VIGRE Algebra Review

These notes are dedicated to the memory of Professor Jack McLaughlin, August 17, 1923 – January 12, 2001, a wonderful teacher, colleague, and friend.

The Moving Finger writes: and, having writ,
Moves on: nor all thy Piety nor Wit
Shall lure it back to cancel half a Line,
Nor all thy Tears wash out a Word of it.

Edward FitzGerald, *The Rubaiyat of Omar Khayyam*

Since Omar Khayyam was an algebraist who studied the problem of solving cubic equations by intersecting conic sections and who gave an algorithm for extracting n th roots, the obvious interpretation of the verse above may not be the correct one. Perhaps[1] Khayyam meant that what one writes when taking the Algebra QR cannot be changed afterwards, no matter how regretful one feels.

In taking the exam it is important to show what you know, and also not to give the impression that you lack understanding. Do not write down something that doesn't really make sense to you in the hope that it will be perceived as a proof. Bluffing is unlikely to get any credit, but might well make a distinctly negative impression on the examiners. You may succeed in getting partial credit by giving the definitions of terms used, to show that you know them, and by stating a theorem or theorems that you think are relevant, to show that you know the theorems, along with a comment to the effect that you believe the results are relevant but do not see how to carry through. You may also get partial credit by doing a special case.

You might, of course, try this one: "I have discovered a truly remarkable proof of this, but the margin is too narrow to contain it." The examiners may or may not be amused.

We shall use the notations $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ for the non-negative integers, the integers, the rational numbers, the real numbers, and the complex numbers, respectively. Of course, $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$.

When a problem consists of a list of assertions, the task intended is proving the assertions.

[1] And then again, perhaps not. There is a great deal of doubt about which of the quatrains attributed to him were written by Khayyam. He does not seem to have been known as a poet at all during his lifetime. The attributions were very largely made long, e.g., two centuries, after his death, and may have been an attempt to place him in a bad light. The quatrains containing the most famous lines are, in fact, among those most suspect. Beyond that, FitzGerald's "translation" is very free, and FitzGerald should get the lion's share (as those who have attempted to share with a lion might testify, this phrase originally meant "all," or "nearly all": the meaning has eroded with time) of the credit for this work.

1. Functions and relations

Functions. A *function* (or *map*, or *mapping*) f from a set A (called the *domain*, or *initial set*) to a set B (called the *target set*, *codomain*, *range*, or *terminal set*) assigns to every element a in A a unique value $f(a)$ in B. A function is usually described by a rule (e.g., double the input, or square the input, etc.) but different rules may give the same function. Our axiom is that two functions f, g are equal provided that they have the same domain A, the same target set B, and for all $a \in A$, $f(a) = g(a)$. Our convention is that there is exactly one function from the empty set $\emptyset$ to any set B (B may be empty as well): it does not need to assign any values. On the other hand, when A is not empty and B is empty there are no functions from A to B.

A function $f : A \to B$ is called *injective* or *one-to-one* if $f(a) = f(a')$ implies that $a = a'$.

The *image* of $f : A \to B$ is the subset of B consisting of all values of f on elements in A, i.e., $\{f(a) : a \in A\}$. Note that in general, the image may be strictly smaller than the target. If the image of the function f is all of B, the function f is called *surjective* or *onto.*

Suppose that A has m elements and B has n elements: in particular, A and B are finite. The number of functions from A to B is then $n \times n \times \cdots \times n$ (m times), or n^m . This is correct even when A or B is empty (with the convention $n^0 = 1$ even if $n = 0$). In this case, there are no one-to-one functions from A to B unless $m \le n$, and then the number of such functions is $n(n-1)(n-2)\cdots\big(n-(m-1)\big)$ (m factors). There are no surjective functions from A to B unless $m \ge n$.

Problem. Show that the number of surjections from a set with m elements to a set with n elements, $m \ge n \ge 1$, can be expressed as

$$\sum_{i=0}^{n-1}(-1)^i\binom{n}{i}(n-i)^m = n^m - \binom{n}{1}(n-1)^m + \binom{n}{2}(n-2)^m - \cdots \pm \binom{n}{n-1}1^m.$$

(For each element u_i in the target set let W_i be the set of maps which do not take on the value u_i. One is trying to measure the complement, in the set of all maps, of the union of the W_i.)

Problem Let $S(m, n)$ denote the number of surjections from a set with m elements to a set with n elements, $m \ge n \ge 1$. Without any reference to the problem above, prove that $S(m, n) = \sum_{i=1}^{m-n+1} \binom{m}{i} S(m-i, n-1)$.

A function is called *bijective* or a *one-to-one correspondence* if it is both one-to-one and onto, i.e., both injective and surjective. When A and B have the same finite number of elements, a function is one-to-one iff it is onto iff it is bijective. But the two sets must be finite, with the same number of elements, before one can make this assertion. The number of bijections between two finite n element sets is $n!$.

A bijection from a set A to itself is called a *permutation* or *set automorphism* of A. Again, the number of permutations of a finite n element set is $n!$.

If $f : A \to B$ and $g : B \to C$ are functions the *composition,* denoted $g \circ f$ or, simply, gf, of the functions f and g is the function from A to C defined by the rule $(g \circ f)(a) = g(f(a))$. If $h : C \to D$ is another function, then it is easy to verify that $h \circ (g \circ f) = (h \circ g) \circ f$: this is called the *associativity* of composition. Even when f and g are both functions from A to A, so that both compositions are functions from A to A, one does not usually have that $f \circ g = g \circ f$. (Example: let f and g be constant functions taking on different values.)

The identity function $\mathbf{1}_A$ on set A is the function $A \to A$ whose value on each element $a \in A$ is a. If $g : A \to B$ then $g \circ \mathbf{1}_A = g$ while if $h : B \to A$ the $\mathbf{1}_A \circ h = h$. The functions $f : A \to B$ and and $g : B \to A$ are called *inverses* (and each is called the *inverse* of the other) if $g \circ f = \mathbf{1}_A$ and $f \circ g = \mathbf{1}_B$. One also writes $g = f^{-1}$ in this situation. A function $f : A \to B$ has an inverse if and only if f is bijective, in which case the inverse g is unique and is defined by the rule that $g(b) = a$ if and only if $f(a) = b$.

If $A \subseteq B$ the function $\iota_{A,B} : A \to B$ whose value on each element $a \in A$ is a is called the *inclusion map* of A into B. Its graph is the same as the graph of $\mathbf{1}_A$, but it is a different function when $B \neq A$ because its target is different: the identity map on A is surjective, while $\iota_{A,B}$ is never surjective unless $B = A$.

The set $A \times B$ is called the *Cartesian product* of A and B, and consists of all pairs[2] $(a,\, b)$ such that $a \in A$ and $b \in B$. If A is finite with m elements and B is finite with n elements, then $A \times B$ has mn elements. $A \times A$ is often written A^2. Thus, $\mathbb{R}^2$ is the set of pairs of real numbers, and these correspond to the points of the plane.

We occasionally deal with infinite Cartesian products. Let Λ be a set (in this context sometimes called an *index set* and let $\{A_\lambda\}_{\lambda \in \Lambda}$ be a family of sets indexed by Λ: this may be thought of as a function on the set Λ whose value on $\lambda \in \Lambda$ is the set A_λ. Let A denote the union of all the sets A_λ. The *Cartesian product* $\prod_{\lambda \in \Lambda} A_\lambda$ may be defined as the set of all functions f from Λ to A such that for every $\lambda \in \Lambda$, $f(\lambda) \in A_\lambda$. Then $f(\lambda)$ is referred to as the λ-coordinate of f. The function π_λ that sends an element of the Cartesian product $\prod_{\lambda \in Lambda} A_\lambda$ is called the *coordinate projection* on the λ coordinate. Note that when $\Lambda = \{1,\, 2,\, \ldots\,,\, n\}$ there is a bijection between functions f from $\{1,\, 2,\, \ldots\,,\, n\}$ to $\bigcup_{i=1}^n A_i$ such that for all i, $f(i) \in A_i$ and n-tuples $(a_1,\, \ldots\,,\, a_n)$ such that $a_i \in A_i$ for all i, $1 \leq i \leq n$, given by sending f to $(f(1),\, f(2),\, \ldots\,,\, f(n))$.

1.1 Problem. If X is a set, $A = \prod_{\lambda \in \Lambda} A_\lambda$ is the Cartesian product, and the π_λ are the coordinate projections as in the preceding paragraph, then given any function $g : X \to A$ one has a function $g_\lambda : X \to A_\lambda$ for every λ, namely $g_\lambda = \pi_\lambda \circ g$. Show that, conversely,

[2] One might ask, what exactly is the ordered pair $(a,\, b)$? From a practical point of view, one only needs that it have the property that $(a, b) = (a', b')$ if and only if $a = a'$ and $b = b'$. In formal treatments of set theory, if one wants to minimize the number of undefined terms, one can define $(a,\, b)$ to be $\{a, \{a,\, b\}\}$. (One may need axioms that guarantee that $c \notin c$ and that $c \in d \Rightarrow d \notin c$. It is then easy to prove that this definition gives the desired property.) Note that once one has ordered pairs, one can formally define ordered triples (a, b, c) by letting $(a, b, c) = ((a, b), c)$, and so forth.

given a function $g_\lambda : X \to A_\lambda$ for every $\lambda \in \Lambda$, there is a unique function $g : X \to A$ such that that $\pi_\lambda \circ g = g_\lambda$ for every $\lambda \in \Lambda$. This fact is referred to as the *universal mapping property of Cartesian products.*

Relations. Let A and B be sets. A relation ρ from A to B specifies, for each $a \in A$ and $b \in B$, whether $a \ \rho \ b$ is true (or holds). The relation is completely determined by knowing the sets A, B and the set Γ of all pairs $(a, b) \in A \times B$ such that $a \ \rho \ b$ holds: this set is called the *graph* of the relation.[3]

Examples. Here are seven examples of relations.

(a) Let $A = B$ be the set of all people. We can let ρ be "is the son of", i.e., $a \ \rho \ b$ iff person a is the son of person b.

(b) In Example (a) use "sister of" instead of "son of".

(c) Let $A = B = \mathbb{R}$, the real numbers, and let the relation be $\leq$.

(d) Let $A = B = \mathbb{Q}$, the rational numbers, and let ρ be be $\leq$.

(e) Let $A = B$ be any set and let ρ be $=$, i.e., equality.

(f) Let A, B be any two sets and let f be a function from A to B. We can define a relation ρ by the rule $a \ \rho \ b$ precisely if $b = f(a)$. This relation uniquely determines f. In fact some authors define a function as a relation f from A to B with the property that for every element $a \in A$ there is one and only one element $b \in B$ such that $a \ f \ b$. They then define $f(a)$ to be the unique b such that $a \ f \ b$. The graph of the relation f is the same as the graph of the function f, i.e., both are $\{(a, f(a)) : a \in A\} \subseteq A \times B$. Note that the graph of the identity function on A and the graph of the equality relation on A are the same.

(g) Let $A = B$ both be the set of real numbers, and define the real numbers x and y to be related if they satisfy the equation $x^2 + y^2 = 1$. (We could do the same with any other equation.) The graph of this relation is the same as the graph of the equation in the elementary sense used in secondary school mathematics or first year college calculus courses.

If ρ is a relation from A to A then ρ is called a relation *on* A. Two types of relations on A are of particular importance: partial orders and equivalence relations.

Partial orders. A relation ρ on A is called a *partial order* (or *partial ordering*) if it satisfies the following three conditions:

(1) (Reflexivity) For all $x \in A$, $x \ \rho \ x$.

(2) (Antisymmetry) For all $x, y \in A$, ($x \ \rho \ y$ and $y \ \rho \ x$) implies $x = y$.

---

[3]Some authors define a relation ρ to be an ordered triple (A, B, Γ) where A and B are sets and $\Gamma \subseteq A \times B$. Then if $a \in A$ and $b \in B$, one defines $a\rho b$ to hold precisely if $(a, b) \in \Gamma$.

(3) (Transitivity) For all $x, y, z \in A$, ($x \ \rho \ y$ and $y \ \rho \ z$) implies $x \ \rho \ z$.

If $A = P(X)$ is the set of all subsets of a fixed set X then $\subseteq$ is a partial order on A. The usual notion of $\leq$ is a partial order on the rational or real numbers. (A partial order is called a *total order* if for all $a, b \in A$, either $a \ \rho \ b$ or $b \ \rho \ a$. Note that $\leq$ is a total order.) The relation $|$ is a partial order on $\mathbb{N}$, where $a|b$ means that $b = ma$ for some choice of $m \in \mathbb{N}$ (i.e., a divides b evenly). The same rule does not give a partial order on $\mathbb{Z}$, however: antisymmetry fails, since each nonzero integer divides its negative and conversely. The relation "is an ancestor of" is a partial order on the set of all people who have ever lived if we agree to include each person among her or his ancestors, but this may not be in accord with the way the word "ancestor" is usually used.

Equivalence relations. A relation ρ on A is called *symmetric* if whenever $a \ \rho \ a'$ then $a' \ \rho \ a$. The relation on points in the plane that specifies that two points are related if the distance between them is at most 1 is symmetric, as is the relation of being siblings, while the relation "is the sister of" is not. Equality is symmetric but $<$ is definitely not.

A relation ρ is called an *equivalence relation* if it is reflexive, symmetric, and transitive. Examples of equivalence relations are equality, having the same family name (on the set of people), congruence of triangles, similarity of triangles, and having the same integer part (on real numbers). For each integer n we can define an equivalence relation on $\mathbb{Z}$ by defining $a \equiv b$ to mean that $a - b$ is divisible by n. (This relation is called *congruence modulo* n. We keep track of hours modulo 12 (modulo 24 in the military) and angles in degrees modulo 360.)

Suppose $\sim$ is some equivalence relation on A. If $x \in A$ we define the equivalence class of x (written $[x]_\sim$ or simply $[x]$) to be the set $\{y \in A : y \sim x\}$. It is easy to see that $[y]$ meets $[x]$ iff $[y] = [x]$ iff $y \sim x$. It follows that the equivalence classes form a partition of A, that is they are a family of nonempty subsets of A such that every element of A is in exactly one of them (said slightly differently: they are a family of nonempty subsets whose union is A such that the intersection of any two that are different is empty). Conversely, given a partition we get an equivalence relation on A by defining $a \sim b$ if a, b are in the same set in the partition. It is not hard to see that this gives a one-to-one correspondence between the partitions of A and the equivalence relations on A.

Problem. Let A be a set and ρ a relation on A with graph Γ. Describe as explicitly as you can the relation on A with the smallest graph containing Γ that is reflexive. Do the same with "symmetric" replacing "reflexive" and then with "transitive" replacing "reflexive." Finally, describe the relation on A with the smallest graph containing Γ that is an equivalence relation. This relation is called the *equivalence relation generated by* Γ. (Note: in each case, one can assert the existence of a smallest such graph because graphs of relations with the stated property are closed under intersection. Do not make use of this fact in your answers.)

Problem. Show that the number of equivalence relations $E(n)$ on a set with n elements satisfies $E(0) = 1$ (the empty relation is an equivalence relation on the empty set: the three requirements are vacuously satisfied because there are no elements to test them on), and $E(n) = \sum_{k=1}^{n} \binom{n-1}{k-1} E(n-k)$ for $n \geq 1$.

