Math 421 : Abstract Algebra II

0. Preliminaries.

You should be prepared to present, to the class at the blackboard and to hand in, the proofs for every result except definitions, the simpler examples, and those statements marked with ♯ (meaning that either the result or the proof takes us beyond the scope of the course, so the result is included only for clearer understanding or enrichment) or with ♭ (meaning that to do the whole proof is not worth the effort, although you are encouraged to do some part of it, or to try a few examples of your own devising, to help you understand the statement; of course, this is always a good idea!). The proofs of the results marked $*$ are particularly difficult. To the usual taxonomy of mathematical results — lemma, proposition, theorem, corollary, remark — I have chosen to add the medieval "scholium": a result included more to provide a broader understanding than to apply to the proofs of later results.

In general, our terminology and notation follows that of Dan Saracino, *Abstract Algebra: A First Course* (Waveland Press, Prospect Heights, Illinois, 1992). In particular, we use $\subseteq$ for "is a subset of" and $\subset$ for "is a proper subset of." Absolute value bars around a finite set means the number of elements in that set. A minus sign between sets (rather than a backslash $\setminus$) denotes set difference. The symbols $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of all integers, rational numbers, real numbers and complex numbers respectively. But throughout our course, the term *ring* means commutative ring with unity (= multiplicative identity), and a ring homomorphism takes the unity to the unity. If we speak of one ring containing another, it is assumed that the smaller is a subring of the larger, i.e., that the operations on the smaller are the restrictions of those on the larger and that the rings share the same unity (so that, by this definition, a proper ideal of a ring is not a subring). *Domain* means integral domain, i.e., a ring in which, if a product is zero, then one of the factors must be zero; or equivalently, if $ab = ac$ and $a \neq 0$, then $b = c$. A domain D has a *field of fractions* (called the "quotient field" in Saracino's text) consisting of all fractions a/b where $a, b \in D$ and $b \neq 0$; it is the smallest field containing D, in the sense that any field containing D also contains (an isomorphic copy of) the field of fractions of D. A *unit* in a ring is an element with a multiplicative inverse, and an *ideal* is a nonempty subset that is closed under addition and "captures" multiplication. If $a_1, a_2, \ldots, a_n$ are elements of the ring R, then the smallest ideal of R that contains these elements is the set

$$a_1R + a_2R + \cdots + a_nR = \{\, a_1r_1 + a_2r_2 + \cdots + a_nr_n : r_1, r_2, \ldots, r_n \in R \,\} ,$$

called the *ideal* (of R) *generated by* $a_1, a_2, \ldots, a_n$. An ideal aR generated by a single element is called a *principal ideal.* The subset of a ring R consisting of only one element, the zero of R, is an ideal; we will often denote this ideal by 0 rather than by $\{0\}$ (by "abuse of language," as the group of mathematicians "Nicolas Bourbaki" say); so $R - 0$ means the set of nonzero elements of R.

We will need:

0.1 Fundamental Theorem of Ring Homomorphisms ♭. (1) *Let I be an ideal in the ring R. Then the set R/I, consisting of the cosets $a + I$ as a varies over R, is a ring with the operations induced by the operations in R (i.e., $(a+I)+(b+I) = (a+b)+I$ and $(a+I)(b+I) = (ab)+I$), called the* factor ring *of R by I (or just "R mod I"); and the* natural *(or* canonical*) map $\eta : R \to R/I$, defined by $\eta(a) = a + I$ for each a in R, is a surjective ring homomorphism.*

(2) *Let R, S be rings and $\varphi : R \to S$ be a ring homomorphism. Then the kernel* $\operatorname{Ker}(\varphi) = \{a \in R : \varphi(a) = 0\}$ *of φ is an ideal of R, and the factor ring $R/\operatorname{Ker}(\varphi)$ is isomorphic (as a ring) to the image $\varphi(R)$ of φ, a subring of S, by the isomorphism $\overline{\varphi} : R/\operatorname{Ker}(\varphi) \to S$ induced by φ as follows: $\overline{\varphi}(a + \operatorname{Ker}(\varphi)) = \varphi(a)$.*

(One often says that φ "factors through" $R/\operatorname{Ker}(\varphi)$ or through $\overline{\varphi}$, since if η denotes the natural map $R \to R/\operatorname{Ker}(\varphi)$, then $\varphi = \overline{\varphi}\eta$, where juxtaposition means composition of functions.)

A final comment about notation: Just as the symbol $\mathbb{Z}$ was inspired by the German "Zahlen" ("numbers"), we will usually use K to denote a field, inspired by the German "Korps". The word "ring" is the same in both English and German; in the literature a ring may be denoted A, for the French "anneau".

1. Special Types of Rings and Ideals.

1.1 Definition. A proper ideal I in a ring R is called

(1) *prime* iff, for a, b in R, $ab \in I$ implies $a \in I$ or $b \in I$.

(2) *maximal* iff there is no proper ideal J of R that properly contains I.

1.2 Examples. If p is a prime integer (2, 3, 5, 7, etc.) then $p\mathbb{Z}$ is both a prime ideal and a maximal ideal in $\mathbb{Z}$. The zero ideal is prime but not maximal in $\mathbb{Z}$. No other ideals in $\mathbb{Z}$ are prime or maximal.

1.3 Fact ♯. In any ring, any element that is not a unit is contained in at least one maximal ideal. But if we were to use the definition of ring in Saracino's text, then there would be rings that have no maximal ideals at all.

1.4 Proposition. *Let I be a proper ideal in a ring R. Then:*

(1) I is prime iff R/I is a domain.

(2) I is maximal iff R/I is a field.

1.5 Corollary. *(1) Every maximal ideal is prime. (2) The zero ideal in a ring is prime iff the ring is a domain.*

1.6 Discussion ♭. Let X be a set, S be a ring, and $Func(X, S)$ denote the set of all functions from X to S. Define addition and multiplication on $Func(X, S)$ "pointwise", i.e., for F, G in $Func(X, S)$, $F + G$ and FG are the functions defined by the equations

$$(F + G)(a) = F(a) + G(a) \ , \qquad (FG)(a) = F(a) \cdot G(a)$$

for all a in X. (Note that $F(a), G(a)$ are elements of S; the addition and multiplication on the right sides of these equations are the ring operations in S.) These operations make $Func(X, S)$ into a ring. An element a of X for which $F(a) = 0$ is called a *zero* or *root* of the function F.

1.7 Warning ♭. Even if X is a ring, almost all the elements of $Func(X, S)$ are <u>not</u> ring homomorphisms; i.e., we almost never have $F(a+b) = F(a)+F(b)$ or $F(ab) = F(a)F(b)$ for all a, b in S. For instance, the only function from $\mathbb{Z}$ to itself that is a ring homomorphism is the identity function, taking each integer to itself.

1.8 Exercise. (1) Define the (pointwise) additive inverse, i.e., the negative, of an element F of $Func(X, S)$.

(2) Verify that the pointwise operations on $Func(X, S)$ satisfy the distributive law.

1.9 Exercise. What are the zeros of the elements $\sin x$, e^x and $x^2 + 4x + 3$ of $Func(\mathbb{R}, \mathbb{R})$?

1.10 Definition. Let R be a ring. A *polynomial* over R is an infinite sequence $f = (f_0, f_1, f_2, \dots)$ of elements of f_i of R for which, for some n, $0 = f_{n+1} = f_{n+2} = f_{n+3} = \dots$. We add and multiply polynomials as follows:

$$f + g = (f_0 + g_0, f_1 + g_1, f_2 + g_2, \dots) \ , \qquad fg = (f_0g_0, f_0g_1 + f_1g_0, f_0g_2 + f_1g_1 + f_2g_0, \dots) \ .$$

(Thus, the n-th "coordinate" $(f + g)_n$ of $f + g$ is $f_n + g_n$, and the n-th coordinate $(fg)_n$ of fg is $\sum_{i=0}^{n} f_i g_{n-i}$.) With these operations, the set of polynomials over R is a ring (♭), and R is isomorphic to the subring of polynomials of the form $(r, 0, 0, 0, \dots)$ as r varies over R (again, ♭).

If we set $x = (0, 1, 0, 0, 0, \dots)$, then by identifying the element r of R with $(r, 0, 0, 0, \dots)$ we can write any polynomial in a more familiar way:

$$(f_0, f_1, f_2, \dots) = f_0 + f_1 x + f_2 x^2 + \dots \qquad .$$

Thus, x is not an "unknown quantity" of some sort (though it is called an *indeterminate* or "variable"); the powers of x just serve in the sum expression as placeholders for the coordinates of the infinite sequence. Though the sum is written as though it had infinitely many terms, it is really finite, since all the "later" f_i's are 0; and when this finiteness is important, we will often write f in the form

$$f = f_0 + f_1 x + f_2 x^2 + \dots + f_n x^n \ ,$$

or in the more familiar descending order of powers of x. If f is written in this way, f_n may or may not be nonzero (i.e., the last nonzero coefficient may have come earlier); whether it is so should be made clear in a separate statement. Hereafter we will usually write polynomials in this way and often incorporate the x into the name of the polynomial: $f(x)$ instead of just f (as if a polynomial were a function instead of a sequence of coefficients; see Discussion 1.12 below). The ring of all polynomials over R in the indeterminate x is denoted $R[x]$. For a given polynomial $f(x) = f_0 + f_1 x + f_2 x^2 + \cdots$, the largest integer n for which $f_n \neq 0$ is called the *degree* of $f(x)$ (we will say that the degree of the zero polynomial is $-\infty$), and that coefficient f_n is called the *leading coefficient* of $f(x)$. If the leading coefficient is 1, $f(x)$ is called *monic.* To include polynomials of more than one indeterminate, we form a polynomial ring over a polynomial ring: $x^2 + y^2 \in (R[x])[y] = R[x, y]$.

The definitions of addition and multiplication of polynomials as sequences were chosen to reflect the familiar addition and multiplication of polynomials as sums. See also Discussion 1.12 below.

1.11 Proposition. *Let D be a domain and $f(x), g(x) \in D[x] - 0$. Then* $\deg(f(x)g(x)) = \deg(f(x)) + \deg(g(x))$.

1.11a Corollary. *If D is a domain, then the only units in the polynomial ring $D[x]$ are the units in D.*

1.12 Discussion. Let $R \subseteq S$ be rings, and let a be an element of S and $f(x)$ be a polynomial in $R[x]$. Then we get an element of S by "substitution":

$$f(a) = f_0 + f_1 a + f_2 a^2 + \dots \ ,$$

where x in the sum expression for $f(x)$ is replaced by a to yield a sum in S, a finite sum because the later f_n's are 0. If we hold $f(x)$ fixed and let a vary over the elements of S, we get a "polynomial function" from S to S, i.e., an element of $Func(S,S)$. It is usually still denoted $f(x)$, though it is possible (see Scholium 1.13; but rare — see Corollary 3.3) that different polynomials will yield the same polynomial function. The definitions of addition and multiplication of polynomials (as sequences or sums) in Definition 1.10 were chosen so that they would make the association of a polynomial in $R[x]$ to the corresponding polynomial function in $Func(S,S)$ into a ring homomorphism $\flat$; i.e., $(f+g)(a) = f(a)+g(a)$ and $(fg)(a) = f(a)g(a)$ for all a in S. (But most polynomial functions, like most functions, are not ring homomorphisms.)

On the other hand, if we hold the element a of S fixed and vary $f(x)$ over the elements of $R[x]$, we get a function from $R[x]$ to S, "evaluation at a," defined by $\varepsilon_a(f(x)) = f(a)$ for each $f(x)$ in $R[x]$. The function ε_a is a ring homomorphism [this statement is the only part of the discussion that requires proof], so its kernel, $\{f(x) \in R[x] : a \text{ is a root of } f\}$, is an ideal in $R[x]$.

1.13 Scholium. *Let K be a field with only finitely many elements; say $|K| = q$. Then every element of K is a zero of the polynomial $x^q - x$. It follows that the distinct polynomials 0 and $x^q - x$ give the same polynomial function in $Func(K,K)$.*

(Suggestion: $K - 0$ is a multiplicative group with $q-1$ elements; apply Lagrange's Theorem.)

1.14 Fact $\sharp$. If in the definition of polynomial we remove the restriction that the f_i's in the sequence $f = (f_0, f_1, f_2, \dots)$ are eventually zero, the definitions of addition and multiplication are still meaningful, since they require only a finite number of additions and/or multiplications in each component. These operations make the set of all sequences $f = (f_0, f_1, f_2, \dots)$ of elements of R, i.e., all possibly infinite sums

$$f(x) = f_0 + f_1 x + f_2 x^2 + \dots ,$$

into a ring, the *ring of formal power series* $R[[x]]$ in x over R. But there is no natural way to interpret evaluation, i.e., substituting a value a for x, into a power series — unless $a = 0$ or $f(x)$ is really a polynomial — since infinite sums are usually not meaningful. (In some cases, an infinite sum can be interpreted as a limit, as in real analysis; but in a general ring, with no sense of "neighborhoods," a limit is meaningless.)

2. Special Types of Domains.

Throughout this section, D denotes a domain. We consider three classes of domains: UFD's, PID's and Euclidean domains. We will see that each is more restrictive than the last, and then we

will see that the last includes both $\mathbb{Z}$ and $K[x]$ for any field K. So we can use all the properties of all three classes later, while we are studying roots of polynomials over a field.

2.1 Definition. Let a, b be elements of D.

(1) We say a *divides* b (in D), and write $a|b$, iff there is an element c of D for which $ac = b$, i.e., iff $b \in aD$ (or equivalently iff $bD \subseteq aD$).

