k]\cdot[k(\alpha_1,\alpha_2):k(\alpha_1)] = [k(\alpha_1,\alpha_2):k]$, (by multiplicativity of degrees of field extensions), and equals that product provided both α_1,α_2 are separable over k. If α_1 is not separable over k, the number of such homomorphisms is less than the degree $[k(\alpha_1,\alpha_2):k]$. Continuing, we get that the number of k-homomorphisms from $L \to \overline{k}$, equals the degree $[L:k]$ if every α_i is separable over k, but is less if α_1 is not separable overk. Since we can reorder the generators, any α_i can be taken as α_1. Thus the separable degree equals the degree iff all the α_i are separable over k. **QED**.

Cor: A finite field extension $k \subset L = k(\alpha_1, .., \alpha_n)$ is separable over k iff every generator α_i is separable over k.
proof: Since all α_i belong to L, if L is separable then all α_i are separable. Coversely, if all α_i are separable over k, then the separable degree of L over k equals the degree. If some element β of L is not separable over k, then we could use β as a generator, i.e. $L = k(\beta,\alpha_1,....\alpha_n)$, and the result just proved then would show that the seprable degree of L is less than the degree. This contradiction shows that there is no such element β. **QED**.

The following corollary contains some very useful inequalities.
Cor: If $k \subset L$ is any finite extension, then $\#G_k(L) \le [L:k]_s \le [L:k]$. Equality holds in the first inequality iff L is normal over k, and equality holds in the second one iff L is separable over k. Thus equality holds in both, i.e. $\#G_k(L) = [L:k]$, iff $k \subset L$ is "Galois".
proof: The lemma and the previous corollary show that $[L:k]_s = [L:k]$ iff $k \subset L$ is separable. On the other hand, since $\#(G_k(L))$ is the

number of k-automorphisms $L \to \bar{k}$, and $[L:k]_s$ is the number of k-homomorphisms $L \to \bar{k}$, it follows that $\# G_k(L) = [L:k]_s$ iff every such homomorphism is an automorphism of L, i.e. iff $k \subset L$ is normal.
QED.

Remarks: Recall that if k has characteristic zero, then every algebraic extension of k is separable. This used the derivative criterion for multiple roots. I.e. if f is the minimal polynomial of an element over k, then f is irreducible, hence cannot divide a non zero polynomial of lower degree. If f is non constant, and irreducible over k of characteristic zero, then the derivative of f is non zero, so f is separable. More generally, a field is called "perfect" if every algebraic extension is separable. We will see later that not only are fields of characteristic zero perfect, but all finite fields are perfect too. The simplest non perfect field is $k = \mathbf{Z}_2(X)$, the field of fractions of the polynomial ring $\mathbf{Z}_2[X]$. An example of a non separable extension is the splitting field L of $T^2 - X$. The polynomial $T^2 - X$ is irreducible over k since it is quadratic and has no root in k. If we write α for a root of f in L, then $T^2 - \alpha^2 = T^2 - 2\alpha + \alpha^2 = (T - \alpha)^2$ since charac(k) = charac(L) = 2. Hence both roots in L of the irreducible polynomial f are equal to α.

Exercise #101) Let k be any field of prime characteristic p.
(i) Prove that if a is in k, and if $r = p^t$ with $t \geq 1$, then $X^r - a = (X - \alpha)^r$, for some α in an extension of k.
(ii) If f is an irreducible polynomial over k, and if f has a multiple root in some splitting field, show that the only terms in f which can have non zero coefficients are those of form X^s where p divides s. Deduce that $f(X) = g(X^r)$ where $r = p^t$, for some $t \geq 1$, where $g(X)$ is irreducible and has no repeated roots.
(iii) If f is irreducible over k, prove that all roots of f have the same multiplicity.
(iv) If $K = k(X)$, the field of fractions of the polynomial ring $k[X]$, prove $f(T) = T^p - X$ is irreducible over K, and that f is not separable. Conclude that $L = K[T]/(f)$ is a non separable extension of K.