Example. If the equivalence relation is "having the same last name" one equivalence class would consist of all people with the last name "Smith." If the relation is similarity of triangles one class contains all equilateral triangles; another contains all 30–60–90 right triangles. If the equivalence relation is congruence modulo 2, one class consists of all even numbers and the other of all odd numbers. If the equivalence relation is having the same integer part then [3,4) is the equivalence class of 3 (or any number in the interval): there is one equivalence class, $[n, n+1)$, for all $n \in \mathbb{Z}$.

An element of an equivalence class is called a *representative* of the equivalence class. Sometimes a function on A defines (or induces) a function on equivalence classes: the idea is to apply the function to representatives of the classes. For this to make sense, the answer must be independent of the choices of representatives. When this happens the original operation is said to be *well-defined* on the equivalence classes. Likewise, a function on $A \times A$ may define a function on pairs of equivalence classes.

It turns out that addition (and multiplication) are well-defined on the integers modulo n for any n (for $n = 2$, we get even + even = even, even + odd = odd, odd + odd = odd), provided that the values are also taken modulo n. However, + on $\mathbb{R}$ is not well-defined on integer parts: adding the representatives 3.1 and 4.2 gives a different integer part, 7, from that gotten by adding the representatives 3.7 and 4.9, i.e. 8.

Choice. We shall assume that, given a set[4] $\mathcal{S}$ of non-empty sets there is a function that assigns to each set $X \in \mathcal{S}$ an element in $x \in X$.

A totally ordered set is called *well-ordered* if every non-empty subset has a least element. A subset of a partially ordered set that happens to be totally ordered in the restricted (or *inherited*) order is called a *chain*. An element of a partially ordered set is called *maximal* (respectively, *minimal*) if there is no strictly larger (respectively, smaller) element. An *upper bound* for a subset $Y \subseteq X$ is an element that is $\geq$ every element of Y: it is called a *strict* upper bound if it is $>$ every element of Y. There is similar terminology for lower bounds. The following two statements are equivalent to each other and to the weak axiom of choice:

Zorn's lemma: if X is a partially ordered set and every chain T of X has an upper bound, then X has a maximal element.

Well-ordering principle: Every set X has a well-ordering.

[4]This is a weak form of the axiom of choice. If one formulates set theory so that there are classes and sets, classes are not allowed to be elements, but there is a class of all sets, then there is a strong form that provides a function that assigns to *every* non-empty set an element of that set. Both these forms of the axiom of choice are known to be consistent with the other axioms of set theory provided that the other axioms are themselves consistent.

Assume the axiom of choice, and let us see that Zorn's lemma holds (but some substantial details are left to the reader). Let f assign to every non-empty subset of X an element of that subset. Consider all subsets Y of X which have a well-ordering that satisfies the following condition: for all $y \in Y$, if Z_y is the set of strict predecessors of y in Y, then y is the value of f on the set of strict upper bounds for Z_y in X. The empty set is one example of such a Y. Given any two such well-ordered sets, Y, Y', one can show that one is an initial segment of the other. One can then show that the union of all such sets is a well-ordered set with the same property. Thus, there is a largest set Y^* with this property. Then Y^* must have a largest element, and it must be maximal in X: if it has no largest element, we can form a strictly larger set with the same property by enlarging Y^* with the value of f on the set of strict upper bounds for Y. Likewise, if the largest element in Y^* is not maximal in X, we can also enlarge Y^* in the same way. This proves Zorn's lemma.

Now assume Zorn's lemma. Consider the set $\mathcal{P}$ of pairs consisting of a subset of X and a well-ordering on it. (The empty set has a well-ordering, so this is non-empty.) Partially order this as follows: one well-ordered subset precedes another if we have an inclusion of the underlying sets and the smaller subset inherits its order from the larger and is an initial segment of it. Every totally ordered subset of $\mathcal{P}$ has an upper bound (take the union of the subsets with the obvious order). Thus, there is a maximal element. The underlying set must be all of X, or else we can find a larger element by increasing the underlying set by one element, which we take to be largest in the new well-ordered set.

The well-ordering principle implies the axiom of choice. Given a family of non-empty sets, well-order the union. One gets the required function by having it assign to every set in the family the least element in it under the well-ordering.

2. Operations

A *binary operation* on a set A is a function from $A \times A$ to A. It follows that the number of binary operations on a finite set A with m elements is m^{m^2}. If the operation is denoted $*$ the value of $*$ on (a, a') is denoted $a * a'$. Addition and multiplication are binary operations on the real numbers, as is subtraction.

A subset $B \subseteq A$ is said to be *closed* under a binary operation $*$ on A if whenever $b, b' \in B$ then $b * b' \in B$.

The set of positive integers is closed (in, for example, the real numbers) under addition and multiplication, but not under subtraction.

A binary operation $*$ on A is called *associative* if for all elements $a, b, c \in A$, $a * (b * c) = (a * b) * c$. $*$ is said to have an identity[5] if there is an element $e \in A$ such that $e * a = a * e = a$

[5]The element e is called a *left* identity if $e * a = a$ for all $a \in A$, and one may define *right* identity similarly. If A has both a left identity and a right identity then they are equal (this does not even need associativity). Suppose we define an operation on A by the rule $a * b = b$ for all $a, b \in A$. Then every element of A is a left identity, and, if A has two or more elements, there is no right identity.

for all $a \in A$. If there is an identity, it is unique. (If e' were another, one has $e = e * e' = e'$: $e = e'$ after all.) Elements a, $b \in A$ are said to *commute* (under $*$) if $a * b = b * a$. The operation $*$ is called *commutative* if for all $a, b \in A$, $a * b = b * a$. $a \in A$ is said to have an *inverse*[6] b under $*$ if there is an element $b \in A$ such that $a * b = b * a = e$. If $*$ is associative and an inverse b exists, it is unique. (If b' were another, note that $b' = e * b' = (b * a) * b' = b * (a * b') = b * e = b$.)

In dealing with associative operations it is a frequent practice to omit the symbol for the operation. Thus, instead of $a * b$ one writes ab for the value of the binary operation. Abstract associative binary operations are often referred to as *products*. Because $(ab)c = a(bc)$ one may simply write abc for the triple product without ambiguity. Likewise, it turns out that all of the interpretations of a 4-fold product like $abcd$ (these are $a(b(cd))$, $a((bc)d)$, $(ab)(cd)$, $((ab)c)d$, and $(a(bc))d$) are the same. One can show that for longer strings, an expression like $a_1 \cdots a_n$ has a unique meaning no matter how one puts in parentheses. For definiteness, one might give a recursive definition like $a_1 \cdots a_n = (a_1 \cdots a_{n-1})a_n$. A careful proof that the result is independent of how one inserts the parentheses can be given using mathematical induction. This is not difficult.

Problem. Let $F(n)$ be the number of ways to insert parentheses in an expression of length n. Thus, $F(1) = F(2) = 1$, $F(3) = 2$, and $F(4) = 5$. Prove that $F(n) = \sum_{k=1}^{n-1} F(k)F(n-k)$. E.g., $F(5) = 1 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 5 \cdot 1 = 14$. Find a "closed form" for $F(n)$. (The usual answer is a binomial coefficient in which both the top and bottom entries involve n.)

When one has an associative operation on A one can define the powers a^n of an element a by the recursive rule $a^1 = a$ while $a^n = a(a^{n-1})$. If A has an identity e the standard convention is to define $a^0 = e$. If a has an inverse b one may define a^{-n} for an integer $n > 0$ as b^n, which also turns out to be the inverse of a^n.

Commutative binary operations are sometimes denoted $+$, and called *addition*. They may be very different from ordinary addition though. In dealing with operations denoted $+$, it is a matter of convention to call the identity element 0, if there is one. If a has an inverse under $+$, the inverse is usually denoted $-a$. When the operation is $+$, instead of writing a^n for $a * \cdots * a$ one writes na.

Note that for any associative operation $*$, one has $a^m a^n = a^{m+n}$ and $(a^m)^n = a^{mn}$ for any integers m, n, for which the powers are defined. When additive notation is used, these equations are written instead as $(ma) + (na) = (m + n)a$ and $n(ma) = (mn)a$.

---

[6] One may define b to be a *left inverse* for a if $b * a = e$, and similarly for right inverse. Then, in the presence of associativity, if a has both a left inverse and a right inverse they are equal. Show that in the set of functions from $\mathbb{N}$ to $\mathbb{N}$ under composition, the function f from $\mathbb{N} \to \mathbb{N}$ that sends every integer n to the integer part of $n/2$ (so that both $2k + 1$ and $2k$ are mapped to k for all $k \in \mathbb{N}$) has infinitely many right inverses but no left inverse.

3. Groups and Rings

A *group* is a set G with an associative binary operation $*$ such that there is an identity and every element has an inverse. Note that $(a * b)^{-1} = b^{-1}a^{-1}$ in any group. Officially, the group is the pair $(G, *)$, but one frequently speaks of "the group G." Nonetheless, one must know what $*$ is.

For abstract group operations the identity is often denoted e or 1 and the inverse of a is denoted a^{-1}. The operation is some times called the *group multiplication.*

A group is called *commutative* or *abelian* if $*$ is commutative. In this case, the operation is sometimes called addition and written $+$, and additive notations take over. The identity is then denoted 0 and the inverse of a with respect to $+$ as $-a$.

A subset H of a group $(G, *)$ is called a *subgroup* if it is closed under the operation $*$, contains the identity element of G, and for every $h \in H$, $h^{-1} \in H$. In this case, restricting $*$ gives a function $H \times H \to H$ that makes H into a group in its own right. A subgroup of a subgroup is a subgroup of the original group. An arbitrary intersection of subgroups is again a subgroup. The subgroup $\langle S \rangle$ *generated* by a set $S \subseteq G$ is the smallest subgroup of G containing S (we can intersect all such). If S is empty, then $\langle S \rangle = \{e\}$. When S consists of one element g then $\langle S \rangle$ is simply $\{g^n : n \in \mathbb{Z}\}$, the set of all powers of g. One often writes $\langle g \rangle$ instead of $\langle \{g\} \rangle$ in this case. More generally, when S is nonempty, $\langle S \rangle$ consists of all products of powers of elements of S. (Because one does not, in general, have commutativity, one has to allow terms like $g^2h^3g^{-2}h^{-3}g^4$: there is no way to simplify such an expression in an abstract group. Of course, one can simplify expressions in which two powers of g are consecutive.) A group that can be generated by one element g is called a *cyclic group.* The number of elements in the cyclic group generated by $g \in G$ is called the *order* of g in G, often denoted $|g|$. There are two cases: if the order is infinite, then all powers of g, both positive, negative, and zeroth power, are mutually distinct. If the order of g is a finite positive integer n, then $\langle g \rangle$ must be $\{e, g, g^2, \ldots, g^{n-1}\}$, and in this case the order of g may also be described as the smallest positive integer n such that $g^n = e$.

Every subgroup of $\mathbb{Z}$ (under $+$) is cyclic: if H is a subgroup other than $\{0\}$, let b be the smallest positive integer in it. Any $h \in H$ can be written $h = qb + r$ with $0 \le r \le b$, with $q \in \mathbb{Z}$. Then $r = h - qb \in H$. Since h is the smallest positive element of H, we must have that $r = 0$, and so $h \in \langle b \rangle$.

3.1 Examples.

(a) Let S be any set. The set $\operatorname{Perm} S$ of permutations of S (bijections of S with itself) is a group under composition. The identity function is the identity element in the group, and the inverse of a permutation as a function gives it inverse as a group element. The set of permutations of $\{1, 2, \ldots, n\}$ is denoted S_n and is called the *symmetric group* on

n letters (this term is sometimes used as well for the permutations of some other set with n elements).

(b) Let F denote the set of rational numbers, or the set of real numbers, or the set of complex numbers. The group $\mathrm{GL}(n, F)$ of all invertible $n \times n$ matrices with entries in F is a group under matrix multiplication. It is called the *general linear group* over F. The identity element is the size n identity matrix. The matrices with determinant one form a subgroup, the *special linear group* over F, denote $\mathrm{SL}(n, F)$. The set of $n \times n$ upper triangular unipotent matrices (i.e., upper triangular matrices all of whose entries on the main diagonal are 1), sometime denoted $\mathrm{U}(n, F)$, is a subgroup of $\mathrm{SL}(n, F)$. (Much more generally, F may be replaced by *any* field or even any associative ring with identity.)

(c) Let G, H be groups with identities e, e', respectively. Then the Cartesian product $G \times H$ forms a group under the operation $(g, h)(g', h') = (gg', hh')$. The identity is (e, e'), and the inverse of (g, h) is (g^{-1}, h^{-1}). The group G contains a "copy" G_1 of G (i.e., a subgroup *isomorphic* with G — *isomorphism* is defined in the next section), namely the set of all elements of the form (g, e'), and also contains a "copy" H_1 of H. Elements of $G \times H$ can be written uniquely as the product of an element of G_1 and an element of H_1, and these two elements commute. We can similarly form larger products of groups, including infinite products, doing operations coordinatewise. The group $(\mathbb{R}^n, +)$ with the operation vector addition is the product of n copies of $(\mathbb{R}, +)$. One may similarly define infinite products of families of groups.

A ring consists of a set R together with two binary operations, often denoted $+$, $\cdot$ and called the *ring addition* and *the ring multiplication*, such that $(R, +)$ is an abelian group, $\cdot$ is associative, and the left and right distributive laws hold: for all r, s, $t \in R$, $r(s+t) = rs+rt$ and $(s+t)r = sr+tr$. R is called a ring with identity if $\cdot$ has an identity, i.e., if there is a multiplicative identity, which is often denoted 1 or 1_R. The elements of a ring with identity that have inverses under multiplication are called *invertible elements* or *units* in the ring: they form a group under multiplication, denoted R^* or $R^\times$. There is always an identity element for $+$, almost always denoted 0 or, if it is necessary to be more precise, 0_R. R is called *commutative* if multiplication is commutative. A commutative ring with identity element 1 is called an *integral domain* if $1 \neq 0$ and whenever $ab = 0$ either $a = 0$ or $b = 0$. This implies that whenever $ab = ac$ and $a \neq 0$ then $b = c$.

A ring with identity $1 \neq 0$ is called a *division ring*, *skew field*, or *sfield* if every nonzero element has an inverse under multiplication. A commutative division ring is called a *field*: a field is always an integral domain.

$R \subseteq S$ is called a *subring* of S and S is called an *overring* or *extension ring* of R if R is a subgroup of S under addition and is closed under multiplication. R is then a ring in its own right under the restricted operations from S. If S is commutative, then so is R. If S is an integral domain and $1 \in R$, then R is an integral domain.

3.2 Examples.

(a) $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$ are integral domains. The rings $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are fields. $\mathbb{N}$ is not a subring of $\mathbb{Z}$, since it is not a subgroup under addition.

(b) If R is a ring and X is a non-empty set, the set of all functions on X taking values in R is a ring, if one defines addition and multiplication of functions by adding and multiplying values, i.e., $(f+g)(x) = f(x) + g(x)$ and $(f \cdot g)(x) = f(x) \cdot g(x)$. E.g., one may form the ring of real-valued or complex-valued functions on $\mathbb{R}$, in which case the continuous functions form a subring. If R is commutative (respectively, has an identity) then the ring of functions from X to R has the same property. The property of being an integral domain is not preserved, however, even if R is a field if X has two or more elements. (Why?)