(2) We call a, b *associates* (in D) if there is a unit u in D for which $au = b$.

2.2 Examples. 3 and -3 are associates in $\mathbb{Z}$. The polynomials $2x + 1$ and $4x + 2$ are associates in $\mathbb{Q}[x]$, but not in $\mathbb{Z}[x]$.

2.3 Remark. "Is associate to" is an equivalence relation on D, so we can speak of "associate classes".

2.4 Lemma. *Let $a, b \in D$. Then TFAE ["The following are equivalent"]:*

(1) a, b are associates.

(2) $a|b$ and $b|a$.

(3) $aD = bD$.

(Remark ♭: We have (1) $\implies$ (2) $\iff$ (3) *in any ring, but* (2) $\implies$ (1) *requires the hypothesis of domain.)*

2.5 Note. If K is a field and $f(x) \in K[x] - 0$, then there is exactly one monic polynomial associate to $f(x)$.

2.6 Definition. A nonzero nonunit element p of D is *irreducible* iff the only elements of D that divide p are associates of p and units. A nonzero nonunit element p of D is *prime* if pD is a prime ideal.

2.7 Exercise. An element associate to an irreducible element is irreducible. An element associate to a prime element is prime.

2.8 Examples. Prime numbers (2, 3, 5, etc.) are irreducible elements of $\mathbb{Z}$, and so are their negatives. The polynomial $2x + 4$ is not irreducible in $\mathbb{Z}[x]$, but it is irreducible in $\mathbb{Q}[x]$. The polynomial $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$, but it is reducible (as $(x - \sqrt{2})(x + \sqrt{2})$) in $\mathbb{R}[x]$.

2.9 Proposition. *If an element of D is prime, then it is irreducible.*

2.10 Example. The converse is not always true: Let D be the subring of $\mathbb{Q}[x]$ consisting of the polynomials with no linear term (i.e., the coefficient of x is 0). Then x^2 is irreducible in D; but it is not prime because $(x^3)^2 = x^2x^4 \in x^2D$, while $x^3 \notin x^2D$.

2.11 Definition. A domain D is called a *unique factorization domain,* or UFD (or "factorial ring"), iff every nonzero nonunit a is a product of irreducible elements, and this factorization is "unique up to order and units", i.e., if $a = p_1 \cdots p_m$ and $a = q_1 \cdots q_n$ where all the p_i's and q_j's are irreducibles, then $m = n$ and the q_j's can be rearranged so that p_i is associate to q_i for each $i = 1, \ldots, m$.

2.12 Proposition. *An irreducible element in a UFD is a prime element.*

2.13 Discussion $\flat$. Let D be a UFD. Pick a set $\{p_\lambda\}_{\lambda \in \Lambda}$ consisting of one element from each associate class of irreducible elements. (E.g., in $\mathbb{Z}$, one might pick all the positive prime numbers; in $K[x]$ where K is a field, one might pick all the monic irreducible polynomials.) Then every nonzero element a of D has a unique expression in the form

$$a = u \prod_{\lambda \in \Lambda} p_\lambda^{e_\lambda} \ , \tag{2-1}$$

where the e_λ's are nonnegative integers, all but finitely many of which are equal to zero, and u is a unit in D. For example, in the UFD $\mathbb{Z}$, we have

$$-240 = (-1) \cdot 2^4 \cdot 3^1 \cdot 5^1 \cdot 7^0 \cdot 11^0 \cdot 13^0 \cdot \ldots \text{ and}$$
$$35 = 1 \cdot 2^0 \cdot 3^0 \cdot 5^1 \cdot 7^1 \cdot 11^0 \cdot 13^0 \cdot \ldots \ .$$

Moreover, any element a of the field of fractions K of D has a unique expression in the form (2-1), but the e_λ's now may be negative integers, still all but finitely many equal to zero. For example, in the field of fractions $\mathbb{Q}$ of $\mathbb{Z}$, we can write

$$-\frac{240}{35} = (-1) \cdot 2^4 \cdot 3^1 \cdot 5^0 \cdot 7^{-1} \cdot 11^0 \cdot 13^0 \cdot \ldots \ .$$

Thus, for any UFD D, we get a family of functions $v_\lambda : (K - 0) \to \mathbb{Z}$, called the *essential valuations of* D, given by $v_\lambda(a) = e_\lambda$, the exponent on p_λ in the expression (2-1) for a, for each a in $K - 0$. Note the following useful fact: An element a of $K - 0$ is in D iff $v_\lambda(a) \geq 0$ for every λ in Λ. For convenience, we extend each v_λ to all of K by setting $v_\lambda(0) = \infty$, yielding functions $v_\lambda : K \to \mathbb{Z} \cup \{\infty\}$.

2.14 Proposition. *If $v : K \to \mathbb{Z} \cup \{\infty\}$ is one of the essential valuations of the UFD D, then v satisfies, for any a, b in K:*

(0) $v(a) = \infty$ *iff* $a = 0$,

(1) $v(ab) = v(a) + v(b)$, *and*

(2) $v(a + b) \geq \min\{v(a), v(b)\}$.

2.15 Proposition. *If a function $v : K \to \mathbb{Z} \cup \{\infty\}$ satisfies conditions* (0)–(2) *in Proposition 2.14, then it also satisfies the following conditions:*

(3) $v(1) = v(-1) = 0$.

(4) $v(-a) = v(a)$ *for all* a *in* K.

(5) *If* $v(a) \neq v(b)$, *then* $v(a + b) = \min\{v(a), v(b)\}$. *More generally, if one of the elements* $a_1, a_2, \ldots, a_n$ *of* K *has* v*-value strictly smaller than that of any of the others, then* $v(a_1 + a_2 + \ldots + a_n) = \min\{v(a_1), v(a_2), \ldots, v(a_n)\}$.

(6) *The subset* $v(K - 0)$ *of* $\mathbb{Z}$ *is the set of all multiples of some nonnegative integer* e. *If* $e \neq 0$ *and we replace* v *by* $(1/e)v$, *then* v *becomes surjective and still satisfies conditions* (0)–(2).

2.16 Definition. A domain D is a *principal ideal domain,* or PID, iff every ideal in D is principal.

2.17 Lemma. *(1) If* $I_1 \subseteq I_2 \subseteq I_3 \subseteq \ldots$ *is a chain of ideals in a ring* R, *then the union* $\bigcup_{n=1}^{\infty} I_n$ *is also an ideal of* R.

(2) If D *is a PID, then any chain* $a_1D \subseteq a_2D \subseteq a_3D \subseteq \ldots$ *of ideals in* D *must terminate, i.e., for some* n *we must have* $a_nD = a_{n+1}D = a_{n+2}D = \ldots$.

(3) An irreducible element in a PID generates a maximal ideal (and hence is prime).

(4) If a prime element of a domain divides a product of irreducibles, then it is associate to one of the irreducibles in the product.

(5) If an element of a domain is a product of prime elements, then up to order and units that factorization is the only factorization of that element into irreducibles.

2.18 Theorem $*$**.** *Every PID is a UFD.*

(Suggestion: Use (2) of the lemma to show by contradiction that every nonzero nonunit element in D has at least one irreducible factor. Then use (2) again to show that every nonzero nonunit is a product of (one or more) irreducibles in at least one way. Then the uniqueness of the factorization follows from (3) and (5).)

2.19 Example. Not every UFD is a PID: $\mathbb{Z}[x]$ is a UFD but not a PID; in particular, the ideal $2\,\mathbb{Z}[x] + x\,\mathbb{Z}[x]$ is not principal.

2.20 Discussion $\sharp$. The *(Krull) dimension* of a ring R is the length of the longest chain of prime ideals in R, $P_0 \subset P_1 \subset P_2 \subset \ldots \subset P_n$ (counting inclusions, not prime ideals, so this chain has length n). By Lemma 2.17(3), a PID has dimension 1 (or dimension 0, if it is the trivial case of a PID, a field). In fact, a UFD of dimension 1 (or 0) is a PID. (The UFD $\mathbb{Z}[x]$ has dimension 2: A chain of primes of length 2 is $0 \subset x\,\mathbb{Z}[x] \subset 2\,\mathbb{Z}[x] + x\,\mathbb{Z}[x]$.)

2.21 Definition. A domain D is called a *Euclidean domain* iff there is a function φ (a "Euclidean function") from $D-0$ into the set of nonnegative integers, satisfying:

(1) If $a, b \in D-0$ and $a|b$, then $\varphi(a) \leq \varphi(b)$; and

(2) (Division Algorithm) If $a, b \in D$ with $a \neq 0$, then there exist elements q, r in D for which $b = aq + r$ and either $r = 0$ or $\varphi(r) < \varphi(a)$.

2.22 Theorem. *A Euclidean domain is a PID.*

It can be argued that the concept of Euclidean domain was invented only so that it was possible to identify certain PID's. Indeed, although there are PID's which are not Euclidean (e.g., $\mathbb{Z}[\sqrt{-19}] = \{a + b\sqrt{-19} : a, b \in \mathbb{Z}\}$), I know of no "important" results that hold for every Euclidean domain but not for every PID.

2.23 Lemma. *(Long division of polynomials) Let R be a ring and $f(x), g(x) \in R[x]$; suppose that $f(x) \neq 0$ and that the leading coefficient of $f(x)$ is a unit in R. Then there are elements $q(x), r(x)$ in $R[x]$ for which $g(x) = f(x)q(x) + r(x)$ and either $r(x) = 0$ or $\deg(r(x)) < \deg(f(x))$.*

2.24 Examples. 1. $\mathbb{Z}$ is a Euclidean domain, with absolute value as its Euclidean function. Note that there is no requirement in the definition that q, r are unique: When $a = 3$ and $b = 10$, we can have either $q = 3$ and $r = 1$ or $q = 4$ and $r = -2$.

2. For any field K, the ring of polynomials $K[x]$ is a Euclidean domain, with degree of a polynomial as its Euclidean function. (Lemma 2.23 shows that "degree" satisfies the Division Algorithm.)

3. ♯ The set of *Gaussian integers*, consisting of the complex numbers of the form $x + yi$ where x, y are integers (and i is as usual the square root of -1), is a Euclidean domain, with the square of complex absolute value (i.e., $|x+yi|^2 = x^2+y^2$) as its Euclidean function. To see that the Division Algorithm holds for this function, identify the complex number $x + yi$ with the point in the plane (x, y). Then the discs of radius 1 centered at the Gaussian integers cover the plane; so if we are given two Gaussian integers a, b with $a \neq 0$, then the complex number b/a is within distance 1 of some Gaussian integer q, so $r = a(b/a - q)$ is a Gaussian integer of smaller absolute value than a.

4. Let K be a field and $v : K \to \mathbb{Z} \cup \{\infty\}$ be a function satisfying conditions (0)–(2) of Proposition 2.14. Then $V = \{a \in K : v(a) \geq 0\}$ is a subring of K, and the restriction of v to $V - 0$ is a Euclidean function on V. (The Division Algorithm holds trivially: If $v(a) \leq v(b)$, then the selection $q = b/a$ and $r = 0$ works; while if $v(a) > v(b)$, then $q = 0$ and $r = b$ works.) If $v(K - 0) = 0$, then $V = K$; we assume this is not the case, and using Proposition 2.15(6) we

assume v is surjective. Then the domain V is called a *discrete (rank-one) valuation ring,* or DVR, and $v : K \to \mathbb{Z} \cup \{\infty\}$ is the *(normalized) valuation* associated to V. Examples of these rings are the subrings $\mathbb{Z}_{(p)}$ of $\mathbb{Q}$ for the various prime integers p, defined by $\mathbb{Z}_{(p)} = \{a/b : a, b \in \mathbb{Z}, p \nmid b\}$. In fact, a DVR V is a very nice domain: Let p be an element of V for which $v(p) = 1$; then p and its associates are the only primes in V, i.e., all the nonzero ideals in V have the form $p^n V$ for some positive integer n, and they are arranged in a chain:

$$V \supset pV \supset p^2V \supset p^3V \supset \dots$$

Moreover, for each nonzero element a of K, either a or a^{-1} is in V.

2.25 Discussion ♯. The emphasis in these notes on essential valuations rather than prime factors is somewhat unusual in introductory presentations. The reason it was selected is that an essential family of valuations exists for many domains that are not UFD's; among them "rings of algebraic integers" like $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$, where $(2+\sqrt{-5})(2-\sqrt{-5})$ and $3 \cdot 3$ are essentially different factorizations of 9 into irreducibles.

3. Criteria for Irreducibility of Polynomials.

Just as it is often difficult to tell whether a large integer is prime, it is often difficult to tell whether a polynomial $f(x)$ over a field, even over $\mathbb{Q}$, is irreducible. Indeed, since the set of polynomials of degree less than that of $f(x)$ is infinite (except over a finite field), it is in principle even more difficult. There are a few elementary results in this direction:

3.1 Proposition. *Let K be a field, $r \in K$, and $f(x) \in K[x]$. Then r is a root of $f(x)$ iff $x - r$ divides $f(x)$ in $K[x]$.*

(Suggestion: Apply Lemma 2.23 with $g(x) = x - r$.)

3.2 Corollary. *Let K be a field and $f(x) \in K[x]$ with* $\deg(f(x)) > 1$. *If $f(x)$ has a root in K, then $f(x)$ is reducible in $K[x]$. If* $\deg f(x) = 2$ *or* 3, *then $f(x)$ is irreducible in $K[x]$ iff $f(x)$ has no root in K.*

3.3 Corollary. *If K is a field and $f(x) \in K[x] - 0$, then $f(x)$ has at most* $\deg(f(x))$ *different roots in K.*

Proposition 3.6 below sometimes appears in elementary texts, but the proof is usually incomplete in that it ignores the intervening results. The first lemma (or something like it) is sometimes credited to Kronecker, but it is usually called Gauss's Lemma.