Lemma: A finite field extension $k \subset L$ is Galois iff L is the splitting field of a separable polynomial f over k.
proof: If $k \subset L$ is Galois, then it is separable, so we can write L =

$k(\alpha_1,\ldots,\alpha_n)$, where every α_j is separable over k. Thus each minimal polynomial f_j of α_j over k is separable. Now suppose two polynomials f_i, f_j have a root say α in common. Then the minimal polynomial for α divides both f_i and f_j. Since all these polynomials are irreducible, and monic, they are all equal. Thus $f_i = f_j$ and hence we may eliminate one of them. Continuing in this way we have a collection of distinct irreducible polynomials $\{f_j\}$, the minimal polynomials of some of the generators α_j, where two distinct f_j do not have common roots, and we let $f=\prod f_j$ be the product of these polynomials Then f is separable and L is the splitting field of f. Coversely, if L is the splitting field of f, a separable k polynomial, then L is generated by the roots of f. Since each root α satisfies f, a minimal polynomial, the minimal polynomial of α is a factor of f, and hence is also separable. Thus every root of f is separable over k, and L is generated by a collection of elements separable over k. By arguments given above, the separable degree of L over k thus equals the degree, so the extension is separable. Since L is a splitting field over k it is also normal, hence Galois. **QED.**

§8) The "Fundamental theorem of Galois Theory"

Now we are ready to begin the arguments leading to the FTGT. We are trying to prove a 1-1 correspondence between intermediate fields of a Galois extension and subgroups of the Galois group. The next corollary proves the easier direction, that every intermediate field is the fixed field of some subgroup of the Galois group.

Lemma: If $k \subset L$ is a finite Galois extension, and α is an element of L such that $\varphi(\alpha) = \alpha$ for all φ in $G_k(L)$, then α is in k. I.e. the subfield of L left fixed by $G_k(L)$ is exactly k.

proof: (We prove the contrapositive statement, that if α is not in k then some k-automorphism of L does not fix α.) If α is not in k then the minimal polynomial of α has degree at least 2, and the extension is separable, so there are other roots different from α. By our extension principle for homomorphisms, we can define a homomorphism $\varphi: L \to \bar{k}$ that sends α to some other root of its minimal polynomial, hence $\varphi(\alpha) \neq \alpha$ Since $k \subset L$ is normal, $\varphi(L) = L$, and hence φ is an element of $G_k(L)$. **QED.**

Cor: Given $k\subset L$ a Galois extension, and an intermediate field $k\subset F\subset L$, the subfield of elements of L left fixed by the subgroup $G_F(L)\subset G_k(L)$ is exactly F.
proof: Note here $G_F(L)$ is the group of k-homomorphisms of L fixing F. Since these also fix $k\subset F$, this is a subgroup of $G_k(L)$. Moreover, if L is the splitting field of a separable k-polynomial f, then f is also an F polynomial, and L is still its splitting field. Thus $F\subset L$ is Galois, with Galois group $G_F(L)$, and this is the result just proved above. QED.

Now consider the following **fundamental correspondence** : φ:{subgroups H of $G_k(L)$}$\rightarrow${subfields F with $k\subset F\subset L$}, where $\varphi(H)\subset L$ is the subset of elements of L left fixed by every element of H, i.e. α is in $\varphi(H)$ iff $\sigma(\alpha) = \alpha$ for every σ in H. Since $H\subset G_k(L)$, every element of H fixes k, so $k\subset\varphi(H)\subset L$.

Terminology: The subfield $\varphi(H)\subset L$ is called the "fixed field" of H.

Cor: For any Galois extension $k\subset L$, the fundamental correspondence φ {subgroups H of $G_k(L)$}$\rightarrow${subfields F with $k\subset F\subset L$} is surjective. Thus #{subfields F with $k\subset F\subset L$} $\leq$ #{subgroups of $G_k(L)$}, and in particular the number of subfields F between k and L is finite.
proof: Let F be any field with $k\subset F\subset L$. We have just proved that F is the fixed field of the subgroup $G_F(L)\subset G_k(L)$. I.e. $\varphi(G_F(L)) = F$. Hence φ is surjective. Since there are only a finite number of subgroups of a finite group (G is finite since $\#(G) = [L:k]$), the domain of the function φ is finite, whence the image is also finite QED.