(c) The Cartesian product of two (or finitely many, or infinitely many) rings with addition and multiplication performed coordinatewise is a ring. E.g., in $R \times S$ the operations are given by the rules $(r_1, s_1) + (r_2, s_2) = (r_1 + r_2, s_1 + s_2)$ and $(r_1, s_1)(r_2, s_2) = (r_1 r_2, s_1 s_2)$. If all the rings occurring in a product are commutative (respectively, have an identity) then the product ring is commutative (respectively, has an identity). The ring of all functions from a set X to R discussed in (b) is isomorphic with a product of copies of R, one for every element in X. [There is also a mapping property for products of rings entirely analogous to the property for sets given in Problem 1.1, and the property for groups stated in Example 3.1(c) in brackets.]

(d) Let $\mathcal{Q}$ denote a real vector space with basis $1, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$, where we think of $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ as the standard basis for $\mathbb{R}^3$. This, every element of $\mathcal{Q}$ can be written uniquely in the form $a + \boldsymbol{v}$, where $a \in \mathbb{R}$ and $\boldsymbol{v} \in \mathbb{R}^3$. We can now define a multiplication on $\mathcal{Q}$ by the rule $(a + \boldsymbol{v})(b + \boldsymbol{w}) = \big(ab - (\boldsymbol{v} \cdot \boldsymbol{w})\big) + (a\boldsymbol{w} + b\boldsymbol{v} + \boldsymbol{v} \times \boldsymbol{w})$, where $\boldsymbol{v} \cdot \boldsymbol{w}$ is the dot product (also called inner product or scalar product) of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, and $\boldsymbol{v} \times \boldsymbol{w}$ is the cross product. This multiplication is not commutative but is associative and distributes over addition on both sides. $\mathcal{Q}$ is called the *quaternion algebra*. It is easy to check that $(a+\boldsymbol{v})(a-\boldsymbol{v}) = a^2 + ||\boldsymbol{v}||^2$, from which it follows that if $a+\boldsymbol{v}$ is nonzero, it has multiplicative inverse $(a^2+||\boldsymbol{v}||^2)^-1(a-\boldsymbol{v})$. The nonzero elements of $\mathcal{Q}$ form a group under multiplication. Thus, $\mathcal{Q}$ is a division ring that is not commutative. Note that, because of the distributive law and the fact that $\mathcal{Q}$ is an $\mathbb{R}$-vector space, multiplication is determined once we know how to multiply any two of the elements 1, $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$. The vector space structure tells us how to multiply by 1 and -1: 1 is an identity, and, therefore, multiplication is determined by the rules $\boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1$, $\boldsymbol{ij} = \boldsymbol{k} = -\boldsymbol{ji}$, $\boldsymbol{jk} = \boldsymbol{i} = -\boldsymbol{kj}$, and $\boldsymbol{ki} = \boldsymbol{j} = -\boldsymbol{ik}$.

(e) Let R be a commutative ring (although this is restriction is not necessary). Let x be an indeterminate or variable. One may construct a formal power series ring consisting of elements of the form $\sum_{j=0}^{\infty} r_j x^j$ with the r_j a sequence of elements of R. The addition and multiplication are defined as follows:

$$\sum_{j=0}^{\infty} r_j x^j + \sum_{j=0}^{\infty} s_j x^j = \sum_{j=0}^{\infty} (r_j + s_j) x^j \qquad \text{and}$$

$$(\sum_{j=0}^{\infty} r_j x^j) \cdot (\sum_{j=0}^{\infty} s_j x^j) = \sum_{j=0}^{\infty} \sum_{i=0}^{j} (r_i s_{j-i}) x^j$$

This ring is denoted $R[[x]]$. The subring of power series in which all but finitely many coefficients are 0 is the polynomial ring $R[x]$ in one variable over R. Both constructions may be iterated, so that there is a power series ring and a polynomial ring in several variables over R. In both cases, the result of iterating with several variables is independent, up to isomorphism, of how one adjoins the variables. The polynomial ring $R[x_1, \ldots, x_n]$ in n variables consists of all R-linear combinations of finitely monomials $x_1^{k_1} \cdots x_n^{k_n}$. Addition is performed by adding corresponding coefficients, while multiplication is performed by "forcing" the distributive law to hold: one knows how to multiply monomials, and one collects terms afterwards.

In the case of power series one allows "formal" linear combinations of all the monomials. Addition is still performed by adding corresponding coefficients. Multiplication is also performed by multiplying all terms from the factors in all possible ways, and then "collecting" terms. A key point here is that only finitely many of the terms will involve any given monomial.

(f) Addition and multiplication on $\mathbb{Z}$ induce well-defined operations on the set of equivalence classes, denoted $\mathbb{Z}/m\mathbb{Z}$ or $\mathbb{Z}_m$, for congruence modulo m. Under these operations, $\mathbb{Z}/m\mathbb{Z}$ is a commutative ring with identity. If $m > 1$, this is ring in which $1 \neq 0$, and it is easy to see that it is an integral domain if and only if m is prime. A finite integral domain is a field (multiplication by $a \neq 0$ is injective and therefore surjective), and so the rings $\mathbb{Z}_p$ are fields when p is a prime integer.

(g) Let R be any ring. For every positive integer n, let $M_n(R)$ denote the set of $n \times n$ matrices with entries in R. This set has the structure of a ring, where the matrix operations are given by the following rules: let $A = (r_{ij})$ and $B = (s_{ij})$ Then $A + B = (c_{ij})$ where $c_{ij} = r_{ij} + s_{ij}$ for all i and j and $AB = (d_{ij})$ where $d_{ij} = \sum_{t=1}^{n} r_{it}s_{tj}$. The zero element of this ring is the matrix all of whose entries are 0. The easiest verification of the ring properties is to use the fact that matrix multiplication corresponds to composition of certain functions (linear transformations of certain free R-modules). We shall consider this viewpoint later. The ring properties can also be verified by calculations that have little entertainment value.

If $r, s \in R$ are such that $rs \neq 0$, then $\begin{pmatrix} 0 & 0 \\ r & 0 \end{pmatrix} \begin{pmatrix} 0 & s \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & rs \end{pmatrix}$, while the product in the other order is 0. This means that $M_n(R)$ is almost never commutative for $n \geq 2$.

If R has an identity, then $M_n(R)$ has an identity, the matrix (r_ij) such that $r_{ij} = 1$ if $i = j$ and $r_{ij} = 0$ otherwise. A matrix (r_{ij}) is called *upper triangular* if $r_{ij} = 0$ for $i > j$ and *lower triangular* if $r_{ij} = 0$ for $i < j$. The upper triangular matrices form a subring of $M_n(R)$, as do the lower triangular matrices. Their intersection, the diagonal matrices, is also a subring.

4. Homomorphisms

If G and H are groups a function $f : G \to H$ is called a *group homomorphism* if for all $g, g' \in G$, $f(g * g') = f(g) * f(g')$. Note that the two operations $*$ are usually different: the one on the left is the group operation in G and the one on the right is the group operation in H. If the operations are omitted from the notation, the condition that one have a homomorphism may be written simply $f(gg') = f(g)f(g')$. It is a theorem that a group homomorphism sends the identity of G to the identity in H, and that $f(g^{-1}) = f(g)^{-1}$ for all $g \in G$, i.e., that f *preserves inverses*. The image $f(G)$ of G in H is a subgroup of H (the same is true for $f(G_0)$, where G_0 is a subgroup of G). The inverse image $f^{-1}(H_0) = \{g \in G : f(g) \in H_0\}$ of any subgroup H_0 of H is a subgroup G. The inverse image of the subgroup of H consisting only of the identity element e_H of H is called the *kernel* of f. It is a subgroup of G. Note that $f(g) = f(g')$ if and only if $f(gg'^{-1} = e$, i.e., if and only if gg'^{-1} is in the kernel of f. It follows that f is injective if and only if its kernel is $\{e\}$.

A subgroup N of G is called *normal* if for all $g \in G$ and $n \in N$, $gng^{-1} \in N$. The kernel of a group homomorphism is easily verified to be normal. We shall discuss this further in the next section.

The inclusion map of a subgroup into a group is a homomorphism.

An injective group homomorphism is sometimes called a *monomorphism*, while a surjective group homomorphism is sometimes called an *epimorphism*. A bijective group homomorphism is called a *group isomorphism*. If $f : G \to H$ is a monomorphism, then one may restrict the image of f to get an isomorphism of G with $f(G)$. Sometimes this is paraphrased as follows: up to isomorphism, a group monomorphism is the same as the inclusion of a subgroup. If there is a group isomorphism of G with G', then G and G' are called *isomorphic* groups.

4.1 Examples.

(a) Any one element set $\{g\}$ can be made into a group in precisely one way: we must have $g * g = g$. All groups with just one element are canonically isomorphic to one another (there is only one way to choose the isomorphism) and all such groups are referred to as *trivial*. Every group G has a unique one element subgroup that contains only the identity element e of G. The nonzero complex numbers form a group under multiplication: $\{1\}$, under multiplication, is a subgroup with just one element. $\mathbb{C}$ is a group under addition: $\{0\}$ is a subgroup.

(b) If G is a group, we can map G into the permutations $\mathrm{Perm}\, G$ of the set G by letting $f(g)$ be the permutation that sends $g' \in G$ to gg'. This map $f : G \to \mathrm{Perm}\, G$ is a monomorphism. This gives Cayley's theorem (possibly, a misnomer): every group is isomorphic with a subgroup of a permutation group.

(c) With notation as in (3.1), the determinant function is a homomorphism of $\mathrm{GL}(n, F)$ onto $F - \{0\}$ under multiplication. This is an epimorphism. The kernel is $\mathrm{SL}(n, F)$. (F may be any field. If F is an associative ring with identity, the statement remains true if we use the set of all elements of F with multiplicative inverses, sometimes denoted F^* or $F^\times$, as the target instead of $F - \{0\}$ (when F is a field, $F - \{0\} = F^*$).)

(d) Let a be a positive real number. The map $\mathbb{R} \to \mathbb{R}^+$ (the positive real numbers) sending $x \in \mathbb{R}$ to a^x is an isomorphism of $(\mathbb{R}, +)$ with $(\mathbb{R}^+, \cdot)$. The inverse isomorphism sends the positive real number y to $\log_a y$. This is reflected by the equations $a^{x+x'} = a^x a^{x'}$ and $\log_a(yy') = \log_a(y) + \log_a(y')$.

(e) If $g \in G$, the function $C_g : G \to G$ that sends h to ghg^{-1} is a group isomorphism of G with itself, called *conjugation* by g. An element of the form ghg^{-1} is called a *conjugate* of h in G. An isomorphism of G with itself is called a *group automorphism* of G, although the word "group" is frequently omitted. If H is a subgroup of G, then the image of H under C_g is also a subgroup: the image is denoted $gHg^{-1} = \{ghg^{-1} : h \in H\}$. This subgroup is called a *conjugate* of H and is isomorphic with H. The automorphisms of G that have the form C_g for some $g \in G$ are called *inner automorphisms of* G. All other automorphisms of G are called *outer*. The group automorphisms of G form a group, denoted $\mathrm{Aut}\, G$, under composition, in which the identity function is the identity element. The map that sends g to C_g is a group homomorphism $G \to \mathrm{Aut}\, G$. The kernel of this group homomorphism is easily verified to consist of all elements $g \in G$ that commute with every element of G: this subgroup is called the *center* of G.

(f) Given a group $(G, *)$ we can define a new group G^{op} with the same underlying set as G in which multiplication $\#$ is defined by the rule $g\#h = h * g$. The identity element for this group is the same as the identity for G. The group properties are easily verified. It turns out that the map that sends $g \to g^{-1}$ is an isomorphism of G with G^{op} (and also an isomorphism of G^{op} with G).

A map $f : R \to S$ of rings is called a *ring homomorphism* if it is a function that preserves both ring operations, i.e., $f(r + r') = f(r) + f(r')$ and $f(rr') = f(r)f(r')$. In particular, a ring homomorphism is a homomorphism of abelian groups. The *kernel* of f is defined to be its kernel as homomorphism of abelian groups, the set of elements in R that map to 0 in S. The kernel of f is an *ideal* of R: a subgroup I of R such that if $i \in I$ and $r \in R$ then both ri and ir are in I. A ring homomorphism is evidently injective if and only if its kernel is $\{0\}$, since one may think of it simply as a group homomorphism of abelian groups. A bijective ring homomorphism is called a ring *isomorphism*. Its inverse is also a ring homomorphism.

4.2 Examples.

(a) The inclusion map of subring into a ring is a ring homomorphism.

(b) The map $\mathbb{Z} \to Z_m$ that sends n to its equivalence class is a ring homomorphism.

(c) Let X be a set and let R be a ring. Let S be the ring of all R-valued functions on X and let $x \in X$. The evaluation map that sends $f \in S$ to $f(x)$ is a ring homomorphism.

(d) Let $f : R \to S$ be a homomorphism of commutative rings. Let $R[x]$ be a polynomial ring in one variable over R, and let $s \in S$. Then there is a unique way of extending f to a ring homomorphism of $R[x]$ to S that maps x to s: it must send $\sum_{j=1}^{n} r_j x^j$ to $\sum_{j=1}^{n} f(r_j)s^j$, and this works. Similarly, if $R[vectxn]$ is the polynomial ring in n variables over R and $s_1, \ldots, s_n \in S$ there is a unique way of extending f to a ring homomorphism from $R[x_1, \ldots, x_n] \to S$ such that x_j maps to s_j for $1 \le j \le n$.

(e) Let X be a non-empty set and let S denote the set of all subsets of X. Define operations $+$ and $\cdot$ on S by the rules $A + B = (A - B) \cup (B - A)$ (symmetric difference) and $A \cdot B = A \cap B$. One can show directly that S is a commutative ring with multiplicative identity X: the empty set $\emptyset$ is the additive identity. This example is a bit mysterious, and the verification of the ring properties is tedious. A good way to look at it is to think about the ring R of $\mathbb{Z}_2$ valued functions on X. There is a bijection β of R to S that sends each function f to $f^{-1}(1)$, the set of points at which it takes on the value one. This bijection preserves operations. From this one can deduce that R is a ring, and that β is an isomorphism.

(f) The quaternion algebra $\mathcal{Q}$ of Example 3.2(d) is isomorphic with a subring of the ring $M_4(\mathbb{R})$ of 4×4 matrices over the real numbers via the map that sends $a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$ to $\begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}$, which is the matrix of the $\mathbb{R}$-linear map $\mathcal{Q} \to \mathcal{Q}$ given by left multiplication by $a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$ with respect to the $\mathbb{R}$-vector space basis 1, $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$.

5. Congruence, cosets, and quotient groups

Let G be a group and H a subgroup. We can define an equivalence relation on G, *left congruence* modulo H, by the rule $g \equiv g'$ if there exists $h \in H$ such that $gh = g'$. If it is necessary to specify which subgroup is being used to define the equivalence relation, we may write $\equiv_H$ for this equivalence relation. The equivalence class of $g \in G$ under this relation is often written gH, and is the same as $\{gh : h \in H\}$. It is called a *left coset* of H. If the operation in the group is addition, one writes $g + H$ for the coset instead. The set of left cosets is denoted G/H.