3.4 Gauss's Lemma (valuation version). *Let K be a field and $v : K \to \mathbb{Z} \cup \{\infty\}$ be a function satisfying conditions* (0)–(2) *of Proposition 2.14. Extend v to $K[x]$ by letting $v(g(x))$ denote the minimum v-value of the coefficients of $g(x)$. Then for $g(x), h(x)$ in $K[x]$, $v(g(x)h(x)) = v(g(x)) + v(h(x))$.*

(Suggestion: Let i, j be the smallest subscripts for which $v(g_i)$ and $v(h_j)$ attain the respective minima. Then the coefficient of x^{i+j} in $g(x)h(x)$ has the correct v-value.)

Gauss's Lemma says that the extension of v to $K[x]$ also satisfies condition (1) of Proposition 2.14. Conditions (0) and (2) also hold, but (1) is the hardest to prove and the one we will need below.

3.5 Corollary. *Let D be a UFD with field of fractions K, and let $f(x), g(x) \in D[x] - 0$ and $h(x) \in K[x]$ for which $f(x) = g(x)h(x)$. If the coefficients of $g(x)$ have no common factor in D (except units), then $h(x) \in D[x]$.*

(Suggestion: It suffices to show that $v_\lambda(h(x)) \geq 0$ for all the essential valuations v_λ of D.)

3.6 Proposition. *Let D be a UFD with field of fractions K, and let $f(x) \in D[x]$. Then any root of $f(x)$ in K has the form: a factor (in D) of the constant term of f over a factor of the leading coefficient of f.*

One of the few general results that yield irreducibility for polynomials of higher degree is:

3.7 Eisenstein's Criterion. *Let D be a UFD with field of fractions K, let p be a prime element of D, and let*

$$f(x) = f_0 + f_1 x + f_2 x^2 + \ldots + f_n x^n \in D[x] - 0 \ .$$

If $p \nmid f_n$ (so that $f_n \neq 0$), if $p | f_i$ for each $i = 0, 1, 2, \ldots, n-1$, and if $p^2 \nmid f_0$, then $f(x)$ is irreducible in $K[x]$.

(Suggestion: Suppose that $f(x) = g(x)h(x)$ in $K[x]$, and let v be the essential valuation of D associated to p. Then we may assume that $v(g(x)) = 0$ (so that $v(h(x)) = 0$ also by Gauss's Lemma) and that $v(g_0) = 1$ and $v(h_0) = 0$. Let i be smallest so that $v(g_i) = 0$, and argue that $h(x)$ must be a constant, i.e., a unit in $K[x]$.)

3.8 Proposition. *For any (positive) prime integer p, the polynomial $\Phi_p(x) = 1 + x + x^2 + \ldots + x^{p-1}$ is irreducible over $\mathbb{Q}$.*

(Suggestion: It is enough to show that $\Phi_p(x+1)$ is irreducible; use the equation $\Phi_p(x) = (x^p-1)/(x-1)$, the Binomial Theorem, and the fact that p divides the binomial coefficients $\binom{p}{j}$ for all $j = 1, 2, \ldots, p-1$.)

On the basis of Gauss's Lemma, we can also show that

3.9 Scholium ♯. *Let D be a UFD. Then $D[x]$ is a UFD. Moreover, let $\{p_\lambda\}_{\lambda\in\Lambda}$ be a complete set of nonassociate primes in D, as in Discussion 2.13, and let K be the field of fractions of D. For $f(x)$ in $K[x] - 0$, the* content *of $f(x)$, i.e., the element*

$$c(f) = \prod_{\lambda\in\Lambda} p_\lambda^{v_\lambda(f(x))}$$

of K, is such that $(c(f))^{-1}f(x) \in D[x]$ and the coefficients of $(c(f))^{-1}f(x)$ have no common factor except units. Let $\{q_\mu(x)\}_{\mu\in M}$ be a complete set of nonassociate irreducible polynomials in $K[x]$. Then a complete set of nonassociate primes in the UFD $D[x]$ is

$$\{p_\lambda\}_{\lambda\in\Lambda} \cup \{(c(q_\mu)^{-1}q_\mu(x)\}_{\mu\in M} \ .$$

3.10 Example. Some of the prime elements in the UFD $\mathbb{Z}[x]$ are 2, x^2+5, and $2x-3$. The content of $f(x) = (10/3)x^2 + (4/5)x + 12$ is $2/15$, and $(c(f))^{-1}f(x) = 25x^2 - 6x + 90$.

4. Field Extensions: Finite, Algebraic, and Transcendental.

Throughout this section, K denotes a field. (It may help to think of it as the field $\mathbb{Q}$, but later we will want to apply these results in more general situations.)

4.1 Definition. Let $K \subseteq L$ be fields.

(1) We call L an *extension field* of K, and we often write L/K. (Note: L/K is a traditional notation for $K \subseteq L$, having nothing to do with factor rings.) Of course, L is a vector space over K; the dimension of L as a K-vector space is denoted by $[L:K]$ and called the *degree* of L over K. If $[L:K]$ is finite, we call L a *finite extension* of K. For elements $a_1, a_2, \ldots$ of L, the smallest subfield of L that contains both K and all the a_i's is denoted $K(a_1, a_2, \ldots)$ and called the *subfield* (of L) *generated* (over K) *by* $a_1, a_2, \ldots$.

(2) An element a of L is *algebraic over* K iff there is a nonzero polynomial in $K[x]$ of which a is a root; if there is no such polynomial over K, then a is *transcendental over* K. If every element of L is algebraic over K, then L is called an *algebraic extension* of K.

4.2 Remark ♭. For $K, L, a_1, a_2, \dots$ as in (1) of this definition, the smallest subring of L containing K and all the a_i's is just the result of evaluating all polynomials at the a_i's, i.e., taking all the polynomials $f(x_1, x_2, \dots)$ in $K[x_1, x_2, \dots]$ and considering all the elements of L of the form $f(a_1, a_2, \dots)$; so we denote this subring by $K[a_1, a_2, \dots]$. The field of fractions of this ring is $K(a_1, a_2, \dots)$; so if $K[x_1, x_2, \dots]$ is a field, then it is equal to $K(a_1, a_2, \dots)$. But $K(a_1, a_2, \dots)$ may be strictly larger than $K[a_1, a_2, \dots]$. In fact, if there are only finitely many a_i's, then we claim that the ring they generate is equal to the field they generate iff all the a_i's are algebraic over K. (A first step toward proving this claim is part of Theorem 4.5.)

4.3 Examples. (1) Since $\sqrt{3}$ is a zero of the polynomial x^2-3 in $\mathbb{Q}[x]$, $\sqrt{3}$ is algebraic over $\mathbb{Q}$. Also, the fact that $(\sqrt{3})^2 = 3$ means that if we have any polynomial expression in $\sqrt{3}$ with coefficients in $\mathbb{Q}$, we can rewrite the relation so that it has no squared or higher powers of $\sqrt{3}$, i.e., it has the form $a+b\sqrt{3}$ where a, b are in $\mathbb{Q}$. Thus, $\mathbb{Q}[\sqrt{3}] = \{a+b\sqrt{3} : a, b \in \mathbb{Q}\}$. Moreover, if we have a quotient of two such expressions, $(a+b\sqrt{3})/(c+d\sqrt{3})$, then we can multiply numerator and denominator by the "conjugate" $c - d\sqrt{3}$ of the denominator, and again reach the form (rational) + (rational)$\sqrt{3}$. Thus, $\mathbb{Q}[\sqrt{3}]$ is a field, so $\mathbb{Q}[\sqrt{3}] = \mathbb{Q}(\sqrt{3})$; and a vector space basis for $\mathbb{Q}(\sqrt{3})$ over $\mathbb{Q}$ is $\{1, \sqrt{3}\}$, so $[\mathbb{Q}(\sqrt{3}) : \mathbb{Q}] = 2$.

(2) Similarly, since $\sqrt[5]{3}$ is a root of $x^5 - 3$, we see a $\mathbb{Q}$-basis for $\mathbb{Q}(\sqrt[5]{3})$ is $1, \sqrt[5]{3}, \sqrt[5]{3}^2, \sqrt[5]{3}^3, \sqrt[5]{3}^4$, and $[\mathbb{Q}(\sqrt[5]{3}) : \mathbb{Q}] = 5$. ("Rationalizing a denominator" is now much harder, but it is still possible.) But the primitive fifth root of unity, $\omega_5 = \cos(2\pi/5) + i\sin(2\pi/5) = e^{2\pi i/5}$, one of the complex roots of the polynomial $x^5 - 1$, is such that $[\mathbb{Q}(\omega_5) : \mathbb{Q}] = 4$, since $x^5 - 1$ factors as $(x-1)(x^4 + x^3 + x^2 + x + 1)$, and ω_5 is a root of the second factor, which is irreducible by Proposition 3.8. A $\mathbb{Q}$-basis for $\mathbb{Q}(\omega_5)$ is $1, \omega_5, \omega_5^2, \omega_5^3$.

(3) $\mathbb{Q}(3^{1/2}, 3^{1/4}, 3^{1/8}, \dots)$ is an algebraic extension of $\mathbb{Q}$ that is not finite.

(4) The same extension field can be generated in many different ways. For example,

$$\mathbb{Q}(\sqrt[4]{2}, i\sqrt[4]{2}, -\sqrt[4]{2}, -i\sqrt[4]{2}) = \mathbb{Q}(\sqrt[4]{2}, i) \ .$$

(5) ♯ It is true, but difficult to prove (cf. Serge Lang, *Algebra,* Addison-Wesley, Reading, MA, 1965, Appendix), that π (the ratio of the circumference of a circle to its diameter) is transcendental over $\mathbb{Q}$. As a result, all of the powers of π are linearly independent over $\mathbb{Q}$, so $\mathbb{Q}(\pi)/\mathbb{Q}$ is not a finite extension. The same is true of e, the base of the natural logarithms.

4.4 Proposition. *A finite field extension is always algebraic.*

(Suggestion: If a is an element of the larger field, then the powers of a cannot all be linearly independent over the smaller field.)

4.5 Theorem. *Let a be an element of an extension field of K such that a is algebraic over K. Then there is a unique monic irreducible polynomial $p(x)$ in $K[x]$ such that, for each $f(x)$ in $K[x]$, $f(a) = 0$ iff $p(x)|f(x)$. Moreover, $K[x]/(p(x)K[x]) \cong K[a] = K(a)$ (the isomorphism being given by a map that takes the coset $x + (p(x)K[x])$ to a and is the identity on K), a K-basis for $K(a)$ is $1, a, a^2, \ldots, a^{(\deg(p))-1}$, and $[K(a) : K] = \deg(p(x))$.*

4.6 Definition. If a is algebraic over K, then the monic irreducible polynomial $p(x)$ in Theorem 4.5 is called the *minimal polynomial of a over K*, and denoted $\mathrm{Irr}(a, K)$.

4.7 Examples. (1) $\mathrm{Irr}(\sqrt{3}, \mathbb{Q}) = x^2 - 3 = \mathrm{Irr}(-\sqrt{3}, \mathbb{Q})$; but $\mathrm{Irr}(\sqrt{3}, \mathbb{R}) = x - \sqrt{3}$.

(2) $\mathrm{Irr}(\sqrt[5]{3}, \mathbb{Q}) = x^5 - 3 = \mathrm{Irr}(\omega_5 \sqrt[5]{3}, \mathbb{Q})$, where ω_5 is as in Example 4.3(2).

(3) $\mathrm{Irr}(\omega_5, \mathbb{Q}) = x^4 + x^3 + x^2 + x + 1$.

(4) $\mathrm{Irr}(\sqrt[4]{3}, \mathbb{Q}(\sqrt{3})) = x^2 - \sqrt{3}$.

(5) $\mathrm{Irr}((1 + \sqrt{3})/2, \mathbb{Q}) = (x - (1 + \sqrt{3})/2)(x - (1 - \sqrt{3})/2) = x^2 - x - (1/2)$.

(6) $\mathrm{Irr}(i, \mathbb{R}) = x^2 + 1 = \mathrm{Irr}(i, \mathbb{Q})$.

4.8 General Example. For a general positive integer n, the *cyclotomic polynomial* $\Phi_n(x)$ is the minimal polynomial of the "primitive" n-th roots of unity in $\mathbb{C}$; one of these is

$$\omega_n = \cos(2\pi/n) + i \sin(2\pi/n) = e^{2\pi i/n} ,$$

and the others are the powers ω_n^m where $m \in \{2, 3, \ldots, n-1\}$ and $\gcd(m, n) = 1$. Thus, for example, $\omega_1 = 1$, $\omega_2 = -1$, $\omega_4 = i$, and $\omega_4^3 = -i$. (The n-th roots of unity form a cyclic group under multiplication, and the primitive ones are the generators of that group — cf. Saracino, Theorem 4.4(ii), page 38). The polynomial $\Phi_n(x)$ is defined by induction: $\Phi_1(x) = x - 1$, and for $n > 1$,

$$\Phi_n(x) = (x^n - 1) / \left(\prod_{d|n, d \neq n} \Phi_d(x) \right) .$$

(In effect, to obtain $\Phi_n(x)$, we take the polynomial $x^n - 1$ whose roots are all the n-th roots of unity, and remove the roots of strictly smaller order.) It follows from Proposition 3.8 that, if p is prime, then $[\mathbb{Q}(\omega_p) : \mathbb{Q}] = p - 1$.