Now we can prove a very useful technical result, the so called "theorem of the primitive element".
Theorem: Every finite separable extension $k\subset L$ is "simple", i.e. there exists an element α in L such that $L = k(\alpha)$.
Proof when k is infinite:
Step (i) If $k\subset L$ is finite and separable, k is infinite, then there are only finitely many fields between k and L.
proof: If we enlarge $k\subset L$ to a normal field extension $k\subset K$, where $k\subset L\subset K$, by adjoining to L all the roots of the minimal polynomials of some finite set of k-generators for L, then $k\subset K$ is a finite Galois extension, (because the new generators have the same minimal polynomials as did the generators of $k\subset L$). Hence there are only

finitely many subfields between k and K by the Cor. above. Since $k \subset L \subset K$, there are only a finite number of subfields between k and L. **QED (i).**

Step (ii) If $k \subset L$ is finite and separable, k is infinite, and α, β are in L, then there exist $\lambda \neq \mu$ in k such that $k(\alpha+\lambda\beta) = k(\alpha+\mu\beta)$. s
proof: Since there are infinitely many elements of k to choose from and only a finite number of subfields, some two elements λ, μ give the same subfield. **QED (ii).**

Step (iii) If $k \subset L$ is finite and separable, k is infinite and α, β are in L, then there is some λ in k such that $k(\alpha+\lambda\beta) = k(\alpha, \beta)$.
proof: If we choose λ, μ in k as above, since $F = k(\alpha+\lambda\beta) = k(\alpha+\mu\beta) \subset k(\alpha,\beta)$, we have $(\alpha+\lambda\beta) - (\alpha+\mu\beta) = (\lambda-\mu)\beta$ in F, and since $(\lambda-\mu) \neq 0$ is in $k \subset F$, dividing by it proves β is in F. Then $\alpha = (\alpha+\lambda\beta) - \lambda\beta$ is in F too. Hence $k(\alpha,\beta) \subset k(\alpha+\lambda\beta) = F$, so they are equal. **QED (iii).**

Step (iv) If $k \subset L$ is finite and separable, and k is an infinite field, then $L = k(\gamma)$ for some γ in L.
proof: Use induction on the proof above: i.e
$k \subset k(\alpha_1, \ldots \alpha_{n-2}, \alpha_{n-1}, \alpha_n) = k(\alpha_1, \ldots \alpha_{n-2})(\alpha_{n-1}, \alpha_n) =$
$k(\alpha_1, \ldots \alpha_{n-2})(\gamma_{n-1}) = k(\alpha_1, \ldots \alpha_{n-3})(\alpha_{n-2}, \gamma_{n-1}) =$
$k(\alpha_1, \ldots \alpha_{n-2})(\gamma_{n-2}) = \ldots = k(\alpha_1, \gamma_2) = k(\gamma_1) = k(\gamma)$.
QED (iv) (when k is infinite).

Proof when k is finite: By the corollary of the next Lemma, in a finite field L, the multiplicative group L^* of units is a cyclic group. If k is finite and $k \subset L$ is a finite extension, then L is a finite field, so the group of units of L is generated as a multiplicative group by one element γ. Since every $\neq 0$ element of L is a power of γ, $L = k(\gamma) =$ the smallest subfield of L containing γ. Thus every finite extensioin of a finite field is a simple extension, even without the hypothesis of separability. [We shall see below however that separability is automatically true in this case as well.] **QED for Thm.**

Digression on finite fields:
Dealing with a finite field k is easy, once we observe that k^* is a finite abelian group, and hence every element satisfies an equation of form $X^n = 1$, where $n = \#(k^*)$, by Cauchy's theorem. Thus essentially a finite field just consists of zero plus some roots of unity.

Combining this with the fact that in a field, an equation like this cannot have more than n solutions, will almost prove k* is cylic. The key concept is that of "annihilator" or "exponent" of a group.

Definition: If G is any finite (multiplicative) group, the "annihilator" of G = ann(G), or exp(G) = "exponent" of G, is the least positive integer m such that $x^m = 1$, for every element x of G
[For a finite additive group G, ann(G) is the least positive integer m such that mx = x+....+x (m terms) = 0 for all elements x in G.]