In an entirely similar way, one may define *right congruence* and *right cosets*. The right coset of $g \in G$ with respect to H is denoted Hg. The set of right cosets of H in G is sometimes denoted $G\backslash H$. Some authors make conventions about the use of the words "left" and "right" in this context that differ from these.

The number of elements in G/H (which may be finite even if G is infinite) is called the *index of* H *in* G, and is denoted $[G : H]$.

Note that the map that sends $g \to g^{-1}$ induces a map on sets that gives a bijection of the left cosets of H with the right cosets of H. (Cf. also the final example (f) of the preceding section.) Thus, the index $[G : H]$ is also the number of right cosets of H in G, and the number of left cosets is finite if and only if the number of right cosets is finite.

The function that sends h to gh is a bijection of H with gH. Thus, when H is finite, all left cosets of H have the same number of elements. (This is also true when H is infinite, if one has the theory of infinite cardinal numbers available: two infinite sets have the same cardinality if and only if there is a bijection between them.)

We write $|S|$ for the cardinality of the set S. Since the elements of G fall into mutually disjoint equivalence classes, each of which has the same cardinality as H, it follows that

$|G| = |H||G/H| = |H|[G : H]$ when G is finite (this is also true when G is infinite). From this we can conclude that the order of any subgroup of a finite group divides the order of the group: this is called *LaGrange's theorem*, not necessarily accurately. Of course, we have similarly that $|G| = |H||G\backslash H|$. Here are some consequences:

(1) If G is finite, the order $|g|$ of any element $g \in G$ divides $|G|$, since $|g| = |\langle g\rangle|$.

(2) If a group has prime order p, then every subgroup either consists of the identity only or is the whole group (its order must divide p and so is either 1 or p). Every element other than the identity generates the group, which is therefore cyclic.

Example. In the permutations S_3 of $\{1, 2, 3\}$ each element of order 2 generates a subgroup of order 2 and index 3. The left cosets of such a subgroup are different from the right cosets.

It is clear that if G is finite, then $|G/H| = |G\backslash H|$, since both must be $|G|/|H|$.

Problem. Let $K \subseteq H \subseteq G$ be subgroups of the group G. Let g_i by a family of left coset representatives for G (i.e., there is precisely one of the g_i in each left coset of H) and let h_j be a family of left coset representatives of K in H. Show that the elements g_ih_j form a family of coset representatives for K in G. Conclude that $[G : K] = [G : H][H : K]$.

If $f : G \to H$ is a group homomorphism and let N be the kernel, i.e., the set $\{g \in G : f(g) = e\}$. We have already noted that N is a subgroup of G. Much more is true: it is a *normal subgroup*, by which we mean that the following equivalent conditions hold:

(1) for all $g \in G$, $gNg^{-1} \subseteq N$.

(2) for all $g \in G$, $gNg^{-1} = N$, i.e., N is the only conjugate of N in G.

(3) for all $g \in G$, $gN = Ng$.

Thus, for a normal subgroup N of G, G/N and $G\backslash N$ are the same.

When N is a normal subgroup of G, one can check that there is a well-defined binary operation on G/N induced by the group operation on G: one defines $gN \cdot g'N$ to be $(gg')N$, and checks that this independent of the choices of coset representative g and g' (which is false when H is not normal). The point is that with $n,\, n' \in N$, $(gn)(g'n') = gg'(g'ng'^{-1}n')$. It is easy to check that G/H with this operation is again a group, and there is a surjective group homomorphism $G \to G/H$ that maps g to gH for all $g \in G$.

All surjective group homomorphisms are of this form, in the following sense: if $f : G \to H$ is a surjection from G onto H with kernel N, then there is a unique isomorphism of H with G/N such that for all $g \in G$, $f(g) \in H$ corresponds to gN in G/N.

Every group homomorphism $f : G \to H$ is the composition of a surjective group homomorphism $G \to f(G)$ with the inclusion $f(G) \subseteq H$. If N is the kernel of f, we have that $f(G)$ is isomorphic with G/N. Thus, every group homomorphism may be thought of as a two-step process: first pass to the quotient group G/N (briefly put, "kill" N), and then identify G/N with an isomorphic subgroup of H.

We note the following fact: when N is a subgroup of index 2 in G, N is automatically normal! This is not true for index 3 or greater. The reason is that if there are only two left cosets, they must be N and $G - N$, the complement of N in G. Then $Ng^{-1} \subseteq G - N = g^{-1}N$, and $gNg^{-1} \subseteq N$ follows. But this is trivially true as well when $g \in N$.

From this discussion, it follows that N is a normal subgroup of G if and only if it is the kernel of some group homomorphism from G to another group.

Examples.

(a) Let $m \in \mathbb{Z}$ be any integer. The subgroup of $(\mathbb{Z}, +)$ generated by m consists of all multiples of m, and is denoted $m\mathbb{Z}$. Congruence modulo $m\mathbb{Z}$ is often referred to as *congruence modulo* m as well. Since m and $-m$ generate the same subgroup of $\mathbb{Z}$, we might as well assume that $m \geq 0$. If $m = 0$, then congruence modulo m is equality. If $m = 1$ the congruence relation is rather trivial: any two integers are equivalent, and there is only one equivalence class. Quite generally, for any group G, G is normal, and G/G is a trivial group having just one element. If $m > 1$ then $m/\mathbb{Z}$ has index m in $\mathbb{Z}$, and every coset in $\mathbb{Z}/m\mathbb{Z}$ has a unique representative in the set $\{0, 1, \ldots, m-1\}$ (if one takes any element in the coset, this representative is the same as its remainder up division by m). $\mathbb{Z}/m\mathbb{Z}$ is often denoted $\mathbb{Z}_m$.

(b) If G is any cyclic group, which we write multiplicatively, and g is a generator, then there is a surjective homomorphism $\mathbb{Z} \to G$ that maps n to g^n. The point is that $g^{n+n'} = g^n g^{n'}$. If g has infinite order this map is an isomorphism, so that every infinite cyclic group is isomorphic with $\mathbb{Z}$. If g has finite order m then the kernel is $m\mathbb{Z}$, and we get an isomorphism $\mathbb{Z}/m\mathbb{Z} \cong G$. Thus, every cyclic group of finite order m is isomorphic with $\mathbb{Z}/mZ$ for some positive integer m, and every cyclic group is isomorphic with $\mathbb{Z}/m\mathbb{Z}$ for a unique non-negative integer m. In particular, every group whose order is a prime integer p is isomorphic with $\mathbb{Z}/p\mathbb{Z}$.

(c) When N is a subgroup of index 2 in G, N is automatically normal! This is not true for index 3 or greater. (Proof: if there are only two left cosets, they must be N and $G-N$, the complement of N in G. If $g \notin N$, then $Ng^{-1} \subseteq G - N = g^{-1}N$, and $gNg^{-1} \subseteq N$ follows. But this is trivially true as well when $g \in N$. $\square$)

(d) **Caution!** If N is a normal subgroup of G and H is a normal subgroup of N, it is *not* necessarily true that H is a normal subgroup of G. We give a specific example. Consider the group of symmetries of a square whose vertices we place in a correspondence with the integers $\{1, 2, 3, 4\}$ moving counterclockwise, say, with 1, 2 corresponding to the bottom edge of the square: ${}^4_1\square^3_2$. Each symmetry is determined uniquely by the manner in which it permutes the vertices, and this enables us to think of the group as a subgroup of S_4. This group D contains eight elements: four of them are the powers of the 90° counterclockwise rotation r that sends each of 1, 2, 3, 4, 1 to the next element in the sequence. The other four elements correspond to reflections in the two diagonals and in the two lines of symmetry perpendicular to the sides. One diagonal reflection, call it a, fixes 1, 3 while switching 2 and 4. One, call it b, fixes 2 and 4 while interchanging 1 and

3. Note that $ab = ba = r^2$. There is a reflection v through a vertical line of symmetry that switches both 1, 2 and 3, 4, while the fourth reflection, h, through a horizontal line of symmetry, interchanges both 1, 4 and 2, 3. Let $H = \{\mathbf{1},\, a\}$, and let $N = \{\mathbf{1}, a, b, r^2\}$ be generated by the two diagonal reflections. N is commutative, and so H is normal in N. (Alternatively, it is normal in N because it has index 2 in N.) N has index 2 in D and so is normal in D. However, H is not normal in D: $rar^{-1} = b \notin H$.

(e) The function π_i from a product group $G_1 \times \cdots \times G_i \times \cdots \times G_s$ to G_i that sends $(g_1, \ldots, g_i, \ldots, g_s)$ to g_i (referred to as the i th *coordinate projection*) is a surjective homomorphism whose kernel is $G_1 \times \cdots \times \{e_i\} \times \cdots \times G_s$, where e_i denotes the identity in G_i. [It is straightforward to show that a Cartesian product of groups, even an infinite one, has the same universal mapping property referred to in Problem 1.1 provided all functions considered are required to be group homomorphisms. If $G = \prod_{\lambda\in\Lambda} G_\lambda$ is a Cartesian product of groups and $\pi_\lambda : G \to G_\lambda$ are the coordinate projections, then for every group H and family of homomorphisms $h_\lambda : H \to G_\lambda$ there is a unique homomorphism $h : H \to G$ such that $h_\lambda = \pi_\lambda \circ h$ for every λ.]

(f) The group $\mathbb{C}^*$ of nonzero complex numbers under multiplication is isomorphic with the set of two by two real matrices of the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ with a, b not both 0, under matrix multiplication. The map sending $a + bi$ to the matrix just specified gives an isomorphism. If we think of $\mathbb{C}$ as a vector space over $\mathbb{R}$ with basis 1, i, the specified matrix is the matrix of the $\mathbb{R}$-linear map $\mathbb{C} \to \mathbb{C}$ given by multiplication by $a + bi$.

(g) Let $\mathbb{R}^+$ be the group of positive real numbers under multiplication, and $\mathbb{R}$ the group of all real numbers under addition. The map $\mathbb{R}^+ \times \mathbb{R} \to \mathbb{C}^*$ (see (f)) that sends $(r,\, \theta)$ to $r(\cos\theta + i\sin\theta)$ (or $re^{i\theta}$) is an epimorphism: the kernel is generated by $(1,\, 2\pi)$ and consists of all elements of the form $(1,\, 2\pi k)$, where $k \in \mathbb{Z}$. From this one sees there is an isomorphism of $\mathbb{R}^+ \times (\mathbb{R}/\langle 2\pi\rangle) \cong \mathbb{C}^*$, where $\cong$ is used to indicate an isomorphism.

(h) The set of elements $\{1, -1, \boldsymbol{i}, -\boldsymbol{i}, \boldsymbol{j}, -\boldsymbol{j}, \boldsymbol{k}, -\boldsymbol{k}\} \subseteq \mathcal{Q}$ [cf. Example 3.2(d)] is closed under multiplication and is a noncommutative group of order 8, called the *quaternion group*. All elements except 1 and -1 have order four.

6. Groups Acting on Sets

By an action of a group G on a set S is meant a function $G \times S \to S$ (the value of the function on (g, s) is denoted $g(s)$) such that for all $g, g' \in G$ and for all $s \in S$ we have that $(gg')(s) = g(g'(s))$ and $e(s) = s$ (where e is the identity in G. It then follows that for fixed $g \in G$ the function $p_g : S \to S$ via $p_g(s) = g(s)$ is a permutation of S whose inverse is $p_{g^{-1}}$, and the function mapping $G \to \operatorname{Perm} S$ by sending $g \to p_g$ is a group homomorphism. The action uniquely determines the group homomorphism. The action is called *faithful* if this group homomorphism is injective, which means that the only element of G fixing every element of S is the identity element.

Conversely, given a group homomorphism $\theta : G \to \operatorname{Perm} S$ one gets an action of G on

S by defining its value on $(g,\, s)$ to be $(\theta(g))(s)$. In this way one can see that there is a bijection between actions of G on S and group homomorphisms $G \to \operatorname{Perm} S$.[7]

Given an action of G on S we may define an equivalence relation on S by the rule that $s \sim s'$ if there exists $g \in G$ such that $g(s) = s'$. The equivalence classes are called the *orbits* of the action of G on S. The orbit of $s \in S$ is sometimes written Gs since it is the same as $\{gs : g \in G\}$. The *stabilizer* H of $s \in S$ under the given action of G (also called the *isotropy subgroup* of s in G) is the subgroup $\{h \in G : h(s) = s\}$. If $g, g' \in G$, then $gs = g's$ iff $s = g^{-1}(g'(s)) = (g^{-1}g')s$ iff $g^{-1}g' \in H$ (the stabilizer) of s iff $g' \in gH$. It follows that the elements of the orbit of $s \in S$ are in bijective correspondence with the left cosets of H in G. Thus, the cardinality of the orbit of s is the same as the index of the stabilizer H of s in G. When G is finite, it follows that the cardinality of any orbit divides $|G|$.

The action of G on S is called *transitive* if S itself is an orbit: in this case, if $|G|$ is finite, then $|S|$ divides $|G|$.

6.1 Examples.

(a) Every group G acts transitively on itself by left multiplication. The map $G \times G \to G$ that gives the action simply takes (g, g') to gg'. The corresponding map $G \to \operatorname{Perm} G$ is the same that is used to prove Cayley's theorem.

(b) Given an action of G on a set S there is an induced action on the set of subsets of S, and also on the set of subsets of a given cardinality. If $X \subseteq S$, $g(X) = \{g(x) : x \in X\}$ is a subset of S of the same cardinality as X. There is likewise an induced action on the set of all k-tuples of elements of S, and on the set of all k-tuples of mutually distinct elements of S. If G acts transitively on S and its action on the set of k-tuples of mutually distinct elements of S is also transitive for a certain positive integer k, then G is said to act *k-ply transitively* on S. If $k = 2$, the word "doubly" is substituted for "2-ply". If the action of the finiite group G on the finite set S with m elements is k-ply transitive then $m(m-1)\cdots(m-k+1)$ divides $|G|$ (why?).

(c) Given an action of G on S, for any subgroup H of G one gets an action of H on S by restricting the map $G \times S \to S$ to get a map $H \times S \to S$.

(d) We may restrict the action of a group G on itself described in (a) to a subgroup $H \subseteq G$. The orbits are the right cosets of H in G. A subset of X of G is stable under the induced action on subsets if and only if it is a union of right cosets.

(e) If we try to let G act on itself by right multiplication, things go slightly wrong. This is actually an action of the opposite group G^{op}, which has the same underlying set as G, but in which the product of g and h is defined to be hg. But the map $G^{\mathrm{op}} \to G$ that sends

[7] Some authors in discussing actions of groups on sets prefer to work with *right actions.* A right action of G on S is a map $S \times G \to S$ (the value on (s, g) is frequently written sg) that satisfies $s(gg') = (sg)g'$ for all $s \in S$ and $g,\, g' \in G$, and $se = s$ for all $s \in S$. This is the same as a left action of G^{op} on S. The same authors are likely to define $\operatorname{Perm} S$ as the opposite of the group that we are using, so that when working in $\operatorname{Perm} S$, fg denotes the composition $g \circ f$.

g to g^{-1} is an isomorphism (it is its own inverse). Another way of looking at this is that there is a (left) action of G on itself corresponding to right multiplication by the inverse of the element of G, i.e., using the map $G \times G \to G$ that sends (g, g') $g'g^{-1}$. If this action is restricted to H, the orbit of g is the left coset gH of H (since $\{h^{-1} : h \in H\} = H$).