4.9 Proposition ♭**.** *Let $f(x)$ be any nonconstant element of $K[x]$. Then there is an extension field L of K in which $f(x)$ has a root r, and $L = K(r)$.*

Proof sketch. Let $p(x)$ be an irreducible factor of $f(x)$ in $K[x]$, and let t be an indeterminate different from x. Then the factor ring $K[t]/(p(t)K[t])$ is a field containing both (an isomorphic copy of) K and an element $t + (p(t)K[t])$ that is a zero of $p(x)$ and hence of $f(x)$. □

The construction of the extension field in Proposition 4.9 is of course completely artificial. The point is just that, if we need a root of a polynomial, we can build a smallest field containing one, and by part of Theorem 4.5, if the polynomial is irreducible, then any such field is isomorphic to any other.

4.10 Scholium. *If a is an element of an extension field of K such that a is transcendental over K, then $K[x] \cong K[a] \subset K(a)$.*

4.11 Theorem. *Let $K \subseteq F \subseteq L$ be fields, and suppose $[F : K], [L : F]$ are finite. Then $[L : K]$ is also finite, and $[L : K] = [L : F][F : K]$ (multiplication of integers).*

4.12 Corollary. *(1) Let a be an element of an extension field of K. Then a is algebraic over K iff $[K(a) : K]$ is finite.*

(2) Let L be an extension field of K. Then the set of elements of L that are algebraic over K is a subfield of L.

4.13 Notes ♭. (1) Given elements a, b of an extension field of K, even if the minimal polynomials of a and b are known, there is no simple way to find a polynomial of which $a + b$ is a root. That is why Corollary 4.12 (2) is set after (1), which in turn is set after Theorem 4.11.

(2) The set (field) of elements of L algebraic over K is called the *algebraic closure of K in L*. The algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$ is called the field of *algebraic numbers*. (Cf. Discussion 6.12 below.)

5. Application: Geometric Constructions.

You probably recall from high school geometry class using "Euclidean tools", i.e., the compass and straightedge, to construct certain figures. (As a refresher, you might recall how to construct an equilateral triangle with a given segment as one side, a perpendicular to a line through a given point on the line, and the bisector of a given angle.) We show in this section that questions of whether it is possible to construct a given figure are really questions about degrees of field extensions.

5.1 Discussion. There are two remarks we must make to connect the geometric context to the algebraic one:

First, we consider the tools themselves. The first Euclidean tool is a "collapsing compass": it can draw a circle with a given center through a given point, but it cannot, *a priori,* pick up a length from one part of one part of the plane and use it as a radius for a circle with a center elsewhere. But the second proposition in Euclid's *Elements* shows how to use a collapsing compass to transfer a length in this way, so we need not worry about this difference between the Euclidean tool and a "real" compass. There is a more important difference in the second Euclidean tool, however: it is not a ruler, to measure distances; it is only a straightedge, to draw the straight line through two points. (More about why this difference is important below.) These tools correspond to three of the "Postulates" set down in Euclid's *Elements:* "Let the following be postulated: 1. To draw a straight line [segment] from any point to any point. 2. To produce [i.e., extend] a straight line [segment] continuously in a straight line. 3. To describe a circle with any center and distance [i.e., radius]"

Second, it will be convenient to speak of "constructible numbers". Most construction problems start with a given line segment, like a segment on which to construct an equilateral triangle, or the radius of a given circle to which a tangent from an exterior point is to be constructed; and when a segment is not given, as in the problem of bisecting a given angle, we can choose one at random. We associate points in the plane with complex numbers by making that given line segment the segment joining 0 and 1 in the usual "complex plane" (or "Argand diagram") representation of the complex numbers: one end of the given segment becomes (0,0) and the other (1,0) to establish a rectangular coordinate system on the plane, and the point (a, b) is associated to the complex number $a+bi$. We describe recursively what it means to "locate" a point, or equivalently a complex number, in the plane: The numbers 0 and 1 are (automatically) located, as the endpoints of the given segment. To locate another point P means to draw two lines, determined by points that have already been located; or two circles, each with a center and some point on the circumference that have already been located; or such a line and such a circle; so that the two curves (lines and/or circles) meet at P. We will call a complex number *constructible* if it can be "located" in this way (in a finite number of steps).

5.2 Lemma. *The complex number $a + bi$ is constructible iff the nonnegative real numbers $|a|$ and $|b|$ are constructible.*

(A proof of this result amounts to a description of how, given $a + bi$, one could construct $|a|$ and $|b|$, and vice versa.)

5.3 Proposition ♭**.** *The set* $\mathbb{E}$ (for Euclidean) *of constructible complex numbers is a subfield of* $\mathbb{C}$ *that contains the square roots of each of its elements.*

Proof sketch. For positive real numbers, which can be interpreted as lengths, the field operations can be done geometrically: addition and subtraction of lengths are easy, and multiplication and division are accomplished by using similar triangles with the original given segment, of length 1, as one side of one triangle. To construct a segment of length $\sqrt{a}$, where a is a positive real, draw a segment of length $a+1$, a perpendicular line 1 unit into the segment, and a semicircle of which the segment is the diameter; the part of the perpendicular from the segment to the semicircle has length $\sqrt{a}$. It is not difficult to see how all this can be generalized to all complex numbers, but there are many cases to consider. In particular, for the square root of a complex number, it is convenient to write it in "polar form": $r(\cos\theta + i\sin\theta)$ where r is the distance from the origin O to the point P in the plane corresponding to this number, and θ is the angle that the ray $\overrightarrow{OP}$ makes with the positive real axis; the square roots of this number are $\pm\sqrt{r}(\cos(\theta/2) + i\sin(\theta/2))$. □

Euclid's *Elements* showed how to perform certain constructions, and it has always been a favorite pastime of amateur mathematicians to repeat these constructions and to find ways to do others. (George Martin, *The Foundations of Geometry and the Non-Euclidean Plane* (Springer-Verlag, New York, 1975) page 479: "The old games are the best games.") But there are three constructions that the Greek mathematicians were unable to do with Euclidean tools:

1. **Duplication of the Cube**: Given (an edge of) a cube, construct (an edge of) the cube with exactly twice the volume of the given cube. (This is also called the "Delian Problem," because, according to legend, it was inspired by the instruction from the oracle of Apollo at Delos that, to end a plague, the cubical altar there must be doubled in size.)
2. **Squaring the Circle**: Given a circle, construct a square with exactly the same area.
3. **Trisecting the Angle**: Given an angle, construct an angle exactly one-third the size (in angle measure).

It is not claimed that these tasks are utterly impossible; the ancients constructed other tools that would perform them. Indeed, if the straightedge were a ruler (or even if it were allowed to make two marks on the straightedge — cf. Howard Eves, *An Introduction to the History of Mathematics, 5th ed.,* CBS College Publishing, New York, 1983; Problem Study 4.6, page 93), it would be possible to trisect any angle. The challenge is to perform these constructions with Euclidean tools or to show that they are impossible with these tools.

5.4 Notation and Remark. Since we know better the equations of lines and circles in the context of real numbers rather than complex numbers, it is convenient to be able to refer just to the real constructible numbers: Set $\mathbb{E}_{\mathbb{R}} = \mathbb{E} \cap \mathbb{R}$. Then $\mathbb{E} = \{a + bi : a, b \in \mathbb{E}_{\mathbb{R}}\}$, or in geometric terms, the constructible points are just those whose rectangular coordinates are elements of $\mathbb{E}_{\mathbb{R}}$, i.e., are the points in the plane $(\mathbb{E}_{\mathbb{R}})^2$. Note that, by Proposition 5.3 and the quadratic formula, if $a, b, c \in \mathbb{E}_{\mathbb{R}}$ and $b^2 - 4ac \geq 0$, then the roots of $ax^2 + bx + c$ are again in $\mathbb{E}_{\mathbb{R}}$.

Lemma 5.5. *Let K be a subfield of $\mathbb{R}$, and $A = (a_1, a_2)$, $B = (b_1, b_2)$, $C = (c_1, c_2)$, $D = (d_1, d_2)$ be points in K^2.*

(1) ♭ If $P = (p_1, p_2)$ is the point where the lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$ meet, then p_1, p_2 are in K; i.e., $[K(p_1, p_2) : K] = 1$.

(2) If $P = (p_1, p_2)$ is a point where the line $\overleftrightarrow{AB}$ meets the circle $C(D)$ though D with center C, then $[K(p_1, p_2) : K] = 1$ or 2. Indeed, unless $\overleftrightarrow{AB}$ is parallel to the vertical axis, we have that p_1 is a root of a polynomial of degree 2 with coefficients in K and $p_2 \in K(p_1)$;

(3) If $P = (p_1, p_2)$ is a point where the circle $A(B)$ meets the circle $C(D)$, then again $[K(p_1, p_2) : K] = 1$ or 2.

Proof. (1) Since the coordinates of A, B, C, D are elements of K, (p_1, p_2) is the solution of a system of two linear equations with coefficients in K. Finding the solution of such a system (if it exists at all, i.e., if the lines are not parallel), by solving one equation for one variable and substituting into the other, involves only the field operations, so the coordinates p_1, p_2 of a solution are elements of K.

(2), (3) Left to you. (Suggestion for (3): Reduce to the kind of system which arises in (2).) □

5.6 Theorem. *A real number p is constructible (i.e., is in $\mathbb{E}_{\mathbb{R}}$) iff there is a chain of subfields $\mathbb{Q} = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ of $\mathbb{R}$ for which $p \in K_n$ and $[K_j : K_{j-1}] = 1$ or 2 for each $j = 1, \ldots, n$.*

5.7 Corollary. *If a real number p is constructible, then $[\mathbb{Q}(p) : \mathbb{Q}]$ is a (finite) power of 2. (The converse is false; see Example 9.8.)*

5.8 Proposition. *It is impossible to Duplicate the Cube (using only Euclidean tools).*

5.9 Proposition. *It is impossible to Square the Circle. (You may assume that π is transcendental over $\mathbb{Q}$.)*

5.10 Note. An angle θ can be constructed iff the complex number $\cos\theta + i \sin\theta$ is constructible, i.e., iff the real numbers $\cos\theta$ and/or $\sin\theta$ are constructible.

5.11 Proposition. *It is impossible to construct an angle of* $20°$*. Hence, it is impossible to Trisect an Angle of* $60°$*.*

(Suggestion: Use the formula $\cos(3\theta) = 4\cos^3\theta - 3\cos\theta$.)

Recall that a *regular* polygon is one in which all the sides have the same length and all the interior angles have the same measure.

5.12 Theorem (Gauss) ♯. *It is possible to construct a regular polygon of* n *sides iff* $n = 2^r p_1 p_2 \dots p_s$ *where* r *is a nonnegative integer and each* p_i *is a Fermat prime, i.e., a prime number of the form* $2^{2^m} + 1$*. In particular, it is possible to construct a regular polygon of 17 sides.*

For a proof, see Ian Stewart, *Galois Theory, 2nd ed.* (Chapman and Hall, London, 1989; Chapter 17).

6. The Galois Group.

6.1 Definition. Let R, S be rings containing the field K. A ring homomorphism $\varphi : R \to S$ is called a *K-homomorphism* iff $\varphi(a) = a$ for every element of K, i.e., iff φ leaves every element of K fixed. (This amounts to saying that φ, in addition to being a ring homomorphism, is also a linear transformation of vector spaces over K.)

6.2 Lemma. *Let* R, S *be rings containing the field* K*, and let* $\varphi : R \to S$ *be a* K*-homomorphism.*

(1) ♭ Let $f(x_1, x_2, \dots, x_n)$ *be a polynomial with coefficients from* K *and* $a_1, a_2, \dots, a_n$ *be elements of* R*. Then* $\varphi(f(a_1, a_2, \dots, a_n)) = f(\varphi(a_1), \varphi(a_2), \dots, \varphi(a_n))$*.*

(2) If a *in* R *is a root of the polynomial* $f(x)$ *from* $K[x]$*, then* $\varphi(a)$ *in* S *is also a root of* $f(x)$*. (For brevity: "Under a* K*-homomorphism, roots go to roots.")*

6.3 Lemma. *(1) If* L/K *is a field extension, then every* K*-homomorphism from* L *to a ring containing* K *is a monomorphism, i.e., is injective.*

(2) If L/K *is algebraic over* K*, then every* K*-homomorphism from* L *to itself is an automorphism.*

(3) If S *is a ring containing* K*, if* $L = K(a_1, a_2 \dots)$*, and if* $\varphi : L \to S$ *is a* K*-monomorphism, then* φ *is uniquely determined by the images of* $a_1, a_2, \dots$*, in the sense that if* $\psi : L \to S$ *is also a* K*-monomorphism and* $\varphi(a_i) = \psi(a_i)$ *for each* $i = 1, 2, \dots$*, then* $\varphi = \psi$*, i.e.,* $\varphi(a) = \psi(a)$ *for every* a *in* L*.*

(Suggestion: For (2), use the fact that, if $a \in L$, then a K-homomorphism from L to itself takes the finite set of roots of $\mathrm{Irr}(a, K)$ in L to itself. For (3), note first that $\varphi(a) = \psi(a)$ for every a in

$K[a_1, a_2, \dots]$.)

6.4 Definition. If L/K is an algebraic extension, then the set of all K-automorphisms of L is called the *Galois group of L over K* and denoted $\mathrm{Gal}(L/K)$.