We will use some simple observations about orders as follows:
Exercise #102) If G is a finite group, prove:
(i) then ann(G) = l.c.m. {ord(x), all x in G}, and ann(G) divides #(G).
(ii) If x is in G and ord(x) = ab, then ord(x^a) = b. In particular, every factor of the order of an element is also the order of some element.
If G is finite and abelian, prove:
(iii) If x,y are elements of g with relatively prime orders a,b, the product xy has order ab.
(iv) If $x_1,\ldots,x_n$ are elements of g with pairwise relatively prime orders a_i, the product $x = \prod x_i$ has order $a = \prod a_i$.
(v) Find an example of two elements x,y in a finite abelian group such that ord(xy) ≠ lcm(ord(x),ord(y)).

Exercise #103) Let $\{n_\alpha\}$ be a collection of non zero integers, and $\{p_i\}$ the set of primes dividing at least one of the n_α. Then m = $\prod p_i^{r_i}$ = l.c.m.$\{n_\alpha\}$ iff, for each i, there is at least one n_α divisible by $p_i^{r_i}$ but no n_β divisible by a higher power of p_i.

Lemma: If G is a finite <u>abelian</u> group, and m = ann(G), then there is an element x of G with ord(x) = m.
proof: We know from Ex. 91(i), above that m = l.c.m. of the orders of all elements of G. From Exs. 91(i), 92, we also know if $m = \prod p_i^{r_i}$, then for each i, $p_i^{r_i}$ divides the order of some element y_i of G. Then by Ex. 91(ii), for each i there is an element x_i of order $p_i^{r_i}$. Then by Ex.91(iv), the product $x = \prod x_i$ has order $m = \prod p_i^{r_i}$. **QED.**

Remark: The previous Lemma is false for non abelian groups, since for instance the icosahedral group $I \cong A_5$ has elements of orders 2, 3, 5, but none of orders 6, 10, 15, or 30.

Corollary: If k is a finite field, the multiplicative group k^* is cyclic.
proof: By the lemma we know there is an element x of order m, where m = ann(k^*), and that every element y of k^* satisfies the equation $Y^m-1=0$. Since k is a field this equation has at most m roots, hence $\#(k^*) \leq m$. On the other hand $(x) \subset (k^*)$, where (x) is the cyclic subgroup generated by the element x. Since $\#(x) = m$, $k^* = (x)$ is cyclic. QED.

Remark: As pointed out above, the previous Corollary completes the proof of the theorem of the primitive element. Now we continue with the proof of the FTGT.

We already know that for a Galois extension $k\subset L$, when $G_k(L)$ acts on L, the roots of an irreducible k polynomial form an orbit of the action. Indeed this is just another way of stating our usual extension theory for homomorphisms. The following lemma is a kind of converse, i.e. any orbit of a G action forms the set of roots of an irreducible polynomial over the fixed field. The method used to construct the polynomial is apparently very old, but the point of view in the Proposition, beginning with a group of automorphisms, is usually attributed to E. Artin.

Proposition (E. Artin): Let $G\subset \mathrm{Aut}(L)$ be any finite group acting on a field L, and let $F\subset L$ be the fixed field of G. Then $F\subset L$ is a (finite) Galois extension, with Galois group $G_F(L) = G$.

Lemma: With the same assumptions as above, if β is any element in L, and $\{\sigma_1,\ldots,\sigma_n\}\subset G$ is a maximal collection of elements of G taking different values on β, then $f(X) = \prod_j (X-\sigma_j(\beta))$ is a separable, irreducible polynomial with coefficients in F, degree(f) $\leq \#(G)$, and $f(\beta) = 0$.
proof of lemma: We know there is a subgroup H of G leaving β fixed, and for each σ in G, the left coset σH consists of those elements of G which send β to $\sigma(\beta)$. We choose one σ_j from each of those cosets. Since σ_1 in H fixes β, we know $(X-\beta)$ is a factor of this