(f) Consider the action in (a) again, and the induced action on subsets. If H is a subgroup of G, its orbit under this action is the set of left cosets of H. The stabilizer of H is H. If H has finite index in G, this gives a homomorphism of G into the permutations of the finite set of left cosets. The kernel N of this homomorphism is then a normal subgroup of G that is contained in H. Then G/N is isomorphic with a subgroup of the permutations of the left cosets, a finite group, and it follows that N has finite index in G. It is easy to see that if H has index 2 in G, then N must be equal to H, giving a slightly different proof that subgroups of index two are normal.

(g) Given a group G, there is an action of G on itself by conjugation: the map $G \times G \to G$ used sends (g, h) to ghg^{-1}.[8] The corresponding map $G \to \operatorname{Perm} G$ actually has an image that consists of group automorphisms of G, the so-called *inner* automorphisms discussed earlier. The orbit of an element $g \in G$ is called its *conjugacy class* and the elements in it are called the *conjugates* of g in G. As elements of G, they all have the same order that g does. The stabilizer of g under this action consists of all elements $u \in G$ such that $ugu^{-1} = g$, which is equivalent to the equation $ug = gu$. This is the subgroup of all elements that commute with g, and is called the *centralizer* of g in G. An element of G that commutes with all other elements in G is called *central*, and the set of all such elements is called the *center* of G. The centralizer of $g \in G$ is the largest subgroup H of G such that g is in the center of H. An element is in the center of G if and only if its conjugacy class contains only one element.

(h) Consider again the action in (g). There is an induced action on subsets of G. The orbit of a subgroup $H \subseteq G$ under this action consists of the conjugate subgroups of H, all of which are subgroups of G isomorphic with H. The set of these subgroups $\{gHg^{-1} : g \in G\}$ is called the *conjugacy* class of H. There is a unique element in the conjugacy class of H if and only if H is a normal subgroup of G. The stabilizer of H under this action consists of all elements $g \in G$ such that $gHg^{-1} = H$, and is called the *normalizer* of H ins G. It may also be characterized as the largest subgroup H' of G such that H is a normal subgroup of H'.

(i) If G acts on the set S and H is the stabilizer of $x \in S$, then for each $g \in G$, gHg^{-1} is the stabilizer of gx. The point is that $f(gx) = gx$ iff $g^{-1}(f(gx)) = x$ iff $(g^{-1}fg)x = x$ iff $g^{-1}fg \in H$ iff $f \in gHg^{-1}$.

6.2 More examples: permutations and cycles.

(a) The group $\operatorname{Perm} S$ acts on S in an obvious way: (g, s) is sent to $g(s)$. In particular, S_n acts on $\{1, \ldots, n\}$. Let $g \in S_n$. Then $\{1, \ldots, n\}$ breaks up into orbits under the

[8] Those who prefer right actions may use $g^{-1}hg$ instead.

action of $\langle g\rangle$. For each orbit one can write down a sequence of the form $a_1, \ldots, a_k$ where $k \geq 1$ is the number of elements in the orbit and for every i, $1 \leq i \leq k$, $a_{i+1} = g(a_i)$, while $g(a_k) = a_1$. The sequence is not quite unique (unless $k = 1$): one can start it with any a_j, and use the sequence $a_j, a_{j+1}, \ldots, a_k, a_1, \ldots, a_{j-1}$ instead. A permutation in S_n which has one orbit of length k and all other orbits of length one is called a *k-cycle*. When $k = 2$ the term *transposition* is also used. If the sequence of elements in the orbit is $a_1, \ldots, a_k$ the notation $(a_1 \ a_2 \ \ldots \ a_k)$ is used to denote the k-cycle. One needs to know what n is as well to understand what this permutation is. E.g., in S_2 the 2-cycle $p = (1 \ 2)$ is the function that maps 1 to 2 and 2 to 1, while in S_3, $(1 \ 2)$ maps 1 to 2 and 2 to 1 while sending 3 to itself. Every permutation g in S_n is uniquely the product (composition) of k-cycles with k varying, with one k-cycle for each orbit of $\langle g\rangle$ on $\{1, \ldots, n\}$.

E.g., the permutation $p \in S_8$ given by the table

$$\begin{array}{ccccccccc} n & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ p(n) & 4 & 6 & 8 & 3 & 5 & 2 & 7 & 1 \end{array},$$

where the value of the function on each element is given just below the element, can be written as $(1 \ 4 \ 3 \ 8)(2 \ 6)(5)(7)$. It is traditional to omit 1-cycles from the notation, so that this could also be written as $(1 \ 4 \ 3 \ 8)(2 \ 6)$, but then one needs to know that this permutation is an element of S_8. If two or several cycles involve mutually disjoint elements, then they commute. The same is true for any two permutations g, h such that every element not fixed by g is fixed by h (i.e., the set of elements not fixed by g is disjoint from the set of elements not fixed by h). The order of a k-cycle is easily checked to be k. The order of a product of mutually disjoint cycles is the least common multiple of the orders of the cycles. Thus $(1 \ 4 \ 3 \ 8)(2 \ 6)(5)(7)$ has order 4, while $(1 \ 4 \ 3 \ 8 \ 5)(2 \ 6 \ 7)$ has order 15. Note that the *cycle decomposition* of a permutation into disjoint cycles is unique except for (1) the order of the cycles occurring, (2) whether one writes down the 1-cycles, and (3) how one writes down a given cycle (i.e., which element in the orbit is written leftmost).

(b) Let notation by as in (a). Let $x_1, \ldots, x_n$ be variables, and let D be the product of all the terms $x_i - x_j$ with $i < j$: there are $\binom{n}{2}$ such terms. Let $p \in S$. Then one can form the product D' of all the terms $x_{p(i)} - x_{p(j)}$. This product actually involves the same terms as before, except that some of them have switched signs, and so is either D or $-D$. The sign can be calculated as $(-1)^R(p)$, where $R(p)$ denotes the number of pairs $1 \leq i < j \leq n$ such that $p(i) > p(j)$ (the number of *reversals* associated with the permutation p). If we assume some ring theory, then we may think about this set-up by letting f_p denote the automorphism of the polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ that sends x_i to $x_{p(i)}$ for every i: then $D' = f_p(D) = (-1)^{R(p)}D$. Let $\operatorname{sgn} p$ denote $(-1)^{R(p)} = f_p(D)/D$. It is easy to see that $f_p \circ f_q = f_{pq}$. Then $\operatorname{sgn}(pq)D = f_{pq}(D) = f_p(f_q(D)) = f_p\big((\operatorname{sgn} q)D\big) = (\operatorname{sgn} q)f_p(D) = (\operatorname{sgn} p)(\operatorname{sgn} q)D$, and so it follows that $\operatorname{sgn}(pq) = (\operatorname{sgn} p)(\operatorname{sgn} q)$. Thus, $\operatorname{sgn} : S_n \to \{-1, 1\}$ is a group homomorphism. The permutations whose sign is 1 (respectively, -1) are called *even* (respectively, *odd*) permutations. The even permutations form a normal subgroup, denoted A_n, in S_n. The group A_n is called *the alternating group* on n elements. The map sgn is surjective if $n \geq 2$ (e.g., $(1 \ 2)$ is odd for $n \geq 2$), and so, for $n \geq 2$, A_n has index 2 in S_n and has order $n!/2$.

One can check easily that all transpositions are odd, and that more generally, k-cycles have sign $(-1)^{k-1}$. Thus, a permutation written as product of cycles (whether disjoint or not) has the same *parity* (evenness or oddness) as the number of cycles of even length occurring. Every permutation in S_n can be written as a product of transpositions (these will not usually be disjoint, and this sort of representation is not unique). E.g., note that $(1\ 2)(2\ 3) = (1\ 2\ 3)$ and that, more generally, if $a_1, \ldots, a_k$ are distinct, $(a_1\ a_2)(a_2\ a_3)\cdots(a_{k-1}\ a_k) = (a_1\ a_2\ a_3\ \ldots\ a_k)$. Evidently, a permutation written as a product of transpositions is even or odd depending on whether the number of transpositions in the product is even or odd.

Problem. Verify that if a_1, a_2, and a_3 are distinct then $(a_1\ a_2)(a_2\ a_3) = (a_1\ a_2\ a_3)$. If, moreover, a_4 is another distinct element then $(a_1\ a_2\ a_3)(a_2\ a_3\ a_4) = (a_1\ a_2)(a_3\ a_4)$. Conclude that the 3-cycles in S_n generate A_n.

Problem. Describe those permutations in S_n (e.g., characterize the lengths of their cycles) whose centralizer is contained in A_n.

Problem. Call a permutation $\pi \in S_{2n}$ *complementing* if there is a partition of $\{1, \ldots, 2n\}$ into two n element sets such that π maps each of them onto the other. Prove that a permutation $\pi \in S_{2n}$ is complementing if and only if all of the cycles in its cycle decomposition have even length. Prove that the number of complementing permutations in S_{2n} is $\bigl(1 \cdot 3 \cdot 5 \cdots (2n-1)\bigr)^2$.

Problem: the 15 puzzle. The usual 15 puzzle consists of a square frame, four units on a side, containing 15 sliding square tiles, numbered 1 through 15, and one empty square. This gives room to slide the tiles in the row or column that contains it, and then the position of the empty square changes. We imagine the empty square as being, initially, in the lower left corner. One can iterate the sliding moves permitted by the empty square. Suppose that one does this repeatedly, rearranging the tiles, but requiring that the empty square eventually be in the lower left corner again. What permutations of the tiles are possible? One can think of the answer as a subgroup of S_{15}. Can one generalize to the case of $n^2 - 1$ tiles in an $n \times n$ frame? [Note: one may view the empty square as just another tile, with a special mark on it. What is allowed may be thought of as switching the marked tile with one of the ones next to it in its row or column — sliding more tiles may be viewed as several iterations of switching the marked tile with one just next to it.]

A game with the number 15 unrelated to group theory. Two participants play a game, alternately taking turns. One of them is chosen arbitrarily to go first. In a given turn, a player chooses a number in the set $\{1, 2, 3, \ldots, 9\}$ that has not already been taken by either player. If one of the players manages to choose 3 distinct numbers that add up to 15, that player wins. The game ends when a player wins or when all 9 numbers have been picked. See whether you can analyze this game.

Problem: an infinite sliding puzzle. Consider an array of square tiles filling an entire plane, each one unit on a side. By a vertical (respectively, horizontal) move, we mean sliding each row (respectively, column) an integer distance, which may be positive or negative. Which permutations of the tiles are possible by a finite sequence of such moves?

Is it possible to interchange two tiles that are next to each other in a column while fixing all the others? What is the answer to the final question of one only allows only vertical and horizontal moves in which finitely many columns (or in the vertical case, columns) move a nonzero distance?[9]

(c) If one knows about even and odd permutations one can use them to define the determinant of an $n \times n$ matrix (r_{ij}) with entries in any commutative ring: the determinant is $\sum_{p \in S_n} \operatorname{sgn} \sigma \prod_{i=1}^{n} r_{i,p(i)}$. If one has already defined determinants by some other method one can recover the sign of a permutation. Let $e_1, \ldots, e_n$ denote the columns of the identity matrix over $\mathbb{Z}$. Then $\operatorname{sgn} p$ is the same as the determinant of the matrix whose columns are $e_{p(1)}, \ldots, e_{p(n)}$. This follows from the fact that the determinant of the identity matrix is 1 coupled with the fact that a transposition of columns reverses the sign of the determinant of a matrix.

(d) Conjugation of cycle decompositions. Let $p \in S_n$ be a permutation for which the cycle decomposition is known and let $g \in S_n$ be another. Then gpg^{-1} has a cycle decomposition in which the cycles have the same lengths as in that for p (the converse is also true: two elements of S_n are conjugate if and only if the lengths of the cycles in their cycle decomposition are the same, counting multiplicities). What is more, it is easy to write down the cycle decomposition for gpg^{-1}: if $(a_1 \ \cdots \ a_k)$ is a cycle occurring for p, then $(g(a_1) \ \cdots \ g(a_k))$ occurs for gpg^{-1}. One gets the entire cycle decomposition for gpg^{-1} by applying this in turn to each of the cycles in the decomposition of p. The assertion that we are making is an immediate consequence of the the fact that if $p(a) = b$, then $(gpg^{-1})\big((g(a)\big) = g(b)$, together with the fact that in a cycle, each entry is the image of its predecessor, while the first element is also the image of the last. Example: in S_5, if $p = (1\ 2\ 3)(4\ \ 5)$ and $g = (3\ 4)$, then $gpg^{-1} = (1\ 2\ 4)(3\ \ 5)$

(e) A group is called *simple* if it has no non-trivial proper normal subgroup. Every finite group G that is not simple has a proper non-trivial normal subgroup N such that G/N is simple. Understanding N and G/N is often easier than understanding G, and so understanding the finite simple groups enhances understanding of arbitrary finite groups. The finite simple groups have been completely classified. There are infinite families, as well as a finitely many that do not fit into an infinite family (the sporadic simple groups). The groups A_n are simple for $n \geq 5$.

Problem. Prove this using (d) above and the problem preceding (c). (This is a problem of moderate difficulty. A very detailed proof is given in [Hall], §5.4.)

Among the non-trivial finite abelian groups, it is easy to see that only the cyclic groups

[9] More formally, we may identify each tile with the coordinates of its center, and we might as well assume that these coordinates are the elements of $\mathbb{Z} \times \mathbb{Z} = \mathbb{Z}^2$. A vertical move is determined by a function $f : \mathbb{Z} \to \mathbb{Z}$ (think of the value of f on n as determining what distance the nth columns slides). From it one gets a permutation $V_f : \mathbb{Z}^2 \to \mathbb{Z}^2$ by the rule $V_f(a, b) = (a, b + f(a))$. Similarly, each such f determines a permutation H_f by $H_f(a, b) = (a + f(b), b)$, corresponding to a horizontal move. In the first problem posed one is working with the group generated by all the V_f and H_f as f varies. In the second, f is required to be zero on all but finitely many integers.

of prime order (each of which is isomorphic to $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ for some positive prime integer p) are simple. [Every subgroup of an abelian group is normal. Therefore, a non-trivial simple group must be generated by one element, say x, and the order of x must have a prime factor p, say $|x| = ph$. Then $y = x^h$ is an element of $\langle x\rangle$ of order p. Now, $\langle y\rangle$ will be a non=trivial proper normal subgroup of $\langle x\rangle$ unless it is all of $\langle x\rangle$.] We shall see that the nonabelian simple group of lowest order has order 60, and that a group of order 60 is simple iff it is isomorphic with A_5.