6.5 Proposition. *Let L/K be an algebraic extension.*

(1) $\mathrm{Gal}(L/K)$ *is a group under the operation of composition of functions.*

(2) *Suppose $L = K(a_1, a_2, \dots)$, and let X denote the set of all roots in L of the polynomials* $\mathrm{Irr}(a_i, K)$. *Then there is a monomorphism of* $\mathrm{Gal}(L/K)$ *into the symmetric group* $\mathrm{Sym}(X)$ *on the set X, i.e., the set of all 1-1 functions from X onto itself, given by restriction of a K-automorphism of L to the set X.*

A good understanding of the following examples is essential to this course. Please study them carefully.

6.6 Examples. (1) Since $\mathbb{C} = \mathbb{R}(i)$, and $\mathrm{Irr}(i, \mathbb{R}) = x^2+1$, any element of $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ is determined by the image of i; and that image must be one of the roots of $x^2 + 1$, i.e., either i or $-i$. Thus, $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ has only two elements, the identity function (which we will denote $id_{\mathbb{C}}$) and the familiar "complex conjugation," i.e., the map taking $a+bi$ (where a, b are real numbers) to $a-bi$. (A complex conjugate is often denoted with an overbar: $\overline{a + bi} = a - bi$.)

(2) Similarly, for any element a of $\mathbb{Q}$ that is not a square of an element of $\mathbb{Q}$ (e.g., $a = -1$ or 2 or 3 or $\dots$), we have $\mathrm{Irr}(\sqrt{a}, \mathbb{Q}) = x^2 - a$, and $\mathrm{Gal}(\mathbb{Q}(\sqrt{a})/\mathbb{Q})$ again has two elements, the identity function $id_{\mathbb{Q}(\sqrt{a})}$ and the function that takes $b + c\sqrt{a}$, where $b, c \in \mathbb{Q}$, to $b - c\sqrt{a}$. (The latter automorphism is also often called "conjugation.")

(3) The polynomial $x^4 - 2$ is irreducible over $\mathbb{Q}$. Its roots are the real numbers $\sqrt[4]{2}$ and $-\sqrt[4]{2}$ and the complex numbers $i\sqrt[4]{2}$ and $-i\sqrt[4]{2}$; consider the field $L = \mathbb{Q}(\sqrt[4]{2}, -\sqrt[4]{2}, i\sqrt[4]{2}, -i\sqrt[4]{2})$. By Example 4.3(4), $L = \mathbb{Q}(\sqrt[4]{2}, i)$, so the images of $\sqrt[4]{2}$ and i determine the element of $\mathrm{Gal}(L/\mathbb{Q})$. The restriction γ of complex conjugation to L gives one element of $\mathrm{Gal}(L/\mathbb{Q})$ of order 2 (i.e., if we apply it twice in succession, we get the identity). But (see Corollary 6.20 below) there is also at least one $\mathbb{Q}$-automorphism of L that takes $\sqrt[4]{2}$ to each of the other roots of its minimal polynomial. In particular, there is one, say φ, for which $\varphi(\sqrt[4]{2}) = i\sqrt[4]{2}$. Since φ is a field automorphism, we get $\varphi^2(\sqrt[4]{2}) = \varphi(i)\varphi(\sqrt[4]{2})$, and we already know what the second factor is. We don't know whether $\varphi(i)$ is i or $-i$, but if it is $-i$, then $\varphi\gamma(i) = i$, and we still have $\varphi\gamma(\sqrt[4]{2}) = i\sqrt[4]{2}$, so we may assume (by replacing φ with $\varphi\gamma$ if necessary) that $\varphi(i) = i$. Then φ is an element of $\mathrm{Gal}(L/\mathbb{Q})$ of order 4, corresponding to the 4-cycle permutation $(\sqrt[4]{2}, i\sqrt[4]{2}, -\sqrt[4]{2}, -i\sqrt[4]{2})$ of the roots of $x^4 - 2$. These four

numbers are arranged in a square set diagonally in the complex plane. The element γ reverses the top and bottom vertices while leaving the ones on the real axis fixed. The element φ rotates the vertices 90°. Thus, these elements generate a group of 8 elements that is isomorphic to the dihedral group of the square, D_4. There are only four possible images of $\sqrt[4]{2}$ under $\mathbb{Q}$-automorphisms, and only two possible images of i; since these images uniquely determine a $\mathbb{Q}$-automorphism of L, there are only 8 possible elements of $\mathrm{Gal}(L/\mathbb{Q})$, so we have them all: $\mathrm{Gal}(L/\mathbb{Q}) \cong D_4$.

The essence of Galois theory is that there is a very close relationship between the family of subgroups of $\mathrm{Gal}(L/K)$ and the family of fields between K and L:

6.7 Definition. Let L/K be an algebraic field extension.

(1) For a subgroup H of $\mathrm{Gal}(L/K)$, we call the subset

$$\mathcal{F}(H) = \{a \in L : \varphi(a) = a \text{ for each } \varphi \text{ in } H\}$$

of L the *fixed field of* H.

(2) For a subfield F of L containing K, we call the subset

$$\mathcal{G}(F) = \{\varphi \in \mathrm{Gal}(L/K) : \varphi(a) = a \text{ for each } a \text{ in } F\} = \mathrm{Gal}(L/F)$$

of $\mathrm{Gal}(L/K)$ the *stabilizer of* F.

6.8 Theorem. *Let L/K be an algebraic field extension.*

(1) *For a subgroup H of* $\mathrm{Gal}(L/K)$*, the set $\mathcal{F}(H)$ is a subfield of L containing K.*

(1′) *For a subfield F of L containing K, the set $\mathcal{G}(F)$ is a subgroup of* $\mathrm{Gal}(L/K)$*.*

(2) *For subgroups $H_1 \subseteq H_2$ of* $\mathrm{Gal}(L/K)$*, we have $\mathcal{F}(H_2) \subseteq \mathcal{F}(H_1)$.*

(2′) *For subfields $F_1 \subseteq F_2$ of L containing K, we have $\mathcal{G}(F_2) \subseteq \mathcal{G}(F_1)$.*

(3) *For a subgroup H of* $\mathrm{Gal}(L/K)$*, we have $H \subseteq \mathcal{G}(\mathcal{F}(H))$ and $\mathcal{F}(\mathcal{G}(\mathcal{F}(H))) = \mathcal{F}(H)$.*

(3′) *For a subfield F of L containing K, we have $F \subseteq \mathcal{F}\mathcal{G}(F)$ and $\mathcal{G}(\mathcal{F}(\mathcal{G}(F))) = \mathcal{G}(F)$.*

(4) *For a subset S of* $\mathrm{Gal}(L/K)$*, let $\langle S\rangle$ denote the smallest subgroup containing S. Then we have $\mathcal{F}(\langle S\rangle) = \{a \in L : \varphi(a) = a \text{ for all } \varphi \text{ in } S\}$.*

(4′) *For a subset T of L, we have $\mathcal{G}(K(T)) = \{\varphi \in \mathrm{Gal}(L/K) : \varphi(a) = a \text{ for all } a \text{ in } T\}$.*

(5) *For subgroups H_1, H_2 of* $\mathrm{Gal}(L/K)$*, the smallest subgroup of* $\mathrm{Gal}(L/K)$ *containing both H_1 and H_2 is $\langle H_1 \cup H_2\rangle$, and we have $\mathcal{F}(\langle H_1 \cup H_2\rangle) = \mathcal{F}(H_1) \cap \mathcal{F}(H_2)$.*

(5′) *For subfields F_1, F_2 of L containing K, the smallest subfield of L containing both F_1 and F_2 is $F_1(F_2) = F_2(F_1)$, and we have $\mathcal{G}(F_1(F_2)) = \mathcal{G}(F_1) \cap \mathcal{G}(F_2)$.*

Parts (2) and (2′) of this theorem show that the correspondences $\mathcal{F}$ and $\mathcal{G}$ are inclusion-reversing functions between the family of subgroups of $\mathrm{Gal}(L/K)$ and the family of fields between K and L. This theorem is a bit confusing but not "deep" in the sense of using hard results. The achievement of Galois was to recognize the hypotheses needed to make $\mathcal{F}$ and G even better behaved. We can improve part (3) of the theorem, at least in the case of a finite subgroup H, without additional hypotheses; but we need some rather technical preliminary results, which use ideas from linear algebra.

6.9 Lemma. *Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be distinct monomorphisms of a field L into another field E. Then:*

(1) *(Dedekind) $\varphi_1, \varphi_2, \ldots, \varphi_n$ are linearly independent over E (i.e., the only list of elements $b_1, b_2 \ldots, b_n$ of E for which*

$$b_1\varphi_1(a) + b_2\varphi_2(a) + \ldots + b_n\varphi_n(a) = 0$$

for every element a of L is $b_1 = b_2 = \ldots = b_n = 0$).

(2) *Let $F = \{a \in L : \varphi_i(a) = \varphi_j(a)$ for all i, j in $\{1, 2, \ldots, n\}\}$. Then F is a subfield of L ($\flat$) and $[L : F] \geq n$.*

(Suggestion: For (1), assume not, BWOC ("by way of contradiction"). Then among the equations

$$b_1\varphi_1 + b_2\varphi_2 + \ldots + b_n\varphi_n = 0 \ , \tag{6-1}$$

with not all the b_i's zero, there is (at least) one in which the fewest number of b_i's are zero. We may assume that (6-1) is this equation and that all the b_i's in it are nonzero. Take an element c of L for which $\varphi_1(c) \neq \varphi_n(c)$. Replace a with ca in (6-1), and multiply (6-1) by $\varphi_1(c)$; then subtract the two resulting equations.

For (2), assume BWOC that $[L : F] = m < n$; take an F-basis $a_1, a_2, \ldots, a_m$ of L. Then the system of m linear equations in n variables

$$\sum_{j=1}^{n} \varphi_j(a_i)x_j = 0 \ , \qquad i = 1, 2, \ldots, m$$

has a solution in L that is not all zeros, say $x_1 = b_1$, $x_2 = b_2$, ..., $x_n = b_n$. Show that this contradicts (1).)

6.10 Proposition $*$**.** *Let G be a finite group of automorphisms of the field L. Then the set $F = \{a \in L : \varphi(a) = a \text{ for all } \varphi \text{ in } G\}$ is a subfield of L, and $[L : F] = |G|$.*

(Suggestion: Assume $[L : F] > |G| = n$; take elements $a_1, a_2, \ldots, a_n, a_{n+1}$ of L that are linearly independent over F. This time, the system of n linear equations in $n + 1$ variables

$$\sum_{j=1}^{n+1} \varphi_i(a_j)x_j = 0 \ , \qquad i = 1, 2, \ldots, n$$

has a solution in L that is not all zeros, but as few as possible are nonzero; say $x_1 = b_1 \neq 0, \ldots,$ $x_r = b_r \neq 0$, and $x_{r+1} = \ldots = x_{n+1} = 0$ is such a solution. Applying an element φ of G to

$$\sum_{j=1}^{r} \varphi_i(a_j)b_j = 0 \ , \qquad i = 1, 2, \ldots, n \tag{6-2}$$

gives a system that can be rearranged into

$$\sum_{j=1}^{r} \varphi_i(a_j)\varphi(b_j) = 0 \ , \qquad i = 1, 2, \ldots, n \tag{6-3}$$

Multiply (6-2) by $\varphi(b_1)$ and (6-3) by b_1, and subtract to reach a contradiction.)

6.11 Corollary. *If L/K is an algebraic extension and H is a finite subgroup of* $\mathrm{Gal}(L/K)$*, then $\mathcal{G}(\mathcal{F}(H)) = H$ and $[L : \mathcal{F}(H)] = |H|$.*

The rest of this section and the next describe the additional hypotheses needed to improve the connection between subfields of L containing K and subgroups of $\mathrm{Gal}(L/K)$. These conditions are "normality," to assure that all K-monomorphisms from L into something possibly larger are in fact into L, and "separability," to assure that there are enough different K-automorphisms of L.

6.12 Discussion. A field E is called *algebraically closed* iff any nonconstant polynomial in $E[x]$ factors into linear factors in $E[x]$, or in other words iff E has no proper algebraic extensions. Starting with any field K and repeating the process in the proof of Proposition 4.9 a possibly infinite number of times, we can reach an algebraic extension $\overline{K}$ of K that is algebraically closed. It can be shown, using Corollary 6.14 to the Extension Lemma below, that any two algebraically closed algebraic extensions of K are isomorphic, so we call $\overline{K}$ the *algebraic closure* of K. Since $\mathbb{C}$ is algebraically closed (a fact that requires results from analysis to prove), the field of all algebraic numbers is the algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$. We will use the fact that if L/K is any algebraic extension, then the algebraic closure $\overline{L}$ of L is also an algebraic closure of K, so it can also be denoted $\overline{K}$, and we may assume $L \subseteq \overline{K}$.

6.13 Extension Lemma. *Let $\varphi : K \to E$ be a monomorphism of fields, and a be an element of an extension field of K such that a is algebraic over K. Applying φ to each of the coefficients of $\mathrm{Irr}(a, K)$ yields a polynomial $p(x)$ in $E[x]$. If there is a root b of $p(x)$ in E, then there is a unique monomorphism $\psi : K(a) \to E$ for which $\psi(k) = \varphi(k)$ for each element k of K (i.e., ψ "extends" φ) and $\psi(a) = b$.*

(Suggestion: $p(x)$ is irreducible in $\varphi(K)[x]$, and $K[x]/(\mathrm{Irr}(a, K)K[x]) \cong \varphi(K)[x]/(p(x)\varphi(K)[x])$ by a map that agrees with φ on K and takes the coset $x + (\mathrm{Irr}(a, K)K[x]$ to the coset $x + (p(x)\varphi(K)[x])$.)