polynomial, hence $f(\beta)=0$. The polynomial is separable since by definition of the σ_j, all the roots $\sigma_j(\beta)$ are distinct. Note that to apply any σ in G to the coefficients of this polynomial, we can simply apply σ to the coefficients of the factors of this product, then multiply out. Since any such σ only permutes the factors of the product, it leaves the polynomial unchanged, and hence every σ in G leaves all the coefficients unchanged By definition of F = the fixed field of G, the coefficients of f are in F. To see the polynomial f is irreducible, we recall that the roots of the minimal F-polynomial of β form a G orbit, by our earlier extension theory. Since the roots of f are precisely the G orbit of β, this polynomial f is the minimal polynomial of β, hence irreducible. Since the number of factors $(X-\sigma_j(\beta))$ in f is at most the number of elements of G, $\deg(f) \leq \#(G)$.
QED.

Proof of Prop: It follows from the definition of F that $G = G_F(L)$, and it follows from the lemma that every element β of L is algebraic and separable over F, with minimal polynomial of degree $\leq \#(G)$.
Step (i) We claim the extension $F\subset L$ is finite.
If not, and $F\subset L$ is an infinite extension, then there would be intermediate extensions $F\subset K\subset L$ with the degree of $F\subset K$ finite but arbitrarily large I.e. we could just keep adjoining elements, $F\subset F(\alpha_1)\subset F(\alpha_1,\alpha_2)\subset \ldots\subset F(\alpha_1,\ldots\ldots,\alpha_n)\subset\ldots\ldots$, with α_j not in $F(\alpha_1,\ldots\ldots,\alpha_{j-1})$, for every j. Then since each α_j has degree ≥ 2 over the previous field, and degree of extensions is multiplicative, the degree of $F(\alpha_1,\ldots\ldots,\alpha_n)$ over F is $\geq 2^n$. But by the theorem of the primitive element, every subextension of L which is finite over k is actually simple, and the previous lemma implies that every simple subextension has degree at most $\#(G)$, so every finite subextension has degree $\leq \#(G)$. This contradiction proves that $F\subset L$ is a finite extension, and that $[L:F] \leq \#(G)$.
Step (ii): We claim the extension $F\subset L$ is Galois.
For this, we have only to apply the useful inequalities proved above, involving the separable degree. I.e. recall that for a finite extension $\#(G) \leq [L:F]$, always holds, with equality iff the extension is Galois. Since we have just proved in step (i) that $[L:F] \leq \#(G)$. we are done.
QED.

Now we can complete the proof that the fundamental Galois correspondence is one to one:
If $k \subset L$ is a Galois extension, consider the **inverse correspondence** γ:{subfields F with $k \subset F \subset L$} → {subgroups of $G_k(L)$} defined by:
$G_k(L) \supset \gamma(F)$ = the subgroup of elements of G fixing every element of F. This is just the Galois group of L over F, $\gamma(F) = G_F(L)$, but we may also call it the "invariant subgroup for F".
As just remarked, we know for $k \subset F \subset L$, that $\varphi(\gamma(F)) = \varphi(G_F(L)) = F$. The next lemma will complete the proof that γ is inverse to φ, by showing that $\gamma(\varphi(H)) = H$.

Cor: For any (finite) Galois extension $k \subset L$, let $H \subset G$ be any subgroup of the Galois group $G = G_k(L)$, and $\varphi(H) = F \subset L$ be the fixed field. Then $\gamma(\varphi(H)) = G_F(L) = H$.
proof: This follows from the previous proposition of E. Artin. **QED**.

To complete the proof of the fundamental theorem of Galois theory, we need to check part (3) about normal subgroups H of G corresponding to normal subextensions of k. Let $k \subset L$ be a Galois extension and let F an intermediate field $k \subset F \subset L$ such that $k \subset F$ is normal. Then the restriction of every k-automorphism of L to F, is a k-automorphism of F, and by our extension theory every k-automorphism of F extends to a k-automorphism of L. Hence the restriction map $G_k(L) \rightarrow G_k(F)$ is surjective, with kernel = the subgroup consisting of those k-automorphisms of L which leave F fixed, i.e. the kernel is precisely $\gamma(F) = G_F(L)$. Since the kernel of a group homomorphism is normal, we see that indeed $\gamma(F) \subset G_k(L)$ is a normal subgroup whenever $k \subset F \subset L$ and $k \subset F$ is a normal field extension. Moreover, by the fundamental homomorphism theorem for groups, we have $G_k(F) \cong G_k(L)/\gamma(F)$.