7. Nilpotence of p-groups and Sylow theorems

7.1 Discussion. Note that if N is normal in G, the subgroups in G/N are in bijective correspondence with the subgroups of G that contain N: given a subgroup of G/N, it corresponds to its inverse image in G, and given a subgroup of G containing N, it corresponds to its image in G/N. This correspondence preserves normality, i.e., if H is a subgroup of G with $N \subseteq H \subseteq G$ then H is normal in G if and only of H/N is normal in G/N. Moreover, if H/N is normal in G/N then the composite surjection $G \to G/N \to (G/N)/(H/N)$ is easily seen to have kernel H, from which it follows that $G/H \equiv (G/N)/(H/N)$.

7.2 Discussion. Note as well that if N is a normal subgroup of G and H is any subgroup of G (or, more generally, if N, H are subgroups of G such that H normalizes N, i.e., $hnh^{-1} \in N$ for all $h \in H$ and $n \in N$, then the subgroup generated by N and H has two particularly simple descriptions: it is simply the set $HN = \{nh : n \in N, h \in H\}$, and it is also the set $NH = \{hn : h \in H, n \in N\}$. (This is also true if H is contained in the normalizer of N, i.e., if H *normalizes* N.) The fact that $NH = HN$ follows because $hn = (hnh^{-1})h$, while $nh = h(h^{-1}nh)$. To see that NH is a group, note that $e = ee$, while $(nh)^{-1} = h^{-1}n{-}1 \in HN = NH$. Finally, that NH is closed under multiplication follows from the fact that $(nh)(n'h') = n(hn')h'$, and hn' can be rewritten in the form $(hn'h^{-1})h = n''h$, so that the product is $(nn'')(hh')$ with $n'' = hn'h^{-1}$. We also note that $NH/N \subseteq G/N$ has the same image as H in G/N. Since the kernel of the map $H \to G/N$ is clearly $H \cap N$, we can conclude that $NH/N \cong H/(H \cap N)$.

A subgroup of G that is mapped isomorphically to itself by every automorphism of G is called a *characteristic* subgroup of G. Such a subgroup is always normal, since it the automorphisms of G include the inner automorphisms. The center of G is always characteristic, as is any subgroup generated by a set of elements clearly stable under the automorphism of G, such as the subgroup generated by all elements of order k for some fixed positive integer k.

Problems.

(1) Prove that if H is a characteristic subgroup of N, and N is a normal subgroup of G, then H is normal in G.

(2) Give an example of where H is a normal subgroup of N, and N is a characteristic subgroup of G, but H is not normal in G.

(3) If $H \subseteq N \subseteq G$ are subgroups such that H is characteristic in G and N/H is characteristic in G/H then N is a characteristic subgroup of G.

Central series and nilpotence. A group G is called *nilpotent* if it has a finite sequence of normal subgroups $\{e\} = G_0 \subseteq G_1 \subseteq \cdots \in G_s = G$ such that for every i, G_{i+1}/G_i is contained in the center of G/G_i. Such a sequence is called a *central series* for G. One way to look for such a sequence is as follows: let Z_1 be the center of G, and, if a normal subgroup Z_i has been constructed, let Z_{i+1} be the inverse image of the center of G/Z_i in G. One of three things happens: either Z_i coincides with G for large i, or $Z_j = Z_{j+1} = \cdots$ from some point but Z_j is not G, or the sequence Z_i strictly increases forever (but this case cannot occur when G is finite). G is nilpotent precisely when the sequence Z_i eventually stabilizes at G. Whether G is nilpotent or not, the sequence Z_i may be defined and is called the *upper central series* for G.

Problems.

(1) Show that if $\{e\} = G_0 \subseteq G_1 \subseteq \cdots \in G_s = G$ is any central series for G, then $G_i \in Z_i$ for all i. This shows that if G_j is eventually G, then so is Z_j. This also explains why the Z_j are referred to as the upper central series for G.

(2) Show that if G is nilpotent, then so is every subgroup of G and every homomorphic image of G.

(3) Show that a finite group G is not nilpotent if and only if it has a non-trivial homomorphic image G/N in which the center is trivial.

(4) Let $G = S_3$. The elements of order 3 generate A_3, which is therefore a characteristic subgroup of S_3. Both $A_3 \cong \mathbb{Z}_3$, and $S_3/A_3 \cong \mathbb{Z}_2$ are abelian, and therefore nilpotent. Show that S_3 has trivial center, and is therefore not nilpotent. *Thus, even when G has a characteristic subgroup H such that H and G/H are both nilpotent, G need not be nilpotent.*

(5) Consider the $n \times n$ matrices over a field F (think of the cases where K is the rational, real or complex numbers if you like). A matrix over F is called *unipotent* if it is the sum of the identity and a nilpotent matrix. The group $U = U_n(F)$ of upper triangular matrices all of whose diagonal entries are equal to 1 consists entirely of unipotent matrices. Prove the group U_n is nilpotent. (Let G_i be the subgroup of U_n consisting of all matrices such that the first $n - i$ diagonals above the main diagonal are all 0, $1 \le i \le n$.) If F is a finite field with q elements, then $U_3(F)$ evidently has q^3 elements. Determine whether, if $F = \mathbb{Z}/2\mathbb{Z}$, $U_3(F)$ is isomorphic with one of the groups of order 8 mentioned earlier (i.e., with the symmetries of a square or with the quaternion group.)

***p*-groups.** A finite group G whose order is a power of p, where p is a positive prime integer, is called a *p-group.* If G is a non-trivial p-group, then the center of G is non-trivial (contains an element other than the identity). [To see this, note that the orbits of the action of G on itself by conjugation (i.e., the conjugacy classes) all have cardinality dividing p^k. Hence, if an element is not central, the number of its conjugates is divisible by p. If the identity is the only element whose conjugacy class has one element, the number of elements in the group must be of the form $pk + 1$, a contradiction. $\Box$] It follows at once that a p-group G is nilpotent.

Problem. Show that a p-group G of order p^k has subgroups of all orders dividing p^k. [Use induction, and consider $G/\langle g\rangle$ where g is an element of order p in the center of G.]

Problem. Suppose that Z is the center of the group G and that $g \in G$. Prove that the group generated by g and Z is abelian. Conclude that if G/Z is cyclic, then G is abelian. Conclude further that every group of order p^2, where p is a positive prime integer, is abelian.

Problem. Show that in a finite nilpotent group G, the Sylow p-subgroup of G is normal for every p that divides $|G|$, and conclude that G is the direct product of its Sylow p-subgroups. This characterizes the finite nilpotent groups as finite products of finite p-groups with p varying. (Suggestions. First show that the normalizer of a Sylow p-subgroup in any finite group G is its own normalizer. Second, show that a proper subgroup H of a nilpotent group G is never its own normalizer: in fact if $Z_i \subseteq Z_{i+1}$ is part of a finite central series for G with $Z_i \subseteq H$ while Z_{i+1} is not contained in H, then H is normal in $Z_{i+1}H$ — one may check this mod Z_i, and, mod Z_i, the image of Z_{i+1} is in the center of G/Z_i. It follows that the normalizer of a Sylow p-subgroup of a nilpotent group must be G.)

7.3 The Sylow theorems. Let G be a group of order $n = p^k m$ where p is a positive prime integer, k is a positive integer, and m is a positive integer not divisible by p. Thus, p^k is the largest power of p that divides n. These hypotheses are in force in (a), (b), (c) below. The proof of each is an application of the fact that the length of an orbit is the index of its stabilizer. What is subtle about these arguments is finding the correct action to look at. Proofs are given in brackets.

(a) (First Sylow theorem.) G has a subgroup of order p^k. [The following proof is due to Wielandt. Consider the action of G by left multiplication on the set of subsets of itself having p^k elements. Let $N = \binom{p^k m}{p^k}$, the number of subsets of G with p^k elements. It is not difficult to show that this binomial coefficient is *not* divisible by p (it is a product of terms $\frac{p^k m - i}{p^k - i}$ for $0 \le i < p^k$, and, in each such term, the highest power of p dividing the numerator is the same as the highest power of p dividing the denominator). It follows that some orbit of the action has cardinality not divisible by p. Let X be a subset in that orbit, and H its stabilizer in G. Since $[G : H]$ is the cardinality of the orbit chosen, it is not divisible by p, which implies that $|H|$ is divisible by p^k. But X is a union of cosets of H (cf. (6.1d)), which shows that $|H|$ divides $|X| = p^k$. Thus $|H| = p^k$. $\Box$]

(b) (Second Sylow theorem.) The Sylow p-subgroups of G are conjugate, and every subgroup of G of order p^h is contained in a Sylow p-subgroup. [Let H by a subgroup of G of G of order p^h and let P be a Sylow p-subgroup. It will suffice to show that H is contained in a conjugate of P (if $h = k$, then H must be equal to that conjugate). Now H acts on the left cosets of P by left multiplication: since this set has cardinality m not divisible by p, and since each orbit has size dividing p^h, there must be an orbit with just one element in it, say gP. Then H stabilizes gP under left multiplication. It follows that $g^{-1}Hg$ stabilizes P under left multiplication (cf. (5.1),(9)), and since $e \in P$ we have

$g^{-1}Hg \subseteq P$ and so $H \subseteq gPg^{-1}$.

(c) (Third Sylow theorem.) The number of Sylow p-subgroups of G is congruent to 1 modulo p, and divides m. [The number of Sylow p-subgroups is the index of the normalizer N of a Sylow p-subgroup P. Since $P \subseteq N$, $[G : N]$ divides m and is not divisible by p. Now P acts on the set S of Sylow p-subgroups by conjugation. The orbit of $P \in S$ contains just one element. No other Sylow p-subgroup Q can have an orbit of length one. (If it did, its normalizer N_Q would contain both P and Q. But since Q is normal in N_Q and the Sylow p-subgroups of N_Q are all conjugate within N_Q, $P = Q$ would be forced.) The orbits other than $\{P\}$ therefore each contain a number of elements that is a power of p other than 1, and the result follows. $\square$]

Note that as a corollary to the Sylow theorems, every finite group G whose order is divisible by a positive prime integer p contains an element of order p: take any element $g \neq e$ in a Sylow p-subgroup. The order of g will be p^t for some positive integer t — but then $g^{p^{t-1}}$ has the required property.

There is a different proof, due to McKay, of this fact, that does not need the Sylow theorems. Let S be the set of all p-tuples $(g_1, \ldots, g_p)$ of elements of G whose product is e. One can choose the first $p-1$ entries in such a p-tuple arbitrarily: one is then forced to choose $g_p = (g_1 \cdots g_{p-1})^{-1}$. It follows that the number of elements in S is $|G|^{p-1}$ and so is divisible by p. The cyclic group of order p acts on S: let the generator of $\mathbb{Z}_p$ send $(g_1, \ldots, g_p)$ to $(g_p, g_1, \ldots, g_{p-1})$. Each orbit either has length 1 or length p. Since the orbit of $(e, \ldots, e)$ has only one element in it, and since the number of orbits with just one element must be divisible by p, there is another orbit with just one element in it, and that must be an element of S of the form $(g, \ldots, g)$ with all entries equal and $g \neq e$. Since the product of the entries is e, we have $g^p = e$, and so $|g| = p$. $\square$

Problems. In all of these problems, G is a finite group, while p, q, and r are distinct positive prime integers. a, k and n are positive integers.

(1) Show that if G has order pq^n, where $p < q$, then a Sylow q-subgroup of G is normal.

(2) Show that if G has order aq^n, where a is not divisible by q, and no divisor of a other than 1 is congruent to 1 mod q, then G has a normal Sylow q-subgroup.

(3) Show that a group of order $3 \cdot 2^n$ cannot be simple. (Either the 2-Sylow subgroup is normal, or one gets a transitive action by conjugation on its 3 conjugates — a nonabelian group with a non-trivial homomorphism to S_3 cannot be simple.)

(4) Suppose that the only factors of p^k that are congruent to 1 modulo q are 1 and p^k. Let G have order $p^k q$. Show that either theSylow q-subgroup of G is normal, or that the Sylow p-subgroup of G is normal. (If the Sylow q-subgroup is not normal, determine the number of elements in G of order other than q, and show that these must form a subgroup, the Sylow p-subgroup.) In particular, a group of order 12 cannot be simple.

(5) Show that if G has a proper subgroup H of index at most 4, then G is not simple. (G acts transitively on the left cosets of H.) Show likewise that if G has a conjugacy class

of either an element not the identity or of a proper subgroup that has 4 or fewer members, then G is not simple. Conclude that a group of order $4 \cdot q^k$ cannot be simple.

(6) Suppose that $p < q < r$, that pq is congruent to 1 modulo r, and that G has order pqr. Show that some Sylow subgroup is normal.

(7) Show that a nonabelian group of order < 60 cannot be simple.

(8) Let G be a simple group of order $60 = 4 \cdot 3 \cdot 5$. Show that G must have six 5-Sylow subgroups (and so 24 elements of order 5), and ten 3-Sylow subgroups (and therefore 20 elements of order 3) — why can there not be only 4? The 15 elements not of order 1, 3, or 5 must break up into conjugacy classes whose sizes divide 60, and there are no classes with 2, 3, or 4 elements (why?), so that the only possible sizes are 5, 6, 10, 12. Since 15 is odd, some conjugacy class must have 5 elements in it. This gives a non-trivial homomorphism $G \to S_5$ (why?) which must be injective (why?). Show that this forces $G \cong A_5$.

The 15 puzzle again, half-solved. Set up coordinates so that the centers of the squares in the puzzle have coordinates in $\{0, 1, 2, 3\} \times \{0, 1, 2, 3\}$, with the coordinates of the empty square (0, 0) initially. Each time one makes a simple move (switching with an adjacent tile), the sum of coordinates of the center of the empty square increase or decreases by one. Since the empty square returns to the $(0, 0)$ position eventually, the total number of moves made is even. It follows that resulting permutation of all the tiles is even, and this is not affected if one omits the empty square, which is fixed, from the set. Thus, only even permutations are possible. This is true no matter what size frame is used. We leave it to the reader to determine whether every even permutation is possible.

The 15 game revisited. Consider the following array of the numbers 1 through 9:

6	1	8
7	5	3
2	9	4

The numbers in the rows, columns, and the two diagonals all add up to 15. These are the only eight ways to write 15 as the sum of three distinct integers in the set $\{1, \ldots, 9\}$. Suppose the first player writes an X in the square of the number chosen after each turn, while the second player writes an O. Then this number game is easily seen to be "essentially" the same game as Tic-Tac-Toe. In fact, one can imagine informally that if one has a good notion of "isomorphism" of games, then this game is isomorphic with Tic-Tac-Toe. This example can be used to explain the notion of isomorphism, even though it is not an example of isomorphism for one of the standard kinds of isomorphism of abstract algebra.

8. Commutators, nilpotent groups, and solvable groups

An element of a group G of the form $xyx^{-1}y^{-1} = (xy)(yx)^{-1}$ is called a *commutator*, and is sometime denoted $(x,\, y)$ Some authors use the notation (x, y) for $x^{-1}y^{-1}xy = (yx)^{-1}(xy)$, which we would write as (x^{-1}, y^{-1}). To a large extent we avoid both notations in the sequel, but note that following facts:

Problems. Prove the following statements.

(1) If N is a normal subgroup of G, then the images of x and y in G/N commute if and only N contains the commutator of x and y.

(2) The inverse of a commutator is a commutator: in fact, $xyx^{-1}y^{-1}$ has as its inverse the element $yxy^{-1}x^{-1}$.