6.14 Corollary ♭. *If L, E are extensions of the field K with L/K algebraic and E algebraically closed, then there is a K-monomorphism of L into E. If the elements a of L and b of E have the same minimal polynomial over K, then the K-monomorphism can be chosen to take a to b.*

Proof sketch. Starting with the K-monomorphism $\varphi : K \to E$ given by $\varphi(k) = k$ for every k in K, and with any element a of L (if one is given, start with that one), note the image of $\mathrm{Irr}(a, K)$ has a root b (if one is given, that one) in the algebraically closed field E; and by the Extension Lemma there is an extension of φ, which we will still denote φ, which is a K-monomorphism $K(a) \to E$ and for which $\varphi(a) = b$. Now if a_1 is an element of $L - K(a)$, then the image under φ of $\mathrm{Irr}(a_1, K(a))$ has a root b_1 in E, and we can extend φ to a K-monomorphism $\varphi : K(a, a_1) \to E$ for which $\varphi(a_1) = b_1$. Continuing in this way, possibly an infinite number of times, until all the elements of L are assigned images in E, we get a K-monomorphism $L \to E$. □

6.15 Definition. Let L/K be a field extension and $f(x) \in K[x]$. Then we say $f(x)$ *splits over* L iff $f(x)$ factors into <u>linear</u> factors in $L[x]$. In this case we can write the factorization in the form

$$f(x) = a(x - r_1)(x - r_2) \cdots (x - r_n)$$

where a is the leading coefficient of $f(x)$ (and hence an element of K), $n = \deg f(x)$, and $r_1, r_2, \ldots, r_n$ are the roots of $f(x)$ in L (not necessarily distinct). If in addition L is generated over K by $r_1, r_2, \ldots, r_n$, then L is called the *splitting field of $f(x)$ over K*. (The use of "the" is again justified by Corollary 6.14 to the Extension Lemma: any two splitting fields of $f(x)$ are K-isomorphic.) Note that a polynomial in $K[x]$ splits over L iff all its irreducible factors in $K[x]$ split over L.

6.16 Examples. (1) $\mathbb{Q}(\sqrt{3})$ is the splitting field of $x^2 - 3$ over $\mathbb{Q}$.

(2) $\mathbb{Q}(\sqrt[4]{2})$ contains a root of x^4-2, but it is not the splitting field of this polynomial over $\mathbb{Q}$, because it does not contain the two imaginary roots, $\pm i\sqrt[4]{2}$. The splitting field is $\mathbb{Q}(\sqrt[4]{2}, i)$.

(3) To say that the field E is algebraically closed is to say that every polynomial in $E[x]$ splits over E.

6.17 Proposition. *Let L/K be an algebraic field extension. Then TFAE:*

(1) *For any extension field E of L and any K-homomorphism $\varphi : L \to E$, $\varphi(L) \subseteq L$.*

(1′) *For any extension field E of L and any K-homomorphism $\varphi : L \to E$, $\varphi(L) = L$.*

(2) *If $p(x)$ is an irreducible polynomial in $K[x]$ having a root in L, then $p(x)$ splits over L.*

(3) *There is a family of irreducible polynomials $\{p_\lambda(x)\}_{\lambda\in\Lambda}$ in $K[x]$ all of which split over L, and L is generated over K by the roots of the $p_\lambda(x)$'s.*

(3′) *There is a family of polynomials $\{f_\lambda(x)\}_{\lambda\in\Lambda}$ in $K[x]$ all of which split over L, and L is generated over K by the roots of the $f_\lambda(x)$'s.*

If L/K is a finite extension, then these conditions are also equivalent to:

(3″) *L is the splitting field of some (not necessarily irreducible) polynomial in K.*

6.18 Definition. An algebraic field extension L/K satisfying the equivalent conditions of Proposition 6.17 is called a *normal* (or "pseudo-Galois") extension.

6.19 Remarks ♭. (1) If $[L:K] = 2$, then L is a normal extension of K, because if $a \in L - K$, then the second root of $\mathrm{Irr}(a, K)$ is $-a - b$, where b is the coefficient of x in $\mathrm{Irr}(a, K)$, so that $b \in K$ and $-a - b \in L$.

(2) Let L/K be an algebraic field extension, say $L = K(a_1, a_2, \dots)$, and let $\overline{K}$ be an algebraic closure of K containing L. Let L_N be the subfield of $\overline{K}$ generated over K by all the roots in $\overline{K}$ of the polynomials $\mathrm{Irr}(a_i, K)$. Then L_N is the smallest normal extension of K that contains L; it is called the *normal closure* of L over K. If L/K is finite, then so is L_N/K.

6.20 Corollary. Let L/K be a normal algebraic extension and $a, b \in L$ such that $\mathrm{Irr}(a, K) = \mathrm{Irr}(b, K)$. Then there is an element φ of $\mathrm{Gal}(L/K)$ for which $\varphi(a) = b$. (In other words, if L/K is normal and $p(x)$ is an irreducible element of $K[x]$, then $\mathrm{Gal}(L/K)$ "acts transitively" on the set of roots of $p(x)$ in L.

7. Separable Extensions.

The second property needed to enhance the connection between the set of subfields of L containing K and the set of subgroups of $\mathrm{Gal}(L/K)$ is separability. In this section we discuss this property, and in the next we see that, in many familiar situations, all algebraic field extensions

are separable. (That is why Galois never mentioned the concept explicitly.) But there are many situations in which nonseparable extensions do arise.

Recall that the *multiplicity* of a root r of a polynomial $f(x)$ is the largest integer m for which $(x-r)^m$ divides $f(x)$. If $m > 1$, then r is called a *repeated root* of $f(x)$.

7.1 Definition. Let L/K be an algebraic extension. An element a of L is called *separable over* K iff $\mathrm{Irr}(a, K)$ has no repeated roots, i.e., has $\deg(\mathrm{Irr}(a, K))$ distinct roots in the algebraic closure of K. If L is generated over K by elements separable over K, then L/K is a *separable extension.*

7.2 Proposition. *Let L/K be an algebraic extension.*

(1) *Let $a \in L$. Then a is separable over K iff there are* $\deg(\mathrm{Irr}(a, K))$ *distinct K-monomorphisms of $K(a)$ into the algebraic closure $\overline{K}$ of K.*

(2) *Let F be a subfield of L containing K and $a \in L$. If a is separable over K, then a is separable over F.*

(3) *Suppose L/K is a finite extension. Then L/K is separable iff there are $[L : K]$ different K-monomorphisms of L into $\overline{K}$.*

(4) *Let F be a subfield of L containing K. Then L/K is separable iff L/F and F/K are both separable. In particular, if L/K is separable, then each element of L is separable over K (and of course conversely).*

7.3 Warning. The statement corresponding to the first part of Proposition 7.2(4) with "normal" in place of "separable" is false. If L/K is normal, then L/F is normal (proof?), but F/K need not be normal (example?). And if L/F and F/K are both normal, L/K need not be normal (example?). (Suggestion: For both these examples, look at subfields of $\mathbb{Q}(\sqrt[4]{2}, i)$.)

7.4 Corollary. *Let L/K be a normal algebraic extension and $a \in L$. If a is separable over K and $a \in \mathcal{F}(\mathrm{Gal}(L/K))$, then $a \in K$.*

7.5 Lemma. *Let L be a field and G be a finite subgroup of the group $L - 0$ under multiplication. Then G is cyclic.*

(Suggestion: Let $|G| = q$. If G is not cyclic, then it follows from the Fundamental Theorem on Finite Abelian Groups (Saracino, Theorem 14.2, page 131) that there is an integer $e < q$ for which $g^e = 1$ for each g in G. Count the roots of $x^e - 1$ in L.)

7.6 Primitive Element Theorem $*$**.** *Let L/K be a finite separable extension. Then there is an element a of L for which $L = K(a)$.*

(Suggestion: For K finite, use Lemma 7.5. For K infinite, it suffices by induction to assume $L = K(b,c)$. Let $b_1, \ldots, b_n$ be the roots of $\mathrm{Irr}(b,K)$ in $\overline{K} \supseteq L$ and $c_1, \ldots, c_m$ the roots of $\mathrm{Irr}(c,K)$ in $\overline{K}$; and take r in K different from all the elements $(b_i - b_j)/(c_k - c_l)$ of $\overline{K}$. Then the element $a = b + rc$ of $K(b,c) = L$ has $[K(b,c) : K]$ different images under K-monomorphisms of $K(b,c)$ into $\overline{K}$.)

8. Characteristic of a Ring and Perfect Fields.

8.1 Definition. For any ring R, we can regard integers as elements of R by regarding the positive integer n as the sum of n copies of the unity of R. This gives a ring homomorphism $\mathbb{Z} \to R$; the nonnegative generator of the kernel is called the *characteristic* of R. (Thus, if no sum of copies of 1_R is 0_R, i.e., if the homomorphism is injective, then $\mathrm{char}(R) = 0$; otherwise, $\mathrm{char}(R)$ is the smallest n for which $1_R + 1_R + \ldots + 1_R (n \text{ terms}) = 0_R$.) The image of $\mathbb{Z}$ in R under this homomorphism is called the *prime subring* of R.

8.2 Proposition. *If R is a domain and* $\mathrm{char}(R) \neq 0$*, then* $\mathrm{char}(R)$ *is a prime number p, the prime subring of R is (isomorphic to) the field $\mathbb{Z}/p\mathbb{Z}$, and the function $\varphi : R \to R$, given by $\varphi(a) = a^p$ for all a in R, is a ring monomorphism, called the* Frobenius homomorphism. *In particular, an element of R has at most one p-th root.*

(Suggestion: Use the fact that p divides the binomial coefficient $\binom{p}{j}$ for $j = 1, 2, \ldots, p-1$.)

8.3 Definition. If K is a field, then the smallest subfield contained in K is call the *prime subfield* of K. If $\mathrm{char}(K) = p > 0$, then by Proposition 8.2 its prime subfield is (isomorphic to) $\mathbb{Z}/p\mathbb{Z}$. If $\mathrm{char}(K) = 0$, then the prime subring of K is (an isomorphic copy of) $\mathbb{Z}$; its field of fractions, (an isomorphic copy of) $\mathbb{Q}$, is the prime subfield of K.

8.4 Definition. For any ring R, the *(formal) derivative* of a polynomial $f(x)$ in $R[x]$ is the element $D(f(x))$ of $R[x]$ given by

$$D(f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \ldots) = f_1 + 2f_2 x + 3f_3 x^2 + \ldots \ .$$

Note that this "differentiation", a function from $R[x]$ to itself, is defined purely algebraically; there is no limit process involved. So the proof of the following result does not involve limits.

8.5 Facts ♭. For polynomials $f(x), g(x)$ over any ring and an element a of that ring, we have

(1) $D(f \pm g) = D(f) \pm D(g)$

(2) $D(fg) = f(D(g)) + (D(f))g$

(3) $D((x-a)^n) = n(x-a)^{n-1}$

8.6 Proposition. *Let K be a field and $f(x) \in K[x]$.*

(1) *An element r of K is a repeated root of $f(x)$ iff r is a root of both f and $D(f)$.*

(2) *f has a repeated root in the algebraic closure $\overline{K}$ of K iff f and $D(f)$ have a common factor in $K[x]$.*

(3) *An element a of $\overline{K}$ is not separable over K iff $D(\mathrm{Irr}(a,K)) = 0$. This is possible only if $\mathrm{char}(K) = p \neq 0$ and $\mathrm{Irr}(a,K) = f(x^p)$ for some (irreducible) $f(x)$ in $K[x]$.*

8.7 Definition. A field K is called *perfect* iff every algebraic extension of K is separable.

8.8 Proposition. *A field of characteristic 0 is perfect.*

In particular, $\mathbb{Q}$ and $\mathbb{R}$ are perfect. So is $\mathbb{C}$, but that is not very interesting, because $\mathbb{C}$ has no proper algebraic extensions.

8.9 Proposition. *A finite field is perfect. In fact, every algebraic extension of a finite field is both separable and normal.*

(Suggestion: It is enough to show that a finite extension L of a finite field K is separable and normal over K. But then L is itself finite, and all its elements are roots of $x^{|L|} - x$.)

8.10 Scholium ♯. *Let K be a field of characteristic $p \neq 0$, and let L be an algebraic extension of K.*

(1) *The set K_s of elements of L separable over K is a subfield of L (called the* separable algebraic closure *of K in L). No element of L not in K_s is separable over K_s; i.e., L/K_s is a* purely inseparable *extension.*

(2) *If a in L is not separable over K, then $\mathrm{Irr}(a,K) = g(x^{p^e})$ for some largest integer e, and then $g(x) = \mathrm{Irr}(a^{p^e}, K)$, a^{p^e} is separable over K, and a is the only p^e-th root of a^{p^e} in $\overline{K}$.*

9. The Fundamental Theorem of Galois Theory.

9.1 Definition. An algebraic extension of fields is called *Galois* iff it is both normal and separable.

9.2 Remark. Let L/K be a separable algebraic field extension, and let L_N be the normal closure of L over K, as in Remark 6.19. Then since L_N is obtained by adjoining more roots of the same irreducible polynomials over K, we have that L_N/K is also separable and hence is a Galois extension.

Please imagine a fanfare sounding: The following two theorems are the climax of the course. What follows them is denoument.

9.3 FTGT, general case. *Let L/K be a Galois field extension, and let F be a subfield of L containing K.*

(1) $\mathcal{F}(\mathcal{G}(F)) = F$.