Conversely, suppose $H \subset G_k(F)$ is a normal subgroup and let $F = \varphi(H) \subset L$ be the fixed field. Then by Artin's proposition, $H = G_F(L)$, whence $\#(H) = [L:F]$. We claim $k \subset F$ is a normal extension. If $\sigma: F \rightarrow \bar{k}$ is any k-homomorphism, we know it extends to a k-homomorphism $\sigma: L \rightarrow \bar{k}$, which maps L into L since $k \subset L$ is normal. I.e. σ is an element of $G_k(L)$. We must show $\sigma(F) = F$. If $\sigma(F) \subset L$ is the image of F under σ, then the subgroup $\sigma H \sigma^{-1}$ conjugate to H leaves $\sigma(F)$ fixed. Thus $\sigma H \sigma^{-1}$ is a subgroup of the invariant subgroup for $\sigma(F)$

= $\gamma(\sigma(F))$ = $G_{\sigma(F)}(L)$. I.e., since σ is a k-homomorphism, hence a k-vector space map and injective, the k-vector dimensions [F:k] and [σ(F):k] are equal, hence also the degrees [L:F] = [L:σ(F)] are equal, and thus the Galois groups $G_F(L)$ and $G_{\sigma(F)}(L)$ have the same order. Since $\#(G_F(L)) = \#(H) = \#(\sigma H\sigma^{-1})$ and $\sigma H\sigma^{-1} \subset G_{\sigma(F)}(L)$, it follows then that $\sigma H\sigma^{-1} = G_{\sigma(F)}(L)$. Finally since H is normal, we see that H = $\sigma H\sigma^{-1}$, so that $G_F(L) = G_{\sigma(F)}(L)$ are the same subgroup of $G_k(L)$. Since the correspondence between subgroups and subfields is 1-1, we must have σ(F) = F. Thus k⊂F is a normal extension.
QED for FTGT.

Remark: Let k⊂L be a finite normal extension, which is not separable, let G=$G_k(L)$ be the Galois group, and let F⊂L be the fixed field of G. It follows from the theory above that k⊂F⊂L, where F⊂L is Galois, the inclusion k⊂F is proper, and an element of F is separable over k iff it belongs to k.
Moreover, G = $G_F(L)$, and $\#(G) = [L:k]_s = [L:F] < [L:k]$. Hence there is a 1-1 correspondence between subgroups of G and subfields of L containing F, hence the Galois group $G_k(L)$ contains no information at all on the structure of the subextension k⊂F. In particular, the result proved above, that $\varphi(\gamma(K))$=K for any k⊂K⊂L [i.e. that the fixed field of the invariant group of K is K again], fails if the extension k⊂L is not separable [since k itself is then not the fixed field of any subgroup of $G_k(L)$].

#104) (i) Prove that if k⊂L is any finite dimensional field extension, not necessarily either normal or separable, then #(G) divides $[L:k]_s$, and $[L:k]_s$ divides [L:k].
(ii) If k⊂L⊂F is any pair of finite dimensional field extensions, prove $[F:k]_s = [F:L]_s[L:k]_s$.

#105) (i) For the polynomial X^3-2, over Q, compute the splitting field, the Galois group, all subgroups, all subfields, and the correspondence between them asserted in the FTGT.
(ii) Do the same for the polynomial X^4-2. [You may assume from last quarter that this group is ≅ D_4, the dihedral group on 4 elements. There are 8 proper intermediate fields!]

§9) Galois theory for finite fields

The finite field case offers a nice easy family of examples of Galois theory. Indeed all finite extensions of finite fields are Galois, and all the Galois groups are cyclic. The FTGT holds, so an extension of degree n has exactly one intermediate field for each factor of n. For each prime p and each natural number n, there is exactly one field of order p^n, the unique extension of $\mathbf{Z}_p$ of degree n.