(3) The image of a commutator under any automorphism of G is a commutator (the same is true for its image under any group homomorphism from G to any other group).

Therefore, the commutators generate a characteristic subgroup of G, called the *commutator subgroup* of G, or the *derived* group of G. It is sometimes denoted G' or $G^{(1)}$.

If A, B, are subgroups of G, let $(A,\, B)$ denote the subgroup of G generated by all commutators of the form $(a,\, b)$ (it does not matter which way one defines this notation for $a \in A$ and $b \in B$). Note that $(A, B) = (B, A)$ (the inverses of the generators of one give the generators of the other). With this notation, the commutator subgroup of G is (G, G).

Problems.

(1) If A, B are subgroups of G, and A is normal, then $(A,\, B) \subseteq A$, while if B is normal, then $(A,\, B) \subseteq B$. Hence, if both A and B are both normal, then $(A,\, B) \subseteq A \cap B$.

(2) If A and B are normal and $A \cap B = \{e\}$, then every element of A commutes with every element of B, and there is an isomorphism of $A \times B$ with the subgroup they generate in G whose value on the ordered pair $(a,\, b)$ (this is not a commutator here) is ab.

(3) If A and B are both normal subgroups of G, then $(A,\, B)$ is a normal subgroup of G.

(4) If A, B are both characteristic subgroups of G, then $(A,\, B)$ is a characteristic subgroup of G.

(5) Define a decreasing sequence of characteristic subgroups of G by letting $\Gamma^0 = G$ and $\Gamma^{i+1} = (\Gamma^i, G)$. Then G is nilpotent if and only if Γ^j is the trivial subgroup of G for all large j, in which case, if s is the smallest integer such that $\Gamma^s = \{e\}$, $\{e\} = \Gamma^s \subseteq \Gamma^{s-1} \geq\subseteq \cdots \subseteq \Gamma^1 \subseteq G$ is a central series for G. [If G has a central series $\{e\} = G^s \subseteq \cdots \subseteq G^0 = G$, then $\Gamma^i \subseteq G^i$ for all i. For this reason, the series Γ^i is called the *lower central series* for G: the lower central series is defined whether G is nilpotent or not.]

The lower central series and upper central series for a nilpotent group G have the same length: both are as short as a central series can be.

8.1 Solvable groups. A sequence of subgroups $e = G^s \subseteq \cdots G^1 \subseteq G^0 = G$ is called a *subnormal* series if G^{i+1} is normal in G_i for all i, while it is called a *normal series* if G^i is normal in G for all i, which is a stronger condition. The groups G^i/G^{i+1} are called the *factor groups* associated with the series in either case. A group G is called *solvable* if it has a finite subnormal series such that all of the factors are abelian groups. This is equivalent to the assertion that G^{i+1} is a normal subgroup of G^i containing (G^i, G^i) for all i. Thus, if G is solvable, we may always get a subnormal series with abelian factors by taking $G^{(0)} = G$ and $G^{(i+1)} = (G^{(i)}, G^{(i)})$. From the fact that $G^i = \{e\}$ one gets that $G^{(i)} = \{e\}$. The sequence $G^{(i)}$ is called the sequence of *commutator subgroups* of G. Since the commutator subgroups are all characteristic subgroups, a solvable group always has a *normal* series in which the factors are abelian, not just a subnormal series. British authors may use the word "soluble" instead of solvable.

Problems.

(1) Show that a subgroup or a homomorphic image of a solvable group is solvable.

(2) Prove that if N is a normal subgroup of G then G is solvable if and only if both N and G/N are solvable. (The "if" part fails if the condition of being solvable is replaced by the condition of being nilpotent.)

(3) Prove that every group of order smaller than 60 is solvable.

(4) Prove that a finite group is not solvable if and only if it has a non-trivial subgroup H that is equal to its commutator subgroup.

(5) Prove that a finite group is not solvable if and only if some subgroup of G can be mapped onto a nonabelian simple group.

(6) Show that the group S_3 is solvable but not nilpotent.

(7) A finite group is solvable if and only if it has a subnormal series in which every non-trivial factor group is prime cyclic.

9. Composition series and the Jordan-Hölder theorem

Consider two subnormal series $e = G_0 \subseteq G_1 \subseteq \cdots \subseteq G_s = G$ and $e = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_t = G$.

These are called *isomorphic* if $s = t$ and there is a permutation p of the integers $1, \ldots, s$ such that for every i, $1 \leq i \leq s$, $G_i/G_{i-1} \cong H_{p(i)}/H_{p(i)-1}$. That is, the factor groups associated with the first subnormal series, G_1/G_0, G_2/G_1, etc. are the same, up to isomorphism, and keeping track of multiplicities, as the factors associated to the second subnormal series.

One finite subnormal series H_j is said to be *a refinement* of another G_i if the G_i form a subsequence of the H_j. Between each $G_{i-1} \subseteq G_i$ there may be a segment of the H's, although there also may not be any additional groups inserted. Repetitions of a subgroup are permitted.

Problem (Zassenhaus Lemma). Prove the assertions below:

Let G have subgroups $G_1 \subseteq G_2$ and $H_1 \subseteq H_2$ such that G_1 is normal in G_2 and H_1 is normal in H_2. Then $G_1(G_2 \cap H_1)$ (this is a group since G_1 is normal in G_2: cf. 7.2) is normal in $G_1(G_2 \cap H_2)$, and the quotient $G_1(G_2 \cap H_2)/G_1(G_2 \cap H_1)$ is isomorphic with $(G_2 \cap H_2)/\big((G_1 \cap H_2)(G_2 \cap H_1)\big)$. By symmetry, the latter is also isomorphic with $H_1(G_2 \cap H_2)/H_1(G_1 \cap H_2)$.

Schreier's Theorem. Any two finite subnormal series $\{e\} = G_0 \subseteq G_1 \subseteq \cdots \subseteq G_s = G$ and $\{e\} = H_0 \subseteq H_1 \subseteq \cdots \subseteq H_t = G$ for G have isomorphic refinements. This result, which is due to Schreier, may be proved as follows: refine the series $\{G_i\}_i$ by inserting between G_{i-1} and G_i the groups $G_{i-1,j} = G_{i-1}(G_i \cap H_j)$ (the product is a group: cf. the Zassenhaus Lemma just above), which increase with j: it is easy to see that $G_{i-1,0} = G_{i-1}$ and that $G_{i-1,t} = G_i$. The Zassenhaus Lemma implies that $G_{i-1,j-1}$ is normal in $G_{i,j}$ for $1 \leq j \leq t$. Similarly, one can refine the $\{H_j\}_j$ by letting $H_{j-1,i} = H_{j-1}(H_j \cap G_i)$. The Zassenhaus Lemma implies that $G_{i-1,j}/G_{i-1,j-1} \cong H_{j-1,i}/H_{j-1,i-1}$ — in fact, both are $(G_i \cap H_j)/\big((G_{i-1} \cap H_j)(G_i \cap H_{j-1})\big)$. The result follows. $\square$

Any finite group has a subnormal series in which the factors are all simple groups. (Why?) Such a subnormal series is called a *composition series*. The Jordan-Hölder theorem asserts that any two composition series are isomorphic, i.e., the two sets of simple groups occurring as factors are the same up to isomorphism, except for their ordering. [Proof. We know that any two subnormal series have isomorphic refinements. But when one refines a composition series for which G_i/G_{i-1} is simple, the new groups introduced between G_{i-1} and G_i must all be either G_i or G_{i-1}. Thus, the process of refinement introduces only new factors that are trivial, while the set of non-trivial factors remains the same. $\square$]

10. Semi-direct products

A surjective group homomorphism $f : G \to H$ is called *split* if it has a right inverse $g : H \to G$ that is a group homomorphism: thus $f \circ g = \mathbf{1}_H$. This implies that $g(H)$ is isomorphic with H, is disjoint from the kernel N, and that $g(H)$ and N generate G. Conversely, if H is a subgroup of G disjoint from a normal subgroup N and H, N together generate G, then $G = NH$ (cf. 7.2) and $G/N \equiv H$. Each element of G can be written uniquely in the form hn with $h \in H$ and $n \in N$, for if $h_1n_1 = h_2n_2$ then $h_2^{-1}h_1 = n_2n_1^{-1} \in H \cap N = \{e\}$, and this implies both $h_2 = h_1$ and $n_2 = n_1$.

In this situation the group structure of G is determined by knowing one additional piece of information, namely, how H acts on N by conjugation. The action of H on N by conjugation gives a homomorphism $\phi : H \to \operatorname{Aut} N$ (the value of $\phi(h)$ on n is hnh^{-1}). If we know ϕ, then we can recover the multiplication in G: $(nh)(n'h') = \big(n(hn'h^{-1})\big)(hh') = (n\phi(h)(n'))(hh')$.

In fact, given any two groups N, H, together with a homomorphism from $\phi : H \to \operatorname{Aut} N$ one can construct a group G with H and N as disjoint subgroups that generate G such that N is normal and the action of H on N by conjugation is given by ϕ. Moreover, the discussion of the paragraphs above shows that there is only one possible candidate for G: introduce a group structure on $N \times H$ by letting the multiplication be given by the rule $(n, h) * (n', h') = (n\phi(h)n'), hh')$. It is easy to check that (e, e) is an identity (to be more precise, one might write (e_N, e_H)), that multiplication is associative, and that (n, h) has inverse (what is it?). It is easy to check that the set of elements of the form (e, h) for $h \in H$ is a subgroup isomorphic with H, and that the set of elements of the form (n, e) for $n \in N$ is a normal subgroup, disjoint from it, isomorphic with n. The group $N \times H$ with multiplication $*$ is called the *semi-direct product* of N by H. The term *normal product* is also used.[10]

Problems.

Let H, N be groups and let $\phi : H \to \operatorname{Aut} N$ be a group homomorphism. Let G be the corresponding semi-direct product. Prove the statements below.

(1) H is normal in G if an only if ϕ is trivial, in which case G is simply $N \times H$.

(2) G is abelian if and only if H, N are both abelian and ϕ is trivial.

(3) The automorphisms of the abelian group $(\mathbb{Z}/n\mathbb{Z}, +)$ may be identified with the multiplicative group of invertible elements of the ring $\mathbb{Z}/n\mathbb{Z}$ (these are the elements that generate $\mathbb{Z}/n\mathbb{Z}$). There is a nontrivial map of $\mathbb{Z}/m\mathbb{Z}$ into $\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})$ if and only if m is not relatively prime to $|\operatorname{Aut}(\mathbb{Z}/n\mathbb{Z})| = \Phi(n)$ (this is the Euler Phi function — its value on $n = \prod_{j=1}^{s} p_j^{k_j}$, where this is the prime factorization of n with mutually distinct p_j, is $\prod_{j=1}^{s}(p_j^{k_j} - p_j^{k_j-1})$.

(4) If $p < q$ are prime then every group of order pq is a semi-direct product with $H = \mathbb{Z}/p\mathbb{Z}$ and $N = \mathbb{Z}/q\mathbb{Z}$. There is a non-trivial ϕ that can be used if and only if $q \equiv 1$ modulo p. If $q-1$ is not divisible by p then every group of order pq is abelian. In particular, evey group of order 15, 33, 35, 51, or 65 is abelian.

(5) If N is any abelian group and $H = \mathbb{Z}/2\mathbb{Z}$ one may form a semi-direct product in which the value of ϕ on the generator of $\mathbb{Z}/2Z$ is the map that takes each element of N to its inverse. The action is non-trivial unless every element of N has order 2. If N is $\mathbb{Z}/n\mathbb{Z}$, $n \geq 3$, this group is called the *dihedral group*, and often denoted D_n. It has $2n$ elements. Of these elements, n lie in the cyclic normal subgroup N. The other n elements are conjugate, and each has order two. This group is the same as the group of symmetries of a regular n-sided polygon in the plane. The group N corresponds to the group of rotations through an angle that is an integer multiple of $360°/n$. Each of the other n elements corresponds to a reflection through one of n lines of symmetry: if n is even these are the perpendicular bisectors of the opposite pairs of sides, and the diagonals joining opposite vertices. If n is

[10] Authors who favor right actions will prefer to describe such groups using the right action of H on N by conjugation. The point is that in the situation of the first paragraph, G is also HN, and one has $(h'n')(hn) = (h'h)(h^{-1}n'hn)$. The details of this formulation are left to the reader.

odd, these are the n perpendicular bisectors of the sides, each of which passes through the vertex opposite the side. Note that D_3 is isomorphic with S_3, and that D_4 is the same as the group of symmetries of a square.

(6) Let $\mathrm{O}(n, \mathbb{R})$ denote the it real orthogonal group, which is the subgroup of matrices M in $\mathrm{GL}(n, \mathbb{R})$ such that M^{-1} is the transpose of M (equivalently, M gives a distance-preserving map from $\mathbb{R}^n$ to itself). Let $\mathrm{SO}(n, \mathbb{R})$ denote the *real special orthogonal group*, consisting of all elements of $\mathrm{O}(n, \mathbb{R})$ with determinant one. Show that $\mathrm{O}(n, \mathbb{R})$ is the semi-direct product of $\mathrm{SO}(n, \mathbb{R})$ by $\mathbb{Z}_2$ (the copy of $\mathbb{Z}_2$ may be thought of as generated by reflection in a hyperplane of codimension 1 in $\mathbb{R}^n$: the diagonal matrix whose diagonal entries are $-1, 1, 1, \ldots, 1$ is an example of such a reflection). When $n = 2$, $\mathrm{SO}(2, \mathbb{R})$ may be identified with $\mathbb{R}/2\pi\mathbb{Z}$ (under $+$): describe the action of $\mathbb{Z}_2$ on this group which gives rise to the orthogonal group in dimension 2.

(7) S_n is the semi-direct product of A_n by $\mathbb{Z}_2$ for a suitable action of $\mathbb{Z}_2$ on A_n for $n \geq 2$.

11. Free Groups

Let B be a set of elements. We want to construct a group that is generated by the elements of B with as few "relations" as possible holding among them. We first consider the set $\mathcal{M}$ of all formal monomials[11] of the form $b_1^{n_1} \cdots b_k^{n_k}$ where $b_1, \ldots, b_k$ are elements of B and $n_1, \ldots, n_k$ are integers. These formal monomials are also called *words* in the symbols from B. We refer to $b_1^{n_1}, \ldots, b_k^{n_k}$ as the *terms* occurring in the word. We make the convention that we may write b instead of b^1 in such an expression. We allow the empty sequence and denote it by e (this assumes, of course, that e is not used in B — in one of the examples in the sequel, where B is the set of letters of the alphabet, we need some other symbol, e.g., "1" or "ϵ," for this element). Note that repetitions of the same element of B are allowed. We can define a binary operation on $\mathcal{M}$ as follows: if $\mu, \nu \in \mathcal{M}$, then $\mu\nu$ is their concatenation, obtained by writing down all the terms of μ and then those of ν. This operation is associative, and e is an identity element. We next define a relation $\sim$ on $\mathcal{M}$ as follows:

(1) If $\mu \in \mathcal{M}$ then $\mu \sim \mu$.

(2) If a formal monomial μ contains $b^m b^n$, and ν is the monomial obtained by replacing these two terms by b^{m+n}, then $\mu \sim \nu$ and $\nu \sim \mu$.