(2) *If F/K is a normal extension, then restriction of domain from L to F gives a group epimorphism from* $\mathrm{Gal}(L/K)$ *to* $\mathrm{Gal}(F/K)$ *with kernel $\mathcal{G}(F)$, so that $\mathcal{G}(F)$ is a normal subgroup of* $\mathrm{Gal}(L/K)$ *and* $\mathrm{Gal}(F/K) \cong \mathrm{Gal}(L/K)/\mathcal{G}(F)$.

(3) *Conversely, if $\mathcal{G}(F)$ is a normal subgroup of* $\mathrm{Gal}(L/K)$*, then F/K is normal.*

(Suggestion for (3): Suppose BWOC that b, c in L have the same minimal polynomial, but $b \in F$ and $c \notin F$. Take ψ in $\mathcal{G}(F)$ such that $\psi(c) \neq c$, and take φ in $\mathrm{Gal}(L/K)$ such that $\varphi(b) = c$.)

9.2 Notation. Let G be a group and H a subgroup of G. The number of left (or right) cosets of H in G is called the *index of H in G*, and denoted $index(G, H)$.

We have chosen to change the notation for index of a subgroup from Saracino's because his looks too much like our notation for the degree of a field extension (although, as we shall see, there is a strong connnection between the two). Recall that by Lagrange's theorem, $|H| \cdot index(G, H) = |G|$.

9.4 FTGT, finite case. *Let L/K be a finite Galois extension.*

(1) $|\mathrm{Gal}(L/K)| = [L : K]$. *If F is a subfield of L containing K, then $|\mathcal{G}(F)| = [L : F]$ and $[F : K] = index(\mathrm{Gal}(L/K), \mathcal{G}(F))$.*

(2) *For any subgroup H of* $\mathrm{Gal}(L/K)$*, $\mathcal{G}(\mathcal{F}(H)) = H$. Thus, $\mathcal{F}$ and $\mathcal{G}$ are inverse 1-1 correspondences between the set of subgroups of* $\mathrm{Gal}(L/K)$ *and the set of subfields of L containing K.*

9.5 Remark ♯. For (1) of this theorem even to make sense, we need L/K to be a finite extension. We have placed part (2) in this theorem, rather than in Theorem 9.4, because if L/K is an infinite Galois extension, then part (2) does not hold. See Paul McCarthy, *Algebraic Extensions of Fields* (Blaisdell, Waltham, MA, 1966), Chapter 2, Sections 10–12, for a discussion of the case where L/K is infinite. (There is a new edition of this book, published by Dover.)

9.6 General Example. If L is a finite field, say of characteristic p, then the prime subfield of L is (isomorphic to) $\mathbb{Z}/p\mathbb{Z}$; and if $[L : K] = n$, then $|L| = p^n$. Since L is the splitting field of $x^{|L|} - x$ over $\mathbb{Z}/p\mathbb{Z}$, L is uniquely determined up to isomorphism by its number of elements, which

is often denoted q (a power of its characteristic, as we have just seen); L is called the *Galois field of q elements* (and sometimes denoted $\mathbb{F}_q$). By Proposition 8.9, $L/(\mathbb{Z}/p\mathbb{Z})$ is a Galois extension. The Frobenius homomorphism $\varphi : L \to L$, given by $\varphi(a) = a^p$ for all a in L, is an element of $\mathrm{Gal}(L/(\mathbb{Z}/p\mathbb{Z}))$, and so are its powers: $\varphi^m(a) = a^{p^m}$. Since the order of φ is easily seen to be $n = [L : (\mathbb{Z}/p\mathbb{Z})] = |\mathrm{Gal}(L/(\mathbb{Z}/p\mathbb{Z}))|$, we conclude that $\mathrm{Gal}(L/(\mathbb{Z}/p\mathbb{Z}))$ is cyclic, generated by the Frobenius homomorphism. Since the subgroups of $\mathrm{Gal}(L/(\mathbb{Z}/p\mathbb{Z}))$ are all cyclic, namely the subgroups $\langle\varphi^d\rangle$ as d varies over the divisors of n, there are exactly as many fields between L and $\mathbb{Z}/p\mathbb{Z}$ as n has divisors.

9.7 Example. Let $L = \mathbb{Q}(\sqrt[4]{2}, i)$ be the splitting field of $x^4 - 2$ over $\mathbb{Q}$. Then $\mathrm{Gal}(L/K) \cong D_4$ by Example 6.6(3). Since we know the family of subgroups H of D_4, we can describe the family of subfields of L containing $\mathbb{Q}$: Recall that in the notation of that example, φ induces a cyclic permutation $(\sqrt[4]{2}, i\sqrt[4]{2}, -\sqrt[4]{2}, -i\sqrt[4]{2})$ on the roots of $x^4 - 2$, and γ induces the 2-cycle $(i\sqrt[4]{2}, -i\sqrt[4]{2})$.

$H = \langle\varphi\rangle$: Since $index(D_4, \langle\varphi\rangle) = 2$, we have $[\mathcal{F}(\langle\varphi\rangle) : \mathbb{Q}] = 2$; so all we need to generate $F(\langle\varphi\rangle)$ over $\mathbb{Q}$ is one element of $F(\langle\varphi\rangle) - \mathbb{Q}$. Two elements of L that φ leaves fixed are the sum $\sqrt[4]{2} + i\sqrt[4]{2} - \sqrt[4]{2} - i\sqrt[4]{2} = 0$ and the product $(\sqrt[4]{2}()i\sqrt[4]{2}) - \sqrt[4]{2})(-i\sqrt[4]{2}) = -2$, but both of these are in $\mathbb{Q}$. However, $\varphi(i) = \varphi(i\sqrt[4]{2}/\sqrt[4]{2}) = -\sqrt[4]{2}/i\sqrt[4]{2} = i$, which is not in $\mathbb{Q}$, so $\mathcal{F}(\langle\varphi\rangle) = \mathbb{Q}(i)$.

$H = \langle\varphi^2\rangle$: Since $index(D_4, \langle\varphi^2\rangle) = 4$, we have $[\mathcal{F}(\langle\varphi^2\rangle) : \mathbb{Q}] = 4$; and since $\langle\varphi^2\rangle \subset \langle\varphi\rangle$, we have $\mathbb{Q}(i) = \mathcal{F}(\langle\varphi\rangle) \subset \mathcal{F}(\langle\varphi^2\rangle)$, and by comparison of degrees and indices, $[\mathcal{F}(\langle\varphi^2\rangle) : \mathbb{Q}(i)] = 2$. Also, $\varphi^2(\sqrt{2}) = (\varphi^2(\sqrt[4]{2}))^2 = (-\sqrt[4]{2})^2 = \sqrt{2} \notin \mathbb{Q}(i)$, so $\mathcal{F}(\langle\varphi^2\rangle) = \mathbb{Q}(i, \sqrt{2})$. Since $\mathbb{Q}$ is perfect, $\mathbb{Q}(i, \sqrt{2})/\mathbb{Q}$ is separable, so by the Primitive Element Theorem there is an element a for which $\mathbb{Q}(i, \sqrt{2}) = \mathbb{Q}(a)$. To find such an element a, it is enough to find an element having $[\mathbb{Q}(i, \sqrt{2}) : \mathbb{Q}] = 4$ different images under elements of $\mathrm{Gal}(L/K)$, and a natural first guess is $a = i + \sqrt{2}$. Since $i + \sqrt{2}$, $i - \sqrt{2}$, $-i + \sqrt{2}$, and $-i - \sqrt{2}$ are all images of $i + \sqrt{2}$, we have $\mathcal{F}(\langle\varphi^2\rangle) = \mathbb{Q}(i + \sqrt{2})$.

$H = \langle\gamma\rangle$: Since γ is the restriction to L of complex conjugation, it leaves all real numbers fixed: $\mathcal{F}(\langle\gamma\rangle) = \mathbb{Q}(\sqrt[4]{2})$.

$H = \langle\varphi^2\gamma\rangle$: Since γ reverses the top and bottom vertices of the square $\sqrt[4]{2}$, $-i\sqrt[4]{2}$, $-\sqrt[4]{2}$, and $-i\sqrt[4]{2}$, $\varphi^2\gamma$ reverses the left and right ones: $\mathcal{F}(\langle\varphi^2\gamma\rangle) = \mathbb{Q}(i\sqrt[4]{2})$.

$H = \langle\varphi\gamma\rangle$: $\varphi\gamma$ reverses the right and top vertices (and the left and bottom ones), so it leaves their sum fixed: $\mathcal{F}(\varphi\gamma) = \mathbb{Q}(\sqrt[4]{2}(1 + i))$.

$H = \langle\varphi^2, \gamma\rangle$: $\mathcal{F}(\langle\varphi^2, \gamma\rangle) = \mathcal{F}(\langle\varphi^2\rangle) \cap \mathcal{F}(\langle\gamma\rangle) = \mathbb{Q}(i, \sqrt{2}) \cap \mathbb{Q}(\sqrt[4]{2}) = \mathbb{Q}(\sqrt{2})$.

For the other choices of H, the determination of $\mathcal{F}(H)$ is left to you.

Example 9.8. We want to find an element a of R that is not constructible but for which $[\mathbb{Q}(a) : \mathbb{Q}]$ is a power of 2. A chain of fields from $\mathbb{Q}$ up to a field containing $\mathbb{Q}(a)$ with "links" of degree 2 would correspond, in the Galois group G of the normal closure $L = \mathbb{Q}(a)_N$ of $\mathbb{Q}(a)$ over $\mathbb{Q}$, to a chain of subgroups from G down to a subroup contained in $\mathcal{G}(\mathbb{Q}(a))$ with links of index 2. So we want a Galois group over $\mathbb{Q}$ which has a subgroup H of index a power of 2 but no chain of subgroups from H to G with links of index 2. Now in the symmetric group S_4 on four letters (say $\{1, 2, 3, 4\}$), there is no chain of subgroups of with links of index 2 down to a subgroup contained in a cyclic group H generated by a 3-cycle, say $(1, 2, 3)$: If the chain that does not end at H, it would have to end at $\{e\}$ and hence we would have that $|S_4| = 24$ is a power of 2), a contradiction. So such a chain, if it exists, must end at H itself. Also, its second link (from the top) must be the alternating group A_4, i.e., the group of even permutations. For, if a subgroup of S_4 is of index 2, then it must be normal with abelian factor group, so it must be A_4 by Proposition 10.8 below. Since $|A_4| = 12$, we must show that $\langle(1, 2, 3)\rangle$ and any other even permutation generate all of A_4 and not a subgroup of order 6. This can be done by "exhaustion" (an apt term). If we can find a normal extension L of $\mathbb{Q}$ for which $\mathrm{Gal}(L/\mathbb{Q}) \cong S_4$, then the fixed field of the subgroup corresponding under this isomorphism to a cyclic subgroup generated by a 3-cycle is $\mathbb{Q}(a)$ for some element a, by the Primitive Element Theorem. And by what we have just seen, even though $[\mathbb{Q}(a) : \mathbb{Q}] = 4$, that element a is not constructible.

Now any S_n is generated by two elements, the 2-cycle $(1, 2)$ and the n-cycle $(1, 2, \ldots, n)$. So QQQ

10. Solvable Groups.

The source of the name "solvable" will become clear in the next section.

10.1 Definition. Let G be a group. Then:

(1) G is called *solvable* (in the United Kingdom, "soluble") iff it has an *abelian tower,* i.e., a chain of subgroups

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \ldots \supseteq G_{n-1} \supseteq G_n = \{e\}$$

such that, for each j in $\{1, 2, \ldots, n\}$, G_j is a normal subgroup of G_{j-1} (though not necessarily normal in G) and the factor group G_{j-1}/G_j is abelian.

(2) The *commutator subgroup* $G^{(1)}$ of G is the subgroup generated by the set of all "commutators", i.e., elements of G of the form $aba^{-1}b^{-1}$ for some a, b in G. (Warning: It happens to be the case that the inverse of a commutator is again a commutator. But it is not always true that

the product of two commutators is a commutator; so not every element of the commutator subgroup is necessarily a commutator.)

(3) The *upper central series* of G is the chain of subgroups

$$G^{(0)} \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \dots \ ,$$

where $G^{(0)} = G$ and for $n \geq 1$, $G^{(n)} = (G^{(n-1)})^{(1)}$ (i.e., each group in the chain is the commutator subgroup of the one before).

10.2 Lemma. *Let G be a group, G_1, G_2, H be subgroups of G, and N be a normal subgroup of G. Suppose G_1 is a normal subgroup of G_2. Then:*

(1) $G_1 \cap H$ is normal in $G_2 \cap H$, and the factor group $(G_2 \cap H)/(G_1 \cap H)$ is isomorphic to a subgroup of G_2/G_1.

(2) $HN = \{hn : h \in H,\ n \in N\}$ is a subgroup of G.

(3) The subgroup G_1N is a normal subgroup of G_2N, and the function

$$\psi : G_2/G_1 \to (G_2N)/(G_1N)$$

defined by $\psi(gG_1) = g(G_1N)$ is a well-defined, surjective group homomorphism.

10.3 Corollary. *If a group G is solvable, then any subgroup of G is solvable and any factor group of G is solvable.*

10.4 Corollary. *Let G be a group and*

$$\{e\} = G_0 \subseteq G_1 \subseteq G_2 \subseteq \dots G_n = G$$

be a chain of subgroups of G such that G_{j-1} is a normal subgroup of G_j for each j in $\{1, 2, \dots, n\}$. Then G is solvable iff all of the factor groups G_j/G_{j-1} are solvable.

10.5 Lemma. *Let S be a subset of a group G, and suppose that, for all s in S and g in G, the element gsg^{-1} is again in S. Then the subgroup $\langle S \rangle$ generated by S is normal in G.*

(Suggestion: Use the fact that any element of $\langle S \rangle$ is a product of elements of S and inverses of elements of S.)