A) A finite field F of characteristic p, has exactly p^n elements, where $n = [F: \mathbf{Z}_p]$.

proof: If F is any finite field of characteristic p, then F contains the prime field, i.e. $\mathbf{Z}_p \subset F$ and thus F is a vector space over $\mathbf{Z}_p$, of dimension n, say. Then we claim F has exactly p^n elements. Let $\alpha_1, \ldots, \alpha_n$ be a basis for F over $\mathbf{Z}_p$. Then each element x of F has a unique expression as a linear combination $x = \Sigma c_i \alpha_i$, where the coefficients c_i are in $\mathbf{Z}_p$. That says the map $F \to (\mathbf{Z}_p)^n$ taking x to the sequence of coefficients $(c_1, \ldots, c_n)$ is a bijection. In fact it is a vector space isomorphism, but we don't need this now. Since $\mathbf{Z}_p$ has p elements, the Cartesian product $(\mathbf{Z}_p)^n$ has p^n elements, and thus so does F. QED.

B) If $k \subset F$ is an inclusion of finite fields, of characteristic p, and if $\#(k) = q = p^s$, then $\#(F) = q^t = p^{st}$, for some $s,t > 0$.

proof: The same argument as above shows F is a vector space over k, so $\#(F) = q^t$ for $t = [F:k]$, and $q = p^s$, where $s = [k:\mathbf{Z}_p]$. Hence $\#(F) = q^t = (p^s)^t = p^{st}$ QED.

Cor: A field of order p^n cannot contain a field of order p^s unless s divides n. For example, a field of order p^n cannot contain a field of order p^{n-1} unless $n = 2$.

C) If F is a finite field with $q = p^n$ elements, then F is the splitting field over $\mathbf{Z}_p$ of the separable polynomial $f(X) = X^q - X$. In particular, F is Galois over $\mathbf{Z}_p$.

proof: Note that f(X) is separable since the derivative is $f'(X) = qX^{q-1} - 1 = -1$ [since $q = p^n$], so f' has no roots in common with f. Next, every element y of F^* satisfies $y^{q-1} = 1$, so every element x of

F, including $x=0$, satisfies $X^q-X = 0$ Thus not only is F generated over $\mathbf{Z}_p$ by the roots of f(X), F consists entirely of the roots of f(X). QED.

D) If $k\subset F$ is any inclusion of finite fields, F is Galois over k.
proof: If p = charac(k) = charac(F), then $\mathbf{Z}_p\subset k\subset F$, and $\mathbf{Z}_p\subset F$ is Galois, so as usual $k\subset F$ is also Galois, since F is the splitting field of X^q-X also over k, where again q = #(F). QED.

E) If k is a finite field, and $k\subset L$ any algebraic extension, even infinite dimensional, then $k\subset L$ is separable. I.e. k is "perfect".
proof: If α in L is any element, the extension $k\subset k(\alpha)$ is finite, hence Galois by part D, so all elements of $k(\alpha)$, including α, are separable over k. QED.

F) If p is any prime integer and n any positive integer, there is exactly one field of order p^n, up to isomorphism.
proof: Consider F = the splitting field over $\mathbf{Z}_p$ of the equation X^q-X, where $q = p^n$. Then F is the "smallest" field extension of $\mathbf{Z}_p$ containing the roots of this equation. Since the equation has been seen to be separable, it has exactly $q = p^n$ roots.
Claim: These roots actually form a field, and hence F consists of precisely those q elements.
proof: Since q is a power of p, we get from the binomial theorem that $(a\pm b)^q = a^q \pm b^q$, hence if $a^q = a$, and $b^q = b$, then $(a\pm b)^q = a\pm b$, so the sum or difference of two roots of X^q-X, is also a root. Even easier, $(ab)^q = a^q b^q = ab$, and if $b \neq 0$, $(a/b)^q = a^q/b^q = a/b$, so the product and quotient of two roots are roots. Since also $1^q = 1$, the roots form a field. F is unique, up to $\mathbf{Z}_p$ - isomorphism, by the uniqueness of splitting fields. QED.

Remark: This result implies that any two finite fields with the same number of elements are isomorphic. This is an astonishing fact, i.e. the forgetful functor from finite fields to sets loses almost no information. By contrast, there is a bijection between the underlying sets of $\mathbb{R}$ and $\mathbb{C}$ but no isomorphism, since one is algebraically closed and the other is not.