(3) If b^0 occurs in $\mu \in \mathcal{M}$ and ν is the element obtained by omitting it from μ, then $\mu \sim \nu$ and $\nu \sim \mu$.

(4) if $\mu_1, \ldots, \mu_s \in \mathcal{M}$ are formal monomials such that for every i, $1 \leq i < k$, μ_i and μ_{i+1} are related as in (1), (2), or (3), then $\mu_1 \sim \mu_k$.

[11] We may make all this more precise by defining $\mathcal{M}$ to be the set of all finite sequences of even length, including the empty sequence, such that the odd numbered terms are elements of B and the even numbered terms are integers.

Then $\sim$ is an equivalence relation. It is not difficult to check that the binary operation on $\mathcal{M}$ defined earlier (concatenation) induces a well-defined binary operation on the equivalence classes. The class of e is an identity, and the induced operation is associative. The class of the word $b_1^{n_1} \cdots b_k^{n_k}$ has an inverse, namely the class of the word $b_k^{-n_k} \cdots b_1^{-n_1}$. Thus, the equivalence classes form a group, called the *free group* on the generators or *basis* B. A word in $\mathcal{M}$ is called *reduced* if it does not contain any term $b^m b^n$ nor any term b^0. It should be clear that a word of minimum length in a given equivalence class must be reduced. We leave it as an exercise (see just below) to show that there is a unique reduced word in each equivalence class. It is natural to use these as class representatives.

Problem. Give a completely explicit algorithm for passing from an arbitrary word to an equivalent reduced word which decreases the number of terms occurring in the word at every step, and, if the word μ one starts with is already reduced, produces μ as the result. Then show that if two words are immediately equivalent as in (2) or (3) above in the definition of $\sim$, then they yield the same reduced word when your algorithm is applied. Conclude that if two reduced words are equivalent, then they are equal.

It is not difficult to show that if G is the free group on B then for every group H and function $f : B \to H$ there is a unique way of extending f to a homomorphism $G \to H$: its value on $b_1^{n_1} \cdots b_k^{n_k}$ will be $f(b_1)^{n_1} \cdots f(b_k)^{n_k}$. This property characterizes the free group on B up to canonical (i.e., choice-free) isomorphism and leads to the following terminology: we say that a group G is *free* on a set $B \subseteq G$ if for every group H and function $f : B \to H$ there is a unique way of extending f to G so that it is a homomorphism. This property is called the *universal mapping property* of free groups.

Problem. Let G and G' be two groups that are free on the same set B. Then there is a unique homomorphism from G to G' that is the identity on B, and a unique homomorphism from G' to G that is the identity on B. These two homomorphisms are inverses, and so there is a unique isomorphism of G with G' that fixes B. Thus, a free group on B exists, by our construction, and is unique up to canonical isomorphism.

Note: later we shall define *free abelian groups*. This is a different notion, and one must be careful not to confuse the two. The two notions agree only when B has just one element.

Generators and relations. One way of specifying a group G is to give generators b_i and a family of equations called *relations* each of which sets one word in the b_i equal to another. One refers to this group as the group generated by the b_i *subject* to the specified relations. When a group is given by generators and relations one is said to have a *presentation* of the group. The group determined by these generators and relations is a quotient of the free group G on the b_i by a normal subgroup N determined as follows. Every equation $\mu = \nu$ can be rewritten as $\mu\nu^{-1} = e$: let N be the normal subgroup generated by all of these words. The group $H = G/N$ has the following property: given any group K with elements u_i satisfying the specified equations, there is a unique homomorphism of H to K that sends every b_i to the corresponding g_i.

Example. Every group has a presentation. (One can map a free group onto H by choosing one element g_h of the basis B for every element $h \in H$: there will be a group homomorphism from the free group G on the g_h that sends every g_h to h. More efficiently but less canonically, one may use a set of generators of H instead of all elements of H. In either case, one can then take as the needed equations all of the equations that set two elements of G equal when they have the same image in H. Again, more efficiently but less canonically it suffices to set equal to e a subset S of the kernel N of the homomorphism $G \to H$ that generates N. In fact, it suffices to use a set S whose conjugates under the action of G generate N (so that N is the least normal of G containing N).

Example. Let H be the group generated by x and y subject to the relations $x^3 = e = y^7$ and $xy = y^2x$. We shall determine H. Evidently, either y becomes e in the quotient or it has order 7. Let B be the subgroup generated by y. The equation $xy = y^2x$ can be rewritten as $xyx^{-1} = y^2$. This implies that $xy^kx^{-1} = (y^2)^k = y^{2k}$, since conjugation is a homomorphism. Note that squaring is an automorphism of the cyclic group of order 7, and has order 3.

Thus, x normalizes $\langle y\rangle$, and x has order 3 or 1. It follows that H is the semi-direct product of the cyclic groups $\langle y\rangle$ by $\langle x\rangle$, and has order at most 21.

Let C_3 be a cyclic group of order 3 (this is the same as $\mathbb{Z}_3$, but it is less confusing to write it multiplicatively) with generator x and let C_7 be a multiplicative cyclic group of order 7 with generator y. Since there is a homomorphism of C_3 to $\operatorname{Aut} C_7$ that sends x to the map that squares each element of C_7, we can construct a semi-direct product of C_7 by C_3. This group has 21 elements and H maps onto it. It follows that the map is an isomorphism.

Example. Let K be the group generated by the x and y subject to the relations $x^3 = e = y^7$ and $xy = y^3x$. Again, we shall determine K. The second relation may be written $xyx^{-1} = y^3$, and cubing is an automorphism of C_7 — however, it does not have order 3. Thus, since $x(y^3)x^{-1} = (y^3)^3 = y^9 = y^2$, and $xy^2x^{-1} = (y^2)^3 = y^6$, we have that $x\big(x(xyx^{-1})x^{-1}\big)x^{-1} = x(xy^3x^{-1})x^{-1} = x(y^2)x^{-1} = y^6$, but this is also $x^3y(x^3)^{-1} = eye^{-1} = y$, and so $y^6 = y$ and $y^5 = e$. But since $y^7 = e$, this forces $y = e$. In this case, K turns out to be the same as the quotient of $\langle x\rangle$ by the subgroup generated by x^3, which is C_3.

Problems. In each case, say as much as you can about the group generated by x and y such to the specified relations.

(1) $xy = yx$, $x^m = e = y^n$, where m, n are given positive integers.

(2) $xy = y^ax$, $x^p = e = y^q$, where $p \neq q$ are positive prime integers, a is a positive integer, and $a^p \equiv 1 \bmod q$.

(3) $x^2 = y^2 = (xy)^2 = e$.

(4) $x^2 = e = y^n$, where $n > 2$ is a positive integer, and $(xy)^2 = e$.

(5) This may be a long term project. Pick a dictionary. Let G be the free group on the twenty-six letters of the Roman alphabet, i.e., a, b, …, z. Two words of the dictionary are said to be *anagrams* of one another if they have the same letters, counting multiplicity.

In this problem we allow only words of the dictionary that are not listed as foreign, that are not usually capitalized, that are not slang, and that do not have any hyphens or apostrophes in them. Suppose that we set two words equal whenever they are anagrams of one another, and form the group A that is subject to these relations. E.g., one has such relations as $eta = eat = ate = tea$, $backswing = wingbacks$, $pyrogenic = recopying$, $intoxicate = excitation$ and $reabsorption = probationers$ (assuming that the dictionary chosen contains these). Note that "prop" and "poor" are not anagrams: they have the same letters, but not if one keeps track of multiplicities. Show that in A, any two letters other than j, q, x, z commute. (Any collegiate dictionary should have enough words for this to be true.) How many letters can you find that commute with x?

12. Direct sums and free abelian groups

If G and H are abelian groups, particularly if their group operations are written additively, then their product $G \times H$ is often written $G \oplus H$ instead. The element (g, h) is written $g \oplus h$ when this sort of notation is used. The group $G \oplus H$ is called the *direct sum* of G and H. Sometimes the term *external* direct sum is used, to distinguish this notion from *internal* direct sums discussed in the next paragraph.

Suppose that $(A, +)$ is an additive abelian group with subgroups G, H such that (1) $G \cap H = \{0\}$ and (2) $A = G + H$, the subgroup generated by G and H. Then it is easy to see that there is an isomorphism of $G \oplus H$ with A that sends $g \oplus h$ to $g + h \in A$. A is called the *internal* direct sum of its subgroups G and H.

These notions extend to finite families of additive abelian groups and to infinite families. In the case of finitely many groups $(G_i, +)$, $1 \le i \le n$, the direct sum $G_1 \oplus \cdots \oplus G_n$ is the same as $G_1 \times \cdots \times G_n$, but the element $(g_1, \ldots, g_n)$ is denoted $g_1 \oplus \cdots \oplus g_n$. Moreover, $G_1 \oplus \cdots G_n \cong (G_1 \oplus \cdots \oplus G_{n-1}) \oplus G_n$, so that a recursive approach is possible.

An additive abelian group A is the *internal* direct sum of subgroups $G_1, \ldots, G_n$ if each meets the sum of its predecessors in $\{0\}$ and together the G_i generate A: this means that $G_1 + \cdots + G_n = A$. A stronger condition on the G_i then holds: each meets the sum of the others in $\{0\}$.

In the case of infinite families, the direct sum of the G_λ, where λ runs through the index set Λ, is *not* the same as the product group. It may be identified with the subgroup of the product group consisting of all elements such that only finitely many coordinates have nonzero entries. For every finite set of mutually distinct elements $\lambda_1, \ldots, \lambda_s$ of Λ, the infinite direct sum, which may be denoted $\bigoplus_{\lambda \in \Lambda} G_\lambda$, contains a subgroup isomorphic with $G_{\lambda_1} \oplus \cdots \oplus G_{\lambda_s}$. It is, moreover, the union of all of these. There is a monomorphism of every G_λ into the direct sum G: if we think of G as the product (or a subgroup of the product when Λ is infinite) this map sends $g \in G_\lambda$ to the element of the product which has g as its λ-coordinate while every other coordinate is 0.

There is a universal mapping property for direct sums of abelian groups (including infinite direct sums). Note that with $G = \bigoplus_{\lambda \in \Lambda} G_\lambda$, we have a monomorphism $\iota_\lambda : G_\lambda \to$

G for every $\lambda \in \Lambda$. Given a homomorphism $h : G \to H$, where H is a abelian, we get a homomorphism $h_\lambda : G_\lambda \to H$ for every λ: in fact, h_λ is simply $h \circ \iota_\lambda$. Conversely, given a family of maps $h_\lambda : G_\lambda \to H$, where H is an abelian group, for all $\lambda \in \Lambda$ there is a unique homomorphism $G \to H$ such that $h_\lambda = h \circ \iota_\lambda$ for all λ. The idea is simply to let the value of h on $g_{\lambda_1} \oplus \cdots \oplus g_{\lambda_s}$, where the λ_j are mutually distinct and $g_{\lambda_j} \in G_{\lambda_j}$ for every j, be $h_{\lambda_1}(g_{\lambda_1}) + \cdots + h_{\lambda_s}(g_{\lambda_s})$.

Appendix: Quotations from Jack McLaughlin's Students

The following quotations are taken from letters from Jack McLaughlin's students over the years, both graduate and undergraduate. The letters are part of the file that was put together when Jack won the Amoco Teaching Award.

"... when I pooh-poohed the work of Sylvester ... McLaughlin delivered a 5-minute lecture on the personal hardships Sylvester overcame in order to do mathematics. My respect for mathematicians dates from that moment. Now I am a mathematician myself, and I wish I could be as effective as McLaughlin was at that moment."

"Jack McLaughlin was the best teacher I had in my undergraduate and graduate student career. ... I never thought of mathematics as 'beautiful' until I saw Jack's lectures on linear algebra ... In my mathematical research, teaching, and administrative career, which has been spent at four major universities, I have never encountered a better teacher of undergraduate mathematics than Jack McLaughlin."

"Despite my incredibly sloppy handwriting and note-taking, my notebooks from his courses present the material with textbook clarity. Apparently I improved my handwriting just for his courses. ... After 21 years, few of my professors at Michigan are as memorable as Prof. McLaughlin. ... There should be no doubt in anybody's mind that he is remembered and appreciated by his students indefinitely."

"'Fast Jack' is a superb teacher, his only competition is Dick Feynman at Cal Tech ... He motivates mathematics so well that even sticky problems become clear, and the clarity remains."

"... my contact with Professor McLaughlin is one of the primary reasons I chose a career in mathematics ..."

"... he made the greatest impression of all my teachers. ... The challenge came in the form of the problems. I loved them. I have kept them ... and assigned them to my own students. ... I remember with delight one of my undergraduates who initially looked undistinguished, but blossomed under the effect of the McLaughlin problems. (He is now a research mathematician.)"

"I was struck at the time and am even more impressed in retrospect at the *creativity* he elicited from his students. ... I am sure that my 'discoveries' in the solution of the presented problems were as exciting to me as they were to the original discoverers."

"Professor McLaughlin at the board is like Bach at the organ. ... Fifteen years later I still have my class notes ... Professor McLaughlin had a profound effect on my career.

... After a while I didn't care about the grade. The whole point was to learn — failing to solve a problem was part of the process. Who could assign graduate level problems to sophomores and have them look forward to it?"

"Jack was unquestionably the most inspiring of a number of excellent teachers I had at Michigan ... the algebra course he taught in my sophomore year was the single most important factor in my decision to become a mathematician ... I have never encountered anyone who had taken a class from Jack who doesn't remember him as one of the very best teachers"

"During my years at Michigan he was the advisor to and mainstay of the honors program in Mathematics. Certainly, he had the strongest influence on my own undergraduate career. It is surely true that McLaughlin was ... an extraordinary teacher. His lectures were, it seemed, perfectly conceived and flawlessly presented. ... the demands he made on the honors students were extraordinary. In particular, the problem assignments were enormously challenging. ... the problems took up most of one's life, but it was a life well spent since the assignments, like the lectures, were masterpieces."

"I still have the notes from the course ... he played a major role in my decision to switch from engineering to mathematics ..."

"He was probably the best teacher I ever had; his courses were the most demanding I ever had; and under his tutelage I learned what it was to be a mathematician."

"... his technique was very motivating ... I have kept all my notebooks and homework problems from Dr. McLaughlin's courses. His were the only courses whose materials I kept ..."

"He was the most influential teacher I had while a student at the University of Michigan ... There were no women on the faculty at Michigan in mathematics, so I was afraid no one would hire a woman mathematician ... When I discussed this with McLaughlin, all he said was 'You should go into mathematics. There are women at other universities.' That was all I needed. ... when I was applying to graduate school the University of Chicago lost my file and all my letters of recommendation, so didn't want to admit me. The situation was salvaged by Professor McLaughlin, who called up Kaplansky, who was then chairman, and spoke to him and John Thompson about me. I was admitted because they trusted McLaughlin's judgment."

"... the most extraordinary thing about his teaching, undergraduate or graduate ... was his ability to deliver lectures of astonishing clarity and insight day after day after day. ... his lectures were seamless narratives, raising questions, examining them from every angle, considering examples, and then resolving the questions. ... There can't be many like Jack, whose lecture notes circulate in photo-copy among mathematicians like a kind of samizdat."