10.6 Proposition. *Let G be a group. Then the commutator subgroup $G^{(1)}$ is normal in G and $G/G^{(1)}$. Moreover, if N is any normal subgroup of G for which G/N is abelian, then $G^{(1)} \subseteq N$.*

10.7 Corollary. *A group G is solvable iff its upper central series reaches $\{e\}$, i.e., iff $G^{(n)} = \{e\}$ for some positive integer n.*

10.8 Proposition. *For $m \geq 5$, the symmetric group S_m on m letters is not solvable. In fact, for $n \geq 1$, $(S_m)^{(n)} = A_m$, the alternating group on m letters (i.e., the set of even permutations).*

(Suggestion: For any $m \geq 3$, A_m is generated by the 3-cycles (Saracino, Exercise 8.17, page 78). If $m \geq 5$, given the 3-cycle (a, b, c), consider the commutator

$$(a, b, d)(a, c, e)(a, b, d)^{-1}(a, c, e)^{-1} \ .)$$

11. Solvability by Radicals.

In this section, the fields are all assumed to have characteristic 0, so that we can restrict the radical sign to its usual meaning, namely, $\sqrt[m]{a}$ denotes any of the roots of the polynomial $x^m - a$. (It is usual to try to specify one of these roots as the meaning of this symbol; e.g., if a is a real number and m is odd, then this symbol is usually assumed to mean the unique real root of this polynomial; and if a is a positive real number and m is even, then this symbol is usually assumed to mean the positive real root of the polynomial. But in the case of a general field, there is no natural way to select one root of this polynomial. We will discuss this ambiguity further below.) If we were to allow fields of positive characteristic p, then we would need a new form of radical, to denote a root of the polynomial $x^p - x - a$, in order to complete the theory. See Lang, *Algebra,* (Addison-Wesley, Reading, MA, 1965), Chapter VII for a discussion of the case of positive characteristic.

The term "solvability by radicals" could conceivably refer to any of (at least) three related concepts:

(1) We will say that a polynomial $f(x)$ with coefficients in a field K of characteristic 0 is solvable by radicals if we can express all the roots of $f(x)$ in terms of the field operations and the taking of m-th roots for various positive integers m.

But a stronger concept would be:

(2) For a given field K and a given positive integer n, there is a single formula, in terms of the field operations and m-th roots for various m, that expresses the roots of any polynomial over K of degree n in terms of the coefficients of the polynomial.

The familiar quadratic formula, stating that the roots of a polynomial $f(x) = ax^2 + bx + c$ of degree 2 (so $a \neq 0$) are given by $x = (-b \pm \sqrt{b^2 - 4ac})/(2a)$, is such a formula. (In fact, it works over any field of characteristic not equal to 2). Similarly, there are formulas for polynomials of

degrees 3 and 4, at least for fields of characteristics not dividing 3! and 4! respectively, discovered during the Italian Renaissance but too complicated for everyday use. (See, for example, the book of Standard Mathematical Tables from the Chemical Rubber Company.) So it is natural to ask whether there are similar formulas for polynomials of degree 5 or higher.

Since any monic polynomial $f(x)$ in $K[x]$ of degree n can be written in the form

$$f(x) = (x - r_1)(x - r_2)\dots(x - r_n)$$

where the r_i's are the roots of $f(x)$ (not necessarily distinct), and then multiplied out to give

$$f(x) = x^n - (r_1 + r_2 + \dots + r_n)x^{n-1} + \dots + (-1)^n(r_1 r_2 \dots r_n) \ ,$$

concept (2) amounts to saying that one can recover, by a formula in the field operations and m-th roots, the numbers $r_1, r_2, \dots, r_n$ from the coefficients $(r_1 + r_2 + \dots + r_n), \dots, (r_1 r_2 \dots r_n)$, and that this formula works whenever these coefficients are in K. Let us look more closely at these coefficients, since it is not obvious what the ones between the first and the last (the ones we have actually written) should be.

11.1 Definition. The *elementary symmetric polynomials* in the indeterminates $x_1, x_2, \dots, x_n$ are $s_1(x_1, x_2, \dots, x_n)$, $s_2(x_1, x_2, \dots, x_n)$, $\dots$, $s(x_1, x_2, \dots, x_n)$, where s_j is the sum of all products of j <u>different</u> x_i's.

So, for example, if $n = 3$, then

$$\begin{aligned} s_1 &= x_1 + x_2 + x_3 \ , \\ s_2 &= x_1x_2 + x_1x_3 + x_2x_3 \ , \\ s_3 &= x_1x_2x_3 \ . \end{aligned}$$

In this notation we can write

$$f(x) = x^n - s_1(r_1, r_2, \dots, r_n)x^{n-1} + s_2(r_1, r_2, \dots, r_n)x^{n-2} + \dots + (-1)^n s_n(r_1, r_2, \dots, r_n) \ ;$$

and concept (2) amounts to saying that there is a formula for which, if $r_1, r_2, \dots, r_n \in \overline{K}$ such that $s_1(r_1, r_2, \dots, r_n), s_2(r_1, r_2, \dots, r_n), \dots, s_n(r_1, r_2, \dots, r_n) \in K$, then the formula produces the roots $r_1, r_2, \dots, r_n$ from $s_1(r_1, r_2, \dots, r_n), s_2(r_1, r_2, \dots, r_n), \dots, s_n(r_1, r_2, \dots, r_n)$. This translation leads us to the last, and strongest, possible meaning of "solvability by radicals":

(3) There is a formula, in terms of the field operations and m-th roots, that yields the indeterminates $x_1, x_2, \ldots, x_n$ from $s_1(x_1, x_2, \ldots, x_n), s_2(x_1, x_2, \ldots, x_n), \ldots s_n(x_1, x_2, \ldots, x_n)$.

Concept (3) amounts to saying that there is a formula that works not only on polynomials of degree n over K, but on polynomials of degree n over <u>all</u> extension fields of K, algebraic or transcendental. This is because, if (3) holds, then the same formula can be applied when any elements in any extension field of K are substituted for the x_i's.

11.2 Definition. Let K be a field of characteristic 0.

(1) We will call an extension L of K a *root extension* if $L = K(a)$ where some power of a is in L; and we will call a chain $K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ a *root chain* if each K_j is a root extension of K_{j-1}.

(2) A polynomial $f(x)$ in $K[x]$ is called *solvable by radicals* (over K) iff there is a root chain of fields $K = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ such that $f(x)$ splits over K_n.

The terms "root extension" and "root chain" are nonstandard. The term "radical extension" is sometimes used, but it is too close to the French term "extension radicielle", which is used only in the context of fields of nonzero characteristic.

Let us return to the problem that the meaning of $\sqrt[m]{a}$ is not uniquely determined. One thing is clear: If this symbol is used more than once in a formula, it should be assumed that it means the same element in both cases; and if the other m-th roots of a are needed in the same formula, they should be obtained by multiplying the original choice by m-th roots of unity. But the m-th roots of unity may not be available in K. Some discussions of solvability by radicals avoid this problem by assuming that K contains all the desired roots of unit; but since $\mathbb{Q}$ does not satisfy this hypothesis, we choose to avoid it. As a result, we will require the following lemma:

11.3 Technical Lemma ♭. *Suppose $K = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n$ is a root chain; say for each j in $\{1, \ldots, n\}$, we have $K_j = K_{n-1}(a_j)$ where $(a_j)^{m_j} \in K_{j-1}$. Let L be the smallest field containing both the normal closure N of K_n over K and a primitive m_j-th root of unity for all j in $\{1, \ldots, n\}$. Then:*

(1) L is a normal extension of K, and

(2) there is a chain $K = L_0 \subseteq L_1 \subseteq L_q = L$ of fields from K to L such that each L_i is the splitting field of a polynomial of the form $x^m - a$ where m is one of the positive integers m_j from the given chain and $a \in L_{j-1}$.

Proof sketch. The field L is normal over K because it is the smallest extension of K over which the polynomials $x^{m_j} - 1$, for j in $\{1, \ldots, n\}$, and $\mathrm{Irr}(a, K)$, for a in K_n, all split. We want to show

that there is a root chain involving only the powers m_j from K to L; the full strength of (2) will then follow because L contains all the necessary roots of unity.

To get a root chain from K just to N, we note that if we denote by Φ the (finite) set of all K-monomorphisms from K_n into $\overline{K}$, then $N = K(\bigcup\{\varphi(K_n) : \varphi \in \Phi\})$. If

$$\Phi = \{\varphi_1(= id), \varphi_2, \varphi_3, \dots, \varphi_t\} ,$$

let $N_i = K(\varphi_1(K_n), \dots, \varphi_i(K_n))$; then a root chain from N_{i-1} to N_i is

$$N_{i-1} \subseteq N_{i-1}(\varphi_i(K_1)) \subseteq N_{i-1}(\varphi_i(K_2)) \subseteq \dots \subseteq N_{i-1}(\varphi_i(K_n)) = N_i ,$$

and $N_{i-1}(\varphi_i(K_j) = N_{i-1}(\varphi_i(K_{j-1}))(\varphi_i(a_j)$ where $\varphi_i(a_j)^{m_j} \in N_{i-1}(\varphi_i(K_{j-1}))$. We can string these chains together to get a root chain from K to N. Finally, if ω_j denotes a primitive m_j-th root of unity, then a root chain from N to L is

$$N \subseteq N(\omega_1) \subseteq N(\omega_1, \omega_2) \subseteq \dots \subseteq N(\omega_1, \omega_2, \dots, \omega_n) = L . \quad \square$$

The L constructed in this lemma may be larger than it needs to be, because, for instance, there may be a smaller power of a_j than the m_j-th that lies in K_{j-1}. But it allows us to achieve our present objective, Corollary 11.7(2), with minimal fuss.

11.4 Proposition. *Let K be a field of characteristic 0, $a \in K$, m be a positive integer, and L be the splitting field of $x^m - a$ over K. Then:*

(1) $L = K(\omega, b)$ *where ω is a primitive m-th root of unity and $b^m = a$.*

(2) $K(\omega)$ *is a normal extension of K, and* $\mathrm{Gal}(K(\omega)/K)$ *is isomorphic to a subgroup of $\{q \in \{1, \dots, m-1\} : \gcd(m, q) = 1\}$, which is a group under multiplication mod m. Thus,* $\mathrm{Gal}(K(\omega)/K)$ *is abelian.*

(3) $\mathrm{Gal}(L/K(\omega))$ *is isomorphic to a subgroup of $\mathbb{Z}/m\mathbb{Z}$ under addition, and hence is abelian.*

(4) $\mathrm{Gal}(L/K)$ *is solvable.*

11.5 Definition. Let K be a field and $f(x) \in K[x]$. Then the *Galois group* of $f(x)$, $\mathrm{Gal}(f(x))$, is the Galois group of the splitting field of $f(x)$ over K.

11.6 Exercise. *If $f(x)$ has degree n, then the Galois group of $f(x)$ has order at most $n!$.*

11.7 Corollary. (1) *For K, L as in Technical Lemma 11.3,* $\mathrm{Gal}(L/K)$ *is solvable.*

(2) *Let K be a field and $f(x) \in K[x]$. If $f(x)$ is solvable by radicals, then* $\mathrm{Gal}(f(x))$ *is solvable.*

To see that the third, strongest possible meaning of solvability by radicals cannot work, we take any field K of characteristic 0 and the indeterminates $x_1, x_2, \ldots, x_n$, and consider the field of fractions $K(x_1, x_2, \ldots, x_n)$ of the polynomial ring $K[x_1, x_2, \ldots, x_n]$. Any permutation φ of the set $\{1, 2, \ldots, n\}$ (i.e., any element φ of S_n) gives rise to a K-automorphism of $K(x_1, x_2, \ldots, x_n)$, which we will also denote by φ, given by, for any "rational function" (= quotient of two polynomials) $g(x_1, x_2, \ldots, x_n)$ in $K(x_1, x_2, \ldots, x_n)$,

$$\varphi(g(x_1, x_2, \ldots, x_n) = g(x_{\varphi(1)}, x_{\varphi(2)}, \ldots, x_{\varphi(n)}) \ .$$

It is easy to see that the elementary symmetric polynomials

$$s_1(x_1, x_2, \ldots, x_n),\ s_2(x_1, x_2, \ldots, x_n),\ \ldots,\ s_n(x_1, x_2, \ldots, x_n)$$

are left fixed by any such automorphism, so $K(s_1, s_2, \ldots, s_n)$ is contained in the fixed field F of this group of K-automorphisms. By Corollary 6.11, $[K(x_1, x_2, \ldots, x_n) : F] = |S_n| = n!$, and since $K(x_1, x_2, \ldots, x_n)$ is the splitting field over F of

$$\begin{aligned} f(x) &= (x - x_1)(x - x_2) \cdots (x - x_n) \\ &= x^n - s_1 x^{n-1} + \ldots + (-1)^n s_n \ , \end{aligned}$$

we have that $[K(x_1, x_2, \ldots, x_n) : K(s_1, s_2, \ldots, s_n)] \leq n!$. It follows that every rational function in $x_1, x_2, \ldots, x_n$ that is fixed by all permutations of these indeterminates can be expressed as a rational function in $s_1, s_2, \ldots, s_n$. Moreover:

11.8 Proposition. *For any field K of characteristic 0 and any integer $n \geq 5$, it is not possible to express $x_1, x_2, \ldots, x_n$ in terms of the elementary symmetric polynomials $s_1, s_2, \ldots, s_n$ using only the field operations and m-th roots for various positive integers m.*