G) The intermediate fields of the extension $\mathbf{Z}_p \subset L$, where $\#(L) = p^n$, are precisely those F of order p^s where s divides n.

proof: We already know from part B) these are the only possible intermediate fields, and we now claim such intermediate fields exist for all s dividing n. I.e. if s divides n, consider the splitting field F of the polynomial X^r-X, where $r = p^s$. Then F consists of all roots in $\overline{\mathbf{Z}}_p$ of this polynomial, and L consists of all roots of X^q-X, where $q = p^n$.

Claim. $F \subset L$, i.e. any root of X^r-X is also a root of X^q-X.

proof of claim: If $n = st$, then $q = p^n = p^{st} = (p^s)^t = (p^s)\ldots\ldots(p^s) = r^t$. Assume $a^r = a$; then $a^q = a^{(r\ldots r)}$, with t factors of r in the exponent. Thus if $f(x)=x^r$, then $a^q = f^r(a)$ [where the exponent now means composition of functions r times] $= f^{r-1}(f(a)) = f^{r-1}(a^r) = f^{r-1}(a) = f^{r-2}(f(a)) = f^{r-2}(a) = \ldots = f(a) = a^r = a$. **QED.**

H) The Galois group of an extension $k \subset L$ of degree n, of finite fields, is isomorphic to $\mathbf{Z}_n$, and is generated by the "Frobenius" automorphism $\varphi: x \mapsto x^r$, where $r = \#(k)$.

proof: Since r is a power of p, the binomial theorem again tells us that $(x\pm y)^r = x^r \pm y^r$, and $(xy)^r = x^r y^r$, $(x/y)^r = x^r/y^r$ as usual, if $y\neq 0$, so that the Frobenius φ is an automorphism of L. Moreover since k consists of the roots of X^r-X, φ is the identity on k. Thus φ belongs to the group $G_k(L)$, which has order t. To see that G is cyclic, with φ as generator, it suffices to show that the powers id, φ, $\varphi^2,\ldots,\varphi^{n-1}$, are all distinct. For this it is enough that these powers act differently on some one element of L. We use the fact that L^* is a cyclic group, with a generator x, and we consider the elements x, $\varphi(x)$, $\varphi^2(x),\ldots\ldots\varphi^{n-1}(x)$. If $\#(k) = r = p^s$, then $\#(L) = q = p^{sn} = r^n$, and the order of the element x is $\#(L^*) = q-1$. Thus $x^q = x$, but for all u with $1 < u < q$, we have $x^u \neq x$. Since $\varphi^j(x) =$ "x raised to the power r^j ", the elements x, $\varphi(x)$, $\varphi^2(x),\ldots\ldots\varphi^{n-1}(x)$ above, are all distinct. QED.

Remarks: i) When no base field is given, the term "Frobenius automorphism" refers to that for the base field $\mathbf{Z}_p$, i.e. to the automorphism $\varphi: x \mapsto x^p$. The Frobenius automorphism used in part G, is the s^{th} power of this one.

ii) Given any extension $k\subset L$ of degree n, of finite fields, part H) and the FTGT imply there is a 1-1 correspondence between intermediate fields and subgroups of $\mathbf{Z}_n$, i.e. factors of the integer n. This correspondence, hence the FTGT for finite fields, can be seen directly by an argument slightly extending the one given above for part G).

I) The Frobenius automorphism $\varphi: x\mapsto x^p$ of a finite field **F** of characteristic p, is bijective, yet has derivative identically zero.
proof: The derivative of φ is px^{p-1}, which equals zero for all x in F. Yet $\varphi(x) = \varphi(y)$ iff $x^p = y^p$ iff $0 = x^p - y^p = (x-y)^p$ iff $(x-y) = 0$ iff $x=y$. Thus φ is injective, and since F is finite φ is also bijective. **QED.**

#106) Let k be any finite field.
(i) If f is any irreducible polynomial in k[X] of degree n, prove the Galois group of f is $\cong \mathbf{Z}_n$.
(ii) Prove, for every positive integer n, there exists an irreducible polynomial f in k[X] of degree n.